



universität
wien

DISSERTATION

Titel der Dissertation

Auswirkung der Bearbeitungsrichtung bei
Raumvorstellungsaufgaben auf die Testleistung:
Eine kognitionspsychologische Studie

Verfasserin

Mag. Veronika Steinhauser

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. rer. nat.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 091 298
Dissertationsgebiet lt. Studienblatt:	Psychologie
Betreuerin / Betreuer:	Ao.Univ.-Prof. Dr. Georg Gittler

Inhaltsverzeichnis

Danksagung	6
1. Einleitung	7
2. Theoretischer Teil	9
2.1 Raumvorstellung im Rahmen von Intelligenztheorien	9
2.1.1 Zwei-Faktoren-Theorie von Spearman	9
2.1.2 Primärfaktormodell von Thurstone	10
2.1.3 Modell der „fluid“ und „crystallized general intelligence“ von Cattell	11
2.1.4 Das „Structure of Intellect“-Modell von Guilford	12
2.1.5 Das Berliner Intelligenzstrukturmodell von Jäger	14
2.1.6 Die Intelligenztheorie von Gardner	15
2.2 Das Konstrukt Raumvorstellung	17
2.2.1 Zur Struktur des Faktors Raumvorstellung	18
2.2.1.1 Ergebnisse aus der faktorenanalytischen Forschung	18
2.2.1.2 Kritik an der faktorenanalytischen Forschung	22
2.2.1.3 Ergebnisse aus einer anderen Forschungsperspektive: Ein kognitiver Zugang	23
2.3 Bearbeitungsstrategien bei Raumvorstellungsaufgaben	26
2.3.1 Erste Untersuchungen zu Bearbeitungsstrategien	26
2.3.2 Studien zum Zusammenhang zwischen Bearbeitungsstrategie und Leistung	28
2.3.3 Strategien beim Mental Rotations Test	32
2.3.4 Strategien bei der Bearbeitung von Würfelaufgaben	33
2.3.5 Kognitive Koordinatensysteme	35
2.4 Geschlechtsunterschiede bei Raumvorstellung	40
2.4.1 Die Größe der Geschlechtsunterschiede	40
2.4.2 Warum treten Geschlechtsunterschiede bei Raumvorstellung auf?	46
2.4.2.1 Körperliche Ursachen	47
2.4.2.2 Evolutionäre Ursachen	48
2.4.2.3 Soziale und soziokulturelle Ursachen	49
2.4.3 Geschlechtsunterschiede bei Varianten der Testvorgabe und Testauswertung	50
2.4.4 Veränderungen bei Geschlechtsunterschieden	51
2.4.5 Geschlechtsunterschiede und Bearbeitungsstrategien	53
2.4.6 Die Berücksichtigung von Anlage- und Umweltaspekten bei der Erforschung von Geschlechtsunterschieden sowie Leistungsunterschieden innerhalb der Subgruppe der Frauen	54
2.5 Händigkeit	57
2.5.1 Händigkeit und Leistung	58

2.5.2 Die Theorie von Geschwind, Behan und Galaburda (GBG)	59
3. Empirischer Teil	61
3.1 Entstehungsgeschichte der Endlosschleifen	61
3.2 Entwicklung der Fragestellung für die vorliegende Arbeit	62
3.3 Beschreibung des verwendeten Testmaterials	64
3.3.1 Endlosschleifenaufgaben	64
3.3.2 Zusätzliches Testmaterial	67
3.4 Weitere Fragestellungen	68
3.5 Stichprobenbeschreibung	69
3.5.1 Geschlecht	69
3.5.2 Alter	70
3.5.3 Bildungsstand	71
3.5.4 Besuch von Unterricht in Darstellender Geometrie	71
3.5.5 Bearbeitungszeit	71
3.5.6 Verteilung der Testpersonen auf die Testversionen	73
3.5.7 Item- und Datenselektion	73
3.5.7.1 Itemselektion	73
3.5.7.2 Datenselektion	74
3.5.8 Verteilung des Scores	75
3.6 Auswertungen mit Methoden der Probabilistischen und Klassischen Testtheorie	79
3.6.1 Analysen nach dem Modell von RASCH	79
3.6.1.1 Einleitende Worte	79
3.6.1.2 Auswertungsvarianten mit dem Rasch-Modell für die vorliegenden Daten	80
3.6.1.3 Likelihood-Quotienten-Tests zwischen den verschiedenen Rasch-Modellen	100
3.6.1.4 Likelihood-Quotienten-Tests auf Ebene der Teilungskriterien	103
3.6.1.5 Weitere Likelihood-Quotienten-Tests	105
3.6.1.6 Schlussfolgerungen aus den Ergebnissen der Likelihood-Quotienten-Tests	108
3.7 Klassisch-testtheoretische Auswertungen	110
3.7.1 Schätzung der Personenparameter nach Warm	110
3.7.2 Überprüfung von Unterschieden zwischen Testpersonengruppen	112
3.7.2.1 Überprüfung der Frage nach Leistungsunterschieden zwischen den Geschlechtern	114
3.7.2.2 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit des Alters	114

3.7.2.3 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit des Besuchs von Unterricht in Darstellender Geometrie	115
3.7.2.4 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der Bearbeitungsdauer.....	116
3.7.2.5 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der Testform.....	118
3.7.2.6 Überprüfung der Frage nach der Auswirkung von Wechselwirkungen zwischen Geschlecht, Alter und Zeit auf die Leistung	120
3.7.2.7 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der gewählten Strategie	123
3.7.3 Zusammenfassung der Ergebnisse der Überprüfung von Unterschieden zwischen Testpersonengruppen.....	126
3.7.4 Fragebogenanalysen	126
3.7.4.1 Fragebogen zur Einschätzung der Raumvorstellungsleistung, den Testaufgaben und dem Bearbeitungsverhalten	126
3.7.4.2 Fragebogen zur Händigkeit.....	151
4. Abschließende Zusammenfassung und Bewertung der Arbeit	156
Literaturverzeichnis	161
Tabellenverzeichnis.....	171
Abbildungsverzeichnis.....	182
Anhänge	183

Danksagung

Ich möchte mich bei all jenen bedanken, die mich bei der Erstellung der vorliegenden Arbeit unterstützt haben. Allen voran gilt mein Dank meinem Betreuer, Ao. Univ.-Prof. Dr. Georg Gittler, der mir eine interessante Fragestellung ermöglichte, mir als Mitarbeiterin in meiner Zeit als Tutorin und Studienassistentin in seiner Abteilung vieles beibrachte und mir mit (auswertungstechnischen) Ratschlägen hilfreich zur Seite stand. Bei meinen Eltern möchte ich mich dafür bedanken, dass sie mir durch ihre nicht nur finanzielle Unterstützung meine Ausbildung(en) ermöglichten. Ein „Dankeschön“ auch an die fast 300 Personen, die sich für die Testbearbeitung zur Verfügung stellten. Schließlich gilt der größte Dank meinem lieben Mann Georg, der mich unerschütterlich motivierte, die Dissertation trotz Berufstätigkeit und Kinder fortzusetzen und abzuschließen.

1. Einleitung

Die vorliegende Arbeit behandelt ein Testmaterial zur Erfassung des räumlichen Vorstellungsvermögens, das sich in Entwicklung befindet. Ausgangspunkt der Arbeit sind Ergebnisse aus zwei Datensätzen, die in Seminaren bei Prof. Gittler an der Fakultät für Psychologie im Sommersemester 1998 und Wintersemester 1998/99 erhoben wurden. In diesen beiden Untersuchungen wurden so genannte *photographische Endlosschleifen* (Gittler & Arendasy, 2003a) als Testmaterial verwendet. Eine *Endlosschleife* ist ein gewundener, in sich geschlossener Schlauch (siehe Abbildung 1) und liegt jeweils in zwei Ansichten vor: der Startansicht und einer um 90 Grad gedrehten Zielansicht. Die Leistung einer Testperson liegt darin, zu entscheiden, in welche von vier Richtungen die Startansicht rotiert werden muss, um mit der Zielansicht deckungsgleich zu sein.



Abbildung 1: *Photographische Endlosschleife*

Bei den oben angesprochenen Untersuchungen bearbeiteten Testpersonen *Endlosschleifenaufgaben*, die in zwei Blöcken vorgegeben wurden. Die beiden Blöcke beinhalteten die gleichen Aufgaben, nur wurden in einem Block Start- und Zielansicht vertauscht – im Folgenden soll von korrespondierenden Items die Rede sein. Theoretische Vorüberlegungen führen dazu, anzunehmen, dass sich die jeweils korrespondierenden Items in ihrer Schwierigkeit nicht unterscheiden.

Die Ergebnisse zeigten allerdings ein anderes Bild, das vermuten lässt, dass sich einige oder alle korrespondierenden Items bezüglich ihrer Schwierigkeit unterscheiden, wenn Start- und

Zielansicht vertauscht werden. Dieses Phänomen ist deshalb interessant, weil sich bei jeder Einzelaufgabe weder die Aufgabenstellung noch die Bilder (Start-/Zielansicht), die diese Aufgabe definieren, ändern, sondern lediglich die Position von Start- und Zielansicht und damit die durch das Itemdesign vorgeschriebene Bearbeitungsrichtung.

Das Ziel der vorliegenden Dissertation ist, diese empirische Beobachtung anhand anderer, nämlich computergenerierter *Endlosschleifenaufgaben* in einem abgeänderten Studiendesign zu überprüfen.

Um die Arbeit flüssig lesbar zu gestalten, wurde auf die Verwendung geschlechtsneutraler Formulierungen verzichtet. Es sind in diesen Fällen jedoch immer beide Geschlechter im Sinne der Gleichbehandlung angesprochen.

2. Theoretischer Teil

2.1 Raumvorstellung im Rahmen von Intelligenztheorien

In den folgenden Unterkapiteln wird ein Überblick über einige wichtige und bekannte Intelligenztheorien und -modelle gegeben. Raumvorstellung wird in den beschriebenen Modellen zum einen als unabhängiger Faktor oder eigene Intelligenz gesehen, zum anderen als untergeordneter Faktor. In manchen Modellen wird nicht der Begriff „Raumvorstellung“ als Faktor verwendet, sondern von figuralen, bildhaften Informationen oder Inhalten gesprochen.

2.1.1 Zwei-Faktoren-Theorie von Spearman

Spearman (1904, zit. nach Asendorpf, 1999) formulierte die Zwei-Faktoren-Theorie der Intelligenz. Er stellte fest, dass Personen, die in einem Leistungstest gut abschnitten, auch in anderen Tests gute Ergebnisse erbrachten. Spearman schloss daraus, dass es einen gemeinsamen Faktor der Intelligenz, den so genannten g-Faktor (Generalfaktor der Intelligenz) gibt, der Leistungen zugrunde liegt. In faktorenanalytischen Untersuchungen extrahierte Spearman neben dem g-Faktor weitere, untergeordnete, voneinander unabhängige Faktoren, die er als s-Faktoren bezeichnete (spezifische Faktoren, beispielsweise Raumvorstellung; siehe Abbildung 2). Aus dieser Theorie ergibt sich, dass Leistungstests positiv miteinander korreliert sein müssen, und zwar in dem Ausmaß, in dem die Tests den g-Faktor messen (Asendorpf, 1999; Williams, Zimmerman, Zumbo & Ross, 2003).

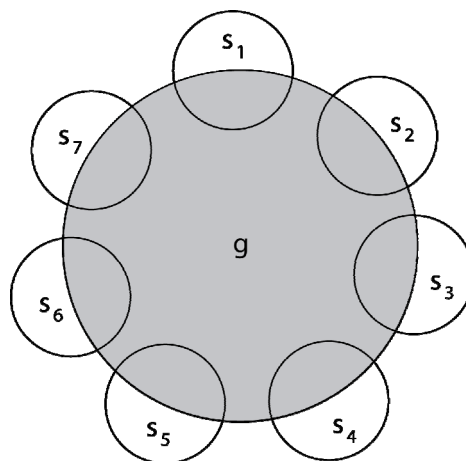


Abbildung 2: Schematische Darstellung der Zwei-Faktoren-Theorie von Spearman

2.1.2 Primärfaktormodell von Thurstone

Thurstone (1938, zitiert nach Amelang & Bartussek, 2001) nahm eine gegensätzliche Position zu Spearman ein, indem er von mehreren unabhängigen Intelligenzfaktoren ausging, die beim Lösen von Aufgaben in wechselnden Gewichtungsverhältnissen zum Tragen kommen. Als Methode verwendete Thurstone die multiple Faktorenanalyse und das Einfachstrukturkriterium (das bedeutet, dass jeder Faktor einige hohe Ladungen und viele um Null liegende Ladungen aufweist). Thurstone konnte anhand einer Stichprobe von 218 College-Studenten, die eine Batterie von 56 Tests bearbeiteten, neun Faktoren, so genannte „Primary Abilities“ (Primärfähigkeiten) identifizieren; bei späteren Untersuchungen resultierten folgende sieben Faktoren (vgl. Thurstone & Thurstone, 1941, zitiert nach Amelang & Bartussek, 2001):

verbal comprehension (v)

Dieser Faktor umfasst die Kenntnis von Wörtern und deren Bedeutung sowie die angemessene Verwendung von Wörtern im Gespräch.

Testitems: z.B. verbale Analogien, Wort-/Textverständnis, Rechtschreibung

word fluency (w)

Unter diesem Faktor versteht man die rasche Produktion von Wörtern, die strukturellen oder symbolischen Anforderungen entsprechen.

Testitems: z.B. Wörter finden, die mit einem bestimmten Buchstaben beginnen, Reime bilden

number (n)

Mit diesem Faktor werden die rechnerischen Fertigkeiten bei einfachen arithmetischen Aufgaben beschrieben.

Testitems: Aufgaben zu den Grundrechnungsarten

memory (m)

Dieser Faktor kennzeichnet die Merkfähigkeit des Kurzzeitgedächtnisses.

Testitems: Behalten und Reproduzieren von paarweise gelernten Assoziationen (z.B. Wort-Zahl-, Bild-Figur-, Wort-Bild-Paare)

perceptual speed (p)

Mit diesem Faktor ist die Wahrnehmungsgeschwindigkeit gemeint, d.h. das rasche Erkennen von relevanten Details in irrelevantem Material. 1944 unterteilte Thurstone (zitiert nach Mayer, 1994) diesen Faktor in zwei Subfaktoren, von denen einer die Geschwindigkeit bei der Herstellung von Beziehungen, der andere die Fähigkeit, umzustrukturieren, kennzeichnet.

Testitems: z.B. bestimmte Symbole anstreichen, versteckte Figuren finden

space (s)

Der Faktor bezieht sich auf das räumliche Vorstellungsvermögen, die räumliche Orientierung und das Erkennen von Objekten aus verschiedenen Perspektiven. Zu einem späteren Zeitpunkt unterteilte Thurstone (1949; 1950, zitiert nach Maier, 1994) diesen Faktor in drei Subfaktoren (siehe auch Kap. 2.2.1 „Zur Struktur des Faktors Raumvorstellung“)

Testitems: z.B. Würfel aus unterschiedlichen Perspektiven vergleichen, Papierfaltaufgaben, mechanische Bewegungen verfolgen

reasoning (r) oder induction (i)

Bei diesem Faktor geht es um schlussfolgerndes Denken bzw. um das Erkennen und Anwenden von Regeln in Zahlen- oder Symbolfolgen.

Testitems: z.B. Reihen fortsetzen, Muster ergänzen

2.1.3 Modell der „fluid“ und „crystallized general intelligence“ von Cattell

Cattells Theorie (Cattell, 1963, zitiert nach Maier, 1994; vgl. auch Amelang & Bartussek, 2001) kann als Zusammenschluss der Intelligenzmodelle von Spearman und Thurstone verstanden werden: Es wird ein g-Faktor (Spearman) angenommen, der aus den interkorrelierenden Sekundär- bzw. Primärfaktoren (Thurstone) erschlossen wird. Interessant sind die ersten beiden Sekundärfaktoren, die als „fluid“ bzw. „crystallized general intelligence“ interpretiert wurden:

Fluide Intelligenz (fluid intelligence, g_f):

Die fluide Intelligenz umfasst die angeborene Leistungsfähigkeit. Sie beschreibt die Fähigkeit, sich bei neuartigen Problemen und in neuartigen Situationen zurechtzufinden, ohne auf Lernerfahrungen zurückgreifen zu müssen. Dieser Faktor ist genetisch bestimmt und nimmt im Alter ab. Die fluide Intelligenz, in deren Bereich auch räumlich-visuelle Faktoren fallen, kann „culture fair“ – unabhängig vom erworbenen Allgemeinwissen – gemessen werden. Nach Horn (1968, zitiert nach Amelang & Bartussek, 2001) ist der Faktor g_f durch „Figural Relations“, „Memory Span“ sowie „Induction“ charakterisiert.

