



universität  
wien

# Diplomarbeit

Titel der Arbeit

Der Raumvorstellungstest EST im Kulturvergleich:  
Italien-Österreich

Verfasserin

Julia Brugger

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juni 2009

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ. Prof. Dr. Georg Gittler



# Danksagung

Diese Diplomarbeit möchte ich meinen Eltern widmen, ohne deren großzügige und liebevolle Unterstützung so Vieles in meinem Leben nie möglich gewesen wäre.

Dank möchte ich auch Univ. Prof. Dr. Georg Gittler und Mag.a Eva-Maria Adlmann aussprechen, die mit ihrer aufmerksamen und tatkräftigen Hilfe das Voranschreiten dieser Arbeit in ihren schwierigen Momenten in die richtigen Bahnen gelenkt haben und immer ein offenes, verständnisvolles Ohr für mich hatten.

Auch meinem Freund, meinem Bruder und meinem gesamten Freundeskreis sei an dieser Stelle dafür gedankt, dass sie meine Stimmungsschwankungen und Verzweiflungsmomente so liebevoll hingenommen und mich immer wieder an die schönen Dinge des Lebens erinnert haben.

Ein herzlicher Dank sei selbstverständlich auch all meinen Probanden ausgesprochen, ohne die ich diese Zeilen nie schreiben hätte können.

Danke an alle, die an der Erlangung meines Magistertitels beteiligt waren.



# Inhaltsverzeichnis

<b>1. Abstract</b> .....	<b>4</b>
<b>2. Einleitung</b> .....	<b>5</b>
<b>Theoretischer Teil</b> .....	<b>7</b>
<b>3. Kulturvergleichende Psychologie</b> .....	<b>8</b>
3.1 Entstehung.....	8
3.2 Definitionen von „Kultur“ .....	10
3.3 Ziele .....	11
3.4 Schwierigkeiten der Kulturvergleichenden Psychologie.....	13
3.5 Möglichkeiten der Übersetzung eines Testinstruments .....	15
3.6 Stichprobe .....	17
3.7 Psychometrische Qualitäten.....	18
3.7.1 Vorausgehende Analysen.....	18
3.7.2 Äquivalenz .....	19
3.7.3 Bias .....	20
3.7.3.1 Möglichkeiten Bias zu verhindern .....	22
3.7.3.1 Möglichkeiten Bias aufzudecken.....	24
<b>4. Die Item Response Theory (IRT)</b> .....	<b>25</b>
4.1 Das dichotome logistische Modell von Rasch .....	26
4.1.1 Voraussetzungen für die Geltung des Rasch-Modells .....	28
4.1.2 Prüfung des Rasch-Modells .....	29
4.2 Exkurs: Besondere Vorteile der IRT im Kulturvergleich .....	30
<b>5. Raumvorstellung</b> .....	<b>31</b>
5.1 Was ist Raumvorstellung .....	31
5.2 Raumvorstellung und der Unterschied der Geschlechter.....	36
5.3 Einflussgrößen auf das räumliche Vorstellungsvermögen .....	38
5.3.1 Biologische Erklärungsansätze .....	39
5.3.1.1 Vererbbarkeit .....	39

5.3.1.2 Hormone .....	40
5.3.1.3 Unterschiede in der Gehirnstruktur.....	41
5.3.1.3.1 Lateralisierung .....	41
5.3.2 Exkurs: Händigkeit .....	43
5.3.3 Die Bedeutung der Umwelt in kulturvergleichenden Untersuchungen .....	44
5.3.3.1 Exkurs: Darstellende Geometrie .....	47
5.3.4 Biopsychosoziale Theorien.....	48
5.3.5 Einfluss von Alter auf räumliche Fähigkeiten .....	50
<b>Empirischer Teil.....</b>	<b>52</b>
<b>6. Stichprobe.....</b>	<b>53</b>
6.1 Gesamtstichprobe.....	53
6.2 Italienische Stichprobe.....	54
6.3 Österreichische Stichprobe .....	56
<b>7. Erhebungsinstrumente .....</b>	<b>58</b>
7.1 Der Endlosschleifentest (EST).....	58
7.2 Das persönliche Datenblatt .....	61
7.3 Übersetzung der Erhebungsinstrumente .....	61
7.4 Testvorgabe.....	62
7.4.1 Aufgabenstellung .....	63
<b>8. Datenerhebung.....</b>	<b>65</b>
<b>9. Fragestellungen und Hypothesen .....</b>	<b>66</b>
<b>10. Rasch-Modell Analysen des EST.....</b>	<b>68</b>
10.1 Rasch-Modell Analysen in der italienischen Stichprobe .....	69
10.2 Rasch-Modell Analysen in der österreichischen Stichprobe .....	72
10.3 Rasch-Modell Analysen über beide Kulturen.....	74
10.4 Item- und Personenparameterschätzung im Rasch-Modell .....	75

<b>11. Hypothesenprüfung der Kulturvergleichenden Untersuchung.....</b>	<b>77</b>
11.1 Leistungsunterschiede im EST zwischen den Kulturen.....	77
11.2. Geschlechtsspezifische Leistungsunterschiede.....	79
11.3 Altersspezifische Leistungsunterschiede .....	81
11.4 Bildungsspezifische Leistungsunterschiede.....	85
11.5 Leistungsunterschiede auf Grund vorausgehender Erfahrungen mit darstellender Geometrie.....	87
11.6 Der Einfluss der Handpräferenz auf die Leistung im EST .....	89
11.7 Bearbeitungszeitspezifische Leistungsunterschiede .....	91
<b>12. Interpretation und Diskussion.....</b>	<b>94</b>
<b>13. Zusammenfassung.....</b>	<b>99</b>
<b>Literaturverzeichnis .....</b>	<b>102</b>
<b>Tabellenverzeichnis.....</b>	<b>112</b>
<b>Abbildungsverzeichnis.....</b>	<b>116</b>
<b>Anhang.....</b>	<b>117</b>
<b>Curriculum Vitae .....</b>	<b>123</b>

# 1. Abstract

This thesis dealt with a cross-cultural comparison of spatial ability between an Italian and an Austrian sample. The psychometric instrument used in this study was the “Endlosschleifentest” (EST; engl.: endless-loops-test) whose application already showed culture-fair results in several different cross-cultural studies (Austria, Slovakia, Mexico and Australia).

In every cross-cultural comparison, attention has to be paid to equivalent meaning of the test scores obtained by the measurement instrument within and between the compared cultures. To verify this assumption, the German version of the EST was translated into Italian language by using the “translation-backtranslation” procedure and administered to a sample of 152 Italian people and 150 Austrians (age range between 13 and 74; 152 males and 150 females). Rasch-homogeneity of the Italian and Austrian EST-version ( $k=21$ ) was examined by using several Likelihood Ratio Tests (Andersen, 1973). After the theory-driven exclusion of two items at the beginning and one item at the end of the test ( $k=18$ ), Rasch-homogeneity has been confirmed within and between both cultures. Therefore the EST once again proved its property as a culture-fair instrument to measure spatial ability and the most important principal of cross-cultural research, “scalar equivalence” has been reached, allowing further investigation.

Moreover this study revealed a non-significant difference of spatial ability between cultures ( $T_{(297)}=.691$ ,  $p=.490$ ); the better ability of mental rotation in men (Mean=.15, SD=1.25,  $T_{(297)}=4.274$ ,  $p<.001$ ); the finding that older participants (age>40) yielded lower test-scores than younger ones ( $p=.002$ ) and the significant positive influence of descriptive geometry on spatial ability ( $p=.001$ ).

Key words: spatial ability, mental rotation, cross-cultural comparison, Rasch-homogeneity, culture-fair.



## 2. Einleitung

Diese Arbeit dient einem interkulturellen Vergleich der Raumvorstellungsfähigkeit zwischen den Ländern Österreich und Italien mittels des *Endlosschleifentests* (EST) von Gittler und Arendasy (2003).

Die „Kulturvergleichende Psychologie“ ist ein bedeutungsvolles Teilgebiet der Psychologie, das sich nicht eindeutig einer ihrer Disziplinen zuordnen lässt, sondern vielmehr ein wichtiger Bestandteil aller ihrer Gebiete ist (Thomas, 2003).

Mit der Suche nach Allgemeingültigkeit und Generalisierbarkeit psychologischen Wissens bietet sie eine Möglichkeit zur verbesserten interkulturellen Kommunikation und die Eröffnung eines weiteren Schrittes in Richtung einer einheitlichen Psychologie.

Ebenso ist die Raumvorstellung, als einer der Primärfaktoren der Intelligenz, seit langer Zeit ein bedeutendes Forschungsgebiet in der Psychologie und der interkulturelle Vergleich ermöglicht es, Ergebnisse aus einer Kultur in eine andere Kultur zu übertragen oder wie in diesem Fall ein Testinstrument, das in einer Kultur entwickelt wurde, auch in anderen Kulturen auf Validität zu untersuchen.

Die Gewährleistung sinnvoller Vergleiche ist jedoch nur dann gegeben, wenn die Items in den untersuchten Kulturen äquivalente Bedeutung vorweisen. Dieser Frage wird hier anhand der drei Hauptfragestellungen Rechnung getragen, indem der EST mittels Rasch-Modell analysiert wird.

Die Hauptfragestellungen dieser Arbeit behandeln die Thematik, ob der EST, der in österreichischen (z. B. Gittler & Arendasy, 2003), slowakischen (Kriser-Fischer, 2006), mexikanischen (Ruiz-Zapata, 2005) und australischen (Waldenmair, 2003) Stichproben die psychometrische Eigenschaft der Rasch-Homogenität aufweist, diese erneut in einer österreichischen sowie in einer italienischen Population zeigt und ob dies auch über beide Kulturen gemeinsam gewährleistet ist.

Neben den Hauptfragestellungen wird in dieser Studie auch den Fragen nach dem Einfluss von Ausbildung, Alter, Geschlecht sowie Handpräferenz und Erfahrungen mit „darstellender

Geometrie“ auf die Raumvorstellungsfähigkeiten nachgegangen. Dazu wurde den Untersuchungsteilnehmern neben dem EST ein Datenblatt zur Erhebung dieser Variablen vorgegeben.

Weiters bietet diese Arbeit dem Leser einen Überblick über „Kulturvergleichende Psychologie“ im Allgemeinen, ihre Ziele, Vorteile und Schwierigkeiten.

Ebenso wird versucht zu erklären, was heute unter dem Begriff „Raumvorstellung“ verstanden wird, wie sich der Begriff mit der Zeit gewandelt hat, welche Faktoren die Raumvorstellung beeinflussen können und warum nicht alle Personen die gleichen Raumvorstellungsleistungen erbringen. Dem Leser wird weiters ein Einblick in das Rasch-Modell, seine theoretischen Annahmen, sowie dessen Vor- und Nachteile, geboten.

Der EST wurde unter Berücksichtigung zahlreicher möglicher Fehlerquellen, mittels einer einfachen „translation-backtranslation“ Methode (Werner & Campbell, 1970) von der Autorin ins Italienische übersetzt und von einer Italienisch-Dolmetscherin rückübersetzt. Im nächsten Schritt wurden die Ergebnisse verglichen, diskutiert und gegebenenfalls abgeändert. Die endgültige Version wurde 152 Italienern und 150 Österreichern vorgegeben, von welchen drei Personen wegen unseriöser Testbearbeitung ausgeschlossen wurden.

Im empirischen Teil dieser Arbeit erhält der Leser einen Einblick in die wissenschaftliche Vorgehensweise der Untersuchung; neben der Beschreibung der Erhebungsinstrumente und der detaillierten Zusammensetzung der Stichprobe, sowie einer klaren Formulierung der Forschungshypothesen, werden die Ergebnisse im Anschluss dargestellt und interpretiert.

In der gesamten Arbeit wird stets die Schreibweise der männlichen Form verwendet; diese ist wertschätzend für beide Geschlechter zu verstehen und dient der Optimierung des Ductus.

# **Theoretischer Teil**

### **3. Kulturvergleichende Psychologie**

“Cross-cultural psychology is the study of similarities and differences in individual psychological functioning in various cultural and ethnic groups; of the relationships between psychological variables and sociocultural, ecological, and biological variables; and of current changes in these variables.” (Berry, Poortinga, Segall & Dasen, 1992, p.2)

Die „Kulturvergleichende Psychologie“ (*cross-cultural psychology*) stellt eine relativ junge Tradition innerhalb der Psychologie dar. Ihre Anfänge werden von vielen Wissenschaftlern in den 1960er Jahren gesehen (Thomas, 1993, a).

Eines der größten Ziele der „Kulturvergleichenden Psychologie“ ist es seit jeher, eine Vergleichbarkeit psychischer Prozesse auf internationaler Ebene zu gewährleisten. Dazu untersucht sie neben den Gemeinsamkeiten auch die Unterschiede zwischen verschiedenen Kulturen. Der Großteil kulturvergleichender Untersuchungen hat seinen Ursprung in der westlichen Welt und Studien hierzu zielen oftmals darauf ab, in diesem Kontext gefundene Gesetzmäßigkeiten auch in nicht-westlichen Kulturen zu untersuchen. Hierbei sind vielfältige Probleme zu beachten, damit ein Vergleich fair ist und das im Rahmen der Untersuchung gegebene Umfeld nicht eine der zu vergleichenden Kulturen bevorzugt.

Nachfolgend wird eine kurze Einführung in die Entwicklung der „Kulturvergleichenden Psychologie“, ihre Definitionen, Ziele und Schwierigkeiten gegeben.

#### **3.1 Entstehung**

Jahoda und Krewer (1997) geben einen sehr detaillierten Überblick über die Entwicklung des Interesses für andere Kulturen, von der Antike und dem Mittelalter, über Renaissance und das Zeitalter der Aufklärung bis hin zu den 1990er Jahren. Sie sehen den Beginn des Interesses an anderen Kulturen bereits in der Antike bei den Griechen und Römern, denen es zu militärischen und administrativen Zwecken zum Vorteil wurde, Informationen über andere Länder und Kulturen zu besitzen.

Während in der Antike das Wort „anders“ jedoch gleichbedeutend mit „niedriger“ war, entwickelte sich in der Renaissance mit ihren Eroberungen, Entdeckungen, Missionierungen und wachsenden Handelsbeziehungen ein breiteres Interesse für die Art der Verschiedenheit anderer Völker (Jahoda & Krewer, 1997).

Erst mit der Ära der Aufklärung begann die Entwicklung in Richtung einer echten wissenschaftlichen Herangehensweise an die Thematik. Es wurde angenommen, dass allen Menschen eine bestimmte mentale Fähigkeit eigen ist und psychische Funktionsweisen mehr oder weniger auf einem natürlichen und universellen Ursprung basieren. Als eine weitere Stufe in der Entwicklung zur „Kulturvergleichenden Psychologie“ sehen Jahoda und Krewer (1997) die Bewegung der Aufklärungsgegner, die davon ausgingen, dass nichts über die menschliche Entwicklung ausgesagt werden kann, solange man nicht auch deren soziokulturellen Hintergrund beachtet. Vertreter dieser Gegenbewegung waren unter anderem Gottfried Herder, Lazarus und Steinthal sowie Wundt, der allgemein als Vater der Völkerpsychologie bekannt wurde.

Die ersten systematischen, empirischen, Versuche menschliche Unterschiede zu untersuchen, traten jedoch erst am Ende des 18. Jahrhunderts durch eine interdisziplinäre Gruppe von Wissenschaftlern, der „Société des Observateurs de l’Homme“, die von Jahoda und Krewer (1997) als „first cross-cultural scientific association“ (p.15) bezeichnet wurde, auf.

1860 eröffneten Lazarus und Steinthal durch die Gründung der „Zeitschrift für Völkerpsychologie und Sprachwissenschaft“ eine weitere Möglichkeit zur Verbreitung der „Kulturvergleichenden Psychologie“. Sie schufen dadurch ein Portal, auf welchem Vertreter verschiedener Fachgebiete die Möglichkeit erhielten, eng zusammenzuarbeiten (Thomas, 1993, b). In diesem Zusammenhang beschäftigte sich auch Wundt (1913, nach Berry et al., 1992) in seinem Werk „Völkerpsychologie“, rege mit kulturellen Einflüssen auf den Menschen.

Die beiden Hauptziele der Völkerpsychologie waren zu jener Zeit, Jahoda und Krewer (1997) zu Folge, denen der heutigen „Kulturvergleichenden Psychologie“ ähnlich: die Suche nach generellen, für alle geltenden Gesetzmäßigkeiten und das Erkennen von Unterschieden zwischen diesen (Jahoda & Krewer, 1997).

In den 1960ern stieg das Interesse an der „Kulturvergleichenden Psychologie“ rasant; dies äußerte sich in einer schnell steigenden Anzahl von Zeitschriften, die Artikel zu diesen Themen veröffentlichten; darunter befanden sich das „Journal of Social Psychology“, das „International Journal of Psychology“ und das „Journal of Cross-Cultural Psychology“. 1972 entstand die „International Association for Cross-Cultural Psychology“ (Jahoda & Krewer, 1997).

Auch Thomas (1993, a) ist der Ansicht, dass sich die „Kulturvergleichende Psychologie“ erst in den 50er und 60er Jahren zu einer eigenständigen Teildisziplin der Psychologie entwickelte. Als wesentlich für die Lösung des Problems, dass es bis dato noch keine einheitliche Definition Kulturvergleichender Psychologie gibt, erachtet Thomas (1993, c), dass jeder Forscher die von ihm verwendete Definition von Kultur seiner Arbeit voranstellen sollte, um damit zu zeigen, welche dominanten Aspekte diese enthält.

### **3.2 Definitionen von „Kultur“**

Lonner (1994) zu Folge, gibt es in den Sozialwissenschaften über 200 Definitionen von Kultur, von denen jedoch keine einzige durch eine Mehrzahl von Wissenschaftern anerkannt und vertreten ist.

An dieser Stelle seien einige dieser Definitionen angeführt, um einen Überblick über ihre verschiedenen Prioritäten zu gewährleisten.

Kroeber und Kluckhohn (1952) definieren Kultur folgendermaßen:

„Kultur besteht aus expliziten und impliziten Verhaltensmustern, die durch Symbole erworben und vermittelt werden, die spezifische Leistung einer menschlichen Gruppe begründen, einschließlich ihrer Verkörperung in Kulturprodukten. Der Wesensgehalt der Kultur besteht aus tradierten (historisch gewachsenen und selektierten) Ideen und damit verbundenen Wertvorstellungen. Kulturelle Systeme können einerseits als Ergebnisse von Handlungen und andererseits als Bedingungs-elemente von Handlungen betrachtet werden.“ (Kroeber & Kluckhohn, 1952, S.181)

„Explizit“ sei hierbei als die Ansammlung beobachtbarer Verhaltensweisen anzusehen, die eine Gruppe von Menschen zu Tage bringt; „implizit“ wird durch jene Faktoren definiert, die der expliziten Form der Kultur zu Grunde liegen (Berry et al. 1992).

Trommsdorff (1993) versteht Kultur als einen „...Variablenkomplex, der die Entwicklung und das Handeln der Person u. a. mit beeinflusst sowie auch selbst vom Handeln der Person mit beeinflusst wird“ (S.108).

Berry et al. (1992) definieren Kultur etwas profaner und allgemeiner durch die Worte „...the shared way of life of a group of people“ (p. 1).

Anschließend seien noch einige Definitionen der „Kulturvergleichenden Psychologie“ angeführt:

Berry (1980) definierte diese anfänglich folgendermaßen:

„...cross-cultural psychology is defined primarily by its method.... Thus, it is our methodology that we must turn to in order to seek our identity as a discipline....“  
(Berry, 1980, p. 1).

17 Jahre später änderte er seine Definition jedoch, in dem er die „Kulturvergleichende Psychologie“ als die systematische Untersuchung zwischen dem kulturellen Kontext der menschlichen Entwicklung und den Verhaltensweisen, die in das Repertoire von Individuen einer bestimmten Kultur aufgenommen werden, bezeichnete (Berry, Poortinga & Pandey, 1997).

### **3.3 Ziele**

Kulturvergleichende Psychologen sind vor allem daran interessiert, die Gründe für kulturelle Unterschiede zu verstehen, mit dem Ziel, sich einer universell akzeptablen Psychologie anzunähern (Lonner und Adamopoulos, 1997). Im Vordergrund steht dabei das Auffinden universell gültiger Gesetzmäßigkeiten einerseits und das Erkennen jener kulturellen Bedingungen, welche die individuellen Merkmalsausprägungen beeinflussen, andererseits (Helfrich, 1993).

Thomas (1993, a) definiert dies folgendermaßen:

„Das Ziel der kulturvergleichenden Forschung ist die kritische Überprüfung der Generalisierbarkeit psychologischer Gesetzmäßigkeiten und die Analyse der spezifischen Wirkungen kultureller Kontextbedingungen auf die Art und Ausprägung psychischer Funktionen“ (S. 13).

Zu den wichtigsten Zielen der „Kulturvergleichenden Psychologie“ zählen nach Berry et al. (1992):

- Das Testen der Allgemeingültigkeit psychologischen Wissens und Theorien. Dieses Ziel dient dem Zweck, aktuelle Erkenntnisse und Wissen über psychische Gegebenheiten aus einer Kultur in eine andere zu übertragen, um deren Validität zu prüfen. Je nach Ergebnis können diese Erkenntnisse entweder auch auf jenen neuen Kontext generalisiert werden, oder eben nicht. Ziel dieser Vorgehensweise ist es, den Aktionsradius psychologischer Gesetzmäßigkeiten zu erweitern und stellt den „etic“ Ansatz dar (Berry et al., 1997).

Sollte dieses Ziel - der Versuch Theorien auf andere Kulturen zu erweitern - nicht erreicht werden, liegt es den Autoren zu Folge nahe, die

- Unterschiede zu untersuchen, welche diese Generalisierbarkeit verhindert haben. Denn aus diesen ergeben sich wertvolle Informationen für den Kulturvergleich und gleichzeitig bietet dies eine Möglichkeit, neue kulturelle Systeme zu erforschen, um Modifikationen psychischer Phänomene zu entdecken (Berry et al., 1997). Weiters dient die Suche nach Unterschieden dem Zweck, Interesse und Offenheit für die andere Kultur zu erwecken (Berry et al., 1992). Dies stellt den „emic“ Ansatz dar (Berry et al., 1992).
- Das dritte Ziel kann als eine „Integration“ der beiden ersten gesehen werden, mit dem Versuch, den Weg für eine universellere Psychologie zu ebnen. Dieser Weg ergibt sich aus der Generalisierung jener Theorien, die es zulassen und der Untersuchung von Gründen für die Unvereinbarkeit anderer Theorien (Berry et al., 1992).



Eine Übereinstimmung mit Berry et al. (1992) lässt sich auch bei Thomas (1993, a) finden. Ihm zu Folge sind die zentralen Aufgaben und Ziele zum einen, Verhaltens- und Erlebensunterschiede zwischen den verschiedenen Kulturen zu analysieren und zu beschreiben, andererseits zu prüfen, ob die daraus gewonnenen Theorien kulturgebunden sind oder auch auf andere Kulturen verallgemeinert werden können. Weiters muss seiner Ansicht nach untersucht werden, wie sich kulturell determinierte, psychologische Phänomene und Prozesse über die Zeit entwickeln und verändern und als letztes Ziel müssen die auf jenem Weg erhaltenen Forschungsergebnisse sinnvoll angewandt werden, um damit einen Schritt in Richtung der Lösung interkultureller Probleme zu beschreiten.

Der Erkenntniswert der „Kulturvergleichenden Psychologie“ liegt dabei vor allem darin, dass „...allgemeine Theorien psychischer Erscheinungen und psychologischer Gesetzmäßigkeiten auf ihre kulturspezifische oder universelle Geltung hin überprüft und gegebenenfalls modifiziert werden“ (Thomas, 1993, a, S.15).

Dies entspricht in etwa der Trennung zwischen dem „emic“ und dem „etic“ Ansatz. Diese Unterteilung geht auf den Linguisten Pike (1967, nach Thomas, 1993, b) zurück und hat seit jeher einen festen Platz in der „Kulturvergleichenden Psychologie“. Der etic-Ansatz versucht, Verhalten oder psychische Gesetzmäßigkeiten zu erforschen, die universell Gültigkeit haben, während der emic-Ansatz die zwischen den Kulturen verschiedenen, und für jede Kultur spezifischen Abwandlungen bestimmter psychischer Phänomene zu erfassen und sie vergleichend gegenüberzustellen sucht (z. B. Berry et al., 1992). Idealerweise ist diese Aufteilung als ein Kontinuum zu sehen, bei dem die Entwicklung des Messinstruments dem etic-Konzept folgen sollte und dessen Anwendung in den emic-Ansatz übergeht (Thomas, 1993, b).

### **3.4 Schwierigkeiten der Kulturvergleichenden Psychologie**

Die „Kulturvergleichende Psychologie“ ist ein bedeutungsvolles Teilgebiet der Psychologie, die mit der Suche nach Allgemeingültigkeit und Generalisierbarkeit psychologischen Wissens die Möglichkeit einer verbesserten interkulturellen Kommunikation und einen weiteren Schritt in Richtung einer einheitlichen Psychologie bietet. Neben ihren zahlreichen Vorzügen bringt sie allerdings auch Schwierigkeiten mit sich, die im Folgenden behandelt werden sollen.

Um einen Vergleich zwischen Kulturen in einem bestimmten Aspekt (hier sei vom Vergleich kognitiver Fähigkeiten ausgegangen) durchführen zu können, muss zunächst ein für beide Kulturen äquivalentes Messinstrument entwickelt werden. Dabei wird meist ein Instrument, das in einer bereits untersuchten Kultur hohe Validität und Reliabilität gezeigt und interessante Ergebnisse gebracht hat, in anderen Kulturen angewandt, um kulturelle Ähnlichkeiten oder Unterschiede zu untersuchen (Van de Vijver & Hambleton, 1996).

Der optimale Fall, ein Erhebungsinstrument von Anfang an, simultan für den Gebrauch in verschiedenen Kulturen zu entwickeln, stellt jedoch in der Forschung eher die Ausnahme als die Regel dar (Van de Vijver & Leung, 1997, b).

Die Verwendung eines in einer Kultur bereits erprobten Messinstruments in einer anderen Kultur ist mit vielfältigen Problemen verbunden. Damit ein Erhebungsinstrument, das nicht simultan entwickelt wurde, dennoch in anderen Kulturen angewandt werden kann, gibt es drei in der Literatur anerkannte Methoden (z. B. Van de Vijver & Leung, 1997 a, b):

- Die Anwendung desselben Instruments, das in einer Ausgangskultur bereits verwendet worden ist, durch einfache Übersetzung des selben (*apply*).
- Adaption des ursprünglichen Instruments für die neue Kultur (*adapt*).
- Erstellung eines völlig neuen Messinstruments für den neuen kulturellen Kontext. Dies ist dann notwendig, wenn das ursprüngliche Instrument dermaßen verändert werden muss, dass es sich quasi um ein neues Instrument handelt (*assemble*). Dies wird vor allem dann angewendet, wenn der Konstruktbias (siehe Kapitel 3.7.3) einen direkten Vergleich zwischen den Gruppen gefährdet.

Jede dieser Methoden hat bestimmte Vor- und Nachteile und für welche man sich entscheidet, hängt von vielfältigen Faktoren ab. Van de Vijver und Leung (1997, a) führen folgende Vorteile an:

Jene Methode, die auf Englisch mit „*apply*“ bezeichnet wird, nimmt an, dass eine einfache Übersetzung des Messinstruments für die Erhebung des Konstrukts in der neuen Zielgruppe genügt und dass der interessierende Faktor dadurch adäquat repräsentiert wird. Die Anwendung desselben Instruments hat den Vorteil, dass man Kulturen, die bereits mit diesem untersucht wurden, leichter vergleichen kann und damit die Skalenäquivalenz, die später in

Abschnitt 3.7.2 erklärt wird, gesichert ist. Ein weiterer Vorteil besteht in dem zeitlich und finanziell, weitaus geringeren Aufwand, der mit dieser Vorgehensweise im Vergleich zu den beiden anderen einhergeht.

Ist jedoch das bestehende Instrument sehr spezifisch auf jene Kultur zugeschnitten, für die es ursprünglich entwickelt wurde und deckt es wichtige Aspekte des interessierenden Merkmals in der anderen Kultur nicht ab, sollte man es den Autoren zu Folge vorziehen, eine andere Variante zu wählen.

In diesem Sinne scheint die Adaption des Instruments dann opportun, wenn das Instrument in beiden Kulturen an sich dasselbe misst und nur einige Items inadäquat sind und einer Anpassung bedürfen. Ist dies der Fall, werden die störenden Items entweder verändert, herausgenommen oder eventuell durch andere ersetzt (Van de Vijver & Leung, 1997, a).

Ein völlig neues Instrument muss dann erstellt werden, wenn die ersten beiden Möglichkeiten keine passende Vorgehensweise darstellen, weil etwa die zu vergleichenden Kulturen eine andere Vorstellung von beispielsweise Intelligenz haben und das ursprüngliche Instrument dermaßen verändert werden müsste, dass es sich quasi um ein neues Instrument handelt. Die Notwendigkeit einer völlig neuen Entwicklung des Testinstruments ist jedoch im Vergleich mit einer Übersetzung oder Adaption des bereits bestehenden Instruments mit einem ungleich höheren Arbeits- und Kostenaufwand verbunden (Hambleton & Kanjee, 1995) und diese Vorgehensweise nimmt auch die Möglichkeit, mit dem ursprünglichen Instrument erhobene normative Daten, zum Vergleich heranzuziehen (Brislin, 1986, nach Hambleton & Kanjee, 1995).

### **3.5 Möglichkeiten der Übersetzung eines Testinstruments**

Nachfolgend werden jene Methoden der Übersetzung eines Messinstruments angeführt, die in der Literatur am häufigsten vorgeschlagen werden.

Vorweggenommen sei noch die in der Literatur weithin anerkannte Tatsache, dass es ungeachtet dessen, welcher Übersetzungsstrategie man sich bedient, eines ausgeglichenen Verhältnisses psychologischer, sprachlicher und kultureller Überlegungen bedarf und keineswegs nur die einfache Übersetzung eines Instruments in eine andere Sprache ausreichend ist (Van de Vijver & Tanzer, 2004).

1. *Translation-backtranslation* Variante (Werner & Campbell, 1970). Hierbei wird ein Test von der Originalsprache in die Zielsprache übersetzt und in einem weiteren Schritt nimmt sich eine oder mehrere andere (im besten Fall bilinguale) Personen der Rückübersetzung aus der Zielsprache in die Ursprungssprache an. Danach werden die beiden Versionen der Originalsprache verglichen und analysiert und eventuelle Änderungen können anhand eines wissenschaftlichen Konsenses vorgenommen werden. Diese Vorgehensart kann zahlreiche Fehler aufdecken und ist die Methode des geringsten Aufwands, welche die beste Möglichkeit für das Erreichen einer hohen Äquivalenz darstellt (Van de Vijver & Leung, 1997, b). Dennoch ist mit einer guten sprachlichen Übersetzung noch keine psychologisch hinreichende Qualität garantiert und im Rahmen professioneller Übersetzungen wird diese Methode kaum noch angewandt, da es oft zu Verständnisproblemen kommt. Manche Begriffe und Begriffskonstrukte können schlichtweg nicht wörtlich übersetzt werden oder es gibt in einer der beiden Sprachen kein passendes Äquivalent.

Diese Methode beinhaltet die Gefahr, dass eine, von einem linguistischen Standpunkt zwar korrekte Übersetzung, in der Zielkultur zu Verwirrungen in Lesbarkeit und Verständlichkeit führen kann (Van de Vijver & Leung, 1997, a).

2. *Committee approach*: dies ist der zweite und häufiger verwendete Ansatz; hierbei wird die Übersetzung des interessierenden Instruments von einer Gruppe kompetenter bilingualer Personen, die oft aus verschiedensten Bereichen (beispielsweise Experten für Kultur, Linguistik oder Psychologie) kommen, vorbereitet. Ein wesentlicher Vorteil dieser Vorgehensweise besteht in der Kooperation zwischen verschiedenen Fachgebieten, die den Einfluss differenzierter Sichtweisen und Prioritäten zulassen (Van de Vijver & Leung, 1997, a, b; Van de Vijver & Tanzer, 2004). Dieser Ansatz wird auch oft mit der „translation-backtranslation“ Methode kombiniert (Van de Vijver & Leung, 1997, b) und führt zu einer gesteigerten Qualität der Übersetzung durch die Expertise der daran beteiligten Personen; als Nachteil sei jedoch angeführt, dass keine externe Kontrolle der übersetzten oder adaptierten Version durch Unabhängige geboten wird (Van de Vijver & Leung, 1997, b).

3. *cultural decentering* (Werner & Campbell, 1970) bietet eine weitere Alternative zur Testübersetzung. Ziel dieser Methode ist es, Wörter oder Ausdrücke der „Ursprungsversion“ zu eliminieren, die schwierig zu übersetzen oder nur für diese Kultur typisch sind. Anhand dieser Überlegung wird nicht die neue Version, sondern die ursprünglich verwendete geändert, um den Vergleich zu erleichtern. Mit dieser Methode ist ein großer Arbeitsaufwand verbunden, da sie nach der Zusammenarbeit multikultureller und mehrsprachiger Mitglieder verlangt, die alle über ein profundes Wissen des interessierenden Konstrukts der Studie verfügen sollten (Van de Vijver & Leung, 1997, b), allerdings ist es jene Methode, die den Zielen einer kulturfairen (siehe Kapitel 4.2) Testung am nächsten kommt (Van de Vijver & Leung, 1997, b).

Um eine professionelle Herangehensweise an die Anwendung von Testinstrumenten im interkulturellen Kontext zu gewährleisten, wurde im Jahr 1978 die „International Test Commission“ (ITC) gegründet. Diese Organisation zielt auf die Verbesserung der Kooperation zwischen Personen, die mit psychodiagnostischen Instrumenten arbeiten, ab. Seit 1993 existiert in diesem Rahmen auch ein internationales Komitee von Psychologen, die Richtlinien zur Übersetzung, Adaption, sowie Vorgabe von Messinstrumenten publizieren (Van de Vijver & Hambleton, 1996). Eine Anführung dieser wäre hier zu ausführlich, der interessierte Leser sei deshalb auf die Website der „International Test Commission“ verwiesen (<http://www.intestcom.org>).

