



universität
wien

Diplomarbeit

Titel der Diplomarbeit

Der Einfluss von Aufgabenstrukturmerkmalen auf die
Eindimensionalität des Raumvorstellungstests
SPACE

Verfasserin

Siquia Santos-Santiago

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, 2013

Studienkennzahl: A 298

Studienrichtung: Diplomstudium Psychologie

Betreuer: Ao. Univ.-Prof. Dr. Georg Gittler

DANK

Diese Arbeit wäre nicht möglich gewesen ohne die wertvolle Unterstützung vieler Menschen, die mich mit ihrem Fachwissen, ihrer Geduld, Zeit, Wärme und Zuspruch unterstützt haben.

In erster Linie danke ich dem akademischen Team Ao. Univ.-Prof. Dr. Georg Gittler, Dipl.-Psych. Anne Milatz, Mag^a. Eva Adelman und Mag^a. Christiane Heider.

Weiters bedanke ich mich von Herzen bei allen Probanden, die sich dem Test, der die Basis meiner Arbeit ist, konzentriert und ernsthaft unterzogen haben.

Besonderer Dank gilt meinen Freundinnen und Freunden und meiner Familie: Sie haben mich in meiner arbeitsintensiven, stressreichen Zeit wohlwollend und verständnisvoll begleitet.

Inhaltsverzeichnis

1. EINLEITUNG	1
THEORETISCHER TEIL	4
2. RAUMVORSTELLUNG	4
2.1 Definition: Raumvorstellung	4
2.2 Raumvorstellung und ihre Subdimensionen	5
3. MENTALER-ROTATIONS-TEST	8
3.1 Shepard und Metzler (1971)	8
3.2 Vandenberg und Kuse-Test	11
3.3 SPACE 2d/3d	13
4. ERKLÄRUNGSANSÄTZE FÜR GESCHLECHTSSPEZIFISCHE UNTERSCHIEDE IN DER RAUMVORSTELLUNG	14
4.1 Bio-psycho-soziale Aspekte	14
4.2 Testspezifische Aspekte	16
4.2.1 Stimulusmaterial	16
4.2.2 Vorgabeweise von Testaufgaben	19
4.3 Strategiewahl	21
4.3.1 Holistische Strategien	23
4.3.2 Analytische Strategie	23
4.3.3 Intermediäre Strategie	24
4.4 Strategien bei Mentaler Rotation (MR)	25
4.5 Zugänge zur Strategieerhebung	26
5. METHODIK UND ERHEBUNGSVERFAHREN	27
5.1 Rasch-Modell	29
5.2 Space 2d-2 Modi: „ohne-Gitter“- und „mit-Gitter“-Aufgaben	34
6. ZIELSETZUNG UND FORSCHUNGSFRAGEN	38
6.1 Zielsetzung	38
6.2 Forschungsfragen	38
EMPIRISCHER TEIL	40
7. PLANUNG UND DURCHFÜHRUNG DER STUDIE	40
7.1 Forschungsfragen und statistische Hypothesen	40
7.2 Untersuchungsteilnehmer und Stichprobenziehung	42
7.3 Stichprobenbeschreibung	43
7.4 Durchführung	44
7.4.1 Space 2d-2 Modi: „ohne-Gitter“- und „mit-Gitter“-Aufgaben	44
7.4.2 Zusatzfragen (Aussagenpaare)	46

8.	ERGEBNISSE DER STATISTISCHEN AUSWERTUNG	53
8.1	Rasch-Analysen	53
8.1.1	Andersen Likelihood Quotiententest	54
8.1.2	Martin-Löf-Test	55
8.1.3	Einzelitems als Teilungskriterium	55
8.1.4.	Personenparameter (Gesamttest) und Gruppenunterschiede	57
8.2	Ergebnisse der klassischen Testtheorie	60
8.2.1	Reliabilitätsanalyse	60
8.2.2	Klassische Itemschwierigkeiten	61
8.3	Überprüfung der Fragestellungen	65
8.3.1	Gibt es leistungsmäßige Unterschiede zwischen den Testteilen (oG/mG) in Abhängigkeit vom Geschlecht?	65
8.3.2	Gibt es Leistungsunterschiede in a) der Bearbeitung zwischen den Link-Items 2,3,4 (oG) und 23,24,25 (mG) und b) in Abhängigkeit vom Geschlecht?	67
8.3.3.	Gibt es Unterschiede in der Bearbeitungszeit zwischen a) den beiden Aufgabenmodi (oG vs. mG) und b) in Abhängigkeit vom Geschlecht?	69
8.3.4	Kann das externe Kriterium Selbsteinschätzung der „Orientierungsfähigkeit im Raum“ die Leistung der Probanden im Test vorhersagen?	72
8.3.5	Zusatzfragen: Unterscheiden sich die subjektiven Ratings bezüglich der Aufgabenmodalität in Abhängigkeit vom Geschlecht?	73
9.	ZUSAMMENFASSUNG UND DISKUSSION	80
9.1	Zusammenfassung	80
9.2	Interpretation	83
9.3	Diskussion	84
10.	ABSTRACT-DEUTSCH	88
11.	ABSTRACT-ENGLISCH	89
12.	LITERATURVERZEICHNIS	90
13.	ABBILDUNGSVERZEICHNIS	96
14.	TABELLENVERZEICHNIS	98
15.	ANHANG	100
	Anhang A: Graphischer Modelltest	100
	Anhang B: Rasch Modell-Analyse	117
	Anhang C: Zusatzfragen	122
	Anhang D: Curriculum vitae	129

1. Einleitung

Die Fähigkeit visuell-räumlich zu denken zählt zu den sieben grundlegenden Bereichen menschlicher Intelligenz, die im Rahmen faktorenanalytischer Verfahren ermittelt wurden (vgl. Pawlik, 1971; zitiert nach Gittler & Arendasy, 2003). Für viele alltägliche Handlungen ist es notwendig, räumliche Beziehungen vorstellen und abschätzen zu können (Schuhbänder binden, Einparken, Kästen zusammenbauen). Auch für viele Schulfächer wie Mathematik, Geographie oder Geometrisches Zeichnen sowie für viele handwerkliche, technische und naturwissenschaftliche Berufe ist räumlich-visuelles Denken erforderlich (Maier, 1999). Eine valide Diagnostik dieser zentralen Fähigkeit ist somit ein wichtiges Forschungsfeld.

Dabei stellt die „Raumvorstellungsfähigkeit“ neben „verbaler Flüssigkeit“ einen der beiden kognitiven Bereiche dar, in denen sich Männer und Frauen unterscheiden. Männer erzielen im Durchschnitt bessere Leistungen in der Raumvorstellung, Frauen in verbaler Flüssigkeit (Hausmann, 2007). Den möglichen Ursachen für den vielfach belegten signifikanten Leistungsunterschied zu Gunsten der Männer im Bereich der Raumvorstellung (im Folgenden von mir als RV abgekürzt) (Linn & Petersen, 1985; Voyer, Voyer & Bryden, 1995) soll in dieser Arbeit nachgegangen werden, wobei das Hauptaugenmerk auf dem Aspekt der Strategiewahl beim Lösen von Raumvorstellungsaufgaben liegt, da diese oft als mögliche Erklärung für den geschlechtsspezifischen Unterschied vermutet wird (Glück, 2001; Glück & Fitting, 2003). So geht aus der bisherigen Literatur hervor, dass Personen Raumvorstellungsaufgaben teilweise auch ohne räumliche Vorstellung lösen (z.B. Putz-Osterloh, 1977).

Im theoretischen Teil wird einleitend eine Begriffsklärung zum Konstrukt der Raumvorstellungsfähigkeit und dessen Subfaktoren gegeben. Da es in dieser Arbeit um den Subfaktor der mentalen Reaktionsfähigkeit gehen wird, folgen eine kurze Schilderung der Entwicklung des bekanntesten Verfahrens zur Erfassung dieser Subfähigkeit und eine Darstellung neuerer Tests.

Für die vorliegende Untersuchung wurden die Aufgaben eines kürzlich entwickelten Rasch-homogenen Itempools von Raumvorstellungsaufgaben („ohne-Gitter“-Aufgaben) (Heider, 2010), hinsichtlich eines Aufgabenmerkmals („Gitternetz“) verändert („mit-Gitter“-Aufgaben). In dem neu zusammengestellten Test wurden sowohl ein paar der ursprünglichen „ohne-Gitter“-Aufgaben als auch „mit-Gitter“-Aufgaben vorgelegt.

Im empirischen Teil soll durch die Bearbeitung der beiden verschiedenen Aufgabenmodi („ohne-Gitter“ vs. „mit-Gitter“) untersucht werden, inwieweit die a-priori angenommene visuell-räumliche Leistung (konkret: mentale Rotation) bei der Bearbeitung der Aufgaben auch tatsächlich erbracht und somit messbar gemacht werden kann. Besonders bei den für diese Arbeit neu erstellten „mit-Gitter“-Aufgaben stellt sich die Frage nach der Eindimensionalität des Aufgabenmaterials im Sinne des dichotom-logistischen Modells von Rasch (kurz: Rasch-Modell, RM; Rasch, 1960/1980; zitiert nach Gittler & Arendasy, 2003, S. 165ff), da auch eine Bearbeitung und Aufgabenlösung mittels Reasoning angenommen werden kann. Reasoning bezeichnet die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge zu erkennen (Kubinger, 2006).

Zwei Zugänge sollen mehr Klarheit schaffen hinsichtlich Strategieeinheit oder Strategievaretät bei der Bearbeitung von mentalen Rotations-Aufgaben vom Shepard-Metzler-Typ in Abhängigkeit von Strukturmerkmalen. Erstens sollen die Modellkontrollen nach Rasch (1960/1980; zitiert nach Gittler & Arendasy, 2003, S. 165ff) das verwendete Aufgabenmaterial dahingehend prüfen, ob für alle Personen essentiell dieselben Fähigkeitsdimensionen a) innerhalb eines Aufgabenmodus („ohne-Gitter“ vs. „mit-Gitter“) und b) über beide Aufgabenmodi hinweg (Gesamttest) gemessen werden. Denn erst wenn sich das Rasch-Modell als gültig erweist, misst es bei allen Personen (somit auch bei Männern und Frauen) essentiell dieselbe Fähigkeitsdimension. In diesem Fall könnte der vielfach geäußerten Kritik entgegen getreten werden, dass Raumvorstellungstests nicht geeignet seien, Leistungsunterschiede zwischen den Geschlechtern, die bedingt sind durch geschlechtsspezifische Itembearbeitungsstrategien, zu erfassen (Gittler, 1999). Zweitens

wurden im Anschluss an die Testung Fragen präsentiert, anhand derer mehr Informationen über die Bearbeitungsstrategien in Abhängigkeit von der Bildgestaltung („ohne-Gitter“ vs. „mit-Gitter“) des Aufgabenmaterials erwartet wurden. Diese beiden Analyseansätze (das Rasch-Modell und die Zusatzfragen) gemeinsam sollen zu valideren Ergebnissen führen als jede für sich, da jede Strategieerfassungsmethode für sich alleine ihre spezifischen Schwächen hat (Glück, 2001). Außerdem erlauben die aus der bisherigen Literatur abgeleiteten Zusatzfragen, mögliche signifikante Rasch-Modell-Analysen besser inhaltlich zu interpretieren (vgl. Gittler, 1986; Gittler, 1992) und einen möglichen Einfluss der Strategiewahl auf den Leistungsunterschied zwischen den Geschlechtern besser zu erfassen.

Im letzten Abschnitt werden die wichtigsten Ergebnisse der vorliegenden Arbeit noch einmal zusammengefasst dargestellt.

Es wird der besseren Lesbarkeit wegen bei allgemeinen Aussagen nur die männliche Form verwendet. Bei der Wahl der männlichen Form ist im Sinne einer genderneutralen Formulierung zugleich auch immer die weibliche Darstellungsform gemeint.

THEORETISCHER TEIL

2. Raumvorstellung

Was ist nach dem heutigen Stand der Wissenschaft unter „Raumvorstellungsfähigkeit“ (kurz: RV) zu verstehen jenem kognitiven Bereich, in welchem im Allgemeinen Männer besser als Frauen abschneiden (Linn & Petersen, 1985; Voyer et al., 1995)?

2.1 Definition: Raumvorstellung

Allgemein kann man unter dem Begriff der Raumvorstellungsfähigkeit eine Leistung im Umgang mit visuell-räumlicher, *nicht-sprachlicher* Information verstehen, die im Gedächtnis gespeichert, meistens in irgendeiner Form transformiert bzw. manipuliert und/oder wieder abgerufen wird (Glück, Kaufmann, Dünser & Steinbügl, 2005). Dabei stellt die Raumvorstellungsfähigkeit kein einheitliches Konstrukt dar, sondern besteht aus unterschiedlichen heterogenen „Subfähigkeiten“, die sich jeweils auf einen anderen Aspekt des räumlichen Denkens beziehen (Glück et al., 2005) (siehe dazu auch Kapitel 2.1). Dabei gehen die meisten Modelle des Raumvorstellungskonstrukts von zwei bis drei Faktoren aus (Glück et al., 2005).

Gittler (1992) führt die unscharfe Definition des Konstrukts und seiner Kategorisierung auf den Umstand zurück, dass es schwer sei, dieses „Spezifische einer nonverbalen, vorstellungsmäßigen Informationsverarbeitung“ (S. 46) fest zu machen. Außerdem stelle die traditionelle Herangehensweise über die Methode der Faktorenanalyse kein geeignetes Verfahren dar, Klarheit in die Unterteilung des Raumvorstellungskonstrukts zu bringen, u.a. wegen der Problematik der Stichprobenabhängigkeit innerhalb der klassischen Testtheorie (Gittler, 1984; vgl. Putz-Osterloh & Lüer, 1979).

Des Weiteren ging man im Rahmen der faktorenanalytisch-psychometrischen Forschungsperspektive von einer Strategiehomonogenität bei Raumvorstellungsaufgaben (kurz: RV-Aufgaben) aus, weswegen unterschiedliche Leistungen als bessere oder schlechtere Fähigkeit in der Anwendung dieser einen Strategie interpretiert wurden

(Grüßing, 2002). Um die angenommene Strategiehomoogenität auch ehestmöglich zu erreichen, versuchte man, die Aufgaben und Instruktionen auf maximale Strategiehomoogenität hin zu konstruieren (Glück, 2001). Dass dies nicht hinreichend gelungen war, zeigte sich bei der Analyse einzelner Testergebnisse, die bei ganz unterschiedlichen Faktoren hohe bzw. niedrige Ladungen aufwiesen, wie dies Putz-Osterloh und Lür (1975) und Putz-Osterloh (1977) in ihren Untersuchungen zeigen konnten (auf diese Untersuchungsergebnisse wird unter Kapitel 5 noch ausführlicher eingegangen). Denn vor allem komplexe RV-Aufgaben ließen sich oft genauso gut durch logisches Denken lösen wie durch reine bildhafte Imagination, wodurch eine einheitliche Faktorenstruktur nicht zu erreichen war (Glück et al., 2005).

2.2 Raumvorstellung und ihre Subdimensionen

Linn und Petersen (1985) arbeiteten im Rahmen ihrer vielzitierten Metaanalyse zu Geschlechterunterschieden in der Raumvorstellungsfähigkeit eine Kategorisierung der RV heraus, auf der die vorliegende Arbeit beruht.

Bei ihrem Kategorisierungsprozess erweiterten sie den klassisch-psychometrischen Zugang der Raumvorstellungsforschung, die Faktorenanalyse, durch drei weitere Zugänge:

- **Differentielle Perspektive:** Diese befasst sich mit den Leistungsunterschieden in der Raumvorstellungsfähigkeit innerhalb unterschiedlicher Gruppen, wie z.B. zwischen Männern und Frauen.
- **Kognitive Perspektive:** Diese befasst sich mit der Identifizierung jener Prozesse, die dem Lösen bestimmter RV-Aufgaben zugrunde liegen, zum Beispiel mit den Bedingungen, unter denen mentale Rotation durchgeführt wird oder nicht (Glück, 1999).
- **Strategische Perspektive:** Hierbei geht es um die Erforschung qualitativ unterschiedlicher Strategien beim Lösen derselben Aufgaben, die verschiedene Personen anwenden.

Durch die Beachtung all dieser Aspekte beabsichtigten Linn und Petersen (1985) Variationen in der Leistung bei ein und derselben Aufgabe besser interpretieren zu

können und das Konstrukt der „Raumvorstellungsfähigkeit“ in möglichst sinnstiftende Kategorien zu unterteilen (Linn & Petersen, 1985). Die drei Kategorien, die sie herausarbeiteten, sind:

1 Räumliche Wahrnehmung (spatial perception)

Fähigkeit zur Identifikation von Horizontalen und Vertikalen. Dabei spielt die Orientierung des eigenen Körpers, das Körperschema, eine wesentliche Rolle (Maier, 1999). Die räumliche Wahrnehmung stellt die Fähigkeit dar, räumliche Beziehungen zwischen Objekten oder in Bezug zum eigenen Körper herzustellen (Hausmann, 2007).

2 Mentale Rotation (mental rotation)

Hierbei geht es um die räumliche Fähigkeit, zwei- oder dreidimensionale Figuren schnell und akkurat (Linn & Petersen, 1985) um eine oder mehrere räumliche Achsen mental zu rotieren, um zwei oder mehrere Objekte miteinander vergleichen zu können (Hausmann, 2007).

3 Räumliche Visualisierung (spatial visualization)

Sie stellt eine Funktion dar, bei der eine komplizierte, mehrstufige mentale Manipulation räumlicher Informationen verlangt wird. Dabei können auch Prozesse, wie sie bei räumlicher Wahrnehmung und mentaler Rotation gefordert werden, beteiligt sein (Hausmann, 2007).

Wie eingangs schon erwähnt, stellt die Raumvorstellungsfähigkeit einen der wenigen kognitiven Bereiche dar, in dem sich Männer und Frauen (immer noch) unterscheiden (Voyer et al., 1995). Eine differenziertere Betrachtung zeigt jedoch, dass sich in den meisten Bereichen des räumlichen Vorstellungsvermögens die geschlechtsspezifischen Unterschiede in den letzten Jahrzehnten beträchtlich reduziert haben und teilweise gar nicht mehr nachweisbar sind (Voyer et al., 1995). Manche Autoren schließen aus diesen

Ergebnissen auf eine sozio-edukative, also gesellschaftliche Bedingtheit dieses geschlechtsspezifischen Unterschieds in räumlicher Vorstellungsfähigkeit. Dieser scheint sich durch die Angleichung der Erziehung von Mädchen und Burschen aufzuheben (Glück et al., 2005). Die Angleichung der Erziehung stellt gleichermaßen eine Angleichung von Erfahrungswelten dar. Durch solche Ergebnisse wird der Umweltaspekt in der Umwelt-Anlage Debatte rund um die Begründung von kognitiven Fähigkeitsdimensionen gestärkt.

Zu den Einflussgrößen der Erfahrung sollen auch die Ergebnisse von Trainingsstudien, hier im konkreten Fall eine Studie zur Darstellenden Geometrie angeführt werden, die einen positiven Trainingseffekt auf den Intelligenzfaktor der Raumvorstellungsfähigkeit ergaben (Gittler & Glück, 1998). In dieser Studie habe das Training die im Prätest erhobenen geschlechtsspezifischen Unterschiede sogar aufgehoben.

Trotzdem zeigen sich im Subfaktor „Mentale Rotation“ (kurz: MR), um den es in der vorliegenden Arbeit hauptsächlich gehen soll, siehe Kapitel 4.4, ziemlich große und konsistente geschlechtsspezifische Unterschiede (Linn & Petersen, 1985; Voyer et al., 1995). Doch auch hier ist eine differenziertere Betrachtung wichtig, da sich die größten Unterschiede in MR-Fähigkeit in dem Mentalen-Rotations-Test (MRT; Vandenberg & Kuse, 1978) zeigten (Linn & Petersen, 1985; Peters, Laeng, Latham, Jackson, Zaiyouna & Richardson, 1995; Voyer et al., 1995).

Dieser Test wurde bei etwa 50% aller Studien zu geschlechtsspezifischen Unterschieden in mentaler Rotationsfähigkeit eingesetzt (Linn & Petersen, 1985). Er ist daher in den Metaanalysen stark vertreten. Somit kann davon ausgegangen werden, dass die bisher vorliegenden Ergebnisse durch dieses Erhebungsinstrument mitbestimmt wurden. Glück und Frabrizii (2010) formulierten diesen Umstand so, dass der MRT einige Charakteristika aufweisen müsse, die geschlechtsspezifische Unterschiede verstärken, ganz abgesehen von der Fähigkeit zur mentalen Rotation (S. 106). Auf den Einfluss testspezifischer Aspekte wird unter Kapitel 4.2. genauer eingegangen. So zeigten Voyer et al. (1995), dass geschlechtsspezifische Unterschiede beim MRT mit 0.94 SD und für chronometrische, auf genaue Zeitmessung bzw. Geschwindigkeit beruhende, mentale

Rotationsaufgaben wie bei Shepard und Metzler (1971) bei nur 0.37 SD lagen. Auf diese beiden Untersuchungsinstrumente wird unter Kapitel 3.1. und 3.2. gesondert Bezug genommen. Somit differieren Leistungsunterschiede in RV-Fähigkeit stark in Abhängigkeit von den verwendeten Tests (Voyer et al., 1995), woraus Glück (2001) ableitete, dass die Aufgabenstellungen in unterschiedlichem Ausmaß strategievariant sein dürften.

Es sei am Rande erwähnt, dass mentale Rotations-Leistungen nicht nur zwischen Geschlechtern unterschiedlich ausfallen, sondern auch zwischen unterschiedlichen Studienrichtungen (vgl. Quaiser-Pohl, Lehmann, Jordan, Schirra, 2001), wodurch allein aufgrund des Geschlechts keine individuellen Leistungsvorhersagen im Bereich der RV-Fähigkeit möglich sind (Hirnstain & Hausmann, 2010).

Im folgenden Kapitel soll ein kurzer Überblick über die Geschichte der Erfassung der mentalen Rotation seit der klassischen und einflussreichen Studie von Shepard und Metzler (1971) gegeben werden.

3. Mentaler-Rotations-Test

Wie im vorhergehenden Kapitel dargestellt, treten die größten und beständigsten geschlechtsspezifischen Unterschiede im kognitiven Bereich (vgl. Hooven, Chabris, Ellison, Kievit & Kosslyn, 2008) in der Fähigkeit zu mentaler Rotation auf, vor allem bei der Vorgabe des MRT von Vandenberg und Kuse (1978) (Linn & Petersen, 1985; Voyer et al., 1995).

3.1 Shepard und Metzler (1971)

Im Rahmen ihrer Studie wollten Shepard und Metzler (1971) prüfen, wie viel Zeit Personen brauchen um zu entscheiden, ob sich zwei Objekte dreidimensionaler Gestalt hinsichtlich ihrer Form, abgesehen von ihrer räumlichen Ausrichtung, gleichen oder nicht. Dafür entwickelten sie Darstellungen geometrischer Figuren, die aus jeweils 10 aneinandergereihten Würfeln zusammengesetzt waren und drei „right-angled elbows“

(Shepard & Metzler, 1971, S. 702) besaßen. Die rotierten Ansichten wurden in 20°-Schritten produziert, wobei diese Rotation in der Bildebene oder in der Bildtiefe vorgenommen wurde.

Den acht erwachsenen Probanden wurden diese Darstellungen perspektivischer „Linienzeichnungen“ (Shepard & Metzler, 1971, S. 702) paarweise präsentiert (siehe Abbildung 1). Die Aufgabenstellung bestand darin, so schnell wie möglich und so richtig wie möglich zu entscheiden, ob es sich bei den zwei dargestellten Figuren um dasselbe Objekt handelte, abgesehen von dessen unterschiedlicher räumlicher Ausrichtung (Rotation eines der Objekte in der Bildebene und in der Bildtiefe). In diesem Fall war es ein „Gleiches-Paar“. Wenn dies nicht der Fall war, handelte es sich um ein „Unterschiedliches Paar“. Die Darstellung der beiden Figuren unterschied sich dahingehend, dass die eine Figur eine gespiegelte (und gedrehte) Ansicht der anderen war, sie also durch eine MR nicht miteinander in „Kongruenz“ gebracht werden konnten (Shepard & Metzler, 1971).

Die Verwendung von gespiegelten Figuren zur Herstellung von „Unterschiedlichen-Paaren“ begründen Shepard und Metzler (1971) folgendermaßen: Hätten sie zwei unterschiedliche geometrische Figuren zum Vergleichen vorgelegt, hätten sich die Probanden eventuell an markanten Detailunterschieden orientieren können, wodurch sie zur Entscheidung „Unterschiedliches Paar“ gelangt wären, ohne wirklich eine mentale Rotation vollzogen zu haben.

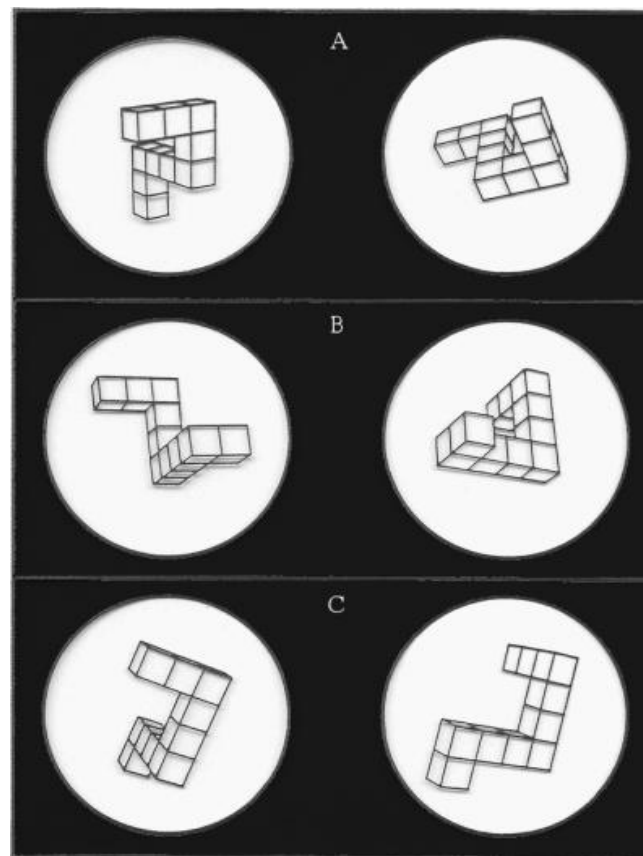


Abbildung 1: Paarweise Darstellung der geometrischen Figuren bei Shepard & Metzler (1971). A: stellt ein "Same-Pair" rotiert in der Bildebene dar; B: "Same-Pair" rotiert in der Bildtiefe; C: Gespiegelte und rotierte Ansicht in der Bildtiefe (Shepard & Metzler, 1971).

Des Weiteren achteten sie bei der Erstellung ihrer Figuren und deren räumlich unterschiedlich ausgerichteten Ansichten darauf, dass sie möglichst Gegenstands- und Bedeutungsloses darstellten. In Anlehnung an diesen Gedanken sei kurz das Thema „Erfahrung“ im Zusammenhang mit bestimmten Test-Stimuli erläutert: So merken Gittler und Arendasy (2003) an, dass immer dann relative Bildungs- und Kulturunabhängigkeit angenommen werden könne, wenn ein Testmaterial nichts Alltägliches und somit durch Vorerfahrung bekanntes Material anbiete. So werde durch das Testmaterial keine Teilpopulation bevorzugt, sondern allen Testpersonen eine faire Erhebungssituation geboten.

Die Autoren entdeckten, zumindest bei Mittelung über mehrere Testpersonen (Glück et al., 2005), einen positiven linearen Zusammenhang zwischen der Winkelgröße der Rotation und der Reaktionszeit bis zur Entscheidungsfällung bei „Gleichen Paaren“,

wobei es keinen Unterschied machte, ob die Figur in der Bildtiefe oder in der Bildebene rotiert worden war. Sie leiteten daraus ab, dass die zweidimensional abgebildeten Objekte der dreidimensionalen Figuren einen räumlichen Eindruck vermitteln konnten (Shepard & Metzler, 1971).

Als mögliche Erklärung dafür, wie Personen zu den „Gleiche Paare“-Entscheidungen gekommen waren, nahmen sie an, dass die Probanden die Objekte vor ihrem geistigen Auge gedreht und somit eine mentale Simulation der physischen Welt vollzogen hatten (Khooshabeh & Hegarty, 2010). Diese Annahme geht auch konform mit den introspektiven Selbstberichten der Probanden, auch wenn solcherart erhobene Daten hinsichtlich ihrer Interpretierbarkeit mit Vorsicht zu handhaben sind, wie Shepard und Metzler (1971) selber zu bedenken gaben (siehe dazu auch 4.5).

3.2 Vandenberg und Kuse-Test

Vandenberg und Kuse (1978) griffen die perspektivischen „Linienzeichnungen“ (Shepard & Metzler, 1971; S. 702) auf und entwickelten ein Instrument, den Mental-Rotation-Test (MRT; Vandenberg & Kuse, 1978), der in der Folge zu den meist verwendeten Verfahren zur Erfassung der Mentalen Rotationsfähigkeit wurde (Linn & Petersen, 1985). Außerdem ist es jenes Testverfahren, das die beständigsten und zuverlässigsten geschlechtsspezifischen Unterschiede in der MR hervorbringt (Linn & Petersen, 1985).

Es handelt sich bei dem Verfahren um einen Paper-Pencil-Test. Die Aufgabenstellung besteht aus einem Ausgangsbild und vier Vergleichsbildern, wie in Abbildung 2 dargestellt. Zwei der vier Antwortalternativen stellen die „richtige“ Antwort (rotierte Ansicht) dar und zwei die falsche (entweder gespiegelte und rotierte Ansicht oder aus zwei strukturell unterschiedlichen Figuren bestehende Ansichten). Hier sei angemerkt, dass die Kritik von Shepard & Metzler (1971) hinsichtlich der Erstellung von „Unterschiedlichen Paaren“ nicht beachtet wurde.

Nur die Aufgaben, bei denen die beiden korrekten Antworten angekreuzt wurden, wurden als richtig bewertet. Durch diesen Verrechnungsmodus erhofften Vandenberg und Kuse, die Ratewahrscheinlichkeit zu reduzieren. Für eine falsche Antwortalternative gab es einen Punkteabzug. Insgesamt hatten die Probanden sechs Minuten Zeit zur Bearbeitung der 24 Aufgaben. Hooven et al. (2008). heben hervor, dass im Gegensatz zu Shepard-Metzler-Aufgaben beim MRT-VK die Reaktionszeit und die Richtigkeit der Testaufgaben nicht als eine Funktion der Winkelgröße bei jedem einzelnen Item erhoben werden könne, da die Leistung nur als Summe richtiger Antworten innerhalb einer limitierten Zeit ausgewertet würde, wodurch eine Überprüfung der ursprünglichen Annahme einer analogen Verarbeitung, wie bei Shepard und Metzler (1971) postuliert, nicht möglich sei.

Vandenberg und Kuse (1978) fanden signifikante Leistungsunterschiede in Abhängigkeit vom Geschlecht zugunsten der Männer. (Heute wird der MRT in seiner revidierten Form nach Peters et al. (1995) verwendet).

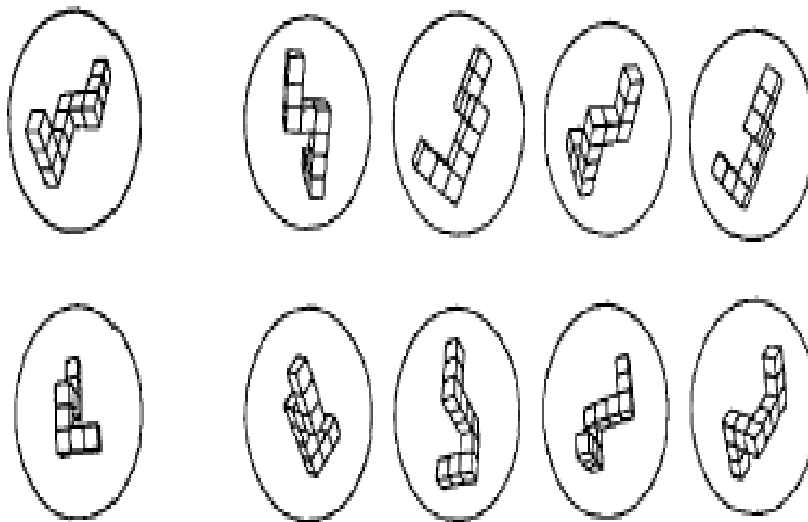


Abbildung 2: Beispielitems des MRT (Vandenberg & Kuse, 1978): In der ersten Reihe sind Figur 1 und 4 ident mit der Ausgangsfigur, 2 und 3 sind Distraktoren und stellen gespiegelte Ansichten der Ausgangsfigur dar. In der zweiten Reihe sind die 2. und 3. Figur die richtigen Alternativen. Die anderen zwei Figuren sind Darstellungen eines anderen Ausgangsbildes.

3.3 SPACE 2d/3d

Heider (2010) konstruierte einen neuen Itempool von Raumvorstellungsaufgaben in Anlehnung an den klassischen Aufgabentyp von Shepard und Metzler (1971). Eine Neuerung bestand in der dreidimensionalen Darstellungsmöglichkeit des Aufgabenmaterials durch eine Rot-Blau-Anaglyphenbrille, wodurch ein alltagsnaher Eindruck des Objektes erzeugt wurde. Eine weitere Modifizierung ihres Aufgabenmaterials gegenüber den klassischen Shepard-Metzler-Figuren stellte das Weglassen des Gitternetzes, also der Linien auf der Figurenoberfläche, dar (siehe Abbildung 1). Eine Testaufgabe bestand darin, eine Ausgangsfigur mit vier weiteren Figuren zu vergleichen, wobei für jede angegeben werden musste, ob sie entweder eine rotierte oder eine gespiegelte und rotierte Ansicht der Ausgangsfigur darstellte (siehe Abbildung 5).

Eine mögliche Erklärung für den signifikanten geschlechtsspezifischen Unterschied und zentrale Hypothese bei Heider (2010) war: Frauen sind genauso gut in ihrer mentalen Rotationsfähigkeit wie Männer, haben aber möglicherweise größere Schwierigkeiten bei der Transformation und Retransformation einer zweidimensionalen Figur in eine dreidimensionale mentale Repräsentation, wie dies bei den meisten klassischen RV-Aufgaben verlangt wird (Strauss, Strasser, Csisinko, Kaufmann & Glück, 2007; vgl. Glück et al., 2005). Tests, die diesen Transformationsprozess abverlangen, fordern somit von den Testteilnehmern eine zusätzliche Leistung, die bei Raumvorstellung im realen Leben nicht erforderlich ist (Kaufmann, Csisinko, Strasser, Strauss, Koller & Glück, 2008). Demnach müssen Aufgabenstellungen, die dreidimensional (stereoskopisch) präsentiert werden, eine höhere ökologische Validität besitzen als die klassischen zweidimensionalen Tests (Kaufmann et al., 2008). Durch den Wegfall des Transformationsprozesses - bedingt durch die dreidimensionale Vorgabeweise der Testaufgaben - könnte man untersuchen, ob dadurch nunmehr Frauen ebenso gut abschneiden würden wie Männer und damit die Hypothese bestätigt würde, dass es die Zweidimensionalität sei, die Frauen mehr Probleme mache als Männern und nicht etwa die mentale Rotation.

Das Weglassen des „Gitternetzes“ basierte auf der Annahme, dass „mit Gitternetz“-Aufgaben eine hohe Wahrscheinlichkeit hätten, auch durch stark analytisch orientierte Strategien gelöst werden zu können. Dadurch würde die beabsichtigte räumliche Verarbeitung und Lösung der Aufgabe jedenfalls nicht bei allen Testpersonen abgerufen und gemessen (auf das Thema der Strategiewahl wird unter Kapitel 4.3 noch ausführlich eingegangen). Außerdem würde ein Aufgabenmaterial, das unterschiedliche Bearbeitungsstrategien und Fähigkeiten zur Lösung der Aufgaben zulasse, die gewünschte Eindimensionalität des Itemmaterials und letztlich die Aussagekraft eines Tests beeinträchtigen.

Heider (2010) konnte für die neuentwickelten Aufgaben (2d-, 3d-Aufgaben und für beide Aufgabentypen zusammen) Eindimensionalität im Sinne von Rasch zeigen.

4. Erklärungsansätze für geschlechtsspezifische Unterschiede in der Raumvorstellung

Nachdem im vorhergehenden Kapitel die Geschichte des bekanntesten Verfahrens zur Erfassung der MR vorgestellt wurde (MRT; Vandenberg & Kuse, 1978), soll im folgenden Kapitel auf die möglichen Ursachen und Einflussfaktoren eingegangen werden, die im Zusammenhang mit geschlechtsspezifischen Unterschieden bei räumlichen Leistungen diskutiert werden. Hauptaugenmerk liegt dabei auf dem Aspekt der Strategiewahl bei Raumvorstellungsaufgaben.

4.1 Bio-psycho-soziale Aspekte

Die Frage, ob die geschlechtsspezifischen Unterschiede bei räumlichen Vorstellungsleistungen biologischer oder sozialer Natur sind, ist von großem Interesse (Gittler & Glück, 1998), da - wie eingangs schon erwähnt - diese Fähigkeit eine hohe praktische Relevanz besitzt (vgl. Maier, 1999). Wäre die RV-Fähigkeit stark von einem „Umweltaspekt“ geprägt, wäre dies vielversprechend hinsichtlich der Veränderbarkeit dieser Fähigkeit durch Training (vgl. z.B. Gittler & Glück, 1998). Als Ursachen für den kognitiven Leistungsunterschied im räumlichen Vorstellungsvermögen zwischen den

Geschlechtern werden heute biologische (z.B. hormonelle Einflüsse; Lateralisation des Gehirns), soziale (z. B. Erfahrung/Erziehung/Training) und psychologische Faktoren (z.B. geschlechtsspezifische Rollenkonzepte, Selbstsicherheit) diskutiert. Für diesbezügliche Information sei der interessierte Leser auf Lautenbacher, Güntürkün und Hausmann (2007) sowie Hirnstein und Hausmann (2010) verwiesen.