Kristalline Intelligenz (crystallized intelligence, g_c):

Dieser Faktor beinhaltet alle Fähigkeiten, die in bisherigen Lernprozessen erworben wurden und zum Problemlösen angewendet werden können. Die kristalline Intelligenz ist kulturab-

hängig und bleibt bis ins Alter stabil (bzw. kann sogar noch ansteigen). Marker-Items der kristallinen Intelligenz sind „Verbal Comprehension“, „Experimental Evaluation“ sowie „Semantic Relations“ (vgl. Amelang & Bartussek, 2001).

Dadurch, dass mehrere Primärfaktoren auf g_f und g_c laden, kann ein Faktor 3. Ordnung mit größerem Allgemeinheitsgrad extrahiert werden – nämlich Spearmans g-Faktor.

In der von Cattell 1987 formulierten Investmenttheorie (zitiert nach Kubinger & Jäger, 2003) wird das Verhältnis von fluider und kristalliner Intelligenz präzisiert: Die kristalline Intelligenz, die von Erfahrungen und kulturellen Einflüssen abhängt, ist ein Produkt der Investition der fluiden Intelligenz in entsprechende Lernmöglichkeiten. Bei Untersuchungen von Schmidt und Crano (1974, zitiert nach Amelang & Bartussek, 2001) konnte der kausale Zusammenhang allerdings nur an Mittelschichtkindern nachgewiesen werden.

2.1.4 Das „Structure of Intellect“-Modell von Guilford

Guilford (1959, 1967, zitiert nach Amelang & Bartussek, 2001) verwendete ebenso die Faktorenanalyse als zentrales Forschungsinstrument, allerdings wurde sie zur Hypothesenprüfung – und nicht zum Auffinden einer Struktur – eingesetzt. In Guilfords Modell kann aufgrund der orthogonalen Rotationen kein g-Faktor gefunden werden.

Guilford formulierte anhand theoretischer Überlegungen und Untersuchungen mit spezifischen Aufgaben ein Modell, in dem Leistungsprozesse beschrieben, klassifiziert und erklärt werden sollen. Er traf im Sinne eines Modells der Informationsverarbeitung eine Unterscheidung zwischen Input-, Operations- und Output-Variablen.

Guilford und Hoepfner (1971, zitiert nach Amelang & Bartussek, 2001) unterschieden zwischen vier Inhaltsbereichen (Input), fünf Vermittlungsprozessen (Operationen) sowie sechs Produkten (Output) (siehe Abbildung 3).

- Die *Inhaltsbereiche* beschreiben die Arten oder Bereiche der Information:

figural (F): Information in konkreter Form, Beteiligung von visuellen, auditiven, kinästhetischen Sinnen

symbolisch (S): Information in Form von Zeichen, z.B. Buchstaben, Zahlen, Musiknoten

semantisch (M): Information in Form von Begriffen oder geistigen Konstrukten

behavioral oder *verhaltensmäßig (B)*: Information, die bei Interaktionen eine Rolle spielt, einschließlich Wünsche, Bedürfnisse, Einstellungen etc.

- Die *Vermittlungsprozesse* beschreiben, was beim Verarbeiten von Information im Organismus geschieht:

Kognition (C): Information wird entdeckt, erkannt, verstanden, begriffen.

Gedächtnis (M): Information wird im Speicher fixiert.

Divergente Produktion (D): Logische Alternativen werden aus der gegebenen Information entwickelt.

Konvergente Produktion (N): Logische Schlussfolgerungen werden aus der gegebenen Information entwickelt, die Betonung liegt auf dem Erreichen der besten Lösung.

Evaluation (E): Information wird verglichen, um zu beurteilen, ob ein Kriterium erreicht ist.

- Die *Produkte* sind jene Formen, die die Information durch die Aktivitäten des Organismus annimmt:

Einheiten (U): Teile von Information, die abgegrenzt sind

Klassen (C): Gruppierungen von Information.

Beziehungen (R): Verbindungen zwischen Informationen

Systeme (S): Ansammlungen von Information, die strukturiert oder organisiert sind

Transformationen (T): Veränderungen bei vorhandener Information

Implikationen (I): Informationen, die zufällig verbunden sind

Durch die Kombinationen von vier Inhalten mit fünf Operationen und sechs Produkten resultieren 120 Faktoren. Kritisiert wird das „Structure of Intellect“-Modell beispielsweise von Horn und Undheim (1977, zitiert nach Amelang & Bartussek, 2001), die eine machbare Überprüfung der 120 Faktoren wegen des zu großen Aufwands ausschließen.

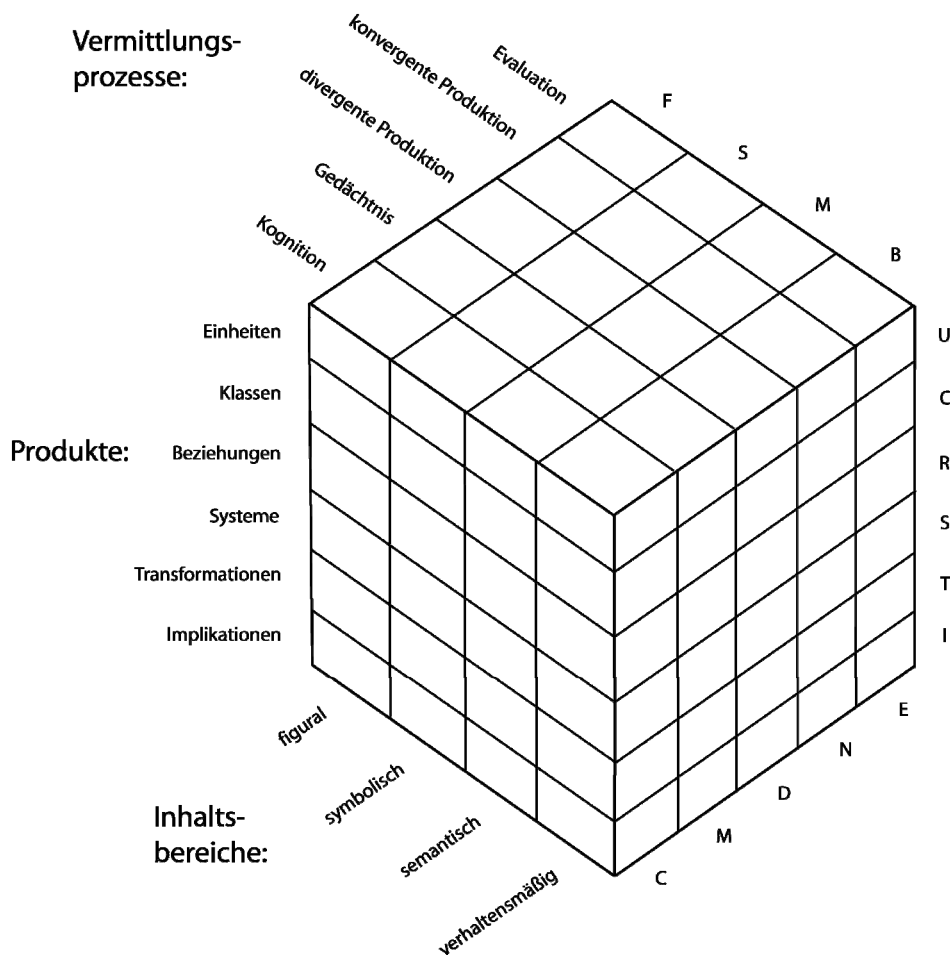


Abbildung 3: „Structure of Intellect“-Modell

2.1.5 Das Berliner Intelligenzstrukturmodell von Jäger

Jäger entwickelte 1984 (zitiert nach Amelang & Bartussek, 2001) aus den Modellen von Spearman, Thurstone und Guilford sowie aus Erkenntnissen der Intelligenzforschung das Berliner Intelligenzstrukturmodell. Ausgangspunkt war ein Pool aus ca. 2000 Itemtypen, der auf 191 Aufgabenblöcke reduziert wurde und einer Stichprobe von Maturanten vorgegeben wurde. Auf Grundlage dieser Daten entstand das integrative, hierarchische, bimodale Berliner Intelligenzstrukturmodell (vgl. Jäger, Süß & Beauducel, 1997). An der Spitze des Modells steht die „Allgemeine Intelligenz“, eine Ebene darunter können vier Operations- und drei Inhaltsklassen unterschieden werden. Die Operationsklassen sind Bearbeitungsgeschwindigkeit, Gedächtnis, Einfallsreichtum sowie Verarbeitungskapazität, die drei Inhaltsklassen sind figural-bildhaft, verbal und numerisch (siehe Abbildung 4).

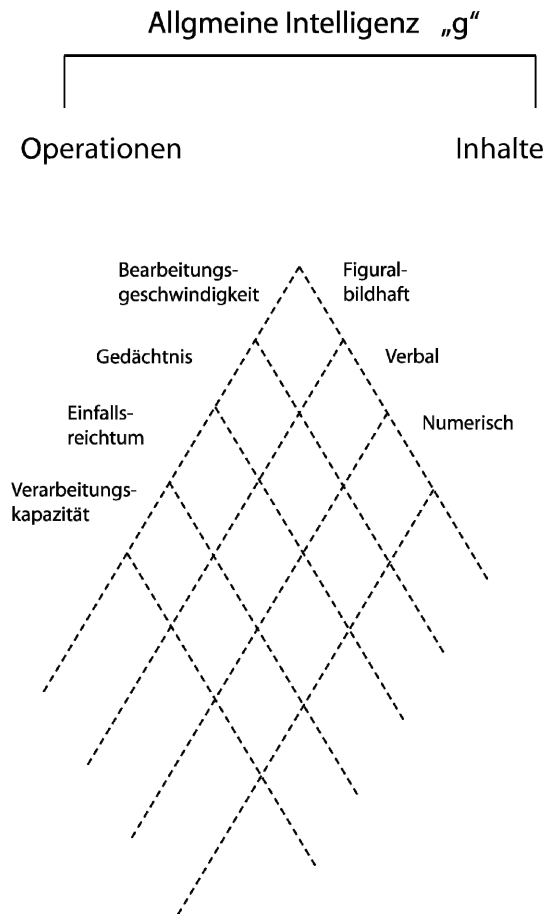


Abbildung 4: Berliner Intelligenzstrukturmodell

2.1.6 Die Intelligenztheorie von Gardner

Gardner (2002), der eine Vielfalt möglicher Intelligenzen finden wollte, verwendete hierfür keine psychometrischen Methoden, sondern nannte acht Kriterien, die auf Fähigkeiten zutreffen müssen, damit diese als menschliche Intelligenz bezeichnet werden können. Gardners acht Kriterien lauten:

- die Möglichkeit der Isolierung durch Hirnschäden
- die Verankerung in der Stammesgeschichte und eine entwicklungsgeschichtliche Plausibilität
- eine erkennbare Grundfunktion
- eine Eignung, sich symbolisch kodieren zu lassen

- die Verankerung in der Entwicklungsgeschichte des Individuums zusammen mit einem Komplex von „Endzuständen“ oder Expertenleistungen
- die Existenz von „idiots savants“, von Wunderkindern und anderen Ausnahmeerscheinungen
- die Bestätigung im psychologischen Experiment
- die Bestätigung durch psychometrische Ergebnisse

Durch Anwendung der Kriterien bilden sich nach Gardner (1994) sieben Intelligenzen heraus, nämlich:

Sprachliche Intelligenz: Fähigkeit, Sprachen zu lernen und sie für bestimmte Zwecke einzusetzen

Logisch-mathematische Intelligenz: Fähigkeit zur logischen Analyse von Aufgaben und zur Durchführung mathematischer Operationen

Musikalische Intelligenz: Fähigkeit, zu musizieren und komponieren

Körperlich-kinästhetische Intelligenz: Fähigkeit, Körperteile beim Problemlösen und Gestalten einzusetzen

Räumliche Intelligenz: Fähigkeit, große Räume (Piloten, Seeleute) sowie kleinere Räume (Chirurgen, Bildhauer) zu strukturieren

Personale Intelligenzen:

- Interpersonale Intelligenz: Fähigkeit, andere Menschen zu verstehen
- Intrapersonale Intelligenz: Fähigkeit, sich selbst zu verstehen.

2.2 Das Konstrukt Raumvorstellung

Die Bedeutung der Raumvorstellungsfähigkeit für unser tägliches Leben ist unumstritten. Wir benötigen räumliches Vorstellungsvermögen, um viele Tätigkeiten im Alltag sowie im Beruf zu bewältigen: um uns in Räumen zurechtzufinden (z.B. in Gebäuden, an verschiedenen Orten), um Entfernungen abzuschätzen (z.B. beim Einparken), um in verschiedenen Berufssparten erfolgreich sein zu können (z.B. in technischen Berufen wie der Architektur, in künstlerischen Berufen wie der Malerei, in naturwissenschaftlichen Berufen wie der Chemie und Physik, in medizinischen Berufen wie der Chirurgie, im Flugwesen) und für vieles mehr. In einem Verzeichnis des U.S. Employment Service (1957, zitiert nach Stumpf und Fay, 1983) von 4000 Industriebberufen wird bei 85 Prozent der erfassten Berufe Raumvorstellung als notwendig aufgelistet.

Eine einheitliche Bestimmung der Begriffe Raumvorstellung, Raumvorstellungsfähigkeit oder räumliches Vorstellungsvermögen gibt es nicht; gemeinsam ist den meisten Auslegungen, dass es bei Raumvorstellung um den Umgang mit räumlichen Informationen geht. Einige Definitionen aus der Literatur seien hier beispielhaft angeführt:

Thurstone (1938, zitiert nach Lohman, 1979) beschreibt den Faktor Raumvorstellung als „facility in spatial or visual imagery“ (S. 8).

Lohman (1979) definiert Raumvorstellungsfähigkeit als „... the ability to generate, retain, and manipulate abstract visual images. At the most basic level, spatial thinking requires the ability to encode, remember, transform, and match spatial stimuli.“ (S.126-127).

Eliots und Smiths (1983) Auslegung zur Raumvorstellungsfähigkeit lautet:

„Some scholars describe spatial ability broadly in terms of individual differences in the processing of non-linguistic information, while others describe it narrowly in terms of individual differences in performance on spatial tests.“ (S. 1).

Linn und Petersen (1985) schreiben, dass Raumvorstellungsfähigkeit „... generally refers to skill in representing, transforming, generating, and recalling symbolic, nonlinguistic information.“ (S. 1482).

Carroll (1993) verfasst folgende Definition über räumliche und visuelle Fähigkeiten:

Spatial and other visual perceptual abilities have to do with individuals' abilities in searching the visual field, apprehending the forms, shapes, and positions of objects as visually perceived, form-

ing mental representations of those forms, shapes, and positions, and manipulating such representations “mentally”. (S. 304)

Nach Maier (1994) geht es bei Raumvorstellung darum, „in der Vorstellung räumlich zu sehen und räumlich zu denken.“ (S. 14). Sinneseindrücke werden verarbeitet, sodass daraus Vorstellungsbilder entstehen, die im Gedächtnis gespeichert werden. Nach Maier gehört zur Raumvorstellung aber auch, mit „... diesen Bildern aktiv umzugehen, sie *mental umzuordnen* und *neue Bilder aus vorhandenen vorstellungsmäßig* zu entwickeln.“ (S. 14).

2.2.1 Zur Struktur des Faktors Raumvorstellung

In diesem Kapitel sollen Ergebnisse zur Struktur der Raumvorstellung dargestellt werden, die einerseits aus faktorenanalytischen Forschungsansätzen entwickelt wurden, andererseits aus einem kognitiven Zugang heraus entstanden.

2.2.1.1 Ergebnisse aus der faktorenanalytischen Forschung

Ähnlich den Definitionen von Raumvorstellung gibt es auch bei der Struktur des Faktors kein einheitliches Konzept. Die meisten Forscher versuchten mittels faktorenanalytischer Methoden eine Strukturierung des Faktors Raumvorstellung zu erzielen. Eine Gemeinsamkeit der hier dargestellten Konzepte ist zumindest die Sichtweise, dass sich Raumvorstellung aus zwei oder mehreren Subfaktoren zusammensetzt. Allerdings herrscht weder über die Anzahl dieser Subfaktoren, noch über deren Inhalt Einigkeit. Der Vergleich von verschiedenen Konzepten und Forschungsansätzen zu Subfaktoren gestaltet sich nicht einfach, da Forscher ähnliche Faktoren unterschiedlich benennen und umgekehrt ähnliche Faktorennamen unterschiedliches beinhalten.