Die Auswahl einer geeigneten Vorgehensweise für die Anwendung eines Messinstruments in einer anderen Kultur ist jedoch erst der Anfang einer kulturvergleichenden Untersuchung; danach muss geprüft werden, ob das Instrument auch tatsächlich dasselbe misst. Dazu dienen die in Kapitel 3.7.1 beschriebenen „vorausgehenden Analysen“.

### **3.6 Stichprobe**

Um einen validen Vergleich zwischen verschiedenen Kulturen gewährleisten zu können, müssen sich die verglichenen Personen im Hinblick auf relevante Hintergrundcharakteristika sehr ähnlich sein. Nur dadurch wird garantiert, dass eventuell gefundene Unterschiede auf kulturelle Faktoren zurückzuführen sind und nicht auf bereits vorher vorhandene

Unterschiede in den Stichproben (Van de Vijver & Leung, 1997, a). Ebenso muss die Stichprobe repräsentativ für die Population sein, aus der sie gezogen wurde. Nur wenn dieser Fall eintritt, können die gefundenen Ergebnisse auf die gesamte Kultur generalisiert werden (Berry et al., 1992).

Van de Vijver und Leung (1997, a) nennen 3 mögliche Arten der Stichprobenauswahl:

- *Convenience sampling*: Bei dieser Vorgehensweise werden die zu vergleichenden Kulturen auf Grund von praktischem Nutzen und oft ohne ausreichende theoretische Vorüberlegungen ausgewählt (Van de Vijver & Leung, 1997, a). Gründe für die Auswahl spezifischer Kulturen können hierbei unter anderem die Zugehörigkeit des Wissenschafters zu der ausgewählten Kultur sein oder dass dieser gute Beziehungen zu einer bestimmten Kultur unterhält und ihm dadurch der Zugang zu Personen derselben erleichtert wird.
- *Systematic sampling*: Hierbei werden die zu vergleichenden Kulturen, anhand detaillierter Vorarbeit, systematisch und theoriegeleitet ausgewählt (Van de Vijver & Leung, 1997, a).
- *Random sampling*: Diese Herangehensweise sieht es vor, dass viele verschiedene Kulturen zufällig ausgewählt werden und stellt damit die schwierigste, kosten- und ressourcenaufwändigste Variante der Stichprobenauswahl dar (Van de Vijver & Leung, 1997, a).

Jede dieser drei Varianten dient unterschiedlichen Zwecken, wobei jedoch den letzten beiden wenn möglich, Vorrang zu geben ist.

## **3.7 Psychometrische Qualitäten**

### **3.7.1 Vorausgehende Analysen**

Diese müssen vor der eigentlichen Datenanalyse durchgeführt werden, um zu prüfen, ob das Messinstrument tatsächlich für den Kulturvergleich geeignet ist. Dabei untersucht man

Intergruppendifferenzen, um zu sehen, ob sich die Gruppen oder Kulturen in bestimmten Fähigkeiten unterscheiden. Sollte derartiges auftreten, kann das zwei Ursachen haben: entweder die Gruppen unterscheiden sich tatsächlich oder die Unterschiede basieren auf verschiedenen Arten von Bias (Van de Vijver & Leung, 1997, a). Auf diese wird später in Kapitel 3.7.3 eingegangen.

### 3.7.2 Äquivalenz

Äquivalenz ist die Voraussetzung für jeden validen Kulturvergleich und bezieht sich auf die Vergleichbarkeit der Daten in den untersuchten Kulturen (Van de Vijver & Leung, 1997, a).

Es gibt drei Arten von Äquivalenz, die von Van de Vijver und Leung (1997, a, b) in eine dreistufige Hierarchie unterteilt wurden:

1. *structural equivalence* (Konstruktäquivalenz): Bedeutet, dass in den verglichenen Kulturen dasselbe Konstrukt gemessen wird, unabhängig davon, ob dieses mit denselben Instrumenten erhoben wurde. Die Konstruktäquivalenz ist eine Grundvoraussetzung jeder vergleichenden Untersuchung (Helfrich, 1993).
2. *measurement unit equivalence* (Messeinheitsäquivalenz): Bedeutet, dass die Scores der verschiedenen Kulturen zwar in derselben Messeinheit erhoben wurden, aber dass diese nicht denselben Ursprung aufzeigen müssen; die den Messungen zu Grunde liegende Skala kann also eine völlig andere sein.  
Folgendes Beispiel kann hier zum Verständnis dienen. Wird etwa eine Temperaturmessung anhand verschiedener Skalen durchgeführt, hier sei angenommen, einmal mittels der Kelvin- und einmal mittels der Celsiusskala, so haben beide dieselbe Messeinheit, aber sie unterscheiden sich um 273 Grad. Dieser Umstand führt dazu, dass die beiden Skalen nicht direkt miteinander verglichen werden können. Da man aber weiß, um wie viel Grad sie sich unterscheiden, ist eine Vergleichbarkeit gewährleistet (Van de Vijver & Tanzer, 2004; Van de Vijver & Leung, 1997, a).
3. *scalar equivalence* (Skalenäquivalenz): Dies ist das höchste erreichbare Level der Äquivalenz. Wenn zwei metrische Messungen nicht nur dieselbe Messeinheit besitzen

sondern auch der Ursprung dieser Identität ist, dann spricht man von Skalenäquivalenz oder *full score comparability*. Diese Art der Äquivalenz erlaubt den Vergleich von Scores sowohl innerhalb einer Kultur, als auch über beide Kulturen (Van de Vijver & Leung, 1997, a).

Die Außerachtlassung der Kontrolle von Datenäquivalenz gefährdet jede interkulturelle Untersuchung und beinhaltet laut Salzberger und Sinkovics (2006) die Gefahr falscher Schlussfolgerungen, wobei

- beobachtete Messunterschiede auf grundsätzliche Unterschiede zwischen den Populationen zurückgeführt werden können, obwohl sie durch unterschiedliches Antwortverhalten und nicht durch tatsächlich vorhandene Unterschiede in den latenten Variablen entstanden sind.
- Ebenso ist der umgekehrte Fall denkbar, in dem echte Unterschiede durch verschiedenes Antwortverhalten verdeckt sind und nicht erkannt werden.

Aus diesen Gründen muss in allen Stufen des Tests (Erstellung, Vorgabe, Datensammlung etc.) penibel auf die Vergleichbarkeit der Daten geachtet werden, da andernfalls ein Vergleich der Populationen nicht aussagekräftig ist.

### **3.7.3 Bias**

Während Äquivalenz mit dem Messlevel in Verbindung gebracht wird, ist Bias die generelle Bezeichnung für Störfaktoren in interkulturellen Scorevergleichen (Van de Vijver & Tanzer, 2004).

Van de Vijver und Leung (1997, a, b) sowie Van de Vijver und Poortinga (1992) unterscheiden drei Arten von Bias:

1. *construct bias* (Konstruktbias): Dieser tritt auf, wenn das gemessene Konstrukt in den verglichenen Kulturen nicht oder nur teilweise dasselbe ist. Beispielsweise wenn ein IQ-Test, der auf westliche Kulturen zugeschnitten ist, unverändert auch in nicht-westlichen Kulturen vorgegeben wird, wobei sich hier etwa das herrschende Intelligenzkonzept der verglichenen Kulturen stark unterscheiden kann und der



Testinhalt somit nicht dasselbe erhebt (Van de Vijver & Tanzer, 2004). Ebenso könnten in einer der verglichenen Kulturen nicht alle relevanten Verhaltensweisen erfasst werden, oder die Konstruktdefinitionen stimmen in den verschiedenen Kulturen nicht ausreichend überein (Van de Vijver & Leung, 1997, b).

Um Konstruktbias zu vermeiden, ist es von Bedeutung, Wörter und Konzepte, die für eine Kultur spezifisch sind, im interkulturellen Vergleich herauszunehmen oder bereits in der Entwicklungsphase des Tests eine zweite Version für die interessierende Kultur mitzugestalten (Van de Vijver & Leung, 1997, b) (siehe Kapitel 3.5).

2. *method bias* (Methodenbias): Diese Art von Bias tritt auf, wenn das psychologische Konstrukt vom Messinstrument zwar gut repräsentiert wird, es aber dennoch zu ungewollten Intergruppendifferenzen kommt, die durch Fehler der Untersuchung auftreten (Van de Vijver & Leung, 1997, a). Der Methodenbias wird wiederum in drei Gruppen unterteilt (Van de Vijver & Tanzer, 2004):

- a) *Sample bias* (Inkompatibilität der Stichproben): Unvergleichbarkeit der Stichproben, weil diese sich zwar nicht im Zielkonstrukt unterscheiden, dafür aber in anderen relevanten Charakteristika wie etwa Unterschieden im Schulsystem, Erfahrung mit Tests und der Neuheit der Testsituation, bzw. der Motivation diese zu bearbeiten (Van de Vijver & Tanzer, 2004).
- b) *Instrument bias*: Dieser kann auftreten, wenn die Testpersonen z. B. über eine unterschiedliche Vertrautheit mit dem Stimulusmaterial, dem Antwortformat oder der Testsituation verfügen.
- c) *Administration bias* „...is a generic term for validity threatening factors that are related to instrument administration“ (Van de Vijver & Hambleton, 1996, p.91).

Diese Definition beinhaltet Störfaktoren, die allein durch die Vorgabe des Messinstruments auftreten. Darunter können generelle Umgebungsfaktoren bei der Testvorgabe (physisch, psychisch oder sozial) fallen, unterschiedliche Vertrautheit der Testpersonen mit eventuellem technischem Equipment, Motivation, Einzel- vs. Gruppenvorgabe, uneindeutige Instruktion, sowie auch Kommunikationsprobleme zwischen dem Testleiter und der Testperson (Van de Vijver & Tanzer, 2004).

Methodenbias kann jedoch verhindert werden, indem man, im Sinne einer multiprofessionellen Herangehensweise, bereits bei der Testentwicklung oder der Übersetzung, mit Personen kollaboriert, die die jeweils interessierende Sprache und Kultur kennen oder noch besser dieser angehören (Van de Vijver & Tanzer, 2004). Dadurch können unnötige Fehler durch spezifische Unterschiede vermieden werden.

3. *item bias* (Itembias) oder *differential item functioning* (DIF) führt zur Verzerrung der Items und tritt dann auf, wenn Personen, die über gleiche Fähigkeiten verfügen, nicht dieselbe Wahrscheinlichkeit haben, eine Aufgabe korrekt zu lösen (Prieto, Novick, Sacristán, Edgell & Alonso, 2003; Van de Vijver & Leung, 1997, a). Dies kann im Falle eines Kulturvergleichs darauf beruhen, dass diese verzerrten Items unterschiedliche psychologische Bedeutung in den verschiedenen Kulturen haben. Wenn etwa ein Item in den verglichenen Kulturen ein unterschiedliches Ausmaß an sozialer Erwünschtheit innehat, wäre ein Vergleich der Gesamtttestscores ungültig, sofern dieses Item einbezogen wird. Daher werden jene Items, die DIF zeigen, meist aus der Ursprungsversion des Erhebungsinstruments entfernt, um die Vergleichbarkeit zu gewährleisten (Prieto et al., 2003; Van de Vijver & Leung, 1997, a). Itembias kann viele Ursachen haben vor allem aber eine schlechte Übersetzung der Items, im Sinne einer inadäquaten Itemformulierung (Van de Vijver & Leung, 1997, a).

Wenn Bias vorhanden ist, verringert dies den Grad erreichbarer Äquivalenz. Die Skalenäquivalenz, als höchste der drei Äquivalenzarten, duldet beispielsweise keinen der drei oben erwähnten Biasarten. Und ein Konstruktbias macht einen direkten Scorevergleich zwischen Kulturen unmöglich, da, sobald er vorhanden ist, nicht dasselbe Konstrukt erhoben wird (Van de Vijver & Tanzer, 2004).

### **3.7.3.1 Möglichkeiten Bias zu verhindern**

Um das Auftreten eines Bias zu verhindern, gibt es vielfältige Methoden, die auch hier wieder je nach Art des Bias unterschieden werden.

Zur Verhinderung eines Konstruktbias ist es sinnvoll, die oben bereits erwähnte Methode des „cultural decentering“ auf Grund ihrer Vorteile anzuwenden (Van de Vijver & Tanzer, 2004).

Um sowohl Konstrukt-, als auch Methodenbias zu verhindern, kann es nützlich sein, sich das Wissen einer Person aus der interessierenden Kultur und Sprache zu Nutzen zu machen und die übersetzte oder adaptierte Version des Erhebungsinstruments durch diese Personen auf sprachliche Korrektheit und Sinnhaftigkeit überprüfen und verbessern zu lassen (Van de Vijver & Tanzer, 2004).

Um die Gefahr des Methodenbias zu umgehen, sollte ein genaues Training der Versuchsleiter stattfinden, um zu sichern, dass sie allen Teilnehmern jeder Kultur dieselbe Instruktion geben und dabei auf kulturspezifische Besonderheiten achten (Van de Vijver & Tanzer, 2004). Ebenso ist die Entwicklung detaillierter Manuale und Administrationsprotokolle von Vorteil, die genau definieren, wie das Messinstrument vorzugeben ist, wie die Testpersonen instruiert werden müssen und wie sich der Versuchsleiter bei Problemen verhalten soll (Van de Vijver & Tanzer, 2004).

Auch um Itembias zu verhindern, ist besonders eine Zusammenarbeit mit Personen der Zielkultur von großer Bedeutung.

Die Testvorgabe ist also wie im vorigen Abschnitt aufgezeigt wurde, mit vielfältigen Problemrisiken behaftet. Van de Vijver und Leung (1997, a) zu Folge reichen diese von einer Gefährdung der Ergebnisvalidität durch die bloße Anwesenheit eines Testleiters, der einer anderen Kultur als die Testperson angehört, über Kommunikationsprobleme zwischen Testleiter und Versuchsperson, bis hin zu Unterschieden in wichtigen Charakteristika der Stichprobe und deren Erfahrung mit situationsbestimmenden Merkmalen (Van de Vijver & Leung, 1997, a).

Van de Vijver und Hambleton (1996) führen im Weiteren folgende Aspekte an:

- Es muss darauf geachtet werden, dass Faktoren, die mit dem Stimulusmaterial, der Vorgabesituation und dem Antwortformat zusammenhängen, für alle Teilnehmer unabhängig von ihrer Kulturzugehörigkeit gleich verstanden werden.
- Auch oder besonders in kulturvergleichenden Untersuchungen ist es wichtig, dass die Umgebungsbedingungen während allen Testungen für alle Personen gleich sind (Lichtverhältnisse, Temperatur und z. B. bei Computergestützten Testungen gleicher Abstand der Probanden zum Bildschirm etc.).

- Das Testmanual soll auf spezifische Aspekte, die sich als problematisch erweisen können, eingehen. Dies soll gewährleisten, dass die jeweiligen Versuchsleiter auf eventuell auftretende Fragen gleich reagieren und gleiche Antworten geben können. Besonders wichtig ist dies im Hinblick auf das vorher erwähnte Risiko der Versuchsleiter-Versuchsperson-Interaktion und um Effekte dieser zu verhindern.

Die Konklusion der Vorgehensweise interkultureller Untersuchungen kann somit folgendermaßen zusammengefasst werden: nur wenn in allen Schritten einer Untersuchung sorgfältig die Gefahren möglicher Probleme berücksichtigt werden, kann ein interkultureller Vergleich valide, fair und sinnvoll durchgeführt werden.

### **3.8.7.1 Möglichkeiten Bias aufzudecken**

Um Itembias oder *differential item functioning* (DIF) und damit invariante Items zu entdecken, gibt es mittlerweile zahlreiche statistische Methoden; darunter finden sich die Varianzanalyse und vor allem die Mantel-Haenszel Statistik (Holland & Thayer, 1988, nach Van de Vijver & Leung, 1997, a) für dichotome Daten. Letztere testet, ob ein einzelnes Rasch-Modell für die Daten in jeder Gruppe passt (Van de Vijver & Leung, 1997, a).

Eine andere Methode, um DIF zu entdecken, bietet die *Item Response Theory* (IRT), die im folgenden Kapitel behandelt wird. Sie nimmt an, dass ein Item, das keinen Bias zeigt, dieselbe Bedeutung für Menschen verschiedener Gruppen oder Kulturen hat.

Um DIF zu testen, untersucht man den Mittelwert der Residuen (z. B. zwischen erwartetem und erreichtem Itemscore einer Person). Wenn kein DIF vorhanden, ist der Mittelwert der Residuen gleich Null und es kann von Skalen- und Datenäquivalenz ausgegangen werden; ist jedoch DIF vorhanden, weicht der Mittelwert systematisch von Null ab (Salzberger & Sinkovics, 2006). Hambleton und Kanjee (1995) weisen darauf hin, dass die Untersuchung von DIF eine statistische Methode ist, die zwar jene Items aufdecken kann, die auf irgendeine Weise problematisch sind, anhand dieser Methode kann man jedoch nichts über ihre Ursachen aussagen.

## 4. Die Item Response Theory (IRT)

„IRT is a field of inquiry in the methodology of measurement, in which the characteristics of the items composing an assessment are the focus of the study, and these characteristics and the underlying individual abilities or traits are described and analysed” (Prieto et al., 2003, p.26).

Die IRT ist eine probabilistische Testtheorie, die sich gewissermaßen aus den Nachteilen der Klassischen Testtheorie<sup>1</sup> (KTT) weiterentwickelt hat, um einige der messtheoretischen Probleme dieser zu lösen (Hambleton, Swaminathan & Rogers, 1991).

Grundidee der IRT ist es, eine Verbindung zwischen Itemantworten und dem Fähigkeitslevel einer Person herzustellen, das diesen zu Grunde liegt. Untersucht werden dabei latente Variablen (beispielsweise Raumvorstellungsfähigkeit), die einer direkten Beobachtung nicht zugänglich sind und aus diesem Grunde über die Antworten auf die einzelnen Testitems, die als Indikatoren fungieren, erhoben werden. Dabei wird der Lösung eines Items eine bestimmte Wahrscheinlichkeit zugeteilt, welche von der Fähigkeit einer Person, das Item zu lösen (der Ausprägung auf der latenten Variable) und von der Itemschwierigkeit abhängig ist.

Diese Wahrscheinlichkeit einer bestimmten Antwort auf ein Testitem wird als (logistische) mathematische Funktion von Personen- und Itemmerkmalen definiert und kann anhand der so genannten *Item Characteristic Curve* (ICC) oder *Itemfunktion* dargestellt werden. Die ICCs werden anhand dreier Parameter definiert; der Diskriminationsfähigkeit des Items, der Itemschwierigkeit und dem Grad, zu dem die Antwort durch Raten beeinflusst wird (Van de Vijver & Leung, 1997, a). Diese Kurven zeigen, dass vor allem bei Leistungstests der Kurvenverlauf monoton steigt und damit die Wahrscheinlichkeit einer richtigen Antwort auf ein Item größer wird, je höher die zu Grunde liegende Eigenschaft oder Fähigkeit ist (Rost, 2004).

---

<sup>1</sup> Die schwerwiegendsten Nachteile der KTT sind in der Stichprobenabhängigkeit von Test-, Item- und Personenstatistiken zu sehen und darin, dass die Annahme der Gleichheit der Messfehler über alle Merkmalsausprägungen empirisch nicht begründbar ist (Hambleton, Swaminathan & Rogers, 1991).

Der größte Vorteil der IRT gegenüber der KTT besteht in der Überwindung der Stichprobenabhängigkeit dieser; mittels Modellen der IRT erhaltene Ergebnisse sind sowohl von der Stichprobe als auch von dem verwendeten Messinstrument unabhängig. Dies macht die IRT besonders geeignet für die Anwendung in kulturvergleichenden Studien.

So wird beispielsweise im einparametrischen logistischen Modell<sup>2</sup> die Itemschwierigkeit auf einer linearen Intervallskala geschätzt, welche die Möglichkeit bietet, die Äquivalenz der Items über verschiedene Stichproben und Kontexte hinweg (dies inkludiert auch verschiedene sprachliche oder kulturelle Settings) zu untersuchen (Prieto et al., 2003). Dabei wird der Zusammenhang zwischen der Wahrscheinlichkeit einer bestimmten Antwort und einer latenten Merkmalsvariable (wahrer Wert) untersucht (Van de Vijver & Leung, 1997, a).

## 4.1 Das dichotome logistische Modell von Rasch

Das Rasch-Modell (Rasch, 1960) ist eine spezielle Form und gleichzeitig auch eines der einfachsten Modelle der IRT, das auf Items mit dichotomem Antwortformat angewandt werden kann (z. B. Item gelöst oder nicht gelöst).

Wenn für ein Testinstrument das Rasch-Modell Gültigkeit hat, bedeutet dies, dass für alle Personen dieselbe Fähigkeitsdimension gemessen wird und dass nur eine latente, zu Grunde liegende Eigenschaft für die Antworten und Leistungen verantwortlich ist (auch hier kann jedoch, wie in der KTT, ein Zufallseinfluss nicht ausgeschlossen werden).

Das Rasch-Modell kann durch eine logistische Formel beschrieben werden (siehe Abb.4.1.1), welche durch Item- und Personenparameter spezifiziert wird. Diese beschreibt die Wahrscheinlichkeit  $P(x_{vi})$ , dass eine Person  $v$  mit der Fähigkeit  $\theta_v$ , ein dichotomes Item  $i$  mit der Schwierigkeit  $\delta_i$ , löst.

---

<sup>2</sup> Dies ist eines der einfachsten Modelle der IRT und gleichzeitig das wohl bekannteste, welches zu Ehren seines Entwicklers, dem dänischen Mathematiker Georg Rasch, unter dem Namen „Rasch-Modell“ bekannt ist.

$$p(x_{vi}) = \frac{\exp(x_{vi}(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)}$$

Abb.4.1.1 Die Formel des Rasch-Modells (nach Rost, 2004, S.121)

Die Antwort einer Person wird also nur durch deren Fähigkeit anhand eines Fähigkeitsparameters und die Schwierigkeit der Aufgabe anhand eines Schwierigkeitsparameters bestimmt. Dabei werden die Item- und Personenparameter auf einer gemeinsamen Skala konzipiert und können dennoch unabhängig voneinander geschätzt werden (dies wird auch als Invarianz-Eigenschaft bezeichnet). Diese Eigenschaft - dass Schätzer von  $\theta_v$  frei von Effekten von  $\delta_i$  sind und umgekehrt - wird nur durch das Rasch-Modell gewährleistet (Wright & Stone, 1979).

Wie aus Abb.4.1.2 ersichtlich, kann die Funktionsgleichung des Rasch-Modells auch graphisch als *Itemcharakteristik* oder *Itemfunktion* dargestellt werden:

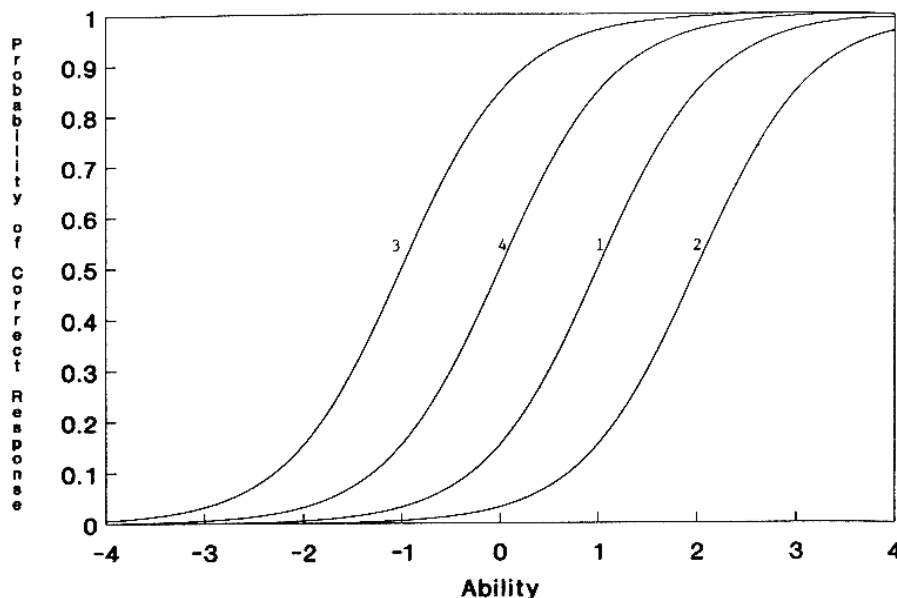


Abb.4.1.2 Eine einparametrische Itemcharakteristik Kurve (nach Hambleton et al., 1991, p. 14)

Anhand dieser Kurven kann man den Verlauf der Lösungswahrscheinlichkeiten eines Items in Abhängigkeit der latenten Variable ablesen.

Dabei werden auf der x-Achse ( $\theta_v$ ) die Personen skaliert und auf der y-Achse wird die Wahrscheinlichkeit, das Item zu lösen, aufgetragen. Die Schnittpunkte zwischen Fähigkeit und Wahrscheinlichkeit zeigen die Itemschwierigkeit auf (Hambleton et al., 1991).

In diesem Fall (siehe Abb.4.1.2) sind die Itemcharakteristik-Kurven für alle vier Items parallel, woraus sich erkennen lässt, dass der Unterschied in der Schwierigkeit der Items für alle Personen gleich ist und dass dieser Vergleich somit nicht von den Personen abhängt, die verglichen werden (Amelang & Bartussek, 2001).

Das einparametrische Rasch-Modell postuliert, dass Itemschwierigkeit die einzige Itemeigenschaft ist, welche die Testleistung beeinflusst und somit dürfen sich die ICCs nur in ihrer Position auf der x-Achse unterscheiden; ihr Verlauf muss parallel sein (Hambleton et al., 1991). Jenes Item, das auf der x-Achse am weitesten links liegt, stellt das leichteste Item und jenes, das am weitesten rechts liegt, das schwierigste dar (Rost, 2004).

#### **4.1.1 Voraussetzungen für die Geltung des Rasch-Modells**

- *Eindimensionalität*: Der Test (und somit jedes Item) erfasst für jede Person nur eine (dieselbe) Fähigkeitsdimension.
- *Lokale stochastische Unabhängigkeit*: Die Wahrscheinlichkeit, ein Item zu lösen, hängt ausschließlich von der Fähigkeit der Person und der Itemschwierigkeit ab (und von dem Zufall, der jedoch nirgends ausgeschlossen werden kann) (Embretson & Reise, 2000). Sie besagt, dass Antworten von Personen, die über die gleiche Fähigkeitsausprägung verfügen, unabhängig sind und dass die korrekte Lösung eines Items nicht von der Lösung eines anderen Items abhängig sein darf.
- *Spezifische Objektivität*: Vergleiche zwischen Objekten müssen über den spezifischen Kontext, in dem sie gemessen wurden, generalisierbar sein (Rasch, 1977, nach Embretson & Reise, 2000). Das bedeutet, dass die Itemparameterschätzer und die Personenparameterschätzer voneinander unabhängig sind (Salzberger & Sinkovics, 2006).

Spezifische Objektivität beinhaltet weiters folgende Annahmen:

- a) Vergleiche zwischen Personen sind über die spezifischen Items, die verwendet wurden, um sie zu messen, invariant; der Fähigkeitsunterschied



zweier Personen kann unabhängig davon, welche Items ihnen vorgegeben werden, erhoben werden (Kubinger, 2006).

b) Vergleiche zwischen Items bleiben gleich über die spezifischen Personen, die verwendet wurden, um sie zu messen; Unterschiede in den Items hängen demnach nicht von den Personen ab, die für den Vergleich herangezogen wurden (Kubinger, 2006).

- *Suffizienz*: Die Gesamtzahl der richtigen Antworten (*total number correct score*) einer Person ist ausreichend, um die unbekannt Parameter zu schätzen und der Itemsummenscore (Summe aller richtigen Antworten auf einem Item) ist suffizient, um unbekannt Itemparameter zu erhalten (Molenaar, 1995).

Die Item- und Personenparameter werden mittels Maximum-likelihood Verfahren geschätzt.

- *Ausschluss von Rateparametern*: Das Rasch-Modell sieht im Gegensatz zu anderen Modellen keinen Parameter zum Ausgleich von Ratewahrscheinlichkeiten vor. Daraus ergibt sich die Notwendigkeit, dass die Wahrscheinlichkeit des Ratens von vornherein durch das Testdesign ausgeschlossen oder minimiert wird<sup>3</sup> (Nenty & Dinero, 1981).

#### 4.1.2 Prüfung des Rasch-Modells

Die Überprüfung der Geltung des Rasch-Modells und somit der Rasch-Homogenität kann unter anderem mittels des asymptotisch  $\chi^2$ -verteilten bedingten Likelihood-Quotienten Test (*Likelihood Ratio Test* (LRT)) nach Andersen (1973) durchgeführt werden. Dieser Test basiert auf der *Conditional Maximum Likelihood* Schätzung (CML) und ist nach Rost „das Kernstück von Tests auf Personenhomogenität“ (Rost, 2004, S.345).

Der LRT prüft, ob sich die Itemparameter der verschiedenen Scoregruppen signifikant voneinander unterscheiden. Hierbei müssen sowohl die Itemparameter der gesamten Stichprobe als auch jene jeder Scoregruppe geschätzt werden (Rost, 1988). Wird in allen Gruppen dieselbe Fähigkeit gemessen, müssen die Itemschwierigkeiten über die Gruppen ident sein, d. h. die Itemparameter der verschiedenen Gruppen unterscheiden sich nicht signifikant voneinander (Rost, 2004).

---

<sup>3</sup> Etwa durch die Antwortalternative „Ich weiß die Lösung nicht“ oder „keine Lösung ist richtig“.

Die LRTs können nach internen und externen Kriterien durchgeführt werden. Die internen Kriterien beziehen sich auf den Test selbst (eine häufige Methode der internen Teilung besteht darin, die Stichprobe anhand Median oder Mittelwert in zwei Gruppen zu teilen, etwa in hohen und niedrigen Score), die externen Kriterien beziehen sich auf die Stichprobe (beispielsweise Geschlecht, Alter, Grad der höchsten abgeschlossenen Schulausbildung, Händigkeit etc.).

## **4.2 Exkurs: Besondere Vorteile der IRT im Kulturvergleich**

Van de Vijver und Leung (1997, b) führen 2 Gründe an, warum die IRT in der Anwendung kulturvergleichender Untersuchungen besonders sinnvoll ist:

1. die geschätzten Parameter (siehe Kapitel 10) hängen, im Gegensatz zu Untersuchungen mittels KTT, nicht von der Ausprägung der Gruppe auf dem latenten trait ab. Vielmehr hängt die Itemschwierigkeit vom durchschnittlichen Fähigkeitslevel der gesamten Gruppe ab (Van de Vijver & Leung, a).
2. Die Schätzung der Personenfähigkeitsparameter ist unabhängig von den verwendeten Items. Dies bedeutet, man kann auch Personen (Kulturen) vergleichen, die nicht dasselbe Itemset bearbeitet haben, solange dieses groß genug ist um die interessierende Eigenschaft angemessen zu erfassen.

Ein weiterer Grund, der die IRT besonders wertvoll für einen interkulturellen Vergleich macht, ist die Möglichkeit einer kulturfairen Testung. Entsprechen alle Items eines psychometrischen Messinstruments dem Rasch-Modell und misst dies somit in den untersuchten Kulturen dieselbe latente Fähigkeitsdimension, kann davon ausgegangen werden, dass zwischen den Kulturen Chancengleichheit besteht.

Dies ist ein wertvoller Hinweis darauf, dass sich eventuell ergebende Unterschiede auf die tatsächlichen Leistungsunterschiede zwischen den Kulturen zurückführen lassen und nicht Artefakte ungleicher, eine Kultur bevorzugender Bedingungen sind (sofern sich die Stichproben in wesentlichen Parametern wie Alter, Geschlechtsverteilung, Bildung etc. nicht zu sehr unterscheiden).

## 5. Raumvorstellung

Die Fähigkeit zum Raumvorstellungsvermögen ist ein wesentlicher Bereich unseres täglichen Lebens und darüber, dass Raumvorstellung einen wichtigen und gesicherten Teil der Intelligenz darstellt, herrscht allgemein Einigkeit. Doch obwohl man sich in der Psychologie bereits seit Anfang des 19. Jahrhunderts mit der Erfassung und Definition von Raumvorstellung auseinandersetzt, ist es nicht gelungen, vorbehaltlos angenommene Ergebnisse zu erzielen (Gittler & Arendasy, 2003). Uneinigkeit herrscht immer noch in der Frage, welche und wie viele Faktoren benötigt werden, um diesen Fähigkeitsbereich angemessen abzudecken (Gittler & Arendasy, 2003).

Zum Teil kann dieses Problem daraus entstanden sein, dass Forscher unabhängig voneinander faktoren-analytisch raumvorstellungsbeschreibende Faktoren extrahierten; die Inhaltsbeschreibungen und ihre Benennungen sind aber nicht immer einheitlich (Gittler & Arendasy, 2003).

### 5.1 Was ist Raumvorstellung

Beginnend mit Untersuchungen um 1920 (McGee, 1979), haben sich unzählige Studien damit beschäftigt, dass es unter den die Intelligenz bestimmenden Faktoren etwas wie „räumliche Fähigkeiten“ geben könnte.

Mit dem Jahr 1931 schien der Durchbruch geschafft, indem es Thurstone (1931, nach Amelang & Bartussek, 2001) gelang, mittels multipler Faktorenanalyse Raumvorstellungsfähigkeit als einen der sieben Primärfaktoren der Intelligenz zu isolieren.

Die Subfaktoren des Faktors „Raumvorstellung“ bezeichnete er als:

- *spatial relations* (räumliche Beziehungen): dies beschreibt die Fähigkeit, sich Beziehungen zwischen Figuren und Teilen daraus vorzustellen.
- *visualization* (Veranschaulichung): stellt die Fähigkeit, ein Objekt in der Vorstellung zu manipulieren, zu rotieren oder zu drehen dar.
- *spatial orientation* (räumliche Orientierung): diesen Faktor fügte er seinen Erkenntnissen erst 7 Jahre später hinzu (Thurstone, 1938). Damit kann die Fähigkeit

bezeichnet werden, sich vorzustellen wie ein Objekt aus einer anderen Perspektive aussieht oder auch die Fähigkeit, sich selbst im Raum zu orientieren.