Halpern und Tan (2001) schildern einen bio-psycho-sozialen Ansatz, der die beiden Variablen Biologie („Anlage“) und Psychosoziales („Umwelt“) als mit einander interagierend versteht, wobei sich Ursache und Wirkung zirkulär beeinflussen. Sie vertreten die Ansicht, dass es unmöglich sei, Variablen in biologische (Natur) und psychosoziale (Umwelt) Kategorien aufzuteilen. Denn aufgrund der Plastizität unseres Gehirns (z.B. Nelson, 1999; zitiert nach Halpern & Tan, 2001) seien unsere neuronalen Strukturen durch Erfahrungen ein Leben lang veränderbar. Andererseits sei jeder von seiner Biologie her prädisponiert, manche Fertigkeiten leichter zu lernen als andere, was wiederum Einfluss habe auf die Umwelten, die man sich sucht, und somit auf das, was man erfährt. Somit sei die Verbindung zwischen Biologischem und Psychosozialem untrennbar. Ein biopsychosoziales Bezugssystem als Erklärungsgrundlage für kognitive Unterschiede führe zu einer ganzheitlichen Betrachtungsweise jener Prozesse, die Gehirnstrukturen und das Verhalten beeinflussen.

Die Behauptung, dass ein Verhalten von mehreren Variablen beeinflusst wird, wird durch Ergebnisse untermauert, die zeigen konnten, dass Testleistungen mit nicht-kognitiven Variablen (Glück, 2001) wie z.B. dem Selbstkonzept (Glück et al., 2005) korrelieren. So wird zum Beispiel die Testleistung indirekt über die Strategiewahl beeinflusst (Glück, 2001). Auch Hooven et al. (2008) schlossen, dass Selbstsicherheit als Mediatorvariable den geschlechtsspezifischen Einfluss auf die Leistung im MRT-VK mitbedingt. So würden sich Personen, die ein schlechtes Selbstkonzept hinsichtlich ihrer räumlichen Fähigkeiten aufweisen (v.a. Frauen und Mädchen), weniger zutrauen, was zu schlechteren Leistungen führen könne (Glück et al., 2005), da selbstunsichere Personen mehr Zeit aufwänden um z.B. ihre Lösungen zu überprüfen (Hirnstein, Bayer & Hausmann, 2009; Glück & Fabrizii, 2010), oder sie wenden besonders sichere, aber sehr zeitaufwändige Strategien (Glück & Fitting, 2003) an. Doch gerade diese

Vorgehensweise stellt bei (klassischen) mentalen Rotationsaufgaben, bei denen es um die Erfassung der Rotationsgeschwindigkeit geht, einen Nachteil dar (Glück & Fitting, 2003; vgl. Heil & Jansen-Osmann, 2008). Somit würde eine unterschiedliche Strategiewahl und nicht der Prozess der mentalen Rotation als solcher die Unterschiede in der Testleistung (mit-) bedingen (vgl. Peters, 2005).

4.2 Testspezifische Aspekte

Neben den bio-psycho-sozialen Aspekten sollen auch testspezifische Eigenschaften des Aufgabenmaterials und der Aufgabenvorgabe erwähnt werden, da diese in unterschiedlichem Ausmaß Einfluss auf die eigentlich zu erfassende Fähigkeit haben können. Letztlich ist nicht nur der quantitative Aspekt, *wie viele* Aufgaben jemand löst, sondern auch *wie* er oder sie dies tut, eine wichtige diagnostische Information. So sollen Piloten Raumvorstellungsaufgaben mittels Visualisierung und nicht durch sehr gute Reasoning-Fähigkeiten lösen (Glück, 2001). Folglich ist bei der Messung von RV-Fähigkeit die Frage der Dimensionalität des Testmaterials entscheidend (Gittler & Glück, 1998), um in weiterer Folge zu validen Fähigkeitsbestimmungen zu gelangen.

4.2.1 Stimulusmaterial

Glück (2001) formulierte, dass je nach Art des Stimulusmaterials verschiedene Bearbeitungsstrategien in unterschiedlichen Ausmaß angewendet werden können, somit Aufgaben mehr oder weniger strategieinvariant gestaltet sind und tendenziell bestimmte Strategien bevorzugt werden (vgl. Gittler, 1984; Putz-Osterloh, 1977). Dieser Umstand kann als Erklärung für die starken Unterschiede hinsichtlich der Leistungen in den verschiedenen RV-Tests zwischen den Geschlechtern dienen (z.B. Linn & Petersen, 1985). Als ein Beispiel für die Strategievarietät von RV-Material seien die Ergebnisse von Gittler (1984) bei der Entwicklung eines Rasch-homogenen Raumvorstellungstests (3DW; Gittler, 1990) erwähnt.

Zu Beginn war nicht klar ersichtlich, warum der neu entwickelte Itempool aus Würfelaufgaben von Gittler (1978; zitiert nach Gittler, 1984) im Sinne von Rasch

Inhomogenitäten zeigte (Gittler, 1984) (Erläuterungen zum Rasch-Modell unter Kapitel 5.1 oder zur vertiefenden Lektüre bei Rost (2004)).

Example b

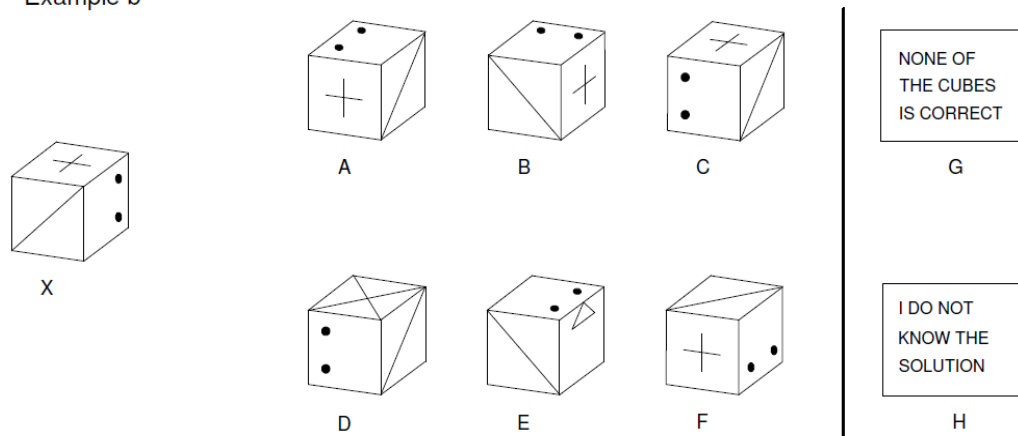


Figure 1: Practice item from the 3DC (solution: D)

Abbildung 3: Übungsbeispiel aus dem 3DW (Gittler, 1990): richtige Antwort ist d (Gittler & Glück, 1998, S. 73).

Erst eine genauere Analyse und inhaltskritische Überlegungen zu möglichen systematischen Unterschieden zwischen den oberflächlich sehr gleichartig aussehenden Testaufgaben (Aufgabenstrukturmerkmale; Gittler, 1984, S. 152; vgl. Gittler, 1986, S. 397) in Zusammenhang mit den angewendeten Bearbeitungsstrategien machten drei verschiedene Lösungsstrategien sichtbar, die im unterschiedlichen Maße visuell-räumliche Bearbeitung verlangten und somit verschiedene latente Dimensionen ansprachen (Gittler, 1984). Bei den Aufgaben wurden Würfel präsentiert, bei denen immer drei Seiten sichtbar waren. Es gab einen Ausgangswürfel und sechs weitere Würfeldarstellungen als Antwortalternativen (siehe Abbildung 3). Es sollte aus diesen sechs Antwortwürfeln derjenige herausgefunden werden, der dem Ausgangswürfel, abgesehen von dessen räumlich veränderter Ansicht, entsprach. Dabei wurden drei verschiedene Herangehensweisen zur Lösungsfindung gefunden (Gittler, 1984):

- Raumstrategie: Vorstellungsmäßiges Drehen und Kippen eines dreidimensionalen Ausgangswürfels hin zum Lösungswürfel.

- Flächenstrategie: Diese Strategie konnte nur dann zur Lösung führen, wenn der Ausgangswürfel dieselben drei Muster zeigte wie der Vergleichswürfel.
Der Würfel wird bei dieser Strategie nicht räumlich wahrgenommen, sondern als „flächige Gebilde“. Man benötigt in der Folge nur eine einfache gedankliche Drehung des Würfels um den Umkreismittelpunkt, um zur Lösung zu gelangen (Gittler, 1984), wodurch eine MR nur im zweidimensionalen Bereich durchgeführt wird (Maier, 1999). Dabei ist die Flächenstrategie mit der Beachtung der Musterrelationen keineswegs ein nicht-räumlicher Zugang, sondern er umgeht dadurch die Imagination von Transformation (Glück, 1999).
- Relationsstrategie: Die drei verschiedenen Muster des Ausgangswürfels werden zueinander in Beziehung gesetzt. Man gelangt zur Lösung, wenn man bei den Antwortalternativen überprüft, ob diese Musterrelationen bei diesen auch stimmen.

Maier (1999) vertritt die Ansicht, dass auch bei einer Relationsstrategie Raumvorstellungsfähigkeiten eingesetzt werden, da die Probanden die „Räumlichkeit“ der Abbildung antizipieren müssen, wenn beim Vergleichswürfel meist ein neues Muster sichtbar wird, während ein anderes verschwindet. Doch sei bei einer solchen Herangehensweise vermutlich das logische Denken als Lösungsstrategie besonders prominent (Maier, 1999) und somit für Aufgaben zur Erfassung von räumlicher Begabung nicht geeignet. So schreibt auch Kubinger (2006), dass das Problematische an der Konstruktion von Papier-Bleistift-RV-Tests sei, dass sie Gefahr laufen, mehr schlussfolgerndes Denken als tatsächlich räumliches Vorstellungsvermögen zu prüfen, wodurch Maier (1999) zufolge RV-Aufgaben auch ohne räumliche Methoden lösbar werden. Dabei scheint sowohl eine inter- als auch intraindividuelle Schwankung in der Anwendung der unterschiedlichen Strategien der wahrscheinlichste Fall zu sein (Gittler, 1984) (siehe auch Kapitel 4.3).

Bei Gittler (1984) waren es jene Aufgaben, die ausschließlich mit einer Raumstrategie lösbar waren, die bei allen Testpersonen die von der Testkonstruktion und vom Aufgabendesign intendierten „räumlichen Kognitionen“ evozierten (Gittler 1990,1992; zitiert nach Gittler, 1999). So seien die gängigsten Methoden der Aufgabenkonstruktion wie hohe Augenscheinvalidität, Konstruktion nach generativen Regeln, Übernahme

„bewährter“ Aufgaben etc. (Gittler, 1992, S. 47) und Verwendung von nonverbalen, bildhaften Aufgabenstellungen allein nicht geeignet, Aufgaben zur Erfassung der RV zu kreieren und Konstruktvalidität zu gewährleisten (Gittler, 1992, S. 47).

4.2.2 Vorgabeweise von Testaufgaben

Bezogen auf den MRT (Vandenberg & Kuse, 1978) meinten Glück und Fabrizii (2010), dass dieser Test einige Charakteristika aufweisen müsse, die zusätzlich zur Erfassung der MR-Fähigkeit den geschlechtsspezifischen Unterschied vergrößern dürften. Ein Grund könnte in der Vorgabeweise der Raumvorstellungsaufgaben liegen. Im Gegensatz zu den Aufgaben bei Shepard und Metzler (1971) (siehe Abbildung 1), bei denen die Versuchspersonen immer nur zwei Bilder gleichzeitig miteinander vergleichen mussten, ist beim MRT (Vandenberg & Kuse, 1978) ein Ausgangsbild mit vier Antwortbildern zu vergleichen, von dem jeweils die zwei richtigen Darstellungen, d.h. die rotierten Ansichten, zu markieren sind.

Hirnstain et al. (2009) beschreiben bei der Bearbeitung dieser Art von Aufgabenstellung zwei Alternativen im Antwortverhalten: Eine sprunghafte versus eine konservative Strategie. Dabei verstehen sie unter einer konservativen Strategie das Betrachten und „Bearbeiten“ aller vier Antwortbilder. Dabei werden nicht nur die zwei richtigen Antworten (rotierte Ansicht) als richtig identifiziert, sondern zusätzlich die zwei falschen (gespiegelte Ansicht) als falsche Antworten verifiziert. Bei der sprunghaften Strategie hingegen begnügt man sich damit, die (nach Ansicht des Bearbeitenden) zwei richtigen Antworten gefunden zu haben.

Da der MRT unter einer Speed-Komponente vorgelegt wird, würden Personen mit einer konservativen Bearbeitungsstrategie im Nachteil sein, da sie aus Zeitgründen nicht dazu kämen, mehrere Aufgaben zu lösen. Hirnstain et al. (2009) halten aufgrund von anekdotenhaften Rückmeldungen fest, dass Männer aufgrund ihrer Selbstsicherheit im Bereich des räumlichen Vorstellungsvermögens eher sprunghafte Strategien anwenden, Frauen hingegen eher konservative Strategien wählen. Außerdem stelle das konservative Antwortverhalten sowie der „Double-Check“ der Antworten (Hirnstain et

al., 2009; Linn & Petersen, 1985) eine weniger effiziente Strategie dar, wenn Testungen unter Zeitdruck erfolgen (Linn & Petersen, 1985). Somit könnte die Selbstsicherheit im Antwortverhalten laut Hirnstein et al. (2009) einen entscheidenden Einfluss auf den geschlechtsspezifischen Unterschied zu Gunsten der Männer haben.

Doch auch eine Verlängerung der Bearbeitungszeit, wie sie Peters (2005) in seiner Studie bot, hielt den signifikanten Leistungsunterschied zwischen den Geschlechtern zu Gunsten der Männer aufrecht. Seine Interpretation bezog sich darauf, dass Frauen länger bräuchten, um eine „falsch“-Antwort (gespiegelte Ansicht) als solche sicher bestätigen zu können. Eine differenzierte Analyse der Bearbeitungsschritte bei MR-Aufgaben ist zu überlegen, da z.B. Peters (2005) zeigte, dass bezüglich des Aspekts der mentalen Rotationsgeschwindigkeit an sich geringe oder keine geschlechtsspezifischen Unterschiede zu finden waren, diese somit in anderen Aspekten der Aufgabenbearbeitung zu suchen seien. So meinen auch Hooven et al. (2008), dass der höhere Testwert der Männer nicht in einer Überlegenheit in der mentalen Rotationsfähigkeit per se zu sehen sei, sondern dass diese mit anderen Aspekten der Aufgabenperformance zusammenhänge, was für einen „ganzheitlichen“ Forschungsansatz im Sinne des bio-psycho-sozialen Modells mit Beachtung der Wechselwirkungen der drei unabhängigen Variablen spräche.

Um den Einfluss des unterschiedlichen Antwortverhaltens bei konservativen bzw. sprunghaften Strategie zu umgehen, haben Hirnstein et al. (2009) Aufgabenstellungen vorgelegt, bei denen 0-4 Antwortmöglichkeiten richtig sein konnten. Somit wurde von allen Teilnehmern die Bearbeitung aller Antwortbilder verlangt. Es zeigte sich, dass sich die Leistungen von Männern und Frauen angleichen, wobei ein größerer Leistungsabfall bei Männern in der modifizierten Version gefunden wurde. Trotzdem blieben die Testleistungen der Männer signifikant besser. Außerdem halten die Autoren am Ende fest, dass durch das veränderte Antwortformat der Leistungsunterschied zwischen den Geschlechtern im MRT nicht berührt werde, dass das unterschiedliche Antwortverhalten von Männern und Frauen den robusten, geschlechtsspezifischen Unterschied in MR nicht bedinge.

Als letzten Aspekt der möglichen Einflussgrößen auf die RV-Leistung soll neben den bio-psycho-sozialen und testspezifischen Aspekten auch noch die Strategiewahl beim Lösen von Raumvorstellungsaufgaben angesprochen werden.

4.3 Strategiewahl

Es gibt zwischen Männern und Frauen Präferenzen in der Art der Informationsverarbeitung speziell im Bereich der räumlichen Informationen (Glück, 1999; vgl. Glück & Fitting, 2003; Glück et al., 2005). So könnte man sich eine Route als räumlich-bildliche Darstellung oder als eine listenartige Aufzählung von bestimmten Orientierungspunkten/Kennzeichen der Route merken (Glück & Fitting, 2003). Dieser Umstand hat auch Implikationen bei der Lösung von RV-Aufgaben, bei der Frauen eher analytische und Männer eher holistische Strategien anwenden (Glück et al., 2005; vgl. Peters et al., 1995). Genauere Erläuterungen zu den Strategien erfolgen ab Kapitel 4.3.

Eine psychologische Erkenntnis im Zusammenhang mit räumlichem Denken ist, dass räumliche Aufgaben auf verschiedene Art gelöst werden können (Glück et al., 2005). Vor allem komplexe RV-Aufgaben lassen sich oft genauso gut durch logisches Denken lösen wie durch Vorstellung (Glück et al., 2005; vgl. Putz-Osterloh, 1977). An diesem Beispiel erkennt man, wie wichtig es ist, die Art der Aufgabengestaltung in Hinblick auf mögliche Bearbeitungsstrategien zu analysieren, wenn man eine bestimmte Fähigkeit erfassen will (wie an früherer Stelle anhand der Entwicklung des 3DW (Gittler, 1990) schon gezeigt wurde).

Wie schon erwähnt, erlauben RV-Aufgaben in verschiedenem Ausmaß unterschiedliche kognitive Lösungsstrategien (Glück, 2001; vgl. Glück et al., 2005). Dadurch können Aufgaben als unterschiedlich schwierig empfunden werden, je nachdem mit welcher Strategie man sie löst (Gittler & Glück, 1998). In Abhängigkeit von der verwendeten Strategie werden verschiedene Fähigkeitsdimensionen gemessen und ein aussagekräftiger Vergleich zwischen Personen, die die Strategie A und solchen, die Strategie B angewandt haben, ist nicht möglich (Gittler & Glück, 1998).

Wie werden nun RV-Leistungen, insbesondere solche in mentaler Rotation, erbracht? Die bisher angeführten Ergebnisse schließen direkt an die Debatte an, wie wir eigentlich visuelle Informationen in unserem Gehirn verarbeiten und repräsentieren (Gittler, 1992). Glück und Fitting (2003) vertreten die Ansicht, dass sowohl die Forschung als auch die Messung der RV durch die Integration des Strategieaspekts und der daraus gewonnenen Informationen profitieren. Außerdem könnte mit der Erfassung der eingesetzten Strategie eine mögliche Erklärung für die geschlechtsspezifischen Unterschiede in RV-Fähigkeit gelingen (Glück, 2001). Strategiedifferenzen scheint es dabei nicht nur zwischen den verschiedenen RV-Tests (Voyer et al., 1995) zu geben. Bedingt durch deren unterschiedlich stark strategie-invariante Gestaltung, (Glück, 2001), treten sie sowohl zwischen verschiedenen Personen als auch innerhalb ein und derselben Person auf (Glück & Fitting, 2003).

Dies hängt einerseits mit den Aufgabencharakteristika zusammen, wobei mit steigendem Komplexitätsgrad der Aufgabenstellung zu analytischen Bearbeitungsstrategien gewechselt wird, z.B. durch Betrachtung von Einzelteilen oder Verbalisierungen (Glück, 2001). Andererseits liegt ein weiterer Grund für die unterschiedlichen Strategieanwendungen darin, dass der Komplexitätsgrad der Aufgaben individuell unterschiedlich empfunden wird. So kann der Fall eintreten, dass bei ein und derselben Aufgabe gute Raumvorsteller die Aufgabe visuell-räumlich lösen, während schwache Raumvorsteller, bei denen die Kapazität des visuellen Arbeitsgedächtnisses, Geschicklichkeit bei der Encodierung der Stimuli etc. (Glück, 1999, S. 247) nicht so ausgeprägt sind, diese Aufgaben mit analytischen Strategien bearbeiten (Glück, 1999; vgl. Glück, 2001).

Da ein Strategiewechsel innerhalb ein und derselben Person im selben Test vorkommen kann, bedingt durch unterschiedlich (empfundene) Aufgabenschwierigkeiten, sollte eine Strategieauswertung nicht personenweise, sondern itemweise erfolgen (Glück, 1999). Das Rasch-Modell kommt durch seine Eigenschaften dieser Forderung nach, indem es systematische Schwierigkeitsunterschiede der Items zwischen verschiedenen Personengruppen bestimmen kann.

Glück und Fitting (2003) beschreiben diese beiden Bearbeitungsweisen als die Endpunkte eines Kontinuums mit den beiden Polen holistische vs. analytisch-verbale. Diese beiden Kategorien schließen sich nicht gegenseitig aus, sondern lassen - ganz im Gegenteil - auch „intermediäre“ Strategien zu.

4.3.1 Holistische Strategien

Holistische Strategien können als die Manipulation räumlicher Informationen in Form einer mentalen Repräsentation verstanden werden (Glück & Fitting, 2003), wobei der Stimulus als Ganzes bearbeitet wird (Linn & Petersen, 1985), so wie es Shepard und Metzler (1971) anhand ihrer Ergebnisse angenommen haben. Bei dieser Strategie wird somit von einem analogen Prozess (Glück, 1999) ausgegangen, wodurch ausschließlich räumlich-visuelle Mittel zur Lösungsfindung eingesetzt werden (Grüßing, 2002). Holistische Strategien erfordern einen höheren kognitiven (Informationsverarbeitungs-) Aufwand, führen aber grundsätzlich schneller zum Ziel, was v.a. bei zeitbegrenzten Tests, wie zum Beispiel dem MRT (Vandenberg & Kuse, 1978), von Vorteil ist (Glück et al., 2005; vgl. Heil & Jansen-Osmann, 2008).

4.3.2 Analytische Strategie

Unter einer analytischen Bearbeitungsweise versteht man die Betrachtung von Einzelteilen bzw. Details der Objekte, die zum Vergleich herangezogen werden, wodurch die Komplexität der räumlichen Information reduziert wird (Glück & Fitting, 2003). Ein Beispiel für eine solche Situation stellt die Studie von Yuille und Stieger (1982) dar. Sie erhöhten die Aufgabenkomplexität der Shepard-Metzler-Figuren (1971), indem sie weitere Würfel hinzufügten, wodurch eine holistische Erfassung (zu) anstrengend und somit nach anderen, alternativen Strategien gesucht wurde. Dabei entdeckten sie die Möglichkeit, redundante Informationen zu identifizieren, wodurch nicht mehr die gesamte Figur visuell erfasst und geistig bearbeitet werden musste. Durch diesen Lösungsweg verlor sich auch der lineare Zusammenhang zwischen Winkelgröße und Zeit, wie ihn Shepard und Metzler (1971) beschrieben hatten.

Auch wenn manchmal analytische als verbale Strategien bezeichnet werden, kommt nicht notwendigerweise Verbalisierung im Lösungsprozess zum Zuge (Glück & Fitting, 2003). Als Beispiel einer verbal-analytischen Strategie sei die „Corridorstrategie“ (Just & Carpenter, 1985) kurz beschrieben. Bei der Anwendung einer Corridorstrategie bei Shepard-Metzler-Figuren (1971) wird statt der Visualisierung der ganzen Figur die Stimulusfigur anhand einer Serie von Drehungen encodiert und sprachlich codiert: drei Würfel rauf, zwei rechts, zwei rauf, vier hinunter etc. Eine solche analytische Repräsentation räumlicher Information braucht mehr Zeit, ist aber weniger anstrengend (Glück & Fitting, 2003).

French (1965, S. 22; zitiert nach Maier, 1999) zeigt den drastischen Umstand auf, dass ein Test nicht in der Lage ist, RV-Fähigkeit zu erfassen, wenn neben visuellen auch analytische Strategien eingesetzt werden können. Das könne so weit gehen, dass ein ursprünglich geplanter RV-Test zu einem „Beweisführungs“-Test umgewandelt wird. Vor allem bei anspruchsvollen Aufgaben würden Ergänzungs- und Ausweichstrategien (Maier, 1999, S. 69), also analytische Strategien, zur Aufgabenlösung statt der intendierten visuell-räumlichen Bearbeitungsstrategie zum Einsatz kommen.

4.3.3 Intermediäre Strategie

Ein Beispiel für eine intermediäre Vorgehensweise, die sowohl analytische als auch räumliche „Informationen“ verwendet, wäre das Unterteilen eines Stimulus und das anschließende serielle Rotieren der Einzelelemente, um zwei Darstellungen miteinander zu vergleichen (Heil & Jansen-Osmann, 2008).

Dies kann dann verwendet werden, wenn das ganzheitliche Erfassen des Stimulus zu schwer fällt. Dann wird der Stimulus untergliedert und jedes Element für sich nacheinander rotiert, wobei bei diesem Zugang und den damit verbundenen Bearbeitungsschritten die Fähigkeit, sich die vorhergehenden Ergebnisse der Teilrotationen zu merken, mit ins Ergebnis einfließt (Glück, 1999). Solch eine intermediäre Strategie (Glück & Fitting, 2003) ist zeitaufwändiger, v.a. wenn man den Überblick verliert und die Teilrotationen noch einmal von Anfang an durchführen muss (Glück, 1999; vgl. Heil & Jansen-Osmann, 2008).

Glück (1999) hält zusammenfassend fest, dass auch bei „Nicht-Rotationsstrategien“, also bei eher analytisch-serieller Encodierung statt einer ganzheitlich-holistischen, räumliche Manipulationen an den encodierten Stimuli durchgeführt werden müssen.

4.4 Strategien bei Mentaler Rotation (MR)

Da das Hauptaugenmerk der vorliegenden Arbeit auf der mentalen Rotationsfähigkeit liegt, wird im Folgenden der Strategieaspekt noch genauer im Zusammenhang mit eben jener betrachtet. In diesem Kontext stellt sich die Frage, wie eigentlich visuelle Informationen im Gehirn verarbeitet und repräsentiert werden (vgl. Gittler, 1992). Bezogen auf die Subdimension der MR-Fähigkeit zentriert sich die Debatte darauf, ob MR einen analogen Prozess zu einer physischen Rotation darstelle, oder ob Probanden analytische Verarbeitungsstrategien bei solcher Art von Aufgabenstellungen verwenden (z.B. Linn & Petersen, 1985).

Wovon hängt es nun ab, ob wir die Strategie der mentalen Rotation bei (vermeintlichen) MR-Aufgaben verwenden oder nicht? Gittler (1992) beschreibt in diesem Kontext, dass Items einen Aufforderungscharakter haben, eine bestimmte Strategie anzuwenden. Glück (1999) führt dazu aus, dass selbst bei MR-Aufgaben nicht immer mentale Rotation als Bearbeitungsstrategie angewendet wird, auch wenn dieser ganzheitliche, analoge Prozess den effektivsten Zugang darstellt (Linn & Petersen, 1985, S. 1485), v.a. bei Vorgabe unter Zeitdruck wie im Fall des MRT (Glück et al., 2005). Gründe hierfür seien:

- Zu komplex gestaltete Aufgaben, da in einem solchen Fall auf analytische Strategien ausgewichen wird. Denn nur einfache oder vertraute Objekte, die im Langzeitgedächtnis bereits repräsentiert sind, können durch ihre ressourcensparende, holistische Verarbeitung auch schnell mental gedreht werden (vgl. Glück, 1999).
- Außerdem dürften Detailvergleiche der Stimuli nicht alleine ausreichen, die Lösung zu finden (Glück, 1999).
- Erlaubt die Aufgabenstellung bzw. die Darstellung der Stimuli die Bearbeitung mit einer vorstellungsfreien Strategie, wird diese eher angewendet, da sie ökonomischer

(leichter) ist als das räumliche Vorstellen (Linn & Petersen, 1985), jedoch beansprucht diese Strategie mehr Zeit (Glück, 1999).

4.5 Zugänge zur Strategieerhebung

Hier sei kurz auf die Güte der erhaltenen Daten im Zusammenhang mit der Art der Erfassung von Bearbeitungsstrategien bei räumlichen Vorstellungsaufgaben hingewiesen. So erhoben z.B. Peters et al. (1995) die Auskünfte ihrer Teilnehmer über die angewendeten Strategien bei Shepard-Metzler-Aufgaben erst nach Beendigung des Tests. Obwohl eine valide Strategieerfassung bedeutsam ist (Glück, 2001), fällt diese sehr schwer, da jede Strategieerfassungsmethode und jeder Zugang Einschränkungen aufweist (Lohman & Kyllonen, 1983; zitiert nach Glück & Fitting, 2003). Deswegen sei eine konvergente Evidenz von mehr als einer Methode für reliable und valide Ergebnisse wichtig (z.B. Lohmann & Kyllonen, 1983; zitiert nach Glück & Fitting, 2003).

Glück (2001) listet als gängigste Methoden zur Strategieerfassung bei RV-Aufgaben die direkten bzw. retrospektiven Befragungen, die Antwortlatenzerfassungen, die Blick-Bewegungs-Analysen und die Antwort-Fehler-Muster-Analysen auf. Dabei wird auf die Einschränkung hingewiesen, dass nicht alle dieser Methoden auf derselben Ebene operieren. So könnten z.B. nicht alle Erhebungsmethoden Informationen auf Einzelitemniveau liefern.

Grüßing (2002) kritisiert bei retrospektiven Interviews zur verwendeten Strategie, dass das tatsächliche Vorgehen verkürzt dargestellt werde. Dabei würden auch Rechtfertigungen für die gefundene Lösung und nachträgliche Überlegungen zum Lösungsprozess mit in die Beschreibung eingehen, wodurch die Validität solcher „Selbstberichte“ fraglich sei (Glück & Fitting, 2003). Des Weiteren erfolgt die Strategieerfassung bei diesem Zugang nicht bei jedem Item, wodurch am Ende eher nur die am häufigsten zum Einsatz gekommene Strategie erhoben wird. Dadurch wird das Erfassen eines möglichen Zusammenhangs zwischen Strategiewahl und bestimmten Aufgabenmerkmalen nicht möglich (vgl. Glück, 1999).

Die Methode des lauten Denkens, so Grüßing (2002), würde das Problemlöseverhalten beeinflussen, in dem sie, bedingt durch den Einsatz von Sprache, zu einer analytischen Strategiewahl verleite.

Bisher gibt es erst wenige Studien, die sich explizit mit Strategiedifferenzen bei Raumvorstellungsaufgaben zwischen Männern und Frauen befassen (Glück & Fitting, 2003, S. 298). Dabei wäre die Ursachenfindung für eine unterschiedliche Strategiewahl eine wichtige Frage in der psychologischen Forschung. Speziell für die Teildisziplin der Differentiellen Psychologie, die sich mit dem Zustandekommen von Gruppenunterschieden (auch mit jenen zwischen Männern und Frauen) befasst, wäre sie von großem Interesse (Glück, 2001).

5. Methodik und Erhebungsverfahren

Neben den eingangs erwähnten Schwierigkeiten, Subfaktoren des RV-Konstrukts auf Basis der Faktorenanalyse zu bestimmen und neben den Erkenntnissen zur mangelnden Eindimensionalität von RV-Tests, soll im Folgenden noch etwas genauer auf die bisherigen messtheoretischen Problematiken eingegangen werden.

So führen Putz-Osterloh und Lüer (1979) als weiteren Nachteil der Faktorenanalyse den Umstand an, dass man im Zuge einer Faktorenanalyse nur Testwerte (Menge richtig gelöster Aufgaben) erfassen könne. Dieser Zugang lasse keine Aussagen über das dahinterliegende Lösungsverhalten bzw. die zum Einsatz gekommenen Strategien zu (Putz-Osterloh & Lüer, 1979). Erst die Möglichkeit, Aussagen auf Einzelitemniveau zu machen, wie dies das Rasch-Modell erlaubt, ermögliche es zu prüfen, ob Items sich in ihrer Dimensionalität (zugrundeliegenden Fähigkeit) unterscheiden (vgl. Gittler, 1984), und ob Items desselben Aufgabentyps bei verschiedenen Personengruppen, wie z.B. bei Männern und Frauen, überhaupt dieselbe latente Fähigkeitsdimension erfassen (Gittler, 1999).

Neben Putz-Osterloh (1977) und Putz-Osterloh und Lüer (1979) konnte auch Gittler (1978; 1983, zitiert nach Gittler, 1984) berechtigte Zweifel an der Homogenität und Validität von Würfelaufgaben (Gittler, 1984) äußern. Dabei gingen Putz-Osterloh (1977) und Putz-Osterloh & Lüer (1979) von einem „denkpsychologischen“

Untersuchungsansatz aus und Gittler von probabilistischen Testmodellen (Gittler, 1984).

Am Anfang eines denkpsychologischen Untersuchungsansatzes steht eine theoretische Analyse des Problems, aus der sich dann mögliches Problemlöseverhalten ableitet (Putz-Osterloh & Lüer, 1979). Durch diesen Zugang entdeckten die Autoren, dass RV-Aufgaben keineswegs selbstverständlich mit räumlicher Vorstellung gelöst werden. So kann ein denkpsychologischer Ansatz mit Problemstrategieanalysen auch zur Beurteilung der Gültigkeit eines Tests beitragen (Putz-Osterloh & Lüer, 1979). Außerdem wirkt ein solcher Zugang dem Misslingen von Kreuzvalidierungen entgegen (Gittler, 1986, vgl. Putz-Osterloh, 1977) und kann als Erfolg bringende Grundlage für eine inhaltliche, theoriegeleitete Itemselektionsstrategie im Rahmen von Rasch-Analysen dienen (vgl. Gittler, 1984).

Eine theoretische Aufgabenanalyse als Grundlage der Beschreibung von Lösungsverhalten reicht aber nicht aus, sondern muss erst durch empirische Beobachtungsdaten überprüft werden (Putz-Osterloh & Lüer, 1979), wobei das Rasch-Modell zur Dimensionalitätsprüfung der Items und somit des Tests (Homogenitätsproblematik; vgl. Gittler, 1984, 1990; zitiert nach Gittler & Arendasy, 2003) essenziell ist. Nur wenn tatsächlich *eine* Lösungsstrategie zur Lösung führt, kann das Ergebnis ohne Bedenken interpretiert werden und bezieht sich eindeutig auf das zu messende Merkmal. Wenn hingegen verschiedene Personen unterschiedliche Fähigkeiten zur Lösung ein und derselben Aufgabe anwenden, wird die Interpretierbarkeit des Testwertes „beeinträchtigt“, da nicht mehr nur Vergleichbares verglichen wird (Glück & Fitting, 2003).

Im Folgenden soll das bisher schon oft zitierte Rasch-Modell vorgestellt werden, da für die Messung der RV-Fähigkeit die Dimensionalität des Testmaterials essenziell ist (Gittler & Glück, 1998).

5.1 Rasch-Modell

Um der Forschungsfrage nachzukommen, ob die beiden Aufgabenmodi einerseits für sich und andererseits zusammen nur eine Fähigkeitsdimension erfassen, wird das dichotom logistische Modell von Rasch (1960/1980; zitiert nach Gittler & Arendasy, 2003, S. 165ff) angewendet. Sollte sich bei den vorliegenden Daten die Gültigkeit des Rasch-Modells zeigen, würde das bedeuten:

- a) dass für alle Personen (jung/ alt, Mann/Frau, begabt/unbegabt etc.) essenziell dieselbe Fähigkeitsdimension erfasst wird und
- b) dass der Testwert (die Anzahl gelöster Aufgaben im Test = Summenscore) ein faires Leistungsmaß darstellt, da es alle relevanten Informationen in Bezug auf eine Fähigkeit enthält (Arendasy & Gittler, 2003). Somit würden die einzelnen Items mit den übrigen Items aus dem Test dieselbe Fähigkeitsdimension erfassen (Gittler & Arendasy, 2003) und man könnte daraus folgern, dass die Testpersonen dieselbe Lösungsstrategie bei den beiden Aufgabenmodi angewendet haben (vgl. Gittler, 1984; für genauere Ausführungen zum RM siehe z.B. Rost, 2004). Außerdem erlauben die Anwendung intervallskalierter Personenparameter parametrische statistische Verfahren (Arendasy & Gittler, 2003).

Beim Rasch-Modell wird der Zusammenhang zwischen Testwert und latenter Fähigkeitsdimension als logistische Funktion dargestellt, was nach Arendasy und Gittler (2003) folgendes bedeutet:

Die Wahrscheinlichkeit für die Person v , eine richtige Antwort („+“) auf die Testaufgabe i zu geben, hängt nur von der Personenfähigkeit (θ_v) und der Itemschwierigkeit (ε_i) ab. Wenn für eine Itemmenge das Rasch-Modell empirisch nicht verworfen werden muss, dann ist dieser Test *eindimensional* (Rasch-homogen) in dem Sinne, dass die Lösungswahrscheinlichkeit für alle Items nur vom Parameter θ_v der jeweiligen Testperson abhängt. Das heißt, dass alle „latenten“ Personenmerkmale, die - neben den Itemeigenschaften - die

Reaktionen einer Testperson bestimmen, vollständig in diesem einen Parameter θ_v „zusammengefasst“ werden können. Eine Verletzung der oben genannten Annahme würde bedeuten, dass gewisse Personenmerkmale bei verschiedenen Aufgaben mit unterschiedlicher Gewichtung zur Reaktion beitragen (S. 262).

Denn würden in einer Erhebung einige Teilnehmer die Aufgaben mit einer holistischen und andere Teilnehmer die Aufgaben mit einer analytischen Lösungsstrategie bearbeiten, würde dies zu unterschiedlichen Itemschwierigkeiten führen (Gittler & Glück, 1998; Glück, Machat, Jirasko & Rollett, 2002). In diesem Fall würde in den zwei Subgruppen jeweils etwas anderes gemessen werden und ein fairer Vergleich der Ergebnisse wäre nicht möglich. Das Rasch-Modell kann bestimmen, ob für alle Untergruppen dieselbe Fähigkeitsdimension anhand der Items erfasst wird.

Gilt das Rasch-Modell (kurz: RM) für einen bestimmten Itempool, können Items trotzdem im Durchschnitt für die jeweiligen Subgruppen noch unterschiedlich schwierig sein, z.B. für Frauen und Männern. Das Entscheidende ist, dass die Rangordnung der Items und die Relationen zwischen den Itemschwierigkeiten für beide Subgruppen gleich sind (Gittler & Glück, 1998).

Der Likelihood-Ratio-Test (LRT) von Andersen (1973; zitiert nach Gittler & Arendasy, 2003) prüft die Itemparameterschätzungen aus vordefinierten Personengruppen (z.B. Männer/Frauen, niedriger Testwert/hoher Testwert usw.) auf ihre statistische Gleichheit (Gittler & Arendasy, 2003). Die Untergruppen erhält man, indem man die Stichprobe anhand sogenannter Teilungskriterien in zwei Gruppen aufgliedert, z.B. am Median des Testwerts (hoher vs. niedriger Testwert). Erhält man nach Durchführung der Modellkontrollen signifikante Ergebnisse, folgt daraus, dass das Rasch-Modell nicht gilt. Die klassische Herangehensweise ist es nun, jene Items, die bei graphischen Modellkontrollen und signifikanten z-Werten auffällig werden, sukzessive zu eliminieren. Auf dieses rein „technische“ Vorgehen bei signifikanten Rasch-Modell-Analysen geht Gittler (1986) kritisch ein (siehe übernächsten Absatz).