In der Literatur wird El Koussy (1935, zitiert nach Eliot & Smith, 1983) als der Forscher genannt, der bei Untersuchungen mit einer Testbatterie, die er elf- bis 13-jährigen Buben vorgab, überhaupt die ersten Hinweise für die Existenz eines Raumvorstellungsfaktors fand. Seine Analysen ergaben keinen Gruppenfaktor, der die ganze Bandbreite von Raumvorstellung abdeckt, aber einen Faktor – von El Koussy mit „K“ bezeichnet –, der sich aus den Scores von sieben Raumvorstellungstests und einem weiteren Test zusammensetzte. Alle Tests erforderten die Fähigkeit, räumliches Material zu verwenden und zu manipulieren.

Der erste Wissenschaftler, der sich intensiv mit der Frage nach der Struktur des Faktors Raumvorstellung beschäftigte, war Thurstone. 1938 (zitiert nach Lohman, 1979) identifizierte er zwei räumliche Faktoren, von denen einer nicht interpretierbar war, da auch verbale Tests auf ihm luden. 1951 veröffentlichte Thurstone (zitiert nach Lohman, 1979) einen Bericht über drei räumliche Subfaktoren, die er mit *S1*, *S2* und *S3* bezeichnete:

Den Faktor *S1* interpretierte Thurstone als „the ability to visualize a rigid configuration when it is moved into different positions“ (zitiert nach Lohman, 1979, S. 67).

Den Subfaktor *S2* definierte Thurstone als „the ability to visualize a configuration in which there is movement or displacement among the parts of the configuration“ (zitiert nach Lohman, 1979, S. 70).

Zum Faktor *S3* schrieb Thurstone in einem früheren Bericht (1950, zitiert nach Lohman 1979), dass er die Fähigkeit beinhaltet, über räumliche Beziehungen unter Berücksichtigung der Position des Beobachters nachzudenken.

Dass nicht nur – wie am Anfang dieses Kapitels erwähnt – über Anzahl und Inhalt der Subfaktoren Uneinigkeit besteht, sondern auch darüber, zu welchen Ergebnissen Thurstone in welchem Jahr kam, zeigt beispielsweise ein Artikel von Grüßing (2002). Die Autorin schreibt, dass Thurstone 1938 in seiner „3-Faktoren-Hypothese“ die Subfaktoren *Spatial Relations*, *Visualization* und *Spatial Orientation* nannte. Diese drei Bezeichnungen finden sich auch in einem umfangreichen Review von Lohman (1979; siehe unten), allerdings als zusammenfassende Bezeichnungen für die drei wichtigsten Raumvorstellungssubfaktoren nach Durchsicht vieler Studien. Lohman (1979) zufolge erforscht Thurstone 1938 zwei Faktoren und 1951 drei, die er mit *S1*, *S2* und *S3* bezeichnet.

Zimmerman beschrieb 1953 (zitiert nach Lohman, 1979) in Anlehnung an Thurstones Ergebnisse aus 1938 zwei Faktoren, die er *Spatial Relations* und *Visualization* nannte. Ersterer ist mit Thurstones interpretierbarem, räumlichem Faktor identisch, letzterer eine Überarbeitung des nicht-interpretierbaren Faktors.

Guilford und Lacey berichteten 1947 (zitiert nach Lohman, 1979) von zwei sicheren Faktoren *Spatial Relations* und *Visualization* sowie von zwei unsicheren, vorläufigen Faktoren *S2* und *S3*. Der Faktor *Spatial Relations* ist als die Fähigkeit, den Zusammenhang zwischen räumlich angeordneten Reizen zu verstehen, beschrieben; der Faktor *Visualization* als Fähigkeit, sich den Positionswechsel von Objekten im Raum, die Rotation von Objekten sowie das Zusammen- und Entfalten von Mustern vorzustellen.

French (1951, zitiert nach Eliot & Smith, 1983) schloss aus der bis dato vorhandenen Forschung, dass zumindest drei getrennte Raumvorstellungsfaktoren existieren, ein *Spatial Factor*, ein *Orientation Factor* und ein *Visualization Factor*. Der *Spatial Factor* beschreibt nach French die Fähigkeit, räumliche Muster wahrzunehmen und sie miteinander zu vergleichen. French nahm an, dass der *Orientation Factor* die Fähigkeit beschreibt, „to remain unconfused by the varying orientations in which a spatial pattern may be represented“ (zitiert nach Eliot & Smith, 1983, S. 4). Der *Visualization Factor* beinhaltet die Fähigkeit, vorgestellte Bewegungen im dreidimensionalen Raum zu verstehen oder Objekte in der Vorstellung zu manipulieren.

McGee berichtete 1979, dass eine Vielzahl an Studien die Existenz von zwei unabhängigen Faktoren unterstützen, nämlich *Spatial Visualization* und *Spatial Orientation*. Unter *Spatial Visualization* fasst McGee die Fähigkeit, Objekte mental zu manipulieren, zusammen. Reizmuster, deren Teile in Bewegung sind und Objekte, die im dreidimensionalen Raum manipuliert werden, müssen wiedererkannt, gespeichert und abgerufen werden.

Zum Faktor *Spatial Orientation* zählen die Fähigkeiten, räumliche Zusammenhänge zwischen Reizen zu verstehen und die Orientierung (auch in Bezug auf den eigenen Körper) beibehalten zu können.

Lohman (1979) bietet in einem umfangreichen Review eine Übersicht über faktorenanalytische Untersuchungen zur Raumvorstellungsfähigkeit und beschreibt zusammenfassend nach Durchsicht vieler Studien drei wichtige Subfaktoren von Raumvorstellung:

Spatial Relations: Der Faktor kann nach Lohman nur dann extrahiert werden, wenn Tests vorgegeben werden, bei denen die mentale Rotation von Figuren erforderlich ist.

Spatial Orientation: Im Mittelpunkt dieses Faktors steht die Fähigkeit, sich einen Reiz aus einer anderen Perspektive vorzustellen. Lohman schreibt, dass Testpersonen bei Aufgaben zu diesem Faktor oft mental rotieren, statt durch Veränderung der eigenen Position die Perspektive zu wechseln.

Visualization: Auf diesen Faktor laden Tests mit räumlich-figuralem Inhalt, die einerseits nicht oder unter relativ wenig Zeitdruck vorgegeben werden und andererseits insofern komplexer sind, als sie mehrere Aufgabenschritte verlangen (z.B. Papierfaltaufgaben, Kopieraufgaben oder Versteckte Figuren). Tests mit hohen Ladungen auf den *Visualization-Faktor* ähneln auch oft jenen Tests, die Spearmans g-Faktor oder Cattells g_f-Faktor messen.

Weiters schrieb Lohman (1979), dass Faktoren wie *Closure Speed* (die Geschwindigkeit beim Vergleich unvollständiger visueller Reize mit ihrer Repräsentation im Langzeitgedächtnis), *Perceptual Speed* (Geschwindigkeit beim Vergleich visueller Vorlagen), *Visual Memory*

(Kurzzeitgedächtnis für visuelle Reize) und *Kinesthetic* (Geschwindigkeit beim Diskriminieren zwischen rechts und links) für individuelle Unterschiede in der Geschwindigkeit oder Effizienz beim Enkodieren, Erinnern, Transformieren von räumlichen Reizen sowie dabei, die Reize in Übereinstimmung zu bringen, verantwortlich sein könnten.

Carroll veröffentlichte 1993 eine umfangreiche Re-Analyse faktorenanalytischer Untersuchungen. Er achtete dabei vor allem auf den Inhalt der Testaufgaben, da ähnliche Testnamen nicht bedeuten, dass auch ähnliches Reizmaterial verwendet wird bzw. ähnliche Anforderungen an die Testpersonen gestellt werden. Die wichtigsten, unterscheidbaren Faktoren sind nach Carroll:

Visualization: Dieser Faktor beschreibt die Fähigkeit, visuelle Muster ohne Geschwindigkeitskomponente zu manipulieren.

Spatial Relations: Bei diesem Faktor geht es um das Manipulieren (z.B. mentale Rotationen) einfacher visueller Muster unter Berücksichtigung der Geschwindigkeit.

Closure Speed: Dieser Faktor beinhaltet die Geschwindigkeit, mit der ein visuelles Muster identifiziert wird, ohne im Voraus zu wissen, welches Muster entstehen wird.

Flexibility of Closure: Mit diesem Faktor wird die Geschwindigkeit bezeichnet, mit der ein im Vorhinein bekanntes visuelles Muster identifiziert wird.

Perceptual Speed: Bei diesem Faktor geht es um die Geschwindigkeit, mit der ein visuelles Muster gefunden wird oder die Geschwindigkeit, mit der Muster verglichen werden.

Carroll nennt weitere Faktoren, für deren unabhängige Existenz es Hinweise gibt, die aber durch zusätzliche Forschungsarbeiten abgesichert werden müssen. Dazu gehören *Serial Perceptual Integration* (die Fähigkeit, ein visuelles Muster zu identifizieren, wenn Teile des Musters seriell/sukzessiv präsentiert werden), *Spatial Scanning* (Geschwindigkeit, mit der einer Route durch ein visuelles Muster gefolgt werden kann), *Imagery* (Fähigkeit, mentale Repräsentationen von visuellen Mustern zu bilden und diese bei der Lösung räumlicher Aufgabenstellungen zu verwenden) sowie *Length Estimation* (Fähigkeit, Längen oder Distanzen einzuschätzen oder zu vergleichen).

Abschließend wurde in Tabelle 1 der Versuch unternommen, alle dargestellten Strukturkonzepte in ein System einzureihen, das keinen Anspruch auf Vollständigkeit sowie gänzlich korrekte Zuordnung erhebt. Anhand der Tabelle wird auch die am Anfang des Kapitels erwähnte uneinheitliche Benennung von Faktoren mit ähnlichen Inhalten ersichtlich.

Tabelle 1: Zusammenfassende Darstellung einiger Faktoren aus verschiedenen Strukturkonzepten von Raumvorstellung mit den Faktorennamen, den jeweiligen Forschern sowie dem Faktorinhalt.

Faktorennamen	Forscher	Inhalt des Faktors
Visualization S1 Visualization Factor Spatial Visualization Spatial Relations	Guilford & Lacey (1947, zit. n. Lohman, 1979) Thurstone (1951, zit. n. Lohman, 1979) French (1951, zit. n. Eliot & Smith, 1983) McGee (1979) Lohman (1979), Carroll (1993)	Positionswechsel von Objekten, Manipulation/Rotation von Objekten (unter Berücksichtigung der Geschwindigkeit), Zusammen-/Entfalten von Mustern/Objekten
Spatial Relations S2 S3 Spatial Factor Orientation Factor Spatial Orientation	Guilford & Lacey (1947, zit. n. Lohman, 1979) Thurstone (1951, zit. n. Lohman, 1979) Thurstone (1950, zit. n. Lohman, 1979) French (1951, zit. n. Eliot & Smith, 1983) French (1951, zit. n. Eliot & Smith, 1983) McGee (1979), Lohman (1979)	Zusammenhänge zwischen räumlich angeordneten Reizen verstehen, die Orientierung beibehalten, sich Reize aus einer anderen Perspektive vorstellen
Visualization	Lohman (1979), Carroll (1993)	Visuelle Muster ohne Zeitdruck manipulieren, komplexe Aufgaben mit mehreren Schritten bearbeiten

2.2.1.2 Kritik an der faktorenanalytischen Forschung

Eine grundlegende Annahme bei der Faktorenanalyse ist, dass Testpersonen bei der Bearbeitung von Testaufgaben dieselben Strategien verwenden. Unterschiede zwischen Testpersonen entstehen, weil manche Personen mehr, andere weniger gut ausgebildete Fähigkeiten haben bzw. sie mehr oder weniger gut einsetzen. Dass die Unterschiede nicht nur quantitativer, sondern auch qualitativer Natur sein können, wird bei der Faktorenanalyse nicht berücksichtigt.

Zwei Forscher, die sich intensiv mit diesem qualitativen Aspekt beschäftigten und – wie viele andere – auf die Probleme faktorenanalytischer Untersuchungen aufmerksam machten, waren Just und Carpenter (1985). Die Wissenschaftler sprechen, die Literatur zum Thema Raumvorstellung zusammenfassend, von zumindest zwei Faktoren, nämlich *Spatial Visualization* und *Spatial Orientation*: der erste Faktor werde häufig mit Rotationsaufgaben in Verbindung gebracht, während es beim zweiten um Relationen zwischen Reizen oder einem Wechsel der Position des Betrachters gehe. Just und Carpenter merken aber an, dass der

Nachweis, dass es sich um zwei getrennte Faktoren handle, schwierig sei. Sie führen das darauf zurück, dass die Anwendung qualitativ unterschiedlicher Bearbeitungsstrategien bewirken kann, dass ein Test ein Mal auf einen, ein anderes Mal auf einen anderen Faktor lädt. Just und Carpenter schlossen aus ihren Überlegungen, dass die Strategien, die in *Spatial Visualization*-Tests verwendet werden, mit jenen, die in *Spatial Orientation*-Tests zur Anwendung kommen, überlappen.

Ein anderer Forscher, der noch vor Just und Carpenter festhielt, dass unterschiedliche Ladungen von Raumvorstellungstests auf die Verwendung unterschiedlicher Strategien bei der Aufgabenbearbeitung zurückgeführt werden kann, war French (1965). Er bemerkte, dass selbst bei Tests, die nur einen Faktor messen (so genannte „Pure Factor Tests“), unterschiedliche Strategien zur Anwendung kommen und dass sich die Faktorenstruktur verändert, wenn die Strategien kontrolliert werden.

Die Ergebnisse einer weiteren Forscherin sollen hier nur kurz, im Kapitel 2.3.2 „Studien zum Zusammenhang zwischen Bearbeitungsstrategie und Leistung“ ausführlicher besprochen werden: Schultz (1991) schließt aus ihren Untersuchungen, dass bei Einbindung der Lösungsstrategien die Faktorenstruktur von Raumvorstellung klarer dargestellt werden kann. Die Autorin schreibt, dass die durch Faktorenanalysen identifizierten Subfaktoren eher widerspiegeln, wie Testpersonen die Aufgaben bearbeiten – anstatt die inhaltlichen Komponenten der Aufgaben zu beschreiben.

Zusammenfassend bedeutet das, dass „reine“ Konstrukte nur unter der Voraussetzung gemessen werden können, dass alle Testpersonen dieselben homogenen Strategien bei der Aufgabenbearbeitung verwenden. Dass dies nicht der Fall ist und der qualitative Aspekt der Bearbeitungsstrategien einen wichtigen Stellenwert einnimmt, wird im Kapitel 2.3 „Bearbeitungsstrategien bei Raumvorstellungsaufgaben“ erörtert.

2.2.1.3 Ergebnisse aus einer anderen Forschungsperspektive: Ein kognitiver Zugang

Einen anderen Zugang als jenen über die Faktorenanalyse wählten Linn und Petersen (1985). Sie beschrieben drei Kategorien von Raumvorstellungsfähigkeit, die sie erhielten, indem sie die bei der Bearbeitung von Testaufgaben ablaufenden Prozesse auf Ähnlichkeit untersuchten und damit einen Schritt in Richtung Berücksichtigung kognitiver, qualitativer

Aspekte gingen. Auf diesem Weg erhielten die Autorinnen die drei Kategorien *Spatial Perception*, *Spatial Visualization* und *Mental Rotation*:

Spatial Perception: In diese Kategorie fallen Tests zur räumlichen Wahrnehmung, wie beispielsweise der Rod and Frame Test (Witkin, Dyk & Faterson, 1962), bei dem Testpersonen trotz ablenkender Informationen räumliche Beziehungen zwischen Reizen und der Position des eigenen Körpers bestimmen müssen. Linn und Petersen stellten fest, dass Testpersonen einerseits das Reizmaterial rotieren, um eine Senkrechte zu erhalten und andererseits kinästhetische Hinweise für die Bestimmung der Senkrechten zu Hilfe nehmen.