Seit Thurstone haben sich unzählige Autoren in einer Neudefinition räumlicher Fähigkeiten versucht, doch die Ergebnisse von Untersuchungen zur räumlichen Vorstellung sind auch heute noch sehr kontrovers und das Ziel einer einheitlichen Definition scheint immer noch in weiter Ferne. Die Schwierigkeit, uniforme Ergebnisse zu finden, besteht unter anderem in der großen Zahl verschiedenster Tests zu ihrer Erfassung und der auf Grund dessen mangelnden Replizierbarkeit der Faktoren in unterschiedlichen Testverfahren (Voyer, Voyer & Bryden, 1995). Voyer et al. (1995) vertreten in der Literatur eine etwas extreme Position, indem sie diese Schwierigkeit auf die Annahme zurückführen, jeder Test stelle eine eigene Definition eines spezifischen Subfaktors der Raumvorstellung dar.

Neben der Vielfalt vorhandener Definitionen besteht ein weiteres Problem in der großen Zahl verschiedener Aufgabentypen, welche Raumvorstellung erfassen sollen und dass zu ihrer Bearbeitung unterschiedliche Fähigkeiten nötig sind, die mehr oder weniger alle etwas mit Raumvorstellung zu tun haben.

Auch wenn in der Literatur immer wieder anders definiert wird, was genau unter Raumvorstellung zu verstehen ist, so herrscht mittlerweile unter den Wissenschaftlern zumindest ein breiter Konsens darüber, dass es sich bei dem Terminus „Raumvorstellung“ nicht um ein einheitliches Konstrukt handelt.

Linn und Petersen (1985) z. B. definieren räumliche Fähigkeiten als „...skill in representing, transforming, generating, and recalling symbolic, nonlinguistic information.“ (p.1482) und trafen eine bedeutende Unterteilung des „Raumfaktors“ in 3 Kategorien:

- *spatial perception* (räumliche Wahrnehmung): darunter fassen sie die Fähigkeit zusammen, räumliche Beziehungen zwischen Objekten herzustellen, unter Berücksichtigung der Ausrichtung des eigenen Körpers zum Objekt.
- *Spatial visualization* (räumliche Veranschaulichung): beschreibt schwierige Manipulationen räumlich dargebotener Information durch mehrere Schritte und

- *mental rotation* (mentale Rotation): diese wird als schnelle und korrekte Drehung zwei- oder dreidimensionaler Objekte in der Vorstellung beschrieben.

Innerhalb dieser Trennung stellt jedoch *spatial visualization* einen Sonderfall dar, da diese Fähigkeit zwar eindeutig der Raumvorstellung zugeordnet werden kann, hierfür aber nicht ausschließlich räumliche Strategien sondern auch andere analytische Prozesse zur Lösung herangezogen werden können (Linn & Petersen, 1985).

Halpern (2000) erweitert die vorausgegangene Definition noch um zwei zusätzliche qualitativ unterschiedliche, visuell-räumliche Fähigkeiten:

- *spatiotemporal ability* (räumlich-zeitliche Fähigkeit): beschreibt die Beurteilung von, und die Reaktion auf dynamische (z. B. sich bewegende) visuelle Reize.
- *generation and maintenance of a spatial image* (Erstellung und Behalten räumlicher Bilder): beschreibt die Fähigkeit, ein Bild aus dem Gedächtnis zu erstellen, diese Information zu behalten und danach weiter zu verwenden.

Ein umfassendes und neueres Modell kognitiver Fähigkeiten bietet das Cattell-Horn-Carroll Modell (CHC-Modell) (McGrew, 2005). Dieses stellt gewissermaßen eine Symbiose der Theorien zur fluiden und kristallinen Intelligenz von Horn und Cattell (Horn, 1991, nach McGrew, 2005) und der „drei Schichten Theorie“ von Carroll (Carroll, 1993, nach McGrew, 2005) dar.

Das CHC-Modell ist ein hierarchisches Intelligenzstrukturmodell, welches die kognitiven Fähigkeiten in drei Schichten einteilt. Die dritte und oberste Schicht stellt die generelle Intelligenz dar (wobei das CHC-Modell einen generellen Faktor „g“ der Intelligenz - im Gegensatz zu Carroll und in Übereinstimmung mit Horn und Cattell - nicht für gültig ansieht) (McGrew, 2005). Die zweite Schicht beinhaltet breite (engl. narrow) kognitive Fähigkeiten wie die fluide und kristalline Intelligenz, Allgemeinwissen, visuell räumliche Fähigkeiten, akustische Verarbeitung, Kurzzeitgedächtnis, Speichern und Abrufen aus dem Langzeitgedächtnis, kognitive Verarbeitungsgeschwindigkeit, Reaktionsgeschwindigkeit, Lesen und Schreiben, psychomotorische Geschwindigkeit, sowie motorische, kinästhetische, olfaktorische und taktile Fähigkeiten (McGrew, 2005).

Die erste und unterste Schicht beinhaltet etwa 100 enge (engl. small) kognitive Fähigkeiten, wobei hier nur jene Fähigkeiten der ersten Schicht erwähnt sein sollen, die dem räumlich-visuellen Bereich zugeteilt sind. Darunter befinden sich auch hier:

- *Visualization*: Dies beinhaltet die Fähigkeit, sich ein Objekt vorzustellen, es zu manipulieren oder transformieren und vorherzusehen, wie es unter veränderten Bedingungen aussieht.

und

- *Spatial relations*: McGrew (2005) bezeichnet sie als die "Ability to rapidly perceive and manipulate (mental rotation, transformations, reflection, etc.) visual patterns or to maintain orientation with respect to objects in space" (McGrew, 2005, p.152).

sowie

- *Closure speed*: Verlangt die Fähigkeit, die Repräsentation eines bekannten Objekts schnell aus dem Langzeitgedächtnis abzurufen, um Teile visueller Stimuli zu einem Ganzen kombinieren zu können.
- *Flexibility of closure*: "Ability to identify a visual figure or pattern embedded in a complex distracting or disguised visual pattern or array, when knowing in advance what the pattern is" (McGrew, 2005, p.152). Hierbei wird die Fähigkeit gefordert, einen Stimulus im Arbeitsgedächtnis zu behalten, solange man versucht, diesen aus einem komplexen Bild herauszufinden.
- *Visual memory*: beschreibt die Fähigkeit, ein visuelles Objekt über kurze Zeit im Gedächtnis zu behalten und später wieder abzurufen.
- *Spatial scanning*: bezieht sich auf die Geschwindigkeit der Analyse eines komplexen Bereichs und das Identifizieren eines spezifischen Stimulus.
- *Serial perceptual integration*: beschreibt die Fähigkeit, einen Gegenstand zu identifizieren, dessen Teile schnell und hintereinander gezeigt werden.
- *Length estimation*: bezieht sich auf die Fähigkeit des Schätzens und Vergleichens von Längen ohne technische Hilfsmittel.
- *Perceptual illusions*: Tendenz, optischen Täuschungen, welche geometrische Figuren enthalten, zu widerstehen.

- *Perceptual alternations*: bezeichnet nach McGrew (2005) die “Consistency in the rate of alternating between different visual perceptions” (McGrew, 2005, p. 153).
- *Imagery*: beschreibt die Fähigkeit, abstrakte räumliche Formen zu manipulieren.

In den letzten Jahren entstanden zahlreiche Untersuchungen kognitiver Fähigkeiten auf Basis des CHC-Modells (vgl., McGrew, 2009). Es scheint, dass diese Aufteilung einen weiteren Schritt in Richtung einer einheitlicheren Auffassung der Intelligenzfaktoren und ihrer psychometrischen Erfassung ermöglicht, wobei McGrew (2009) betont, der derzeitige Entwicklungsstandpunkt des Modells stelle keine endgültige Faktorenaufteilung dar, sondern biete einen erweiterbaren und zu vereinheitlichenden Ausgangspunkt zur Erreichung kongruenter Ergebnisse.

Diese Arbeit beschränkt sich auf die Anführung dieser, nur wenigen der überaus zahlreichen, Definitionsansätze räumlichen Vorstellungsvermögens, da eine ausführliche Auflistung verschiedener Definitionen nicht sinngemäß erscheint und dies an anderer Stelle (z. B. Halpern, 2000; McGee, 1979; Voyer et al., 1995) bereits sehr umfangreich dargestellt wurde.

Abschließend sei noch gesagt, dass zum heutigen Zeitpunkt die meisten Raumvorstellungsmodelle davon ausgehen, dass mindestens zwei getrennte räumliche Faktoren der Intelligenz existieren (McGee, 1979). Dabei handelt es sich um:

- *räumliche Visualisierung*, welche die Fähigkeit beschreibt, zwei- oder dreidimensionale bildhafte Stimuli zu manipulieren oder zu rotieren (dem Leser wird hier die Ähnlichkeit zu der Definition mentaler Rotation von Linn und Petersen (1985), auffallen). Eine weitere Definition bezeichnet diese Fähigkeit als mentale Manipulation und Integration von Stimuli, die aus mehr als einem Teil oder aus beweglichen Teilen bestehen (Olkun, 2003).
- *räumliche Orientierung* betrifft das Erfassen der Anordnung visueller Stimuli, sowie die Fähigkeit, sich durch eine veränderte Sichtweise auf ein Objekt nicht verwirren zu lassen und räumliche Beziehungen bestimmen zu können, bei denen die Position des Betrachters einen wesentlichen Teil des Problems darstellt.

## 5.2 Raumvorstellung und der Unterschied der Geschlechter

Die Literatur zu geschlechtsspezifischen Unterschieden auf dem Gebiet der räumlichen Vorstellungsfähigkeit ist immens und die Ergebnisse sind auch hier bei weitem nicht einheitlich. Halpern (2000) führt als mögliche Gründe dafür unter anderem die Verwendung unterschiedlicher Erhebungsinstrumente und auch die Unterschiede in der Definition dessen, was „räumliche Fähigkeit“ nun wirklich ist, an.

Heute weiß man, dass nicht ein Geschlecht allgemein intelligenter ist als das andere, sondern dass Unterschiede nur in bestimmten Bereichen der Intelligenz auftreten. Pauschalisierend kann nach Voyer et al. (1995), angeführt werden, dass Frauen über bessere Fähigkeiten in Bereichen der Sprache, Feinmotorik und Wahrnehmung verfügen, während Männer bezüglich räumlicher Fähigkeiten und mathematischem Schlussfolgern bessere Leistungen erbringen. Unter genauerer Betrachtung scheint diese Aufteilung jedoch als zu ungenau. Während Männer in vielfältigen Aufgaben im Bereich der Raumvorstellung bessere Leistungen erzielen als Frauen, so gilt dies dennoch nicht für den gesamten Bereich der räumlich-kognitiven Fähigkeiten; durchschnittlich erzielen beispielsweise Frauen in bestimmten Aufgaben, die „Objektlokalisierung“ erfordern, bessere Leistungen als Männer (Kimura, 1999). Auch die weit verbreitete Meinung, Frauen verfügten über bessere verbale Fähigkeiten, ist unter dieser allgemeinen Aussage nicht haltbar. Das weibliche Geschlecht ist besonders in Aufgaben, die das Gedächtnis für sprachliches Material fordern besser und in Aufgaben, die „Wortflüssigkeit“ verlangen (Kimura, 1999). Ebenso haben Männer nicht generell bessere mathematische Fähigkeiten; während Männer durchschnittlich über bessere Leistungen im Bereich des „logischen Schlussfolgerns“ verfügen, schneiden Frauen bei Aufgaben die allgemeines Rechnen erfordern, besser ab (Kimura, 1999).

Im Folgenden soll besonders auf den Bereich der besseren Leistungen des männlichen Geschlechts im Bereich der „mentalen Rotation“ eingegangen werden.

Geschlechtsunterschiede zu Gunsten der Männer in räumlichen Fähigkeiten (bis zu einer Standardabweichung über den Leistungen der Frauen) wurden bis jetzt in der Literatur sehr häufig bestätigt (Bischof-Köhler, 2006; Cooke-Simpson & Voyer, 2007; Halpern, 2000; Linn & Petersen, 1985; Maccoby & Jacklin, 1974; Voyer et al., 1995) und die Annahme eines



männlichen Vorteils in Raumvorstellungsaufgaben zählt zu einem der meist untersuchten Phänomene in der Literatur der Geschlechtsdifferenzen.

Eine der bekanntesten Arbeiten zu diesem Thema stammt von Maccoby und Jacklin (1974). Obwohl ihre Meta-Analyse über mehr als 1400 Studien auf Grund methodischer Probleme oft kritisiert wurde, ist sie dennoch eine der wichtigsten Arbeiten zu geschlechtsspezifischen Unterschieden geblieben. Den beiden Autorinnen ist es gelungen, drei kognitive Fähigkeiten und einen Persönlichkeitsfaktor aufzuzeigen, in denen Geschlechtsunterschiede tatsächlich vorhanden sind, wobei Frauen über stärker ausgeprägte verbale Fähigkeiten verfügen, während Männer im mathematischen Denken und in visuell-räumlichen Fähigkeiten überlegen scheinen. Der einzig auffindbare Unterschied zwischen den Geschlechtern im Bereich der Persönlichkeitsvariablen liegt den Autorinnen zu Folge in der Aggression<sup>4</sup>, bei der das männliche Geschlecht deutlich höhere Werte erreicht.

Auch Linn und Petersen (1985) konnten in den oben erwähnten Kategorien der Raumvorstellung verschieden starke Unterschiede zwischen den Geschlechtern feststellen, wobei homogene Unterschiede zu Gunsten der Männer nur in den Sparten der *räumlichen Wahrnehmung* (sofern die Probanden über 18 Jahre alt waren) und in *mentaler Rotation* gefunden wurden.

Aufgaben zur mentalen Rotation wiesen die größten Differenzen auf; diese reichten von einer viertel Standardabweichung bis zu fast einer ganzen Standardabweichung, mit einer durchschnittlichen Effektgröße von  $d=.73$  ( $p>.05$ ) (Linn & Petersen, 1985) beziehungsweise einer durchschnittlichen, gewichteten Effektgröße von  $d=.56$  (Voyer et al., 1995). Aufgaben zur räumlichen Wahrnehmung oder Visualisierung zeigten hingegen keine oder nur sehr geringe Unterschiede (Geiser, Lehman & Eid, 2006).

Jener Test, der die größten Unterschiede zu Tage bringt, ist der *Mental Rotation Test* (MRT) von Vandenberg und Kuse, 1978 (Cooke-Simpson & Voyer, 2007; Geiser et al., 2006; Goldstein, Haldane & Mitchell 1990; Hirnstein, Bayer & Hausmann, in press; Linn & Petersen, 1985).

---

<sup>4</sup> Dieses Ergebnis muss aus heutiger Sicht jedoch differenzierter betrachtet werden. Zahlreiche Studien beschäftigten sich unterschiedlichen Formen der Aggression; während Männer bei physischer Aggression größere Ausprägungen zeigten ( $d=.74$ ), überwiegt bei Frauen die „Beziehungsaggression“ (Asendorpf, 2007).

Der MRT ist ein häufig angewandter paper-pencil-Test zur Messung räumlicher Fähigkeiten (genauer gesagt, der mentalen Rotation<sup>5</sup>). Seine Beliebtheit ist vermutlich auf die Tatsache zurückzuführen, dass durch seine Verwendung die größten Geschlechtsdifferenzen erhoben werden können, was ihn für die Untersuchung dieser zu großer Bedeutung erhebt (Cooke-Simpson & Voyer, 2007).

Mögliche Gründe für die unterschiedlichen Leistungen der Geschlechter in diesem Aufgabengebiet werden in der Literatur zahlreich angeführt. Unter anderem finden sich darunter Studien zu

- der Verwendung unterschiedlicher Bearbeitungsstrategien der Geschlechter (Hirnstein et al., in press)
- verschiedenem Bearbeitungstempo der Aufgaben (Geiser et al., 2006; Goldstein et al., 1990; Maccoby & Jacklin, 1974)
- Einfluss von Faktoren wie Erfahrung mit räumlichen Aufgaben (Casey, Nuttall & Pezaris, 1997), Zeitbeschränkung oder Aufgabenschwierigkeit (Goldstein et al., 1990)

Als Erklärung für die Leistungsunterschiede der Geschlechter in räumlichen Fähigkeiten können die im Folgenden angeführten, die Raumvorstellung beeinflussenden Faktoren angesehen werden.

### **5.3 Einflussgrößen auf das räumliche Vorstellungsvermögen**

Faktoren, die das räumliche Vorstellungsvermögen beeinflussen, sind sehr facettenreich. Zur Identifikation dieser leistet die „Kulturvergleichende Psychologie“ einen wesentlichen Beitrag.

Es wird oft postuliert, dass Menschen verschiedener kultureller Hintergründe Dinge unterschiedlich wahrnehmen und verschieden denken. Dieser Ansicht zu Folge entwickeln sich kognitive Strukturen und Prozesse (wie etwa auch die Raumvorstellungsfähigkeit) in Abhängigkeit verschiedener Erfahrungs- und Sozialisationsprozesse in unterschiedlichen Kulturen anders (Montello, 1995).

---

<sup>5</sup> Als mentale Rotation wird, wie bereits erwähnt, der kognitive Prozess bezeichnet, in der Vorstellung ein Objekt zu rotieren (Shepard & Metzler, 1971) oder auch zwei- oder dreidimensionale Objekte in der Vorstellung schnell und richtig zu rotieren (Linn & Petersen, 1985).

Untersuchungen kultureller Unterschiede in raumvorstellungsbezogenen Konstrukten sind nicht immer einheitlich; während manche Aspekte räumlicher Fähigkeiten universell in verschiedenen Kulturen gefunden werden können, scheinen andere diese Generalisierbarkeit nicht aufzuzeigen (Montello, 1995).

Im Folgenden werden verschiedene biologische und soziokulturelle Faktoren angeführt, welche die Raumvorstellungsfähigkeit beeinflussen können. Die Anlage - Umwelt Kontroverse kann dabei als Hintergrund für Erklärungsansätze individueller Unterschiede verstanden werden (Pervin, Cervone & John, 2005). Sie zeigt das Kontinuum zwischen den Vertretern einer biologischen Determination menschlichen Verhaltens und jenen, die das soziale Umfeld mit seinen spezifischen Erfahrungen und Erwartungen als ausschlaggebend für Unterschiede zwischen Menschen ansehen, auf.

### **5.3.1 Biologische Erklärungsansätze**

Die Vertreter eines biologischen Standpunktes schreiben der Vererbbarkeit und der Wirkung von Hormonen auf die räumliche Vorstellung eine wesentliche Rolle zu. Bis heute ist jedoch nicht geklärt, wie viel tatsächlich durch biologische Determination erklärt werden kann; sicher ist nur, dass sie bis zu einem gewissen Grad nachweisbar ist (McGee, 1979).

Biologische Theorien gehen davon aus, dass Unterschiede im räumlichen Vorstellungsvermögen in allen Kulturen existieren; Daten aus kulturvergleichenden Untersuchungen wird in diesem Zusammenhang großer Wert beigemessen, da sie biologische Theorien stützen können und der Generalisierbarkeit der Ergebnisse dienlich sind (Feingold, 1994).

#### **5.3.1.1 Vererbbarkeit**

Unter den Theorien über die Vererbbarkeit der Raumvorstellungsfähigkeit ist die bekannteste jene des „X-gebundenen, rezessiven Gens“ (z. B. Bock & Kolakowski, 1973).

Diese Theorie geht davon aus, dass die Fähigkeit einer guten Raumvorstellung von einem rezessiven Allel auf dem X-Chromosom getragen und durch dieses an die folgende Generation vererbt wird. Demnach würde eine Person nur dann eine gute räumliche Vorstellung aufweisen, wenn alle X-Chromosomen das rezessive Allel, welches die Fähigkeit

trägt, besitzen. Als Folge hätten Mädchen (XX) auf Grund der zwei X-Chromosomen eine geringe Chance, zwei rezessive Gene zu erhalten, was allerdings Voraussetzung für die Manifestierung wäre. Buben (XY) hingegen benötigen nur eines, da sie nur ein X-Chromosom besitzen und das Gen nicht durch ein anderes dominantes X-Chromosom unterdrückt werden könnte. Einer Studie von Bock und Kolakowski (1973) zu Folge, sollte dieses rezessive Gen bei 50% der Männer und 25% der Frauen vorhanden sein.

Obwohl die Theorie des „X-gebundenen, rezessiven Gens“ nach einer viel versprechenden Erklärung für die höhere männliche Fähigkeit der Raumvorstellung klingt, konnte Boles bereits 1980 feststellen, dass die von Bock et al. (1973) und anderen Wissenschaftern gelieferten Ergebnisse nicht in dieser Form reproduzierbar sind (für einen Vergleich siehe Boles, 1980). Ein erneuter Versuch, diese Theorie zu bestätigen, wurde 1994 von Gittler und Vitouch vorgenommen. Die Theorie des „X-gebundenen, rezessiven Gens“ konnte jedoch in ihrer Studie, welche anhand eines Rasch-homogenen Raumvorstellungstests durchgeführt worden war, nicht bestätigt werden.

Für den Einfluss der Vererbbarkeit auf räumliche Fähigkeiten sprechen jedoch unter anderem auch die Ergebnisse von Vandenberg (1968, nach Maccoby & Jacklin, 1974), der anhand von Zwillingsstudien<sup>6</sup> feststellte, dass sowohl räumliche als auch verbale Fähigkeiten stark mit Vererbung zusammenhängen.

Die hohe Vererbbarkeit und auch die unterschiedliche Leistungsfähigkeit der Geschlechter in räumlichen Fertigkeiten legen auf jeden Fall einen genetischen Zusammenhang nahe (Maccoby & Jacklin, 1974), auch wenn man die exakten Vorgänge, die dies bewirken, noch nicht in allen Bereichen herausgefunden hat.

### **5.3.1.2 Hormone**

Neben der Vererbbarkeit spielen auch die prä- und perinatalen Hormone ohne Zweifel eine bedeutende Rolle in der Entwicklung des Gehirns und somit, im Sinne einer biologischen

---

<sup>6</sup> Anhand von Zwillingsstudien versucht man, den Einfluss der Genetik und der Vererbung bestimmter Merkmale von jenem der Umwelt, in ihrer Bedeutung für die Entwicklung bestimmter Merkmale zu trennen.

Sichtweise, auch bei der späteren Entwicklung geschlechtstypischer Kognitionen und Verhaltensweisen.

Noch im Mutterleib bestimmen die Hormone einen wesentlichen Teil der embryonalen Entwicklung. Bischof-Köhler (2006) ist der Ansicht, dass bereits ab der siebten Schwangerschaftswoche die Geneinwirkung terminiert und alles Folgende von Hormonen bestimmt wird, wobei dem Androgen im Vergleich mit dem Östrogen eine größere Wirkung bei der Bestimmung des Geschlechts und auch bei der Ausbildung von Gehirnstrukturen zugesprochen wird (Bischof-Köhler, 2006).

Auch das weibliche Sexualhormon Östrogen hat jedoch große Bedeutung für die Entwicklung von Geschlechtsunterschieden.

Nyborg (1990) entdeckte durch die Untersuchung von Patientinnen, die am Turner-Syndrom<sup>7</sup> leiden, dass für gewisse räumliche Fähigkeiten ein optimales Level an Estradiol (einer Form von Östrogen) notwendig ist. Er vermutet, dass gesunde Frauen dieses Hormon in zu hohem Maße produzieren, wobei eine Verringerung zu einer verbesserten Raumvorstellung führen könnte. Anhand der kleinen Stichprobe, die von ihm verwendet wurde, lässt sich diese Annahme jedoch bis heute nicht verallgemeinern.

### **5.3.1.3 Unterschiede in der Gehirnstruktur**

Nachfolgend wird ein Bereich angeführt, der eine unterschiedliche strukturelle Organisation des menschlichen Gehirns nahe legt und als Ursache für Differenzen im Raumvorstellungsvermögen der Menschen herangezogen werden kann (McGee, 1979).

#### **5.3.1.3.1 Lateralisierung**

Dieser Begriff bezeichnet die unterschiedliche Verteilung bestimmter Funktionen auf die beiden Gehirnhemisphären beziehungsweise die Spezialisierung dieser für bestimmte Aufgaben. Zahlreiche Studien zur Aktivierung verschiedener Hirnregionen während der

---

<sup>7</sup> eine genetische Anomalie, bei der nur ein X- Chromosom vorhanden ist, oder zumindest das zweite stark defizitär und die Östrogenproduktion stark unzureichend ist (für genauere Informationen siehe z. B. Sybert & McCauley, 2004).

Bearbeitung diverser Aufgabentypen (sprachlich vs. räumlich) zeigen, dass die linke Gehirnhälfte stärker mit Fähigkeiten des schlussfolgernden Denkens und analytischen Fähigkeiten sowie verbalen Leistungen (Sprachwahrnehmung und Sprachproduktion) in Verbindung gebracht wird, während die rechte Hemisphäre für räumliche Fähigkeiten und ganzheitliches Erfassen, sowie Emotionen verantwortlich ist (Clements et al., 2006; Gur et al., 2000; Halpern, 2000; Maccoby & Jacklin, 1974).

Auf Grund der Leistungsdifferenzen der Geschlechter bei diesen Aufgaben, begann man sich in der Forschung für mögliche Unterschiede in der Lateralisierung zerebraler Aktivitäten zu interessieren und stellte, je nachdem ob die zu bearbeitende Aufgabe sprachliche oder räumliche Fähigkeiten erfordert, Unterschiede zwischen den Geschlechtern fest (Gur et al., 2000).

So scheint es, dass Männer bei sprachlichen Aufgaben stärker links lateralisiert und bei Aufgaben, die räumlich-visuelle Fähigkeiten fordern, eher bilateral<sup>8</sup> organisiert sind. Bei Frauen hingegen verhält es sich genau umgekehrt; sie zeigen eine stärkere bilaterale Aktivierung bei sprachlichen, und eine Lateralisierung der rechten Hemisphäre bei visuell-räumlichen Aufgaben (Clements et al., 2006; Gur et al., 2000). Eine Hypothese (vgl. Gur et al., 2000) besagt, dass es genau diese bilaterale Aktivierung sei, die für die besseren Leistungen des jeweiligen Geschlechts in den verschiedenen Aufgabentypen verantwortlich sein könnte.

Andere Studien ergaben im Gegensatz dazu eine stärkere Involvierung der rechten Hemisphäre während der Ausführung räumlicher Aufgaben und, dass diese (da bei Frauen nicht auftretende) für die unterschiedlichen Leistungen verantwortlich sein könnte (Gur et al., 2006). Diese Ergebnisse brachten Gur et al. (2006) zu der Annahme, dass eine optimale Leistung die einseitige Aktivierung primärer Regionen (linkshemisphärisch für Sprache und rechtshemisphärisch für räumliche Fähigkeiten) und die bilaterale Aktivierung assoziierter Regionen verlangt. Diese Hypothese muss allerdings noch genaueren Untersuchungen unterzogen werden (Gur et al., 2006).

---

<sup>8</sup> Bilaterale Aktivierung bedeutet, dass beide Hemisphären die interessierende Aufgabe lösen können (Clements et al., 2006), während Lateralisierung die Spezialisierung einer Hemisphäre für eine spezifische Aufgabe meint.

### 5.3.2 Exkurs: Händigkeit

Auch zu der Frage nach dem Zusammenhang von Handpräferenz und kognitiven Fähigkeiten haben sich mannigfaltige Theorien entwickelt und auch hier kann, ähnlich der Frage nach den Geschlechtsdifferenzen im räumlichen Vorstellungsvermögen, die Lateralisierung als mögliche Erklärung herangezogen werden.

Spekulationen über diesen Zusammenhang stützen sich mitunter auf die Begründung, dass die rechte Hemisphäre stärker für die Verarbeitung räumlicher Aufgaben verantwortlich ist als die linke und dass sich Rechts- und Linkshänder (ebenso wie Männer und Frauen) in ihrer zerebralen Organisation zu unterscheiden scheinen (Peters, Reimers & Manning, 2006).

Die Annahme, dass bei Rechtshändern die linke Gehirnhälfte dominant ist und vice versa (Halpern, 2000), erschließt sich aus der Tatsache, dass die rechte Hemisphäre Bewegungen der linken Körperhälfte koordiniert und umgekehrt die linke Hemisphäre für Bewegungen der rechten Körperseite verantwortlich ist. Diese Dominanz bietet wiederum Implikationen für den Einfluss der Händigkeit auf räumliche Fähigkeiten.

Untersuchungen zu Unterschieden in der Organisation von Gehirnstrukturen zwischen Links- und Rechtshändern konnten zeigen, dass Linkshänder in sprachlichen (Bourne, 2008; Levy & Reid, 1978) sowie in räumlichen Aufgaben (Brodie & Dunn, 2005) weniger stark lateralisiert sind und über geringere räumliche Fähigkeiten verfügen als Rechtshänder (McGee, 1978). Für diese Theorie spricht auch die Annahme von Levy (1976, nach Halpern, 2000), die postuliert, dass Frauen in ihren kognitiven Fähigkeiten jenen linkshändiger Männer ähneln, da beide Gruppen zu einer bilateralen Hemisphärenorganisation tendieren.

Widersprochen wird dieser Annahme durch andere Studien (Annett, 2002, nach Peters et al., 2006), die von tendenziell besseren räumlichen Fähigkeiten in Linkshändern und besseren verbalen Fähigkeiten in Rechtshändern ausgehen.

Peters et al. (2006) fordern in ihrem Artikel dazu auf, die kontrastierenden Ergebnisse in der Literatur zu diesem Thema vor dem Hintergrund zu sehen, dass Händigkeit nur weniger als 1% der Varianz in Leistungen mentaler Rotation zu erklären vermag. Ebenso machen sie darauf aufmerksam, dass auf Grund der Vielfalt der in derartigen Untersuchungen angewandten Methoden zur Erhebung von Händigkeit und der darin angewandten

verschiedenen Aufgaben zur Messung räumlicher Fähigkeiten, eine einfache und eindeutige Antwort auf diese Frage derzeit unmöglich ist (Peters et al., 2006).

Auf Grund ihrer Bedeutsamkeit im kulturvergleichenden Sinn soll hier dennoch kurz die Studie von Peters et al. (2006) erwähnt sein. Die Autoren verglichen sieben ethnische Gruppen verschiedener Nationen hinsichtlich ihrer Handpräferenz und konnten bei vier der sieben Gruppen unter Beachtung mehrerer Faktoren deutliche Unterschiede feststellen. In Hinblick auf die mentale Rotationsfähigkeit (erhoben anhand des MRT) zeigten sich unterschiedliche Leistungen in Abhängigkeit der Stärke der Handpräferenz (gemessen anhand einer fünfstufigen Skala von „links“ bis „rechts“). Dabei zeigten „beidhändige“ Personen signifikant ( $p < .0001$ ) schlechtere Ergebnisse. Wie oben bereits erwähnt, sind diese Resultate allerdings mit gebotener Vorsicht zu interpretieren.

Einen unter vielen, möglichen Ansätzen zur Entstehung von Linkshändigkeit bietet die Theorie von Geschwind und Galaburda (1987). Sie besagt, dass ein hoher Einfluss pränatalen Testosterons das neuronale Wachstum der linken Gehirnhälfte verlangsamen kann, was zu einer Dominanz der rechten Hemisphäre führt und eine mögliche Erklärung für die Ursachen von Linkshändigkeit darstellt. Kongruent zu dieser Annahme existiert eine Vielzahl von Studien (Gladue & Bailey, 1995), die einen größeren Anteil von Linkshändern in der männlichen Population nachweisen konnten.

Ein größerer Überblick über die verschiedenen kontrastierenden Ergebnisse der Studien zu diesem Thema kann auf Grund deren Umfangs an dieser Stelle nicht gegeben werden. Der interessierte Leser sei z. B. auf Peters et al. (2006) verwiesen.

### **5.3.3 Die Bedeutung der Umwelt in kulturvergleichenden Untersuchungen**

Wie bereits erwähnt, gibt es immer noch keinen Konsens bezüglich der Anlage-Umwelt-Kontroverse. Auch wenn sich eine Tendenz zu einer ganzheitlicheren Sichtweise in den letzten Jahren immer mehr herauszukristallisieren beginnt, vertreten die Anhänger des Umwelt-Ansatzes dennoch den Standpunkt, dass sich Differenzen zwischen den Geschlechtern durch das soziale Umfeld der Menschen und nicht durch biologische Determination entwickeln.



Besonders in kulturvergleichenden Untersuchungen ist es unumgänglich, dem soziokulturellen Hintergrund besondere Bedeutung zuzumessen (Friedlmeier, 1994).

Kultur kann, wie in Kapitel 3.2 bereits erwähnt, als Zusammenspiel vielfältiger Variablen gesehen werden, welches sich in der wechselseitigen Beeinflussung von Personen und deren Handeln widerspiegelt (Trommsdorff, 1993).

Ein Bereich des Lebens, der besonders vom soziokulturellen Hintergrund eines Individuums beeinflusst wird, ist jener der Sozialisation. Diese „...lässt sich auch als einen Prozess verstehen, bei dem eine Person Werthaltungen entwickelt, die in dem jeweiligen kulturellen Wertesystem ihren Ausgangspunkt nehmen“ (Friedlmeier, 1994, S.44).

Bei der Sozialisation kommt vor allem den Erziehungspersonen eine wesentliche Rolle zu, welche Werthaltungen und Deutungsrahmen an die Kinder weitergeben.

Ein Beispiel des Zusammenhangs kultureller Werthaltungen und Erziehungsstilen bietet eine durch ihren kulturvergleichenden Charakter für die Sozialisationsrolle der Erwachsenen bedeutsame Untersuchung von Friedlmeier (1994). Diese Studie zielte darauf ab, die Bedeutung subjektiver Erziehungstheorien auf den Sozialisationsprozess anhand eines Vergleichs von Brasilien und Deutschland aufzuzeigen. Die Ergebnisse zeigen, dass, während die Erziehungsvorstellungen der Brasilianer eher eine kollektivistische Werthaltung vermitteln, von Erziehern in Deutschland verstärkt Individualismus angestrebt wurde.

Neben dem Erziehungsstil stellt das Geschlecht in fast allen Kulturen eine wesentliche Einflussgröße auf die Sozialisierungsprozesse dar (Kienbaum, 1995). Die meisten theoretischen Ansätze dazu, dass geschlechtstypisches Verhalten erlernt werde, spiegeln sich im Feld der *Lerntheorie* wieder. Hier sei die *Imitationstheorie* erwähnt, in der davon ausgegangen wird, dass geschlechtsspezifisches Verhalten anhand von Modellen und deren Imitation gelernt wird (Asendorpf, 2007).