Bei den graphischen Modellkontrollen wird jedem Item ein Punkt in einem rechtwinkligen Koordinatensystem zugewiesen. Wenn das Modell gilt, liegen alle Items auf einer 45°-Geraden, wobei ein gewisser Verwerfungsbereich entlang der Geraden toleriert wird (vgl. z. B. Formann, 2004, S. 89).

In einer Studie von Gittler (1978; zitiert nach Gittler, 1984) wurden an einem neu entwickelten Testmaterial Itemanalysen mit dem RM durchgeführt. Nachdem der LRT nach Andersen (1973; zitiert nach Gittler & Arendasy, 2003) nach dem trennschärfsten Teilungskriterium, dem Testwert (Gittler, 1984), signifikant wurde, wurden gemäß der damaligen Praxis die Items entfernt, die die Homogenität am höchsten zu beeinträchtigen schienen. Die mit den übrig gebliebenen Aufgaben neuerlich durchgeführten Modelltests zeigten keine Signifikanzen mehr, die Geltung des RM konnte angenommen werden. Doch nach neuerlicher Überprüfung dieses vermeintlich Rasch-homogenen Itempools an einer neuen Stichprobe, erkannte man das Problem der artifiziellen Modellanpassung, eine Scheinhomogenität, wenn man ausschließlich technische Kriterien beim Herstellen homogener Skalen verwendete (Gittler, 1984).

Gittler (1986) griff die Kritik an der artifiziellen Modellanpassung des Rasch-Modells auf und beschreibt einen Ansatz, wie man diesem Problem entgegentreten könne und müsse, wolle man relevante, valide und sinnvolle Ergebnisse erhalten. Die Problematik beim Rasch-Modell tritt dann zu Tage, wenn bei den Modellgeltungskontrollen auffällig gewordene Items lediglich aufgrund von „technischen“ Kriterien (z-Werte, Modell-Graphiken) ausgeschieden werden, da diese keine geeigneten Ausscheidungskriterien zur Erlangung einer eindimensionalen Skala (Gittler 1986, S. 391) darstellen.

Diesem Umstand stellt Gittler (1986) eine psychologisch-inhaltlich begründete und somit theoriegeleitete Selektionsstrategie gegenüber. Demnach sollen auffällige Items nicht einfach auf Basis der „technischen“ Kriterien ausgeschieden werden, sondern durch inhaltliche Vorüberlegungen bezüglich möglicher „systematischer Unterschiede“ (Gittler, 1984) zwischen den Item- und/oder Personengruppen reflektiert werden. Zeigt sich Inhomogenität, kann diese anhand der Vorüberlegungen besser interpretiert werden. Geht man mit inhaltlichen Vorannahmen an eine Modellprüfung heran, kann

bei insignifikanten Ergebnissen nicht nur Modellgeltung gezeigt werden, sondern man erhält dadurch auch gleichzeitig wesentliche Informationen über inhaltliche Gegebenheiten des Aufgabenmaterials (Gittler, 1992).

Gittler (1986) betont, dass die im Rahmen von Validierungsstudien vorgefundenen artifiziellen Modellanpassungen nicht einen Mangel des Rasch-Modells darstellen, sondern dass die Itemselektionsstrategien (Gittler & Arendasy, 2003) fehlgeschlagen seien.

Somit soll die Frage, *warum* ein Testmaterial Heterogenität aufweist, durch Hinzunahme weiterer Informationen (theoriegeleitete Vorüberlegungen hinsichtlich möglicher systematischer Unterschiede des Aufgabenmaterials; Gittler, 1984), befriedigender gelöst werden. Das Wissen um die Möglichkeit der Anwendung unterschiedlicher Bearbeitungsstrategien (statt der in diesem Fall intendierten ausschließlich visuell-räumlichen Leistung), bedingt durch Strukturmerkmale der Aufgaben (in dieser Arbeit ist dies das „Gitternetz“), stellt die Grundlage inhaltlicher Selektionsstrategien dar (Gittler, 1986).

Das Ziel von Heider (2010) war es, einen Rasch-homogenen RV-Test zu entwickeln. Dem ging die theoriegeleitete Überlegung voraus, das Gitternetz im Aufgabenmaterial wegzulassen, um so die intendierte Bearbeitungsstrategie, also eine visuell-räumliche und holistische Verarbeitung, ehest möglich zu evozieren und damit eine gelungene Operationalisierung des „RV-Konstrukts“ zu gewährleisten (vgl. Gittler, 1984, 1999). Dadurch sollte theoriegeleitet eine (mögliche) Quelle für Inhomogenitäten (das Vorhandensein des „Gitternetzes“) schon im Vorhinein vermieden werden, ähnlich dem Weglassen der Flächenaufgaben im Rahmen der Entwicklung des 3DW (Gittler, 1984). Im Rahmen der vorliegenden Arbeit dienen auch die Zusatzfragen dazu, einerseits die Ergebnisse besser inhaltlich interpretieren zu können und andererseits mögliche a-priori Hypothesen für zukünftige Untersuchungen (Gittler, 1992) im Sinne der theoriegeleiteten Itemselektion zu generieren.

Ein weiterer Kritikpunkt an der Leistungsfähigkeit des „dichotom logistischen Modells von Rasch“ (1960/1980; zitiert nach Gittler, & Arendasy, 2003, S. 165ff) wurde von Stelzl (1979) aufgezeigt. So sei der LQT von Andersen (1973; zitiert nach Gittler & Arendasy, 2003) nicht sensitiv gegenüber Inhomogenitäten im Itempool, wenn dieser aus zwei in sich Rasch-homogenen Skalen bestünde, die somit jeweils Verschiedenes messen (Formann, 1981).

Als Lösung für die von Stelzl (1979) aufgezeigte Problematik schlägt Formann (1981) vor, im LQT von Andersen (1973; zitiert nach Gittler & Arendasy, 2003) einzelne Items als Teilungskriterien zu verwenden:

„Denn ohne genauere Vorkenntnis sind nur die einzelnen Items unmittelbarer Ausdruck all dessen, was an verschiedenen Eigenschaften durch einen Test gemessen wird, so dass durch eine derartige Modellprüfung alle Inhomogenitäten aufdeckbar sein müssten, die auf das Zusammenspiel mehrerer Dimensionen zurückführbar sind“ (S. 553f).

Der Martin-Löf-Test (1973; zitiert nach Gittler, 1999) stellt einen weiteren Zugang zur Überwindung der von Stelzl (1979) aufgezeigten Problematik dar, indem die Items, geleitet durch eine Hypothese über möglicherweise vorliegende strukturelle Unterschiede, in zwei Gruppen geteilt werden (Gittler, 1999; „systematische Unterschiede“; vgl. Gittler, 1984, 1986). Dadurch kann die Dimensionalität von Aufgabensubgruppen direkt geprüft werden (Gittler & Arendasy, 2003).

Da Rasch-Homogenität mit inhaltlicher Eindimensionalität der Aufgaben gleichgesetzt wird, müssten inhaltlich-psychologische Theorien das Ausscheiden von Items erklären können (Gittler, 1986, S. 391). Dabei weist Gittler (1999) darauf hin, dass Rasch-Modell-Gültigkeit, also die Eindimensionalität von Testaufgaben, nicht notwendigerweise bedeutet, dass die Personenfähigkeit sich aus einer einzigen zugrundeliegenden psychologischen Dimension zusammensetzt. So können Leistungen in einem spezifischen Leistungstest durch eine Kombination aus primärer Fähigkeit und sogenannten Ko-Faktoren erbracht werden: Z.B. Raumvorstellungsfähigkeit als Primärfähigkeit und Konzentrationsvermögen als Ko-Faktor. Die Rasch-Modell-Gültigkeit

bedeutet, dass die relativen Itemparameter (Itemschwierigkeiten) durch diese beiden Komponenten nicht differenziell beeinflusst werden dürfen.

Erst durch eindimensionale Testaufgaben ist gewährleistet, dass für alle Personengruppen und Personen essenziell dasselbe gemessen wird (Gittler & Arendasy, 2003) und in weiterer Folge auch eine Evaluation der (Konstrukt-)Validität eines Tests möglich ist. Denn lediglich aufgrund der Gültigkeit des Rasch-Modells sind keine Aussagen zur Validität eines Tests, also hinsichtlich dessen, was der Test letztlich misst, ableitbar (Gittler, 1999).

Zur Klärung der Frage, welche Dimension (bzw. welches theoretische Konstrukt) bei Gültigkeit des Rasch-Modells tatsächlich erfasst wird, bedarf es theoriegeleiteter Hypothesen, die den Rasch-Modell-Analysen zugrunde liegen. Das sind im vorliegenden Falle Erkenntnisse zum Thema Strategiewahl bei RV-Aufgaben, und bei Gültigkeit des Rasch-Modells eine anschließende Überprüfung des Aufgabenmaterials mittels z.B. dem linear-logistischen Testmodell LLTM (Fischer, 1973; zitiert Gittler, 1984).

Das LLTM erlaubt die empirische Überprüfung zuvor hypothetisch formulierter Item-Komplexitätsdifferenzen (Gittler & Glück, 1998). Es ermöglicht, die einzelnen kognitiven Operationen bei der Aufgabenbearbeitung herauszukristallisieren (Gittler, 1992), um z.B. Schwierigkeitsunterschiede innerhalb eines Rasch-homogenen Itempools zu identifizieren. Damit erhält man eine brauchbare psychologische Theorie über das Lösungsverhalten der Probanden (Gittler, 1984).

5.2 Space 2d-2 Modi: „ohne-Gitter“- und „mit-Gitter“-Aufgaben

Ausgehend von der Frage, ob die Eindimensionalität der „oG“-Aufgaben nach Hinzufügen des Gitternetzes ebenso gegeben sein würde, wurde zu deren Beantwortung dem Aufgabenpool von Heider (2010) ein Gitternetz hinzugefügt. Durch die gleichzeitige Vorgabe von „ohne-Gitter“-Aufgaben (kurz: oG-Aufgaben) und „mit-Gitter“-Aufgaben (kurz: mG-Aufgaben) wäre eine Überprüfung des Zusammenhangs

zwischen Strukturmerkmalen und Bearbeitungsstrategien (Gittler, 1986) möglich, wodurch eine Erklärung für die Leistungsunterschiede zwischen den Geschlechtern ableitbar wäre.

Diese Herangehensweise ist angelehnt an den Ansatz eines theoriegeleiteten Vorgehens zur Skalenkonstruktion mittels RM, bei dem Items nach signifikanten Rasch-Modell-Analysen auf der Basis von primär formulierten, inhaltlich begründeten Hypothesen über „systematische Unterschiede“ (Gittler, 1984) ausgeschieden werden (Gittler & Arendasy, 2003; vgl. Gittler, 1984, 1986).

Da eine genaue Gradangabe über die Ausrichtung der einzelnen Figuren im Source-Code von Heider (2010) nicht vorgegeben war, wurden die Aufgaben manuell per Maus ihrem Vorbild nach in die entsprechende Position und Ausrichtung gedreht. War eine bestmögliche Angleichung der neuen Figur mit dem Vorgabebild erreicht, wurden folgende Schritte durchgeführt:

- 2d-Gitterversion erstellt (2D.mG)
- 2d-Gitterversion gespiegelt (wenn diese Ansicht im Itempool enthalten war) (2D.mG.m)
- 3d-Gitterversion erstellt (3D.mG.)
- 3d-Gitterversion gespiegelt (wenn diese Ansicht im Itempool enthalten war) (3D.mG.m.)

In Abbildung 4 sind die drei Varianten der Figurendarstellung (2d-Aufgaben normal und gespiegelt; 2d-Aufgaben normal und gespiegelt mit „Gitternetz“, 3d-Aufgaben normal und gespiegelt mit „Gitternetz“) anhand eines Beispiels veranschaulicht.

Standardisiert wurden die neu entwickelten Bilddarstellungen über ein 900x900 Pixel großes Fenster, anschließend wurden sie in das Bildbearbeitungsprogramm Snag IT 9.13 überführt, wo sie einheitlich auf 270 x 270 Pixel skaliert und als Bitmap in die Testentwicklungssoftware Teach And Test (TNT; Brugger, 1993; zitiert nach Heider, 2010) eingebunden wurden.

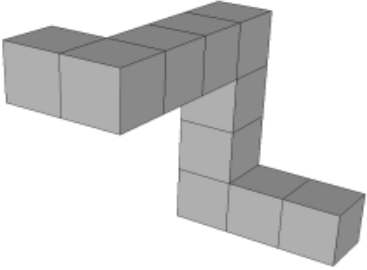
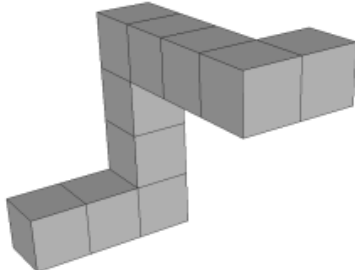
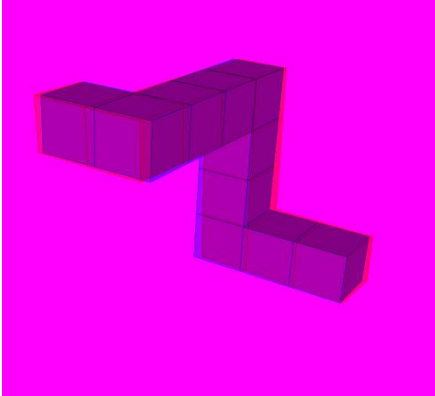
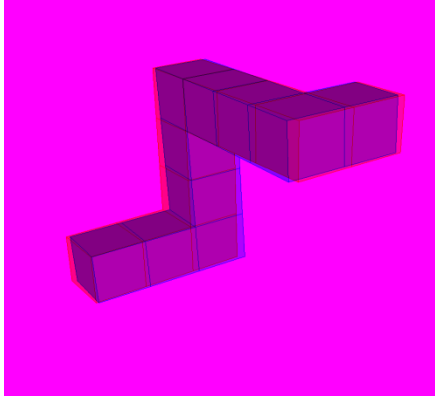
Heider-Version ohne Gitternetz (2D.oG)	Heider-Version ohne Gitternetz, gespiegelt (2D.oG.m)
	
Santos-Version mit Gitternetz (2D.mG)	Santos-Version mit Gitternetz, gespiegelt (2D.mG.m)
	
Santos-Version mit Gitternetz (3D.mG)	Santos-Version mit Gitternetz, gespiegelt (3D.mG.m)
	

Abbildung 4: Beispiele für zweidimensionale Figurendarstellungen in den Varianten „ohne“- und „mit-Gitternetz“ je in der Ausgangsposition und in der gespiegelten Ansicht. Und Beispiele für eine dreidimensionale, „mit-Gitternetz“-Darstellung - ebenfalls in der Ausgangsposition und in der gespiegelten Ansicht.

Somit war eine „mit-Gitternetz“-Itempool-Version vom SPACE 2d/3d erstellt, auch wenn in der vorliegenden Arbeit nur 2d-Aufgaben zum Einsatz kamen. (Anmerkung: Aufgrund von technischen Gegebenheiten ließen sich nicht alle 3d-Ansichten mit dem farbigen Vordergrund herstellen. Im Rahmen der vorliegenden Arbeit wurde darauf aber nicht weiter eingegangen).

Abbildung 5 zeigt eine Beispielaufgabe (ohne Gitternetz): Die Ausgangsfigur in der oberen Bildmitte und die vier Antwortbilder in der unteren Bildhälfte.

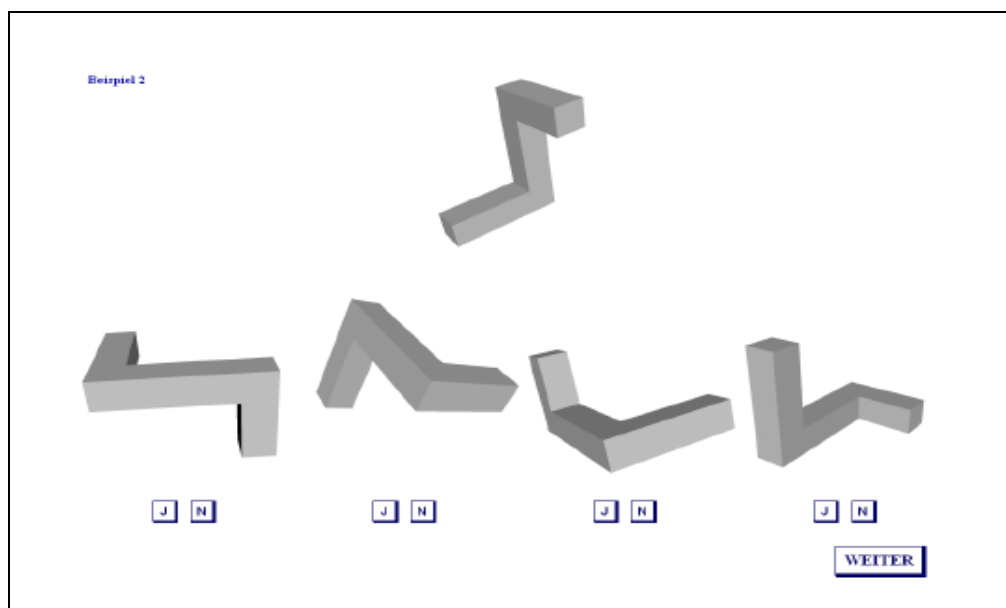


Abbildung 5: Beispielaufgabe (oG-Version).

6. Zielsetzung und Forschungsfragen

Was würde mit der Eindimensionalität des neuen Rasch-homogenen Itempools passieren, wenn man den Abbildungen das Gitternetz der klassischen Aufgaben wieder hinzufügen würde? Würden die von der Testkonstruktion und vom Aufgabendesign her intendierten räumlichen Kognitionen evoziert werden? (Gittler, 1999; vgl. Gittler, 1990,1992).

6.1 Zielsetzung

Ziel dieser Arbeit ist es zu prüfen, ob

- a) die oG-Aufgaben auch in der vorliegenden Stichprobe in sich Rasch-homogen sind, ob dies auch für die Gruppe der mG-Aufgaben gilt, und
- b) ob diese Aufgaben auch gemeinsam eine homogene Skala bilden.

Die Interpretation möglicher nicht Rasch-konformer Ergebnisse soll durch die Information aus den Zusatzfragen unterstützt werden, da Inhomogenität auf Strategiedifferenzen bei der Lösung der Aufgaben zurückgeführt werden kann (Bearbeitungsstrategien als Grundlage inhaltlicher Itemselektion; vgl. Gittler, 1984), wobei dies v.a. für mG-Aufgaben, basierend auf Ergebnissen aus der Literatur (z. B. Peters et al., 1995), angenommen wird.

6.2 Forschungsfragen

Inwieweit evozieren Strukturmerkmale von Aufgaben, wie in diesem Fall die Oberflächengestaltung der Shepard-Metzler-Figuren („Gitternetz“), unterschiedliche Bearbeitungsstrategien, die in der Folge als eine mögliche (Teil-) Erklärung (Hirnstein et al., 2009) für die bestehenden Leistungsunterschiede zwischen Männern und Frauen bei mentalen Rotationsaufgaben herangezogen werden können?

Im folgenden empirischen Teil sollen durch die Untersuchung von 200 Testpersonen mittels des neu zusammengestellten Tests und der anschließenden Zusatzfragen Antworten auf die vorliegenden Forschungsfragen gegeben werden.

EMPIRISCHER TEIL

7. Planung und Durchführung der Studie

Im folgenden Abschnitt werden die Fragestellungen der vorliegenden Studie, die eingesetzten Untersuchungsinstrumente sowie die Durchführung der Testungen erläutert. Des Weiteren erfolgt eine Beschreibung der Stichprobe.

7.1 Forschungsfragen und statistische Hypothesen

Im Rahmen der vorliegenden Studie sollen folgende Fragen beantwortet werden:

Forschungsfragen zu Rasch-Homogenität

- *Sind Aufgaben ohne Gitternetz Rasch-homogen?*

Bleibt der Itempool auch in der vorliegenden Stichprobe Rasch-homogen?

- *Sind Aufgaben mit Gitternetz Rasch-homogen?*

Bleibt der Rasch-homogene Itempool mit verändertem Strukturmerkmal (Gitternetz) Rasch-homogen?

- *Ist der gesamte Test (oG- und mG-Aufgaben) Rasch-homogen, und messen demnach beide Aufgabenmodi dieselbe latente Fähigkeit?*

Forschungshypothese:

H0: „Das Modell gilt“ (Formann, 2004, S. 88). Bei der Überprüfung der Rasch-Homogenität wird von einer Nullhypothese ausgegangen, d.h. Itemparameter der Gruppen unterscheiden sich nicht signifikant voneinander.

Zur Beantwortung dieser Fragen wurden die statistischen Verfahren LRT nach Andersen, Graphische Modelltests und der Martin-Löf-Test verwendet (vgl. Gittler, 1986, 1999; Gittler & Arendasy, 2003)

2. *Gibt es leistungsmäßige Unterschiede zwischen den Testteilen (oG/mG) in Abhängigkeit vom Geschlecht?*

Forschungshypothese:

H1a: Es gibt leistungsmäßige Unterschiede zwischen den Testteilen.

H1b: Es gibt leistungsmäßige Unterschiede zwischen den Testteilen in Abhängigkeit vom Geschlecht.

Zur Beantwortung dieser Fragen wurde eine zweifaktorielle ANOVA (Aufgabenmodi/Geschlecht) gerechnet.

3. Gibt es Schwierigkeitsunterschiede bzw. Leistungsunterschiede bei a) der Bearbeitung zwischen den Link-Items 2,3,4 (oG) und 23,24,25 (mG) und b) in Abhängigkeit vom Geschlecht?

Forschungshypothese:

H1a: Es gibt Schwierigkeitsunterschiede zwischen den Link-Items.

H1b: Es gibt Schwierigkeitsunterschiede zwischen den Link-Items in Abhängigkeit vom Geschlecht.

Zur Beantwortung dieser Fragen wurde eine zweifaktorielle ANOVA (Geschlecht/Aufgabenmodi) mit Messwiederholung gerechnet.

4. Gibt es bearbeitungszeitspezifische Unterschiede zwischen a) den beiden Aufgabenmodi (oG vs. mG) und b) in Abhängigkeit vom Geschlecht?

Forschungshypothese:

H1a: Es gibt bearbeitungszeitspezifische Unterschiede zwischen den Modi.

H1b: Es gibt bearbeitungszeitspezifische Unterschiede zwischen den Modi in Abhängigkeit vom Geschlecht.

Zur Beantwortung dieser Fragen wurde eine zweifaktorielle ANOVA (Geschlecht/Aufgabenmodi) gerechnet.

5. Kann das externe Kriterium Selbsteinschätzung der „Orientierungsfähigkeit im Raum“ die Leistung der Probanden im Test vorhersagen?

Forschungshypothese:

H1: Es gibt einen Zusammenhang zwischen der Selbsteinschätzung der „Orientierungsfähigkeit im Raum“ und der Testleistung.

Zur Beantwortung dieser Frage wurde eine Produkt-Moment-Korrelation berechnet.

6. *Unterscheiden sich die subjektiven Ratings bezüglich a) der beiden Aufgabenmodalitäten und b) in Abhängigkeit vom Geschlecht?*

Forschungshypothese:

H1a: Es gibt Unterschiede im subjektiven Rating der beiden Aufgabenmodi.

H1b: Es gibt Unterschiede im subjektiven Rating der beiden Aufgaben in Abhängigkeit vom Geschlecht.

Zur Beantwortung dieser Fragen wurde eine zweifaktorielle ANOVA (Geschlecht/Aufgabenmodi) mit Messwiederholung gerechnet.

Für die Berechnungen wird allgemein ein globales Signifikanzniveau $\alpha = .05$ festgesetzt. Für die RM-Modellprüfungen gilt ein Signifikanzniveau $\alpha = .01$.

7.2 Untersuchungsteilnehmer und Stichprobenziehung

Ausgehend von der eingeschränkten Generalisierbarkeit der Ergebnisse bei Heider (2010) (es wurden fast ausschließlich Studenten getestet), sowie in Anlehnung an die Ansicht von Kubinger, Rop, Knoll und Wurst (1983; zitiert nach Gittler, 1986), dass ein Intelligenz(sub)test sowohl aus praktischen als auch theoretischen Gründen für einen größeren Altersbereich eindimensional messen sollte, wurde in der vorliegenden Arbeit ein Altersbereich vom 18. bis 65. Lebensjahr gewählt. Die untere Altersgrenze soll dazu dienen, mögliche diskutierte, geschlechtsspezifische hormonelle Einflüsse aus der Pubertät als abgeschlossen annehmen zu können. Die Obergrenze basierte primär auf der Überlegung, dass die Erfahrung im Umgang mit PCs gegeben sein soll, um diese nicht als Störvariable mit zu erheben. Die angestrebte Stichprobengröße betrug $n = 200$ Testpersonen, um sinnvolle Rasch-Modell-Analysen durchführen zu können.

Die Stichprobe setzte sich aus Familie, Verwandten, Bekannten und Freunden zusammen, die mehrheitlich in einem Einzeltestungssetting untersucht wurden. Bei kleineren Gruppentestungen wurde darauf geachtet, dass die Gestaltung des Settings jener einer Einzeltestung nahe kam, z.B. indem die Testpersonen so gesetzt wurden, dass sie sich nicht gegenseitig ansehen konnten, um keine Konkurrenzsituation zu provozieren etc.

7.3 Stichprobenbeschreibung

Die vorliegende Arbeit beruht auf den Daten von insgesamt 199 Testpersonen (von 200 getesteten Personen konnte ein Datensatz nicht verwendet werden), wodurch sich ein Geschlechterverhältnis von 119 Frauen (59.8%) zu 80 Männer (40.2%) ergibt.

Die Altersverteilung erstreckt sich vom 18.-65. Lebensjahr (Männer: *Median* = 29.50, *SD* = 11.68; Frauen: *Median* = 28, *SD* = 12.96), wie Abbildung 6 verdeutlicht.

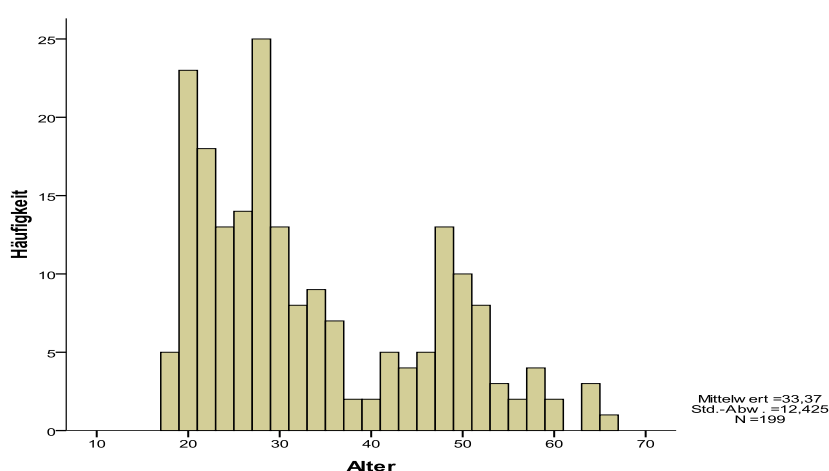


Abbildung 6: Altersverteilung der Gesamtstichprobe.

Bezüglich der Muttersprache gaben 189 (95%) an, Deutsch als Muttersprache zu haben. Des Weiteren gaben 174 (87.5%) der Testpersonen an, Rechtshänder, 17 (8.5%) Linkshänder und 8 (4.0 %) Beidhänder zu sein. Die Bildungsverteilung der Gesamtstichprobe stellte sich wie folgt dar: Neunte Schulstufe (Pflichtschule) gaben 6 Testpersonen an (3%), Berufsbildende Mittlere Schule/Fachhochschule/Lehre wurde von 11 Personen angegeben (5.5%), Matura (AHS, BHS)/Abitur gaben 95 Personen an (47.7%), Fachhochschule/Akademie gaben 28 Personen (14%) und Universität 59 Personen (29.5%) der Teilnehmer an, wie in Abbildung 7 veranschaulicht.

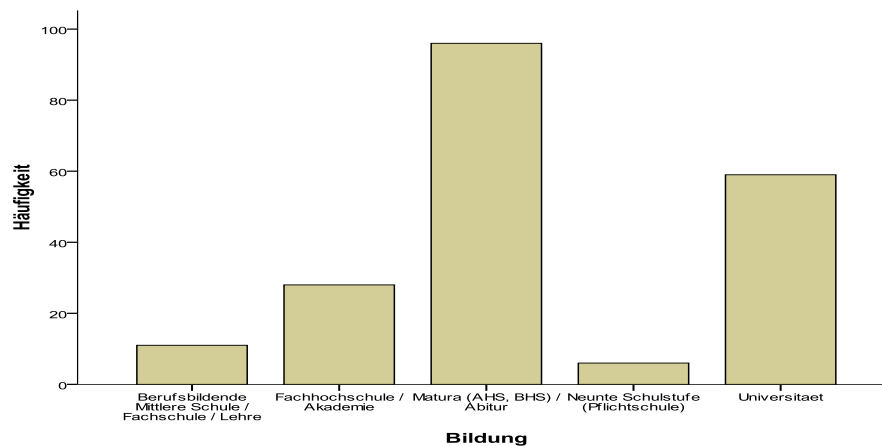


Abbildung 7: Häufigkeiten der Bildungsstufen.

7.4 Durchführung

7.4.1 Space 2d-2 Modi: „ohne-Gitter“- und „mit-Gitter“-Aufgaben

Nachdem der Itempool bei Heider (2010) als mG-Version erstellt war (wie unter Kapitel 5.2 beschrieben ist) und die Aufgaben auf die im Rahmen dieser Diplomarbeit benötigte Weise zu einen Test zusammengefügt und programmiert worden waren, erfolgte die Datenerhebung im Zeitraum von Dezember 2011 bis April 2012.

Jeder Teilnehmer wurde zu Beginn der Testung gebeten, Angaben zu seinen soziodemographischen Daten zu machen (Alter, Geschlecht, Bildung, Händigkeit, Muttersprache, Beruf) und sein Einverständnis zur Testung und zur anonymisierten Nutzung seiner Angaben für wissenschaftliche Zwecke zu geben.

Planlesefähigkeit

Nach Erhebung der demographischen Daten durch die Versuchspersonen wurde eine Frage zur Selbsteinschätzung der eigenen „Orientierungsfähigkeit im Raum“ präsentiert (siehe unten). Anhand eines Schiebereglers konnten sich die Probanden auf einer Analogskala selber positionieren. Diese Frage soll als ein mögliches externes Kriterium zur Vorhersage der Testleistungen überprüft werden.

„Stellen sie sich vor, Sie sind in einer fremden Stadt und sie wollen einen ganz bestimmten Platz finden. Um diesen zu finden benutzen Sie einen Stadtplan. Schätzen Sie bitte ein, wie gut Sie einen Weg mittels Stadtplan bzw. Landkarte finden können. “

Danach folgte auf einer neuen Bildschirmseite folgende Instruktion:

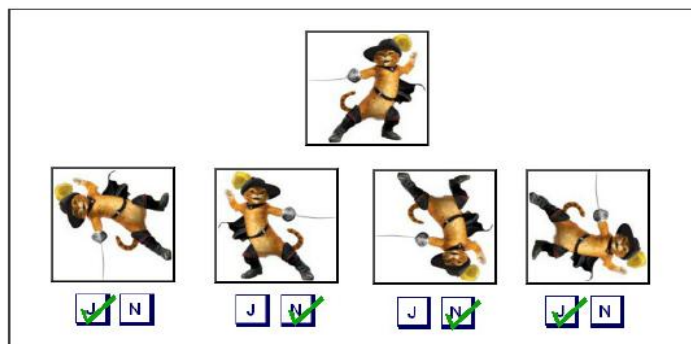
In den nun folgenden Aufgaben geht es darum, eine geometrische Ausgangsfigur mit vier dargebotenen Antwortfiguren zu vergleichen.

Diese vier Antwortfiguren zeigen die Ausgangsfigur aus einer anderen Perspektive oder aber ein Spiegelbild der Ausgangsfigur ebenfalls aus einer anderen Perspektive.

Ihre Aufgabe ist es, für jede der vier Alternativen zu entscheiden, ob es sich um die Ausgangsfigur handelt oder um ein Spiegelbild.

Anschließend erschien folgender Text:

Im unten dargestellten Fenster sehen Sie in der ersten Zeile die Ausgangsfigur, darunter befinden sich vier Antwortfiguren. Ihre Aufgabe besteht darin, für jede der vier Antwortfiguren zu entscheiden, ob sie der Ausgangsfigur gleicht oder nicht. Im Beispiel entsprechen die Antworten eins und vier der Ausgangsfigur, deshalb ist „J“ für JA markiert. Die Antworten zwei und drei entsprechen nicht der Ausgangsfigur (Spiegelbilder), deshalb ist „N“ für NEIN markiert.



Auf einer weiteren Bildschirmseite folgten weitere Erklärungen: Anhand von zwei aktiven Übungsbeispielen konnten die Testpersonen überprüfen, ob sie die Aufgabenstellung verstanden hatten. Dazu wurde nach der Bearbeitung einer Übungsaufgabe gleich eine Rückmeldung hinsichtlich der Richtigkeit ihrer Eingaben geboten. Löste eine Versuchsperson zweimal hintereinander ein Übungsbeispiel nicht, erschien wieder ein Fenster mit der Lösung für dieses Aufgabenbeispiel und einer Erklärung dazu.

Nach der erfolgreichen Bearbeitung der beiden Übungsbeispiele erschien eine weitere Information:

*Bitte vergewissern Sie sich, ob das von Ihnen **gewünschte Antwortkästchen** mit einem „✓“ **markiert** ist. Haben Sie dabei gegebenenfalls etwas **Geduld**, damit es nicht zu unerwünschten **Fehlmarkierungen** (falschen Antworten) kommt.*

Insgesamt sind es 25 Aufgaben. Als Lösungen können alle Antwortkombinationen in Frage kommen (z.B. 4-mal Ja, 4-mal Nein oder alle anderen Kombinationen).

Im Test wird Ihnen nicht mehr rückgemeldet, ob Sie eine Aufgabe richtig oder falsch bearbeitet haben.

7.4.2 Zusatzfragen (Aussagenpaare)

Im Anschluss an die PC-Testung wurden Zusatzfragen (Aussagenpaare) als Paper-Pencil-Version vorgelegt. Sinn und Ziel dieser Zusatzfragen war es, Informationen über Bearbeitungsstrategien (holistisch vs. analytisch) (Aussagenpaare 1, 2 und 13) und mögliche qualitativ unterschiedliche Empfindungen zwischen den beiden Aufgabenmodi (oG- vs. mG-Aufgaben) (Aussagenpaar 3, 4, 5 und 6) zu erheben.

Diese Zusatzfragen sollten einerseits zu einer besseren inhaltlichen Interpretierbarkeit der Ergebnisse beitragen. Andererseits könnten durch sie Hypothesen für zukünftige Untersuchungen hinsichtlich dessen generiert werden, was letztlich erhoben und gemessen wird (siehe „inhaltliche Selektionsstrategien“; Gittler, 1986), da die

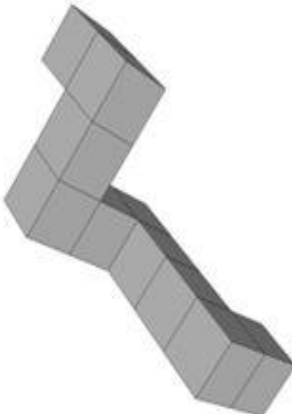
Gültigkeit des Rasch-Modells keine Aussagen zur Validität des Test machen kann (Gittler, 1999).

Die Beleuchtung eines Sachverhaltes über verschiedene Zugänge erhöht die Validität der erhaltenen Ergebnisse, da jede Untersuchungsmethode für sich allein limitiert ist (Glück, 2001), wie im Kapitel 4.5 schon genauer erläutert wurde. In der vorliegenden Arbeit werden statistisch Analysen durch Rasch-Modell-Analysen, die sich auf Antwortmuster konzentrieren, mit einem beschreibenden Ansatz (Zusatzfragen) kombiniert, bei dem verbale Informationen der Testpersonen bezüglich ihrer angewendeten Strategien erhoben werden (vgl. Quaiser-Pohl, Rohe & Amberger, 2010).

Die Aussagen sollten auf einer vierstufigen Likert-Skala mit (4) *trifft sehr zu*, (3) *trifft eher zu*, (2) *trifft eher nicht zu* bis (1) *trifft gar nicht zu* bewertet werden; je kleiner der Wert, desto geringer die Zustimmung zu der Aussage.

Als Beispiel sei das erste Aussagenpaar mit der Instruktion in Abbildung 8 angeführt. Es wurden immer dieselben zwei Bilder als stellvertretende Beispiele für den jeweiligen Aufgabenmodus nebeneinander präsentiert (Figur ohne Gitter vs. Figur mit Gitter). Unter denen befanden sich die jeweiligen Aussagen und Felder zu deren Bewertung. Im Anhang werden alle Aussagenbeispiele im Originalformat, nachfolgend jedoch aus Platzgründen nur als Satzpaare angeführt.

In dem von Ihnen soeben abgeschlossenen Testverfahren wurden Ihnen zwei verschiedene Würfeltypen präsentiert. Der Unterschied lag in der Darstellung der Würfel: Es waren Würfel ohne Gitternetz und Würfel mit Gitternetz abgebildet. Sie sollen nun noch ein paar Aussagen zu diesen verschiedenen Würfeltypen beantworten. Bitte geben Sie bei den folgenden Aussagen jeweils an, wie sehr diese für Sie zutreffen!

Würfel <u>ohne</u> Gitternetz	Würfel <u>mit</u> Gitternetz
	

1. Würfel ohne Gitternetz habe ich gedanklich als Ganzes gedreht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

2. Würfel mit Gitternetz habe ich gedanklich als Ganzes gedreht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

Abbildung 8: Instruktion und erstes Fragenpaar der Zusatzfragen

1. Würfel ohne Gitternetz habe ich gedanklich als **Ganzes** gedreht.
2. Würfel mit Gitternetz habe ich gedanklich als **Ganzes** gedreht.

3. Bei Würfel ohne Gitternetz habe ich mich an bestimmten **Teilaspekten** der Würfel Orientiert.
4. Bei Würfel mit Gitternetz habe ich mich an bestimmten **Teilaspekten** der Würfel orientiert.
5. Bei Würfel ohne Gitternetz hatte ich einen starken Eindruck von **räumlicher Tiefe**.
6. Bei Würfel mit Gitternetz hatte ich einen starken Eindruck von **räumlicher Tiefe**.
7. Bei Würfeln ohne Gitternetz fiel mir das Erfassen der **Gesamtfigur** leicht.
8. Bei Würfeln mit Gitternetz fiel mir das Erfassen der **Gesamtfigur** leicht.
9. Bei Würfeln ohne Gitternetz fiel mir insgesamt gesehen **das Bearbeiten** der Aufgaben leicht.
10. Bei Würfeln mit Gitternetz fiel mir insgesamt gesehen **das Bearbeiten** der Aufgaben leicht.
11. Würfel ohne Gitternetz empfand ich **optisch** ansprechend.
12. Würfel mit Gitternetz empfand ich **optisch** ansprechend.
13. Bei Würfeln mit Gitternetz habe ich die einzelnen **Kästchen** abgezählt.