Spatial Visualization (Veranschaulichung): Unter dieser Bezeichnung werden jene Tests mit räumlichen Inhalten zusammengefasst, die komplexer sind und bei denen mehrere Aufgabenschritte zu erledigen sind, wie beispielsweise Papierfaltaufgaben, Oberflächenabwicklungsaufgaben (Surface Development von French, Ekstrom & Price, 1963, zitiert nach Linn und Petersen, 1985) oder Versteckte Figuren. Linn und Petersen gehen davon aus, dass eine analytische Strategie zur Aufgabenbearbeitung notwendig ist, zum Teil aber auch Elemente von mentaler Rotation und räumlicher Wahrnehmung einfließen. Testpersonen, die bei *Spatial-Visualization*-Tests gut abschneiden, haben ein Repertoire an Lösungsstrategien zur Verfügung und sind beim Einsatz dieser Strategien je nach Aufgabenstellung flexibel.

Mental Rotation: Bei Tests dieser Kategorie müssen Testpersonen zwei- oder dreidimensionale Figuren – oft auch unter Zeitmessung – mental rotieren. Linn und Petersen meinen, dass analytische Prozesse für Rotationen dann zur Anwendung kommen, wenn die mentale Rotation Teil einer komplexeren Aufgabe wie beispielsweise bei Oberflächenabwicklungsaufgaben ist.

Beim Versuch, die über einen kognitiven Zugang erhaltenen Kategorien von Linn und Petersen (1985) mit den Ergebnissen aus der faktorenanalytischen Forschung (Tabelle 1) in Einklang zu bringen, resultiert folgendes Bild: Die Kategorie *Spatial Visualization* von Linn und Petersen ähnelt dem Subfaktor *Visualization* von Lohman (1979) und Carroll (1993). Die Kategorie *Mental Rotation* von Linn und Petersen entspricht in etwa dem Subfaktor *Visualization* (Guilford & Lacey, 1947, zitiert nach Lohman, 1979), *S1* (Thurstone, 1951, zitiert nach Lohman, 1979), dem *Visualization Factor* (French, 1951, zitiert nach Eliot & Smith, 1983), *Spatial Visualization* (McGee, 1979) sowie *Spatial Relations* (Lohman, 1979; Carroll, 1993). Schließlich bleibt noch Linn und Petersens Kategorie *Spatial Perception*, die sich in etwa mit *Spatial Relations* (Guilford & Lacey, 1947, zitiert nach Lohman, 1979), *S3* (Thurstone, 1950,

zitiert nach Lohman, 1979), dem *Orientation Factor* (French, 1951, zitiert nach Eliot & Smith, 1983) sowie *Spatial Orientation* (McGee, 1979) überschneidet.

Zusammenfassend zeigen sich trotz der unterschiedlichen Herangehensweise inhaltliche Übereinstimmungen zwischen den Ergebnissen aus der faktorenanalytischen Forschung und dem kognitiven Zugang.

2.3 Bearbeitungsstrategien bei Raumvorstellungsaufgaben

Die Verwendung unterschiedlicher Strategien bei der Bearbeitung von Testaufgaben ist – wie bereits weiter oben erwähnt wurde – von großer Bedeutung. Einerseits kann eine Testperson bei der Bearbeitung eines Tests mit nur einem Aufgabentypus von Item zu Item entscheiden, welche Strategie sie anwendet; ob sie bei einer Strategie bleibt oder in Abhängigkeit des vorgegebenen Items ihre Strategie anpasst und wechselt. Andererseits können viele Personen, die ein einzelnes Item vorgegeben bekommen, mit ganz unterschiedlichen Strategien an die Bearbeitung dieser einen Aufgabe herangehen. Es kann also davon ausgegangen werden, dass bei der Bearbeitung von Testaufgaben intra- und interindividuelle Unterschiede bei der Auswahl und Anwendung von Strategien vorliegen.

In den folgenden Unterkapiteln wird von den Anfängen der Untersuchungen zu Bearbeitungsstrategien sowie von Zusammenhängen zwischen Bearbeitungsstrategien und Leistung berichtet. Zwei weitere Unterkapitel beschäftigen sich mit Strategien bei in der Raumvorstellungsforschung häufig eingesetztem Testmaterial – dem Mental Rotations Test (Vandenberg & Kuse, 1987) und Würfelaufgaben. Schließlich wird ein Kapitel den Forschungsarbeiten von Just und Carpenter (1976, 1985) und ihren kognitiven Koordinatensystemen gewidmet.

2.3.1 Erste Untersuchungen zu Bearbeitungsstrategien

Ein Forscher, der sich früh mit Bearbeitungsstrategien bei Testaufgaben auseinandersetzte, war Barratt (1953). Er befragte 84 Testpersonen retrospektiv nach ihren Bearbeitungsstrategien bei verschiedenen Raumvorstellungs- und Verbalaufgaben, indem er ihnen verschiedene Beschreibungen möglicher Strategien vorlegte. Die anschließend durchgeführten Faktorenanalysen ergaben einen verbalen Faktor sowie drei Raumvorstellungsfaktoren, die inhaltlich den drei „großen“, später bei Lohman (1979) beschriebenen Faktoren *Spatial Relations*, *Visualization* und *Spatial Orientation* entsprechen.

Spatial-Relations-Aufgaben (Rotationsaufgaben) definiert Barratt folgendermaßen: Eine Figur oder Teile einer Figur müssen in einer Ebene oder um eine imaginäre Achse gedreht bzw. rotiert werden, um die Figuren auf Deckungsgleichheit zu überprüfen. 82 von 84 Testpersonen stimmten zu, eine Methode analog der Beschreibung für die Aufgabenbearbeitung gewählt zu haben.

Barratt unterscheidet bei *Spatial-Relations*-Aufgaben zusätzlich zwischen zwei Vorgehensweisen, nämlich zwischen *part approach* und *whole approach* (andere Forscher wie z.B.

Cooper, 1976, 1980 wählten die Bezeichnungen *analytische* und *holistische Strategie*). Beim *part approach* versucht die Testperson, einzelne Teile zum Vergleich von Figuren heranzuziehen, beim *whole approach* werden die Figuren als Ganzes verglichen. Die 43 Testpersonen, die die Aufgaben analytisch (*part approach*) bearbeiteten, schnitten signifikant besser ab. Barratt stellte auch fest, dass einige Testpersonen versuchten, die Vorgabefiguren mit ihnen bekannten Gegenständen in Zusammenhang zu bringen, während andere die Aufgaben abstrakt lösten. Testpersonen, die letzteren Zugang wählten, erzielten bessere Testergebnisse.

Bei Aufgaben zum Faktor *Visualization* legte Barratt den Testpersonen die Beschreibung vor, dass es um die Bestimmung der räumlichen Beziehungen zwischen Objekten geht, wenn ein Objekt oder mehrere Objekte ihre Position im Raum verändern. 76 bis 83 Testpersonen konnten bestätigen, Testaufgaben analog dieser Formulierung bearbeitet zu haben.

Bei *Spatial-Orientation*-Aufgaben muss nach Barratt die Entscheidung getroffen werden, wo sich eine Person im Vergleich zu einem Objekt befindet. Bei einem der zwei vorgegebenen Tests dieser Kategorie konnten sich 58 Testpersonen mit Barratts Beschreibung identifizieren. Barratt unterscheidet bei *Spatial-Orientation*-Tests zusätzlich zwei Varianten der Aufgabenbearbeitung: entweder bewegt die Testperson das Objekt und verändert ihre eigene Position nicht oder umgekehrt, d.h. die Testperson ändert ihre Position, während das Objekt nicht bewegt wird (diese Strategie beschreiben Just & Carpenter, 1985, unter dem Namen *Perspektivwechsel*, siehe Kapitel 2.3.5 „Kognitive Koordinatensysteme“).

Kail, Carter und Pellegrino (1979) legten in ihrer Untersuchung als Testmaterial eine Vorlagefigur (Buchstaben, Zahlen sowie buchstabenähnliche Formen) und ein Abbild der Figur, das entweder eine Rotation (zwischen 0 und 150 Grad) oder eine rotierte Spiegelung der Vorlagefigur war, vor. Circa ein Drittel der weiblichen Testpersonen benötigte zum Rotieren doppelt so viel Zeit wie die übrigen Testpersonen, woraus die Autoren schlossen, dass diese Gruppe eine andere mentale Rotationsstrategie verwendete. Kail et al. unterscheiden daher zwischen zwei generellen Strategien, nämlich einer *holistic strategy*, bei der ein Stimulus als Ganzes rotiert wird (analog dem *whole approach* nach Barratt, 1953, bzw. der *holistischen Strategie* nach Cooper, 1976, 1980) und einer Strategie, bei der Eigenschaften des Reizmaterials in einer iterativen Vorgehensweise separat rotiert werden (ähnlich dem *part approach* nach Barratt bzw. der *analytischen Strategie* nach Cooper).

Auch Lohman (1979) beschäftigt sich in seinem Übersichtsartikel mit Bearbeitungsstrategien und berichtet, dass Testpersonen bei einfacheren Aufgaben eher dieselben Strategien ein-

setzen als bei schwierigeren Aufgaben; weiters stellt er fest, dass mit zunehmender Itemschwierigkeit die Anzahl der Strategien steigt. Vor allem bei komplexen Aufgaben unter Power-Bedingungen herrscht eine größere Variabilität an Strategien als bei weniger komplexen Tests unter Speed-Bedingungen. Lohman fasst zusammen, dass Testpersonen generell bei schwierigen Aufgaben vermehrt auf analytische Strategien (wie z.B. logisches Schlussfolgern) zurückgreifen als auf holistische Manipulationsstrategien.

Bei der Auswahl der Bearbeitungsstrategie spielt auch Begabung eine Rolle: Lohman beschreibt, dass sich besser begabte Testpersonen eine Antwort überlegen, bevor sie die Antwortalternativen durchsehen und dass sie zumeist in Worten umschreiben können, wie sie zu ihrer Antwort gekommen sind. Weniger begabte Testpersonen tendieren dazu, die vorgegebenen Antwortalternativen zu überprüfen und versuchen, die Items unter Zuhilfenahme von Sprache („Verbalisierungen“) zu lösen; sie raten öfter und haben weniger Vertrauen in ihre Antworten.

Lohman unterstreicht den Nutzen retrospektiver Befragungen als ersten Schritt, Hinweise über denkbare Bearbeitungsstrategien zu bekommen; er merkt dennoch kritisch an, dass retrospektiv erhobene Berichte von Testpersonen aufgrund möglicher Verzerrungen hinterfragt werden müssen. Lohman schlägt vor, Testpersonen kurz vor Lösung des Items, währenddessen oder gleich danach zu befragen.

Zusammenfassend stimmen die vorgestellten Arbeiten darin überein, dass die Strategien in Abhängigkeit der Aufgabenart, der Schwierigkeit des Items sowie der Fähigkeit der Person variieren. Konsens herrscht auch bei der groben Unterteilung in analytische und holistische Bearbeitungsweisen.

2.3.2 Studien zum Zusammenhang zwischen Bearbeitungsstrategie und Leistung

Neben der Überprüfung der Existenz verschiedener Bearbeitungsstrategien stellt sich auch die Frage, ob die Anwendung unterschiedlicher Strategien Auswirkungen auf die Leistung haben kann. Wichtige Beiträge zu diesem Thema lieferten Schultz (1991) sowie Kyllonen, Lohman und Woltz (1984), deren Untersuchungsergebnisse in diesem Kapitel dargestellt werden.

Schultz (1991) beschreibt in ihrer Arbeit insgesamt vier Studien. In der ersten Untersuchung ging es der Wissenschaftlerin um die Erforschung verschiedener Strategien. Zu diesem Zweck wurden Testpersonen 30 Sekunden nach der Vorgabe von Raumvorstellungsitems unterbrochen und aufgefordert, niederzuschreiben, welche Gedanken ihnen zur Lösung der Aufgaben durch den Kopf gingen. Durch die Sammlung, Zusammenfassung und Kombination aller Beschreibungen entstanden drei generelle Bearbeitungsstrategien:

- *move object* (MO) = Bewegung des Objekts im Kopf
- *move self* (MS) = Bewegung der (Test-)Person
- *key feature* (KF) = Identifizierung von Schlüsseigenschaften des Objekts und Überprüfung, ob diese Eigenschaften vorhanden, nicht vorhanden oder verändert sind.

Das Ziel der zweiten Studie war, die Kategorisierung in die drei Bearbeitungsstrategien zu evaluieren. Die Testpersonen wurden gebeten, im Anschluss an die Aufgabenbearbeitung immer jene Beschreibung zu markieren, die ihrer Bearbeitungsstrategie am nächsten kam. Die durchgeführten Analysen ergaben wieder die oben genannten Strategien *move object*, *move self* und *key feature*. Schultz stellte bei dieser Untersuchung zusätzlich fest, dass Testpersonen ihre Strategie von Aufgabe zu Aufgabe wechselten, dass viele Testpersonen alle drei Strategien verwendeten, dass keine Testperson nur eine Strategie für alle Aufgaben gebrauchte und dass jedes Item mit unterschiedlichen Strategien bearbeitet wurde.

Schultz führte eine dritte Studie mit sieben Raumvorstellungstests zu zwei Testzeitpunkten mit zehnwöchigem Abstand durch. Drei der sieben Tests wurden bereits in den ersten beiden Studien verwendet, vier neue kamen hinzu. In dieser dritten Studie standen die Überprüfung der Anwendbarkeit der drei Bearbeitungsstrategien bei anderen Raumvorstellungstests sowie die Reliabilität der Strategien über die Zeit im Vordergrund. Nach jeder Testsitzung füllten die Testpersonen den entwickelten Strategiefragebogen (SSQ = *Spatial Strategy Questionnaire*) aus. Schultz berichtete, dass bei fünf der sieben Tests die Strategie, die in der zweiten Testsitzung gewählt wurde, mit jener der ersten Testsitzung konsistent war. Bestimmte Strategien wurden mit größerer Wahrscheinlichkeit für spezifische Aufgaben verwendet und die Verwendung dieser Strategien war positiv mit Leistung korreliert. Die Testpersonen verwendeten keine weiteren Strategien neben den drei genannten.

Mit der vierten Untersuchung zielte Schultz darauf ab, mittels hierarchischer Regressionsanalysen den Beitrag der Bearbeitungsstrategie auf die Raumvorstellungsleistung zu schätzen. Die Testpersonen bearbeiten drei Raumvorstellungstests, die unterschiedliche Facetten von Raumvorstellung erfassen sollten (*Spatial Visualization*-Aufgaben, *Spatial Orientation*-

Aufgaben sowie *Form Matching*-Aufgaben). Die Instruktionen der Tests wurden umgestaltet, so dass sich darin keine Hinweise mehr fanden, wie die Aufgaben zu bearbeiten waren. Für jeden Test füllten die Testpersonen den SSQ mit den drei Bearbeitungskategorien sowie einer vierten Antwortmöglichkeit (sinngemäß: „Aufgabe anders gelöst“) aus.

Schultz konnte signifikante Zusammenhänge zwischen der Strategiewahl und den Leistungen beobachten: Die Verwendung der *move object*-Strategie führte zu besseren Scores bei *Spatial Visualization*-Aufgaben und niedrigeren Scores bei *Spatial Orientation*-Aufgaben, während die Verwendung der *move self*-Strategie mit höheren Scores bei *Spatial Orientation*-Aufgaben und die Verwendung der *key feature*-Strategie mit niedrigeren Scores bei *Spatial Visualization*-Items einherging. Bei *Spatial Visualization*- und *Spatial Orientation*-Aufgaben standen folglich die Verwendung einer Strategie und die Vermeidung einer anderen signifikant mit Leistung in Zusammenhang (nicht aber bei *Form-Matching*-Aufgaben).

Ein anderes interessantes Ergebnis dieser vierten Studie zeigte, dass ca. 20 Prozent der Testpersonen die Aufgaben mit einer anderen Strategie bearbeiteten, als von der Original-Testinstruktion vorgesehen gewesen wäre; das bedeutet, dass für ein Fünftel aller Testpersonen die Aufgaben nicht die Fähigkeit maßen, die sie zu messen vorgaben.

Kyllonen, Lohman und Woltz (1984) versuchten, unterschiedliche Raumvorstellungsleistungen mit Hilfe von Informationsverarbeitungsmodellen zu erklären. Sie legten ihren Annahmen das Schema von Snow (1978, zitiert nach Kyllonen et al., 1984) zu Grunde, das drei Ursachen für individuelle Leistungsunterschiede beschreibt:

- *sequence differences*: Testpersonen unterscheiden sich durch die Reihenfolge, in der sie Bearbeitungsschritte durchführen.
- *route differences*: Testpersonen unterscheiden sich durch die Bearbeitungsschritte, die sie durchführen.
- *summation* oder *strategy differences*: Alle Bearbeitungsschritte sind von Testperson zu Testperson unterschiedlich oder sind innerhalb einer Testperson von Aufgabe zu Aufgabe unterschiedlich.