Sie geht von drei Nachahmungsszenarien aus (Trautner, 1979, nach Asendorpf, 2007). Zum einen kann Imitation durch „differentielle Beobachtung Gelegenheiten“ entstehen, zu denen die Kinder laut Theorie öfter bei gleichgeschlechtlichen als bei gegengeschlechtlichen Modellen Gelegenheit haben. Zum anderen wird postuliert, dass eine „selektive Imitation“ dahingehend stattfindet, dass Kinder bei der Beobachtung beider Modelle (des gleichen und des anderen Geschlechts) eher das gleichgeschlechtliche Modell imitieren und als drittes

Szenario wird die „Elternidentifikation“ mit jenem Elternteil postuliert, der das gleiche Geschlecht innehat wie das Kind (Trautner, 1979, nach Asendorpf, 2007).

Asendorpf (2007) zeigt jedoch auf, dass diese Theorien keineswegs bewiesen sind, da in westlichen Gesellschaften, in denen Kinder die Gelegenheit haben beide Geschlechter gleich häufig zu beobachten, nicht immer bevorzugt Modelle des eigenen Geschlechts imitiert werden und dass Kinder, wenn sie größer werden, nicht unbedingt dem gleichgeschlechtlichen Elternteil ähnlicher sind als dem anderen (Asendorpf, 2007).

Andere Erkenntnisse, die im Zusammenhang mit unterschiedlichen Raumvorstellungsleistungen in Abhängigkeit der Kultur bedeutungsvoll sind, lieferte eine Studie von Tanzer, Gittler und Sim (1994). In ihrer groß angelegten (n= 1327) Untersuchung der Raumvorstellungsfähigkeit zwischen Jugendlichen aus Singapur und Österreich zeigte sich, dass Mädchen aus Singapur deutlich schlechtere Leistungen erzielten als Buben ihrer Kultur und auch als Mädchen des österreichischen Kulturkreises (als Erhebungsinstrument der Raumvorstellung wurde der dreidimensionale Würfeltest (3DW) von Gittler (1990) verwendet).

Als eine mögliche Erklärung für dieses Ergebnis hypothetisieren die Autoren einen Einfluss der in Singapur gegebenen unterschiedlichen geschlechtsspezifischen Erziehungsstile und dass die weiblichen Testpersonen dieser Kultur keinen Unterricht in Geometrie aufwiesen.

Ihnen zu Folge könnten somit diese Faktoren dazu beitragen, dass Mädchen kein Feedback über ihre aktuelle räumliche Leistungsfähigkeit besitzen und dies in einer unterschiedlichen Selbstwirksamkeitserwartung hinsichtlich der Leistungen der Mädchen resultiere (Tanzer et al., 1994). Dies wiederum könnte als Ursache für das oberflächlichere Testbearbeitungsverhalten gesehen werden, das in dieser Studie bei den Mädchen aus Singapur festgestellt wurde (Mädchen tendierten eher dazu, sich von einer bestimmten Art der möglichen Lösungswürfel irreführen zu lassen. Dieser zeigte drei dem Zielwürfel gleiche Seiten, aber in einer anderen Ausrichtung).

### 5.3.3.1 Exkurs: Darstellende Geometrie

Wie in der eben erwähnten Studie von Tanzer et al. (1994) berichtet, werden im Folgenden Forschungsergebnisse beschrieben, die sich mit einem Einfluss von darstellender Geometrie auf die Raumvorstellungsfähigkeit beschäftigen.

In den letzten Jahren zeigte sich in der Literatur ein steigendes Interesse an der Frage, ob durch Erfahrungen mit beziehungsweise Unterricht in „darstellender Geometrie“ die räumliche Vorstellungsfähigkeit verbessert werden kann. Die darstellende Geometrie (kurz: DG und auch unter dem Namen „Konstruktive Geometrie“ bekannt) „...soll geometrische Formen und Vorgänge verstehen, vorstellen, gestalten und zeichnen lehren“ (Hohenberg, 1966, S. III aus Vorwort zur 1. Auflage).

Das Hauptziel des Geometrieunterrichtes kann im Allgemeinen darin gesehen werden, räumliche Fähigkeiten von Schülern zu trainieren und zu verbessern, wobei die Schüler Verhältnisse, Eigenschaften und Beziehungen zwei- und dreidimensionaler Figuren analysieren.

Eine generelle Tendenz der Ergebnisse deutet daraufhin, dass tatsächlich eine Verbesserung der Raumvorstellung anhand diverser Trainingsmaßnahmen erreicht werden kann. Untersuchungen zu diesem Thema werden häufig mit einem einfachen Prä- Posttestdesign angelegt und prüfen oftmals nur eine kurzfristige Verbesserung der spezifischen Fähigkeit, die mit dem jeweiligen Test gemessen wird, nicht aber eine Übertragung auf andere Bereiche. Diese Art der Übertragung wird als „trivialer Transfer“ bezeichnet, während der Transfer eines nicht-spezifischen Trainings - wie etwa dem Unterricht in DG - auf die Leistung in einem Raumvorstellungstest als „nicht-trivialer Transfer“<sup>9</sup> bekannt ist (Gittler, 1994).

Hinter Untersuchungen zum Einfluss von DG steht unter anderem die Frage, ob es als Unterrichtsfach signifikante Vorteile für den weiteren Bildungsweg der Kinder bieten kann (Gittler & Glück, 1998) oder auch um zu demonstrieren, dass räumliche Fähigkeiten

---

<sup>9</sup> Transfer bedeutet in diesem Zusammenhang, dass das Gelernte in anderen Kontexten angewandt werden kann (Gittler & Glück, 1998).

keineswegs nur genetisch bedingt sind, sondern, dass sie durch Übung verbessert werden können (Németh, 2007).

Gittler und Glück (1998) konnten anhand einer Langzeitstudie positive Auswirkungen von darstellender Geometrie auf die Bearbeitung eines Raumvorstellungstests - genauer des 3DW (Gittler, 1990) - finden. Die Verbesserung erfolgte in gleichem Ausmaß für beide Geschlechter, wobei sich für die Mädchen ein additiver Effekt von Geschlecht und Versuchsbedingung (Teilnahme am Unterricht der darstellenden Geometrie) zeigte, sodass diese danach gleiche Werte erreichten wie die Buben. Auch Németh (2007) ist es gelungen, diesen positiven Effekt darstellender Geometrie zu zeigen; allerdings profitierte in ihrer Untersuchung das männliche Geschlecht mehr als das weibliche.

Die hier angeführten Studien zeigen, dass räumliche Fähigkeiten durch Erfahrungen mit DG durchaus gesteigert werden können, wohingegen die Frage nach den genauen Wirkmechanismen dahinter bis jetzt noch nicht eindeutig geklärt ist.

Roorda (1994) bietet eine mögliche Erklärung dafür wie DG die Raumvorstellungsleistung verbessern kann, indem er von derselben Problematik im Geometrieunterricht und den meisten Raumvorstellungstests ausgeht: in beiden besteht die Schwierigkeit darin, von Objekten, die in 2D vorgegeben sind, auf den dreidimensionalen Raum umzudenken.

### **5.3.4 Biopsychosoziale Theorien**

Im Vergleich zu biologischen und soziokulturellen Erklärungsansätzen, bieten die biopsychosozialen Theorien eine bessere Herangehensweise an die Anlage-Umwelt-Kontroverse. Diese gehen von einer reziproken Beeinflussung biologischer, sozialer und psychischer Variablen in der Ausbildung menschlichen Verhaltens aus.

Die Grundannahme besteht darin, dass biologische Faktoren die Voraussetzung für kognitive Fähigkeiten wie z. B. Lernen sind, sich das Gehirn aber dennoch in Abhängigkeit sozialer Faktoren wie etwa intellektueller Förderung und dem sozialen Umfeld als Ganzes weiterentwickelt.

Halpern (2000) drückt dies folgendermaßen aus:

„Biology responds to the environment, and people adjust and select their environment to make it compatible with their biological properties“ (p. 16).

Die Sichtweise zu der biologischen Determinierung von individuellen Unterschieden hat sich in den letzten Jahren stark verändert. Es erfolgte die Einsicht, dass Menschen ihr Umfeld ebenso sehr beeinflussen wie vice versa und dass eine klare Trennung zwischen dem Einfluss von Vererbung und Umwelt auf den Menschen nicht mehr durchführbar ist (Halpern, 2000).

Auch bei Maccoby (2000) findet man eine sehr treffende Zusammenfassung dieser Thematik:

“Menschliche Säuglinge sind für das Erlernen der Sprache vorprogrammiert, aber welche Sprache sie lernen, hängt davon ab, in welcher Sprache die Menschen ihrer direkten Umwelt mit ihnen sprechen“ (S. 116).

Abschließend sei noch eine Studie von Casey, Nuttall und Pezaris (1999) erwähnt, die sich mit dem Einfluss biologischer und umweltbedingter Faktoren auf die Raumvorstellungsfähigkeit beschäftigte. Sie postulierten, dass Mädchen von der Existenz eines Bruders, auf Grund des so entstehenden, der Raumvorstellung dienlichen, spielerischen Umfelds, profitieren können. Allerdings reicht die bloße Existenz eines Bruders nicht aus, um als Mädchen bessere Raumvorstellungsleistungen zu erreichen. Die Autoren stellten auch einen Zusammenhang zwischen Händigkeit (des Mädchens und innerhalb der Familie) und dem Grad der Nutzung des vorher erwähnten, die Raumvorstellung fördernden, spielerischen Umfelds ab.

Wie auch in anderen Bereichen der Wissenschaft scheint in der Frage nach der Entstehung und Bedeutung von Geschlechtsunterschieden ebenso eine ganzheitliche, holistische Herangehensweise an dieses Problem vorteilhaft.

### 5.3.5 Einfluss von Alter auf räumliche Fähigkeiten

Piaget postulierte eine kontinuierliche Entwicklung der Raumvorstellung beim Kind, die nicht von Anfang an gegeben ist (vgl. Piaget & Inhelder, 1975).

Zu Beginn (etwa bis zum fünften Lebensmonat) sind die räumlichen Wahrnehmungsschemata des Säuglings noch nicht koordiniert. Dies tritt erst ab einem Alter von ca. 5 Monaten auf und zeigt sich in der Auge-Hand Koordination, durch die ein Kind die Möglichkeit erhält nach einem Gegenstand zu greifen. Durch auf diese Weise erhaltene Informationen beginnt das Kind Begriffe der Konstanz von Form und Größe zu entwickeln. Diese Entwicklung scheint erst ab einem Alter von 12 bis 15 Jahren abgeschlossen zu sein (Piaget & Inhelder, 1975).

Thurstone (1938) sieht die Entwicklung der Raumvorstellung ebenfalls in Abhängigkeit des Alters eines Individuums. Ihm zu Folge erreicht eine Person bis zu ihrem neunten oder zehnten Lebensjahr bis zu fünfzig Prozent ihrer Raumvorstellungsfähigkeit, innerhalb des zwölften bis vierzehnten Lebensjahres gelten bis zu achtzig Prozent als erreicht; Den Leistungshöhepunkt sieht Thurstone (1938) als im frühen Erwachsenenalter abgeschlossen.

Jedoch ist nicht nur die Frage nach dem Höhepunkt der Raumvorstellungsleistung bedeutend sondern auch die Frage danach wann die Fähigkeit zum räumlichen Vorstellen wieder abnimmt.

In diesem Zusammenhang publizierte nun Salthouse (2009) neueste Ergebnisse. In seiner Untersuchung kombinierte er eine Querschnittsstudie (n=2359) mit Wiederholungstestungen nach 1 bis 7 Jahren (n=729) und Wiederholungstestungen nach 1 bis 14 Tagen (n=139).

Seine Stichprobe bestand aus gesunden, gebildeten Personen im Alter zwischen 18 und 60 Jahren, die er in Altersgruppen zwischen 22-27, 28-33, 34-39, 40-45, 46-51, 52-57 teilte.

Untersucht wurden ihre Leistungen anhand 12 verschiedener Variablen, welche folgenden vier kognitiven Bereichen, zugeteilt werden können:

- induktives Schlussfolgern
- räumliche Visualisierung
- Wahrnehmungsgeschwindigkeit und
- episodisches Gedächtnis

Die Ergebnisse der Studie zeigen, dass bei den drei erstgenannten Fähigkeiten (SD~1 bei Wahrnehmungsgeschwindigkeit und räumlicher Visualisierung und SD~.7 für induktives Schlussfolgern) ein Höhepunkt der Leistungen in einem Altersbereich um 22 Jahre erreicht wird, sich ab einem Alter von 27 Jahren diese Leistungen jedoch bereits zu verschlechtern beginnen ( $p < .01$ ). Leistungen des episodischen Gedächtnisses blieben seiner Untersuchung zu Folge bis zu einem Alter von 37 Jahren intakt.

Salthouse (2009) konnte hiermit zeigen, dass Verringerungen kognitiver Fertigkeiten bereits in Personen auftreten, die in einem Altersrahmen zwischen 18 und 60 Jahren liegen und nicht erst wie von anderen Wissenschaftern (vgl. Salthouse, 2009) postuliert nach dem Überschreiten der 60 Jahre (Salthouse, 2009).

Auch zu geschlechtsspezifischen Unterschieden der Verringerung räumlicher Leistungen mit zunehmendem Alter gibt es zahlreiche Untersuchungen, deren Ergebnisse jedoch nicht einheitlich sind (für eine Übersicht siehe Lacreuse & Kim, 2005).

Meinz und Salthouse (1998) führen dazu einige Studien an, darunter jene von Gur et al (1991, nach Meinz & Salthouse, 1998), die feststellten, dass Männer im Alter eine größere Gehirnatrophie<sup>10</sup> erleiden und dass der altersbezogene Verfall nur in jenen Gebieten stärker ist als bei Frauen, die mit der linken Hemisphäre in Verbindung gebracht werden.

Auf Grund der konträren Ergebnisse auf diesem Gebiet, führten Meinz und Salthouse (1998) eine meta-analytische Untersuchung über 25 Studien mit einer Gesamtzahl von 5201 Teilnehmern im Alter zwischen 18 und 94 Jahren durch, die ebenfalls das Ergebnis besserer männlicher Leistungen in räumlichen Fähigkeiten stützt. Weiters konnten die Autoren jedoch auch feststellen, dass sich zunehmendes Alter auf räumliche Fähigkeiten (sowie auf alle erhobenen Variablen, mit der Ausnahme von Wissen und Wortflüssigkeit) negativ auswirkt, ohne bei diesem negativen Effekt einen Unterschied zu Gunsten eines Geschlechts zu finden. Zu gleichen Ergebnissen kamen auch Willis und Schaie (1988).

---

<sup>10</sup> Gehirnatrophie bezeichnet die „Verminderung der Gehirns substanz durch Gewebeschrumpfung.“ (<http://lexikon.meyers.de/wissen/Hirnatrophie> ).

# **Empirischer Teil**



## 6. Stichprobe

### 6.1 Gesamtstichprobe

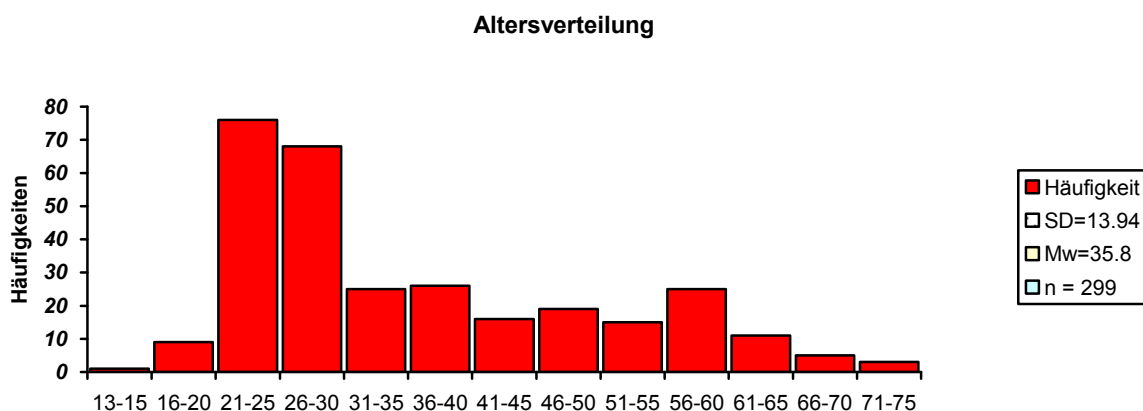
Die Gesamtstichprobe dieser Studie besteht aus 302 Testpersonen. Hierbei mussten jedoch drei Probanden auf Grund unseriöser Testbearbeitung ausgeschlossen werden. Dies war dann der Fall, wenn eine Testperson mehr als 15 Mal die Antwortmöglichkeit „Ich weiß nicht“ ankreuzte und dadurch als nicht ausreichend motiviert angesehen wurde. Nach Ausschluss von drei Testpersonen anhand der soeben erwähnten Kriterien wurden für die Berechnungen nur mehr 299 Personen herangezogen.

Die Gesamtstichprobe weist ein Geschlechtsverhältnis von 152 Männern (durchschnittliches Alter=35.82, SD=13.86) zu 147 Frauen (durchschnittliches Alter=35.87, SD=14.08) auf; dies entspricht einem prozentuellen Verhältnis von 51% zu 49%.

	Häufigkeit	Prozent
männlich	152	50.8
weiblich	147	49.2
Gesamt	299	100.0

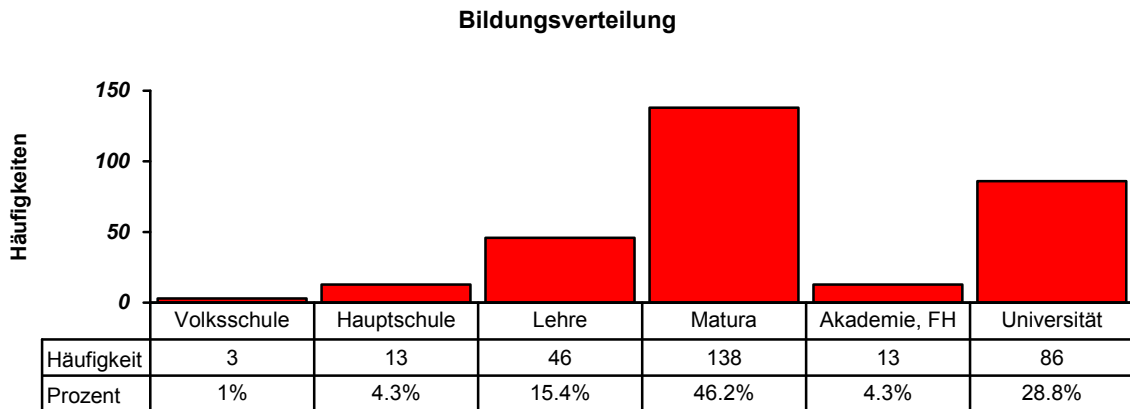
**Tab.6.1.1:** Geschlechterverteilung der Gesamtstichprobe in Fallzahlen (Häufigkeiten) und Prozentangaben.

Das Alter der Versuchspersonen variierte zwischen 13 und 74 Jahren mit einem Mittelwert von 35.84 und einer Standardabweichung von 13.94.



**Abb.6.1.1:** Altersverteilung der Gesamtstichprobe mit Mittelwert (Mw) und Standardabweichung (SD)

Das durchschnittliche Bildungsniveau der Stichprobe ist sehr hoch; 46.2% der Testpersonen wiesen eine abgeschlossene Matura auf, und 28.8% hatten einen Universitätsabschluss.



**Abb.6.1.2:** Bildungsverteilung der Gesamtstichprobe; Häufigkeiten und Prozentangaben pro Ausbildungsstufe

Innerhalb der Gesamtstichprobe gaben 96 Personen (32.1%) an, Unterricht in darstellender Geometrie gehabt zu haben, während 203 Personen (67.9%) dies verneinten.

Weiters setzt sich die Gesamtstichprobe aus 262 (87.6%) Rechtshändern, 17 (5.7%) Linkshändern und 20 (6.7%) Personen zusammen, die angeben im Alltag beide Hände in gleichem Ausmaß zu benutzen.

## 6.2 Italienische Stichprobe

Um mit der Stichprobe eine möglichst weit gefächerte Bandbreite der Population abzudecken, wurden die Versuchspersonen auf verschiedenste Weise gesucht. Teilweise wurden sie auf der Strasse angesprochen, teilweise wurde an verschiedenen Instituten der Universität von Rom eine Stichprobe gezogen; andere Personen nahmen aus persönlichen Gründen (Interesse oder Bekanntschaft mit der Versuchsleiterin) an der Untersuchung teil und brachten wiederum ihnen bekannte Personen zu der Testung. Ebenso wurde die Hilfe der italienischen Kongregation „Madonna della neve“ in Wien in Anspruch genommen.

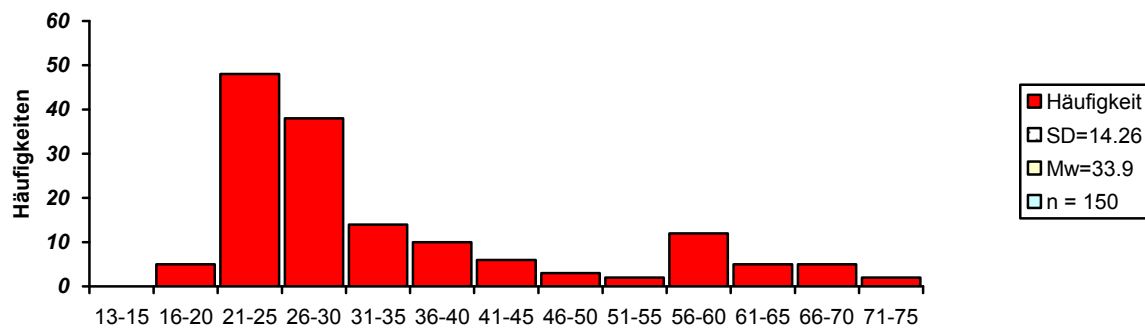
Die italienische Stichprobe besteht aus insgesamt 150 Personen (77 Männern und 73 Frauen); dies entspricht einem prozentuellen Verhältnis von 51.3% zu 48.7%.

	Häufigkeit	Prozent
männlich	77	51.3
weiblich	73	48.7
Gesamt	150	100.0

**Tab.6.2.1:** Geschlechterverteilung der italienischen Stichprobe in Fallzahlen (Häufigkeiten) und Prozentangaben.

Das Alter der Versuchspersonen erstreckte sich von 16 bis 72 Jahren mit einem Mittelwert von 33.93 und einer Standardabweichung von 14.26.

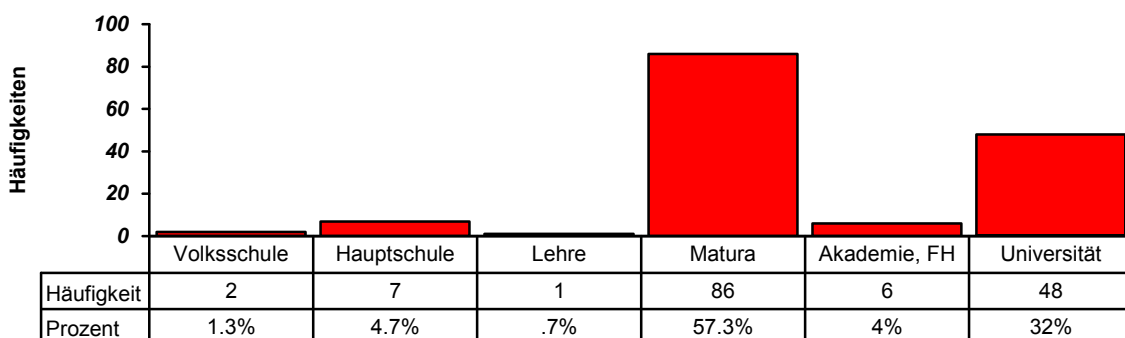
#### Altersverteilung



**Abb.6.2.1:** Altersverteilung der italienischen Stichprobe mit Mittelwert (Mw) und Standardabweichung (SD).

Das durchschnittliche Bildungsniveau ist auch in der italienischen Stichprobe hoch. 57.3% der Versuchsteilnehmer wiesen eine abgeschlossene Matura auf und 32% einen Universitätsabschluss.

#### Bildungsverteilung



**Abb.6.2.2:** Bildungsverteilung der italienischen Stichprobe; Häufigkeiten und Prozentangaben pro Ausbildungsstufe

46 Personen der italienischen Stichprobe gaben an, Unterricht in darstellender Geometrie gehabt zu haben (das entspricht 30.7%) und 104 Personen (69.3%) verneinten dies.

In der italienischen Stichprobe gab es 129 Rechtshänder (86%), 10 Linkshänder (6.7%) und 11 (7.3%) Personen, die beide Hände in gleichem Ausmaß benützen.

### 6.3 Österreichische Stichprobe

Auch bei der Erfassung der österreichischen Stichprobe wurde darauf geachtet ein möglichst breites Spektrum unterschiedlicher Altersgruppen und Bildungsklassen zu gewährleisten. Auch die österreichischen Probanden wurden auf ähnliche Weise (Universität, Bekanntenkreis etc.) gefunden, wie bereits oben im Zusammenhang mit der italienischen Stichprobe erwähnt wurde.

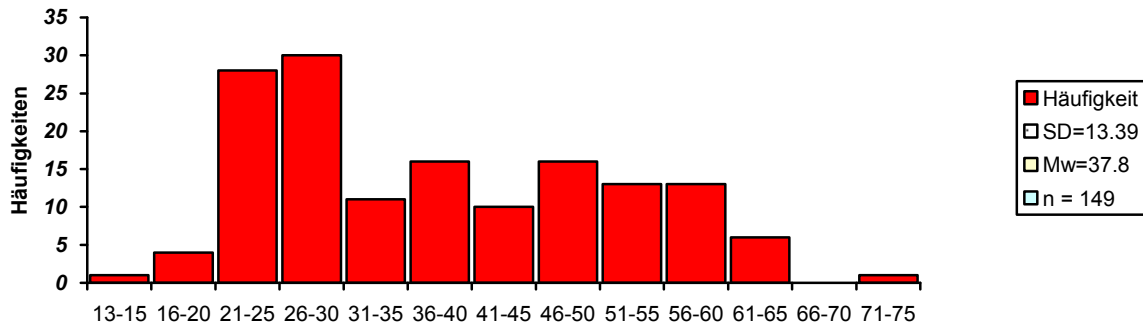
Die österreichische Stichprobe besteht aus insgesamt 149 Personen, davon 75 Männer (dies entspricht 50.3%) und 74 Frauen (dies entspricht 49.7%).

	Häufigkeit	Prozent
männlich	75	50.3
weiblich	74	49.7
Gesamt	149	100.0

**Tab.6.3.1:** Geschlechterverteilung innerhalb der österreichischen Stichprobe in Fallzahlen und Prozentangaben.

Die Versuchspersonen befanden sich in einem Altersbereich von 13 bis 74 Jahren mit einem Mittelwert von 37.77 und einer Standardabweichung von 13.39.

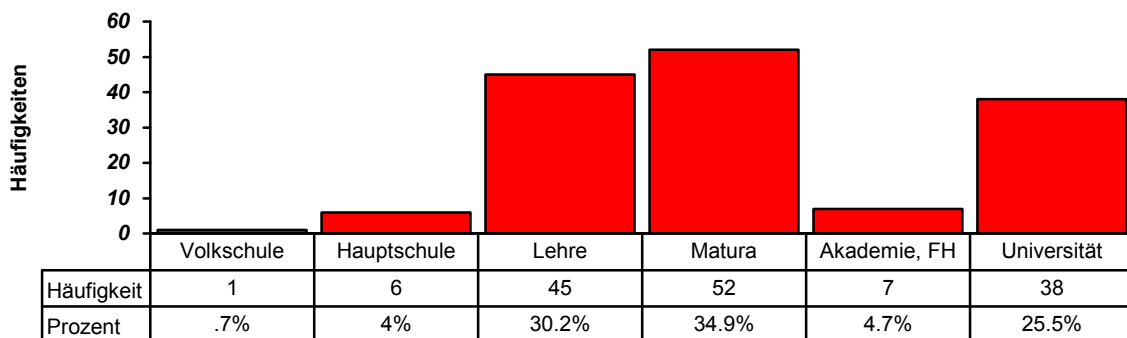
### Altersverteilung



**Abb.6.3.1:** Altersverteilung der österreichischen Stichprobe mit Mittelwert (Mw) und Standardabweichung (SD).

Das durchschnittliche Bildungsniveau in der österreichischen Stichprobe ist immer noch hoch, allerdings haben hier, im Vergleich zu den 57.3% der italienischen Stichprobe nur 34.9% der Versuchsteilnehmer eine abgeschlossene Matura und 25.5% im Gegensatz zu den 32% der italienischen Stichprobe einen Universitätsabschluss.

### Bildungsverteilung



**Abb.6.3.1:** Bildungsverteilung der österreichischen Stichprobe; Häufigkeiten und Prozentangaben pro Ausbildungsstufe

Ähnlich der italienischen Stichprobe hatten 50 der österreichischen Versuchspersonen Unterricht in darstellender Geometrie (33.6%) gehabt während 99 Testpersonen (66.4%) keine Erfahrungen damit angaben.

Die österreichische Stichprobe setzte sich aus 133 (89.3%) Rechts-, und 7 (4.7%) Linkshändern zusammen. 9 Personen (6%) gaben an vorwiegend beide Hände zu nutzen.

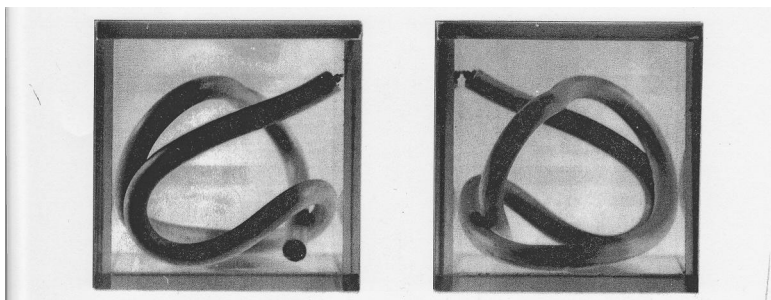
## 7. Erhebungsinstrumente

Zur Erfassung der Raumvorstellungsfähigkeit und zur Beantwortung der drei Hauptfragestellungen wurde in dieser Arbeit der Endlosschleifentest (EST) (Gittler & Arendasy, 2003) verwendet. Diesem wurde ein Datenblatt vorangestellt, welches neben soziodemographischen Daten auch Informationen über die Handpräferenz einer Person, sowie Erfahrungen mit darstellender Geometrie erhob und es dadurch erlaubt, die Unterhypothesen zu untersuchen.

### 7.1 Der Endlosschleifentest (EST)

Der EST entstand in langjähriger Entwicklung in Anlehnung an den Schlauchfigurentests (SFT) von Stumpf und Fay (1983). Das bedeutendste Ziel, welches der Entwicklung des EST zu Grunde lag, war es einen Aufgabentyp zu schaffen, der im Sinne des Rasch-Modells eindimensional messe (Gittler & Arendasy, 2003); eine testtheoretische Eigenschaft die dem SFT nicht eigen ist.

Bei der Bearbeitung des Schlauchfigurentests werden den Versuchspersonen zwei Photos derselben Schlauchfigur, welche in einem durchsichtigen Würfel eingeschlossen ist, aus zwei verschiedenen Perspektiven vorgegeben (Abb.7.1.1). Ihre Aufgabe ist es festzustellen, ob der rechte Würfel in Relation zum linken (dieser stellt immer die Frontansicht dar), von oben, von hinten, von rechts oder von links zu sehen ist.



**Abb.7.1.1** Beispielaufgabe des SFT (nach Stumpf & Fay, 1983)

Ausgehend von den Schlauchfiguren (Stumpf & Fay, 1983) entwickelten Gittler und Arendasy (2003) den EST, der räumliches Vorstellungsvermögen im Sinne der Rasch-Homogenität erhebt. Um dieses Ziel zu erreichen, setzten sie die Lösung der, die Rasch-Homogenität beeinträchtigenden Faktoren des SFT folgendermaßen um:

- a) da im Schlauchfigurentest die offenen Enden der Schläuche als Orientierungshilfe dienen konnten, wurden die Endlosschleifen in sich geschlossen.

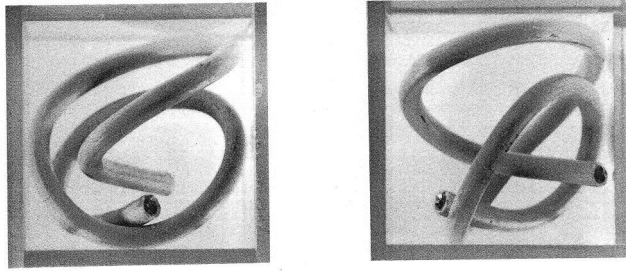


Abb.7.1.2 Item 7 des SFT (nach Stumpf & Fay, 1983, S.4)



Abb.7.1.3 Beispielim des EST

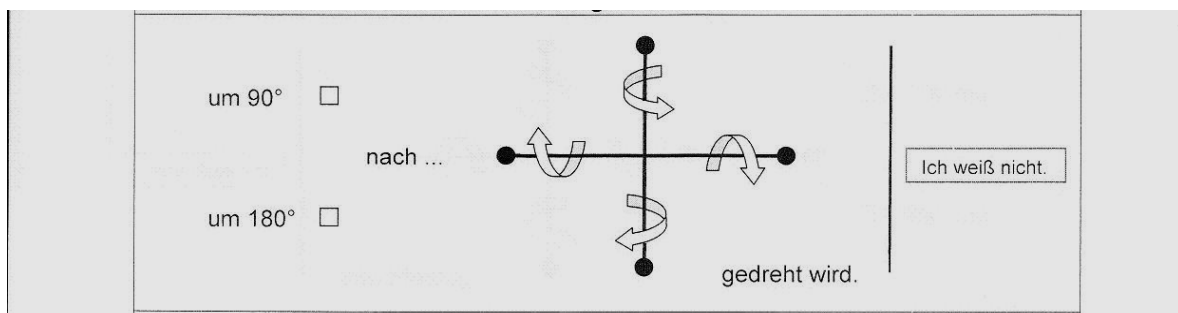
- b) manche Items des SFT beinhalten eine spiegelbildliche Lösung oder eine Lösung „von hinten“ (Gittler & Arendasy, 2003, p. 167), was die Rasch-Homogenität beeinflussen kann (das in Abb.7.1.1 dargestellte Item stellt eine derartige Lösung dar). Jene

Aufgaben, die im Test von Stumpf und Fay (1983) „von hinten“ gelöst werden konnten, werden im EST ausgeschlossen.