7.5 Versuchsplan

Die Abfolge der vorgelegten Aufgaben ist in Tabelle 1 dargestellt.

Tabelle 1: Darstellung des Versuchsplans für die vorliegende Untersuchung und der Bezug zum Versuchsplan von Heider (2010).

Heider	Heider Modi		Santos – Abfolge	Santos		
Item Nr.	Vers. A	Vers. B	Item Nr.	Aufgaben (2d)	n von Heider	n von Santos (2d)
---	---	---	1 Warm up	oG	----	---
0 Warm up	<u>2d</u>	<u>2d</u>	2	LoG	<u>227</u>	199
3	<u>2d</u>	<u>2d</u>	3	LoG	<u>227</u>	199
4	<u>2d</u>	<u>2d</u>	4	LoG	<u>227</u>	199
5	<u>2d</u>	3d	5	LoG	<u>116</u> / <u>111</u>	199
6	<u>2d</u>	3d	6	LoG	<u>116</u> / <u>111</u>	199
10	3d	<u>2d</u>	7	LoG	<u>116</u> / <u>111</u>	199
11	3d	<u>2d</u>	8	LoG	<u>116</u> / <u>111</u>	199
7	<u>2d</u>	3d	9	mG	<u>116</u> / <u>111</u>	199
8	<u>2d</u>	3d	10	mG	<u>116</u> / <u>111</u>	199
9	<u>2d</u>	3d	11	mG	<u>116</u> / <u>111</u>	199
12	3d	<u>2d</u>	12	mG	<u>116</u> / <u>111</u>	199
13	3d	<u>2d</u>	13	mG	<u>116</u> / <u>111</u>	199
14	3d	<u>2d</u>	14	mG	<u>116</u> / <u>111</u>	199
1	3d	3d	15	mG	227	199
2	3d	3d	16	mG	227	199
18	3d	3d	17	mG	227	199
19	3d	3d	18	mG	227	199
20	3d	3d	19	mG	227	199
15	<u>2d</u>	<u>2d</u>	20	LmG	<u>227</u>	199
16	<u>2d</u>	<u>2d</u>	21	LmG	<u>227</u>	199
17	<u>2d</u>	<u>2d</u>	22	LmG	<u>227</u>	199
0a	<u>2d</u>	<u>2d</u>	23	LmG	<u>227</u>	199
3a	<u>2d</u>	<u>2d</u>	24	LmG	<u>227</u>	199
4a	<u>2d</u>	<u>2d</u>	25	LmG	<u>227</u>	199

Anmerkung: n = Fallzahl, mG = „mit-Gitter“-Aufgaben; oG = „ohne-Gitter“-Aufgaben; LmG / LoG = Link-Items

Allen Testpersonen wurde der idente Test am PC vorgegeben. Dieser bestand aus insgesamt 25 Items zweidimensionaler Darstellungen dreidimensionaler geometrischer Figuren, wobei die ersten acht Items oG-Aufgaben waren, analog zu den Aufgaben bei Heider (2010). Da das erste Item als Warm-up diente, wurde es nicht in die Analysen miteinbezogen; somit ergaben sich letztlich 7 oG-Aufgaben und 17 mG-Aufgaben

(siehe Tabelle 1). Anhand der beiden Aufgabenmodi sollte eine möglicherweise unterschiedliche Bearbeitungsstrategie in Abhängigkeit vom (Nicht-)Vorhandensein des Gitternetzes genauer erfasst werden.

Die Items 2,3,4 (LoG) und 22,23,24 (LmG) sind Link-Items (mit unterschiedlichem Lösungsstring, um möglichen Erinnerungseffekten entgegen zu wirken), d.h. diese Aufgaben kamen sowohl als oG-, als auch mG-Version vor. Dadurch sollte eine direktere Gegenüberstellung von oG- und mG-Aufgaben ermöglicht werden (die restlichen Items mit „L“ stellen Link-Items zum Itempool bei Heider (2010) dar, die in der vorliegenden Arbeit aber nicht aufgegriffen wurden).

Außerdem wurde in der Instruktion nichts von „mentaler Rotation“ als Bearbeitungsstrategie erwähnt, um die Testpersonen unbeeinflusst zu lassen hinsichtlich einer bestimmten Lösungsstrategie. Auch wurde keine Information über unterschiedliche Aufgabenmodi gegeben, um besser auswerten zu können, welchen Einfluss das Strukturmerkmal „Gitternetz“ tatsächlich auf die Strategiewahl hat.

Im Gegensatz zu den klassischen Erhebungsverfahren der MR-Fähigkeit wurde dieser Test als reiner Powertest vorgegeben, es gab demnach keine Zeitbegrenzung, weder für jede einzelne Aufgabe noch für den Gesamttest. Der Grund dafür liegt darin, dass ein Test mit Zeitbegrenzung mit einiger Wahrscheinlichkeit die Homogenitätsvoraussetzungen des Rasch-Modells (Rasch, 1960/1980; zitiert nach Gittler & Arendasy, 2003) verletzt, da es zu einer Vermengung von Speed- und Powerkomponente kommt (Gittler & Arendasy, 2003), wodurch die Ergebnisse durch die Zeitbegrenzung beeinflusst werden können (Gittler & Glück, 1998). Außerdem benötigen die möglicherweise zur Anwendung kommenden analytischen Strategien mehr Zeit.

Bezüglich des Verrechnungsmodus der Testleistung wird auf Ergebnisse aus der Studie von Peters und Battista (2008) eingegangen. Diese hatte festgestellt, dass bei paarweiser Bearbeitung der Aufgaben wie bei Shepard und Metzler (1971) kein oder nur ein sehr schwacher geschlechtsspezifischer Unterschied zu finden war, während bei

Untersuchungen mit dem Vandenberg-und-Kuse-Test (Vandenberg & Kuse, 1978) signifikante Unterschiede auftraten. Sie leiteten daraus ab, dass sich die Anzahl von vorgegebenen Vergleichsfiguren erschwerend auf die Aufgabenbearbeitung auswirkte, da unter anderem das Kurzzeitgedächtnis bei Mehrfachvergleichen stärker beansprucht würde.

Heider (2010) entschied sich für eine Vorgabeform, bei der die Probanden eine Ausgangsfigur hatten, die sie mit vier Antwortfiguren vergleichen mussten, wobei zwischen 0 und 4 Antworten richtig sein konnten (vgl. Glück & Fabrizii, 2010; Hirnstein et al., 2009). Dadurch waren die Testpersonen angehalten, alle 4 Antwortalternativen zu beantworten, wodurch eine sprunghafte Bearbeitungsstrategie (Hirnstein et al., 2009) nicht möglich wurde.

Einen Punkt erhielt man, wenn man alle 4 Antworten richtig hatte, wobei alle Antwortkombinationen (1 Ja, 3 Nein; 2 Ja, 2 Nein etc.) möglich waren. Es konnten somit 24 Punkte insgesamt erreicht werden. Personen, die keine Aufgabe ($RW = 0$) oder alle Aufgaben ($RW = 24$) lösten, stellen keine statistische Information zur Parameterschätzung zur Verfügung und gehen damit nicht in die Rasch-Analysen ein. Anhand der restlichen „informativen“ Versuchspersonen können die Parameter für die Modellkontrollen geschätzt werden (Gittler, 1999).

8. Ergebnisse der statistischen Auswertung

8.1 Rasch-Analysen

Da die zentrale Frage im Rahmen dieser Arbeit die Dimensionalitätsbestimmung des vorgelegten Testmaterials, v.a. der mG-Aufgaben war, sollte dies anhand des dichotom-logistischen Modells von Rasch (1960/1980; zitiert nach Gittler & Arendasy, 2003) genauer untersucht werden.

Zuerst wurde eine Rasch-Analyse über den gesamten Test gerechnet, also über beide Aufgabenmodi zusammen. Die Gültigkeit des Rasch-Modells wurde mit dem Likelihood-Ratio-Test (LRT) von Andersen (1973; zitiert nach Gittler & Arendasy, 2003, S. 165) überprüft. Dazu wurde die Stichprobe anhand von Teilungskriterien [mehrere externe Kriterien (z. B: Geschlecht, Alter etc.), ein internes Kriterium (Median: hoher vs. niedriger Rohscore)] in zwei Gruppen geteilt (vgl. Gittler, 1999; Formann, 2004) und jeweils deren Itemparameterschätzungen auf statistische Gleichheit geprüft (Andersen, 1973; zitiert nach Gittler & Arendasy, 2003, S. 165) bzw. die Gleichheit der Itemparameter in den Personenuntergruppen geprüft (Gittler, 1999). Die durch den Modelltest erhaltenen empirischen χ^2 -Werte wurden mit den kritischen χ^2 -Werten verglichen. Ist der kritische χ^2 -Wert größer als der empirische χ^2 -Wert bei einer Irrtumswahrscheinlichkeit $\alpha = .01$, dann ist das Ergebnis nicht signifikant und die Gültigkeit des Rasch-Modells kann angenommen werden.

Die Berechnungen wurden mit dem Programmpaket LpcM-WiN (Fischer & Ponocny-Seliger, 1998) durchgeführt. Die Modelltests für die klassischen Teilungskriterien sind in Tabelle 2 angeführt und zeigen jeweils ein nicht signifikantes Ergebnis.

8.1.1 Andersen Likelihood Quotiententest

Tabelle 2: Klassische Teilungskriterien.

Teilungskriterium		$\chi^2_{\text{empirisch}}(df = 23)$ $\chi^2_{\text{krit.}} = 41.66 (\alpha_{.01})$	p	Sig. ($\alpha_{.01}$)
Median	intern	27.00	.256	n. s.
Geschlecht	extern	28.02	.215	n. s.
Alter	extern	35.51	.046	n. s.
Orientierungsfähigkeit	extern	25.50	.325	n. s.
Bearbeitungszeit	extern	26.17	.293	n. s.
Bildung	extern	35.05	.051	n. s.

Anmerkung: p = Wahrscheinlichkeit, χ^2 = Andersen Chi-Quadrat(Freiheitsgrade)

Als ein Beispiel für die graphischen Modelltests ist der graphische Modelltest nach dem internem Teilungskriterium Median (hoher vs. niedriger Testwert) angeführt (siehe Abbildung 9), der als trennschärfstes Kriterium (Gittler, 1984) gilt. (Alle weiteren graphischen Modelltests sind im Anhang angeführt).

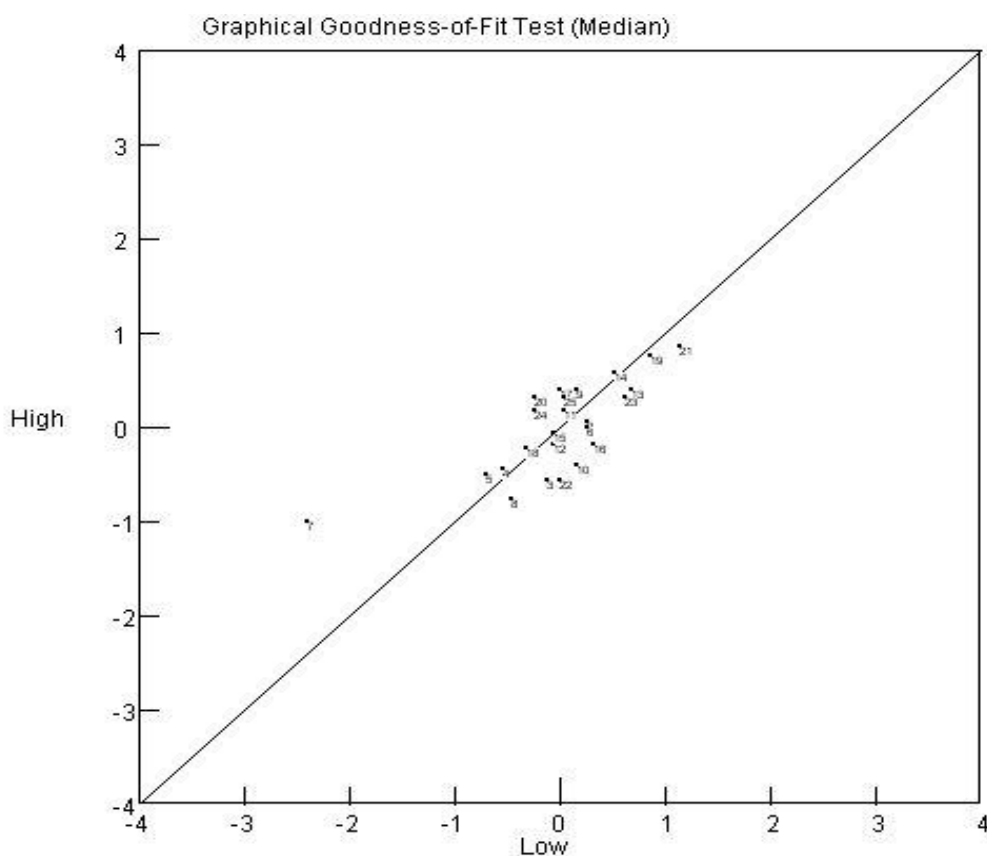


Abbildung 9: Graphischer Modelltest (Teilungskriterium: Median).

Um sicher zu gehen, dass beide Aufgabenmodi ein und dieselbe latente Dimension erfassen, wurde anschließend ein Martin-Löf-Test gerechnet.

8.1.2 Martin-Löf-Test

Da Rasch-Homogenität für den Gesamttest gegeben war und eine Hypothese hinsichtlich möglicher systematischer Unterschiede zwischen den beiden Aufgabenmodi (oG vs. mG) vorlag (Gittler, 1999), wurde zur Überprüfung der Eindimensionalität der Items ein Martin-Löf-Test berechnet, um der Frage, ob beide Aufgabenmodi tatsächlich dieselben latenten Dimensionen erfassen, genauer zu prüfen.

Tabelle 3: Modellgeltungskontrollen (LQT) für das RM nach Martin-Löf (n = 191).

	Log-Likelihood	Martin-Löf-Statistic (df = 118) $\chi^2_{\text{krit.}} = 156.66$	p	Sig. ($\alpha_{.01}$)
oG-Aufgaben	-423.077	90.56	.971	n.s.
mG-Aufgaben	-1290.801			

Anmerkung: χ^2 = Martin-Löf-Statistik (Freiheitsgrade), p = Wahrscheinlichkeit

Die Berechnung der entsprechenden Prüfgröße im Martin-Löf-Test ($\chi^2_{(118)} = 90.56$) fällt unter Berücksichtigung des kritischen χ^2 -Werts von 156.66 nicht signifikant aus (siehe Tabelle 3). Somit kann die Nullhypothese, dass beide Testteile zusammen dieselbe latente Fähigkeitsdimension erfassen, beibehalten werden.

8.1.3 Einzelitems als Teilungskriterium

Da das vorliegende Aufgabenmaterial zwei verschiedene Modi besitzt und aus inhaltlichen Überlegungen aus zwei separaten Rasch-konformen Skalen bestehen könnte, sollen zur Absicherung der Rasch-Modell-Gültigkeit für diesen Itempool noch unter Berücksichtigung der Kritik von Stelzl (1979) und der dazu angestellten Überlegungen von Formann (1981) Rasch-Analysen mit jedem Item als Teilungskriterium ergänzend berechnet werden. Die Ergebnisse sind in Tabelle 4 angeführt.

Tabelle 4: LRT nach Andersen für alle Items als Teilungskriterium.

Item (Test)	Item (eval)	$\chi^2_{\text{empirisch}}(df = 22)$ $\chi^2_{\text{krit.}} = 40.31(\alpha_{.01})$	p	Sig. ($\alpha_{.01}$)
Item1	eval0 (Warm-up)	--	--	--
Item2	eval1	27.41	.195	n. s.
Item3	eval2	23.15	.392	n. s.
Item4	eval3	14.40	.877	n. s.
Item5	Eval4	41.97	.006	n. s.
Item6	eval5	20.97	.521	n. s.
Item7	eval6	22.03	.457	n. s.
Item8	eval7	20.26	.565	n. s.
Item9	eval8	20.37	.557	n. s.
Item10	eval9	21.94	.462	n. s.
Item11	eval10	14.91	.861	n. s.
Item12	eval11	35.54	.034	n. s.
Item13	eval12	25.18	.287	n. s.
Item14	eval13	34.67	.042	n. s.
Item15	eval14	32.54	.067	n. s.
Item16	eval15	18.20	.691	n. s.
Item17	eval16	26.09	.247	n. s.
Item18	eval17	39.61	.012	n. s.
Item19	eval18	27.57	.189	n. s.
Item20	eval19	23.91	.351	n. s.
Item21	eval20	32.37	.070	n. s.
Item22	eval21	28.23	.167	n. s.
Item23	eval22	28.81	.149	n. s.
Item24	eval23	24.40	.326	n. s.
Item25	eval24	35.36	.035	n. s.

Anmerkung: p = Wahrscheinlichkeit, χ^2 = Martin-Löf-Statistik (Freiheitsgrade)

Das Rasch-Modell zeigt für alle Modelltests (auch für den Modelltest von Item5 nach einer Bonferroni-Holm-Korrektur) keine signifikanten Ergebnisse. Die Bonferroni-Holm-Korrektur dient dazu, dem Problem der Alpha-Kumulierung im Rahmen multipler Testungen an einem Datensatz entgegenzuwirken. Dabei wird der p-Wert der Einzeltestung mit dem lokalen α -Niveau (Verhältnis zwischen globalem α -Niveau und Anzahl der Tests) verglichen. Ist der p-Wert (hier .006) höher als das lokale α -Niveau (hier $0.01/24 = .00042$), kann die Null-Hypothese beibehalten werden (Holm, 1979).

8.1.4. Personenparameter (Gesamttest) und Gruppenunterschiede

Anhand der Tabelle 5 wird der Zusammenhang zwischen den ordinalskalierten Testrohwerten (1-23, da ein Testwert von 0 und 24 nicht in die Analyse fallen) und den intervallskalierten Personenparametern (Theta θ), die aus den Gesamtdaten ($n = 191$) geschätzt wurden, gezeigt. Die Abbildung 10 veranschaulicht diesen Zusammenhang noch einmal graphisch.

Tabelle 5: Anzahl der gelösten Items mit ihren entsprechenden Personenparametern (Theta θ).

Testwert	Personenparameter (Theta θ)	Häufigkeit des Antwortpattern
0	---	1
1	-3.2425	5
2	-2.4948	14
3	-2.0329	10
4	-1.6867	11
5	-1.4027	7
6	-1.1568	7
7	-0.9360	9
8	-0.7320	5
9	-0.5409	8
10	-0.3572	6
11	-0.1785	11
12	-0.0021	6
13	0.1741	7
14	0.3533	8
15	0.5374	10
16	0.7295	4
17	0.9337	9
18	1.1553	11
19	1.4023	6
20	1.6877	20
21	2.0356	2
22	2.4996	8
23	3.2498	7
24	----	7
Gesamt		191

Wie man Tabelle 5 entnehmen kann, haben ein Teilnehmer keine einzige Aufgabe und sieben Teilnehmer haben alle Aufgaben gelöst; somit wurden die Modelltests auf Grundlage von $n = 191$ Fällen berechnet. Je größer der Personenparameter ist, desto fähiger ist die Person.

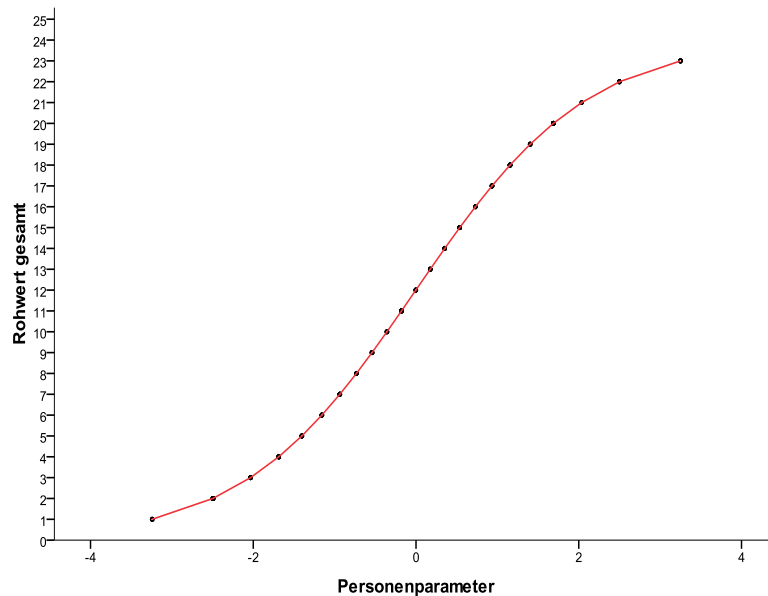


Abbildung 10: Zusammenhang zwischen Testwerten und Personenparametern im Gesamttest

Im Durchschnitt lösten die Männer im Gesamttest 16.10, und die Frauen 9.69 von 24 Aufgaben. Bei den oG-Aufgaben lösten die Männer im Durchschnitt 4, und Frauen 2.42 Aufgaben von 7. Bei den mG-Aufgaben ergab die Verteilung durchschnittlich 12.8 gelöste Aufgaben bei den Männern und 7.27 bei den Frauen von 17 möglichen. Somit erreichten Männer im Durchschnitt in allen Bedingungen beinahe doppelt so viele Aufgaben wie die Frauen (siehe Tabelle 6).

Tabelle 6: Deskriptivstatistik der Testwerte für die Gesamtstichprobe ($n = 199$) im Gesamttest und getrennt für oG- und mG-Aufgaben in Abhängigkeit vom Geschlecht.

Geschlecht		Testwert Gesamttest	Testwert oG	Testwert mG
m ($n = 80$)	<i>M</i>	16.10	4.0	12.8
	<i>Median</i>	18	4.5	13
	<i>SD</i>	6.30	2.13	4.47
	Min	2	0	1
	Max	24	7	17
w ($n = 119$)	<i>M</i>	9.69	2.42	7.27
	<i>Median</i>	9	2	7
	<i>SD</i>	6.35	1.98	4.70
	Min	0	0	0
	Max	24	7	17

Anmerkung: n = Fallzahl, M = Mittelwert, SD = Standardabweichung, Min = Minimum, Max = Maximum

In Tabelle 7 sind die Deskriptivstatistiken für die Personenparameter dargestellt. Da Testpersonen, die keine oder alle Aufgaben lösten, nicht in die Berechnungen einfließen, ergibt sich ein $n = 191$. Je fähiger die Teilnehmer, desto größer war der Personenparameter: Die Gruppe der männlichen Teilnehmer erreichte im Gesamttest durchschnittlich einen Personenparameter von 0.80, die weiblichen Teilnehmer einen Personenparameter von -0.55.

Tabelle 7: Deskriptivstatistik der Personenparameter (θ) im Gesamttest und getrennt für oG- und mG-Aufgaben in Abhängigkeit vom Geschlecht.

Geschlecht			θ Gesamttest	θ_{oG}	θ_{mG}
m (n = 74)	N	Gültig	74	63	69
		Fehlend	6	17	11
	<i>M</i>		.80	.27	.89
	Median		.93	.30	1.22
	<i>SD</i>		1.45	1.14	1.42
	Min		-2.49	-1.86	-2.84
	Max		3.25	1.86	2.83
w (n = 117)	N	Gültig	117	96	110
		Fehlend	2	23	9
	<i>M</i>		-.55	-.51	-.36
	Median		-.54	-.96	-.37
	<i>SD</i>		1.45	1.22	1.34
	Min		-3.24	-1.86	-2.84
	Max		3.25	1.86	2.83

Anmerkung: n = Fallzahl, *M* = Mittelwert, *SD* = Standardabweichung, Min = Minimum, Max = Maximum

Aufgrund des Stichprobenumfanges > 30 , können auf Basis des Zentralen Grenzwertsatzes in der Folge parametrische Verfahren zur Hypothesenprüfung herangezogen werden (Bortz & Döring, 2006).

Mit den durch die Rasch-Modell-Analysen erhaltenen intervallskalierten Personenparametern (Tabelle 7) wird ein t-Tests für unabhängige Stichproben berechnet, der prüft, ob die Personenparameter zwischen Männern und Frauen im Gesamttests einen signifikanten Unterschied aufweisen. Die Berechnung der entsprechenden Prüfgröße (Varianzhomogenität kann mit $p = .870$ angenommen werden) ergibt mit $t(189) = 6.28$, $p < .001$ ($d = 0.89$) einen signifikanten Mittelwertsunterschied hinsichtlich der Leistung in Abhängigkeit vom Geschlecht. Männer erreichten durchschnittlich den höheren Wert.

Um auch den konfundierenden Einfluss der beiden unabhängigen Variablen Lebensalter (metrisch) und Bildung (Nicht-Akademiker/Akademiker) berücksichtigen zu können, wird eine einfaktorielle, univariate Kovarianzanalyse für den Faktor Geschlecht berechnet. Die Homogenität der Varianzen kann mit $p = .898$ angenommen werden.

Bezüglich der Kovariate Lebensalter kann mit $F(1, 187) = 16.92, p < .001 (\eta^2 = .083)$ ein konfundierender Einfluss auf die abhängige Variable Leistung (Personenparameterniveau) angenommen werden, während für die Bildung mit $F(1, 187) = 0.70, p = .402$ kein signifikanter Einfluss auf die Leistung zu beobachten ist.

Des Weiteren ergibt sich ein negativer Korrelationskoeffizient zwischen Lebensalter und Leistung mit $r = .262 (b = -.036), p < .001$ - mit zunehmendem Lebensalter ist eine Abnahme der Leistung zu erkennen. Unter Berücksichtigung dieser Kovariaten fällt die Prüfgröße für das Geschlecht mit $F(1, 187) = 43.19, p < .001 (\eta^2 = .188)$ nach wie vor signifikant aus: Männer erreichen bessere Leistungen).

8.2 Ergebnisse der klassischen Testtheorie

8.2.1 Reliabilitätsanalyse

Das Ergebnis der Reliabilitätsanalysen (siehe Tabelle 8) zeigt, dass die innere Konsistenz der beiden Aufgabenmodi jeweils mit einem Cronbach-Alpha $> .70$ ausreichend hoch ausfällt. Mit Hilfe der Testlängenkorrektur nach Spearman-Brown kann die Reliabilität des Aufgabenmodus „ohne-Gitter“ mit der Reliabilität des Aufgabenmodus „mit-Gitter“ verglichen werden.

Tabelle 8: Innere Konsistenzen der beiden Aufgabenmodi ($n = 199$).

Aufgabenmodus	Cronbach's Alpha (KR20)	Anzahl der Items
mit Gitter	.90	17
ohne Gitter	.75	7
ohne Gitter (korrigiert nach Spearman-Brown)	.88	17
Gesamt 24	.92	24

Anmerkung: Cronbach's Alpha (Kuder-Richardson-Formel 20) = Reliabilität.

8.2.2 Klassische Itemschwierigkeiten

In der klassischen Testtheorie stellt die Lösungswahrscheinlichkeit (Mittelwert) eines Items seine Itemschwierigkeiten dar (siehe Tabelle 9). Tabelle 10 zeigt die durchschnittlich erreichte Leistung in den beiden Aufgabenmodalitäten.

Tabelle 9: Testitems und ihre Schwierigkeitsindizes (Mittelwert).

Itemnummer im Test	Itemnummer bei Auswertung	Modi	n	M
Item1 (Warm up)	eval0	oG	199	---
Item2	eval1	oG	199	.54
Item3	eval2	oG	199	.45
Item4	eval3	oG	199	.43
Item5	eval4	oG	199	.41
Item6	eval5	oG	199	.53
Item7	eval6	oG	199	.31
Item8	eval7	oG	199	.40
Item9	eval8	mG	199	.55
Item10	eval9	mG	199	.49
Item11	eval10	mG	199	.53
Item12	eval11	mG	199	.49
Item13	eval12	mG	199	.60
Item14	eval13	mG	199	.60
Item15	eval14	mG	199	.50
Item16	eval15	mG	199	.52
Item17	eval16	mG	199	.54
Item18	eval17	mG	199	.46
Item19	eval18	mG	199	.64
Item20	eval19	mG	199	.51
Item21	eval20	mG	199	.68
Item22	eval21	mG	199	.46
Item23	eval22	mG	199	.59
Item24	eval23	mG	199	.50
Item25	eval24	mG	199	.54

Anmerkung: Itemnummer im Test (Itemxx), Itembezeichnung für die Auswertung, Modus (oG vs. mG), n = Anzahl, M = Itemlösungswahrscheinlichkeit

Tabelle 10: Deskriptivstatistik der durchschnittlich erreichten Leistungen in den beiden Aufgabenmodalitäten (Gesamtstichprobe).

	n	M
Testwert mG	199	.54
Testwert oG	199	.44

Anmerkung: n = Anzahl, M = Mittelwert

Der neue Test soll schwierig genug sein, um als Power-Tests eingesetzt werden zu können, da Speedtestungen mit Abbruchkriterium streng genommen für RM-Analysen ungeeignet sind (Gittler, 1986). Denn auch wenn sich Items als Rasch-homogen herausstellen sollten, können diese nicht informativ und somit für eine Test ungeeignet sein, wenn sie zu niedrige oder zu hohe Lösungswahrscheinlichkeiten besitzen (Gittler, 1986).

Dies ist hier nicht der Fall, denn mit mehrheitlich rund 50% Lösungswahrscheinlichkeit sind sie relativ informativ (siehe Tabelle 9). Abbildung 11 und Abbildung 12 veranschaulichen die Itemschwierigkeiten für die Gesamtstichprobe und darüber hinaus in Abhängigkeit vom Geschlecht.

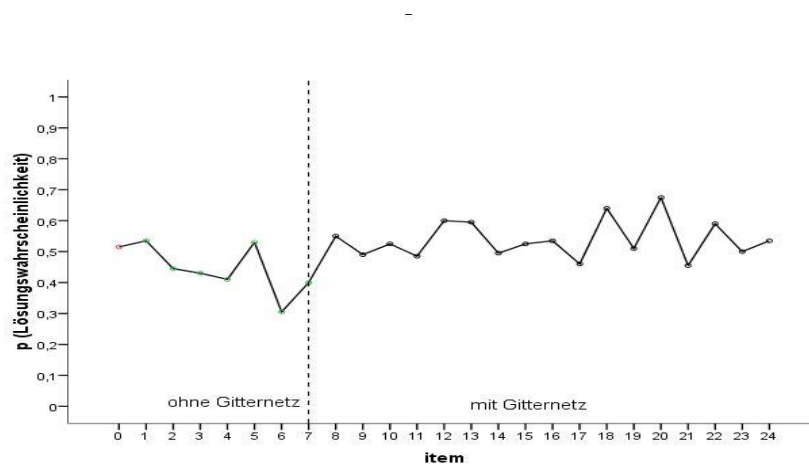


Abbildung 11: Darstellung der klassischen Lösungswahrscheinlichkeiten für jedes Testitem im Gesamttest (0-Item ist Warm-up-Item; oG-Item 0-7, mG-Aufgaben Item 8-24).

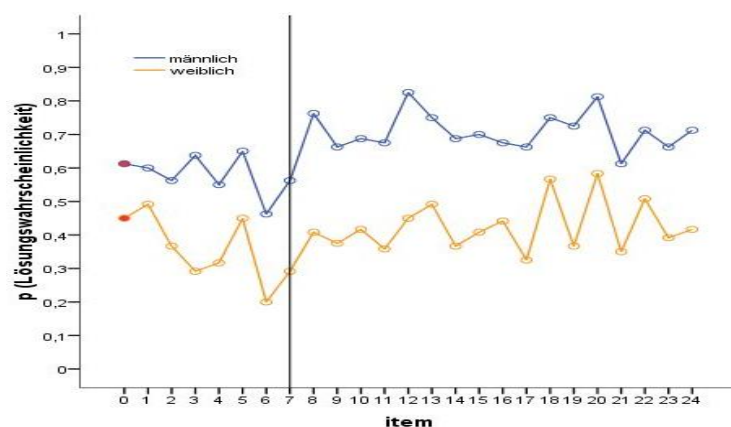


Abbildung 12: Darstellung der klassischen Lösungswahrscheinlichkeiten je Testitem im Gesamttest, getrennt nach Geschlecht.

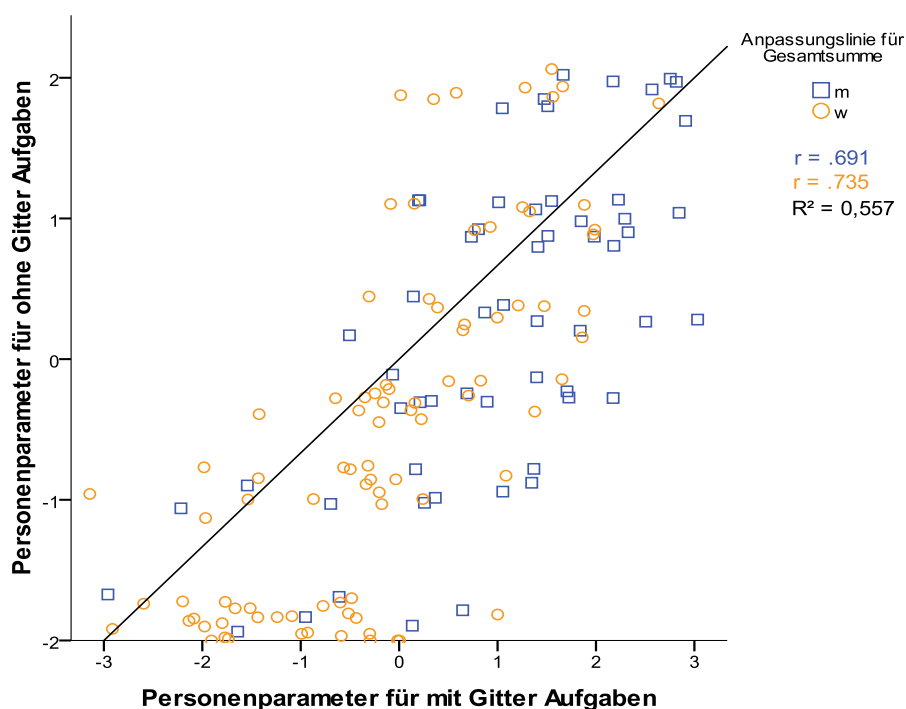


Abbildung 13: Bivariates Streudiagramm zur Darstellung des Zusammenhangs zwischen den Personenparametern der „ohne-Gitter“- vs. „mit-Gitter“-Aufgaben, getrennt nach Geschlecht.

Die Korrelation zwischen den zwei Testteilen ermittelt anhand der intervallskalierten Personenparameter θ_{oG} und Personenparameter θ_{mG} erreicht mit $r = .747$, $p < .001$ ein signifikantes Ergebnis und weist dementsprechend auf einen deutlich positiven Zusammenhang hin (erklärter Varianzanteil 55.7 %); somit erfassen die beide Aufgabenmodi Ähnliches, was auch schon mit der Eindimensionalität nach Rasch

gezeigt werden konnte. Die Abbildung 13 veranschaulicht den Zusammenhang zwischen oG- und mG-Aufgaben anhand der intervallskalierten Personenparameter. In Tabelle 11 sind die den jeweiligen Testwerten zugeordneten Personenparameter für den Gesamttest, oG-Aufgaben und mG-Aufgaben aufgelistet.

Der Zusammenhang zwischen den Leistungen (Personenparameterniveau) in den beiden Aufgabenmodalitäten getrennt nach Geschlecht weist auf hohe, signifikante Korrelationen hin ($r_m = .691, p < .001$; $r_w = .735, p < .001$).

Tabelle 11: Personenparameter in Abhängigkeit der Anzahl richtig gelöster Aufgaben (Gesamttest, nur „mit-Gitter“-Aufgaben, nur „ohne-Gitter“-Aufgaben).

Testwert	Personenparameter Gesamttest	Personenparameter mG-Aufgaben	Personenparameter oG-Aufgaben
0	---	---	---
1	-3.2425	-2.8396	-1.8599
2	-2.4948	-2.0715	-0.9583
3	-2.0329	-1.5869	-0.3030
4	-1.6867	-1.2154	0.2998
5	-1.4027	-0.9028	0.9567
6	-1.1568	-0.6244	1.8617
7	-0.9360	-0.3661	---
8	-0.7320	-0.1187	
9	-0.5409	0.1250	
10	-0.3572	0.3719	
11	-0.1785	0.6291	
12	-0.0021	0.9059	
13	0.1744	1.2164	
14	0.3533	1.5853	
15	0.5374	2.0667	
16	0.7295	2.8311	
17	0.9337	---	
18	1.1553		
19	1.4023		
20	1.6877		
21	2.0356		
22	2.4996		
23	3.2498		
24	----		

8.3 Überprüfung der Fragestellungen

8.3.1 Gibt es leistungsmäßige Unterschiede zwischen den Testteilen (oG/mG) in Abhängigkeit vom Geschlecht?

Anhand einer zweifaktoriellen ANOVA (Modi x Geschlecht) soll geprüft werden, ob sich einerseits Leistungen zwischen den beiden Aufgabenmodi (oG vs. mG) und andererseits zwischen Männern und Frauen signifikant unterscheiden. Auch eine mögliche Wechselwirkung zwischen Aufgabenmodi und Geschlecht soll überprüft werden. Dabei stellen die Aufgabenmodi den Innersubjektfaktor bzw. Messwiederholungsfaktor (oG- vs. mG-Aufgaben) und das Geschlecht den Zwischensubjektfaktor dar.

In die Berechnung gingen die Daten von $n = 147$ ($m = 58$, $w = 89$) Testpersonen ein, da im Rahmen des RM Personen, die keine einzige bzw. alle Aufgaben gelöst haben, nicht in die Analysen eingehen, und nur jene Testpersonen genommen werden konnten, die in beiden Testteilen zumindest eine Aufgabe gelöst haben. Tabelle 12 zeigt die Personenparameter für beide Aufgabenmodalitäten, getrennt nach Geschlecht.

Tabelle 12: Personenparameter in Abhängigkeit von Aufgabenmodus und Geschlecht.