Die Autoren erwarteten zwei Möglichkeiten bei der Itembearbeitung, nämlich, dass Testpersonen die Reihenfolge variieren, in der sie Bearbeitungsschritte durchführen (*sequence shifting*) und dass sie qualitativ unterschiedliche Bearbeitungsschritte für verschiedene Items (*route shifting*) verwenden.

Bei der empirischen Untersuchung, die Kyllonen et al. aufgrund der kleinen Stichprobengröße von 30 männlichen Testpersonen als vorerst exploratorisch bezeichneten, wurde den Testpersonen eine Figur A zur Betrachtung vorgelegt (*Enkodierung*) und wieder entfernt. Anschließend wurde den Testpersonen eine Figur B präsentiert bzw. bei zwei Drittel der Items zusätzlich eine Figur C (*Synthese*). Diese Figur B (bzw. die Figuren B und C) musste(n) mental mit der Figur A kombiniert werden. Im nächsten Schritt wurde die Figur B (bzw. wurden die Figuren B und C) entfernt und eine Ergebnisfigur präsentiert. Die Testpersonen sollten entscheiden, ob die aus den Figuren A und B (bzw. und C) produzierte Figur mit der präsentierten Ergebnisfigur übereinstimmte oder nicht (*Vergleichsphase*).

(Bei einigen Items wurden die Testpersonen zwischen der *Synthese*- und *Vergleichsphase* aufgefordert, ihr mentales Bild aus den Figuren A und B bzw. C um 90 oder 180 Grad zu drehen und anschließend mit der Ergebnisfigur zu vergleichen. Die Autoren ignorierten diesen *Rotationsschritt* allerdings bei den Analysen).

Beim *Enkodieren* konnten Kyllonen et al. drei verschiedene Strategien ausmachen: Manche Testpersonen merkten sich die Figur A, indem sie die Figur als Set einfacher Eigenschaften speicherten. Einige andere Testpersonen zerlegten die Figuren in Einheiten (Subfiguren) wie Rechtecke und Dreiecke und speicherten diese Zerlegungen. Eine Gruppe von Personen wechselte je nach vorgegebener Figur ihre Strategie zwischen verbalem Benennen und Zerlegen. Die günstigste Strategie in der *Enkodierphase* war – da sie sich als wenig fehleranfällig und schnell erwies – die Zerlegung der Figur in Subfiguren. Das Speichern von Eigenschaften war zwar eine genaue, aber langsame Strategie. Der Strategiewechsel zwischen verbalen Benennungen und Zerlegungen erwies sich als schnell, allerdings als fehleranfällig; wahrscheinlich, weil die verbalen Benennungen nur zu einer ungenauen, vagen Speicherung der Figur führten.

Bei der *Synthesephase* zeigte sich, dass einige Testpersonen alle Figuren (A und B bzw. C) zu einer Figur zusammensetzten und diese speicherten. Andere schafften die Zusammensetzung nicht und speicherten daher die Figuren separat. Die meisten Testpersonen wechselten zwischen diesen beiden Strategien oder zwischen einer der beiden Strategien und einer weiteren, bei der sie von drei vorgegebenen Figuren nur zwei kombinierten und speicherten. Manche Testpersonen verwendeten auch eine zu sehr schlechten Ergebnissen führende „Rückwärtsstrategie“, bei der sie sich von der sichtbaren Figur auf die zuvor präsentierte, nicht mehr sichtbare Figur zurückarbeiteten. Die besten Ergebnisse erzielten Personen mit der Strategie, die Figuren kontinuierlich zusammenzusetzen.

In der *Vergleichsphase* überprüften manche Testpersonen ihr mentales Abbild und die Ergebnisfigur, indem sie eine Eigenschaft nach der anderen verglichen; andere verwendeten größere Einheiten oder wechselten zwischen diesen beiden Strategien. Ein paar Testpersonen verwendeten zwar prinzipiell den Vergleich einzelner Eigenschaften, jedoch nicht bei großen Diskrepanzen zwischen dem mentalen Abbild und der Ergebnisfigur. Hier antworteten die Testpersonen sehr rasch, dass die beiden Figuren unterschiedlich waren. In der Vergleichsphase konnten die Autoren keine signifikanten Leistungsunterschiede je nach Strategiewahl ausmachen.

Zusammenfassend belegen sowohl die Untersuchungen von Schultz (1991) als auch von Kyllonen et al. (1984), dass unterschiedliche, inter- und intraindividuelle Bearbeitungsstrategien bei Raumvorstellungsaufgaben vorliegen und dass sich die Anwendung bzw. die Vermeidung bestimmter Strategien auf die Leistung auswirkt.

2.3.3 Strategien beim *Mental Rotations Test*

Ein in der Raumvorstellungsforschung sehr häufig eingesetztes Testmaterial zur Messung der mentalen Rotationsleistung ist der *Mental Rotations Test (MRT)* von Vandenberg und Kuse (1978), der auf die aus mehreren „leeren“ Würfeln zusammengesetzten, dreidimensionalen Figuren von Shepard und Metzler (1971) zurückgreift. Der *Mental Rotations Test* besteht aus fünf Testreihen mit jeweils vier Aufgaben, bei denen eine Vorlagefigur mit vier anderen Figuren verglichen werden muss. Zwei der vier Antwortalternativen sind Rotationen der Vorlagefigur, zwei Figuren sind Distraktoren (zum Teil Spiegelungen, zum Teil andere Figuren). Der Frage, ob die Items des *MRT* auch mit alternativen Bearbeitungsstrategien außer mentaler Rotation gelöst werden können, widmeten sich Geiser, Lehmann und Eid (2006).

Die Autoren verwendeten in ihrer Studie die Form A einer neu gezeichneten, deutschen Version des *MRT* von Peters, Laeng, Latham, Jackson, Zaiyouna und Richardson (1995) mit 24 Items, die in zwei Testteilen zu je zwölf Items und einer Bearbeitungszeit von je drei Minuten pro Testteil vorgegeben wurden. Bei der „klassischen“ Testauswertung nach Vandenberg und Kuse wird ein Item dann als „gelöst“ erachtet und mit einem Score von 1 versehen, wenn beide rotierten Versionen der Vorlagefigur markiert wurden. Andere Auswertungsmöglichkeiten sind einerseits, jede richtige Markierung einzeln zu zählen, so dass sich der Gesamtscore verdoppelt, die Ratewahrscheinlichkeit aber auch steigt; andererseits können so genannte „ratio scores“ berechnet werden, bei denen die Anzahl der richtig bearbeiteten

Items durch die Anzahl der bearbeiteten Items dividiert wird, was der Speed-Komponente des Tests entgegenwirkt.

Geiser et al. (2006) analysierten 1695 *MRT*-Datensätze von Schülern und Studenten mittels Latent-Class-Analyse (LCA, z.B. Rost, 1988). Eine Fünf-Klassen-Lösung beschrieb die Daten beider Testteile am besten. Die Autoren identifizierten jeweils drei Gruppen, die sich in der Anzahl der bearbeiteten Items innerhalb des Zeitlimits unterschieden sowie eine Gruppe, die sich aus Testpersonen mit schwachen Leistungen zusammensetzte. Die letzte Klasse zeigte ein spezielles Antwortmuster, nämlich niedrige Lösungswahrscheinlichkeiten für die Items, deren Distraktoren Spiegelbilder waren und hohe Lösungswahrscheinlichkeiten für die Items, deren Distraktoren andere Figuren waren. Geiser et al. schlossen daraus, dass diese Gruppe eine andere Strategie als Rotation anwendete – nämlich eine analytische Strategie, die den Testpersonen ermöglichte, beim Vergleich der Antwortalternativen jene Figuren auszuschließen, die keine Spiegelbilder waren.

Die Ergebnisse von Geiser et al. (2006) zeigen, dass auch beim *MRT* unterschiedliche Strategien angewendet werden. Bei circa einem Drittel der Testaufgaben führen andere als Rotationsstrategien zur erfolgreichen Lösung. Das bedeutet, dass der *MRT* nicht für alle Testpersonen als Messung der mentalen Rotationsleistung interpretiert werden kann. In die gleiche Richtung gehen die Schlussfolgerungen von Schultz (1991, siehe oben), die – für anderes Testmaterial – angibt, dass 20 Prozent der Testpersonen nach Weglassen der Instruktion Items anders als intendiert bearbeiten.

2.3.4 Strategien bei der Bearbeitung von Würfelaufgaben

Bei der Erforschung räumlicher Fähigkeiten spielen neben den Figuren von Shepard und Metzler (1971) Würfelaufgaben eine große Rolle; ihnen soll dieses Kapitel gewidmet sein.

Eine Wissenschaftlerin, die sich mit der Analyse von Würfelaufgaben beschäftigte, ist Putz-Osterloh. Sie analysierte (1977) die Items des Subtests „Würfelaufgaben“ aus dem *IST* und *IST-70* von Amthauer (1953, 1970) und fand heraus, dass es drei Arten von Würfeln gibt, die unterschiedliche Bearbeitungsstrategien erfordern, nämlich *Flächenwürfel*, *Flächenwürfel plus* und *Raumwürfel*.

Flächenwürfel können gelöst werden, indem die drei sichtbaren Flächen des Vorgabewürfels mit den sichtbaren Flächen der fünf Würfelalternativen verglichen werden. Bei *Flächenwürfeln plus* muss neben den sichtbaren Flächen auch die Relation zwischen zwei Flächen

überprüft werden, um zu einer richtigen Lösung zu kommen. *Raumwürfel* können nur dann korrekt bearbeitet werden, wenn sich Testpersonen durch mentale Rotation vorstellen, wie beim Vorgabewürfel neben zwei sichtbaren Flächen eine dritte, in der Darstellung „unsichtbare“ Fläche erscheint und eine bislang sichtbare Fläche verschwindet.

Putz-Osterloh vermutete, dass nur *Raumwürfel* Raumvorstellung erfassen, da *Flächenwürfel* mit Flächenstrategien erfolgreich bearbeitet werden können. In empirischen Untersuchungen überprüften Putz-Osterloh (1977) sowie Putz-Osterloh und Lürer (1979) das Blickverhalten von Testpersonen bei der Bearbeitung von Würfelaufgaben und trafen die Unterscheidung in Raum- und Flächenstrategen. Erstere sind Personen, die eine holistische Bearbeitungsstrategie wählen und daher längere Fixationszeiten bei der Würfelbetrachtung aufweisen, letztere verwenden eine analytische Strategie mit Aspekten logisch-schlussfolgernden Denkens. Die längeren Zeiten wurden mit dem Hinweis begründet, dass Rotationen die aufwendigeren mentalen Transformationen sind. Die beobachteten Strategien sind mit den Ergebnissen von Barratt (1953) bzw. Cooper (1976, 1980) vergleichbar.

Mit Hilfe probabilistischer Testmodelle untersuchten Köller, Rost und Köller (1994) einige Hypothesen, die auf den Ergebnissen von Putz-Osterloh (1977) sowie Putz-Osterloh und Lürer (1979) aufbauen. Eine erste empirische Untersuchung mit dem Mixed-Rasch-Modell (Rost, 1990) hatte das Ziel, analytische und holistische Strategien bei Würfelaufgaben zu identifizieren. Die Ergebnisse zeigten, dass eine Lösung mit zwei Klassen, deren Mitglieder Köller et al. aufgrund ihres Lösungsverhaltens „Analytiker“ bzw. „Holistiker“ nannten, am besten zu den Daten passte.

In einer zweiten Untersuchung, die mittels Latent-Class-Analyse (Rost, 1988) ausgewertet wurde, gingen Köller et al. davon aus, dass Personen mit holistischen Bearbeitungsstrategien bei *Raumwürfeln* höhere Lösungswahrscheinlichkeiten zeigen, bei *Flächenwürfeln* und *Flächenwürfeln plus* aber niedrigere Lösungswahrscheinlichkeiten, da bei der kognitiv aufwendigeren holistischen Strategie leichter Fehler im Transformationsprozess passieren. Sie erwarteten bezüglich der Personen mit analytischen Strategien, dass diese bei *Raumwürfeln* niedrigere Lösungswahrscheinlichkeiten erreichen, bei *Flächenwürfeln* und *Flächenwürfeln plus* aber höhere Wahrscheinlichkeiten. Zusätzlich nahmen die Autoren an, dass es eine Personengruppe gibt, die bei *Flächenwürfeln plus* systematisch falsch antwortet, da sie die Relationen der Würfelflächen zueinander außer Acht lässt. Die beste Modellanpassung zeigte eine Lösung mit sechs Klassen. Köller et al. konnten neben drei Gruppen mit motivationalen Schwierigkeiten (Personen, die bei der Aufgabenbearbeitung rieten, den Test vorzeitig abbrechen bzw. zu langsam arbeiteten) drei Gruppen mit unterschiedlichen Bearbeitungsstrategien finden:

- Flächenstrategen, die *Flächenwürfel* und *Flächenwürfel plus* größtenteils richtig lösten, aber auch eine Flächenstrategie bei *Raumwürfeln* verwendeten;
- reine Flächenstrategen, die die Flächenrelationen nicht beachteten und daher bei *Flächenwürfel plus* systematisch Fehler machten und auch *Raumwürfel* fehlerhaft bearbeiteten sowie
- eine Gruppe von strategieflexiblen Personen, die aber keine holistische Strategie für die *Raumwürfel* verwendete.

Personen mit holistischen Strategien vermuteten die Autoren am ehesten in der Gruppe der Strategieflexiblen. Köller et al. merkten kritisch an, dass bei ihren Stichproben mit Schülerinnen und Schülern zwischen 13 und 16 Jahren eventuell kognitive Prozesse zum Lösen der *Raumwürfel* noch nicht entsprechend entwickelt waren.

Zusammenfassend zeigen die Ergebnisse von Putz-Osterloh (1977), Putz-Osterloh und Lürer (1979) sowie Köller et al. (1994), dass es unterschiedliche Arten von Würfelaufgaben gibt, die unterschiedlich bearbeitet werden (analytisch versus holistisch). Die Lösungswahrscheinlichkeit variiert dabei in Abhängigkeit der Würfelart und der angewendeten Strategie.

2.3.5 Kognitive Koordinatensysteme

Die Arbeiten von Just und Carpenter (1976, 1985), die in diesem Kapitel ausführlich dargestellt werden, bilden einen zentralen Stellenwert bei der Erforschung von Bearbeitungsstrategien.

Die beiden Wissenschaftler verwendeten als Forschungsmaterial häufig Würfelaufgaben, bei denen zwei Würfel mit jeweils einem Buchstaben oder einer Zahl pro Würfelfläche in unterschiedlichen Ansichten präsentiert wurden und die Testpersonen entscheiden mussten, ob die Würfel einander entsprachen oder nicht. 1976 beschrieben sie drei Prozesse, die bei der Bearbeitung von Würfelaufgaben ablaufen: An erster Stelle steht ein *Suchprozess*, bei dem zusammenpassende Buchstaben (oder jeweils auch Zahlen) auf beiden Würfeln gesucht werden. Anschließend folgt die *Transformations- und Vergleichsphase* (von den Autoren auch *Rotationsphase* genannt), in der ein Buchstabe mental rotiert wird, um ihn in Kongruenz mit dem gleichen Buchstaben auf dem zweiten Würfel zu bringen. Sind die beiden Buchstaben beim anschließenden Vergleich nicht deckungsgleich, wird der Buchstabe weiter rotiert und wieder verglichen, bis Deckungsgleichheit erreicht ist. Als letztes folgt die *Bestäti-*

gungsphase, in der überprüft wird, ob alle übrigen Buchstaben nach Anwendung derselben Transformationen zum jeweiligen „Partner“ auf dem zweiten Würfel passen.