- c) das Vorgabeformat wird im EST gegenüber dem SFT abgeändert, so dass es weniger leicht zu Verwechslungen der Bearbeitungsrichtung bei der Itemlösung kommen kann (siehe Abb.7.1.2).

Neben der Beschriftung durch „Start-Ansicht“ und „Ziel-Ansicht“ über jeder Endlosschleife, wurde auch die Positionierung der Figuren im Vergleich zum SFT verändert. Diese diagonale Vorgabeform (im Gegensatz zu der parallelen im SFT) bringt einen weiteren Vorteil mit sich; während Personen durch die parallele Vorgabe im SFT besonders auffällige Schleifenteile als Orientierungshilfe benutzen können, wird diese Vorgehensweise durch die diagonale Darbietung im EST etwas erschwert (siehe Abb. 7.1.3).

- d) im Gegensatz zum SFT wird der EST ohne Zeitlimitierung vorgegeben, wobei die Items schwierig genug sein müssen, um dennoch ausreichend zwischen den Personen differenzieren zu können.
- e) um der Ratewahrscheinlichkeit entgegenzuwirken, wurde dem Antwortformat eine „Ich weiß nicht“ Möglichkeit hinzugefügt. Die naive Ratewahrscheinlichkeit des EST beträgt somit nur mehr 1/6 (siehe Abb. 7.1.4).



**Abb.7.1.4** Antwortformat des EST mit der Lösungsmöglichkeit „Ich weiß nicht“

Die Entwicklung des EST erfolgte über viele Jahre hinweg in zahlreichen Schritten, wobei jede Entwicklungsphase Neuerungen und Verbesserungen hinsichtlich der vorangegangenen Version mit sich brachte (für eine detaillierte Übersicht der einzelnen Entwicklungsschritte siehe Arendasy, 2000).



## 7.2 Das persönliche Datenblatt

Dieses wurde jeweils vor Beginn der Testung vorgegeben und dient der Erhebung sowohl soziodemographischer Daten wie Geschlecht, Alter und höchste abgeschlossene Ausbildung (Volksschule, Hauptschule oder AHS-Unterstufe, Lehre oder Berufsbildende mittlere Schule, Matura, Akademie oder Fachhochschule, Universität), als auch der Erhebung von Informationen über die Handpräferenz einer Person, sowie über etwaige vorangegangene Erfahrungen mit darstellender Geometrie (siehe Anhang III und IV: Erhebung der Personendaten).

## 7.3 Übersetzung der Erhebungsinstrumente

Der EST wurde mittels einer einfachen „translation-backtranslation“ Methode (Werner & Campbell, 1970) übersetzt.

Dabei wurde in einem ersten Schritt der EST von der Autorin ins Italienische übersetzt und einer zweiten Person, einer Italienisch-Dolmetscherin, Helga Kurzweil<sup>11</sup>, vorgelegt. Diese nahm die Rückübersetzung der italienischen Version ins Deutsche vor und gemeinsam wurden etwaige Verständnisprobleme sowie unpassende Formulierungen geprüft. Nachdem eine Übereinstimmung erzielt war, wurde der Test zwei Italienern vorgelegt, die ihn nochmals auf Verständnis und Korrektheit (auch auf etwaige sprachliche Besonderheiten) analysierten. Diese endgültige Version wurde dann für alle weiteren Testungen verwendet.

Obwohl der Autorin bewusst ist, dass diese Methode der Testübersetzung in der Wissenschaft nicht mehr als der „goldene Weg“ angesehen wird, wurde sie dennoch aus verschiedenen Gründen gewählt: Zum ersten ist der EST ein Test, der verhältnismäßig wenig sprachliche Komponenten beinhaltet. Neben einer kurzen Einleitung, in der den Personen mit großteils bildlichem Material die Aufgabenstellung erklärt wird, beschränken sich seine sprachlichen Komponenten auf die Wörter „Start-Ansicht“ und „Ziel-Ansicht“, um die Bearbeitungsrichtung vorzugeben.

Auf eine Übersetzung der Bezeichnung „Endlosschleifentest“ wurde verzichtet, da „Schleife“ im Italienischen verschiedenartigste Bedeutungen haben kann, die sich vor allem auf

---

<sup>11</sup> Sowohl die Autorin als auch die Dolmetscherin haben lange Zeit in Italien gelebt, sprechen die Sprache fließend und sind mit der Kultur als auch deren Besonderheiten vertraut.

spezifische Schleifen (wie etwa die Biegung eines Flusses oder die Schleife einer Masche) beziehen. Es wurde die Abkürzung „EST“ beibehalten und bei Nachfrage wurde die Bedeutung umschrieben.

Auf Grund des geringen Sprachanteils und der präzisen Formulierung, wurde demnach die Gefahr einer unangemessenen Übersetzung (aufgrund von Metaphern oder komplizierter - kulturspezifischer Konstrukte) als gering erachtet und gab Anlass, die oben erwähnte einfache Methode der Übersetzung anzuwenden.

Auch für die Übersetzung des Datenblatts wurde dieselbe Vorgehensweise gewählt wie für die Testübersetzung und auch die mitwirkenden Personen waren dieselben. Für die Sicherstellung der Äquivalenz der Ausbildungsstufen wurde hierfür jedoch zusätzlich der aus Deutschland stammende Prof. Thomas Hünefeldt hinzugezogen, der an der Universität „la Sapienza“ in Rom als PhD für Philosophie und Psychologie tätig ist.

## **7.4 Testvorgabe**

In dieser Studie wurde den Versuchspersonen eine paper-pencil-Version von 22 Items, denen 5 Beispielitems vorausgingen, vorgelegt.

Nach einem bereits gelösten Item, das dazu dient, die Aufgabenstellung transparenter zu machen, mussten die Personen 5 Beispielaufgaben lösen, die sie danach selbst kontrollieren konnten. Diese Kontrolle bietet den Personen die Möglichkeit, das Aufgabenverständnis für sich selbst zu überprüfen. Nach der Kontrolle war die Uhrzeit des Testbeginns einzutragen und die Testpersonen begannen mit der Bearbeitung.

Die Items dieser Studie wurden in steigender Schwierigkeit vorgegeben, wobei das erste Item als „warming-up-Aufgabe“ nicht in die Berechnung einging. Item zwei bis acht (Poolnummer: p204, p234, p239, p237, p251, p224, p208, p210) zählen zu der Schwierigkeitsstufe 1 und stellen damit das leichteste der drei Schwierigkeitslevels dar. Item neun bis fünfzehn (Poolnummer: p215, p231, p218, p232, p221, p243, p207) gehören dem nächst höheren Schwierigkeitslevel an und Item sechzehn bis einundzwanzig (Poolnummer: p245, p250, p246, p222, p241, p216) sind als die schwierigsten Aufgaben anzusehen.

## 7.4.1 Aufgabenstellung

Der Testung ging eine, für alle Testpersonen standardisierte, mündliche Instruktion voraus, die auch in dem Arbeitsheft schriftlich abgebildet wurde.

**Compito:** Quando sono identiche l'immagine iniziale e quella finale?

Le immagini iniziali e finali sono identiche quando:

Immagine iniziale



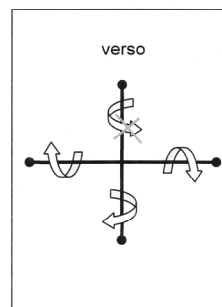
Immagine finale



l'immagine iniziale viene girata:

di 90°

di 180°



© G. Gittler & M. Arendasy 1995-2008

**Abb.7.4.1.1** Instruktionsbeispiel der italienischen Version des EST

Der EST beinhaltet eine sehr klare, präzise und kurze Aufgabenstellung, dennoch wurden die Untersuchungsteilnehmer ermutigt, bei Unklarheiten nachzufragen und dann möglichst genau aber auch zügig zu arbeiten. Besonderer Wert wurde darauf gelegt, die Versuchspersonen darauf aufmerksam zu machen, die Bearbeitungsrichtung nicht zu verwechseln. Ebenso wurde den Untersuchungsteilnehmern die Anonymität ihrer Teilnahme versichert.

Jedes Item des EST beinhaltet eine Endlosschleife, aus zwei verschiedenen Perspektiven. Aufgabe der Testpersonen war es, festzustellen ob dieser Gegenstand von der „Start-Ansicht“ in die „Ziel-Ansicht“ (von links nach rechts) entweder um 90° oder um 180° gedreht werden musste, um dieselbe Ansicht zu erhalten, sowie die Richtung anzukreuzen in die der Gegenstand gedreht werden muss, um von der Start-Ansicht in die Ziel-Ansicht zu gelangen.

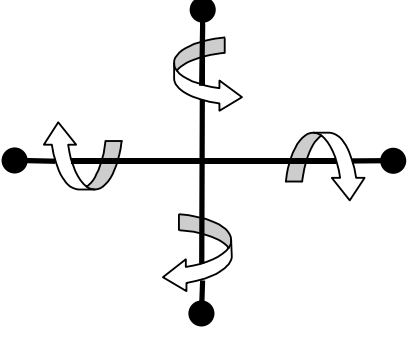
Das Antwortformat des EST in dieser Version beinhaltet anstelle von Formulierungen Pfeile, welche die jeweilige Richtung indizieren (siehe Abb.7.4.1.1). Um den Einfluss von

Ratetendenzen der Versuchspersonen zu verringern, wurde dem Antwortformat weiters die Möglichkeit „ich weiß nicht“ hinzugefügt.

Als richtig wurde ein Item dann gewertet, wenn sowohl die angekreuzte Richtung als auch der Winkel, um den das Objekt rotiert wurde, korrekt gelöst wurden.

Nach der Bearbeitung wurde, falls von den Testpersonen gewünscht, der Test sofort ausgewertet, um eine Rückmeldung der Ergebnisse zu ermöglichen.

Le immagini iniziale e finale sono identiche, quando l'immagine iniziale viene girata...

<b>Compito 8</b>		
di 90° <input type="checkbox"/>	verso 	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Non lo so.</div>
di 180° <input type="checkbox"/>		

Start-Ansicht und Ziel-Ansicht sind gleich, wenn die Start-Ansicht ...

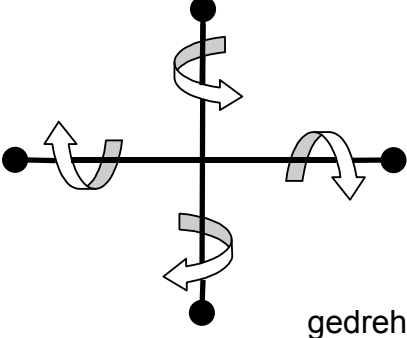
<b>Aufgabe 8</b>		
um 90° <input type="checkbox"/>	nach ... 	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Ich weiß nicht.</div>
um 180° <input type="checkbox"/>		

Abb.7.4.1.2 Beispiele für das Antwortformat des EST der italienischen und der österreichischen Version

## 8. Datenerhebung

Die Datenerhebung erfolgte im Zeitraum von August bis September in Italien und von Oktober bis Dezember in Österreich.

Auf Grund der geographischen Verhältnisse konnten die Untersuchungen nicht immer im selben Testraum durchgeführt werden, doch es wurde dafür Sorge getragen, dass alle Versuchspersonen unter denselben Bedingungen getestet wurden und dass für die Bearbeitung des EST immer ein ruhiger Raum mit guten Lichtverhältnissen zur Verfügung stand. Die Testung wurde meist als Einzeltestung durchgeführt; manchmal waren Gruppentestungen möglich, die eine Testung von bis zu vier Personen zuließen.

Die Versuchspersonen der italienischen Stichprobe wurden hauptsächlich in Rom und Umgebung, sowie in Süditalien rekrutiert.

Da die Versuchspersonen nicht randomisiert zu der unabhängigen Variable „Kultur“, zugeteilt werden können, handelt es sich bei dieser Untersuchung um ein Quasi-experimentelles Design. Wie Van de Vijver und Leung (1997, a) richtigerweise anmerken, wird in kulturvergleichenden Untersuchungen, Kultur als unabhängige Variable (UV) angesehen. Doch während in experimentellen Designs die UV vom Versuchsleiter kontrolliert werden kann (z. B. Medikamenten- oder Placeboeinnahme), ist die zufällige Zuteilung zu einer Kultur nicht möglich. Diese Tatsache impliziert schwerwiegende Folgen für die Ergebnisinterpretation; während ein eventuell gefundener Unterschied in einem experimentellen Design unter Konstanthaltung der Rahmenbedingungen nahezu eindeutig auf die UV zurückzuführen ist, kann ein derartiges Ergebnis in kulturvergleichenden Untersuchungen vielfältige Interpretationen zulassen. „Kultur“ muss dazu in zahlreiche Einzelfaktoren aufgespalten werden, wie etwa personenbezogene (Alter, Geschlecht, etc.) oder kulturbezogene (Schulsystem, Gesundheitssystem etc.) (Van de Vijver & Leung, 1997, a), welche die Personen über lange Zeit beeinflussen (Berry et al., 1992).

## 9. Fragestellungen und Hypothesen

### **Rasch-Modell Analysen des EST**

- Ist der EST in der italienischen Stichprobe Rasch-homogen, d.h. misst er eindimensional?
- Ist der EST innerhalb der österreichischen Stichprobe Rasch-homogen?
- Ist der EST über beide Kulturen gemeinsam Rasch-homogen beziehungsweise erfasst er in beiden Kulturen dieselbe latente Fähigkeitsdimension der Raumvorstellung?

### **Gibt es einen kulturspezifischen Leistungsunterschied?**

- Unterscheiden sich die beiden verglichenen Kulturstichproben in ihren Fähigkeiten zum räumlichen Vorstellungsvermögen?

### **Gibt es geschlechtsspezifische Leistungsunterschiede?**

- Unterscheiden sich die Geschlechter in Aufgaben zum räumlichen Vorstellungsvermögen über beide Kulturen?
- Unterscheiden sich die Geschlechter in Aufgaben zum räumlichen Vorstellungsvermögen innerhalb beider Kulturen?

### **Gibt es einen Einfluss von Erfahrungen in „darstellender Geometrie“ auf die Leistung?**

- Unterscheiden sich jene Personen, die Erfahrungen mit „darstellender Geometrie“ im Rahmen von Unterricht in dieser, hatten, von Personen ohne diese Erfahrung?

### **Gibt es einen Einfluss von Händigkeit auf die Leistung?**

- Unterscheiden sich rechts-, links- und beidhändige Personen in Anbetracht ihres Raumvorstellungsvermögens?

### **Gibt es altersspezifische Leistungsunterschiede?**

- Unterscheiden sich jüngere und ältere Probanden bezüglich ihrer EST-Leistung in Abhängigkeit ihres Alters?

### **Gibt es bearbeitungszeitspezifische Leistungsunterschiede?**

- Unterscheiden sich Personen hinsichtlich der von ihnen aufgewandten Testbearbeitungszeit?

## 10. Rasch-Modell Analysen des EST

Eine grundlegende Fragestellung dieser Arbeit galt dem Umstand, ob der EST erneut in einer österreichischen Stichprobe und zum ersten Mal in einer italienischen Stichprobe, die Eigenschaft der Rasch-Homogenität (siehe Kapitel 4.1) aufweist.

Sollte dies gegeben sein, wird in dieser Untersuchung weiterhin überprüft ob der EST auch über beide Kulturen gemeinsam Rasch-homogen misst und somit ein kulturfreies und kulturfares Maß der Raumvorstellungsfähigkeit (im speziellen der mentalen Rotation) darstellt. Im folgenden Kapitel werden diese Hypothesen untersucht.

Wie in Kapitel 4.1.2 erwähnt, wurde die Überprüfung der Geltung des Rasch-Modells anhand *Likelihood Ratio Tests* (LRT) nach Andersen (1973) durchgeführt.

Als internes Kriterium diente hierbei für beide Kulturen die Trennung der Scores in niedrige und hohe Leistung (am Median der Testpersonen), wodurch die Stichproben in zwei Gruppen aufgeteilt wurden. Als externe Kriterien wurden für beide Kulturen die Stichproben-typischen Faktoren Geschlecht, Alter, Unterricht in darstellender Geometrie und Bearbeitungszeit des EST gewählt. Bei der Überprüfung der Rasch-Homogenität im Kulturvergleich wurde zusätzlich „Kultur“ als externes Kriterium herangezogen.

Die Irrtumswahrscheinlichkeit für die LRTs wurde mit  $\alpha=.01$  festgesetzt. Die Ergebnisse der Modelltests werden mit empirischen Andersens  $\chi^2$ -Werten angegeben ( $\chi^2_{(emp)}$ ). Ist dieser Wert kleiner als der kritische  $\chi^2$ -Wert ( $\alpha=.01$ ;  $\chi^2_{(krit)}$ ), ist das Ergebnis nicht signifikant und es kann die Geltung des Rasch-Modells angenommen werden.

Die Modelltests zur Überprüfung der Rasch-Homogenität wurden mit Hilfe des Programms „LPCM-Win 1.0“ durchgeführt, welches der Autorin für den Zeitraum der Auswertung von der Universität Wien zur Verfügung gestellt wurde.



## 10.1 Rasch-Modell Analysen in der italienischen Stichprobe

Wie bereits erwähnt, besteht die italienische Stichprobe aus insgesamt 150 Personen im Alter von 16 bis 72 Jahren ( $M_w=33.93$ ,  $SD=14.26$ ,  $Median=28$ ), davon waren 77 Personen männlich und 73 weiblich.

Zunächst wurde die Rasch-Homogenität der 21 Aufgaben des EST (unter Ausschluss der ersten Aufgabe als „warming-up“) überprüft. Um die Modellpassung zu untersuchen, wurde die Stichprobe anhand eines internen und vier externer Kriterien geteilt (siehe Tab. 10.1.1).

Der erste Modelltest diente dem Vergleich der Itemparameter der verschiedenen Scoregruppen, dabei wurden die Personen am Median der erzielten Testleistungen ( $Median_{Italien}=9$ ) in zwei Gruppen geteilt (hohe vs. niedrige Testleistung) und anhand eines LRT verglichen. Sein Ergebnis zeigt mit einem empirischen Andersen  $\chi^2=37.54$  bei einem kritischen  $\chi^2(\alpha=.01)=37.59$  ein zwar nicht signifikantes, aber dennoch grenzwertiges Resultat für das interne Kriterium „Testleistung“ ( $df=20$ ,  $p=.0101$ <sup>12</sup>).

Analog zu der Trennung nach dem internen Kriterium wurden auch für die externen Kriterien die Personen jeweils in zwei Gruppen geteilt (siehe Tab.10.1.1). Dabei gilt auch hier, dass ein nicht-signifikantes Ergebnis vorliegt wenn der empirische  $\chi^2$ -Wert kleiner ist als der kritische  $\chi^2$ -Wert von 37.59. Der erste Modelltest für die externen Kriterien teilte die Personen nach dem **Geschlecht**, wobei die Itemparameter der EST-Werte der Männer jenen der Frauen gegenübergestellt wurden. Das Ergebnis zeigt mit einem empirischen Andersen  $\chi^2(df=20, p=.14)=26.88$ , ein insignifikantes Ergebnis.

Als ein weiteres externes Kriterium wurde das **Alter** der Probanden herangezogen. Dabei wurde die Versuchsgruppe am Median ihres Alters ( $Median_{Italien}=28$ ) in jüngere vs. ältere Personen geteilt. Das empirische  $\chi^2(df=20, p=.34)=22.05$  zeigt auch hier ein nicht-signifikantes Ergebnis.

Als drittes externes Kriterium wurde die **Vorkenntnis in darstellender Geometrie** gewählt, wobei die Werte jener Personen, die über Vorkenntnisse in DG verfügten, jenen

---

<sup>12</sup> Bei der Überprüfung der Rasch-Homogenität wird von einer Nullhypothese ausgegangen (die Itemparameter der Gruppen unterscheiden sich nicht signifikant voneinander).

gegenübergestellt wurden die nicht über diese berichtet hatten. Mit einem empirischen  $\chi^2_{(df=20, p=.99)}=7.46$  zeigte sich auch hier kein signifikantes Ergebnis.

Der letzte Modelltest bezog sich auf die unterschiedliche **Bearbeitungszeit** der Probanden. Die Personen wurden am Median der Bearbeitungszeit ( $\text{Median}_{\text{Italien}}=15$ ) in zwei Gruppen geteilt und es wurden die Itemparameter jener Personen, die kürzer für die Bearbeitung aufwandten mit jenen verglichen, die mehr Zeit für die Testbearbeitung aufgewendet hatten. Dieses Ergebnis wurde bei einem empirischen  $\chi^2_{(df=20, p=.48)}=19.60$  nicht signifikant.

Teilungskriterien	n	Mw	SD	LRT ( $\chi^2_{\text{emp}}$ ) bei $\chi^2_{\text{krit}}=37.59$	Signifikanz (p)
<b>EST-Rohscore</b>	150	9.15	4.45	37.54	.0101 n.s (aber <b>grenzwertig</b> )
Niedrig 0-8	67	5.00	2.01		
Hoch 9-21	83	12.51	2.70		
<b>Geschlecht</b>				26.88	.14 n.s
Männlich	77	9.97	4.63		
Weiblich	73	8.29	4.11		
<b>Alter</b>				22.05	.34 n.s
Jung (<28)	71	9.35	4.06		
Alt ( $\geq 28$ )	79	8.79	4.80		
<b>DG</b>				7.46	.99 n.s
Nein	104	8.37	3.93		
Ja	46	10.93	5.06		
<b>Bearbeitungszeit</b>				19.60	.48 n.s
Kurz (<15)	65	8.40	4.16		
Lang ( $\geq 15$ )	85	9.73	4.60		

**Tab.10.1.1:** Rasch-Modell Analyse Italien (k=21) mit kritischem  $\chi^2_{(\alpha=0.1)}$  von 37.59 ( $\chi^2_{\text{krit}}$ ),  $df=20$  und empirischen Andersen  $\chi^2$ -Werten (LRT ( $\chi^2_{\text{emp}}$ )); es werden Mittelwert (Mw<sub>r</sub>), Standardabweichung (SD<sub>r</sub>) der Rohscores sowie die Signifikanz (p) angegeben.

Auf Grund des grenzwertigen Ergebnisses des LRT für das interne Teilungskriterium (Tab.10.1.1) wurde folgende theoretische Überlegung vorgenommen:

In der Psychologie ist es eine mögliche Vorgehensweise, die erste Aufgabe eines Tests als „warming-up“ Aufgabe zu behandeln und diese nicht in die Bewertung der Testergebnisse einfließen zu lassen. Als Grund hierfür kann angenommen werden, dass Personen unterschiedlich von den (einem psychometrischen Test meist vorangestellten) Übungsbeispielen profitieren.

In Anlehnung an Studien von Tanzer, Gittler und Sim (1994) sowie von Gittler (1999) wurde auch hier im Sinne der „Selektionshypothese“ vorgegangen. Diese besagt, dass eine „verlängerte Test-Eingangsphase“ (durch Weglassen einiger Items am Testanfang) notwendig sein kann um systematische Verzerrungen (wie beispielsweise Unterschiede darin wie Probanden von den Übungsbeispielen profitieren oder anfängliche Konzentrationsprobleme der Testpersonen) zu eliminieren. Eben dieses kann auch durch eine „verlängerte Test-Ausgangsphase“, eine so genannte „fading-out“-Phase (durch Weglassen einiger Items am

Test-Ende) erreicht werden um den eventuellen Einfluss unterschiedlich starker Motivation der Probanden am Testende zu kontrollieren.

Auf Grund dieser Überlegungen wurde angenommen, dass eine längere Test-Eingangs-, sowie Test-Ausgangsphase der Rasch-Homogenität dienlich sein könnte, weshalb im nächsten Schritt der Berechnungen zusätzliche zwei Items zu Beginn des Tests als „warming-up-Items“ behandelt, und eine Aufgabe am Testende unter Berücksichtigung der „fading-out“- Phase nicht in die Berechnung einbezogen wurden.

Mit dem, nach diesen theoretischen Überlegungen veränderten Testmaterial wurde eine erneute Überprüfung der Rasch-Homogenität anhand von 18 Items durchgeführt. Alle der in Folge durchgeführten LRTs zeigten bei einem kritischen  $\chi^2(\alpha=.01)=33.43$  insignifikante Ergebnisse. Dies bedeutet, dass alle Items dieselbe latente Fähigkeitsdimension erheben und Rasch-Homogenität der 18 Items des EST in der italienischen Stichprobe angenommen werden kann (siehe Tab.10.1.2).

Die Teilung nach dem internen Kriterium „niedriger vs. hoher Score“ ( $\text{Median}_{\text{Italien}}=7$ ) zeigte bei den verbleibenden 18 Items mit einem empirischen  $\chi^2(\text{df}=17, p=.11)=24.35$  kein signifikantes Ergebnis. Ebenso wenig ergab die Teilung der Personen anhand des externen Kriteriums „Geschlecht“ bei einem empirischen  $\chi^2(\text{df}=17, p=.16)=22.68$  ein signifikantes Resultat. Auch der Modelltest zur Teilung der Probanden nach dem externen Kriterium „Alter“ erbrachte bei einem empirischen  $\chi^2(\text{df}=17, p=.85)=11.07$  ein insignifikantes Ergebnis. Gleiches gilt für die Teilung nach den externen Kriterien „DG“ bei  $\chi^2(\text{df}=17, p=.98)=6.90$  und die aufgewendete „Bearbeitungszeit“ bei  $\chi^2(\text{df}=17, p=.76)=12.64$ .

Teilungskriterien	n	Mw	SD	LRT ( $\chi^2_{\text{emp}}$ ) bei $\chi^2_{\text{krit}}=33.43$	Signifikanz (p)
<b>EST-Rohscore</b>					
Niedrig 0-6	150	7.43	3.71	24.35	.11 n.s
Hoch 7-18	63	3.87	1.57		
	87	10.01	2.46		
<b>Geschlecht</b>				22.68	.16 n.s
Männlich	77	8.09	3.88		
Weiblich	73	6.74	3.40		
<b>Alter</b>				11.07	.85 n.s
Jung (<28)	71	7.56	3.40		
Alt (≥28)	79	7.32	3.98		
<b>DG</b>				6.90	.98 n.s
Nein	104	6.73	3.25		
Ja	46	9.02	4.20		
<b>Bearbeitungszeit</b>				12.64	.76 n.s
Kurz (<15)	65	6.80	3.48		
Lang (≥15)	85	7.92	3.82		

**Tab.10.1.2:** Rasch-Modell Analyse Italien (k=18) mit kritischem  $\chi^2(\alpha=.01)$  von 33.43 ( $\chi^2_{\text{krit}}$ ), df=17 und empirischen Andersen  $\chi^2$ -Werten (LRT ( $\chi^2_{\text{emp}}$ )); es werden Mittelwert ( $Mw_r$ ), Standardabweichung ( $SD_r$ ) der Rohscores sowie die Signifikanz (p) angegeben.

## 10.2 Rasch-Modell Analysen in der österreichischen Stichprobe

Wie bereits erwähnt, besteht die österreichische Stichprobe aus insgesamt 149 Personen in einem Alter von 13 bis 74 Jahren ( $Mw=37.77$ ,  $SD=13.39$ ,  $Median=36$ ), davon 75 Männer und 74 Frauen.

Auch in der österreichischen Stichprobe wurden analog zu Abschnitt 10.1 vorgegangen. Zunächst wurden die 21 Items (unter Ausschluss der ersten Aufgabe als „warming-up“) auf Rasch-Homogenität untersucht.

Um die Modellpassung zu prüfen, wurde auch für die österreichische Stichprobe als erster Modelltest die Teilung der Personen in zwei Gruppen am Median der Testleistung ( $Median_{\text{Österreich}}=9$ ) durchgeführt. Die Teilung nach dem internen Kriterium (hohe vs. niedrige Testleistungen) ergab hier mit einem empirischen Andersen  $\chi^2_{(df=20, p=.07)}=29.81$  bei einem kritischen  $\chi^2_{(a=.01)}=37.59$  ein insignifikantes Ergebnis.

Analog zu der Trennung nach dem internen Kriterium wurden auch für die externen Kriterien, die Personen wie in Abschnitt 10.1 jeweils in zwei Gruppen geteilt (siehe Tab.10.2.1). Der Modelltest der Teilung nach dem externen Kriterium **Geschlecht** ergab mit einem empirischen Andersen  $\chi^2_{(df=20, p=.14)}=26.66$  ein insignifikantes Ergebnis.

Ein weiterer Modelltest prüfte die Teilung nach dem externen Kriterium **Alter**, auch hier wurden die Probanden am Median des Alters ( $Median_{\text{Österreich}}=36$ ) in zwei Gruppen geteilt. Das Resultat dieses LRT ist ebenso nicht-signifikant bei einem empirischen  $\chi^2_{(df=20, p=.56)}=18.36$ .

Die Teilung nach dem externen Kriterium **Erfahrung mit darstellender Geometrie** zeigte sich mit einem empirischen  $\chi^2_{(df=20, p=.39)}=21.09$  gleichfalls insignifikant.

Der letzte Modelltest bei dem die Stichprobe am Median der **Bearbeitungszeit** ( $Median_{\text{Österreich}}=17$ ) in zwei Gruppen geteilt wurden, ergab hingegen bei einem kritischen  $\chi^2_{(df=20, p=.006)}=39.54$  einen signifikanten Wert.

Teilungskriterien	n	Mw	SD	LRT ( $\chi^2_{emp}$ ) bei $\chi^2_{krit}=37.59$	Signifikanz (p)
<b>EST-Rohscore</b>	149	9.54	4.54	29.81	.07 n.s
Niedrig 0-8	66	5.26	1.87		
Hoch 9-21	83	12.94	2.83		
<b>Geschlecht</b>				26.66	.14 n.s
Männlich	75	11.01	4.60		
Weiblich	74	8.04	3.98		
<b>Alter</b>				18.36	.56 n.s
Jung (<36)	74	11.68	4.25		
Alt (≥36)	75	7.43	3.79		
<b>DG</b>				21.09	.39 n.s
Nein	99	9.01	4.40		
Ja	50	10.58	4.69		
<b>Bearbeitungszeit</b>				39.54	<b>.006 signifikant</b>
Kurz (<17)	61	8.72	4.64		
Lang (≥17)	88	10.10	4.41		

**Tab.10.2.1:** Rasch-Modell Analyse der österreichischen Stichprobe (k=21) mit kritischem  $\chi^2_{(\alpha=.01)}$  von 37.59 ( $\chi^2_{krit}$ ), df=20 und empirischen Anderson  $\chi^2$ -Werten (LRT ( $\chi^2_{emp}$ )); es werden Mittelwert ( $Mw_r$ ), Standardabweichung ( $SD_r$ ) der Rohscores sowie die Signifikanz (p) angegeben.

Auf Grund des signifikanten Ergebnisses des LRT bei der Teilung nach dem externen Kriterium „Bearbeitungszeit“ (siehe Tab.10.2.1) wurde auch hier, nach der bereits oben erwähnten, „Selektionshypothese“ vorgegangen.

Diesen theoretischen Überlegungen zu Folge wurden für die nächste Rasch-Modell-Analyse auch in der österreichischen Stichprobe zwei Items am Testbeginn und eines am Testende herausgenommen und die erneute Überprüfung der Rasch-Homogenität wurde anhand von k=18 Items wiederholt (siehe Tab.10.2.2).

Teilungskriterien	n	Mw	SD	LRT ( $\chi^2_{emp}$ ) bei $\chi^2_{krit}=33.43$	Signifikanz (p)
<b>EST-Rohscore</b>	149	7.74	3.79	16.24	.51 n.s
Niedrig 0-6	63	4.13	1.54		
Hoch 7-18	86	10.40	2.54		
<b>Geschlecht</b>				22.26	.17 n.s
Männlich	75	8.91	3.89		
Weiblich	74	6.57	3.31		
<b>Alter</b>				15.61	.55 n.s
Jung (<36)	74	9.51	3.61		
Alt (≥36)	75	6.00	3.11		
<b>DG</b>				16.91	.46 n.s
Nein	99	7.35	3.65		
Ja	50	8.52	3.97		
<b>Bearbeitungszeit</b>				21.12	.22 n.s
Kurz (<17)	61	7.14	3.82		
Lang (≥17)	88	8.16	3.73		

**Tab.10.2.2:** Rasch-Modell Analyse der österreichischen Stichprobe (k=18) mit kritischem  $\chi^2_{(\alpha=.01)}$  von 33.43 ( $\chi^2_{krit}$ ), df=17 und empirischen Anderson  $\chi^2$ -Werten (LRT ( $\chi^2_{emp}$ )); es werden Mittelwert ( $Mw_r$ ), Standardabweichung ( $SD_r$ ) der Rohscores sowie die Signifikanz (p) angegeben.

Keiner der durchgeführten Modelltests (k=18, df=17,  $\alpha=.01$ ) zeigte signifikante Ergebnisse, womit auch in der österreichischen Stichprobe für die 18 Items des EST Rasch-Homogenität

angenommen werden kann (die Werte sind in Tab.10.2.2 dargestellt). Da alle Items sowohl in der italienischen als auch in der österreichischen Stichprobe dieselbe latente Fähigkeitsdimension erheben, kann im nächsten Schritt die Rasch-Homogenität über beide Kulturen geprüft werden.

### 10.3 Rasch-Modell Analysen über beide Kulturen

Nachdem die Eindimensionalität in beiden Kulturen gegeben ist, kann in weiterer Folge überprüft werden ob dies auch gemeinsam über beide Kulturen gilt.

Erweisen sich die im Folgenden durchgeführten Modelltests ebenso als nicht signifikant, kann davon ausgegangen werden, dass der EST in der österreichischen und der italienischen Stichprobe dieselbe latente Fähigkeitsdimension erhebt und dass somit ein Leistungsvergleich zwischen den beiden Kulturen ein faires Maß dieser darstellt. Als zusätzliches externes Kriterium wurde hier „Kultur“ verwendet. In diesem Fall werden die Rasch-homogenen Versionen des EST mit k=18 Items für weitere Berechnungen herangezogen.