	<i>N</i>	<i>M</i> _{θ_{oG}}	<i>SD</i> _{θ_{oG}}	<i>M</i> _{θ_{mG}}	<i>SD</i> _{θ_{mG}}
Männer	58	.20	1.14	.99	1.29
Frauen	89	-.54	1.20	-.13	1.27

Anmerkung: n = Fallzahl, M = Mittelwert, SD = Standardabweichung; θ_{oG} = Personenparameter für die oG-Aufgaben, θ_{mG} = Personenparameter für die mG-Aufgaben.

Die Gleichheit der Kovarianzenmatrizen auf Basis des Box-M-Tests kann mit $F(3, 728525.194) = 0.21$, $p = .893$ angenommen werden.

Der Levene-Test zur Überprüfung der Gleichheit der Varianzen fällt sowohl für θ_{oG} mit einer Prüfgröße von $F(1, 145) = 0.24$, $p = .622$ und für θ_{mG} mit einer Prüfgröße von $F(1, 145) = 0.002$, $p = .967$ nicht signifikant aus; die Varianzhomogenität kann somit angenommen werden.

Der Interaktionseffekt zwischen Aufgabenmodus und Geschlecht weist mit $F(1, 145) = 5.90$, $p = .016$ ($\eta^2 = .039$) ein signifikantes Ergebnis auf. Wie aus Abbildung 14 ersichtlich, schneiden sowohl Männer als auch Frauen bei mG-Aufgaben ($M_{m \theta mG} = 0.99$, $M_{w \theta mG} = 0.13$; $\eta^2 = .287$) signifikant besser ab als bei oG-Aufgaben ($M_{m \theta oG} = 0.20$; $M_{w \theta oG} = 0.54$). Außerdem zeigt sich, dass der Unterschied zwischen den beiden Aufgabenmodi bei den Männern signifikant größer ausfällt als bei den Frauen, wodurch das Gitternetz Männern offenbar entgegenkommt.

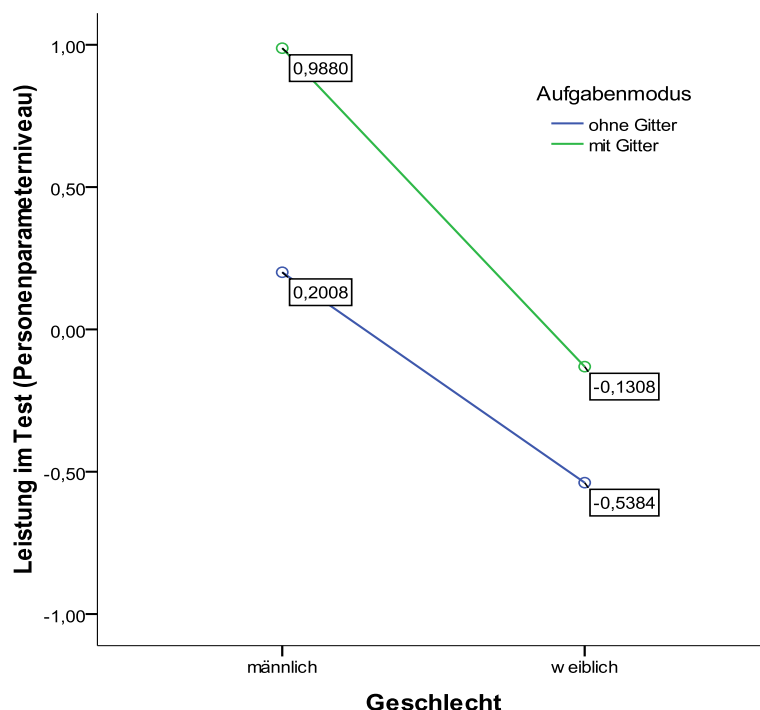


Abbildung 14: Interaktionsdiagramm in Abhängigkeit von Aufgabenmodus und Geschlecht.

Es zeigt sich mit $F(1, 145) = 58.45$, $p < .001$ ($\eta^2 = .287$) ein signifikanter Haupteffekt des Aufgabenmodus, wobei mG-Aufgaben signifikant öfter gelöst wurden. Der Haupteffekt Geschlecht weist mit $F(1, 145) = 23.42$, $p < .001$ ($\eta^2 = .139$) ebenfalls ein signifikantes Ergebnis auf, wobei Männer über beide Aufgabenmodi hinweg bessere Leistungen erzielten.

8.3.2 Gibt es Leistungsunterschiede in a) der Bearbeitung zwischen den Link-Items 2,3,4 (oG) und 23,24,25 (mG) und b) in Abhängigkeit vom Geschlecht?

Bei den Link-Items handelt es sich um strukturell gleiche Aufgaben, die sich nur hinsichtlich des (Nicht-)Vorhandenseins des Gitternetzes voneinander unterscheiden. Anhand einer zweifaktoriellen ANOVA (2x2 Link-Items x Geschlecht) mit Messwiederholung wird geprüft, ob sich die Leistungen (Testwerte) in Abhängigkeit von den beiden Aufgabenmodi (oG vs. mG) einerseits und zwischen Männern und Frauen andererseits zwischen den Link-Items signifikant voneinander unterscheiden. Auch eine mögliche Wechselwirkung zwischen Aufgabenmodus und Geschlecht wird geprüft. Die Aufgabenmodalität (oG- vs. mG-Aufgaben) entspricht hierbei dem Innersubjekt- bzw. Messwiederholungsfaktor und das Geschlecht dem Zwischensubjektfaktor ($m = 80$, $w = 119$). In Tabelle 13 sind die Deskriptivstatistiken für die Link-Paare (2-23, 3-24, 4-25), getrennt nach Geschlecht, angeführt.

Tabelle 13: Lösungswahrscheinlichkeit der Link – Items in Abhängigkeit von Aufgabenmodus und Geschlecht.

		Geschlecht	n	<i>M</i>	<i>SD</i>
1.Link-Paar	Item2 oG	m	80	.60	.49
		w	119	.50	.50
	Item23 mG	m	80	.71	.46
		w	119	.51	.50
2. Link-Paar	Item3 oG	m	80	.56	.50
		w	119	.37	.49
	Item24 mG	m	80	.66	.48
		w	119	.40	.49
3.Link-Paar	Item4 oG	m	80	.64	.48
		w	119	.29	.45
	Item25 mG	m	80	.71	.46
		W	119	.42	.50

Anmerkung: n = Anzahl, M = Mittelwert, SD = Standardabweichung, oG = „ohne-Gitter“-Aufgaben, mG = „mit-Gitter“-Aufgaben

Die Ergebnisse können trotz der Verletzung der Homogenität der Varianzen ($p < .05$) unter Berücksichtigung der etwa gleich großen Zellenbesetzungen aufgrund der Robustheit der Varianzanalysen ohne Einschränkung interpretiert werden.

Die Homogenität der Kovarianzenmatrizen, geprüft mittels Box-M-Test, kann mit $p = .337$ angenommen werden.

Es zeigen sich keine Wechselwirkungen in allen drei Link-Paaren zwischen Geschlecht und Aufgabenmodus ($F_{1, \text{Link-Paar}}(1, 197) = 1.66, p = .199 (\eta^2 = .008)$; $F_{2, \text{Link-Paar}}(1, 197) = 0.907, p = .342 (\eta^2 = .005)$; $F_{3, \text{Link Paar}}(1, 197) = 0.494, p = .483 (\eta^2 = .003)$).

Die Prüfgröße für den Haupteffekt Aufgabenmodus (oG vs. mG) fällt für das erste Link-Paar mit einer Prüfgröße von $F(1, 197) = 3.03, p = .083$ und für das zweite Link-Paar mit $F(1, 197) = 2.542, p = .112$ nicht signifikant aus.

Für das dritte Link-Item ergibt sich mit $F(1, 197) = 6.13, p = .014, (\eta^2 = .030)$ ein signifikantes Ergebnis, mG-Aufgaben wurden signifikant öfter gelöst.

Die Prüfgröße für den Haupteffekt Geschlecht fällt für alle Link-Paare signifikant aus (1. Link-Paar mit $F(1, 197) = 6.31, p = .013 (\eta^2 = .031)$, 2. Link-Paar $F(1, 197) = 15.43, p < .001 (\eta^2 = .073)$, 3. Link-Paar $F(1, 197) = 35.90, p < .001 (\eta^2 = .154)$). Männer lösten in allen drei Link-Paaren signifikant mehr Aufgaben.

Zusammenfassend zeigen sich in allen drei Link-Paaren keine Wechselwirkungen aus Geschlecht und Aufgabenmodus. Ungeachtet des Geschlechts ergibt sich ein signifikanter Leistungsunterschied beim dritten Link-Paar, wobei die mG-Aufgaben signifikant öfter gelöst wurden ($\eta^2 = .030$). Zudem lösen Männer bei allen Link-Paaren signifikant mehr Aufgaben als Frauen, mit einem großen Effekt ($\eta^2 = .154$) beim dritten Link-Paar.

Aus der Bearbeitung der letzten beiden Fragestellungen lässt sich festhalten, dass sowohl der hoch signifikante Geschlechtsunterschied ($\eta^2 = .139$) als auch der hoch signifikante Unterschied hinsichtlich des Aufgabenmodus ($\eta^2 = .287$) bei der direkten Gegenüberstellung strukturell gleicher oG- und mG-Aufgaben (Link-Aufgaben) bezüglich des Aufgabenmodus verschwindet bzw. praktisch irrelevant wird ($\eta^2 = .030$), so dass sich hinsichtlich des Geschlechts kein klares Bild ergibt, da alle Effektstärken ($\eta^2_{\text{Link-Paar1}} = .031$; $\eta^2_{\text{Link-Paar2}} = .073$, $\eta^2_{\text{Link-Paar3}} = .154$) vertreten sind.

8.3.3. Gibt es Unterschiede in der Bearbeitungszeit zwischen a) den beiden Aufgabenmodi (oG vs. mG) und b) in Abhängigkeit vom Geschlecht?

Mittels einer zweifaktoriellen ANOVA (Modus x Geschlecht) wird geprüft, ob sich Männer und Frauen signifikant in ihrer Bearbeitungszeit in Abhängigkeit vom Aufgabenmodus unterscheiden. Als Zwischensubjektfaktor gilt hier wieder das Geschlecht ($m = 80$, $w = 119$) und als Innersubjektfaktor der Aufgabenmodus (oG vs. mG). Es sollen auch mögliche Wechselwirkungen aus Geschlecht und Aufgabenmodus untersucht werden.

Abbildung 15 zeigt in einem Histogramm die Verteilung der Bearbeitungszeit (in Minuten) für den Gesamttest über die gesamte Stichprobe hinweg. Dies ist in Tabelle 14 auch numerisch für den Gesamttest und für beide Testteile getrennt dargestellt.

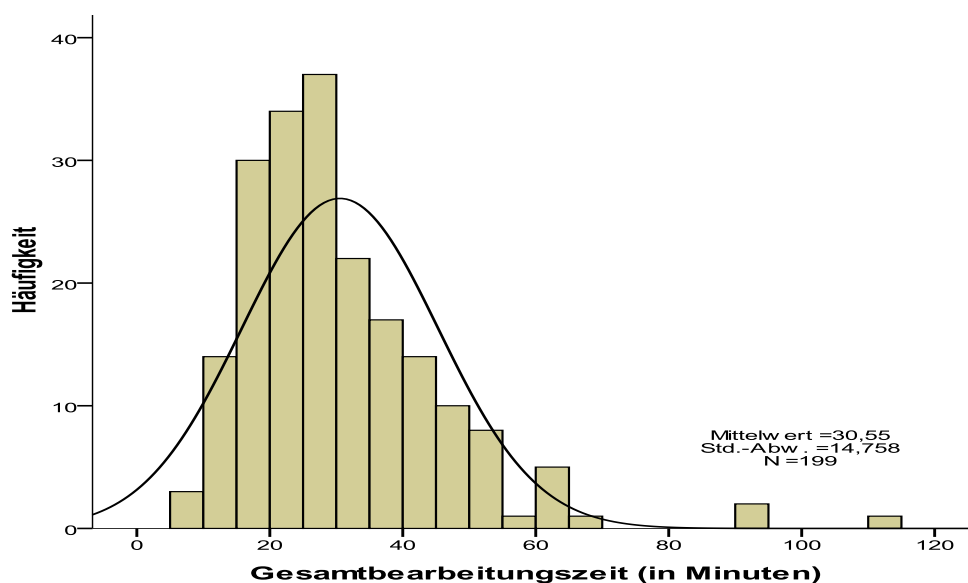


Abbildung 15: Histogramm zur Bearbeitungszeit im Gesamttest (in Minuten).

Tabelle 14: Gesamtbearbeitungszeit (in Minuten) für den Gesamttest und je Testteil (oG.- vs. mG-Aufgaben) getrennt nach Geschlecht.

		Gesamttest (Minuten)	oG Teil (Minuten)	mG Teil (Minuten)
m (n = 80)	<i>M</i>	31.07	11.07	19.99
	<i>Median</i>	27.65	9.92	18.32
	<i>SD</i>	15.52	6.21	10.14
	<i>Min</i>	11.06	3.3	7.76
	<i>Max</i>	110.05	34.51	77.43
w (n = 119)	<i>M</i>	30.20	10.90	19.30
	<i>Median</i>	26.97	9.28	16.96
	<i>SD</i>	14.28	5.91	9.22
	<i>Min</i>	8.97	2.59	5.38
	<i>Max</i>	90.66	31.79	58.88

Anmerkung: n = Fallzahl, *M* = Mittelwert, *SD* = Standardabweichung, Min = Minimum, Max = Maximum

Die mittlere Itembearbeitungszeit (in Sekunden) für oG-Aufgaben beträgt bei den Männern $M = 94.92$ Sekunden ($SD = 53.24$) bei Frauen $M = 93.44$ ($SD = 50.61$), bei den mG-Aufgaben brauchten Männer im Schritt $M = 70.57$ Sekunden ($SD = 35.78$) und Frauen $M = 68.13$ Sekunden ($SD = 32.55$).

Die Varianzenhomogenität kann für mG-Aufgaben mit $p = .495$ und für oG-Aufgaben mit $p = .786$ angenommen werden. Auch die Homogenität der Kovarianzenmatrizen, geprüft mittels Box-M-Test, kann mit $p = .827$ angenommen werden.

Mit $F(1, 197) = 0.04$, $p = .841$ ist keine signifikante Wechselwirkung zwischen dem Aufgabenmodus und dem Geschlecht anzunehmen (siehe Abbildung 16). Für den Haupteffekt Aufgabenmodus ergibt sich mit $F(1, 197) = 108.27$, $p < .001$, ($\eta^2 = .355$) ein signifikanter Effekt, und zwar wurden mG-Aufgaben schneller bearbeitet als oG-Aufgaben. Für den Haupteffekt Geschlecht ergibt sich mit $F(1, 197) = 0.11$, $p = .737$ kein signifikanter Effekt. Somit unterscheiden sich die Bearbeitungszeiten in beiden Aufgabenmodi in Abhängigkeit vom Geschlecht nicht voneinander.

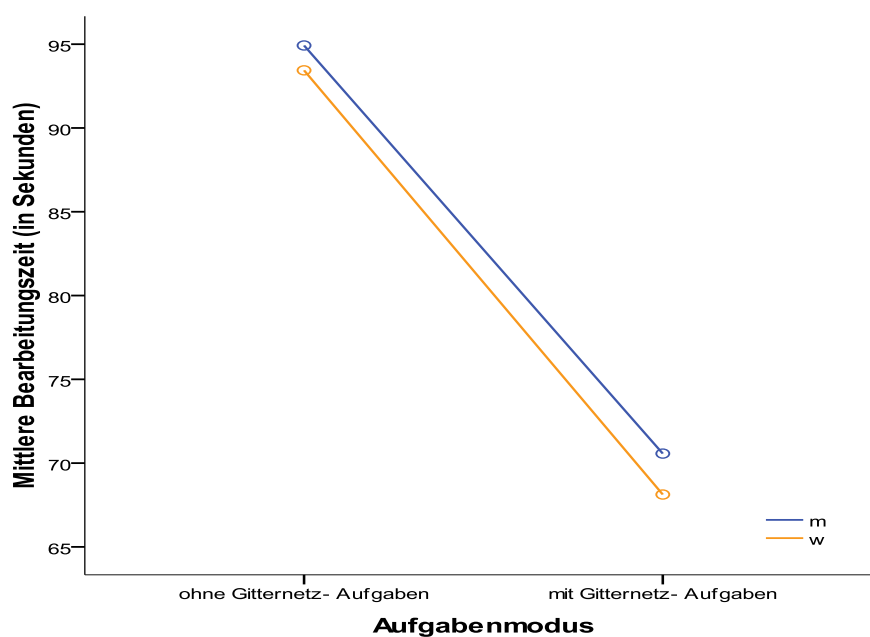


Abbildung 16: Bearbeitungszeit in Abhängigkeit von Aufgabenmodus und Geschlecht.

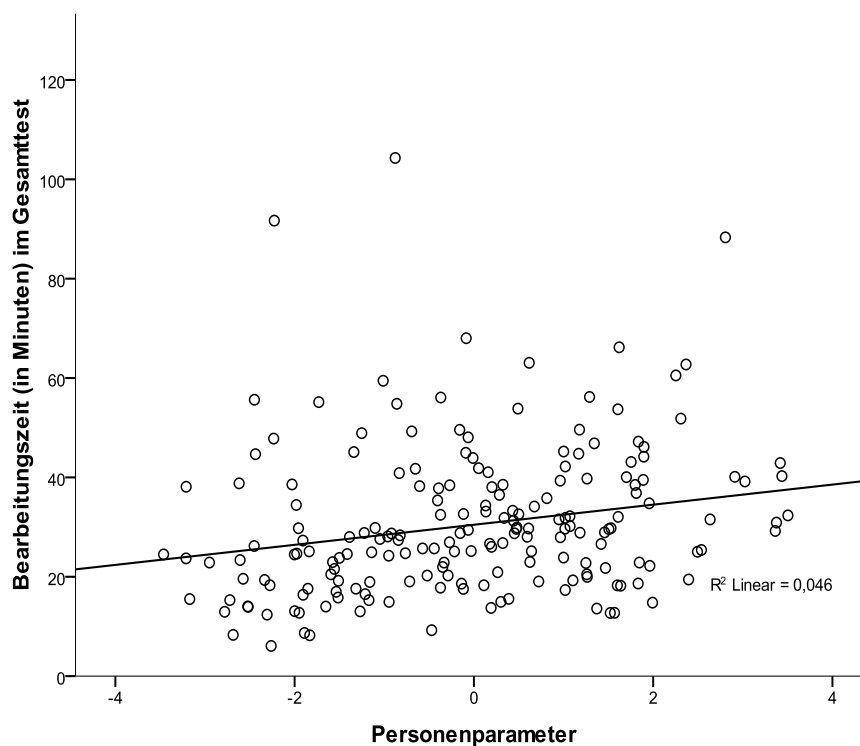


Abbildung 17: Bivariates Streudiagramm zwischen Leistung (Parameterniveau) und Bearbeitungszeit (in Minuten) im Gesamttest.

Die Korrelation zwischen den erbrachten Leistungen und der benötigten Bearbeitungszeit weist auf einen schwach signifikant positiven Zusammenhang hin ($r = .215$, $p = .003$), wie in der Abbildung 17 ersichtlich.

8.3.4 Kann das externe Kriterium Selbsteinschätzung der „Orientierungsfähigkeit im Raum“ die Leistung der Probanden im Test vorhersagen?

Wie in Tabelle 15 ersichtlich, schätzen Männer ihre Orientierungsfähigkeit im Raum höher ein als Frauen. Das bivariate Streudiagramm in Abbildung 18 stellt den Zusammenhang zwischen den Personenparametern und der Selbsteinschätzung dar.

Tabelle 15: Deskriptivstatistik zum Selbstrating mittels Analogskala (0) *sehr schlecht* bis (100) *sehr gut* der eigenen Orientierungsfähigkeit im Raum.

	Geschlecht	<i>n</i>	<i>M</i>	<i>SD</i>
Orientierung	Weiblich	119	69.24	26.20
	Männlich	80	79.10	21.28

Anmerkung: *n* = Anzahl, *M* = Mittelwert, *SD* = Standardabweichung

Die Produkt-Moment-Korrelation nach Pearson für den Personenparameter und dem Selbstrating ($n = 191$) hinsichtlich der eigenen „Orientierungsfähigkeit im Raum“ weist auf einen schwach signifikant positiven Zusammengang von $r = .225$ ($p = .002$, $R^2 = .051$) hin.

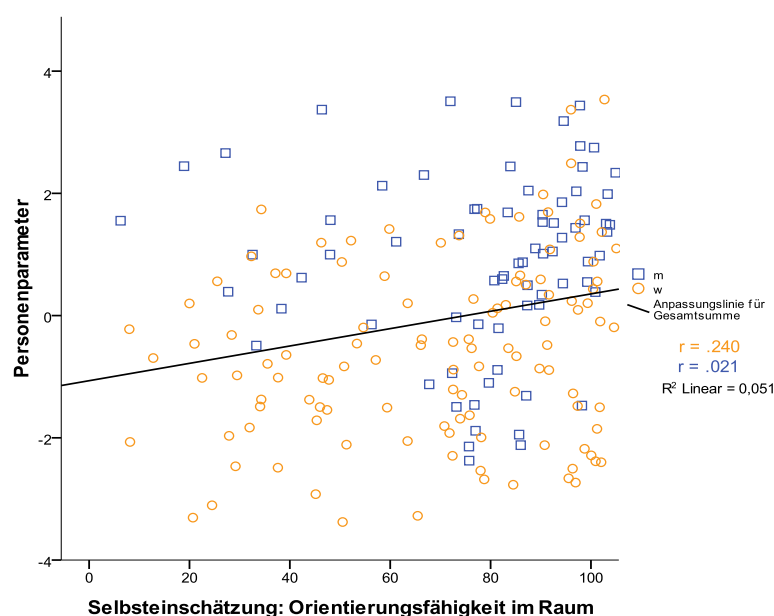


Abbildung 18: Bivariates Streudiagramm für den Zusammenhang zwischen Personenparameter im Gesamtttest und der Selbsteinschätzung "Orientierungsfähigkeit im Raum" (Analogskala 0-100).

Die Korrelationen getrennt für Männer und Frauen weisen ebenfalls auf schwach signifikant positive Zusammenhänge ($r_m = .021$, $p = .856$; $r_w = .240$, $p = .009$) hin.

Der t-Test für unabhängige Stichproben ergibt mit $t(197) = 2.8$, $p = .006$ einen signifikanten Mittelwertunterschied hinsichtlich der Selbsteinschätzung in der Orientierungsfähigkeit zwischen Männern und Frauen, wobei Männer ihre Fähigkeit signifikant höher einschätzen als Frauen ($d = 0.30$).

8.3.5 Zusatzfragen: Unterscheiden sich die subjektiven Ratings bezüglich der Aufgabenmodalität in Abhängigkeit vom Geschlecht?

Anhand einer zweifaktoriellen ANOVA (Modi x Geschlecht) mit Messwiederholung soll überprüft werden, ob sich Männer und Frauen in ihrem Rating hinsichtlich der Zusatzfragen signifikant voneinander unterscheiden. Den Innersubjektfaktor stellen die Aussagenpaare (oG-vs. mG-Aussage) dar und Zwischensubjektfaktor ist das Geschlecht ($m = 78$, $w = 113$). Am Ende der Datenerhebung konnten 191 Fragebögen vollständig ausgewertet werden. Weitere acht Fälle konnten auf Grund von einzelnen „missing

values“ nicht analysiert werden. In Tabelle 16 sind die Aussagenpaare (siehe Abbildung 8) und ihre durchschnittlichen Bewertungen, getrennt nach Männern und Frauen, angeführt.

Die Verletzung der Homogenität der Varianzen bei der Aussage4 mit $F(1, 189) = 6.57$, $p = .011$ ist aufgrund der etwa gleich großen Zellenbesetzungen in Anbetracht der Robustheit der Varianzanalyse ohne Einschränkung interpretierbar.

Es zeigen sich keine Wechselwirkungen zwischen den Fragenpaaren in Abhängigkeit vom Geschlecht ($F_{\text{Aussagenpaar1}}(1, 189) = 0.352$, $p = .554$, $F_{\text{Aussagenpaar2}}(1, 189) = 2.175$, $p = .142$; $F_{\text{Aussagenpaar3}}(1, 189) = 3.065$, $p = .082$, $F_{\text{Aussagenpaar4}}(1, 189) = 0.001$, $p = .971$, $F_{\text{Aussagenpaar5}}(1, 189) = 0.417$, $p = .519$, $F_{\text{Aussagenpaar6}}(1, 189) = .054$, $p = .816$).

Die Prüfgröße für den Haupteffekt Geschlecht fällt bei dem ersten, vierten und fünften Aussagenpaar signifikant aus mit den Prüfgrößen $F_{\text{Aussagenpaar1}}(1, 189) = 4.07$, $p = .045$ ($\eta^2 = .021$), $F_{\text{Aussagenpaar4}}(1, 189) = 3.97$, $p = .044$ ($\eta^2 = .021$) und $F_{\text{Aussagenpaar5}}(1, 189) = 9.14$, $p = .003$ ($\eta^2 = .046$), und zwar haben Männer diesen Aussagenpaaren signifikant mehr Zustimmung gegeben.

Konkret bedeutet das, dass Männer sowohl bei oG- als auch bei mG-Aufgaben die Figur signifikant stärker als Ganzes gedreht haben und ihnen die Erfassung der Gesamtfigur und die Bearbeitung der Aufgaben insgesamt signifikant leichter gefallen ist als Frauen.

Tabelle 16: Deskriptivstatistik des Ratings der Zusatzfragenpaare getrennt nach Geschlecht (Skala: 4 trifft sehr zu bis 1 trifft gar nicht zu).

		Geschlecht	M	SD
Paar 1	Aussage 01 oG	m (n = 78)	3.29	.93
		w (n = 113)	3.09	.95
„Als Ganzes mental gedreht“	Aussage 02 mG	m	3.17	.90
		w	2.88	1.05
Paar 2	Aussage 03 oG	m	3.19	.84
		w	3.17	.90
„An Teilaspekten orientiert“	Aussage 04 mG	m	3.42	.80
		w	3.61	.65
Paar 3	Aussage 05 oG	m	2.76	.91
		w	2.85	.91
„Eindruck von räumlicher Tiefe“	Aussage 06 mG	m	3.27	.80
		w	3.08	.89
Paar 4	Aussage 07 oG	m	2.68	.88
		w	2.48	.93
„Erfassen der Gesamtfigur fiel leicht“	Aussage 08 mG	m	3.35	.77
		w	3.15	.85
Paar 5	Aussage 09 oG	m	2.36	.76
		w	2.05	.77
„Bearbeitung insgesamt fiel leicht“	Aussage 10 mG	m	3.21	.67
		w	2.99	.81
Paar 6	Aussage 11 oG	m	2.33	.86
		w	2.38	.92
„Empfand es optisch ansprechend“	Aussage 12 mG	m	2.99	.89
		W	2.99	.82

Anmerkung: n = Anzahl, M = Mittelwert, SD = Standardabweichung

Für die restlichen drei Aussagenpaare ergaben sich keine signifikanten Unterschiede im Beantworten der Aussagen in Abhängigkeit vom Geschlecht ($F_{\text{Aussagenpaar2}}(1,189) = 0.78, p = .379$; $F_{\text{Aussagenpaar3}}(1, 189) = 0.51, p = .636$, $F_{\text{Aussagenpaar6}}(1, 189) = 0.08, p = .774$).

Für den Haupteffekt Aussagen zeigen sich für alle Aussagenpaare signifikante Unterschiede im Antwortverhalten zwischen oG- und mG-Aussagen unabhängig vom Geschlecht, wobei der Mittelwertsunterschied des ersten Aussagenpaares in eine andere Richtung geht als alle anderen: Nur beim ersten Aussagenpaar bekommt die oG-Aussage signifikant mehr Zustimmung $F(1, 189) = 5.76, p = .017$ ($\eta^2_{\text{Aussagenpaar1}} = .030$), d.h. oG-Aufgaben wurden signifikant häufiger geistig als Ganzes gedreht.

Bei alle weiteren Aussagenpaaren erhalten die mG-Aussagen signifikant mehr Zustimmung. Die Testpersonen haben sich bei mG-Aufgaben signifikant stärker an Teilaspekten der Figur orientiert ($F_{\text{Aussagenpaar2}}(1, 189) = 22.00, p < .001 (\eta^2 = .104)$); sie hatten bei mG-Aufgaben einen signifikant stärkeren Eindruck von räumlicher Tiefe ($F_{\text{Aussagenpaar3}}(1, 189) = 21.16, p < .001 (\eta^2 = .101)$), ihnen sei das Erfassen der Gesamtfigur bei mG-Aufgaben leichter gefallen ($F_{\text{Aussagenpaar4}}(1, 189) = 68.54, p < .001 (\eta^2 = .266)$); ihnen sei insgesamt die Bearbeitung der mG-Aufgaben signifikant leichter gefallen ($F_{\text{Aussagenpaar5}}(1, 189) = 157.11, p < .001, (\eta^2 = .454)$) und sie hätten die Darstellung mit Gitter als optisch ansprechender empfunden ($F_{\text{Aussagenpaar6}}(1, 189) = 46.40, p < .001, (\eta^2 = .197)$).

8.3.5.1 Unterscheiden sich Männer und Frauen im Rating der 13. Aussage?

Anhand eines t-Tests für unabhängige Stichproben wird geprüft, ob sich Männer und Frauen in der Beantwortung der 13. Aussage „Bei Würfeln mit Gitternetz habe ich die einzelnen Kästchen abgezählt“ voneinander unterscheiden. Die Prüfgröße fällt mit $t(189) = 2.19, p = .029$ signifikant aus (standardisierte Effektgröße nach Cohen's $d = 0.29$). Somit unterscheiden sich Männer und Frauen hinsichtlich des Ratings der 13. Aussage, wobei Frauen dieser Aussage mehr zustimmen (Männer $M = 1.83$, Frauen $M = 1.55$), trotzdem befinden sich die Ratings im ablehnenden Bereich („*trifft eher nicht zu*“ (2)).

Zusätzlich wird mittels einer explorativen Faktorenanalyse - eine Hauptkomponentenanalyse mit nachfolgender orthogonaler Varimax Rotation nach Kaiser - untersucht, ob sich Eindrücke zum Aufgabenmaterial zu einzelnen Dimensionen zusammenfassen lassen. Es können zwei vergleichbare Dimensionen für oG- und mG-Fragen identifiziert werden. Der erste Faktor umfasst vier Fragen zu oG-Aufgaben und der zweite Faktor die entsprechenden vier Fragen für die mG-Aufgaben. Die restlichen Items (1-4 und 13) müssen aus dieser Faktorenanalyse ausgeschlossen werden, da sie nicht eindeutig auf diesen beiden Faktoren laden. Das Kaiser-Meyer-Oklun-Kriterium (KMO) = .705 weist darauf hin, dass die Daten für eine Hauptkomponentenanalyse brauchbar sind (Abbildung 17). Der erklärte Varianzanteil

an der Gesamtvarianz dieser beiden Faktoren zusammen beträgt 53%. In der Tabelle 17 sind die Deskriptivstatistiken und in Tabelle 18 die Ladungen für die beiden extrahierten Faktoren dargestellt. Abbildung 19 veranschaulicht den Eigenwerteverlauf der extrahierten Faktoren noch einmal bildhaft anhand eines Screeplots.

Tabelle 17: Deskriptivstatistiken der beiden extrahierten Faktoren.

	<i>M</i>	<i>SD</i>	<i>n</i>
Zusatzfrage 05 oG	2.81	.910	191
Zusatzfrage 06 mG	3.16	.856	191
Zusatzfrage 07 oG	2.56	.909	191
Zusatzfrage 08 mG	3.23	.820	191
Zusatzfrage 09 oG	2.18	.774	191
Zusatzfrage 10 mG	3.08	.760	191
Zusatzfrage 11 oG	2.36	.895	191
Zusatzfrage 12 mG	2.99	.846	191

Anmerkung: *M* = Mittelwert, *SD* = Standardabweichung, *n* = Anzahl

Tabelle 18: Ladungsmatrix - Ergebnisse des Rotationsverfahrens nach der Varimax-Methode – Extraktion zweier Faktoren.

Rotierte Ladungsmatrix^a		
	Komponente	
	1	2
Zusatzfrage 07 oG	.837	.089
Zusatzfrage 09 oG	.758	.151
Zusatzfrage 11 oG	.706	-.075
Zusatzfrage 05 oG	.668	.065
Zusatzfrage 10 mG	.087	.751
Zusatzfrage 08 mG	.118	.693
Zusatzfrage 12 mG	-.237	.672
Zusatzfrage 06 mG	.180	.620

Die aus den aggregierten Items resultierenden Indizes („ohne-Gitter“- vs. „mit-Gitter“-Index) weisen mit $\alpha = .741$ (oG-Fragen) und $\alpha = .627$ (mG-Fragen) auf eine ausreichende Reliabilität dieser beiden Skalen hin. Abbildung 19 zeigt den Verlauf der Eigenwerte mit dem Knick beim 3. Faktor.

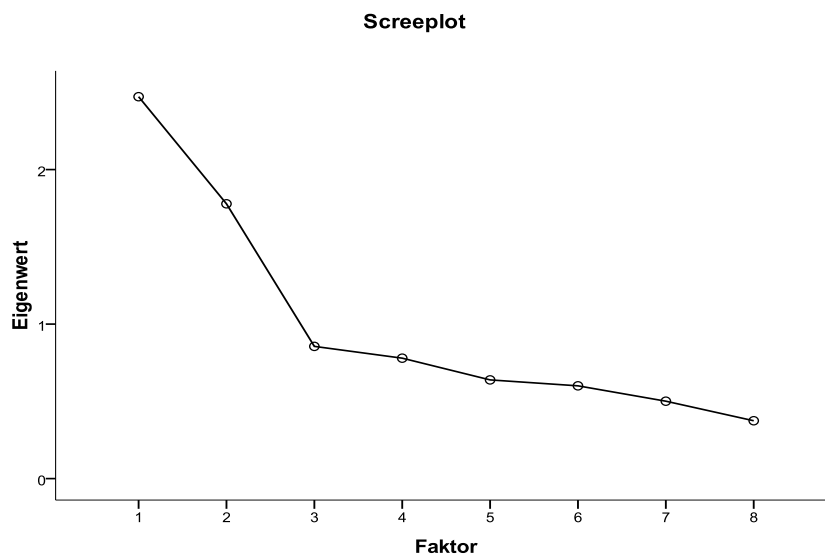


Abbildung 19: Screepplot: Graphische Darstellung des Eigenwerteverlaufs.

Die Frage, ob es Unterschiede zwischen den Empfindungen von „ohne-Gitter“- und „mit-Gitter“-Darstellungen gibt, wird anhand der gemittelten Empfindungsangaben beider Indizes mit einem t-Test abhängig berechnet. Die Berechnung der Prüfgröße fällt mit $t(190) = -10.88$, $p < .001$ signifikant aus. Es kann angenommen werden, dass Figuren mit Gitter von den Personen als entgegenkommender beurteilt werden als die ohne Gitter ($d = -1.58$). Der Zusammenhang zwischen oG- und mG-Empfindungen fällt sehr niedrig und nicht signifikant aus ($r = .126$, $p = .083$). Es zeigt sich kein Zusammenhang insofern, als oG unangenehmer empfunden werden (Abbildung 20).

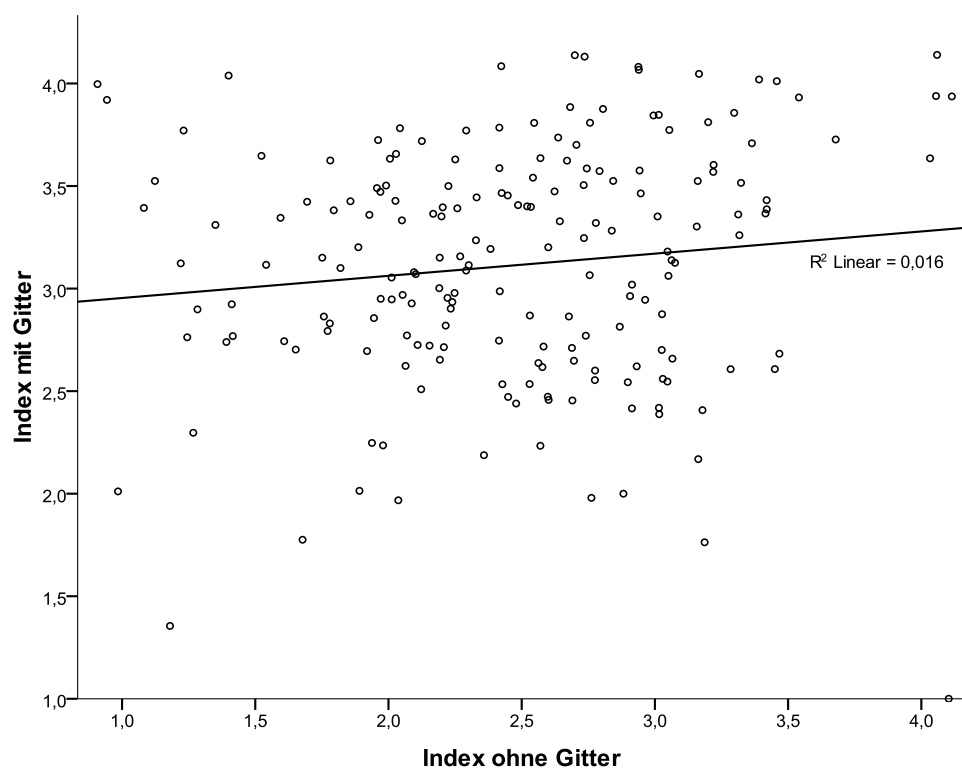


Abbildung 20: Zusammenhang zwischen den oG- und mG-Indizes.

9. Zusammenfassung und Diskussion

9.1 Zusammenfassung

Im Rahmen dieser Arbeit konnte die Gültigkeit des Rasch-Modells für den vorgelegten Test und seine zwei Teile (oG vs. mG-Aufgaben) bestätigt werden: Trotz der beiden Aufgabenmodalitäten scheint nur eine zugrundeliegende Fähigkeitsdimension gemessen zu werden.

Dabei weisen die Ergebnisse der Männer im Gesamttest bereits auf einen deutlichen Unterschied gegenüber den Ergebnissen der Frauen hin (standardisierte Effektgröße Cohen's $d = 0.89$; Bortz & Döring, 2006, S. 606). Auch nach der Berechnung einer ANCOVA kann der Unterschied in der Testleistung in Abhängigkeit vom Geschlecht mit einer Effektstärke von $\eta^2 = .188$ als deutlich bezeichnet werden. Beim Alter zeigt sich ein knapp mittelgroßer Effekt ($r = -.262$) dahingehend, dass die Leistung mit dem Alter abnimmt.