1985 publizierten Just und Carpenter in einem umfangreichen Artikel Theorien darüber, wie Menschen Raumvorstellungsaufgaben bearbeiten, wie sie Reize mental repräsentieren und welche Strategien sie bei verschiedenen Aufgabenstellungen anwenden. Die Autoren wollten vor allem auch Erklärungen für die Unterschiede zwischen Personen mit guten und schlechten Raumvorstellungsfähigkeiten finden. Ausgangspunkt ihrer Überlegungen war, dass Menschen räumliche Informationen in kognitive Koordinatensysteme einbetten. Diese kognitiven Koordinatensysteme können sich von Person zu Person unterscheiden und Auswirkungen auf die Bearbeitungsstrategie – den qualitativen Aspekt – und die Leistung, den quantitativen Aspekt, haben. Just und Carpenter (1985) formulierten vier mögliche Strategien bei der Bearbeitung von Würfelaufgaben:

1. Bei der *mentalen Rotation um Standardachsen* rotiert die Testperson den Würfel um eine vertikale oder horizontale Achse, die durch den Mittelpunkt des Würfels geht.
2. Bei der *mentalen Rotation um eine aufgabenspezifische Achse* rotiert die Testperson den Würfel um eine willkürliche Achse, die einen möglichst kurzen Rotationsweg mit sich bringt.
3. Beim *Vergleich orientierungsfreier Beschreibungen* kodiert die Testperson die Relationen der Buchstaben auf einem Würfel (z.B.: „Die Unterseite des P zeigt zur Oberseite des L.“) und vergleicht sie mit den Buchstabenrelationen auf dem anderen Würfel – ohne Referenz zu einem größeren Bezugsrahmen außerhalb des Würfels. Die Repräsentationen der beiden Würfel werden innerhalb eines objektspezifischen kognitiven Koordinatensystems gebildet und ohne mentale Transformation direkt miteinander verglichen.
4. Beim *Perspektivwechsel* wechselt die Testperson ihre Position, während die des Objekts konstant bleibt und stellt sich vor, wie das Objekt aussieht, wenn sie es beispielsweise von unten betrachten würde. Objekt und Testperson befinden sich innerhalb des kognitiven Koordinatensystems, das Objekt stellt den Ursprung dar. Just und Carpenter (1985) nahmen an, dass der Perspektivwechsel vor allem bei großen, nicht bewegten Objekten wie einem Haus angewendet wird. So könnte die Strategie bei der Frage nach der Anzahl der Fenster im Haus sein, mental durch das Haus zu spazieren und nicht das Haus selbst zu rotieren.

Die Wissenschaftler (Just & Carpenter, 1985) überprüften ihre Theorien auch in drei verschiedenen Experimenten: Im ersten Experiment bearbeiteten vier Personen mit gutem und

vier Personen mit schlechtem Raumvorstellungsvermögen Würfelaufgaben. Es wurden die Blickbewegungen aufgezeichnet, die Antwortzeiten gemessen und die Personen zu ihren verwendeten Strategien befragt. Die Auswertungen ergaben, dass drei der vier Testpersonen mit guten Raumvorstellungsfähigkeiten bei den Aufgaben, bei denen es möglich war, eine Rotationsstrategie um willkürliche Achsen verwendeten; die vierte Testperson bearbeitete die Aufgaben mit Hilfe des Vergleichs orientierungsfreier Beschreibungen.

Die Testpersonen mit weniger guten Raumvorstellungsfähigkeiten beschrieben als Strategie Rotationen um Standardachsen; sie benötigten wesentlich mehr Zeit für die Aufgabenbearbeitung als die guten Raumvorsteller. Die Bearbeitungszeiten der Testperson, die den Vergleich orientierungsfreier Beschreibungen verwendete, waren länger als die der guten Raumvorsteller, aber kürzer als die der schlechten Raumvorsteller.

Die Strategie des Perspektivwechsels wurde von keiner Testperson verwendet.

Mit Hilfe der erhobenen Daten entwickelten Just und Carpenter (1985) – getrennt für die drei oben beschriebenen Prozesse bei der Würfelbearbeitung – Simulationsmodelle für Personen mit guten und Personen mit schlechten Raumvorstellungsfähigkeiten:

In der *Suchphase* merkten sich die Testpersonen meistens den Buchstaben auf der Vorderseite des Vorgabewürfels und suchten diesen auf dem Vergleichswürfel. War dieser Buchstabe auf keiner Fläche des Vergleichswürfels abgebildet, suchten die Testpersonen den Buchstaben, der auf der rechten Würfelfläche oder der Oberseite des Vorgabewürfels abgebildet war, auf dem Vergleichswürfel. In der *Suchphase* unterschieden sich die Personen mit guten und schlechten Raumvorstellungsleistungen nicht voneinander.

Nach erfolgreicher Entdeckung zweier gleicher Buchstaben wurden in der *Rotationsphase* die Würfelflächen rotiert. Just und Carpenter (1985) gingen in Anlehnung an Ergebnisse von Cooper und Shepard (1973) sowie aufgrund eigener Forschungsarbeiten (Just & Carpenter, 1976) davon aus, dass Objekte bei mentaler Rotation Zwischenschritte durchlaufen. Die Personen mit schlechtem Raumvorstellungsvermögen rotierten die Würfelflächen um Standardachsen und in zwei getrennten Prozessen, sodass einerseits die Position der Würfelfläche, andererseits die Ausrichtung der Buchstaben übereinstimmte. Die Personen mit gutem Raumvorstellungsvermögen rotierten bei Aufgaben, bei denen es möglich war, um andere als die Standardachsen. Sie benötigten auch nur einen Rotationsprozess, um die Position der Würfelflächen und die Ausrichtung der Buchstaben anzugleichen. Jeder Zwischenschritt wurde in der so genannten *Transformationsliste* gespeichert.

Just und Carpenter (1985) nahmen in ihrem Modell für die schlechten Raumvorsteller Schrittgrößen von 15 Grad an, für die guten Raumvorsteller Schrittgrößen von 30 Grad. Dieses Verhältnis von 1:2 entsprach dem Verhältnis der Rotationszeiten von schlechten und guten Raumvorstellern – gute Raumvorsteller rotierten doppelt so schnell.

In der *Bestätigungsphase* versuchten die Testpersonen, nachdem sich die ersten beiden Buchstaben auf der gleichen Würfel­fläche befanden, die gespeicherten Rotationen aus der Transformationsliste auf die anderen Buchstaben anzuwenden. Waren alle auf dem Vorgabe- und Vergleichswürfel abgebildeten Buchstaben überprüft, wurden die Buchstaben untersucht, die nur auf einem Würfel zu sehen waren. Bei gleichen Würfeln mussten diese Buchstaben nach Anwendung der Transformationen auf einer verdeckten Fläche verschwinden. Nach Durchführung aller Transformationen wurden die Positionen und Ausrichtungen der Buchstaben des Vorgabe- und Vergleichswürfels verglichen. Die Würfel wurden als „verschieden“ eingestuft, wenn nicht alle Buchstabenpositionen und Ausrichtungen identisch waren. Die Anwendung der Transformationen auf alle übrigen Buchstaben lief in der Gruppe der guten Raumvorsteller wesentlich rascher ab als in der Gruppe der schlechten Raumvorsteller.

Bei der genauen Analyse der Fehler, die den Testpersonen unterliefen, fanden Just und Carpenter (1985) heraus, dass sowohl den guten, als auch den schlechten Raumvorstellern die meisten Fehler in der *Bestätigungsphase* unterliefen. Die schlechten Raumvorsteller vergaßen beispielsweise, alle notwendigen Transformationen auf einen Buchstaben anzuwenden; oder sie vergaßen, die Transformationen für alle Buchstaben durchzuführen, weil sie sich nicht merkten, welche Buchstaben sie schon überprüft hatten und welche nicht. Der Gruppe unterliefen auch Fehler, wenn ein Buchstabe bei der Rotation auf eine sichtbare Fläche zwischenzeitlich auf einer verdeckten Fläche abgebildet war.

Bei Aufgaben, bei denen die Würfel verschieden waren, variierten die Reaktionszeiten unsystematisch. Ein Grund dafür war die unterschiedliche Vorgehensweise der Testpersonen: Manchmal beendeten sie die Aufgabenbearbeitung gleich nach Entdeckung einer Inkonsistenz, manchmal wiederholten sie die Bearbeitung der Aufgabe, manchmal probierten sie andere Kippbewegungen oder starteten mit der Rotation eines anderen Buchstabens.

Ein zweites Experiment bestätigte die Ergebnisse des ersten Experiments, nämlich dass gute Raumvorsteller kürzere Bewegungsbahnen um willkürliche Achsen wählen und schneller rotieren. In der *Bestätigungsphase* unterliefen beiden Gruppen die meisten Fehler. Als

Strategien kamen wieder *mentale Rotationen* und der *Vergleich orientierungsfreier Beschreibungen* vor.

In einem dritten Experiment bearbeiteten vier Personen mit gutem und drei Personen mit schlechtem Raumvorstellungsvermögen Aufgaben von Shepard und Metzler (1971). Just und Carpenter (1985) hypothesisierten, dass diese Aufgaben bei allen Testpersonen ähnliche Rotationsstrategien hervorrufen. Die Autoren nahmen an, dass die schlechten Raumvorsteller langsamer rotieren als die guten und Zwischenergebnisse schlechter behalten, sodass sie einige Prozesse wiederholt durchführen müssen. Die Wissenschaftler konnten diese Annahmen nach Analyse der Daten bestätigen: schlechte Raumvorsteller brauchen länger, da sie langsamer rotierten und weniger effizient beim Behalten von Zwischenergebnissen sind. Mit steigendem Winkel der Ungleichheit stieg auch die Reaktionszeit.

Die wichtigsten Erkenntnisse der Arbeiten von Just und Carpenter (1976, 1985) sind, dass es vier mögliche Strategien bei der Bearbeitung von Würfelaufgaben gibt, nämlich Rotationen um willkürliche oder Standardachsen, den Vergleich orientierungsfreier Beschreibungen, der mit einer analytischen Strategie verglichen werden kann sowie den Perspektivwechsel, der allerdings in den beschriebenen Experimenten von keiner Testperson angewendet wurde. Hier kann auch ein Querverweis auf die Arbeiten von Schultz (1991; siehe oben) und ihre berichteten Strategien gezogen werden: *move object* entspricht den Rotationsstrategien, *move self* der Perspektivwechselstrategie und *key feature* dem Vergleich orientierungsfreier Beschreibungen.

Weiters beschreiben Just und Carpenter (1976, 1985) drei Phasen bei der Bearbeitung von Würfelaufgaben: die Such-, die Rotations- und die Bestätigungsphase. Gute und schlechte Raumvorsteller unterscheiden sich in der Rotations- und der Bestätigungsphase; begabte Personen verwenden kognitive Koordinatensysteme, die Rotationen in größeren Schritten und um willkürliche Achsen erlauben im Vergleich zu kleineren Schritten und Standardachsen bei weniger begabten Personen. Gute Raumvorsteller arbeiten auch in der Bestätigungsphase schneller.

2.4 Geschlechtsunterschiede bei Raumvorstellung

In der wissenschaftlichen Literatur zum Thema Raumvorstellung gelten Geschlechtsunterschiede bei räumlichen Leistungen zugunsten der Männer als unumstritten und gesichert. In diesem Kapitel werden Untersuchungen zur Größe der Geschlechtsunterschiede sowie Ursachen für Geschlechtsunterschiede näher durchleuchtet. In weiterer Folge geht es darum, wie und ob sich Geschlechtsunterschiede durch verschiedene Arten der Testvorgabe und -auswertung, im Verlauf des Alters, durch Training und die Wahl der Bearbeitungsstrategie verändern. Abschließend sollen Forschungsansätze, die Anlage- und Umweltaspekte berücksichtigen, diskutiert werden.

2.4.1 Die Größe der Geschlechtsunterschiede

Anfangs soll erwähnt werden, dass Hyde (1981) zum Thema „Größe der Geschlechtsunterschiede“ auf einen Interpretationsfehler hinweist: Die Aussage von Maccoby und Jacklin (1974), dass der Geschlechtsunterschied zugunsten der Männer gut abgesichert ist, wird immer wieder in die Richtung interpretiert, dass der Geschlechtsunterschied zugunsten der Männer groß ist. Hyde hält in seiner Publikation fest, dass der Mittelwertsunterschied bei Raumvorstellungsleistungen zwischen Männern und Frauen kleiner als eine halbe Standardabweichung ist und dass das Geschlecht nur ca. fünf Prozent der Varianz erklärt. Voyer, Voyer und Bryden (1995) meinen, dass Hyde möglicherweise den Prozentsatz der Varianz, den das Geschlecht erklärt, unterschätzt habe, da er nicht bedachte, dass verschiedene Raumvorstellungstests unterschiedlich große Geschlechtsunterschiede hervorrufen können. Wichtige Erkenntnisse zur Größe der Geschlechtsunterschiede lieferten die Metaanalysen von Linn und Petersen (1985) sowie Voyer, Voyer und Bryden (1995), deren Ergebnisse im Folgenden ausführlicher beschrieben werden.

Linn und Petersen (1985) stützen sich in ihrer Metaanalyse auf vier Forschungsperspektiven:

- die differentielle Perspektive, bei der es um den Vergleich von Raumvorstellungsleistungen in verschiedenen Personengruppen geht,
- die psychometrische Perspektive, bei der Korrelationen zwischen verschiedenen Raumvorstellungsaufgaben berechnet und analysiert werden,
- die kognitive Perspektive, die die Identifizierung der Prozesse beinhaltet, die zur Lösung von Raumvorstellungsaufgaben notwendig sind und

- die strategischen Perspektive, bei der es um qualitativ unterschiedliche Strategien bei der Aufgabenbearbeitung geht.

Die differentielle Perspektive kommt insofern zum Tragen, als es bei der Metaanalyse um den Mittelwertsunterschied zwischen den Raumvorstellungsscores von Männern und Frauen geht. Die psychometrische und die kognitive Perspektive helfen, eine Einteilung der in die Metaanalyse eingehenden Studien in homogene Gruppen vorzunehmen. Die strategische Perspektive hilft, Leistungsvariationen bei manchen Raumvorstellungsaufgaben zu erklären.

Linn und Petersen berücksichtigten in ihrer Metaanalyse Ergebnisse aus Studien zwischen 1974 und 1982 mit mehr als 200 Effektgrößen, von denen 172 in die Metaanalyse eingingen. Sie bezogen auch nicht veröffentlichte Studien und Studien zu anderen Themen als Geschlechtsunterschieden, bei denen aber Geschlechtsunterschiede berichtet wurden, ein, um das sogenannte „file drawer“-Problem auszugleichen: Nach Rosenthal (1979) bleiben Studien mit nicht signifikanten Ergebnissen in der Schublade, was zur Folge hat, dass veröffentlichte Studien in Richtung signifikanter Effekte verzerrt sind.

Linn und Petersen testeten die 172 Effektgrößen auf Homogenität, um sicherzustellen, dass sie sich auf eine Population stützen. Homogene Effektgrößen zeigen an, dass die in die Metaanalyse einbezogenen Studien als Wiederholungen jeder anderen Untersuchung gewertet werden können. Da die Effektgrößen nicht homogen waren, teilten Linn und Petersen diese in die drei im Kapitel 2.2.1.3 beschriebenen Kategorien *Spatial Perception*, *Mental Rotation* und *Spatial Visualization*. Nach neuerlicher Überprüfung der Effektgrößen auf Homogenität innerhalb der Kategorien mussten Linn und Petersen in ersteren beiden weitere Unterteilungen vornehmen.

Zur Kategorie *Spatial Perception* passten 62 Effektgrößen, die wegen Heterogenität in drei Altersklassen unterteilt wurden (unter 13 Jahre, 13 bis 18 Jahre sowie über 18 Jahre); die mittlere Altersklasse brachte homogene, die anderen beiden Gruppen nahezu homogene Effektgrößen. Als Ergebnis in dieser Kategorie berichteten Linn und Petersen von Geschlechtsunterschieden zugunsten der Männer, die ab dem Alter von acht Jahren auftreten und über die Lebensspanne hinweg bestehen bleiben. In den beiden Gruppen der unter 18-Jährigen schnitten Männer jeweils um eine Drittel Standardabweichung besser ab als Frauen (gewichtete Schätzung der Effektgröße = 0,37), in der Gruppe der über 18-Jährigen um zwei Drittel einer Standardabweichung (gewichtete Schätzung der Effektgröße = 0,64; signifikant); die mittlere Effektgröße über alle Altersgruppen hinweg betrug 0,44.