Bei einem kritischen  $\chi^2(\alpha=.01)=33.43$  zeigte der Modelltest bei Teilung nach dem internen Kriterium (niedriger vs. hoher Score;  $\text{Median}_{\text{gesamt}}=7$ ) ein nicht signifikantes Ergebnis mit einem empirischen  $\chi^2(\text{df}=17, p=.03)=29.26$ . Ebenso ergaben die LRTs bei der Teilung nach den externen Kriterien **Geschlecht** mit einem empirischen  $\chi^2(\text{df}=17, p=.09)=25.42$  und **DG** mit einem empirischen  $\chi^2(\text{df}=17, p=.82)=11.64$  nicht-signifikante Ergebnisse. Die Teilung der Personen nach dem externen Kriterium **Alter** ( $\text{Median}_{\text{gesamt}}=30$ ) zeigte mit einem empirischen  $\chi^2(\text{df}=17, p=.20)=21.57$  ebenso wenig ein signifikantes Resultat wie die Teilung der Versuchspersonen am Median der **Bearbeitungszeit** ( $\text{Median}_{\text{gesamt}}=16$ ) mit einem empirischen  $\chi^2(\text{df}=17, p=.013)=32.37$  (siehe Tab.10.3.1). Der Modelltest, für die, bei dem Kulturvergleich zusätzlich hinzugefügte, Teilung nach dem externen Kriterium **Kultur**, zeigt mit einem empirischen  $\chi^2(\text{df}=17, p=.84)=11.32$  ebenso ein insignifikantes Ergebnis.

Teilungskriterium	n	Mw	SD	LRT ( $\chi^2_{emp}$ ) bei $\chi^2_{krit}=33.43$	Signifikanz (p)
<b>EST-Rohscore</b>	299	7.95	3.75	29.26	.03 n.s
Niedrig 0-6	126	4.00	1.55		
Hoch 7-18	173	10.20	2.50		
<b>Geschlecht</b>				25.42	.09 n.s
Männlich	152	8.49	3.89		
Weiblich	147	6.65	3.35		
<b>Alter</b>				21.57	.20 n.s
Jung (<30)	146	8.62	3.71		
Alt (≥30)	153	6.61	3.52		
<b>DG</b>				11.64	.82 n.s
Nein	203	7.03	3.46		
Ja	96	8.76	4.07		
<b>Bearbeitungszeit</b>				32.37	.013 n.s
Kurz (<16)	138	7.00	3.58		
Lang (≥16)	161	8.09	3.82		
<b>Kultur</b>				11.32	.84 n.s
Italien	150	7.43	3.71		
Österreich	149	7.74	3.79		

**Tab.10.3.1:** Rasch-Modell Analyse für Österreich (n=149) und Italien (n=150), bei k=18 mit kritischem  $\chi^2_{(a=.01)}$  von 33.43 ( $\chi^2_{krit}$ ), df=17 und empirischen Andersen  $\chi^2$ -Werten (LRT ( $\chi^2_{emp}$ )); es werden Mittelwert ( $Mw_t$ ), Standardabweichung ( $SD_t$ ) der Rohscores sowie die Signifikanz (p) für die Gesamtstichprobe angegeben.

Aus Tab.10.3.1 mit keinem signifikanten Ergebnis ist ersichtlich, dass nach Verlängerung der Test-Eingangs- und der Test-Endphase, ein fairer Leistungsvergleich durch den EST zwischen den beiden Kulturen Österreich und Italien gewährleistet werden kann. Durch diese Ergebnisse kann bestätigt werden, dass der EST für Italiener und Österreicher, Frauen und Männer, junge und ältere Probanden, Personen die Unterricht in darstellender Geometrie hatten und solche, die keine Erfahrungen damit haben, sowie für Personen, die lange und für solche die kurz für die Bearbeitung des EST brauchten, dieselbe Fähigkeitsdimension erhebt.

Auf Basis dieser Erkenntnisse können nun die in Kapitel 9 erwähnten kulturspezifischen Fragestellungen dieser Arbeit untersucht werden.

## 10.4 Item- und Personenparameterschätzung im Rasch-Modell

Zur Überprüfung der inferenzstatistischen Ergebnisse wurden zu Gunsten einer erhöhten Präzision nicht die Rohscores der Daten verwendet sondern die ermittelten, intervallskalierten *weighted-likelihood-estimates* (WLE) nach Warm (1989). Im dichotomen Rasch-Modell gibt es verschiedene Methoden, um die Itemparameter<sup>13</sup> zu schätzen (für eine detaillierte

<sup>13</sup> Der Begriff „Parameter“ bedeutet, dass es sich hier um zunächst unbekannte Kenngrößen handelt, deren Werte erst anhand der Daten geschätzt werden müssen (Rost, 2004).

Übersicht siehe Rost, 1988; Rost, 2004). Im Folgenden soll nur auf die, in der vorliegenden Arbeit verwendeten WLE-Parameter, eingegangen werden.

Die *weighted-likelihood-estimates* basieren auf dem Bayes-Ansatz für Parameterschätzung, in welchem „...die Wahrscheinlichkeit der Personenparameter unter der Bedingung der Daten und der Itemparameter...“ (Rost, 2004, S.313) maximiert wird. Unter jenen Methoden, die zur Schätzung der Parameter herangezogen werden, ist es diese, die den kleinsten Bias zeigt. Ein wesentlicher Vorteil ist darin zu sehen, dass es die WLE ermöglichen, Parameterschätzer für die Randbereiche der Testleistungen (alle oder kein Item richtig gelöst) zu finden.

Tab.10.4.1 zeigt die berechneten WLE-Parameter<sup>14</sup> für die, in dieser Arbeit verwendeten k=18 Testitems auf.

Rohscores	Häufigkeit der Rohscores Österreich	Häufigkeit der Rohscores Italien	WLE	SE(WLE)
0	1	0	-3.8775	1.8476
1	3	3	-2.6692	0.9691
2	7	14	-2.0447	0.7606
3	9	10	-1.5903	0.6632
4	11	8	-1.2166	0.6072
5	20	17	-0.8890	0.5716
6	12	11	-0.5900	0.5483
7	12	15	-0.3094	0.5330
8	8	15	-0.0405	0.5239
9	18	8	0.2220	0.5199
10	12	17	0.4827	0.5210
11	11	12	0.7463	0.5275
12	5	6	1.0183	0.5405
13	9	6	1.3059	0.5620
14	3	2	1.6198	0.5963
15	4	2	1.9778	0.6517
16	4	4	2.4152	0.7490
17	0	0	3.0217	0.9577
18	0	0	4.2115	1.8308
<b>Gesamt</b>	<b>149</b>	<b>150</b>		

**Tab.10.4.1:** WLE-Parameter (WLE) und ihre Standardfehler (SE), sowie die Auftrittshäufigkeiten der Rohscores in der österreichischen und italienischen Stichprobe für die (k=18) Items des EST.

<sup>14</sup> Die WLE-Parameter wurden der Autorin vom Betreuer dieser Arbeit übergeben. Ihm sei an dieser Stelle nochmals besonders gedankt.



# 11. Hypothesenprüfung der kulturvergleichenden

## Untersuchung

Nachdem anhand der Rasch-Modell-Analysen mittels LPCMwin 1.0 festgestellt wurde, dass 18 Items des EST sowohl in beiden Kulturen als auch über beide Kulturen hinweg, Rasch-homogen messen und daher ein und dieselbe latente Fähigkeitsdimension erfassen, werden im Folgenden detaillierte Unterschiede zwischen den Kulturen untersucht.

Überprüft wird, ob Unterschiede zwischen den Kulturen in den Leistungen des EST auftreten, ob das Geschlecht oder das Alter der Testpersonen einen signifikanten Einfluss auf die Leistungen im EST haben und, wenn ja, in welche Richtung dieser geht; ob auch in dieser Studie gezeigt werden kann, dass Erfahrungen mit darstellender Geometrie eine Verbesserung der Raumvorstellungsfähigkeit mit sich bringen; ob eine unterschiedliche Bearbeitungszeit signifikante Effekte auf die Leistungen im EST zeigt; ob es einen Einfluss von Bildung auf die Fähigkeit des räumlichen Vorstellens gibt und ob die Handpräferenz einer Person mit den Leistungen im EST zusammenhängt.

Alle nachfolgenden Hypothesen werden zu Gunsten einer erhöhten Genauigkeit gegenüber den Rohscores der Daten anhand der in Kapitel 10.4 beschriebenen WLE nach Warm (1989) überprüft. Das Signifikanzniveau wird generell mit  $\alpha=.05$  festgelegt. Die Tabellen mit der Überschrift „Deskriptive Statistik“ enthalten sowohl Mittelwert und Standardabweichung der WLEs ( $MW_{par}$ ,  $SD_{par}$ ) mit denen die Berechnungen durchgeführt wurden, als auch Mittelwert und Standardabweichung der Rohscores ( $MW_r$ ,  $SD_r$ ). Dies dient einer besseren Nachvollziehbarkeit der Ergebnisse für den Leser.

Für alle im Folgenden durchgeführten statistischen Analysen wurde das Programm SPSS 10.0 verwendet.

### 11.1 Leistungsunterschiede im EST zwischen den Kulturen

Eine der interessantesten Fragestellungen dieser Untersuchung bezieht sich auf das Vorhandensein eventueller Leistungsunterschiede zwischen den beiden Kulturen. Von besonderer Bedeutung ist diese Fragestellung deshalb, da der Autorin keine Studie bekannt

ist, die sich mit diesem Thema beschäftigt. Auf Grund geographischer, politischer und kultureller Ähnlichkeiten der untersuchten Kulturen wird davon ausgegangen, dass keine signifikanten Unterschiede zwischen diesen Kulturen aufgedeckt werden.

Die Normalverteilung der Personenparameter (WLE nach Warm, 1989) innerhalb der jeweiligen Kultur konnte mittels Kolmogorov-Smirnov-Test für eine Stichprobe, bestätigt werden (siehe Tabelle 11.1.1). Dies ermöglichte die Untersuchung des Leistungsvergleichs der Kulturen durch einen t-Test für unabhängige Stichproben.

### Deskriptive Statistik

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
<b>Gesamt</b>	299	7.59	3.75	-.22	1.14	1.147	.144
<b>Männer</b>	152	8.49	3.88	.05	1.17	1.123	.160
<b>Frauen</b>	147	6.65	3.35	-.50	1.05	1.240	.092
Italien	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
<b>Gesamt</b>	150	7.43	3.71	-.26	1.13	.887	.411
<b>Männer</b>	77	8.09	3.89	-.07	1.17	.849	.466
<b>Frauen</b>	73	6.74	3.40	-.47	1.05	.924	.361
Österreich	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-s z	K-S p
<b>Gesamt</b>	149	7.75	3.79	-.17	1.16	.925	.359
<b>Männer</b>	75	8.91	3.89	.18	1.16	.888	.410
<b>Frauen</b>	74	6.57	3.32	-.52	1.05	.938	.343

**Tab.11.1.1:** Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Das Ergebnis des t-Tests für unabhängige Gruppen (Tab.11.1.2) zeigte bei Homogenität der Varianzen ( $p=.725$ ) ein nicht-signifikantes Ergebnis ( $p=.490$ ); bei zweiseitiger Fragestellung traten keine signifikanten Unterschiede zwischen den beiden Kulturen in der EST-Leistung auf.

Mit einem durchschnittlichen Rohscore von 7.43 (SD=3.71) war die Leistung der italienischen Stichprobe nicht wesentlich unterschiedlich von jener der österreichischen, mit einem Testleistungs-Mittelwert von 7.75 (SD=3.79). Damit kann die Hypothese, dass es keine Unterschiede zwischen den beiden untersuchten Kulturen gibt, wie erwartet, beibehalten werden.

	n	Levene Test		t-Test		
		F	Sig.	t	df	Sig. (2-seitig)
<b>Testleistung Italien vs. Österreich</b>	299	.124	.725	.691	297	.490
<b>Testleistung italienische vs. österreichische Männer</b>	152	.084	.772	1.287	150	.200
<b>Testleistung italienische vs. österreichische Frauen</b>	147	.048	.826	-.314	145	.754

**Tab.11.1.2.:** Ergebnisse der t-Tests für unabhängige Stichproben mit t-Werten (t), Freiheitsgraden (df) & p-Werten (Sig. (2-seitig)) sowie Ergebnisse des Levene-Test auf Varianzhomogenität mit F-Werten (F) und Signifikanz (Sig.).

Um differenzierte Aussagen tätigen zu können, wurden in einem weiteren Schritt die Geschlechter getrennt über beide Kulturen untersucht (Tab.11.1.2). Dabei zeigten sich keine signifikanten Unterschiede zwischen italienischen ( $M_{w_r}=8.09$ ,  $SD_r=3.89$ ) und österreichischen Männern ( $M_{w_r}=8.91$ ,  $SD_r=3.89$ ) hinsichtlich ihrer Leistung im EST. Selbes gilt für Frauen beider Kulturen; italienische Frauen ( $M_{w_r}=6.74$ ,  $SD_r=3.40$ ) unterschieden sich nicht signifikant von österreichischen Frauen ( $M_{w_r}=6.57$ ,  $SD_r=3.32$ ) (siehe Tab.11.1.1).

## 11.2. Geschlechtsspezifische Leistungsunterschiede

Auf Grund der zahlreichen Untersuchungen zu Geschlechtsdifferenzen in Raumvorstellungsleistungen, wurde auch in dieser Studie untersucht ob Männer und Frauen unterschiedliche Leistungen in der Bearbeitung des EST erbringen. Dieser Frage wird im Folgenden sowohl über die Kulturen als auch innerhalb jeder Kultur nachgegangen.

Wurden die Geschlechter unter Außerachtlassung ihrer Kulturzugehörigkeit verglichen, konnte ein Mittelwertsvergleich anhand eines t-Tests für unabhängige Stichproben durchgeführt werden, da in beiden Geschlechtern die mittels Kolmogorov-Smirnov-Test untersuchten Personenparameter hinreichend normalverteilt waren. In Tab.11.1.1 werden diese innerhalb jeder und über beide Kulturen angezeigt.

Die Ergebnisse der t-Tests für unabhängige Stichproben (Tab.11.2.1) ergaben (mit  $p<.001$  für die Gesamtstichprobe,  $p=.029$  für die italienische und  $p<.001$  für die österreichische Stichprobe) signifikante Ergebnisse.

	n	Levene Test		t-Test		
		F	Sig.	t	df	Sig. (2-seitig)
<b>Testleistung Männer vs. Frauen über beide Kulturen</b>	299	2.341	.127	4.274	297	<b>&lt;.001</b>
<b>Testleistung Männer vs. Frauen Italien</b>	150	1.127	.290	2.209	148	<b>.029</b>
<b>Testleistung Männer vs. Frauen Österreich</b>	149	.867	.353	3.850	147	<b>&lt;.001</b>

**Tab.11.2.1:** Ergebnisse des t-Test für unabhängige Stichproben mit t-Werten (t), Freiheitsgraden (df) & p-Werten (Sig. (2-seitig)) sowie Ergebnisse des Levene-Test auf Varianzhomogenität mit F-Werten (F) und Signifikanz (Sig.).

Genauer gesagt bedeutet dies, dass Männer beider Kulturen ( $M_{w_r}=8.49$ ;  $SD_r=3.89$ ) signifikant bessere Leistungen erbrachten als Frauen beider Kulturen ( $M_{w_r}=6.65$ ;  $SD_r=3.35$ ). Die geschätzte Effektstärke zeigt mit  $d=.51$  einen mittleren Effekt.

In der *italienischen Stichprobe* erzielten die männlichen Probanden ( $n=77$ ) signifikant bessere Ergebnisse ( $M_{w_r}=8.09$ ,  $SD_r=3.88$ ) als die Frauen ( $n=73$ ), die mit einem Rohscore-Mittelwert von 6.74 und einer Standardabweichung von 3.40 deutlich schlechter abschnitten ( $d=.37$ ). Kongruent dazu erreichten Männer in *Österreich* mit einem durchschnittlichen Rohscore von 8.91 ( $SD_r=3.89$ ) eine bessere Leistung im EST als die Frauen mit einem durchschnittlichen Wert von 6.57 ( $SD_r=3.32$ ). Die Schätzung der Effektstärke beläuft sich hierbei auf  $d=.65$  und weist auf einen mittleren Effekt hin. Somit konnte auch diese Untersuchung die Hypothese bestätigen, dass Männer über ein besseres Raumvorstellungsvermögen verfügen als Frauen.

Alle nachfolgenden Hypothesen werden im Anschluss über die Gesamtstichprobe und nicht mehr einzeln über die Kulturen berechnet, da für einen angemessenen Vergleich dieser, teilweise zu ungleiche Verteilungen vorliegen. Da bereits aufgezeigt werden konnte, dass sich die Kulturen hinsichtlich ihrer Raumvorstellungsfähigkeit nicht unterscheiden, wurde eine Auswertung über die Gesamtstichprobe als ausreichend erachtet.

## 11.3 Altersspezifische Leistungsunterschiede

Salthouse (2009) publizierte neueste Daten, in denen er einen Unterschied zwischen jüngeren und älteren Personen in ihren Raumvorstellungsfähigkeiten beschrieb. In Anlehnung an seine Ergebnisse soll auch hier untersucht werden, ob ein Unterschied in den EST-Leistungen in Abhängigkeit des Alters der Probanden aufgezeigt werden kann.

Da die Altersmittelwerte der beiden Stichproben auffällig weit auseinander liegen (siehe Tab.11.3.1), wurde ein statistischer Vergleich dieser durchgeführt.

### Deskriptive Statistik

Alter	Mw	SD	K-S z	p
Österreich	37.77	13.39	1.91	.001
Italien	33.93	14.26	2.63	<.001

Tab.11.3.1: Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Altersmittelwerte, von diesen werden jeweils Mittelwert (Mw) und Standardabweichung (SD) angegeben.

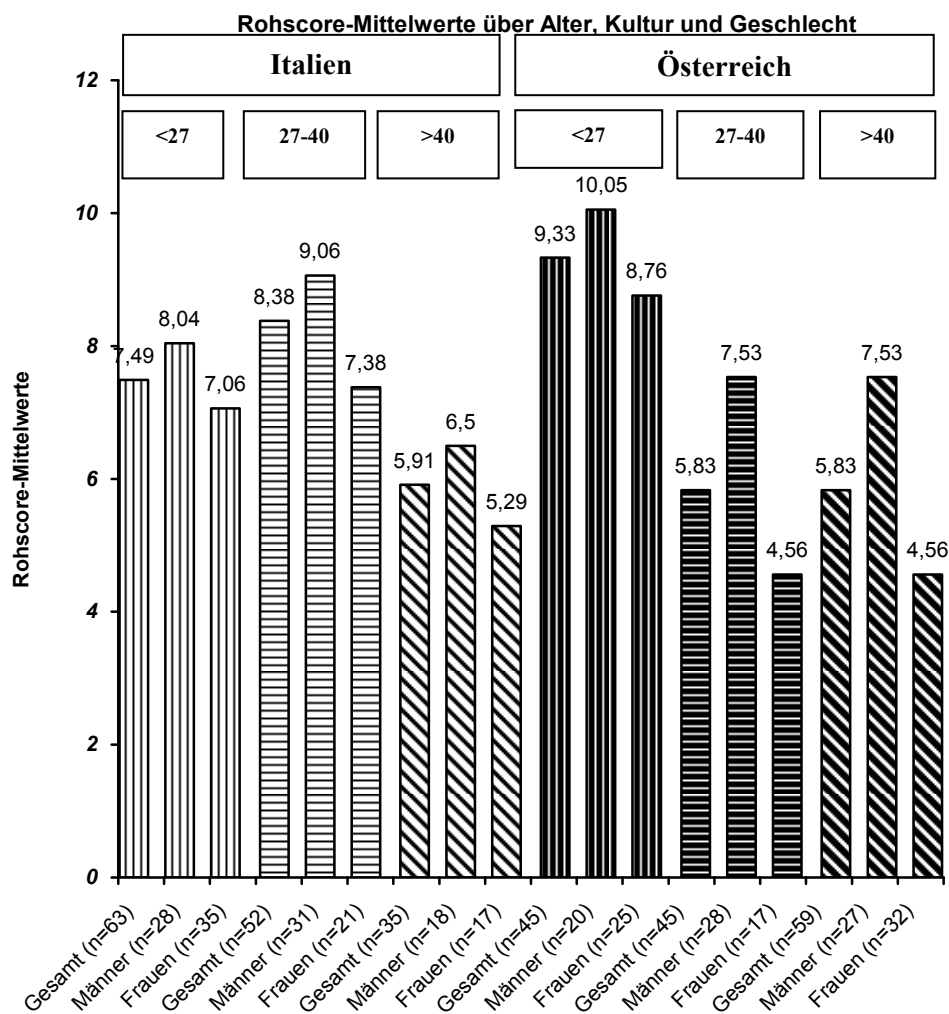
Basierend auf der nicht gegebenen Normalverteilung der Altersmittelwerte (Tab.11.3.1) in beiden Kulturen, wurde anhand des U-Tests von Mann und Whitney ein Mittelwertsvergleich vollzogen; sein Ergebnis zeigt, mit  $p=.003$ , dass, wie angenommen, signifikante Altersunterschiede zwischen den Kulturen vorliegen (Tab.11.3.2).

Dies bedeutet, dass die in der vorliegenden Arbeit vorhandenen Stichproben in unterschiedlichen Altersgruppen erhoben wurden. Da jedoch die Rasch-Modell Analyse des externen Kriteriums „Alter“ ein insignifikantes Ergebnis zeigte, und somit angenommen werden kann, dass der EST sowohl für jüngere als auch für ältere Probanden beider Kulturen dieselbe Fähigkeit erhebt, können diese Unterschiede in der weiteren Interpretation relativiert werden.

	Alter
Mann-Whitney-U	8976.000
Wilcoxon-W	20301.000
Z	-2.944
Asymptotische Signifikanz (2-seitig)	.003

Tab.11.3.2: Ergebnisse des U-Tests nach Mann und Whitney für die Altersunterschiede zwischen den Kulturen.

Für detaillierte Analysen hinsichtlich der altersspezifischen Leistungsunterschiede bei der Bearbeitung des EST wurden die Versuchspersonen in drei Altersklassen eingeteilt. Salthouse (2009) publizierte einen Leistungsabfall räumlicher Fähigkeiten mit zunehmendem Alter, welcher seinen Erkenntnissen zu Folge bereits ab einem Alter von 27 Jahren erkennbar sei. In Anlehnung an diese Ergebnisse erfolgte die Einteilung der Altersgruppen in Personen, die jünger als 27 Jahre alt waren; in eine zweite Gruppe, die aus Personen in einem Altersbereich von 27 bis 40 Jahren bestand und in eine dritte Gruppe von Personen, die älter als 40 Jahre waren. Die Versuchspersonen wurden ungeachtet ihrer Kultur in diese drei Gruppen geteilt. In Abb.11.3.1 wird die Verteilung der Rohscore-Mittelwerte dieser Gruppen getrennt für die beiden Kulturen und Geschlechter dargestellt.



**Abb.11.3.1:**Übersicht der Rohscore-Mittelwerte über die drei Altersgruppen getrennt für die österreichische und die italienische Stichprobe, sowie für Männer und Frauen.

Für die Gesamtstichprobe setzen sich die Altersgruppen folgendermaßen zusammen: die „jünger als 27 Jahre-Gruppe“ bestand aus insgesamt 108 Personen (davon 45 Österreicher und 63 Italiener) in einem Geschlechterverhältnis von 48 Männern zu 60 Frauen. Die Gruppe der 27 bis 40 Jährigen setzte sich aus 45 Österreichern und 52 Italienern (n=97) zusammen, davon 59 Männer und 38 Frauen. Die Gruppe jener Personen, die älter als 40 Jahre waren, bestand aus 59 Österreichern und 35 Italienern, davon waren 45 Testpersonen männlich und 49 weiblich.

### Deskriptive Statistik

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
<27 Gesamt	108	8.26	3.54	-.01	1.06	.968	.306
<27 Männer	48	8.88	3.52	.17	1.03	.743	.640
<27 Frauen	60	7.77	3.51	-.16	1.06	.723	.673
27-40 Gesamt	97	8.52	3.71	.08	1.10	.750	.627
27-40 Männer	59	9.32	3.97	.31	1.20	.794	.554
27-40 Frauen	38	7.26	2.89	-.28	.83	.709	.696
>40 Gesamt	94	5.86	3.45	-.76	1.10	1.09	.188
>40 Männer	45	7.00	3.83	-.41	1.15	.910	.379
>40 Frauen	49	4.82	2.69	-1.08	.95	1.19	.119

Tab.11.3.3: Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Für einen Leistungsvergleich zwischen den unterschiedlichen Altersgruppen wurde eine Kovarianzanalyse mit der Variable „Geschlecht“ als Kovariate herangezogen<sup>15</sup> (deren Voraussetzungen können als gegeben angesehen werden: Varianzhomogenität  $F_{(2,296)}=.104$ ,  $p=.901$ ; Homogenität der Regressionsgeraden (engl.: *regression slopes*)  $F_{(2)}=.737$ ,  $p=.479$ ; Normalverteilung der Testleistungen in allen Gruppen (siehe Tab.11.3.3); die Beobachtungen sind unabhängig und die abhängige Variable ist intervallskaliert).

Das Ergebnis der Kovarianzanalyse (Tab.11.3.4) zeigt, dass das Geschlecht der Probanden einen signifikanten Einfluss auf die Testleistung hatte und dass unter Kontrolle dieses Einflusses die Gruppen unterschiedliche Leistungen in Abhängigkeit ihres Alters erbrachten ( $p=.002$ ).

<sup>15</sup> Die Kovarianzanalyse (ANCOVA) dient einem Vergleich mehrerer Mittelwerte und kontrolliert zusätzlich die Effekte einer oder mehrerer anderer Variablen (Kovariaten). Ist das Ergebnis für eine Kovariate signifikant, kann dies dahingehend gedeutet werden, dass sie einen Einfluss auf die abhängige Variable hat. Zeigt auch die unabhängige Variable ein signifikantes Ergebnis, unterscheiden sich die Mittelwerte der Gruppen unter Kontrolle des Einflusses der Kovariate(n) auf die AV (Bortz, 1999).

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
<b>Korrigiertes Modell</b>	61.609	5	12.322	11.040	<b>&lt;.001</b>
<b>Intercept</b>	.084	1	.084	.075	.784
<b>Alter 3 Gruppen</b>	14.005	2	7.003	6.274	<b>.002</b>
<b>Geschlecht</b>	20.417	1	20.417	18.293	<b>&lt;.001</b>
<b>Alter*Geschlecht</b>	1.645	2	.822	.737	.479
<b>Fehler</b>	327.019	293	1.116		
<b>Gesamt</b>	402.714	299			
<b>Korrigierte Gesamtvariation</b>	388.628	299			

Tab.11.3.4: Ergebnisse der Kovarianzanalyse mit „Alter“ als unabhängige Variable, „Leistungen der Personen im EST“ als abhängige Variable (berechnet mittels Personenparametern) und „Geschlecht“ als Kovariate

Dabei unterschied sich, wie aus Tab.11.3.5 ersichtlich, die Leistung jener Personen, die über 40 Jahre alt waren signifikant von der Leistung der anderen beiden Altersgruppen. Personen über 40 erzielten deutlich schlechtere Leistungen im EST ( $M_{wT}=5.86$ ,  $SD_T=3.45$ ) als die unter 40 Jährigen.

(I) Alter	(J) Alter	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95% Konfidenzintervall für die Differenz	
					Untergrenze	Obergrenze
<27	27-40	-.010	.150	1.000	-.370	.350
	>40	.744	.150	<b>&lt;.001</b>	.384	1.103
27-40	<27	.010	.150	1.000	-.350	.370
	>40	.754	.155	<b>&lt;.001</b>	.383	1.125
>40	<27	-.744	.150	<b>&lt;.001</b>	-1.103	-.384
	27-40	-.754	.155	<b>&lt;.001</b>	-1.125	-.383

Abb.11.3.5: Paarweise Vergleiche für die drei Altersgruppen. Die Signifikanz beruht auf der Anpassung für Mehrfachvergleiche nach Sidak.



## 11.4 Bildungsspezifische Leistungsunterschiede

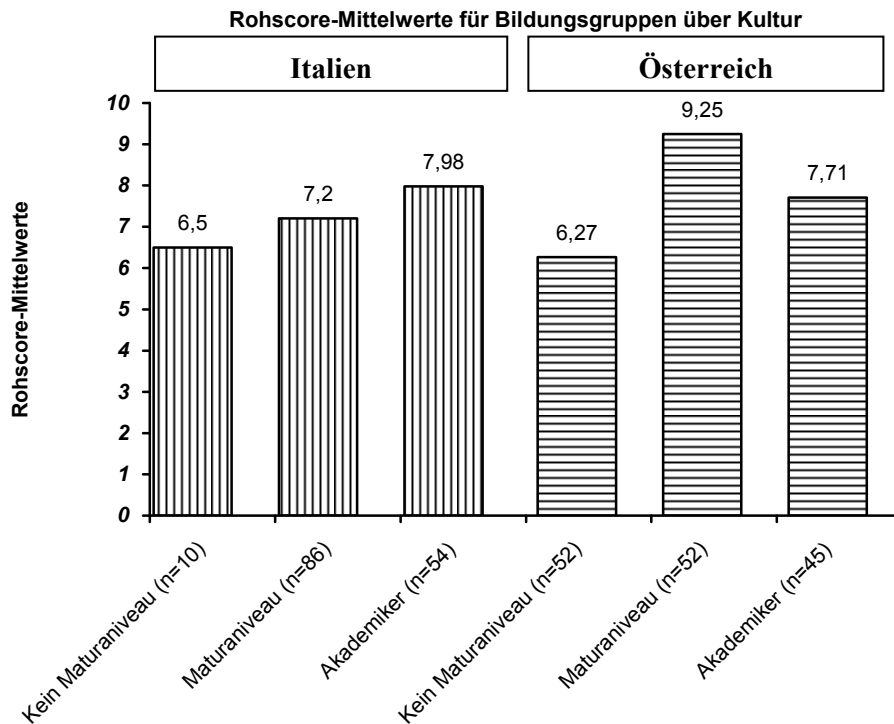
Eine weitere Hypothese dieser Arbeit ging der Frage nach ob der Grad der höchsten abgeschlossenen Schulbildung einen Einfluss auf die Leistungen im EST hat. Auf Grund der stark unterschiedlichen Zellbesetzungen in den erhobenen Bildungsgruppen (siehe Tab.11.4.1) wurden diese zu drei übergeordneten Kategorien zusammengefasst.

### Deskriptive Statistik

Bildung	Gesamtstichprobe		Italien		Österreich	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Volksschule	3	1.0	2	1.3	1	.7
Hauptschule	13	4.3	7	4.7	6	4
Lehre	46	15.4	1	.7	45	30.2
Matura	138	46.2	86	57.3	52	34.9
Akademie, FH	13	4.3	6	4	7	4.7
Universität	86	28.8	48	32	38	25.5
<b>Gesamt</b>	<b>299</b>	<b>100.0</b>	<b>150</b>	<b>100.0</b>	<b>149</b>	<b>100.0</b>

Tab.11.4.1: Bildungsverteilung in den ursprünglich erhobenen Gruppen der Gesamtstichprobe.

Jene Personen deren höchste abgeschlossene Schulbildung „Volksschule“, „Hauptschule“ oder „Lehre“ war, wurden zu der Gruppe „kein Maturaniveau“ (n=62) zusammengefasst. Die Personen der Gruppe „Matura“ wurden in dieser Form beibehalten und werden in der neuen Aufteilung unter der Bezeichnung „Maturaniveau“ weitergeführt (n=138). Jene Personen, die über den Abschluss einer Universität, Fachhochschule oder Akademie verfügten, wurden zur Gruppe „Akademiker“ zusammengefasst (n=99). Abb.11.4.1 zeigt die Verteilung der Rohscore-Mittelwerte der einzelnen Bildungsgruppen dieser neuen Aufteilung getrennt für die beiden untersuchten Kulturen.



**Abb.11.4.1:** Rohscore-Mittelwerte für die Bildungsgruppen in der Gesamtstichprobe, sowie getrennt für Männer und Frauen.

Die gegebene Normalverteilung in allen Gruppen (Tab.11.4.2) ermöglichte einen Vergleich der Mittelwerte über die Kulturen anhand einer univariaten Varianzanalyse (Varianzhomogenität bei  $F_{(2, 296)}=.630, p=.533$  gegeben).

### Deskriptive Statistik

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
Kein Maturaniveau	62	6.31	3.60	-.64	1.18	.674	.754
Maturaniveau	138	7.97	3.59	-.10	1.07	.856	.456
Akademiker	99	7.86	3.91	-.12	1.17	.782	.574

**11.4.2:** Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Die univariate Varianzanalyse ergab ein signifikantes Resultat für den Haupteffekt der unterschiedlichen Bildungsgruppen (Tab.11.4.3). Anhand der Mehrfachvergleiche nach Scheffé (Tab.11.4.4) ist ersichtlich, dass sich die Leistung jener Gruppe, die über kein Maturaniveau verfügte, deutlich von jener der anderen beiden Bildungsgruppen unterschied. Mit einem Rohscore-Mittelwert von 6.31 (SD<sub>r</sub>=3.60) schnitten diese Personen deutlich schlechter ab als jene mit Maturaniveau (Mw<sub>r</sub>=7.97, SD<sub>r</sub>=3.59) und die Gruppe der

Akademiker ( $MW_r=7.86$ ,  $SD_r=3.91$ ). Die beiden letztgenannten Gruppen unterschieden sich nicht hinsichtlich ihrer Testleistung im EST.

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	13.915	2	6.957	5.496	.005
Intercept	21.917	1	21.917	17.313	<.001
Bildung 3 Gruppen	13.915	2	6.957	5.496	.005
Fehler	374.713	296	1.266		
Gesamt	402.714	299			
Korrigierte Gesamtvariation	388.628	298			

Tab.11.4.3: Ergebnisse der Varianzanalyse mit „drei Bildungsgruppen“ als unabhängige Variable und „Leistungen der Personen im EST“ als abhängige Variable (berechnet mittels Personenparametern).

		Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95% Konfidenzintervall für die Differenz	
(I) Bildungsgruppen	(J) Bildungsgruppen				Untergrenze	Obergrenze
Kein Maturaniveau	Maturaniveau	-.5430	.1720	.007	-.9662	-.1198
	Akademiker	-.5148	.1822	.019	-.9631	-.0664
Maturaniveau	Kein Maturaniveau	.5430	.1720	.007	.1198	.9662
	Akademiker	.0282	.1482	.982	-.3363	.3928
Akademiker	Kein Maturaniveau	.5148	.1822	.019	.0664	.9631
	Maturaniveau	-.0282	.1482	.982	-.3928	.3363

Tab.11.4.4: Mehrfachvergleiche für die drei Bildungsgruppen. Die Signifikanz beruht auf der Anpassung für Mehrfachvergleiche nach Scheffé.