Auch innerhalb der beiden Aufgabenmodi zeigen Männer signifikant bessere Leistungen als die Frauen ($M_{m\ oG} = 0.20$, $SD_{m\ oG} = 1.14$; $M_{w\ oG} = -0.54$, $SD_{w\ oG} = 1.20$ und $M_{m\ mG} = 0.98$, $SD_{m\ mG} = 1.29$; $M_{w\ mG} = -0.13$, $SD_{w\ mG} = 1.27$; $\eta^2_{\text{Haupteffekt Geschlecht}} = .139$): Sie lösten in allen Bedingungen beinahe doppelt so viele Aufgaben wie die Frauen. Der hier beobachtete Leistungsvorsprung entspricht den schon vielfach in der Literatur berichteten Ergebnissen (Voyer et al., 1995). Insgesamt und ungeachtet des Geschlechts erscheinen mG-Aufgaben signifikant leichter ($\eta^2 = .287$); wodurch der Aufgabenmodus einen Einfluss auf die Schwierigkeit der Items zu haben scheint.

Die direkte Gegenüberstellung der beiden Aufgabenmodi zeigte nur beim dritten Link-Paar signifikante Unterschiede, und zwar wurden mG-Aufgaben signifikant öfter gelöst, doch weist die dazugehörige sehr kleine Effektgröße ($\eta^2 = .030$) auf keine praktische Bedeutsamkeit dieses Unterschieds hin. Hinsichtlich des Geschlechts zeigte sich auch bei den Link-Aufgaben der signifikante Leistungsvorsprung der Männer, wobei sich nur

für das zweite Link-Paar eine mittlere ($\eta^2 = .073$) und für das dritte Link-Paar ($\eta^2 = .154$) eine sehr große Effektstärke ergaben.

Des Weiteren benötigen die mG-Aufgaben signifikant weniger Bearbeitungszeit ($\eta^2 = .355$) als oG-Aufgaben, was auch anhand ihrer mittleren Effektstärke verdeutlicht wird. Somit kann man von einem erleichternden Effekt durch das Gitternetz für die Bearbeitung dieses Aufgabentyps sprechen. Dabei scheint das Gitternetz Männern stärker entgegen zu kommen als Frauen, jedoch weist der Interaktionseffekt ($\eta^2 = .039$) auf eine kleine Effektstärke hin.

Obwohl hinsichtlich der Bearbeitungszeit (Gesamttest, oG-Aufgaben-Teil, mG-Aufgaben-Teil) zwischen den Geschlechtern keine signifikanten Unterschiede auftraten, lösten in derselben Zeit Männer beinahe doppelt so viele Aufgaben wie die Frauen.

Bei den Zusatzfragen zeigt sich bezüglich möglicher Unterschiede in Abhängigkeit vom Geschlecht folgendes Bild: Männer haben sowohl bei oG- als auch bei mG-Aufgaben die ganze Figur signifikant öfter gedreht ($M_{m\ oG} = 3.29$, $SD_{m\ oG} = .93$; $M_{w\ oG} = 3.09$, $SD_{w\ oG} = .95$; $M_{m\ mG} = 3.17$, $SD_{m\ mG} = .90$; $M_{w\ mG} = 2.88$, $SD_{w\ mG} = 1.05$) als die Frauen, was den referierten Ergebnissen in der Literatur, Männer würden Raumvorstellungsaufgaben eher holistisch bearbeiten, entspricht (Glück et al., 2005). Jedoch weist dieses signifikante Ergebnis eine kleine entsprechende Effektgröße ($\eta^2 = .021$) auf. Da die Durchschnittswerte im oberen Ratingbereich liegen, könnte man daraus ableiten, dass eine hohe Zustimmung hinsichtlich einer ganzheitlichen Bearbeitung vorliegt. In Anlehnung an Gittlers (1999) Primär- und Co-Faktor könnte eine ganzheitliche Bearbeitung, also eine räumlich-visuelle Leistung, den Primärfaktor darstellen, der durch Ko-Faktoren unterstützt wird, ohne dabei die relativen Itemschwierigkeiten unterschiedlich zu beeinflussen.

Männern fiel sowohl das Erfassen der Gesamtfigur als auch die Bearbeitung der beider Aufgabenmodalitäten (oG- und mG-Aufgaben) signifikant leichter als Frauen, doch zeigen die jeweils dazu gehörenden Effektgrößen $\eta^2_{Aussagenpaar4} = .021$, $\eta^2_{Aussagenpaar5} = .046$ nur kleine Effekte. Die Ratings bewegen sich hierbei im Durchschnitt wieder im oberen Ratingbereich („trifft eher zu“ (3)).

Für die Aussagenpaare „an Teilaspekten orientiert“, „räumlicher Tiefeneindruck“ und „optisch ansprechend“ ergaben sich keine signifikanten geschlechtsspezifischen Unterschiede im Rating.

Nur bei dem Aussagenpaar „als Ganzes gedreht“ erhielt die oG-Aussage signifikant mehr Zustimmung, und zwar von beiden Geschlechtern gleichermaßen, wobei es sich aber mit $\eta^2 = .030$ nur um einen kleinen Effekt handelt. Auch hier fiel das Rating im zustimmenden Bereich aus („trifft eher zu“ (3)).

Bei allen anderen Aussagenpaaren wurde den mG-Aufgaben mehr Zustimmung gegeben. Somit empfanden sowohl Männer als auch Frauen, dass mG-Aufgaben einen stärkeren Eindruck von räumlicher Tiefe ($\eta^2 = .101$) vermittelten und optisch ansprechender ($\eta^2 = .197$) waren. Außerdem orientierte man sich stärker an Teilaspekten ($\eta^2 = .104$). Dabei wiesen alle mittlere Effektstärken auf. Das Rating lag auch bei diesen Fragen im zustimmenden Bereich („trifft eher zu“ (3)).

Große Effektstärken gab es für die Aussagen, dass das Erfassen von mG-Aufgaben in ihrer Gesamtgestalt leichter war ($\eta^2 = .266$) wie auch ihre Bearbeitung ($\eta^2 = .454$) leichter empfunden wurde.

Frauen gaben signifikant öfter an, die einzelnen Kästchen bei den mG-Aufgaben gezählt zu haben. Auch dies deckt sich mit der Literatur, wo den Frauen eher eine analytische Herangehensweise an Raumvorstellungsaufgaben zugeschrieben wird (Glück et al., 2005), doch zeigt sich mit $d = 0.29$ nur ein kleiner Effekt und das Rating im ablehnenden Bereich.

Die durch eine Faktorenanalyse ermittelten zwei Indizes für oG- und mG-Fragen weisen mit einer sehr großen Effektstärke von $d = -1.58$ ebenfalls auf einen empfundenen Unterschied zu Gunsten der „mit-Gitter“-Darstellung hin.

Hinsichtlich der Einschätzung der eigenen Orientierungsfähigkeit schätzten Männer ihre Fähigkeit signifikant höher ($d = 0.30$) ein als Frauen. Aufgrund der kleinen positiven Korrelation zwischen Selbsteinschätzung in der Orientierungsfähigkeit und dem

Testergebnis (Personenparameter) kann dieses Selbstrating nicht als ein entscheidendes externes Kriterium oder als ein signifikanter Prädiktor für die Leistung im Test betrachtet werden.

Noch einmal kurz zusammengefasst gibt es hinsichtlich möglicher Unterschiede im Rating der Abschlusss Aussagen zwischen den Geschlechtern keine praktisch relevanten Differenzen. Bezüglich unterschiedlicher Empfindungen gegenüber den beiden Aufgabenmodi zeigt sich eine Bevorzugung der mG-Aufgaben.

9.2 Interpretation

Da den mG-Aufgaben ein stärkerer Eindruck von räumlicher Tiefe zugeschrieben wurde (mittlere Effektstärke), dürfte hier ein möglicher Transformationsprozess (2d-3d) weniger ins Gewicht fallen.

Dass mit dem vorgelegten Testmaterial nur eine einzige latente Fähigkeitsdimension erfasst wird, konnte mit der Gültigkeit des Rasch-Modells gezeigt werden. Dieses Ergebnis stimmt auch mit den Zusatzfragen überein, da sich hier keine geschlechtsspezifischen Unterschiede in der Bearbeitungsstrategie zwischen den oG- und mG-Aufgaben zeigten: Männer und Frauen vollzogen gleichermaßen die ganzheitliche Bearbeitung als auch die Beachtung von Teilaspekten bei der Aufgabenlösung. Das Abzählen der Würfel schien bei mG-Aufgaben nicht die primäre Strategie zu sein.

Obwohl Frauen signifikant öfter angaben, die einzelnen Kästchen abgezählt zu haben, erweist die kleine Effektgröße diesen signifikanten Unterschied als praktisch irrelevant, wodurch dieser Aspekt und diese Herangehensweise nicht die primär durch diese Aufgaben angesprochene Fähigkeitsdimension zu sein scheint (mit einem durchschnittlichen Rating im ablehnenden Bereich „*trifft eher nicht zu*“ (2), was der Annahme einer einheitlichen holistischen Strategieanwendung den Vorzug gibt.

9.3 Diskussion

In Anlehnung an die Aussage, dass die erhaltenen Itemhomogenitäten erst nach einer Validierungsstudie anhand neuer Daten als ausreichend erwiesen angesehen werden können (Gittler & Arendasy, 2003), sollte im Rahmen einer zukünftigen Untersuchung anhand einer neuen Personenstichprobe eine Kreuzvalidierung (Validierungsdaten) durchgeführt werden (Gittler, 1984). Die dadurch möglicherweise erhaltene Rasch-Kalibrierung würde eine wichtige Voraussetzung für die Validität dieses Testmaterials bedeuten (Gittler & Glück, 1998).

Durch die Methode des LLTM (Linear Logistisches Testmodell) könnte man in zukünftigen Untersuchungen die Komponenten und Aspekte herausarbeiten, die für die unterschiedlichen Schwierigkeitsgrade homogener Items verantwortlich sind (vgl. Gittler, 1999), wie sich dies z.B. am erleichternden Effekt durch das Gitternetz gezeigt hat. Ein Grund dafür könnte darin liegen, dass das Gitternetz eine bessere Orientierung erlaubt, welche Seite oder Länge der Ausgangsfigur jener in der Vergleichsfigur entspricht, wie dies manche Versuchspersonen rückmeldeten.

Durch diese Analyse könnte man Strukturfaktoren der Aufgaben und damit verbundene elementare (kognitive) Operationen finden, um z.B. für das Item Nr.7 inhaltliche Erklärungen ableiten zu können, warum es bei manchen graphischen Modellkontrollen, u.a. beim trennschärfsten Teilungskriterium dem Median, trotz gegebener Rasch-Konformität auffällig weit weg von der Geraden liegt (Gittler, 1984): Welche Eigenschaften könnte dieses Item haben, dass es für die beiden Untergruppen (Personen mit hohem Testwert vs. Personen mit niedrigem Testwert) unterschiedlich schwierig ist (vgl. Gittler, 1999)? Dieser Zugang könnte auch mögliche systematische Schwierigkeitsunterschiede zwischen den oG- und mG-Aufgaben klären.

Des Weiteren ist interessant, dass bei gegebener Eindimensionalität des Aufgabenmaterials sich anhand der Abschlussfragen zwei Faktoren extrahieren lassen (einmal für oG- und einmal für mG-Fragen), die jedoch inhaltlich dasselbe erfassen.

Für zukünftige Rasch-Modell-Prüfungen könnte das Programm R (<http://www.r-project.org>) insofern von Vorteil sein, als es in den graphischen Modelltests die Konfidenzintervalle für die Items mit ausgibt. Dies würde die Interpretation erleichtern, ob ein Item im graphischen Modelltest auffällig weit weg von der 45°-Geraden liegt. Eine Zeiterfassung auf „Bild-Paar“-Niveau könnte vielleicht mehr Informationen liefern und theoretische Analysen ermöglichen hinsichtlich der Fragen, weshalb manche Probanden länger für die Bearbeitung bestimmter Items brauchten, oder weshalb manche Darstellungen seltener gelöst wurden. Des Weiteren würde eine Einzelbildpaar-Zeiterfassung auch erlauben zu prüfen, ob Frauen und Männer sich hinsichtlich der Bearbeitungszeit ausschließlich bei echten „Seitigkeitsaufgaben“ (Glück, 1999) unterscheiden oder nicht (vgl. Peters, 2005).

Zusätzlich zur jeweils angewandten Strategie floss auch das Antwortverhalten (z.B. der Doublecheck der Antworten, wie es ebenfalls einige der Teilnehmer rückgemeldet haben) in die Zeiterfassung mit ein, wobei das Antwortverhalten wiederum in Wechselwirkung mit anderen Variablen steht, wie z.B. der Selbstsicherheit bei räumlichen Aufgaben.

In der vorliegenden Studie konnte also gezeigt werden, dass eine zugrundeliegende Fähigkeit erfasst wird. Warum Männer in derselben Zeit doppelt so viele Aufgaben lösen konnten wie Frauen, sollte in weiteren Studien spezifiziert werden. Ist die gemessene Fähigkeit tatsächlich bei Frauen schwächer ausgeprägt, oder gibt es andere Variablen (z.B. Selbstvertrauen), die neben der Strategiewahl einen entscheidenden Einfluss auf die Leistung in Raumvorstellungstests haben könnten (Hirnstain et al., 2009; vgl. Hausmann 2007; Halpern & Tan, 2001).

Auffällig ist die große Bandbreite innerhalb der Stichprobe, was die Bearbeitungszeit (9-110 Minuten) betrifft. Einerseits könnten die im Absatz zuvor angesprochene „Unsicherheit“ und damit einhergehende Mehrfachüberprüfung anhand mehrerer Strategien mit zusätzlichen Bildvergleichen unter den vier Antwortalternativen eine Erklärung dafür bieten (vgl. Gittler & Arendasy, 2003). Dieses Verhalten wird sicher

begünstigt durch das Wegfallen einer Speed-Komponente, wie anekdotische Rückmeldungen von Teilnehmern nahe legen.

Die signifikant längere Bearbeitungszeit für oG-Aufgaben kann durch hohe Schwierigkeit erklärt werden ($\eta^2 = .355$). Ein Grund dafür könnte sein, dass durch die optische Verzerrung keine so klare Orientierung an den Längen von Ausgangsbild und Vergleichsbild möglich ist. Dieses Problem falle durch die Anzahl der Kästchen bei mG-Aufgaben weg.

Darauf deuten auch anekdotische Rückmeldungen von Frauen bei der letzten Zusatzfragen hin, da sie meinten, nur ein paar Kästchen „abgezählt“ zu haben um herauszufinden, welche Länge im Ausgangsbild welcher Länge im Vergleichsbild entsprach. Somit würde sich das Zählen der Kästchen eher auf einen Orientierungsaspekt beziehen, damit sie in weiterer Folge wüssten, welche Seite mit welcher durch ein ganzheitliches Drehen in Übereinstimmung gebracht werden musste, und nicht so sehr um die Anwendung einer analytisch-verbalen Bearbeitungsstrategie. Abgesehen davon wurde jedoch diese Herangehensweise zumeist verneint.

Da der verwendete Verrechnungsmodus nur einen Punkt für alle vier richtigen Antworten pro Aufgabe vorsah, wird nicht erkennbar, woran genau das Nicht-Erreichen der richtigen vier Antworten pro Aufgabe gelegen haben könnte. Würde man die Antworten pro „Bilderpaar“ erfassen, könnte man die Leistung getrennt nach rotierten und gespiegelten Ansichten prüfen.

Dieser Zugang könnte mehr Informationen darüber liefern, ob wirkliche Seitigkeitsentscheidungen die größere Rolle spielten oder doch eher die gespiegelten Ansichten für die längeren Beantwortungszeiten bzw. falschen Antworten verantwortlich sind. Denn wie Kerkman, Wise und Harwood (2000) vertreten, handelt es sich bei gespiegelten Aufgaben nicht um die Erfassung einer MR-Leistung. So könnte der geschlechtsspezifische Unterschied nur bei den „impossible figures“ (Kerkman et al., 2000) auftreten und nicht bei den Aufgaben, die echte mentale Rotation verlangen.

Außerdem wird eine Überprüfung einer analogen Verarbeitung, verstanden als ein linearer Zusammenhang von Zeit und Winkeldisparität - wie in der Studie von Shepard und Metzler (1971) für „Gleiche Paare“ gezeigt - nicht mit dem angewendeten Verrechnungsmodus (Zeit pro Aufgabe und nicht pro Bilderpaar) erfassbar und prüfbar, was auch schon Hooven et al. (2008) kritisierten.

Eine weitere interessante Frage steht im Zusammenhang mit dem Identifizieren einer gespiegelten Ansicht: Versucht sich der Proband die gespiegelte Ansicht vorzustellen? Dies würde mehr Zeit in Anspruch nehmen als wenn bloß bei einem Nicht-Gelingen einer Überführung der Ausgangsfigur in die Vergleichsfigur automatisch auf „gespiegelt“ geschlossen würde.

Ein möglicher Grund für den geringen Zusammenhang zwischen Selbsteinschätzung der räumlichen Orientierungsfähigkeit und den Leistungen im Test könnte, wie einzelne Testpersonen rückmeldeten, daran liegen, dass sie beim Orientieren im Raum ganz andere gedankliche Strategien verwendeten, als sie dies beim Bearbeiten der Testaufgaben gemacht hatten. So meinte eine Testperson, bei Orientierung im Raum stelle sie sich gedanklich gleichsam in den Raum und gehe die „Straßen“ entlang. Ein anderer meinte, dass das Planlesen zweidimensional sei und die Karte gedreht werden könne, was bei den PC-Aufgaben nicht möglich war. Doch erst weitere Validierungsstudien zu den hier vorgelegten Testaufgaben werden mehr Einsicht in die zugrundeliegenden kognitiven Operationen liefern.

10. Abstract-Deutsch

Der Leistungsunterschied in mentaler Rotationsfähigkeit zu Gunsten der Männer ist ein beständiges Ergebnis der Raumvorstellungsforschung (Voyer, Voyer & Bryden, 1995). Als eine mögliche Ursache wird die unterschiedliche Strategiewahl (Glück & Fitting, 2003) bei Raumvorstellungsaufgaben angenommen.

Zwei verschiedene Aufgabenmodi („ohne-Gitter“ vs. „mit-Gitter“-Aufgaben) wurden anhand des Rasch-Modells auf ihre Eindimensionalität geprüft und mittels Zusatzfragen wurden Informationen über Bearbeitungsstrategien und qualitative Unterschiede erhoben.

Die Stichprobe umfasste 199 Versuchspersonen. Der Test umfasste 7 „ohne-Gitter“- und 17 „mit-Gitter“-Aufgaben. Jede einzelne Aufgabe bestand aus der Darstellung einer Ausgangsfigur und vier Antwortfiguren, für die angegeben werden musste, ob es sich um eine rotierte oder gespiegelte Ansicht der Ausgangsfigur handelte. Nach Darstellung der Geltung des Rasch-Modells für die vorliegende Untersuchung, zeigten sich signifikante Leistungsunterschiede ($d = 0.89$) zu Gunsten der Männer und „mit-Gitter“-Aufgaben stellten sich in der Bearbeitung als signifikant leichter heraus ($\eta^2 = .287$).

Die Zusatzfragen ergaben, dass die beiden Aufgabenmodi signifikant unterschiedlich erlebt wurden. Als primäre Bearbeitungsstrategie kann eine visuell-räumliche Leistung über beide Aufgabenmodi angenommen werden.

Key Words: Mentale Rotation, Strategiewahl, Geschlechtsunterschied, Rasch-Modell

11. Abstract-Englisch

Sex differences in mental rotation ability in favour of men are a steady result in spatial ability research (Voyer, Voyer & Bryden, 1995). The varying choice of strategy in spatial tasks is discussed as one possible reason (Glück & Fitting, 2003).

Two different types of tasks (“without-grid” vs. “with-grid”) were analysed based on the Rasch-Model as to their one-dimensionality. Additional questions generated information on solution strategies und qualitative differences.

The sample consisted of 199 subjects. The test was composed of 7 “without-grid”- and 17 “with-grid”-items. Each item consisted of one target image and four comparison images. For each of them the task was to decide whether the comparison image was rotated or mirrored in relation to the target image. After demonstrating the validity of the Rasch-Model for the presented analysis, the results showed significant differences in ability in favor of men ($d = 0.89$) and “with-grid”-items proved significantly easier to solve ($\eta^2 = .287$).

The additional questions showed significant differences between the two kinds of tasks concerning the manner in which they were answered. Regarding the primary strategy of processing the performance of a visual-spatial approach can be assumed in both task categories.

Key words: Mental Rotation, Choice of Strategy, Sex difference, Rasch-Model

12. Literaturverzeichnis

- Arendasy, M., & Gittler, G. (2003). IRT-basierter Vergleich zweier Varianten automatisiert erstellter Matrizentestaufgaben. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24(4), 261-275.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Auflage). Berlin: Springer.
- Brugger, C. (unpublished). Teach and Test (TNT). Vienna
- Fischer, G. H., & Ponocny-Seliger, E. (1998). *Structural Rasch modeling. Handbook of the usage of LpcM-WiN 1.0*. Groningen: PROGRAMMA.
- Formann, A. K. (1981). Über die Verwendung von Items als Teilungskriterium für Modellkontrollen im Modell von Rasch. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 28, 541-560.
- Formann, A. K. (2004). *Testtheorie und Testkonstruktion*. Institut für Psychologie der Universität Wien. Wien: Facultas Verlags- und Buchhandels AG.
- Gittler, G. (1984). Entwicklung und Erprobung eines neuen Testinstruments zur Messung des räumlichen Vorstellungsvermögens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 5, 141-165.
- Gittler, G. (1986). Inhaltliche Aspekte bei der Itemselektion nach dem Modell von Rasch. *Zeitschrift für experimentelle und angewandte Psychologie*, 23(3), 386-412.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest. Ein Rasch-skaliertes Test zur Messung des räumlichen Vorstellungsvermögens. Theoretische Grundlagen und Manual*. Weinheim: Beltz.
- Gittler, G. (1992). *Testpsychologische Aspekte der Raumvorstellungsforschung- Kritik, Lösungsansätze und empirische Befunde* (Unveröffentlichte Habilitationsschrift). Universität Wien, Wien.

- Gittler, G. (1999). Sind Raumvorstellung und Reasoning separierbare Fähigkeitsdimensionen? Dimensionalitätsanalysen zweier Rasch-skalierten Tests: 3DW und WMT. *Diagnostica*, 45(2), 69-81.
- Gittler, G., & Arendasy, M. (2003). Endlosschleifen: Psychometrische Grundlagen des Aufgabentyps E^P. *Diagnostica*, 49, 164-175.
- Gittler, G., & Glück, J. (1998). Differential transfer of learning: Effects of instruction in Descriptive Geometry on spatial test performance. *Journal for Geometry and Graphics*, 2, 71-84.
- Glück, J. (1999). *Spatialstrategies: Kognitive Strategien bei Raumvorstellungsleistungen* (Unveröffentlichte Dissertation). Universität Wien, Wien.
- Glück, J. (2001, September). Strategieaspekte: Ein unterschätzter Faktor in der Raumvorstellungsforschung? [Abstract]. Vortrag auf der Fachgruppentagung Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Leipzig, Deutschland. Abgerufen auf http://www.dgps.de/fachgruppen/diff_psy/Abstracts/Abstracts/glueck.htm
- Glück, J., & Fabrizii, C. (2010). Gender differences in the Mental Rotation Test are partly explained by response format. *Journal of Individual Differences*, 31, 106-109.
- Glück, J., & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing*, 3, 293-308.
- Glück, J., Kaufmann, H., Dünser, A., & Steinbügl, K. (2005): "Geometrie und Raumvorstellung - Psychologische Perspektiven"; Informationsblätter der Geometrie (IBDG), 24(1); 4-11. Abgerufen auf <https://www.ims.tuwien.ac.at/publications/tuw-139926>
- Glück, J., Machat, R., Jirasko, M., & Rollett, B. (2002). Training- related changes in solution strategy in a spatial test: An application of item response models. *Learning and Individual Differences*, 13, 1-22.
- Grüßing, M., (2002). Wieviel Raumvorstellung braucht man für Raumvorstellungsaufgaben? Strategien von Grundschulkindern bei der

- Bewältigung räumlich-geometrischer Anforderungen. *Zentralblatt für Didaktik der Mathematik* 34 (2), 37-45. Abgerufen auf: <http://subs.emis.de/journals/ZDM/zdm022a1.pdf>
- Halpern, D. F., & Tan, U. (2001). Stereotypes and Steroids: Using a Psychobiosocial Model to Understand Cognitive Sex Differences. *Brain and Cognition*, 45, 392-414.
- Hausmann, M. (2007). Kognitive Geschlechtsunterschiede. In S. Lautenbacher, O. Güntürkün & M. Hausmann (Hrsg.), *Gehirn und Geschlecht. Neurowissenschaft des kleinen Unterschieds zwischen Frau und Mann* (S. 106-121). Heidelberg: Springer Science. Abgerufen auf: http://bilder.buecher.de/zusatz/22/22624/22624493 lese_1.pdf
- Heider, C. (2010): *SPACE 2d/3d. Entwicklung eines Rasch-homogenen Raumvorstellungstest* (Unveröffentlichte Diplomarbeit). Universität Wien, Wien. Abgerufen auf <http://othes.univie.ac.at/14547/>
- Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *The Quarterly Journal of Experimental Psychology*, 61, 683-689.
- Hirnstien, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, 19, 225-228.
- Hirnstien, M., & Hausmann, M. (2010). Neuropsychologie. In G. Stein (Hrsg.), *Handbuch Psychologie und Geschlechterforschung* (S. 69-85). Wiesbaden: VS Verlag für Sozialwissenschaften. Abgerufen auf: http://download.springer.com/static/pdf/781/chp%253A10.1007%252F978-3-531-92180-8_5.pdf?auth66=1385973723_11d3a1dd1c512e023b894582e9ef76f0&ext=.pdf
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70. Abgerufen auf: <http://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf>

- Hooven, C. K., Chabris, C. F., Ellison, P. T., Kievit, R. A., & Kosslyn, S. M. (2008). The sex difference on mental rotation tests is not necessarily a difference in mental rotation ability. Manuscript submitted for publication. Abgerufen auf: www.wjh.harvard.edu/~cfc/Hooven2008.pdf
- Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, 92, 137-172.
- Kaufmann, H., Csisinko, M., Strasser, I., Strauß, S., Koller, I. & Glück, J. (2008, August): Design of a Virtual Reality Supported Test for Spatial Abilities. In *Proceedings of the 13th International Conference on Geometry and Graphics (ICGG)*, Dresden, Germany. Abgerufen auf http://publik.tuwien.ac.at/files/PubDat_170660.pdf
- Kerkman, D. D., Wise, J. C., & Harwood, A. E. (2000). Impossible `mental rotation` problems: A mismeasure of women`s spatial abilities? *Learning and Individual Differences*, 12, 253-269.
- Khooshabeh, P., Hegarty, M. (2010). Representations of shape during mental rotation. In T. Barkowsky, S. Bertel, C. Hölscher, T. F. Shipley (Hrsg.), *Cognitive Shape Processing: Papers from the AAAI Spring Symposium, 2010* (S. 15-20). Menlo Park, California: AAAI. Abgerufen auf: <http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/view/1166/1377>
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Lautenbacher, S., Güntürkün, O., & Hausmann, M. (2007). *Gehirn und Geschlecht. Neurowissenschaft des kleinen Unterschieds zwischen Frau und Mann*. Heidelberg: Springer Science.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development*, 56(6), 1479-1498.
- Maier, P. H. (1999). *Räumliches Vorstellungsvermögen*. Donauwörth: Auer Verlag.

- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, 57, 176-184.
- Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, 66, 260-264.
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, 28, 39-58.
- Putz-Osterloh, W. (1977): Über Problemlöseprozesse bei dem Test Würfelaufgaben aus dem Intelligenzstrukturtest IST und IST-70 von Amthauer. *Diagnostika*, 23, 252-265.
- Putz-Osterloh, W., & Lüer, G. (1975). Informationsverarbeitung bei einem Test zur Erfassung der Raumvorstellung. *Diagnostika*, 21, 166-181.
- Putz-Osterloh, W., & Lüer, G. (1979): Wann produzieren Probanden räumliche Vorstellungen beim Lösen von Raumvorstellungsaufgaben? *Zeitschrift für experimentelle und angewandte Psychologie*, 26, 138-156.
- Quaiser-Pohl, C., Lehmann, W., Jordan, K., & Schirra, J. (2001). Zum Einfluss des Computers auf die Raumvorstellung - eine differenzielle Analyse bei Studierenden von Computerwissenschaften [Abstract]. In B. B. Seiwald, J. Guthke, H. Petermann, J. F. Beckmann, & M. Roth (Hrsg.), 6. *Arbeitstagung der Fachgruppe für Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik der Deutschen Gesellschaft für Psychologie: Abstractband und Programmheft* (S. 111). Leipzig: Leipziger Universitätsverlag. Abgerufen auf: http://www.dgps.de/fachgruppen/diff_psy/Abstracts/Abstracts/quaiser.htm
- Quaiser-Pohl, C., Rohe, A., & Amberger, T. (2010). The solution strategy as an indicator of the developmental stage of pre-school children's mental-rotation ability. *Journal of Individual Differences*, 31, 95-100.
- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion* (2. Auflage) Bern: Hans Huber.

- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Stelzl, I. (1979). Ist der Modelltest des Rasch-Modells geeignet, Homogenitätshypothesen zu prüfen? Ein Bericht über Simulationsstudien mit inhomogenen Daten. *Zeitschrift für experimentelle und angewandte Psychologie*, 16 (4), 652-672.
- Strauss, S., Strasser, I., Csisinko, M., Kaufmann, H., & Glück, J. (2007, September). *Entwicklung eines 3d- Raumvorstellungs- Lerntests mit Augmented Reality*. Posterpräsentation auf der Tagung der Fachgruppe Differenzielle Psychologie in der DGPs, Wien. Abgerufen auf: <https://www.ims.tuwien.ac.at/publications/tuw-141221.pdf>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250-270.
- Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. *Perception and Psychophysics*, 31(3), 201-209.

13. Abbildungsverzeichnis

ABBILDUNG 1: PAARWEISE DARSTELLUNG DER GEOMETRISCHEN FIGUREN BEI SHEPARD & METZLER (1971). A: STELLT EIN "SAME-PAIR" ROTIERT IN DER BILDEBENE DAR; B: "SAME-PAIR" ROTIERT IN DER BILDTIEFE; C: GESPIEGELTE UND ROTIERTE ANSICHT IN DER BILDTIEFE (SHEPARD & METZLER, 1971).	10
ABBILDUNG 2: BEISPIELITEMS DES MRT (VANDENBERG & KUSE, 1978): IN DER ERSTEN REIHE SIND FIGUR 1 UND 4 IDENT MIT DER AUSGANGSFIGUR, 2 UND 3 SIND DISTRAKTOREN UND STELLEN GESPIEGELTE ANSICHTEN DER AUSGANGSFIGUR DAR. IN DER ZWEITEN REIHE SIND DIE 2. UND 3. FIGUR DIE RICHTIGEN ALTERNATIVEN. DIE ANDEREN ZWEI FIGUREN SIND DARSTELLUNGEN EINES ANDEREN AUSGANGSBILDES.	12
ABBILDUNG 3: ÜBUNGSBEISPIEL AUS DEM 3DW (GITTLER, 1990): RICHTIGE ANTWORT IST D (GITTLER & GLÜCK, 1998, S. 73).	17
ABBILDUNG 4: BEISPIELE FÜR ZWEIDIMENSIONALE FIGURENDARSTELLUNGEN IN DEN VARIANTEN „OHNE“- UND „MIT- GITTERNETZ“ JE IN DER AUSGANGSPOSITION UND IN DER GESPIEGELTEN ANSICHT. UND BEISPIELE FÜR EINE DREIDIMENSIONALE, „MIT-GITTERNETZ“-DARSTELLUNG - EBENFALLS IN DER AUSGANGSPOSITION UND IN DER GESPIEGELTEN ANSICHT.	36
ABBILDUNG 5: BEISPIELAUFGABE (OG-VERSION).	37
ABBILDUNG 6: ALTERSVERTEILUNG DER GESAMTSTICHPROBE.	43
ABBILDUNG 7: HÄUFIGKEITEN DER BILDUNGSSTUFEN.	44
ABBILDUNG 8: INSTRUKTION UND ERSTES FRAGENPAAR DER ZUSATZFRAGEN	48
ABBILDUNG 9: GRAPHISCHER MODELLTEST (TEILUNGSKRITERIUM: MEDIAN).	54
ABBILDUNG 10: ZUSAMMENHANG ZWISCHEN TESTWERTEN UND PERSONENPARAMETERN IM GESAMTTEST	58

ABBILDUNG 11: DARSTELLUNG DER KLASSISCHEN LÖSUNGSWAHRSCHEINLICHKEITEN FÜR JEDES TESTITEM IM GESAMTTEST (0-ITEM IST WARM-UP-ITEM; OG-ITEM 0-7, MG- AUFGABEN ITEM 8-24).....	62
ABBILDUNG 12: DARSTELLUNG DER KLASSISCHEN LÖSUNGSWAHRSCHEINLICHKEITEN JE TESTITEM IM GESAMTTEST, GETRENNT NACH GESCHLECHT.....	63
ABBILDUNG 13: BIVARIATES STREUDIAGRAMM ZUR DARSTELLUNG DES ZUSAMMENHANGS ZWISCHEN DEN PERSONENPARAMETERN DER „OHNE-GITTER“- VS. „MIT-GITTER“-AUFGABEN, GETRENNT NACH GESCHLECHT.	63
ABBILDUNG 14: INTERAKTIONSDIAGRAMM IN ABHÄNGIGKEIT VON AUFGABENMODUS UND GESCHLECHT.	66
ABBILDUNG 15: HISTOGRAMM ZUR BEARBEITUNGSZEIT IM GESAMTTEST (IN MINUTEN).....	69
ABBILDUNG 16: BEARBEITUNGSZEIT IN ABHÄNGIGKEIT VON AUFGABENMODUS UND GESCHLECHT.	71
ABBILDUNG 17: BIVARIATES STREUDIAGRAMM ZWISCHEN LEISTUNG (PARAMETERNIVEAU) UND BEARBEITUNGSZEIT (IN MINUTEN) IM GESAMTTEST.	71
ABBILDUNG 18: BIVARIATES STREUDIAGRAMM FÜR DEN ZUSAMMENHANG ZWISCHEN PERSONENPARAMETER IM GESAMTTEST UND DER SELBSTEINSCHÄTZUNG "ORIENTIERUNGSFÄHIGKEIT IM RAUM" (ANALOGSKALA 0-100).	73
ABBILDUNG 19: SCREEPLOT: GRAPHISCHE DARSTELLUNG DES EIGENWERTEVERLAUFS	78
ABBILDUNG 20: ZUSAMMENHANG ZWISCHEN DEN OG- UND MG-INDIZES.	79

14. Tabellenverzeichnis

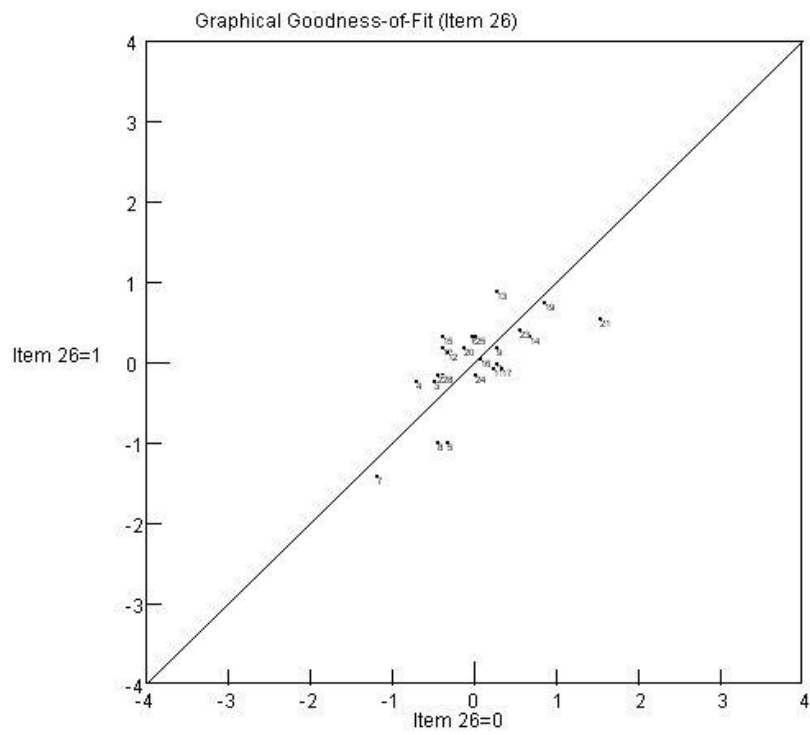
TABELLE 1: DARSTELLUNG DES VERSUCHSPLANS FÜR DIE VORLIEGENDE UNTERSUCHUNG UND DER BEZUG ZUM VERSUCHSPLAN VON HEIDER (2010).	50
TABELLE 2: KLASSISCHE TEILUNGSKRITERIEN.	54
TABELLE 3: MODELLGELTUNGSKONTROLLEN (LQT) FÜR DAS RM NACH MARTIN-LÖF (N = 191).	55
TABELLE 4: LRT NACH ANDERSEN FÜR ALLE ITEMS ALS TEILUNGSKRITERIUM.	56
TABELLE 5: ANZAHL DER GELÖSTEN ITEMS MIT IHREN ENTSPRECHENDEN PERSONENPARAMETERN (THETA Θ).....	57
TABELLE 6: DESKRIPTIVSTATISTIK DER TESTWERTE FÜR DIE GESAMTSTICHPROBE (N = 199) IM GESAMTTEST UND GETRENNT FÜR OG- UND MG-AUFGABEN IN ABHÄNGIGKEIT VOM GESCHLECHT.	58
TABELLE 7: DESKRIPTIVSTATISTIK DER PERSONENPARAMETER (Θ) IM GESAMTTEST UND GETRENNT FÜR OG- UND MG-AUFGABEN IN ABHÄNGIGKEIT VOM GESCHLECHT.....	59
TABELLE 8: INNERE KONSISTENZEN DER BEIDEN AUFGABENMODI (N = 199).	60
TABELLE 9: TESTITEMS UND IHRE SCHWIERIGKEITSINDIZES (MITTELWERT).....	61
TABELLE 10: DESKRIPTIVSTATISTIK DER DURCHSCHNITTlich ERREICHTEN LEISTUNGEN IN DEN BEIDEN AUFGABENMODALITÄTEN (GESAMTSTICHPROBE).....	61
TABELLE 11: PERSONENPARAMETER IN ABHÄNGIGKEIT DER ANZAHL RICHTIG GELÖSTER AUFGABEN (GESAMTTEST, NUR „MIT-GITTER“- AUFGABEN, NUR „OHNE-GITTER“-AUFGABEN).....	64
TABELLE 12: PERSONENPARAMETER IN ABHÄNGIGKEIT VON AUFGABENMODUS UND GESCHLECHT.....	65
TABELLE 13: LÖSUNGSWAHRSCHEINLICHKEIT DER LINK – ITEMS IN ABHÄNGIGKEIT VON AUFGABENMODUS UND GESCHLECHT.....	67