In die Kategorie *Mental Rotation* gingen 29 Effektgrößen ein, die auch nach Teilung in die drei Altersklassen nicht homogen waren, da die verwendeten Tests systematisch verschiedenen große Geschlechtsunterschiede erzeugten. Die Autorinnen teilten daher die Effektgrößen in Abhängigkeit der vorgegebenen Tests in zwei Gruppen, die zu homogenen bzw. nahezu homogenen Effektgrößen führten. Eine Gruppe beinhaltete den *MRT* von Vandenberg und Kuse (1978), die andere Gruppe drei Tests, darunter auch die *Primary Mental Abilities (PMA) space* (Thurstone & Thurstone, 1941, zitiert nach Linn & Petersen, 1985). Linn und Petersen fanden in beiden Gruppen über alle Altersstufen hinweg ab dem Zeitpunkt, ab dem die Tests eingesetzt wurden (ca. 13 Jahre beim *MRT* und zehn bis elf Jahre bei *PMA space*), Geschlechtsunterschiede zugunsten der Männer. Bei den *PMA space*-Aufgaben schnitten Männer um eine Viertel Standardabweichung besser ab als Frauen (gewichtete Schätzung der Effektgröße = 0,26), bei der Testversion von Vandenberg und Kuse um eine Standardabweichung (gewichtete Schätzung der Effektgröße = 0,94); die mittlere Effektgröße betrug 0,73.

In der Kategorie *Spatial Visualization* gingen 81 homogene Effektgrößen in die Metaanalyse ein. Als Ergebnis zeigten sich keine signifikanten Geschlechtsunterschiede und keine Unterschiede über die Lebensspanne hinweg (gewichtete Schätzung der Effektgröße = 0,13).

Die zweite große Metaanalyse zu Geschlechtsunterschieden bei Raumvorstellungsleistungen, die die Ergebnisse von Linn und Petersen im Großen und Ganzen untermauerte und zum Teil noch weiter differenzierte, publizierten Voyer, Voyer und Bryden (1995). Sie analysierten 286 (aus insgesamt 310) Effektgrößen verschiedener Raumvorstellungsmessungen aus Studien zwischen 1974 und 1993. Als Messung der Effektgröße verwendeten sie Cohens *d* (Cohen, 1977), das entspricht in ihrer Metaanalyse dem standardisierten Mittelwertsunterschied zwischen Männern und Frauen. Voyer et al. berücksichtigten auch das „file drawer“-Problem (Rosenthal, 1979; siehe oben), indem sie eine von Rosenthal (1980) vorgeschlagene Berechnungsformel für die „fail safe number“ verwendeten: das ist die Anzahl der Studien mit einer durchschnittlichen Effektgröße von Null, die erforderlich wäre, um die Signifikanz der Ergebnisse auf dem Level von $p = 0,05$ auszusetzen. Die „fail safe number“ ergab, dass knapp 180 000 Studien mit nicht-signifikanten Ergebnissen notwendig wären - eine Anzahl, die (wahrscheinlich) nicht erreicht werden kann.

Voyer et al. gingen in Anlehnung an Linn und Petersen (1985) davon aus, dass Geschlechtsunterschiede bei *mentalen Rotationstests* sowie bei Tests mit *Spatial Perception*-Aufgaben auftreten. Außerdem interessierten sie sich für die zeitliche Stabilität der Ge-

schlechtsunterschiede und hypothetisierten, dass bei signifikanten Geschlechtsunterschieden die Größe der Unterschiede nicht mit dem Untersuchungsjahr korreliert.

Die Analyse der 286 Effektgrößen, die allerdings nicht homogen waren, ergab signifikante Geschlechtsunterschiede zugunsten der Männer (mittleres gewichtetes d von 0,37; $p < 0,01$). Um Homogenität zu erreichen, teilten Voyer et al. die Effektgrößen in die drei von Linn und Petersen (1985) vorgeschlagenen Kategorien und konnten mit diesem Schritt deren Ergebnisse replizieren: Voyer et al. fanden die größte (und signifikante) Effektgröße von $d = 0,56$ ($p < 0,05$) in der *Mental Rotation*-Kategorie, die kleinste (und nicht-signifikante) Effektgröße von $d = 0,19$ in der *Spatial Visualization*-Kategorie. In der Kategorie *Spatial Perception* ergab die Effektgröße ein d von 0,44 (signifikant, $p < 0,05$). Die Einordnung der Tests analog zu Linn und Petersen reduzierte zwar die Heterogenität, führte allerdings nicht zu homogenen Effektgrößen, so dass Voyer et al. weitere Teilungen vornehmen mussten.

Als nächsten Schritt gliederten Voyer et al. die Effektgrößen innerhalb der drei Kategorien in die Altersgruppen „unter 13 Jahre“, „zwischen 13 und 18 Jahre“ und „älter als 18 Jahre“. Die Effektgrößen stiegen mit dem Alter der Testpersonen und waren nur in ein paar Subgruppen homogen bzw. nahezu homogen: Für die Gruppe der unter 13-Jährigen in allen drei Kategorien und für die Gruppe der 13- bis 18-Jährigen in der Kategorie *Spatial Perception*. Da die Effektgrößen weiterhin in vielen Gruppierungen heterogen waren, betrachteten Voyer et al. jeden Raumvorstellungstest als Operationalisierung einer Raumvorstellungskomponente. Die Autoren entschieden sich auch für diesen Weg, da die Testzuteilung von Linn und Petersen in die drei Kategorien ihrer Ansicht nach nicht für alle Tests geglückt war.

Bei sechs von insgesamt zwölf Tests waren die Effektgrößen homogen. In Tabelle 2 sind die Tests mit homogenen Effektgrößen samt Autoren und einer kurzen inhaltlichen Beschreibung aufgelistet. Weiters kann der Tabelle die Kategorie nach Linn und Petersen (1985) entnommen werden (MR = *Mental Rotation*, SV = *Spatial Visualization*) sowie ob signifikante Geschlechtsunterschiede vorlagen oder nicht.

Tabelle 2: Sechs Tests mit homogenen Effektgrößen samt Autoren, einer kurzen Testbeschreibung, der Kategoriezugehörigkeit (Kat.; MR = *Mental Rotation*, SV = *Spatial Visualization*) sowie den Ergebnissen bei der Prüfung auf Geschlechtsunterschiede (GU).

Test	Autor	Testinhalt	Kat.	GU
Spatial Relations Subtest aus dem Primary Mental Abilities Test	Thurstone, 1958 (zit. n. Voyer et al. 1995)	mentale Rotation 2-dimensionaler buchstabenähnlicher Objekte	MR	sign.
Cards Rotation Test	Ekstrom, French & Harman, 1976	mentale Rotation 2-dimensionaler Objekte	MR	sign.
Generic mental rotation tasks (Zusammenfassung verschiedener Rotationsaufgaben)		verschiedene Varianten der Figuren von Shepard und Metzler (1971)	MR	sign.
Paper Form Board	Likert & Quasha, 1941 (zit. n. Voyer et al., 1995)	Erkennen, welche 2-dimensionale Figur aus einzelnen Teilen entstehen wird	SV	sign.
Spatial Relations Subtest aus dem DAT	Bennett, Seshore & Wesman, 1947 (zit. n. Voyer et al., 1995)	Erkennen, wie eine nicht gefaltete Figur aussieht, wenn sie zusammengefoldet wird	SV	n.s.
Paper Folding	Ekstrom et al. 1976	Erkennen, welches auseinander gefaltete Papierstück dasselbe wie die zusammengefoldete Vorgabe mit ausgestanzten Löchern ist	SV	n.s.

Bei den übrigen sechs Tests waren die Effektgrößen heterogen, sodass weitere Teilungen notwendig waren. In Tabelle 3 sind die sechs Tests mit heterogenen Effektgrößen samt Autoren und einer kurzen inhaltlichen Beschreibung aufgelistet. Weiters kann der Tabelle die Kategorie nach Linn und Petersen entnommen werden (MR = *Mental Rotation*, SP = *Spatial Perception*, SV = *Spatial Visualization*), die verschiedenen Teilungen, um homogene Effektgrößen zu erreichen sowie das Ergebnis dieser Teilungen. Schließlich ist in der Tabelle erfasst, ob bzw. in welchen Untergruppen die Geschlechtsunterschiede signifikant waren oder nicht.

Tabelle 3: Sechs Tests mit heterogenen Effektgrößen samt Autoren, einer kurzen Testbeschreibung, der Kategoriezugehörigkeit (Kat.; MR = *Mental Rotation*, SP = *Spatial Perception*, SV = *Spatial Visualization*), den Teilungskriterien samt Ergebnis sowie den Ergebnissen bei Prüfung auf Geschlechtsunterschiede (GU).

Test	Autor	Testinhalt	Kat.	Teilung in	Ergebnis Teilung	GU
Mental Rotations Test	Vandenberg & Kuse, 1978	mentale Rotation 3-dimensionaler Objekte	MR	Auswertungsmodus: Score = 20, Score = 40, andere Scores	hom.	sign., am größten bei 20, am kleinsten bei „andere Scores“
rod-and-frame test	Witkin & Asch, 1948 (zit. n. Voyer et al., 1995)	eine Stange trotz ablenkender Informationen in die Vertikale bringen	SP	3 Altersgruppen	homogen bzw. nahezu hom.	sign. bei 13-18
Water Level Test	Piaget & Inhelder, 1956 (zit. n. Voyer et al., 1995)	die Ausrichtung einer Flüssigkeit in einem geeigneten Behälter erkennen	SP	verschied. Auswertungskriterien (Abweichungsmessung in Grad od. andere)	hom. bzw. nahezu hom.	teilw. sign., teilw. nicht, je nach Auswertungskrit.
Embedded Figures Test (verschiedene Versionen)	Witkin, 1950 (zit. n. Voyer et al. 1995) und Ekstrom et al., 1976	eine in ein komplexes Muster eingebettete Figur finden	SV	Einzeltest, Gruppentest, Kindertest	hom. bzw. nahezu hom.	Einzel- und Kinder: n.s., Gruppe: sign.
Identical Blocks Test	Stafford, 1961 (zit. n. Voyer et al., 1995)	Erkennen, welcher Würfel mit Buchstaben und Zahlen derselbe wie der Vorlagewürfel ist	SV	Einzel- und Gruppentest	hom.	sign., größer bei Einzeltest
Block Design Subtest aus verschiedenen Fassungen der Wechsler Intelligenzskalen	z.B. Wechsler (1981)	eine Figur aus 3-dimensionalen Würfeln konstruieren	SV	3 Altersgruppen	homogen außer unter 13 Jahre (n. hom.)	n. sign., 13-18J. sign.

Aus den Tabellen 2 und 3 kann entnommen werden, dass bei allen vier Tests aus der Kategorie *Mental Rotation* (MR) signifikante Geschlechtsunterschiede auftraten. Bei den Tests der Kategorien *Spatial Perception* sowie *Spatial Visualization* traten teilweise signifikante Geschlechtsunterschiede auf, teilweise nicht. Linn und Petersen berichteten in der Kategorie *Spatial Visualization* (SV) von keinen signifikanten Geschlechtsunterschieden; dem wider-

sprechen manche Ergebnisse von Voyer et al., die mit den Analysen der einzelnen Tests aus dieser Kategorie bei *Paper Form Board* und beim *Identical Blocks Test* in allen Subgruppen, beim *Embedded Figures Test* sowie beim *Block Design Subtest* in jeweils einer Subgruppe signifikante Geschlechtsunterschiede erhielten.

Voyer et al. merken kritisch an, dass die Kategorie *Spatial Visualization* von Linn und Petersen (1985) eine „catchall“-Gruppierung (S. 262) ist („Schublade für Krimskrams“), da *Paper Form Board* und der *Identical Blocks Test* Anteile von mentaler Rotation enthalten und in der *Mental Rotation*-Kategorie ihrer Meinung nach besser aufgehoben wären. Auch bei anderen Tests der *Spatial Visualization*-Kategorie, die aber keine Geschlechtsunterschiede zeigen, sind Rotationsfähigkeiten notwendig. Der Unterschied könnte nach Voyer et al. darin liegen, dass beim *Identical Blocks Tests* sowie bei *Paper Form Board* zweidimensionale Präsentationen in dreidimensionale Lösungen transformiert werden müssen („dimensionality crossing“ nach Horan & Rossler, 1984). Voyer et al. betonen aber, dass mehr empirische Untersuchungen notwendig sind, weil beispielsweise beim *Block Design Subtest* auch „dimensionality crossing“ involviert ist und trotzdem in zwei Subgruppen keine signifikanten Geschlechtsunterschiede auftreten.

Zusammenfassend konnten Linn und Petersen (1985) sowie Voyer et al. (1995) in ihren Metaanalysen empirisch nachweisen, dass Geschlechtsunterschiede bei Raumvorstellungsaufgaben zugunsten der Männer in der Kategorie *Mental Rotation* existieren. Weiters berichteten Linn und Petersen von beobachteten Geschlechtsunterschieden zugunsten der Männer bei *Spatial Perception*-Aufgaben, nicht aber bei *Spatial Visualization*-Aufgaben. Voyer et al. hingegen zeichneten bei letzteren beiden Kategorien ein differenzierteres Bild: Sie fanden bei *Spatial Perception*- und *Spatial Visualization*-Aufgaben in Abhängigkeit der vorgegebenen Tests, der Vorgabe- sowie Auswertungsmodi teilweise signifikante Geschlechtsunterschiede und teilweise nicht.

2.4.2 Warum treten Geschlechtsunterschiede bei Raumvorstellung auf?

Auf die Frage, warum Geschlechtsunterschiede bei Raumvorstellungsleistungen vorkommen, werden unzählige Erklärungen angeboten. Sie reichen von körperlichen, biologischen Ursachen über evolutionäre Theorien bis hin zu sozialen und soziokulturellen Merkmalen.

2.4.2.1 Körperliche Ursachen

Als körperliche Ursachen für Geschlechtsunterschiede werden in der Literatur der Einfluss genetischer Faktoren (z.B. McGee, 1982), der Einfluss von Geschlechtshormonen (z.B. McGee, 1979), die Reifungsrate (z. B. Sanders & Soares, 1986), hirnstrukturelle Unterschiede (z.B. Hausmann, 2007) und vieles mehr genannt.

Studien zum hormonellen Einfluss zeichnen (noch) kein eindeutiges Bild. Während Shute, Pellegrino, Hubert und Reynolds (1983) in zwei Experimenten eine Beziehung zwischen Androgenlevel und Raumvorstellung in die Richtung belegen konnten, dass Männer mit niedrigem und Frauen mit hohem Androgenspiegel die besten Leistungen erzielten, berichteten McKeever, Rich, Deyo und Conner (1987) von keinen Zusammenhängen zwischen Testosteronspiegel und verschiedenen Fähigkeitsmessungen bei beiden Geschlechtern. In einer Studie von Janowsky, Oviatt und Orwoll (1994) konnten wiederum signifikante Leistungssteigerungen beim räumlichen Denken durch die Gabe von Testosteron festgestellt werden. Die Autoren berichteten, dass sich die Raumvorstellungsleistungen bei älteren Männern nach einer dreimonatigen Steigerung der „Testosteron-Baseline“ auf 150 Prozent verbesserten, andere kognitive Bereiche aber unverändert blieben. Als zugrunde liegenden hormonellen Mechanismus vermuten die Autoren, dass sich die Testosterongabe auf die körpereigene Östradiolproduktion auswirkt, die wiederum das räumliche Denken beeinflusst.

Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis und Güntürkün (2000) untersuchten den Zusammenhang zwischen hormonellen Schwankungen bei Frauen (mit regelmäßigem Menstruationszyklus) und Raumvorstellungsleistungen. Es zeigte sich ein signifikanter Leistungsunterschied je nach Zyklusphase mit höheren Scores während der Menstruation, bei der die Östradiolkonzentrationen niedrig sind, und niedrigeren Scores während der Mittluteal-Phase (hoher Östradiolspiegel). Je höher die Testosteron- und je niedriger die Östradiolkonzentrationen waren, desto besser waren die Raumvorstellungsleistungen bei dreidimensionalem Reizmaterial – ein ähnliches Ergebnis wie bei Janowsky et al. (1994).

Als andere mögliche Ursache für Geschlechtsunterschiede werden hirnstrukturelle Abweichungen zwischen Männern und Frauen genannt, die sich nach Hausmann (2007) vor allem auf den Neokortex, aber auch auf andere subkortikale Strukturen beziehen. Bedeutsam für die Geschlechtsunterschiede dürfte auch die Hemisphärenasymmetrie sein. Die Forschungslage deutet in die Richtung, dass Männer meistens mehr kognitive Funktionen in einer Hemisphäre lokalisiert haben, während Frauen zu einer bilateralen Hirnorganisation tendieren (Halpern, 1996). Diese Unterschiede in der Hirnorganisation können bereits in der

frühen Kindheit nachgewiesen werden - ein Hinweis darauf, dass sie angeboren sind und nicht durch unterschiedliche Lebenserfahrungen erworben werden (Witelson, 1987, 1991).