## 11.5 Leistungsunterschiede auf Grund vorausgehender Erfahrungen mit darstellender Geometrie

Eine weitere Fragestellung, die in dieser Arbeit interessierte, bezieht sich auf die Frage des Einflusses von Erfahrungen mit darstellender Geometrie auf die räumliche Vorstellungsfähigkeit. Im Folgenden wird diese Hypothese in der Gesamtstichprobe überprüft.

Der Vergleich der Raumvorstellungsleistungen von Testpersonen, die Erfahrungen mit DG hatten ( $MW_r=8.76$ ,  $SD_r=4.07$ ) und jenen, die über keine Erfahrung berichteten ( $MW_r=7.03$ ,  $SD_r=3.46$ ), wurde unter gegebenen Voraussetzungen - Normalverteilung der Daten (siehe Tab.11.5.1), Varianzhomogenität ( $F_{(1,197)}=1.626$ ,  $p=.203$ ), Homogenität der

Regressionsgeraden ( $F_{(1)}=3.530$ ,  $p=.087$ ) - unter Außerachtlassung der Kulturzugehörigkeit der Probanden, anhand einer Kovarianzanalyse berechnet<sup>16</sup>.

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
DG nein gesamt	203	7.03	3.46	-.37	1.06	1.149	.142
DG nein Männer	91	7.66	3.66	-.19	1.10	.761	.609
DG nein Frauen	112	6.53	3.21	-.52	1.00	1.229	.097
DG ja gesamt	96	8.76	4.07	.11	1.25	1.167	.131
DG ja Männer	61	9.74	3.93	.41	1.18	.985	.287
DG ja Frauen	35	7.06	3.78	-.41	1.21	.662	.773

Tab.11.5.1: Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Um den Einfluss der Erfahrung mit DG auf die Leistung im Raumvorstellungstest zu überprüfen, wurde der Einfluss der Variable „Geschlecht“ konstant gehalten. Die Ergebnisse der Kovarianzanalyse zeigen, dass das Geschlecht der Probanden einen signifikanten Einfluss auf die Leistungen im EST hatte ( $p<.001$ ). Unter Kontrolle dieses Einflusses zeigte sich ein signifikantes Ergebnis für die Variable „Erfahrung mit darstellender Geometrie“ (Tab.11.5.2), wobei jene Gruppe, die über Erfahrung mit DG verfügte deutliche bessere Testleistungen erzielte als die andere Gruppe ( $p=.001$ ).

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	35.700	3	11.900	9.947	<.001
Intercept	1.810	1	1.810	1.513	.220
DG	12.837	1	12.837	10.730	.001
Geschlecht	20.559	1	20.559	17.184	<.001
DG*Geschlecht	3.530	1	3.530	2.951	.087
Fehler	352.928	295	1.196		
Gesamt	402.714	299			
Korrigierte Gesamtvariation	388.628	298			

Tab.11.5.2: Ergebnisse der Kovarianzanalyse mit „Erfahrung in DG“ als unabhängige Variable, „Leistungen der Personen im EST“ als abhängige Variable (berechnet mittels Personenparametern) und „Geschlecht“ als Kovariate

<sup>16</sup> die ungleichen Stichprobengrößen der zu vergleichenden Gruppen können anhand der gegebenen Voraussetzungen unberücksichtigt bleiben (Levy, 1980, nach Bortz, 1999).

Abb. 11.5.1 zeigt die Rohscore-Mittelwerte der beiden DG-Gruppen getrennt für die Kulturen und das Geschlecht.

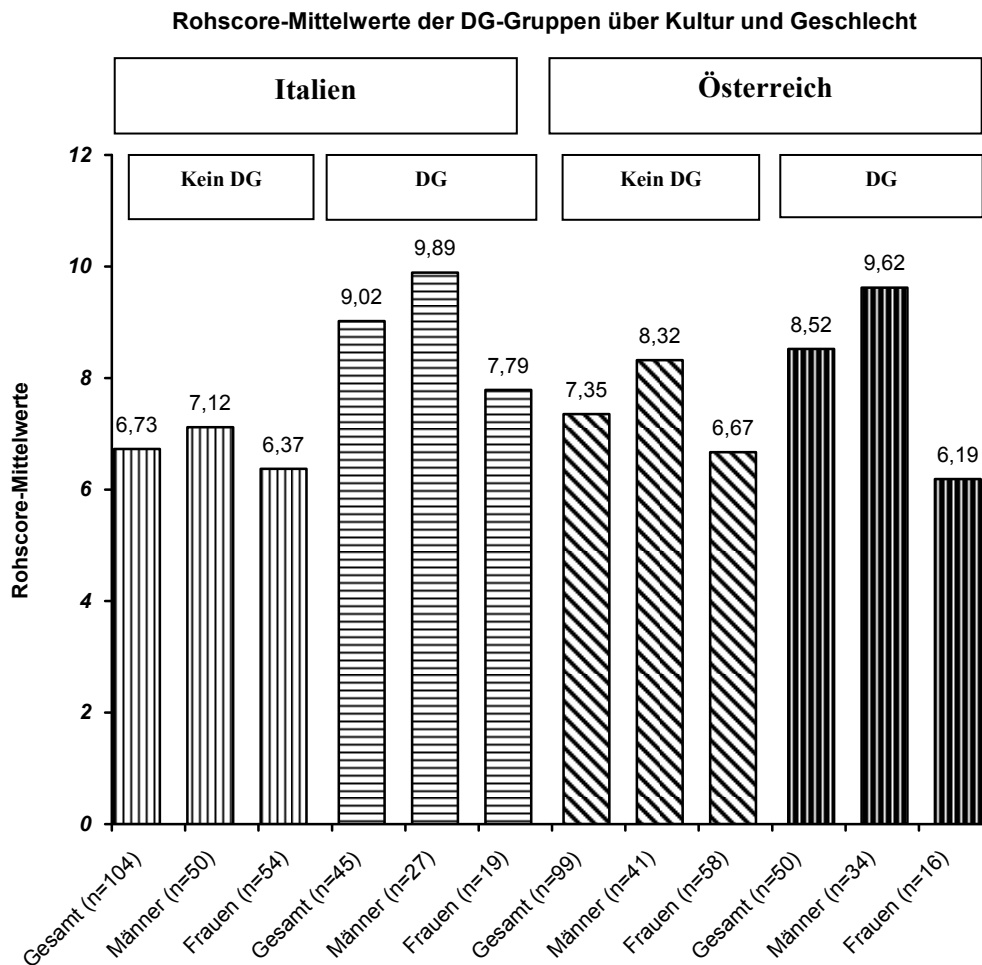


Abb.11.5.1: Rohscore-Mittelwerte der DG-Gruppen für Italien und Österreich sowie getrennt für das Geschlecht.

## 11.6 Der Einfluss der Handpräferenz auf die Leistung im EST

Unterschiedliche Leistungen in Raumvorstellungstests in Abhängigkeit der Handpräferenz wurden in der Literatur häufig berichtet, die Ergebnisse dazu sind allerdings kontrovers. Manche Studien konnten zeigen, dass Linkshänder zu besseren Raumvorstellungsleistungen tendieren als Rechtshänder (Annett, 2002, nach Peters et al., 2006), während andere Wissenschaftler einen Zusammenhang zwischen höherer Raumvorstellungsleistung und Rechtshändigkeit feststellten (McGee, 1978).

Der folgende Absatz beschäftigt sich mit diesem Phänomen in der Gesamtstichprobe. Die Ergebnisse sind allerdings nur als richtungsweisende Tendenzen zu verstehen, da die links-,

und beidhändigen Personen in dieser Studie stark unterrepräsentiert waren und ihre Anzahl als zu klein für eine kausale Interpretation anzusehen ist.

Die Gesamtstichprobe setzt sich aus 262 Rechtshändern, 17 Linkshändern sowie 20 Personen, die beide Hände gleichermaßen für die Aufgaben des Alltags verwenden, zusammen.

Die Verteilung der Personenparameter war, wie aus Tab.11.6.1 ersichtlich, für alle Händigkeitgruppen normal; auf Grund der extrem ungleichen Stichprobengrößen wurde jedoch der Vergleich der Gruppen in Bezug auf ihre Leistungen im EST anhand des Kruskal-Wallis-Test durchgeführt.

### Deskriptive Statistik

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
Rechtshändige Personen	262	7.53	3.73	-.24	1.14	1.181	.123
Linkshändige Personen	17	9.41	3.92	.32	1.23	.687	.732
Beidhändige Personen	20	6.85	3.45	-.39	1.03	.751	.625

Tab.11.6.1: Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Bei einer Untersuchung der Probanden unter Außerachtlassung ihrer Kulturzugehörigkeit ergab der Kruskal-Wallis-Tests bei zweiseitiger Fragestellung keine signifikanten Unterschiede in der Leistung des EST in Abhängigkeit der Handpräferenz einer Person (Tab. 11.6.2).

Kruskal-Wallis-Test				
		Händigkeitsgruppe	n	Mittlerer Rang
Chi-Quadrat	5.218	Rechtshänder	262	73.77
df	2	Linkshänder	17	110.20
Asymptotische Signifikanz	.074	Beidhändig	20	64.27

Tab.11.6.2: Ergebnisse des Kruskal-Wallis Test (Testgröße  $\chi^2$ ), Freiheitsgrade (df) und Signifikanz ( $\alpha=.05$ ); ebenso werden die mittleren Ränge, sowie die jeweilige Probandenanzahl (n) pro Händigkeitsgruppe angegeben.

## 11.7 Bearbeitungszeitspezifische Leistungsunterschiede

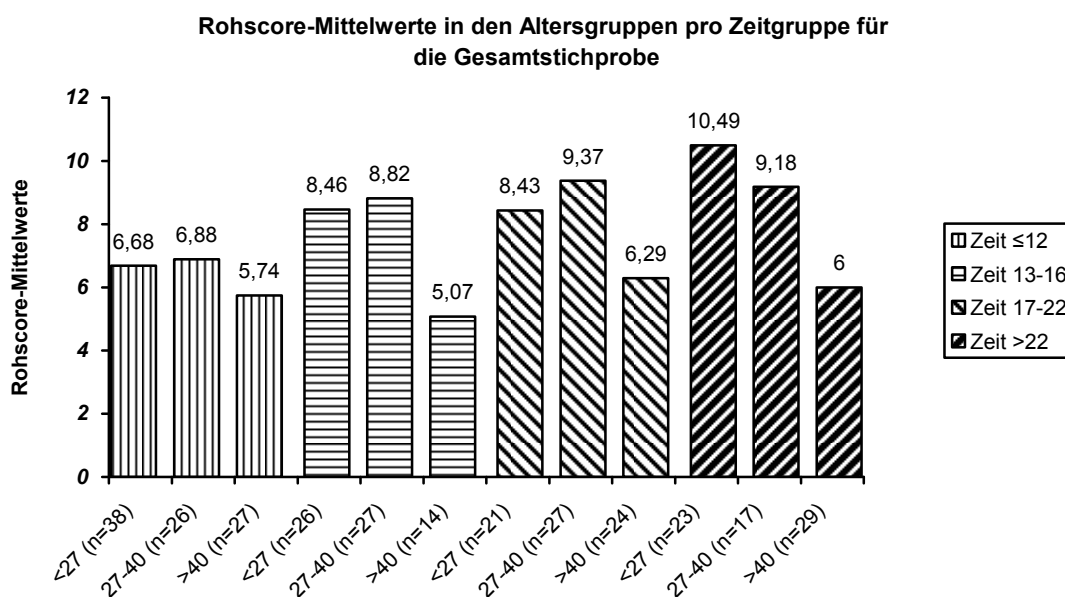
Eine weitere Fragestellung dieser Arbeit beschäftigte sich mit dem Einfluss der Zeit, die Personen für die Bearbeitung des EST aufwendeten, auf ihre Leistung. In diesem Zusammenhang soll untersucht werden ob sich Unterschiede in der Leistung der Versuchspersonen in Abhängigkeit zu der von ihnen aufgewandten Bearbeitungszeit zeigen.

Die durchschnittliche Bearbeitungszeit des EST der Gesamtstichprobe lag bei einem Mittelwert von 17.94 Minuten (SD=8.81).

Um genaue Aussagen über eventuelle Unterschiede zwischen Probanden tätigen zu können, die sich unterschiedlich lang mit der Bearbeitung des EST Zeit ließen, wurde die Testbearbeitungszeit, anhand der Quartile der Gesamtstichprobe in vier Gruppen aufgeteilt.

Die vier Gruppen wurden in Personen, die 12 Minuten oder kürzer (n=91, davon 42 Männer und 49 Frauen), in Personen, die länger als 12 oder kürzer gleich 16 Minuten benötigten (n=67, davon 33 Männer und 34 Frauen), in eine Gruppe, deren Probanden den EST länger als 16 Minuten und kürzer oder gleich 22 Minuten bearbeiteten (n=72, davon 39 Männer und 33 Frauen) und in eine Gruppe, die länger als 22 Minuten zur Testbearbeitung aufwandte (n=69, davon 38 Männer und 31 Frauen) geteilt.

Abb.11.7.1 bietet einen Überblick über die Rohscore-Mittelwerte der Altersgruppen über die Bearbeitungszeitgruppen der Gesamtstichprobe.



**Abb.11.7.1:** Rohscore-Mittelwerte der verschiedenen Altersgruppen pro Bearbeitungszeitgruppe. Die Werte in der Abbildung repräsentieren die Gesamtstichprobe.

## Deskriptive Statistik

Gesamtstichprobe	n	Mw <sub>r</sub>	SD <sub>r</sub>	Mw <sub>par</sub>	SD <sub>par</sub>	K-S z	K-S p
≤ 12 Minuten	91	6.46	3.32	-.55	1.02	.676	.750
13-16 Minuten	67	7.90	3.98	-.15	1.26	.612	.848
17-22 Minuten	72	8.07	3.61	-.05	1.06	.888	.409
>22 Minuten	69	8.28	3.92	-.03	1.19	1.105	.174

Tab.11.7.1: Kolmogorov-Smirnov-Test (z- und p-Werte) zur Prüfung der Normalverteilung der Personenparameter, von diesen werden jeweils Mittelwert (Mw<sub>par</sub>) und Standardabweichung (SD<sub>par</sub>) angegeben; ebenso Mittelwert und Standardabweichung der Rohscores (Mw<sub>r</sub> und SD<sub>r</sub>).

Die Normalverteilung der Personenparameter sowie die Anzahl der Probanden pro Untergruppe in der Gesamtstichprobe können in Tab.11.7.1 abgelesen werden. Diese Werte gemeinsam mit der Erfüllung der weiteren Voraussetzungen (Varianzhomogenität ( $F_{(3,295)}=.760$ ,  $p=.518$ ), Homogenität der Regressionsgeraden ( $F_{(4)}=1.633$ ,  $p=.166$ )) ermöglichten eine Untersuchung anhand einer Kovarianzanalyse, deren Resultate in Tab.11.7.2 dargestellt sind. Da der Einfluss der Bearbeitungszeit auf die Leistungen im EST, frei von Effekten des Geschlechts und des Alters<sup>17</sup> der Probanden, untersucht werden sollte, wurde der Einfluss dieser beiden Variablen konstant gehalten, indem sie als Kovariaten in die Analyse eingingen. Die Ergebnisse zeigen, dass in der Gesamtstichprobe bei zweiseitig formulierter Fragestellung das Alter einen signifikanten Einfluss auf die Testleistung im EST hatte ( $p=.016$ ) und dass Unterschiede in den verschiedenen Bearbeitungszeitgruppen bezüglich der Leistung im EST auftraten ( $p<.001$ ).

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	73.272	9	8.141	7.461	<.001
Intercept	6.168	1	6.168	5.653	.018
Bearbeitungszeitgruppen	20.351	3	6.784	6.217	<.001
Geschlecht	2.526	1	2.526	2.314	.129
Alter	6.393	1	6.393	5.858	.016
Bearbeitungszeitgruppen*Geschlecht*Alter	7.128	4	1.782	1.633	.166
Fehler	315.356	289	1.091		
Gesamt	402.714	299			
Korrigierte Gesamtvariation	388.628	298			

Tab. 11.7.2: Ergebnisse der Kovarianzanalyse mit „Bearbeitungszeitgruppen“ als unabhängige Variable, „Leistungen der Personen im EST“ als abhängige Variable (berechnet mittels Personenparametern), „Geschlecht“ und „Alter“ als Kovariaten

Dabei erzielte jene Gruppe, die weniger als 12 Minuten zur Testbearbeitung aufwendete deutlich schlechtere Leistungen (Mw<sub>r</sub>=6.46, SD<sub>r</sub>=3.32) als die Gruppe jener Personen, die 17 bis 22 Minuten (Mw<sub>r</sub>=8.07, SD<sub>r</sub>=3.61) oder länger (Mw<sub>r</sub>=8.28, SD<sub>r</sub>=3.92) dafür in Anspruch

<sup>17</sup> Die Kontrolle eines eventuellen Einflusses von Alter wurde gewählt da mit zunehmendem Alter eine langsamere Testbearbeitungszeit zu erwarten wäre.



nahmen. Zwischen den anderen Bearbeitungszeitgruppen zeigte sich jedoch kein Unterschied (siehe Tab.11.7.3).

(I) Bearbeitungszeit- gruppe	(J) Bearbeitungszeit- gruppe	Mittlere Differenz (I- J)	Standard- fehler	Signifikanz	95% Konfidenzintervall für die Differenz	
					Untergrenze	Obergrenze
≤12	13-16	-.383	.171	.145	-.836	-.070
	17-22	-.533	.168	<b>.010</b>	-.978	-.089
	>22	-.560	.169	<b>.006</b>	-1.007	-.113
13-16	≤12	.383	.171	.145	-.070	.836
	17-22	-.150	.178	.954	-.623	.322
	>22	-.177	.181	.909	-.656	.302
17-22	≤12	.533	.168	<b>.010</b>	.089	.978
	13-16	.150	.178	.954	-.322	.623
	>22	-.027	.177	1.000	-.494	.441
>22	≤12	.560	.169	<b>.006</b>	.113	1.007
	13-16	.177	.181	.909	-.302	.656
	17-22	.027	.177	1.000	-.441	.494

Tab.11.7.3: Paarweise Vergleiche für die vier Bearbeitungszeitgruppen. Die Signifikanz beruht auf der Anpassung für Mehrfachvergleiche nach Sidak.

Eine letzte Fragestellung, die in dieser Arbeit interessant scheint, bezieht sich darauf ob sich die Geschlechter in der durchschnittlich von ihnen aufgewandten Testbearbeitungszeit unterscheiden. Dies wurde von diversen Forschern postuliert (Geiser et al., 2006; Goldstein et al., 1990; Maccoby & Jacklin, 1974). Um diese Hypothese zu untersuchen, wurden die Geschlechter hinsichtlich verschiedener Bearbeitungszeiten anhand einer univariaten Varianzanalyse untersucht.

Die Ergebnisse aus Tab.11.7.4 zeigen, dass das Geschlecht keinen signifikanten Einfluss ( $p=.208$ ) auf die Bearbeitungszeit im EST hatte. Dies bedeutet, dass Männer und Frauen durchschnittlich gleich viel Zeit zur Bearbeitung des EST aufwandten.

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	2.093	1	2.093	1.595	.208
Intercept	583.029	1	583.029	444.516	<.001
Geschlecht	2.093	1	2.093	1.595	.208
Fehler	389.546	297	1.312		
Gesamt	976.000	299			
Korrigierte Gesamtvariation	391.639	298			

Tab.11.7.4: Ergebnisse der univariaten Varianzanalyse mit „Geschlecht“ als unabhängige Variable und „Bearbeitungszeitgruppen“ als abhängige Variable (berechnet mittels Personenparametern).

## 12. Interpretation und Diskussion

Die bedeutendsten Fragestellungen dieser Arbeit beschäftigten sich mit der Überprüfung des Endlosschleifentests (EST) auf Rasch-Homogenität. Von Interesse war dabei, ob diese erneut für eine österreichische, sowie erstmals für eine italienische Stichprobe und über beide Kulturen gemeinsam gezeigt werden kann.

In der *italienischen Stichprobe* (n=152) wurden die EST-Daten nach Ausschluss zweier Testpersonen (wegen unseriöser Testbearbeitung - bereinigtes n=150) mittels LPCM-win 1.0 auf Rasch-Homogenität untersucht.

Da die Eindimensionalität des Messinstruments jedoch mit der ursprünglichen Version von 21 Items (ein Item wurde als „warming-up Aufgabe“ von vornherein nicht in die Analysen inkludiert) durch ein grenzwertiges Ergebnis für das interne Kriterium (hohe vs. niedrige Testleistung) gefährdet war, wurde anhand theoretischer Überlegungen die Test-Eingangsphase um zwei Items und die Test-Endphase um ein Item verlängert.

Erneute Modelltests der verbleibenden 18 Aufgaben ergaben Rasch-Homogenität innerhalb der italienischen Stichprobe. Somit kann der EST in dieser, als faires Testinstrument angesehen werden, da er für alle Probanden dieselbe latente Fähigkeitsdimension der Raumvorstellung erhebt.

Auch aus der *österreichischen Stichprobe* (n=150) musste eine Person wegen unseriöser Testbearbeitung ausgeschlossen werden. Der EST zeigte mit den ursprünglichen 21 Items keine Rasch-Homogenität (n=149), da der Modelltest des externen Kriteriums „Bearbeitungszeit“ signifikant ausfiel ( $p=.006$ ).

Nach demselben Vorgehen wie in der italienischen Stichprobe wurden auch hier Test-Eingangs- und Test-Endphase verlängert und erneute Analysen anhand von k=18 Items durchgeführt. Mit dieser veränderten Item-Anzahl zeigten alle Modelltests, wie erwartet, insignifikante Ergebnisse und die Erfassung derselben latenten Fähigkeitsdimension für alle Probanden ist somit in der österreichischen Stichprobe erneut gewährleistet.

Die gegebene Rasch-Homogenität innerhalb beider Kulturen ermöglichte es, in einem weiteren Schritt zu überprüfen, ob der EST auch über beide Kulturen gemeinsam dieselbe

latente Fähigkeitsdimension erfasst und dem zu Folge für inferenzstatistische Vergleiche dieser in leistungsbezogenen Variablen herangezogen werden konnte. Die insignifikanten Ergebnisse aller Modelltests bestätigten dies.

Nachdem diese Voraussetzung überprüft und die Rasch-Homogenität angenommen werden konnte, war die Basis für weitere statistische Analysen gewährleistet. Diese erfolgten nicht anhand der Rohscores der Testpersonen (Anzahl korrekt gelöster Items) sondern auf Grund erhöhter Präzision durch die *Weighted Likelihood Estimates* (WLEs).

Die Untersuchung etwaiger *leistungsspezifischer Unterschiede* zwischen den beiden Kulturen zeigte, kongruent mit den im Vorfeld getätigten Annahmen, keine signifikanten Ergebnisse. Dies bedeutet, dass keine der beiden Kulturen wesentlich bessere Leistungen in der Bearbeitung des EST zeigte als die andere. Auch ein detaillierter Vergleich der Geschlechter über die Kulturen wies keine Unterschiede auf. Männer beider Kulturen erreichten vergleichbare Testleistungen, gleiches gilt für die Frauen.

Die Frage, ob es *geschlechtsspezifische Leistungsunterschiede* der Probanden im EST gibt, konnte in allen drei Stichproben bestätigt werden.

Sowohl über beide Kulturen als auch innerhalb der jeweiligen Kultur zeigten Männer signifikant bessere Leistungen als Frauen. Diese Ergebnisse stehen in Einklang mit zahlreichen in dieser Arbeit (Kapitel 5) angeführten Studien, welche eine männliche Dominanz in Raumvorstellungsleistungen publizierten.

Zu der Frage warum dies der Fall ist, kann jedoch in dieser Arbeit nur bedingt Aufschluss gegeben werden. Die in Kapitel 5.2 angeführte Theorie, dass geschlechtsspezifische Leistungsunterschiede auf verschiedene Bearbeitungsstrategien von Männern und Frauen zurückzuführen sind, (Hirnstein et al., in press) kann als Erklärung ausgeschlossen werden, da der EST dahingehend konstruiert ist, dass raumvorstellungsrelevante Strategien zur Bearbeitung notwendig sind, während verbal-analytische Strategien nicht zu einer geeigneten Lösung führen. Auch die Untersuchung einer eventuell unterschiedlichen Bearbeitungszeit zwischen Männern und Frauen zeigte insignifikante Ergebnisse. Dies bedeutet, dass eine unterschiedliche Testbearbeitungszeit (Geiser et al., 2006; Goldstein et al., 1990; Maccoby &

Jacklin, 1974) nicht als Erklärung für die geschlechtsspezifischen Leistungsunterschiede herangezogen werden kann.

Eine mögliche Erklärung könnten hingegen geschlechtsspezifische Faktoren im Sozialisationsprozess darstellen, wie etwa der unterschiedliche Umgang der Eltern mit ihrem weiblichen und männlichen Nachwuchs in Hinsicht auf verschiedene Erziehungsstile. Ebenso können umweltbedingte Faktoren, die der unterschiedlichen Raumvorstellungsfähigkeit der Geschlechter zu Grunde liegen als Interpretationsgrundlage dieser Funde angenommen werden (Casey et al., 1999).

Am wahrscheinlichsten ist jedoch ein Zusammenwirken all dieser Faktoren im Sinne eines bio-psycho-sozialen Erklärungsmodells der geschlechterspezifischen Leistungsunterschiede, nicht nur im Bereich der Raumvorstellung sondern auch in anderen Bereichen des Lebens und in verschiedenen Kulturen. Obwohl in diesem Bereich der Forschung bereits sehr viele Ergebnisse vorliegen, könnten hier zukünftige Untersuchungen, im Sinne eines integrativen Ansatzes nützlich sein, um eventuell noch präzisere kausale Aussagen über die Wechselwirkungen all dieser Faktoren zu ermöglichen.

Alle nachfolgenden Hypothesen wurden im Anschluss über die Gesamtstichprobe und nicht mehr einzeln über die Kulturen berechnet, da für einen angemessenen Vergleich dieser, teilweise zu ungleiche Zellbesetzungen vorlagen. Da jedoch bereits aufgezeigt werden konnte, dass sich die Kulturen hinsichtlich ihrer Raumvorstellungsfähigkeit nicht unterscheiden, wurde eine Auswertung über die Gesamtstichprobe als ausreichend erachtet.

In Bezug auf einen eventuellen *Einfluss des Alters* der Probanden auf die EST-Leistung ergaben sich in der Gesamtstichprobe signifikante Unterschiede in Abhängigkeit des Alters. Dieses Resultat zeigt sich teilweise kongruent mit den Ergebnissen von Salthouse (2009), der einen Leistungsabfall der Raumvorstellungsfähigkeit mit zunehmendem Alter feststellte. Die von Salthouse (2009) publizierte Verringerung räumlicher Fähigkeiten beginnend mit 27 Jahren konnte in dieser Untersuchung jedoch nicht repliziert werden; jene Probanden die jünger als 27 Jahre alt waren, unterschieden sich nicht signifikant von jenen die sich in einem Altersbereich zwischen 27 und 40 Jahren befanden.

Hinsichtlich der höchsten abgeschlossenen Schulbildung der Probanden zeigten sich signifikante *bildungsspezifische Leistungsunterschiede*. Auf Grund einer Unterrepräsentation in den verschiedenen Bildungsgruppen wurden „Volksschule“, „Hauptschule“ und „Lehre“ zu der Gruppe „kein Maturaniveau“ zusammengefasst; Probanden die als höchste abgeschlossene Schulbildung „Matura“ angaben, wurden in dieser Gruppe beibehalten. Jene Personen, die über den Abschluss einer „Universität“, „Fachhochschule“ oder „Akademie“ verfügten, wurden zu der Gruppe „Akademiker“ zusammengefasst. Dabei zeigte sich ein signifikanter Unterschied zwischen den Personen der Gruppe „kein Maturaniveau“ und den anderen beiden Gruppen. Erstere erzielte dabei durchschnittlich schlechtere EST-Leistungen als die anderen beiden.

Für zukünftige Untersuchungen sei jedoch angemerkt, dass bei der Datenerhebung verstärkt auf eine gleichmäßige Bildungsverteilung in den Kulturen geachtet werden sollte. Diese Arbeit erlaubte es auf Grund der geringen Probandenzahl in den Bildungsstufen „Volksschule“, „Hauptschule“ und „Lehre“ nicht, detaillierte Bildungsunterschiede zwischen diesen drei Gruppen zu untersuchen.

Die Beantwortung der Frage nach *unterschiedlichen Leistungen in Abhängigkeit von früheren Erfahrungen mit DG* zeigte eindeutige Ergebnisse. In der Gesamtstichprobe erzielten jene Personen deutlich bessere Leistungen, die angaben, Erfahrungen mit DG gemacht zu haben. Mit diesen Ergebnissen konnte auch in dieser Arbeit kongruent mit Gittler (1994) sowie Gittler und Glück (1998) ein positiver Einfluss von DG-Erfahrung auf die Raumvorstellung bestätigt werden.

Der *Einfluss der Handpräferenz* einer Person auf ihre Leistung im EST konnte nicht bestätigt werden.

Die statistische Bedeutsamkeit dieser Ergebnisse ist jedoch als stark eingeschränkt anzusehen da der Anteil von Linkshändern in der Gesamtpopulation Rickheit, Herrmann und Deutsch (2003) zu Folge, bei etwa 15% liegt. In dieser Untersuchung waren insgesamt über beide Stichproben jedoch nur etwa 6% der Versuchspersonen Linkshänder. In Anbetracht der sehr geringen Anzahl links- und beidhändiger Personen sind kausale Interpretationen der Ergebnisse daher nicht sinnvoll. In zukünftigen Studien sollte verstärkt darauf Wert gelegt

werden, eine das Populationsverhältnis in etwa widerspiegelnde Verteilung dieser Personen in der Stichprobe zu gewährleisten um aussagekräftige Ergebnisse bieten zu können.

Eine letzte Hypothese ging der Frage von *bearbeitungszeitspezifischen Unterschieden* in der EST-Leistung der Probanden nach. Dabei wurden die Probanden an den Quartilen der Gesamtstichprobe in vier Gruppen getrennt: in Personen, die 12 Minuten oder kürzer für die Bearbeitung des EST aufwandten; in eine Gruppe, die länger als 12 oder kürzer gleich 16 Minuten benötigte; in eine Gruppe, deren Probanden länger als 16 Minuten und kürzer oder gleich 22 Minuten brauchten und in eine Gruppe, die länger als 22 Minuten zur Testbearbeitung aufwendete.

Unter Konstanthaltung möglicher Einflüsse von Geschlecht und Alter auf die Bearbeitungszeit zeigten sich bei zweiseitig formulierter Fragestellung Unterschiede in Abhängigkeit der von den Probanden aufgewandten Bearbeitungszeit des EST. Jene Gruppe, die weniger als 12 Minuten für die Testbearbeitung benötigte, unterschied sich dabei deutlich von jenen Personen, die 17 bis 22 Minuten oder länger dafür in Anspruch nahmen.

Diese Ergebnisse entsprechen der von Geiser et al. (2006), Goldstein et al. (1990) sowie Maccoby & Jacklin (1974) postulierten Hypothese, dass eine kürzere Bearbeitungszeit mit schlechteren Leistungen einhergeht.

Eine Überprüfung hinsichtlich der Zeit, die die Probanden für die Bearbeitung des EST aufwandten, konnte keine signifikante Effekte bezüglich des Geschlechts zeigen. Männer und Frauen benötigten in dieser Studie durchschnittlich gleich lang für die Testbearbeitung; eine unterschiedliche Bearbeitungszeit kann somit als Erklärung für die geschlechtsspezifischen Leistungsunterschiede ausgeschlossen werden.

## 13. Zusammenfassung

Diese Arbeit dient einem interkulturellen Vergleich der Raumvorstellungsfähigkeit zwischen den Ländern Österreich und Italien anhand des *Endlosschleifentests* (EST) von Gittler und Arendasy (2003).

Raumvorstellung stellt, als einer der Primärfaktoren der Intelligenz, ein überaus wichtiges Forschungsgebiet der Psychologie dar. Im theoretischen Teil dieser Arbeit (Kapitel 5) erhält der Leser einen Einblick über diverse Definitionen von Raumvorstellung und von welchen Faktoren die Fähigkeit des Menschen zum räumlichen Vorstellungsvermögen beeinflusst werden kann. Demnach kann festgehalten werden, dass der aktuelle Stand der Wissenschaft von mindestens zwei getrennt existierenden Raumvorstellungsfaktoren ausgeht. Diese sind einerseits *räumliche Visualisierung*, welche die Fähigkeit beschreibt, zwei- oder dreidimensionale bildhafte Stimuli zu manipulieren oder zu rotieren und andererseits *räumliche Orientierung*, welche das Erfassen der Anordnung visueller Stimuli, sowie die Fähigkeit, sich durch eine veränderte Sichtweise auf ein Objekt nicht verwirren zu lassen, darstellt (McGee, 1979).

Ebenso wichtig wie das Konzept der „Raumvorstellung“ für dieser Arbeit sind auch die Beiträge der „Kulturvergleichenden Psychologie“, die ein bedeutungsvolles Teilgebiet der Psychologie darstellt (Thomas, 2003). Durch die Suche nach Allgemeingültigkeit und Generalisierbarkeit psychologischen Wissens bietet sie eine Möglichkeit zur verbesserten interkulturellen Kommunikation und die Eröffnung eines weiteren Schrittes in Richtung einer einheitlichen Psychologie.

Der interkulturelle Vergleich ermöglicht es, ein Testinstrument, das in einer Kultur entwickelt wurde, auch in einer anderen Kultur auf Validität zu untersuchen. Die Gewährleistung sinnvoller Vergleiche ist jedoch nur dann gegeben, wenn die Daten eine äquivalente Bedeutung in den untersuchten Kulturen aufweisen. Dieser Frage wird hier anhand der drei Hauptfragestellungen Rechnung getragen, indem der EST durch das Rasch-Modell analysiert wird.

Das zentrale Thema dieser Arbeit behandelt die Frage, ob der Endlosschleifentest wiederholt in einer österreichischen Stichprobe und zum ersten Mal in einer italienischen Stichprobe die Eigenschaft der Rasch-Homogenität aufweisen könne und zu dem auch über beide Kulturen dieselbe latente Fähigkeitsdimension erhebe.