TABELLE 14: GESAMTBEARBEITUNGSZEIT (IN MINUTEN) FÜR DEN GESAMTTEST UND JE TESTTEIL (OG.- VS. MG-AUFGABEN) GETRENNT NACH GESCHLECHT.....	70
TABELLE 15: DESKRIPTIVSTATISTIK ZUM SELBSTRATING MITTELS ANALOGSKALA (0) <i>SEHR SCHLECHT</i> BIS (100) <i>SEHR GUT</i> DER EIGENEN ORIENTIERUNGSFÄHIGKEIT IM RAUM.	72
TABELLE 16: DESKRIPTIVSTATISTIK DES RATINGS DER ZUSATZFRAGENPAARE GETRENNT NACH GESCHLECHT (SKALA: 4 <i>TRIFFT SEHR ZU</i> BIS 1 <i>TRIFFT GAR NICHT ZU</i>).	75
TABELLE 17: DESKRIPTIVSTATISTIKEN DER BEIDEN EXTRAHIERTEN FAKTOREN.	77
TABELLE 18: LADUNGSMATRIX - ERGEBNISSE DES ROTATIONSVERFAHRENS NACH DER VARIMAX-METHODE – EXTRAKTION ZWEIER FAKTOREN.	77

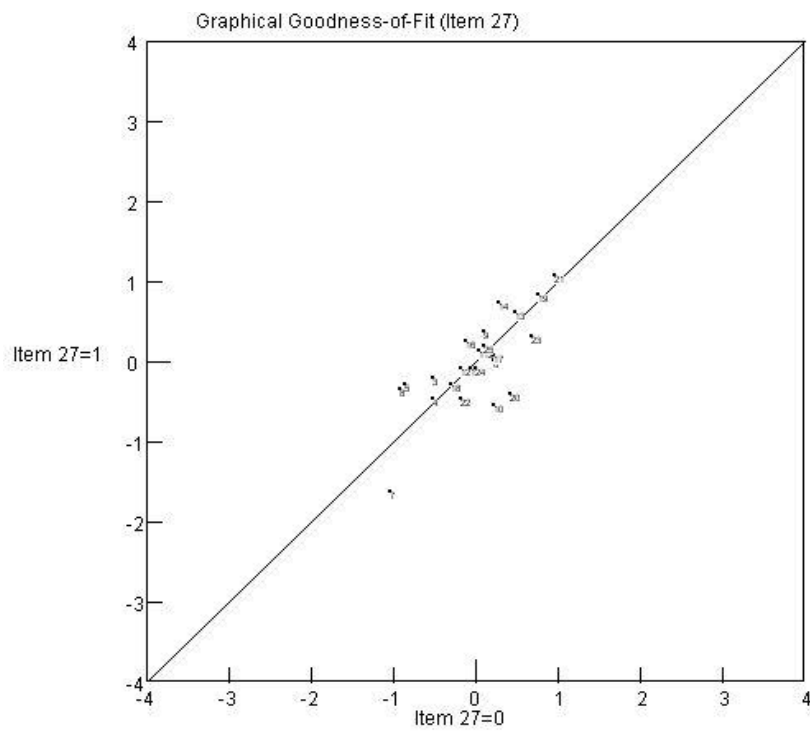
15. ANHANG

Anhang A: Graphischer Modelltest

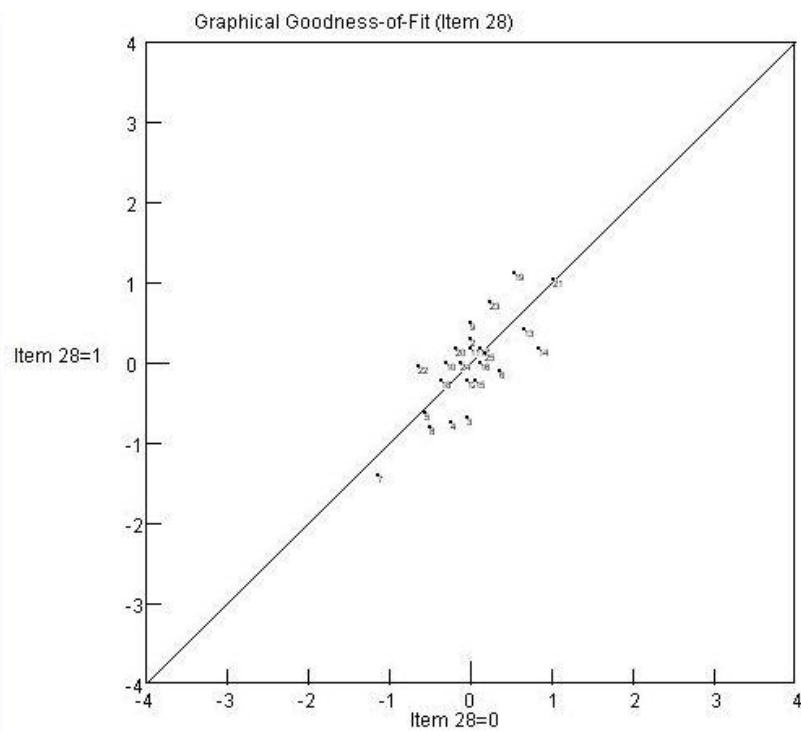
Teilungskriterium: Alter



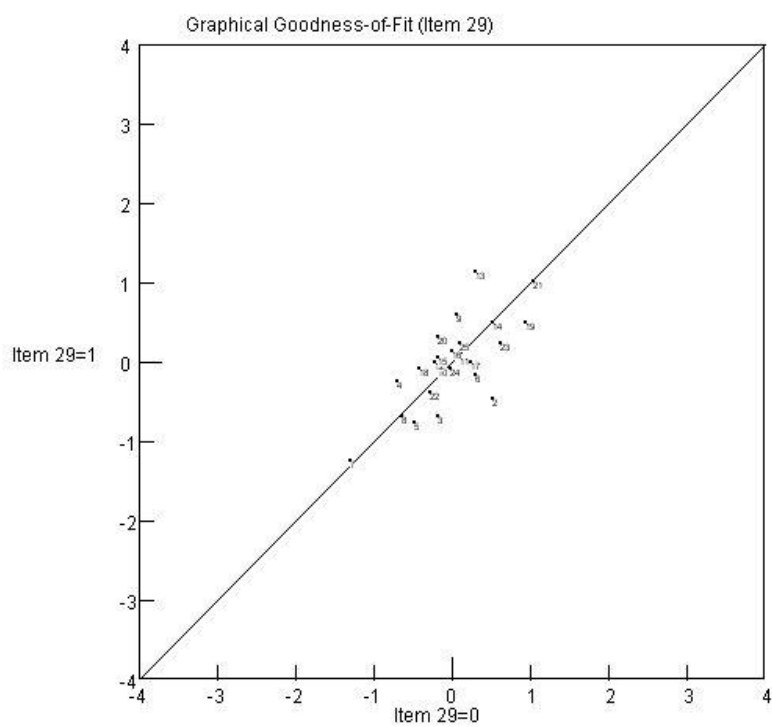
Teilungskriterium: Selbsteinschätzung „Orientierungsfähigkeit“



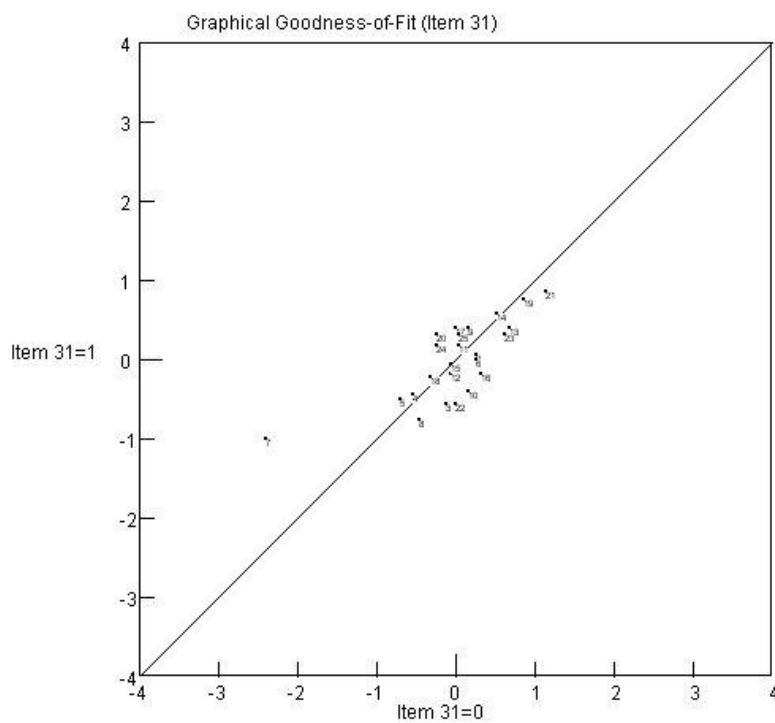
Teilungskriterium: Bearbeitungszeit



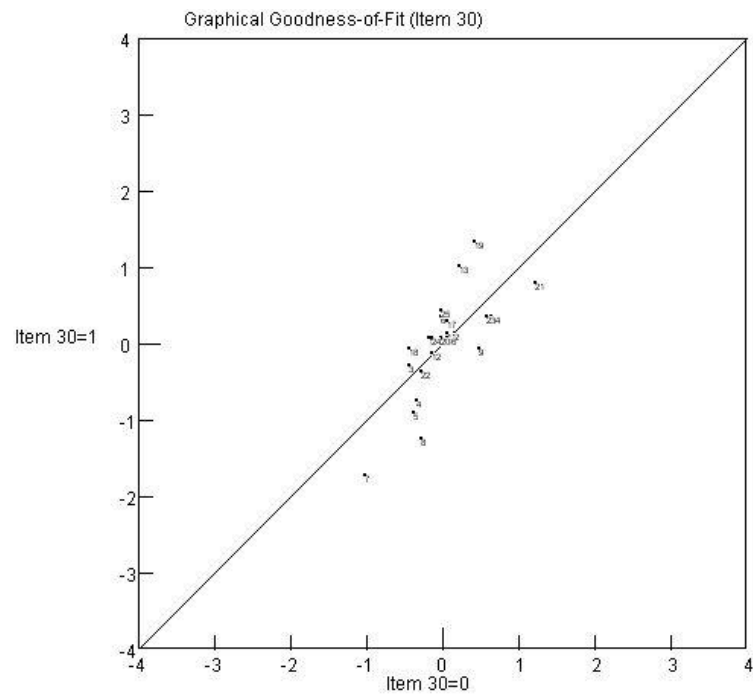
Teilungskriterium: Geschlecht



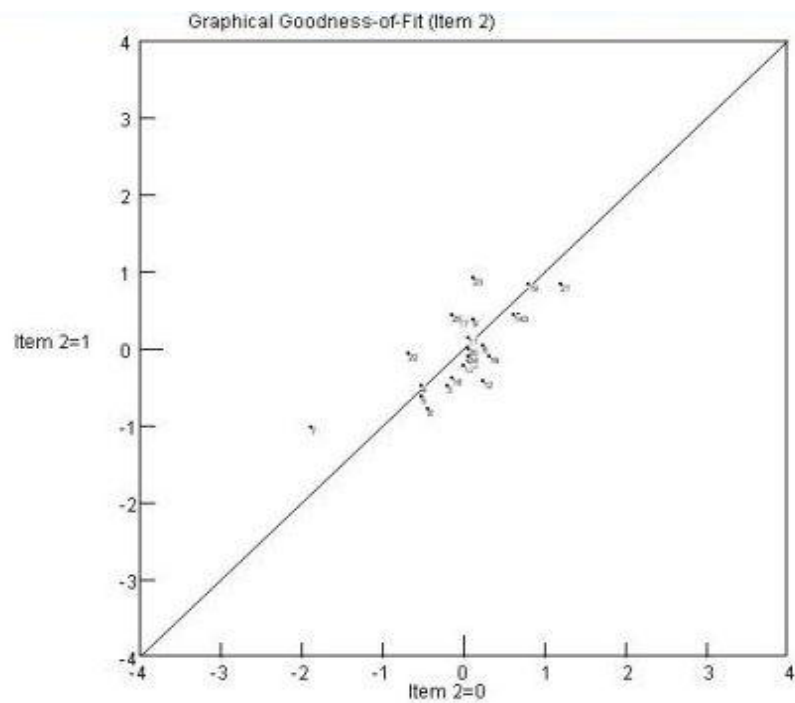
Teilungskriterium: Median (Gesamttest)



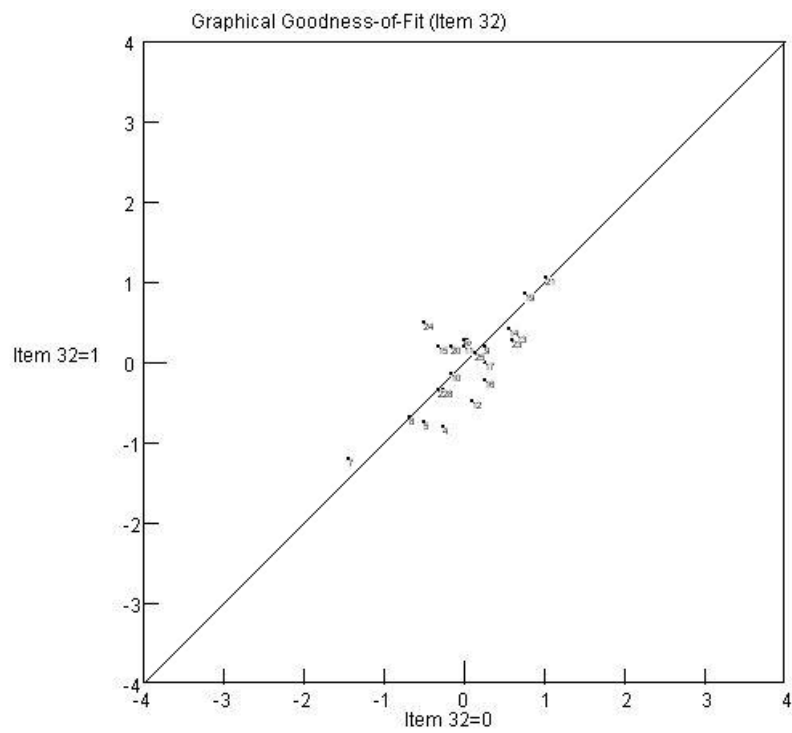
Teilungskriterium: Bildung



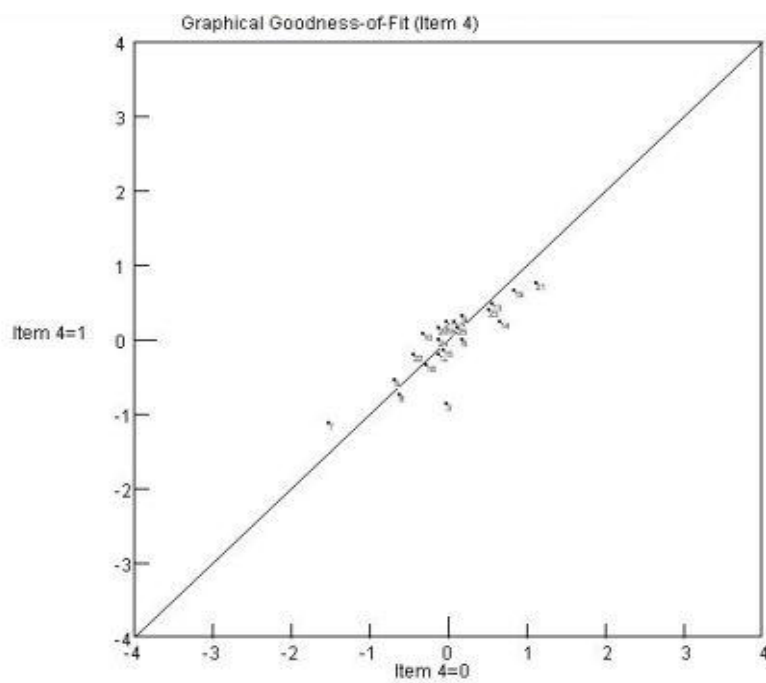
Teilungskriterium: Item 2

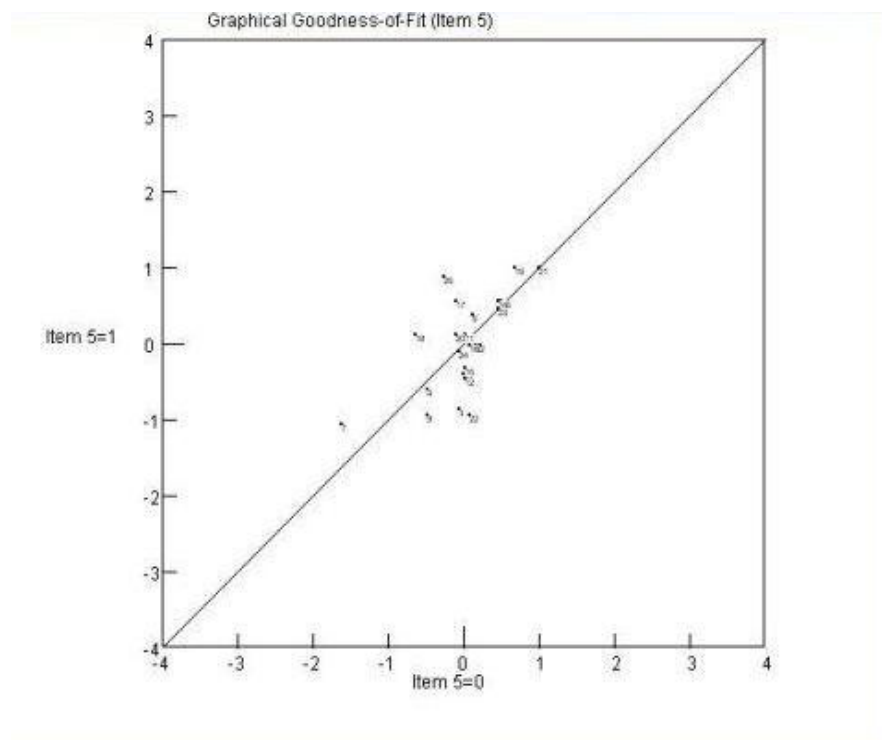
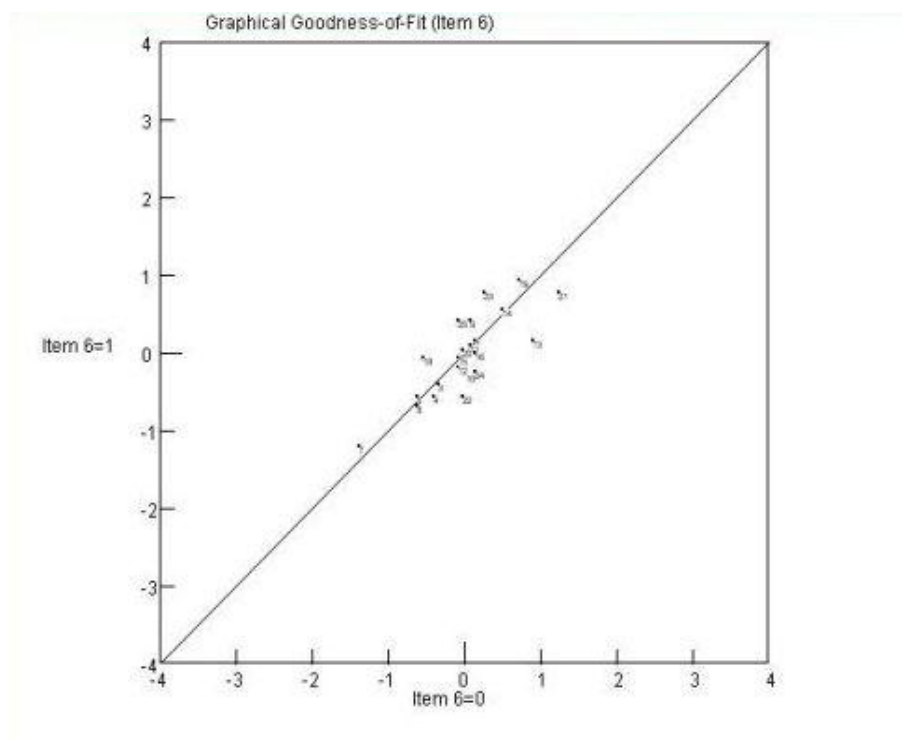


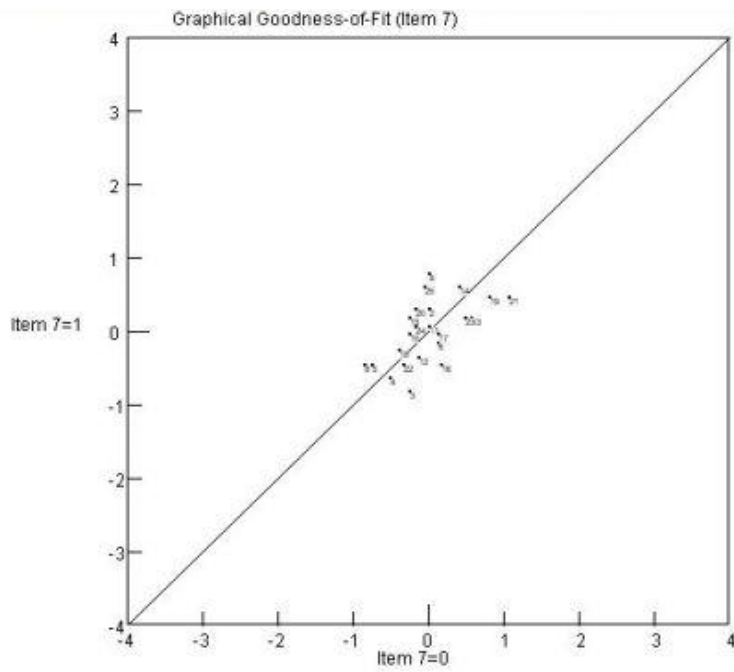
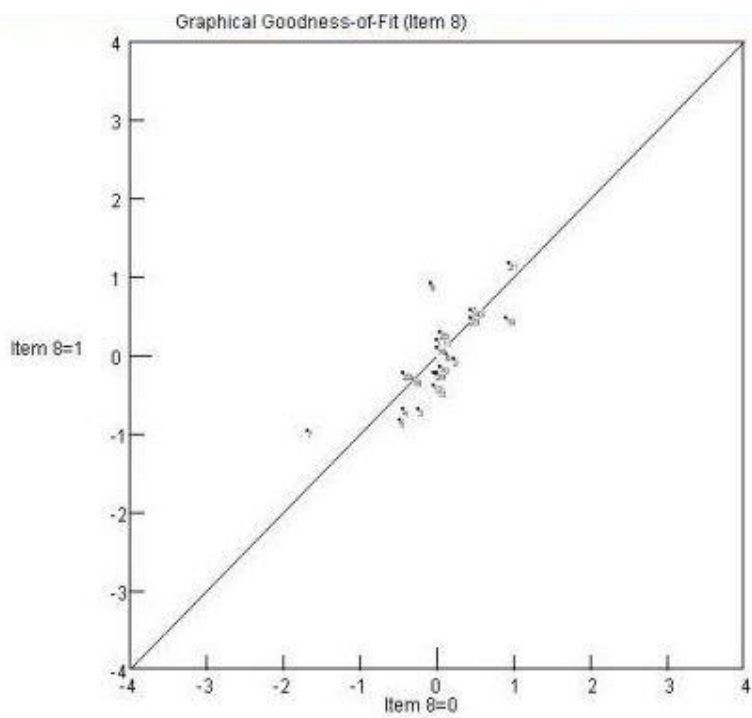
Teilungskriterium: Item 3

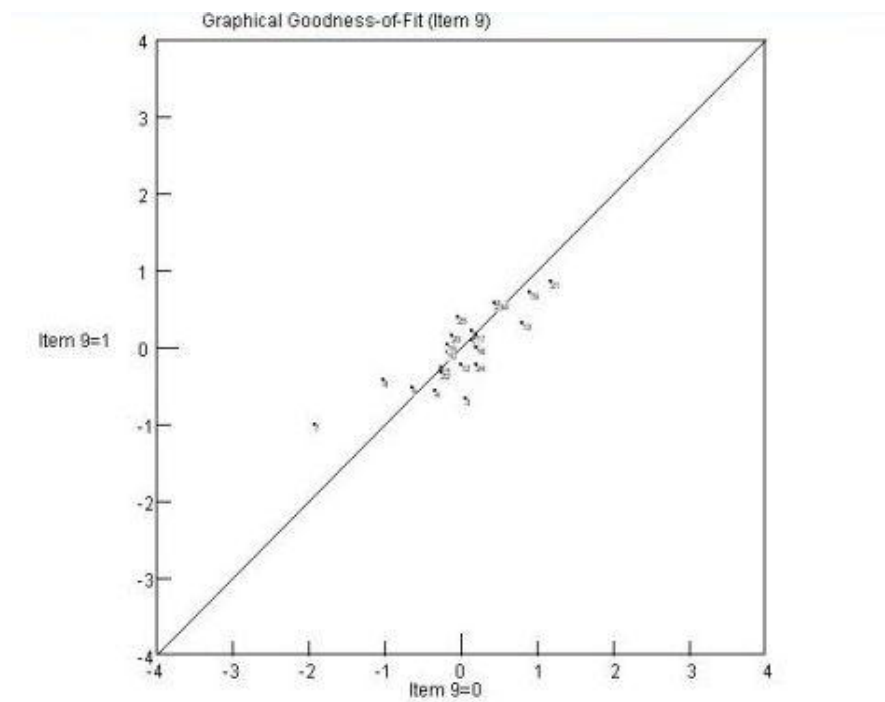
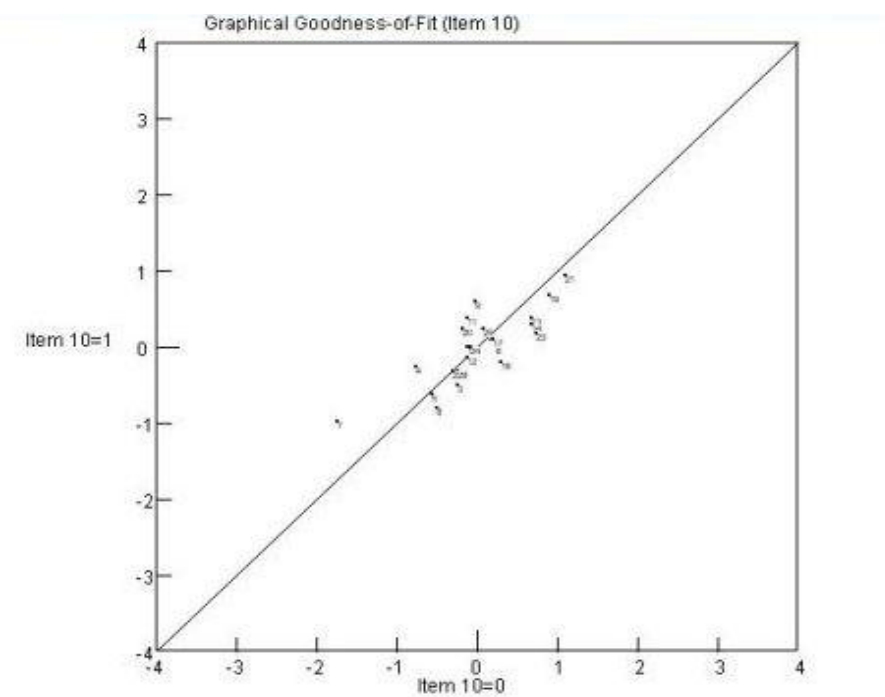


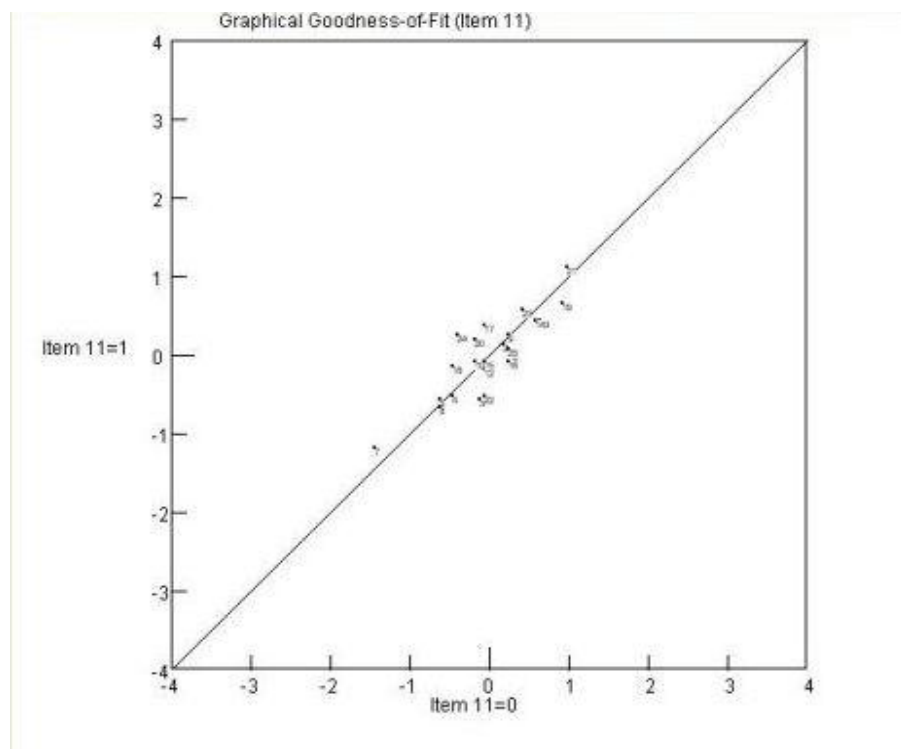
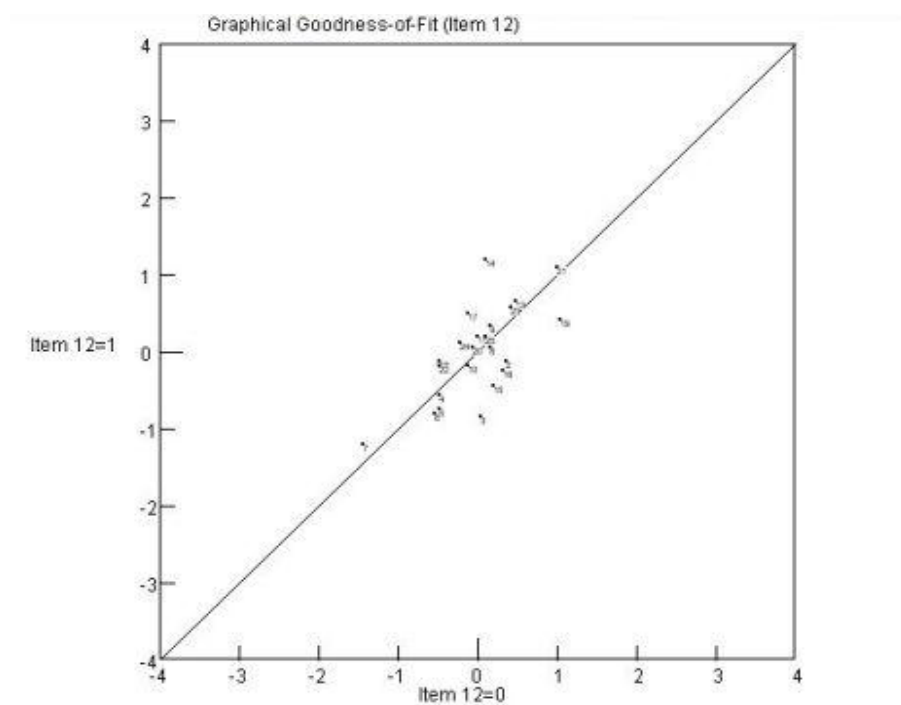
Teilungskriterium: Item 4

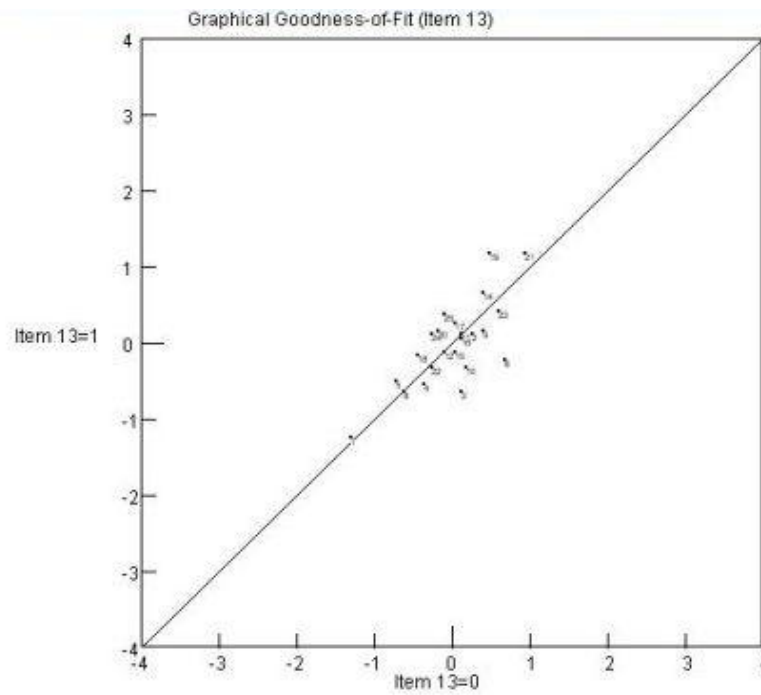
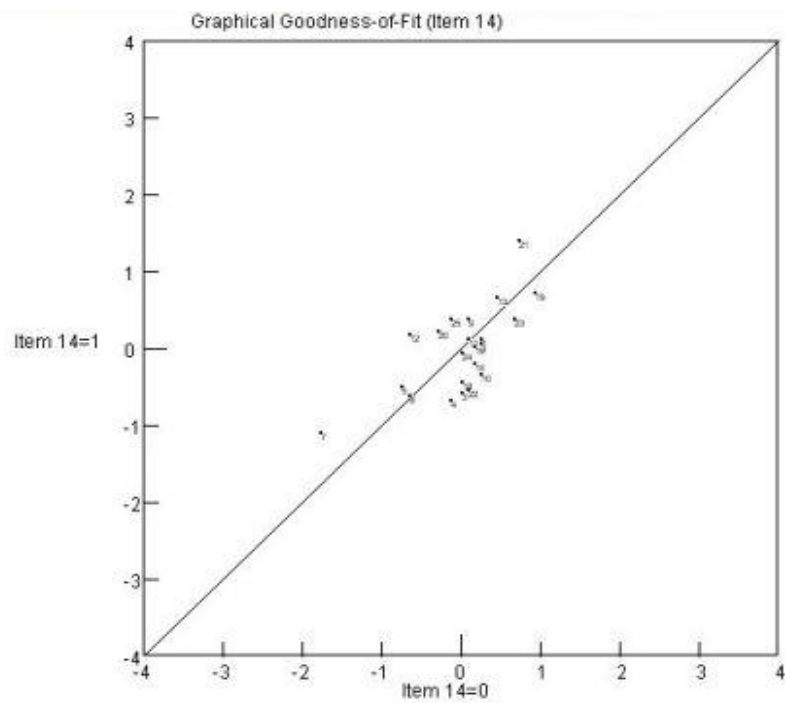


Teilungskriterium: Item 5**Teilungskriterium: Item 6**

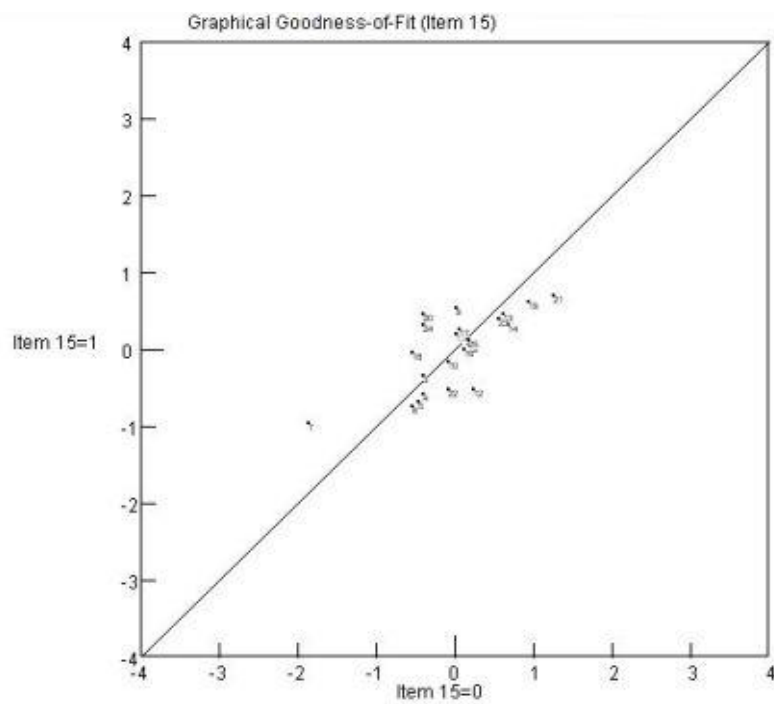
Teilungskriterium: Item 7**Teilungskriterium: Item 9**