Alle Modelltests der Rasch-Modell Analysen ergaben, nach theoriegeleiteter Exklusion einiger Items, insignifikante Ergebnisse und bestätigen somit die Rasch-Homogenität des EST in allen drei Bedingungen (innerhalb der österreichischen, innerhalb der italienischen und über beide Kulturen gemeinsam). Auf Basis der, durch diese Prüfung bestätigten Voraussetzung, ein äquivalentes Messinstrument zu verwenden, können in weiterer Folge inferenzstatistische Untersuchungen über beide und innerhalb beider Kulturen durchgeführt werden.

Neben den Hauptfragestellungen wird in dieser Studie auch einem eventuellen Einfluss von Ausbildung, Alter, Geschlecht sowie Handpräferenz, Bearbeitungszeit und Erfahrungen mit „darstellender Geometrie“ auf die Raumvorstellungsfähigkeiten nachgegangen. Das zusätzlich zum EST vorgegebene Datenblatt (siehe Anhang III und IV) diente der Erfassung dieser Variablen.

Ein Unterschied in den Raumvorstellungsleistungen zwischen Italienern und Österreichern wird, wie angenommen, nicht gezeigt. Diese Erwartung entstand auf Grund der Zugehörigkeit zu einem gemeinsamen Kulturkreis, der geographischen Nähe beider Länder und der damit verbundenen vielfältigen Ähnlichkeiten, welche auf wirtschaftlicher, politischer und kultureller Ebene zu erkennen sind.

Kongruent mit zahlreichen Untersuchungen (vgl. Kapitel 5) zeigt auch diese Arbeit in allen drei Konstellationen (über und innerhalb beider Kulturen) geschlechtsspezifische Leistungsunterschiede zu Gunsten der männlichen Teilnehmer auf.

Unterschiedliche Leistungen in Abhängigkeit des Alters der Probanden zeigen, dass jüngere Probanden (jünger als 27 & 27 bis 40) deutlich bessere Leistungen erzielte, als die älteren Personen (älter als 40). Nicht repliziert werden kann die Hypothese von Salthouse (2009), der



eine Verringerung raumvorstellungsspezifischer Leistungen bereits ab einem Alter von 27 Jahren aufzeigte.

Untersuchungen hinsichtlich eines Einflusses von Erfahrungen mit „darstellender Geometrie“ auf die EST-Leistungen der Probanden zeigen, dass Personen, die Erfahrung mit DG hatten, signifikant bessere Leistungen erzielten als Personen, die über keine Erfahrung darin verfügten.

Die Ergebnisse der statistischen Analysen zeigen einen Einfluss der höchsten abgeschlossenen Schulbildung auf die Leistungen des EST, wobei Personen, die über „kein Maturaniveau“ verfügten und deren höchste abgeschlossene Schulbildung entweder in Volksschule, Hauptschule oder Lehre bestand, über eine deutlich schlechtere Raumvorstellungsleistung verfügten als Personen mit „Maturaniveau“ oder „Akademiker“.

Hinsichtlich der Testbearbeitungszeiten erzielten Personen, die sich 17 Minuten oder länger für den EST Zeit ließen, deutlich bessere Ergebnisse als Personen, die weniger als 12 Minuten für seine Bearbeitung aufwendeten.

Der nicht bestätigte Einfluss der Handpräferenz einer Person auf ihre EST-Leistung ist in der vorliegenden Arbeit stark relativiert zu sehen. Auf Grund der sehr geringen Anzahl linkshändiger Personen in der Stichprobe (6% zu der angenommenen Verteilung von 15% in der Gesamtpopulation) ist dieses Ergebnis mit der gebotenen Vorsicht zu interpretieren.

Das bedeutendste Ergebnis dieser Studie für weitere Untersuchungen kann in der, auch für die italienische Stichprobe gezeigten, Rasch-Homogenität des EST gesehen werden. Wie bereits in kulturvergleichenden Untersuchungen mit der Slowakei (Kriser-Fischer, 2006), Mexiko (Ruiz-Zapata, 2005) und Australien (Waldenmair, 2003) gezeigt, stellt der EST auch für die italienische Stichprobe erneut ein faires Maß der Raumvorstellungsleistung dar.

## Literaturverzeichnis

- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung*. (5., aktualisierte und erweiterte Aufl.). Stuttgart: Kohlhammer.
- Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38 (1), 123-140.
- Arendasy, M. (2000). *Psychometrischer Vergleich computergestützter Vorgabeformen bei Raumvorstellungsaufgaben: Stereoskopisch-dreidimensionale und herkömmlich-zweidimensionale Darbietung*. Unveröffentlichte Dissertation, Universität Wien.
- Asendorpf, J. B. (2007). *Psychologie der Persönlichkeit*. (4., überarbeitete und aktualisierte Aufl.). Heidelberg: Springer.
- Berry, J. W. (1980). Introduction to methodology. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology*, (Vol. 2, pp. 1-28). Boston: Allyn and Bacon.
- Berry, J. W., Poortinga, Y. H. & Pandey, J. (Eds.) (1997). *Handbook of cross-cultural psychology. Theory and Method*, (Vol.1, 2<sup>nd</sup> ed.) Boston: Allyn and Bacon.
- Berry, J. W., Poortinga, Y. H., Segall, M. H. & Dasen, P. R. (1992). *Cross-cultural psychology. Research and applications*. Cambridge: Cambridge University Press.
- Bischof-Köhler, D. (2006). *Von Natur aus anders. Die Psychologie der Geschlechtsunterschiede*. (3., überarbeitete und erweiterte Aufl.). Stuttgart: Kohlhammer.
- Bock, R. D. & Kolakowski, D. (1973). Further evidence of a sex-linked major-gene influence on human spatial visualizing ability. *American Journal of Human Genetics*, 25, 1-14.
- Boles, D. B. (1980). X-Linkage of spatial ability: a critical review. *Child development*, 51, 625-635.

- Bourne, V. J. (2008). Examining the relationship between degree of handedness and degree of cerebral lateralization for processing facial emotion. *Neuropsychology*, 22 (3), 350-356.
- Bortz, J. (1999). Statistik für Sozialwissenschaftler. (5., vollständig überarbeitete und aktualisierte Aufl.). Berlin: Springer.
- Brodie, E. E. & Dunn, E. M. (2005). Visual line bisection in sinistrals and dextrals as a function of hemisphere, hand, and scan direction. *Brain and Cognition*, 58, 149-156.
- Casey, M. B., Nuttall, R. L. & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: a comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, 33 (4), 669-680.
- Casey, M. B., Nuttall, R. L. & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology*, 35 (5), 1237-1247.
- Clements, A.M., Rimrodt, S. L., Abel, J. R., Blanker, J. G., Mostofsky, S. H., Pekar, J. J., Denckla, M. B. & Cutting, L. E. (2006). Sex differences in cerebral laterality of language and visuospatial processing. *Brain and Language*, 98, 150-158.
- Cooke-Simpson, A. & Voyer, D. (2007). Confidence and gender differences on the Mental Rotation Test. *Learning and Individual Differences*, 17, 181-186.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Feingold, A. (1994). Gender differences in variability in intellectual abilities: a cross-cultural perspective. *Sex Roles: A Journal of Research*, 30 (1-2), 81-92.
- Friedlmeier, W. (1994). Subjektive Erziehungstheorien im Kulturvergleich. In G. Trommsdorff (Hrsg.), *Kindheit und Jugend in verschiedenen Kulturen. Entwicklung und Sozialisation in Kulturvergleichender Sicht*. (S. 43-64). Weinheim: Juventa.

- Geiser, C., Lehman, W. & Eid, M. (2006). Separating „rotators“ from „nonrotators“ in the Mental Rotation Test: a multigroup latent class analysis. *Multivariate Behavioral Research*, 41 (3), 261-293.
- Geschwind, N. & Galaburda, A. M. (1987). *Cerebral lateralization: Biological mechanisms, associations, and pathology*. Cambridge, MA: MIT Press.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest (3DW). Ein Rasch-skaliertes Test zur Messung des räumlichen Vorstellungsvermögens*. Weinheim: Beltz Test.
- Gittler, G. (1994). Intelligenzförderung durch Schulunterricht: Darstellende Geometrie und räumliches Vorstellungsvermögen. In G. Gittler, M. Jirasko, U. Kastner-Koller, C. Korunka & A. Al-Roubaie (Hrsg.), *Die Seele ist ein weites Land. Aktuelle Forschung am Wiener Institut für Psychologie*. (S. 105-122). Wien: Wiener Universitätsverlag.
- Gittler, G. (1999). Sind Raumvorstellung und Reasoning separierbare Fähigkeitsdimensionen? Dimensionalitätsanalysen zweier Rasch-skaliertes Tests: 3DW und WMT. *Diagnostica*, 45 (2), 69-81.
- Gittler, G. & Arendasy, M. (2003). Endlosschleifen: Psychometrische Grundlagen des Aufgabentyps Ep. *Diagnostica*, 49 (4), 164-175.
- Gittler, G. & Glück, J. (1998). Differential transfer of learning: effects of instruction in descriptive geometry on spatial test performance. *Journal of Geometry and Graphics*, 2 (1), 71-84.
- Gittler, G. & Vitouch, O. (1994). Empirical contribution to the question of sex-dependent inheritance of spatial ability. *Perceptual and motor skills*, 78, 407-417.
- Gladue, B. A. & Bailey, J.M. (1995). Spatial ability, handedness, and human sexual orientation. *Psychoneuroendocrinology*, 20 (5), 487-497.
- Goldstein, D., Haldane, D. & Mitchell, C. (1990). Sex differences in visual-spatial ability: the role of performance factors. *Memory and Cognition*, 18 (5), 546-550.

- Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, J. A., Turetsky, B. I., Detre, J. A., Gee, J. & Gur, R. E. (2000). An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language*, 74, 157-170.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum.
- Hambleton, R. K. & Kanjee, A. (1995). Increasing the validity of cross-cultural Assessments: Use of improved methods for test adaptations. *European Journal of Psychological Assessment*, 11(3), 147-157.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications.
- Helfrich, H. (1993). Methodologie kulturvergleichender psychologischer Forschung. In A. Thomas (Hrsg.), *Kulturvergleichende Psychologie. Eine Einführung*. (S. 81-102). Göttingen: Hogrefe.
- Hirnstein, M., Bayer, U. & Hausmann, M. (in press). Sex-specific response strategies in mental rotation. *Learning and individual differences* (2008), doi: 10.1016/j.lindif.2008.11.006
- Hohenberg, F. (1966). *Konstruktive Geometrie in der Technik* (3., ergänzte Aufl.). Wien: Springer.
- International Test Commission. *Richtlinien zur Übersetzung, Adaption und Vorgabe von Messinstrumenten*. Retrieved March 15, 2009, from <http://www.intestcom.org/guidelines/index.php>
- Jahoda, G. & Krewer, B. (1997). History of cross-cultural and cultural psychology. In J. W. Berry, Y. P. Poortinga & J. Pandey (Eds.), *Handbook of cross-cultural psychology. Theory and Method* (Vol. 1, 2<sup>nd</sup> ed., pp. 1-42). Boston: Allyn and Bacon.
- Kienbaum, J. (1995). Sozialisation von Mitgefühl und prosozialem Verhalten: Ein Vergleich deutscher und sowjetischer Kindergarten Kinder. In G. Trommsdorff (Hrsg.), *Kindheit und Jugend in verschiedenen Kulturen. Entwicklung und Sozialisation in Kulturvergleichender Sicht*. (S. 83-108). Weinheim: Juventa.

- Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kriser-Fischer, E. (2006). *EST und 3DW im interkulturellen Vergleich Österreich – Slowakei*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Kroeber, A. A. & Kluckhohn, C. (1952). *Culture: A critical review of concepts and definitions*. Cambridge, MA: Papers of the Peabody Museum of American Archaeology and Ethnology, Vol. 47, No.1.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Wien: Hogrefe.
- Lacreuse, A. & Kim, C. B. (2005). Sex, age, and training modulate spatial memory in the Rhesus monkey (*Macaca mulatta*). *Behavioral Neuroscience*, 119 (1), 118-126.
- Levy, J. & Reid, M. (1978). Variations in cerebral organization as a function of handedness, hand posture in writing, and sex. *Journal of Experimental Psychology*, 107(2), 119-144.
- Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child development*, 56, 1479-1498.
- Lonner, W. J. (1994). Culture and human diversity. In E. J. Trickett, R. J. Watts & D. Birman (Eds.), *Human diversity. Perspectives on people in context* (pp. 230-243). San Francisco: Jossey-Bass.
- Lonner, W. J. & Adamopoulos, J. (1997). Culture as antecedent to behaviour. In J. W. Berry, Y. P. Poortinga & J. Pandey (Eds.), *Handbook of cross-cultural psychology. Theory and Method* (Vol. 1, 2<sup>nd</sup> ed., pp. 43-83). Boston: Allyn and Bacon.
- Maccoby, E. E. & Jacklin, C. N. (1974). *The Psychology of sex differences*. Stanford, CA: Stanford University Press.
- Maccoby, E. E. (2000). *Psychologie der Geschlechter. Sexuelle Identität in den verschiedenen Lebensphasen*. Stuttgart: Klett-Cotta.
- McGee, M. G. (1978). Handedness and mental rotation. *Perceptual and Motor Skills*, 47, 641-642.

- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86 (5), 889-918.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of cognitive abilities: past, present and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment. Theories, tests, and issues* (2<sup>nd</sup> ed., pp. 136-182). New York: Guilford Press.
- McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1-10.
- Meinz, E. J. & Salthouse, T. A. (1998). Is age kinder to females than to males? *Psychonomic Bulletin & Review*, 5 (1), 56-70.
- Meyers online Lexikon (2009). *Gehirnatrophie*. Abgefragt 07.02.2009, von <http://lexikon.meyers.de/wissen/Hirnatrophie>
- Molenaar, I. W. (1995). Some Background for Item Response Theory and the Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models. Foundations, recent developments, and applications*. (pp. 3-14). New York: Springer.
- Montello, D. R. (1995). How significant are cultural differences in spatial cognition? In A. U. Frank & W. Kuhn (Eds.), *Spatial information theory: a theoretical basis for GIS*. (pp. 485-500). Lecture notes in computer science 988. Berlin: Springer-Verlag.
- Németh, B. (2007). Measurement of the development of spatial ability by Mental Cutting Test. *Annales Mathematicae et Informaticae*, 34, 123-128.
- Nenty, H. J. & Dinero, T. E. (1981). A cross-cultural analysis of the fairness of the Cattell culture fair intelligence test using the Rasch model. *Applied Psychological Measurement*, 5 (3), 355-368.
- Nyborg, H. (1990). Sex hormones, brain development, and spatio-perceptual strategies in Turner syndrome. In D. B. Berch & B. G. Bender (Eds.), *Sex chromosome*

- abnormalities and human behavior* (pp. 100–128). Washington, DC: American Association for the Advancement of Science.
- Olkun, S. (2003). Making connections: improving spatial abilities with engineering drawing activities. *International Journal of Mathematics Teaching & Learning*. Retrieved January 10, 2009, from <http://www.ex.ac.uk/cimt/ijmtl/ijabout.htm>
- Piaget, J. & Inhelder, B. (1975). *Die Entwicklung des räumlichen Denkens beim Kinde*. Band 6 der gesammelten Werke. Studienausgabe. Stuttgart: Ernst Klett Verlag.
- Pervin, L.A., Cervone, D. & John, O. P. (2005). *Persönlichkeitstheorien*. (5., vollständig überarbeitete und erweiterte Aufl.). München: Ernst Reinhardt Verlag.
- Peters, M., Reimers, S. & Manning, J. T. (2006). Hand preference for writing and associations with selected demographic and behavioral variables in 255,100 subjects: the BBC internet study. *Brain and cognition*, 62, 177-189.
- Prieto, L., Novick, D., Sacristán, J. A., Edgell, E.T. & Alonso, J. (2003). A Rasch model analysis to test the cross-cultural validity of the EuroQoL-5D in the schizophrenia outpatient health outcomes study. *Acta Psychiatrica Scandinavica*, 107 (Suppl., 416), 24-29.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rickheit, G., Herrmann, T. & Deutsch, W. (2003). *Psycholinguistik. Ein internationales Handbuch*. Berlin: Walter de Gruyter
- Roorda, J. (1994). Visual perception, spatial visualization and engineering drawing. *Engineering Design Graphics Journal*, 58 (2), 12-21.
- Rost, J. (1988). *Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. (2., vollständig überarbeitete und erweiterte Aufl.). Bern: Huber.



- Ruiz-Zapata, E. (2005). *Raumvorstellung im Kulturvergleich: Mexiko – Österreich*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin? [Electronic version]. *Neurobiology of Aging*, 30, 507-514.
- Salzberger, T. & Sinkovics, R. R. (2006). Reconsidering the problem of data equivalence in international marketing research. Contrasting approaches based on CFA and the Rasch model for measurement. *International Marketing Review*, 23 (4), 390-417.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Stumpf, H. & Fay, E. (1983). Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens. Göttingen: Verlag für Psychologie, Hogrefe.
- Sybert, V. P. & McCauley, E. (2004). Turner's syndrome. *The New England Journal of Medicine* 351(12), 1227-1238.
- Tanzer, N. K., Gittler, G. & Sim, C. Q. E. (1994). A cross-cultural comparison of a Rasch calibrated spatial ability test between Austrian and Singaporean adolescents. In A. M. Bouvy, F. Van de Vijver, P. Boski & P. Schmitz (Eds.), *Journeys into Cross-Cultural Psychology*. (pp. 96-110). Lisse: Swets & Zeitlinger.
- Thomas, A. (1993, a). *Kulturvergleichende Psychologie. Eine Einführung*. Göttingen: Hogrefe.
- Thomas, A. (1993, b). Entwicklungslinien und Erkenntniswert kulturvergleichender Psychologie. In A. Thomas (Hrsg.), *Kulturvergleichende Psychologie. Eine Einführung*. (S. 27-51). Göttingen: Hogrefe.
- Thomas, A. (1993, c). Psychologie interkulturellen Lernens und Handelns. In A. Thomas (Hrsg.), *Kulturvergleichende Psychologie. Eine Einführung*. (S. 377-424).
- Thomas, A. (2003). *Kulturvergleichende Psychologie. Eine Einführung*. Göttingen: Hogrefe.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

- Trommsdorff, G. (1993). Entwicklung im Kulturvergleich. In A. Thomas (Hrsg.), *Kulturvergleichende Psychologie. Eine Einführung*. (S. 103-143). Göttingen: Hogrefe.
- Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotation: a group test of three-dimensional spatial visualisation. *Perceptual and Motor Skills*, 47, 599-604.
- Van de Vijver, F. J. R. & Hambleton, R. K. (1996). Translating tests: some practical guidelines. *European Psychologist*, 1(2), 89-99.
- Van de Vijver, F. J. R. & Leung, K. (1997, a). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga & J. Pandey (Eds.), *Handbook of cross-cultural psychology. Theory and Method*, (Vol. 1, 2<sup>nd</sup> ed., pp. 257-300). Boston: Allyn and Bacon.
- Van de Vijver, F. J. R. & Leung, K. (1997, b). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.
- Van de Vijver, F. J. R. & Poortinga, Y. H. (1992). Testing in culturally heterogeneous populations: When are cultural loadings undesirable? *European Journal of Psychological Assessment*, 8(1), 17-24.
- Van de Vijver, F. J. R. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée*, 54(2), 119-135.
- Voyer, D., Voyer, S. & Bryden, P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117(2), 250-270.
- Waldenmair, A. (2003). *Raumvorstellung und Geschlechtsrollenidentität im Kulturvergleich Australien – Österreich*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.

- Werner, O. & Campbell, D.T. (1970). Translating, working through interpreters, and the problem of decentering. In R., Naroll & R. Cohen (Eds.), *A handbook of cultural anthropology* (pp. 398-419). New York: American Museum of Natural History.
- Willis, S. L. & Schaie, K. W. (1988). Gender differences in spatial ability in old age: longitudinal and intervention findings. *Sex Roles, 18*, 189-203.
- Wright, B. D. & Stone, M. H. (1979). *Best test design. Rasch measurement*. Chicago: Mesa Press.

## Sekundärliteraturverzeichnis

- Annett, M. (2002). *Handedness and brain asymmetry: the right shift theory*. Hove: Psychology Press.
- Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural psychology* (pp. 137–164). Newbury Park, CA: Sage Publishers.
- Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor analytic studies*. New York: Cambridge University Press.
- Gur, R. C., Mozley, P.D., Resnick, S. M., Gottlieb, G. E., Kohn, M., Zimmerman, R., Herman, G., Atlas, S., Grossman, R., Berretta, D., Erwin, R. & Gur, R. E. (1991). Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, 88, 2845-2849.
- Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Horn, J. L. (1991). Measurement of intellectual capabilities: a review of theory. In K. S. McGrew, J. K. Werder & R.W. Woodcock, (Eds.) *Woodcock-Johnson Technical Manual* (pp. 197-232). Chicago: Riverside.
- Levy, J. (1976). Cerebral lateralization and spatial ability. *Behavior Genetics*, 6, 171-188.
- Levy, K.L. (1980). A Monte Carlo study of analysis of covariance under violations of the assumptions of normality and equal regression slopes. *Educational and Psychological Measurement*, 40, 835-840.
- Pike, K. L. (1967). *Language in relation to a unified theory of the structure of human behaviour*. The Hague: Mouton.

- Rasch, G. (1977). On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements. In M. Glegvad (Ed.). *The Danish Yearbook of Philosophy* (pp. 58-94). Copenhagen: Munksgaard.
- Thurstone, L. L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 26, 249-269.
- Trautner, H. M. (1979). Psychologische Theorien der Geschlechtsrollenentwicklung. In A. Degenhardt & H. M. Trautner (Hrsg.), *Geschlechtstypisches Verhalten*. (S. 50-84). München: Beck.
- Vandenberg, S. G. (1968). *Progress in human behavior genetics*. Johns Hopkins University Press: Baltimore.
- Wundt, W. (1913). *Elemente der Völkerpsychologie* (Zweite Auflage). Leipzig: Alfred Kroner Verlag.

## Tabellenverzeichnis

Tab.6.1.1: Geschlechterverteilung der Gesamtstichprobe .....	53
Tab.6.2.1: Geschlechterverteilung der italienischen Stichprobe. ....	55
Tab.6.3.1: Geschlechterverteilung der österreichischen Stichprobe.....	56
Tab.10.1.1: Rasch-Modell Analyse Italien (k=21). ....	70
Tab.10.1.2: Rasch-Modell Analyse Italien (k=18). ....	71
Tab.10.2.1: Rasch-Modell Analyse Österreich (k=21).....	73
Tab.10.2.2: Rasch-Modell Analyse Österreich (k=18).....	73
Tab.10.3.1: Rasch-Modell Analyse Österreich (n=149) und Italien (n=150).....	75
Tab.10.4.1: WLE-Parameter des EST.....	76
Tab.11.1.1: Kolmogorov-Smirnov-Test. ....	78
Tab.11.1.2: Ergebnisse der t-Tests.....	79
Tab.11.2.1: Ergebnisse der t-Tests.....	80
Tab.11.3.1: Kolmogorov-Smirnov-Test .....	81
Tab.11.3.2: Ergebnisse des U-Tests.....	81
Tab.11.3.3: Kolmogorov-Smirnov-Test .....	81
Tab.11.3.4: Ergebnisse der ANCOVA .....	81
Tab.11.3.5: Mehrfachvergleiche nach Sidak .....	81
Tab.11.4.1: Bildungsverteilung.....	85
Tab.11.4.2: Kolmogorov-Smirnov-Test .....	86
Tab.11.4.3: Ergebnisse der ANOVA .....	87
Tab.11.4.4: Mehrfachvergleiche nach Scheffé .....	87
Tab.11.5.1: Kolmogorov-Smirnov-Test .....	88
Tab.11.5.2: Ergebnisse der ANCOVA .....	88
Tab.11.6.1: Kolmogorov-Smirnov-Test .....	90
Tab.11.6. 2: Ergebnisse des Kruskal-Wallis Test.....	90

Tab.11.7.1: Kolmogorov-Smirnov-Test .....	92
Tab.11.7.2: Ergebnisse der ANCOVA .....	92
Tab.11.7.3: Mehrfachvergleiche nach Sidak .....	93
Tab.11.7.4: Ergebnisse der univariaten Varianzanalyse .....	93

## Abbildungsverzeichnis

Abb.4.1.1 Die Formel des Rasch-Modells (nach Rost, 2004, S.121) .....	27
Abb.4.1.2 Eine einparametrische Itemcharakteristik Kurve (nach Hambleton, 1991, p. 14)..	27
Abb.6.1.1 Altersverteilung der Gesamtstichprobe .....	53
Abb.6.1.2 Bildungsverteilung der Gesamtstichprobe .....	54
Abb.6.2.1 Altersverteilung der italienischen Stichprobe .....	55
Abb.6.2.2 Bildungsverteilung der italienischen Stichprobe .....	55
Abb.6.3.1 Altersverteilung der österreichischen Stichprobe .....	57
Abb.6.3.2 Bildungsverteilung der österreichischen Stichprobe.....	57
Abb.7.1.1 Beispielaufgabe des SFT (nach Stumpf & Fay, 1983).....	58
Abb.7.1.2 Item 7 des SFT (nach Stumpf & Fay, 1983, S.4).....	59
Abb.7.1.3 Beispielitem des EST .....	59
Abb.7.1.4 Antwortformat des EST .....	60
Abb.7.4.1.1 Instruktionsbeispiel der italienischen Version des EST.....	63
Abb.7.4.1.2 Beispiele für das Antwortformat des EST .....	68
Abb.11.3.1 Rohscore-Mittelwerte über Alter, Kultur und Geschlecht .....	82
Abb.11.4.1 Rohscore-Mittelwerte für Bildung .....	86
Abb.11.5.1 Rohscore-Mittelwerte für DG, Kultur und Alter .....	89
Abb.11.7.1 Rohscore-Mittelwerte Bearbeitungszeitgruppen und Alter in Gesamtstichprobe .....	91



# Anhang

## Anhang I:

Die bildhaft unterstützte und verbale Erklärung der italienischen Version des EST.

**Compito:** Quando sono identiche l'immagine iniziale e quella finale?

Le immagini iniziali e finali sono identiche quando:

**Immagine iniziale**



**Immagine finale**

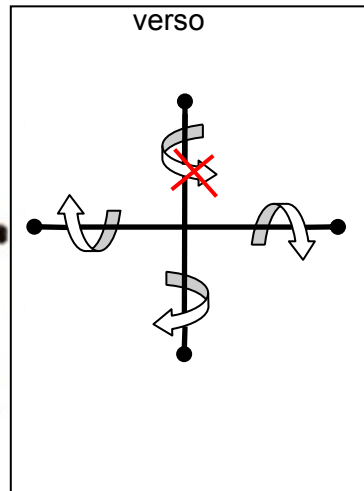


l'immagine iniziale viene girata:

di 90°

di 180°

verso



Per lo svolgimento del seguente test Lei dispone di tempo a sufficienza.

Lavori con calma e concentrazione, ma anche con speditezza. Nel caso non riuscisse a risolvere un compito, indichi  sul fascicolo delle risposte.

Per favore adesso scriva l'orario di inizio del test nel fascicolo delle risposte e cominci con il test.

**Anhang II:**

**Die bildhaft unterstützte und verbale Erklärung der österreichischen Version des EST.**

**Aufgabenstellung:** Wann sind Start- und Ziel-Ansicht gleich?

**Start-  
Ansicht**



**Ziel-  
Ansicht**



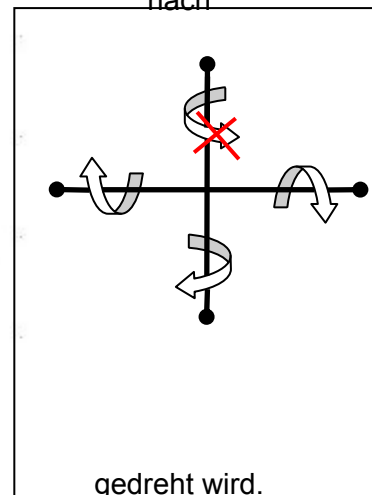
Start- und  
Zielansicht sind  
gleich, wenn

die Start-Ansicht

um 90°

um 180°

nach



gedreht wird.

Für die Bearbeitung des folgenden Tests haben Sie ausreichend Zeit zur Verfügung. Arbeiten Sie ruhig und konzentriert, aber auch zügig. Sollten Sie eine Aufgabe nicht lösen können, kreuzen Sie im Answerheft  Ich weiß nicht an.

Bitte tragen Sie nun die Uhrzeit im Answerheft ein und beginnen Sie mit dem Test.

**Anhang III:**

**Datenblatt zur Erhebung soziodemographischer Daten für die italienische Stichprobe**

***Fascicolo delle risposte***  
**EST**

**Codice personale:**


Ultima lettera del  
mio nome

unità del giorno del  
mio compleanno

prima e ultima lettera  
del nome della madre

seconda lettera del  
nome del padre

❶ **Data:** \_\_\_\_\_

❷ **Genere:**  maschile  
 femminile

❸ **Età:** \_\_\_\_\_

❹ **Istruzione:** Indicare il più alto livello d'istruzione scolastica raggiunto.

- Scuole elementari
- Scuole medie
- Apprendistato
- Maturità
- Corsi di specializzazione professionale
- Università triennale  
Indirizzo: \_\_\_\_\_
- Università specialistica o quinquennale  
Indirizzo: \_\_\_\_\_
- Altro: \_\_\_\_\_

❺ **Ha o ha avuto degli insegnamenti in geometria descrittiva?**

si       no

❻ **Lei si ritiene prevalentemente**

- destro
- mancino
- entrambi

**Anhang IV:**

**Datenblatt zur Erhebung soziodemografischer Daten für die italienische Stichprobe**

*Antwortheft*  
**Endlosschleifentest**

**Persönlicher Code:**


letzter Buchstabe meines Vornamens      Einerstelle meines Geburtstages      1. und letzter Buchstabe d. Vornamens d. Mutter      2. Buchstabe des Vornamens d. Vaters

❶ Datum: \_\_\_\_\_

❷ Geschlecht:       männlich  
                          weiblich

❸ Alter (in Jahren): \_\_\_\_\_

❹ Ausbildung: Kreuzen Sie die höchste abgeschlossene Schulausbildung an.

- Volksschule
- Hauptschule, AHS Unterstufe
- Lehre, Berufsbildende mittlere Schule
- Matura
- Akademie, Fachhochschule
- Universität
- Studienrichtung: \_\_\_\_\_
- Andere: \_\_\_\_\_

❺ Haben bzw. hatten Sie Unterricht in Darstellender Geometrie (DG)?

ja       nein

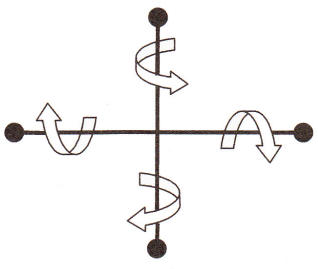
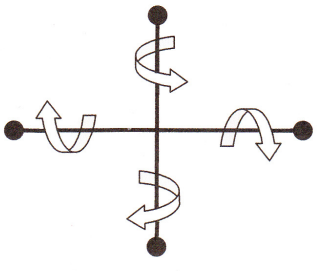
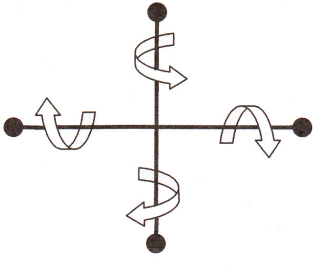
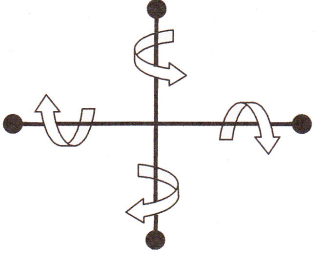
❻ Schätzen Sie sich selbst vorwiegend als

- Rechtshänder
- Linkshänder
- Beidhändig      ein?

**Anhang V:**

**Antwortblatt der italienischen Version des EST**

Le immagini iniziali e finali sono identiche, quando l'immagine iniziale viene girata...

<b>Esempio 1</b>	
di 90° <input type="checkbox"/>	
verso	
di 180° <input type="checkbox"/>	
<b>Esempio 2</b>	
di 90° <input type="checkbox"/>	
verso	
di 180° <input type="checkbox"/>	
<b>Esempio 3</b>	
di 90° <input type="checkbox"/>	
verso	
di 180° <input type="checkbox"/>	
<b>Esempio 4</b>	
di 90° <input type="checkbox"/>	
verso	
di 180° <input type="checkbox"/>	

**Anhang VI:**

**Antwortblatt der österreichischen Version des EST**

Start-Ansicht und Ziel-Ansicht sind gleich, wenn die Start-Ansicht ...

Beispiel 1	
um 90° <input type="checkbox"/>	
nach ...	
um 180° <input type="checkbox"/>	
	gedreht wird.
Beispiel 2	
um 90° <input type="checkbox"/>	
nach ...	
um 180° <input type="checkbox"/>	
	gedreht wird.
Beispiel 3	
um 90° <input type="checkbox"/>	
nach ...	
um 180° <input type="checkbox"/>	
	gedreht wird.
Beispiel 4	
um 90° <input type="checkbox"/>	
nach ...	
um 180° <input type="checkbox"/>	
	gedreht wird.

# Curriculum Vitae

## Persönliche Daten

Name: Julia Brugger  
Adresse: Franz-Gruber Strasse 18b  
5020 Salzburg  
Geburtsdatum: 15. April 1983, Salzburg

## Ausbildung

1989 –1993 Volksschule Herrnau , Salzburg  
1993 –2001 Akademisches Gymnasium am Rainberg, Salzburg  
2002 Beginn des Psychologiestudiums an der Universität Wien  
30.Juni 2008 Beginn der Diplomarbeit  
10/2005 - 08/2006 Erasmus-Aufenthalt im Rahmen des Psychologiestudiums an „la Sapienza“ in Rom

## Praktika

03/2007 - 06/2007 Praktikum bei „Wiener Sozialdienste FÖBE GmbH“ für teilbetreutes Wohnen  
10/2007 - 04/2008 Praktikum bei der „Österreichischen Autistenhilfe“; Klientenbetreuung.

## Auslandsaufenthalte

05/1999 Sprachaufenthalt in Frankreich  
04/2002 – 07/2002 Sprachaufenthalt in Rom