Teilungskriterium: Item 8**Teilungskriterium: Item 10**

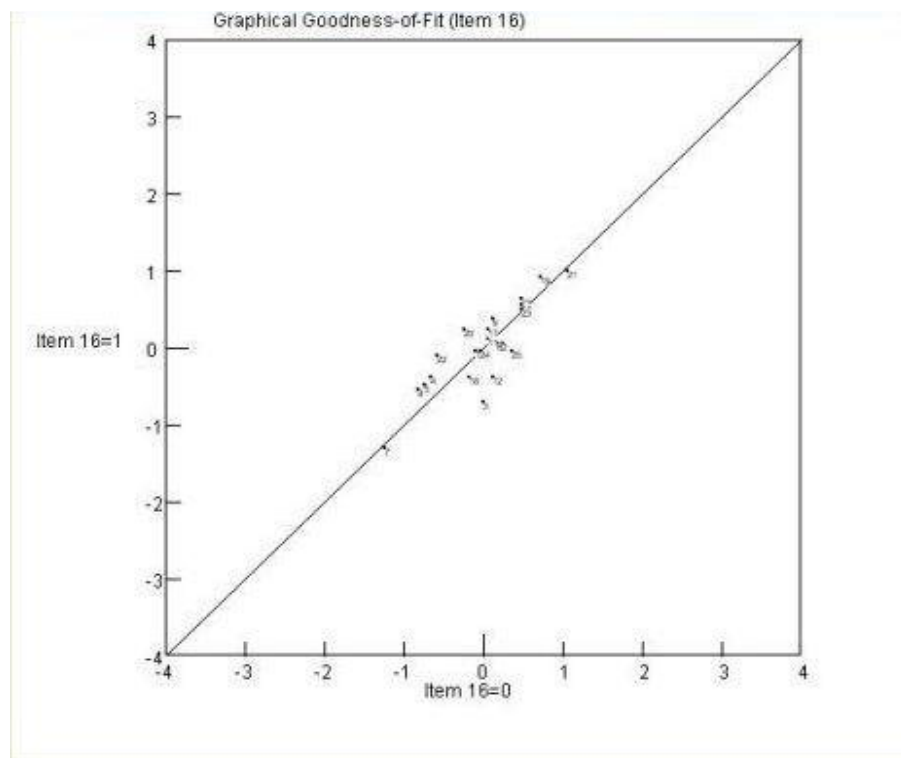
Teilungskriterium: Item 11**Teilungskriterium: Item 12**

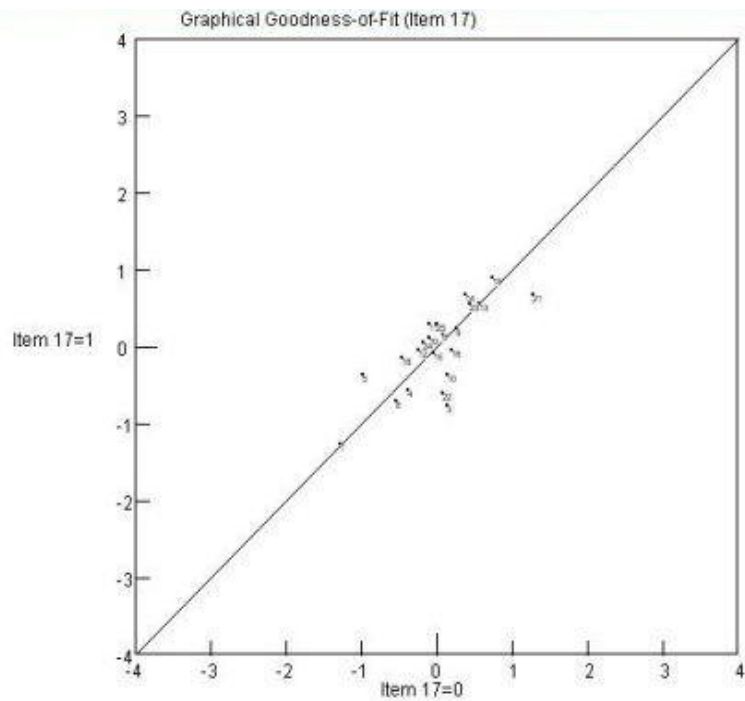
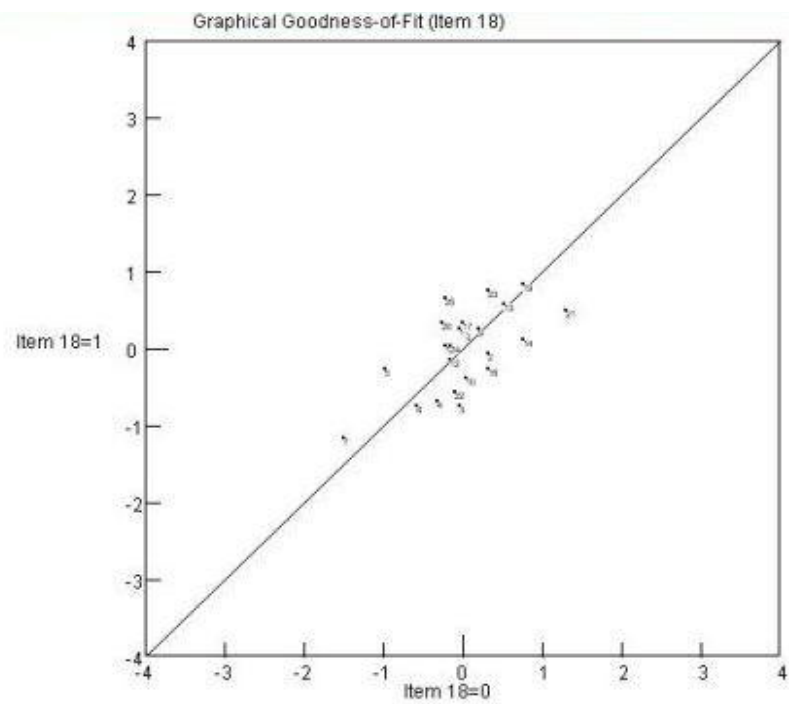
Teilungskriterium: Item 13**Teilungskriterium: Item 14**

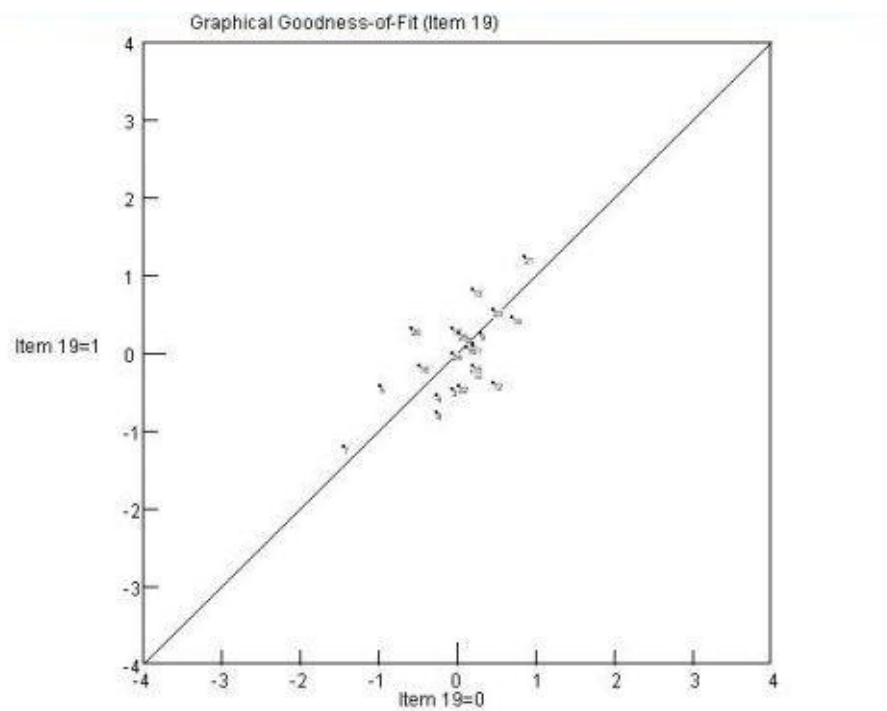
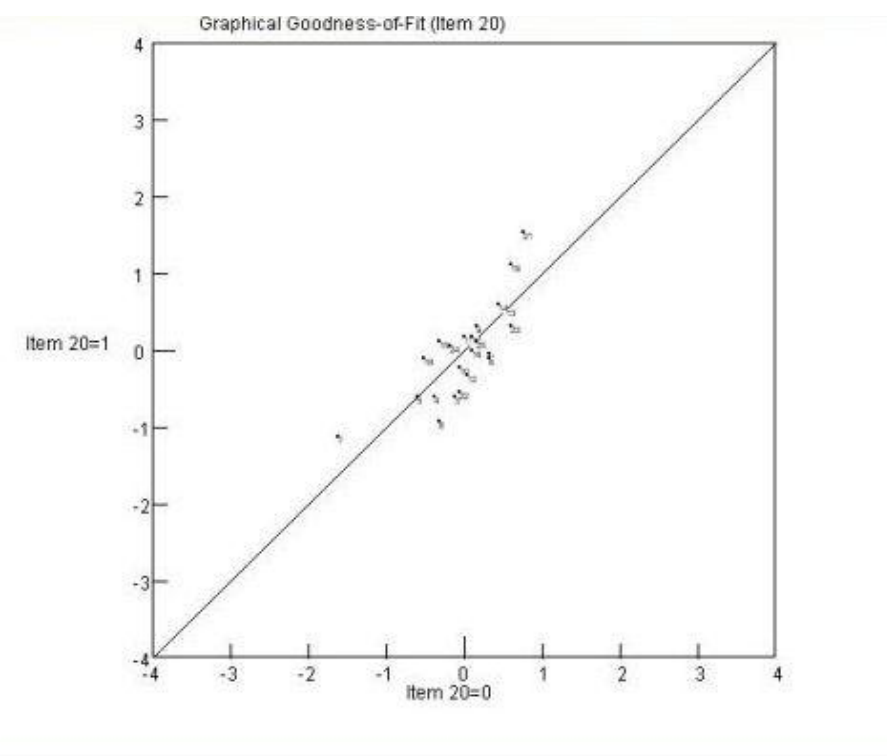
Teilungskriterium: Item 15

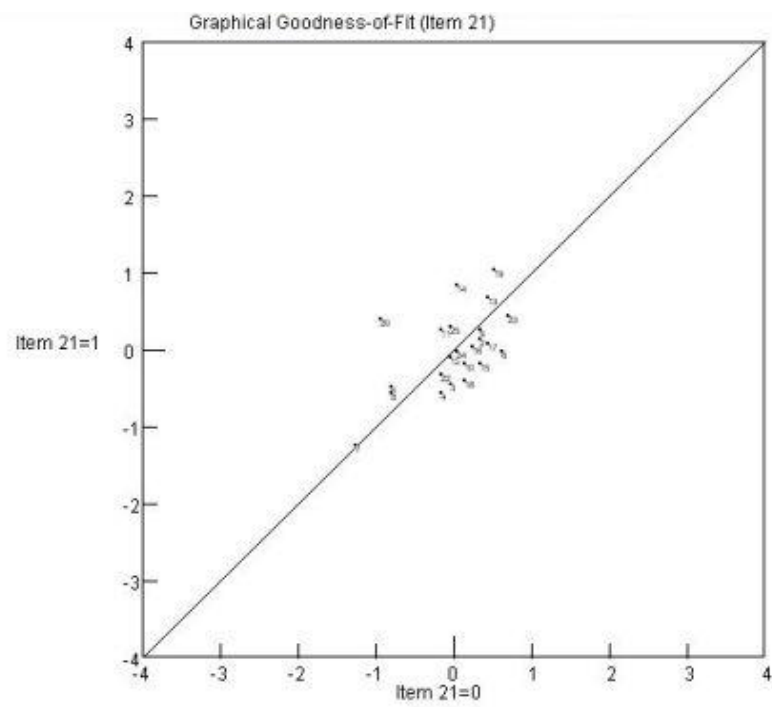
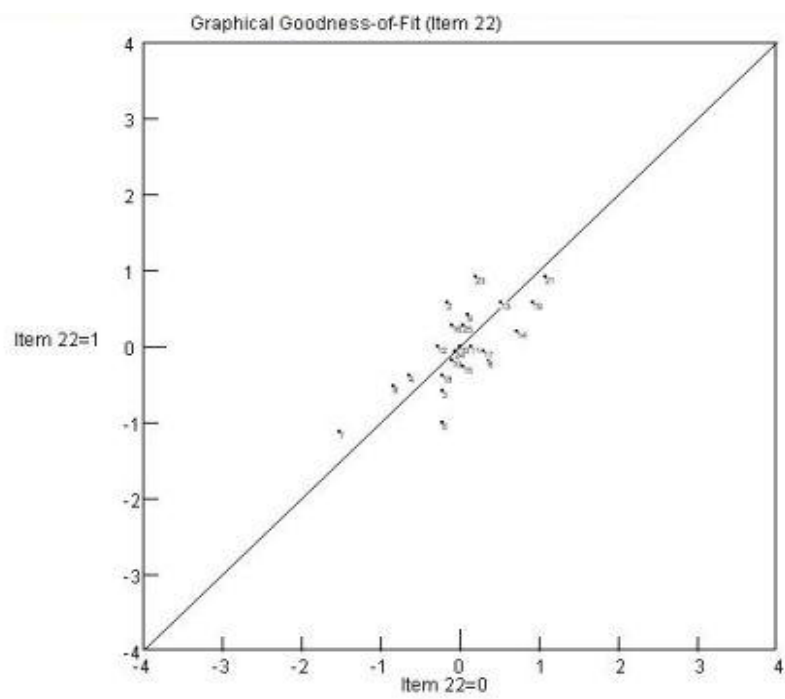


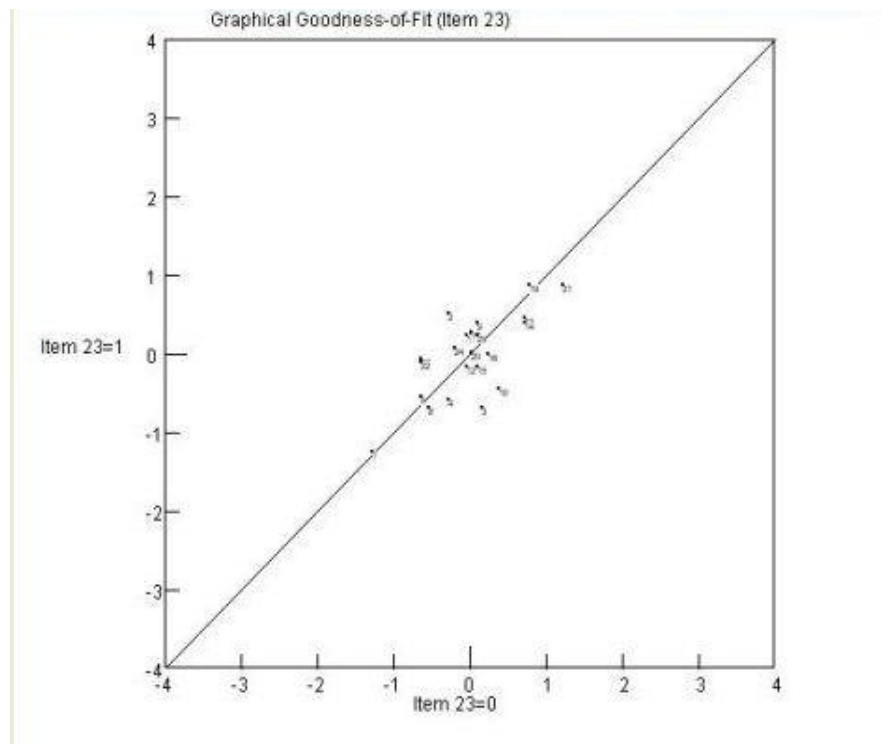
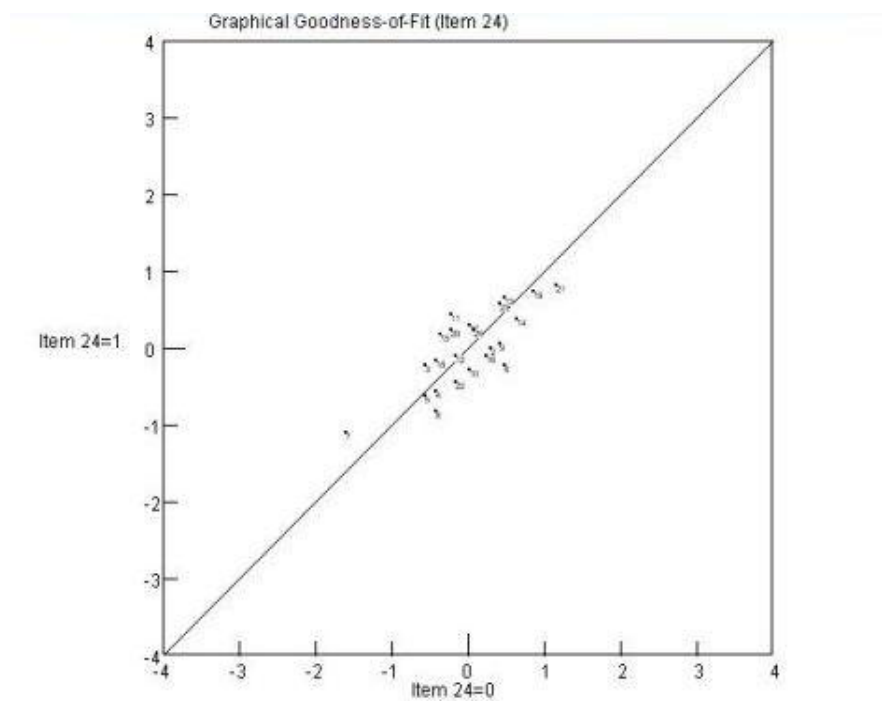
Teilungskriterium: Item 16

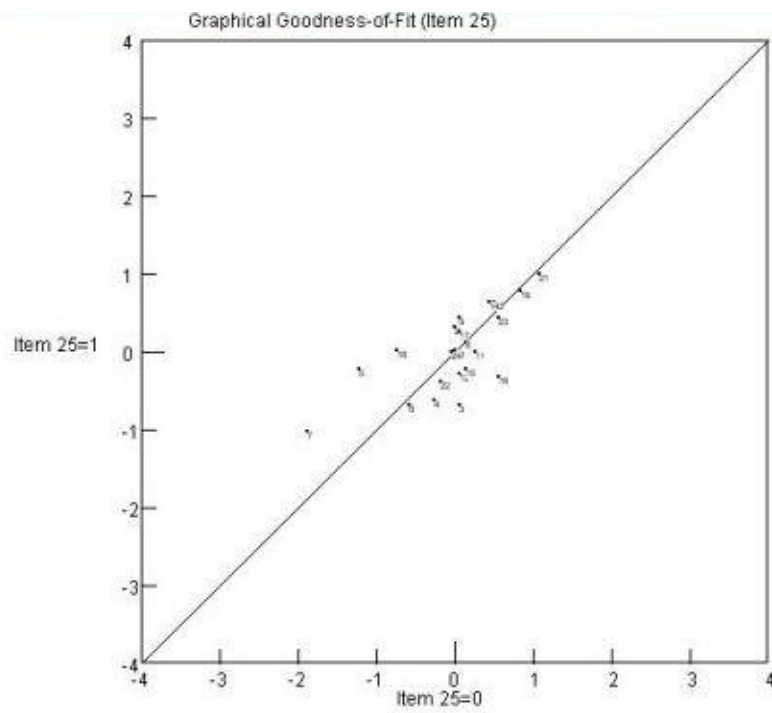


Teilungskriterium: Item 17**Teilungskriterium: Item 18**

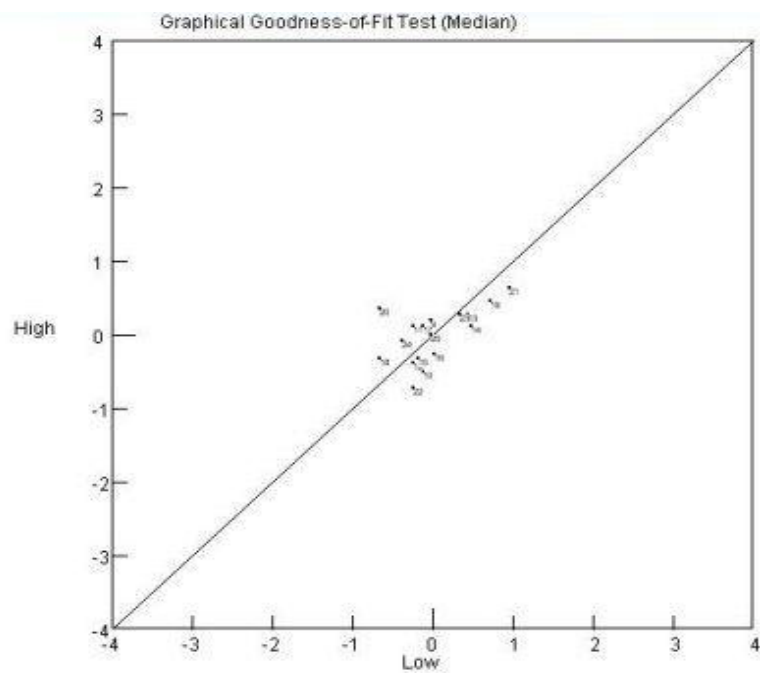
Teilungskriterium: Item 19**Teilungskriterium: Item 20**

Teilungskriterium: Item 21**Teilungskriterium: Item 22**

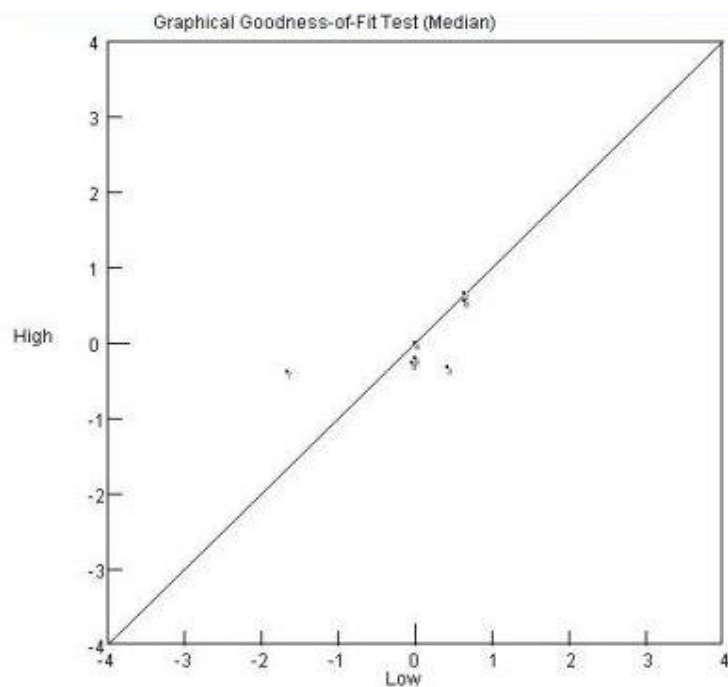
Teilungskriterium: Item 23**Teilungskriterium: Item 24**

Teilungskriterium: Item 25

Teilungskriterium: Median (mG-Aufgaben)



Teilungskriterium: Median (oG-Aufgaben)



Anhang B: Rasch Modell-Analyse

Teilungskriterium: Median_Gesamttest

Results for data file Rasch199.dat
obtained by means of program LpcM-WiN 1.0

Calculation of a Rasch Model
Total Sample

199 Persons
1 Treatment Group
24 Items (out of 33 Items)
24 Item Parameter Estimates (sum-normalized)
11 Iterations
Convergence reached in 0.0000 seconds

Total Log-Likelihood -1961.241245

	Item Parameters	Standard Errors
1	Not used!	
2	0.1603	0.1695
3	-0.3849	0.0846
4	-0.5073	0.0876
5	-0.5998	0.0909
6	0.1300	0.0786
7	-1.2800	0.1782
8	-0.6619	0.0960
9	0.2513	0.0854
10	-0.1422	0.0721
11	0.0997	0.0767
12	-0.1422	0.0721
13	0.5584	0.0852
14	0.5273	0.0850
15	-0.0817	0.0706
16	0.0694	0.0748
17	0.1603	0.0806
18	-0.2936	0.0804
19	0.8100	0.1138
20	0.0089	0.0718
21	1.0368	0.1373
22	-0.3240	0.0820
23	0.4964	0.0853
24	-0.0515	0.0705
25	0.1603	0.0806
26	Not used!	
27	Not used!	
28	Not used!	
29	Not used!	
30	Not used!	
31	Not used!	
32	Not used!	
33	Not used!	

Median per Subgroup: 12.00

Calculation of a Rasch Model
Low Rawscore according to Median

100 Persons

1 Treatment Group

24 Items (out of 33 Items)

24 Item Parameter Estimates (sum-normalized)

20 Iterations

Convergence reached in 0.0000 seconds

Total Log-Likelihood -1001.617294

	Item Parameters	Standard Errors
1	Not used!	
2	0.2784	0.2187
3	-0.1242	0.1250
4	-0.5423	0.2199
5	-0.7073	0.2858
6	0.2784	0.1368
7	-2.4077	0.5413
8	-0.4656	0.1788
9	0.1697	0.1236
10	0.1697	0.1236
11	0.0564	0.1217
12	-0.0624	0.1212
13	0.6812	0.1427
14	0.5350	0.1427
15	-0.0624	0.1212
16	0.3313	0.1446
17	-0.0022	0.1213
18	-0.3216	0.1567
19	0.8700	0.1673
20	-0.2535	0.1485
21	1.1451	0.2030
22	-0.0022	0.1213
23	0.6330	0.1384
24	-0.2535	0.1485
25	0.0564	0.1217
26	Not used!	
27	Not used!	
28	Not used!	
29	Not used!	
30	Not used!	
31	Not used!	
32	Not used!	
33	Not used!	

Calculation of a Rasch Model
High Rawscore according to Median

99 Persons
 1 Treatment Group
 24 Items (out of 33 Items)
 24 Item Parameter Estimates (sum-normalized)
 20 Iterations
 Convergence reached in 0.0000 seconds

Total Log-Likelihood -946.122719

	Item Parameters	Standard Errors
1	Not used!	
2	0.0619	0.2466
3	-0.5619	0.1323
4	-0.4580	0.1265
5	-0.5103	0.1267
6	-0.0017	0.1452
7	-1.0031	0.2135
8	-0.7621	0.1662
9	0.4159	0.1339
10	-0.4049	0.1301
11	0.1955	0.1448
12	-0.1824	0.1320
13	0.4159	0.1339
14	0.5807	0.1915
15	-0.0635	0.1366
16	-0.1824	0.1320
17	0.4159	0.1339
18	-0.2397	0.1346
19	0.7654	0.2583
20	0.3393	0.1259
21	0.8673	0.3094
22	-0.5619	0.1323
23	0.3393	0.1259
24	0.1955	0.1448
25	0.3393	0.1259
26	Not used!	
27	Not used!	
28	Not used!	
29	Not used!	
30	Not used!	
31	Not used!	
32	Not used!	
33	Not used!	

 Conditional Likelihood-Ratio-Test
 (Median)

Andersen chi_2 = 27.0025
 Degrees of Freedom = 23
 chi_2 at alpha(5 percent) = 35.1651 (Wilson-Hilferty
 approximation)

chi_2 at alpha(1 percent) = 41.6620 (Wilson-Hilferty
approximation)

Wald z-Values
(Median)

Item No.	z-Value	Significance
----------	---------	--------------

2	0.6570	n.s.
3	2.4045	sign.(5%)
4	-0.3320	n.s.
5	-0.6302	n.s.
6	1.4043	n.s.
7	-2.4141	sign.(5%)
8	1.2146	n.s.
9	-1.3511	n.s.
10	3.2014	sign.(5% and 1%)
11	-0.7355	n.s.
12	0.6698	n.s.
13	1.3560	n.s.
14	-0.1915	n.s.
15	0.0065	n.s.
16	2.6233	sign.(5% and 1%)
17	-2.3143	sign.(5%)
18	-0.3961	n.s.
19	0.3400	n.s.
20	-3.0454	sign.(5% and 1%)
21	0.7507	n.s.
22	3.1172	sign.(5% and 1%)
23	1.5699	n.s.
24	-2.1652	sign.(5%)
25	-1.6161	n.s.

Pearson Correlation between parameters = 0.7604

Person Parameters

Group 1

Raw	Score	Theta	Xi
1	-3.2425	0.0391	
2	-2.4948	0.0825	
3	-2.0329	0.1310	
4	-1.6867	0.1851	
5	-1.4027	0.2459	
6	-1.1568	0.3145	
7	-0.9360	0.3922	
8	-0.7325	0.4807	
9	-0.5409	0.5822	
10	-0.3572	0.6996	
11	-0.1785	0.8365	
12	-0.0021	0.9979	
13	0.1744	1.1905	
14	0.3533	1.4238	
15	0.5374	1.7115	

16	0.7295	2.0741
17	0.9337	2.5439
18	1.1553	3.1750
19	1.4023	4.0646
20	1.6877	5.4072
21	2.0356	7.6572
22	2.4996	12.1773
23	3.2498	25.7840

*

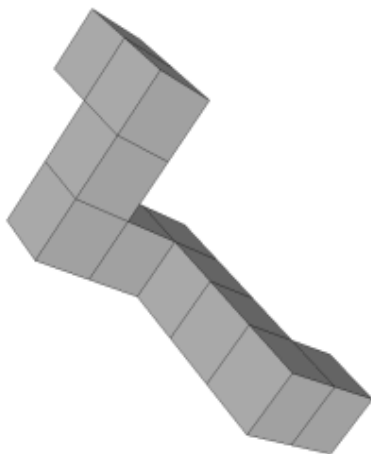
Martin-Löf-Test

Group 1: 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,
24, 25,
Group 2: 2, 3, 4, 5, 6, 7, 8,
LogLikelihood (Group 1): -1290.800916
LogLikelihood (Group 2): -423.077091
Martin-Löf-Statistic: 90.5563 n.s. (5% und 1%)
(chi_2 (5%) = 144.3523, chi_2 (1%) = 156.6598, and df=118)

Anhang C: Zusatzfragen

Datum:
 Probandencode:

In dem von Ihnen soeben abgeschlossenen Testverfahren wurden Ihnen zwei verschiedene Würfeltypen präsentiert. Der Unterschied lag in der Darstellung der Würfel: Es waren Würfel ohne Gitternetz und Würfel mit Gitternetz abgebildet. Sie sollen nun noch ein paar Aussagen zu diesen verschiedenen Würfeltypen beantworten. Bitte geben Sie bei den folgenden Aussagen jeweils an, wie sehr diese für Sie zutreffen!

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

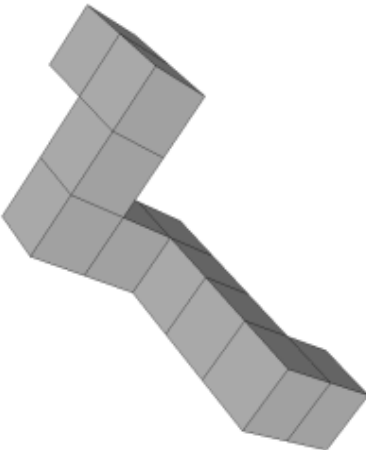
1. Würfel ohne Gitternetz habe ich gedanklich als Ganzes gedreht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

2. Würfel mit Gitternetz habe ich gedanklich als Ganzes gedreht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

Datum:
Probandencode:

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

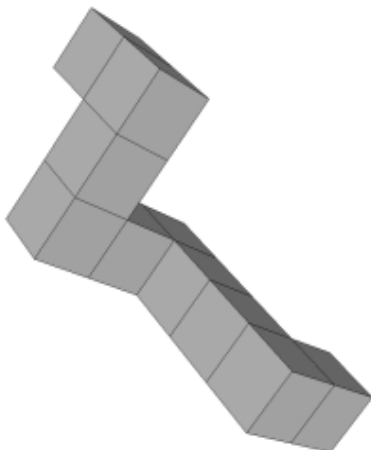
3. Bei Würfeln ohne Gitternetz habe ich mich an bestimmten Teilaspekten der Würfel orientiert.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

4. Bei Würfeln mit Gitternetz habe ich mich an bestimmten Teilaspekten der Würfel orientiert.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

Datum:
 Probandencode:

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

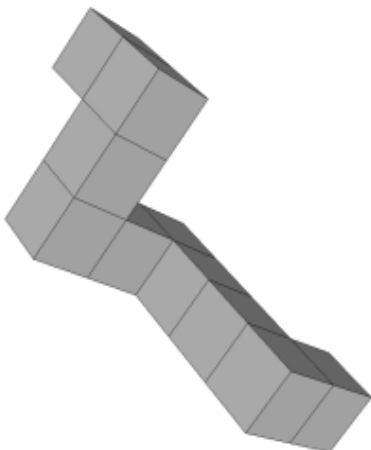
5. Bei Würfeln ohne Gitternetz hatte ich einen starken Eindruck von räumlicher Tiefe.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

6. Bei Würfeln mit Gitternetz hatte ich einen starken Eindruck von räumlicher Tiefe.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

Datum:
Probandencode:

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

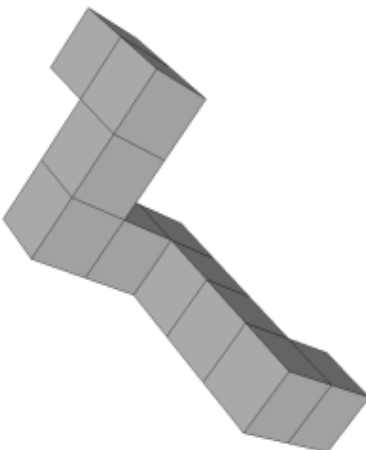
7. Bei Würfeln ohne Gitternetz fiel mir das Erfassen der Gesamtfigur leicht.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

8. Bei Würfeln mit Gitternetz fiel mir das Erfassen der Gesamtfigur leicht.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

Datum:
Probandencode:

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

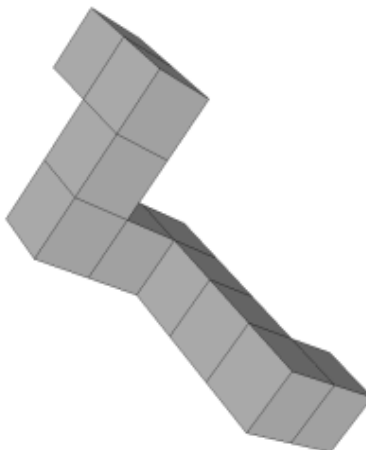
9. Bei Würfeln ohne Gitternetz fiel mir insgesamt gesehen das Bearbeiten der Aufgaben leicht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

10. Bei Würfeln mit Gitternetz fiel mir insgesamt gesehen das Bearbeiten der Aufgaben leicht.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

Datum:
Probandencode:

Würfel ohne Gitternetz	Würfel mit Gitternetz
	

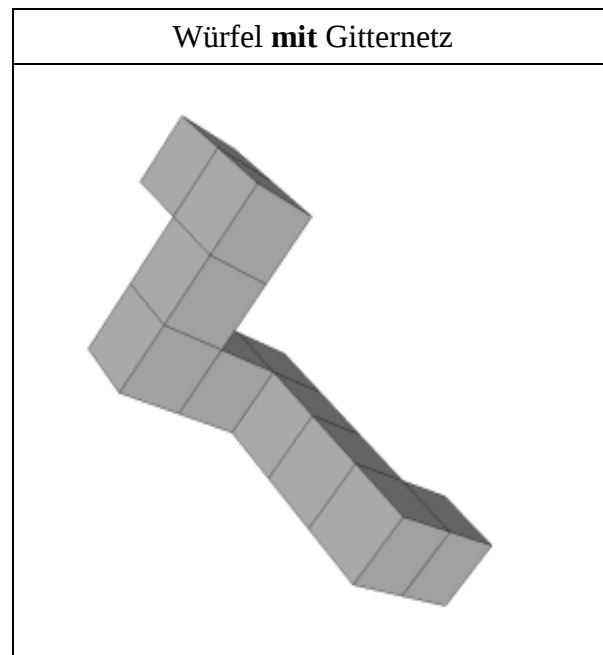
11. Würfel ohne Gitternetz empfand ich optisch ansprechend.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

12. Würfel mit Gitternetz empfand ich optisch ansprechend.

- ☐ Trifft sehr zu
☐ Trifft eher zu
☐ Trifft eher nicht zu
☐ Trifft gar nicht zu

Datum:
Probandencode:



13. Bei Würfeln mit Gitternetz habe ich die einzelnen Kästchen abgezählt.

- ☐ Trifft sehr zu
- ☐ Trifft eher zu
- ☐ Trifft eher nicht zu
- ☐ Trifft gar nicht zu

Anhang D: Curriculum vitae

Persönliche Daten

Nachname	SANTOS-SANTIAGO
Vorname	Siquia
Nationalität	Österreich

Bildungsweg

Seit September 2010	Psychotherapeutisches Propädeutikum, Universitätslehrgang, HoPP, Wien
Seit September 2003	Psychologiestudium, Universität Wien; 1. Diplom 2007 Schwerpunkt Studium: Klinische- und Gesundheitspsychologie Schwerpunkt Freier Wahlfächer: Klinisch-Medizinische Psychologie
09.01.-26.03.2003	«Español para extranjeros, Nivel Superior», Fakultät der Philologie Universidad Complutense de Madrid, Spanien
1989-2002	Waldorfschule Pötzleinsdorf, Wien; Externistenmatura 2002

Praktika

02.07.-10.08.2012	Praktikum am AMEOS Klinikum Bad Aussee
17.08.-25.09.2009	Praktikum am Landesklinikum St. Pölten, Neurologie
17.02.-12.05.2009	Praktikum im „Kolpinghaus Gemeinsam Leben“

Seminare

01.02.2011	Zusammen sind wir mehr! Klinische Psychologie als wesentlicher Bestandteil interdisziplinärer Zusammenarbeit im Krankenhaus; 3. Fachtagung des Klinisch-Psychologischen Fachgremiums AKH Wien und MUW
24.-28.08.2010	Ferienuniversität Kritische Psychologie, Freie Universität Berlin
Okt. 2009-Juni 2010	Vertiefte Ausbildung in medizinischer Psychologie: Psychologie von Tabu, Stress und Krankheit, Individualpsychologische Gruppenselbsterfahrung; Leitung Prof. Ulrich Kropiunigg
28.04.2009	„Angewandte Gerontopsychologie: das Langenloiser Modell“, Informationsabend der Fachsektion Klinische- und

Gesundheitspsychologie und Gerontopsychologie des
Berufsverbandes Österreichischer Psychologinnen und
Psychologen (BÖP)

- Okt. 2008-Juni 2009 Praktikum, Anamnesegruppe-Transdisziplinäre Kleingruppe unter
Leitung von TutorInnen; Seminar mit PatientInnen; Leitung Prof.
Klaus Spiess
- 16.-18.11.2007 BestNET-Kongress, Erster interdisziplinärer Kongress Beratung,
Therapie, Training, J. Kepler-Universität, Linz
- 14.-18.08.2006 Gruppendynamikseminar nach Kurt Lewin am IFF, Fakultät für
Interdisziplinäre Forschung und Fortbildung, Wien

Sonstiges

- 21.3.-31.12.2012 Honorartrainerin bei BBRZ-Österreich (Berufliches Bildungs- und
Rehabilitationszentrum), Wien
- Seit März 2012 Assistentin im **zimd** (Zentrum für Interaktion, Medien & soziale
Diversität), Wien
- Seit 2006 Ordinationsgehilfin in Arztpraxis für Allgemeinmedizin und
Psychotherapie, Wien