2.4.2.2 Evolutionäre Ursachen

Neben der Untersuchung körperlicher Ursachen wird versucht, mit Hilfe von Evolutionsthesen Leistungsunterschiede bei Raumvorstellungsaufgaben zwischen Männern und Frauen zu erklären. Jones, Braithwaite und Healy (2003) kritisieren, dass die meisten Hypothesen unlogisch sind und bislang nicht oder zu wenig überprüft wurden. Die Hypothese, die den Autorinnen nach von der vorhandenen Datenlage am meisten unterstützt wird, bezieht sich auf die „Größe des Aktionsradius“, die besagt, dass sich Männer mit einer größeren Reichweite mit mehr Frauen paaren können. Eine Zusammenfassung der bekanntesten Evolutionsthesen (siehe Jones et al., 2003, S. 408) kann Tabelle 4 entnommen werden.

Tabelle 4: Zusammenfassung einiger Evolutionsthesen samt Spezies, für die sie gelten, der Vorhersage, die sie treffen sowie dem Selektionsdruck.

These	Spezies	Vorhersage	Selektionsdruck
Dispersal (Zerstreuung, Verbreitung)	Alle	verbreitetes Geschlecht besser	Geburtszerstreuungsdistanz variiert zwischen Geschlechtern
Fruchtbarkeit und elterliche Pflege	Alle	Frauen weniger als Männer	Frauen reduzieren Mobilität, um die Sterblichkeit während der Reproduktionsphasen zu verringern
Frauen suchen Futter	Menschen	Frauen mehr als Männer	Teilung der Futtersucharbeit: Frauen erinnern sich an Orte mit Futter
Männer suchen Futter	Menschen	Männer mehr als Frauen	Teilung der Futtersucharbeit: Männer verwenden Navigationsfähigkeiten für das Jagen
Größe des Aktionsradius	Alle	Männer mehr als Frauen	Polygyne Männer haben größere Aktionsradien, um sich mit mehr Frauen zu paaren
Männliche Kriegskunst	Menschen	Männer mehr als Frauen	Männer reisen lange Distanzen, um Gegner zu töten und Frauen zu erobern
Wahl der Frauen	Menschen	Männer mehr als Frauen	Frauen wählen Männer auf Basis ihres Jagderfolges

Silverman und Eals (1992) argumentieren, dass Männer nicht generell bessere Raumvorstellungsfähigkeiten haben, sondern dass sie aus der Evolutionsgeschichte heraus bei bestimmten Raumvorstellungsaufgaben besser abschneiden und Frauen bei anderen. In Anlehnung an eine Jäger-Sammler-Hypothese ist es für Jäger (Männer) wichtig, sich in Relation zu Ob-

jekten oder Orten orientieren zu können, um in unbekannten Gebieten jagen zu können, während für Sammler (Frauen) die Erinnerung an Orte (Futterplätze) wichtig ist.

McBurney, Gaulin, Devineni und Adams (1997) untersuchten Silvermans und Eals Argumentationslinie an einer Stichprobe von 103 Testpersonen, die das kommerziell erhältliche Spiel „Memory“ spielten und den *MRT* (Vandenberg & Kuse, 1978) bearbeiteten. „Memory“ sollte den Autoren nach die Effizienz beim Sammeln messen, da der Spielerfolg vom Erinnerungsvermögen an Orte abhängt. Wie erwartet schnitten Männer bei den Raumvorstellungsaufgaben besser ab als Frauen ($d=0,67$); Frauen zeigten hingegen bei „Memory“ bessere Leistungen ($d=0,89$). McBurney et al. zogen aus den Ergebnissen die Schlussfolgerung, dass für Männer und Frauen verschiedene Raumvorstellungsfähigkeiten bei der evolutionären Anpassung wichtig waren.

2.4.2.3 Soziale und soziokulturelle Ursachen

Neben körperlichen und evolutionären Ursachen werden soziale und soziokulturelle Faktoren als Erklärung für die Geschlechtsunterschiede herangezogen. Obwohl sich die Erziehung in den letzten Jahrzehnten sicherlich in eine „geschlechtsneutralere“ Richtung entwickelt hat, kann nicht davon ausgegangen werden, dass Mädchen und Buben genau denselben Erfahrungen und Entwicklungsbedingungen ausgesetzt sind.

Baenninger und Newcombe (1989, 1995) berichteten, dass mathematische und räumliche Fähigkeiten durch Erfahrung beeinflusst werden und dass Raumvorstellungsleistungen und die Teilnahme an räumlichen Aktivitäten zusammenhängen. Quaiser-Pohl und Lehmann (2002) untersuchten die Zusammenhänge zwischen Erfahrungs-, Einstellungs- und Leistungsvariablen und stellten fest, dass Frauen weniger Erfahrungen mit Aktivitäten haben, die für die Entwicklung von Raumvorstellung als wichtig erachtet werden. Außerdem schätzten Frauen ihre räumlichen Fähigkeiten im täglichen Leben niedriger ein als Männer und ihre Raumvorstellungsleistungen hingen eng mit Einstellungsfaktoren zusammen.

Baenninger und Newcombe (1989, 1995) betonen auch die unterschiedlichen Erwartungen, denen Buben und Mädchen ausgesetzt sind, die das Selbstvertrauen und die Zuschreibungen an die eigenen Leistungen beeinflussen können. In dieselbe Richtung gehen die Aussagen von Shih, Pittinsky und Ambady (1999), die berichten, dass der Einfluss von Selbstvertrauen klar belegt ist und zwar in die Richtung, dass die kognitiven Leistungen mit größerem Selbstvertrauen besser sind.

Zusätzlich können die Geschlechtsrollenidentifikation und Geschlechtsrollenstereotype bei Leistungen eine Rolle spielen (z.B. Hausmann, 2007): Personen strengen sich beispielsweise bei kognitiven Aufgaben mehr an, wenn sie im Glauben sind, dass gutes Abschneiden bei der Aufgabe mit der eigenen Geschlechtsrolle übereinstimmt. Bezüglich Raumvorstellungsleistungen schnitten in einer Untersuchung von Signorella und Jamison (1978) Personen mit einem maskulinen Selbstkonzept besser ab als Personen mit einem femininen Selbstkonzept.

Spencer, Steele und Quinn (1999) beschreiben, dass Geschlechtsunterschiede bei manchen mathematischen Leistungen verschwinden, wenn ein Test in der Instruktion als nicht geschlechtssensitiv dargestellt wird und dadurch bestimmte Geschlechtszuschreibungen abgeschwächt werden. Es stellt sich daher die Frage, ob Geschlechtsrollenstereotype auch eine Ursache für die Schwankungsbreite bei der Größe der Geschlechtsunterschiede sein könnten (Hausmann, 2007).

2.4.3 Geschlechtsunterschiede bei Varianten der Testvorgabe und Testauswertung

Die Größe der Geschlechtsunterschiede kann variieren bzw. Geschlechtsunterschiede können sogar verschwinden, wenn Modifikationen bei der Testvorgabe und/oder Testauswertung vorgenommen werden. Zu diesem Ergebnis kommen Goldstein, Haldane und Mitchell (1990), die empirisch belegen, dass Geschlechtsunterschiede beim *MRT* (Vandenberg & Kuse, 1978) zugunsten der Männer nur gefunden werden, wenn der Test mit dem Zeitlimit von drei Minuten pro Subset vorgegeben wird und der Rohscore als Kennwert für die Leistung herangezogen wird. Bei Wegfall des Zeitlimits und Berechnung von „ratio scores“ (Anzahl der gelösten im Verhältnis zur Anzahl der bearbeiteten Aufgaben) konnten die Autoren keinen Geschlechtsunterschied entdecken. Goldstein et al. nahmen (wie auch z.B. Linn & Petersen, 1985) an, dass Frauen die Aufgaben „vorsichtiger“ und daher langsamer bearbeiten. Auch in der Metaanalyse von Voyer et al. (1995) finden sich einige Hinweise, dass sich die Größe der Geschlechtsunterschiede je nach Vorgabe- oder Auswertungsmodus verändert oder sie gänzlich verschwinden (siehe Tabelle 3).

Von einem anderen Ergebnis als Goldstein et al. berichtet Peters (2005), der die Leistung beim *MRT* in Abhängigkeit der Zeit untersuchte. Als Ergebnis zeigte sich, dass weniger Frauen als Männer alle Items innerhalb des Zeitlimits bearbeiten konnten und dass der Ge-

schlechtsunterschied im Lauf der Testung anstieg. Wurde das Zeitlimit bei der Testvorgabe auf das Doppelte der Standardbedingungen erhöht, verbesserte sich die Leistung bei beiden Geschlechtern, der große Geschlechtsunterschied blieb aber bestehen.

Eine Untersuchung von McWilliams, Hamilton und Muncer (1997) mit realen Modellen führte auch zu interessanten Ergebnissen. Einer Gruppe von Testpersonen wurden reale dreidimensionale Holzmodelle der Figuren von Shepard und Metzler (1971) vorgegeben, der anderen Gruppe wurden die dreidimensionalen Figuren in einer Papier-Bleistift-Version vorgelegt. In der Gruppe mit den dreidimensionalen Figuren auf Papier ergaben die Auswertungen einen signifikanten Geschlechtsunterschied zugunsten der Männer; in der anderen Gruppe konnte kein signifikanter Geschlechtsunterschied nachgewiesen werden. McWilliams et al. schlossen aus den Ergebnissen, dass die Unterschiede zwischen Männern und Frauen bei traditionellen Rotationsaufgaben entstehen, weil Bilder von dreidimensionalen Objekten und nicht die Objekte selbst verwendet werden. Die Testpersonen müssen aus einem zweidimensionalen Abbild auf Papier eine dreidimensionale Repräsentation erstellen („dimensionality crossing“ nach Horan & Rossler, 1984; siehe oben), was Frauen schwerer fallen könnte als Männern.

2.4.4 Veränderungen bei Geschlechtsunterschieden

In diesem Kapitel stehen Veränderungen bei Geschlechtsunterschieden im Vordergrund; konkreter sollen Untersuchungen zur Veränderung der Größe der Geschlechtsunterschiede, aber auch Studien, die sich mit Veränderungen der Geschlechtsunterschiede aufgrund des Alters oder durch Trainingsmaßnahmen befassen, vorgestellt werden.

Die Annahme, dass Geschlechtsunterschiede in den letzten Jahrzehnten kleiner geworden sind, ist aufgrund der großen gesellschaftlichen Veränderungen, die in der jüngeren Vergangenheit stattgefunden haben, einleuchtend. Feingold (1988) untersuchte verschiedene Testfassungen von allgemeinen Begabungstests aus unterschiedlichen Jahren und konnte größtenteils geringer werdende Geschlechtsunterschiede belegen.

Voyer et al. (1995) überprüften Feingolds Ergebnisse und hypothesisierten, dass die Größe der Geschlechtsunterschiede und die Erscheinungsjahre der Publikationen negativ korreliert sein müssten. Es zeigte sich ein kleiner positiver, nicht signifikanter Zusammenhang. Die Autoren änderten ihre Vorgangsweise dahingehend, dass sie als wichtigste Variable das Geburtsjahr der Testpersonen (statt des Erscheinungsjahres der Publikation) heranzogen. Die Auswertungen ergaben einen negativen, nicht signifikanten Zusammenhang zwischen

Geburtsjahr und Größe der Geschlechtsunterschiede, was bedeutet, dass bei später Geborenen kleinere Geschlechtsunterschiede auftreten als bei früher geborenen Personen. Bei der Überprüfung der Annahmen getrennt für alle Raumvorstellungstests zeigte sich beim *MRT* (Vandenberg & Kuse, 1978) ein unerwartetes Ergebnis, nämlich ein positiver, signifikanter Zusammenhang, der für einen Anstieg der Geschlechtsunterschiede sprechen würde. Glück (1999) vermutet, „daß dieser Effekt durch eine untypische Korrelation zwischen Alter der Versuchspersonen und Publikationsjahr zustandegekommen sein könnte, wodurch dann beim Auspartialisieren des Alters ein wichtiger Teil der relevanten Information mit ausgeschaltet worden wäre.“ (S. 120-121).

Stumpf und Klieme (1989) berichten in ihrer Metaanalyse mit Daten aus 18 Studien zwischen 1978 und 1988 aus der deutschsprachigen Kultur eine größere Annäherung der Geschlechter als Feingold.

Mit Veränderungen der Geschlechtsunterschiede mit dem Alter befassten sich auch Geiser, Lehmann und Eid (2008). Sie analysierten Daten von ca. 1600 Testpersonen zwischen 9 und 23 Jahren, die mentale Rotationsaufgaben bearbeitet hatten. In allen 15 Altersgruppen erzielten Männer höhere Scores, in neun Gruppen zeigten sich große Effektgrößen ($d > 0,8$), in allen anderen mittlere (zwischen 0,5 und 0,8). Die Leistung bei mentalen Rotationsaufgaben stieg signifikant mit dem Alter, der Leistungsanstieg war für Männer größer als für Frauen. Geiser et al. hypothesieren, dass die mit dem Alter größer werdenden Geschlechtsunterschiede damit zusammenhängen könnten, dass Männer mehr in tagtägliche räumliche Aktivitäten eingebunden sind und sie ihre Fähigkeiten daher von der Kindheit bis ins Erwachsenenalter mehr verbessern als Frauen.

Dass Trainingsmaßnahmen zur Verbesserung von räumlichen Leistungen beitragen können, berichten neben Forscherinnen wie beispielsweise Baenninger und Newcombe (1989) auch Feng, Spence und Pratt (2007), die sich mit räumlicher Aufmerksamkeit, einer wichtigen Basiskapazität für räumliches Denken auf höheren Ebenen, beschäftigten. Ihre Untersuchungen belegen, dass ein zehnstündiges Training mit einem „Action-Videospiel“ den Geschlechtsunterschied bei räumlicher Aufmerksamkeit zugunsten der Männer fast eliminieren und den Unterschied bei mentalen Rotationsleistungen verringern kann. Das Training mit einem anderen als einem Action-Spiel führte zu keinen Veränderungen.

2.4.5 Geschlechtsunterschiede und Bearbeitungsstrategien

Dass die Anwendung bzw. die Vermeidung bestimmter Strategien mit der Raumvorstellungsleistung zusammenhängt, wurde schon im Kapitel 2.3.2 (Studien zum Zusammenhang zwischen Bearbeitungsstrategie und Leistung) beschrieben. Der Frage, ob die Wahl unterschiedlicher Strategien die Geschlechtsunterschiede bei Raumvorstellungsaufgaben erklären kann, widmeten sich Hosenfeld, Strauss und Köller (1997). Sie formulierten in Anlehnung an Kyllonen et al. (1984) zwei Hypothesen, nämlich, dass erstens die Strategiewahl den Geschlechtsunterschied erzeugt und zweitens die Effektivität der Strategieanwendung den Geschlechtsunterschied erzeugt. Hosenfeld et al. erwarteten, dass männliche Testpersonen aufgrund der angemesseneren holistischen Strategie bessere Leistungen bei mentalen Rotationsaufgaben erzielen und sie die holistische Strategie effizienter einsetzen als weibliche Testpersonen.

Als Testmaterial kamen beide Parallelformen des Subtests Würfelaufgaben aus dem *IST* (Amthauer, 1953) mit insgesamt 40 Items, darunter 20 Raumwürfel zum Einsatz, wobei wegen Deckeneffekten nur die zehn schwierigsten Flächenwürfel und die zehn schwierigsten Raumwürfel analysiert wurden. Bei den Raumwürfeln zeigte sich ein signifikanter Geschlechtsunterschied zugunsten der männlichen Testpersonen, bei den Flächenwürfeln hingegen nicht. Auswertungen mit dem Mixed-Rasch-Modell (Rost, 1990) ergaben eine Zwei-Klassen-Lösung: In der einen Klasse wurde eine analytische Strategie angewendet, in der anderen wurden die Items holistisch bearbeitet.

Kein signifikantes Ergebnis, sondern nur einen Trend in die erwartete Richtung, brachte die Überprüfung der Hypothese, dass männliche Testpersonen bei Rotationsaufgaben höhere Leistungen erzielen, da sie die angemessenere holistische Strategie anwenden. Die Überprüfung der zweiten Hypothese, dass männliche Testpersonen die angemessene Strategie effizienter einsetzen, zeigte in der Klasse der Holistiker einen signifikanten Geschlechtsunterschied zugunsten der Männer, in der Klasse der Analytiker nicht.

Auch Geiser, Lehmann und Eid (2006) stellten die Frage, ob die Verwendung unterschiedlicher Bearbeitungsstrategien bei einer deutschen Version des *MRT* (Peters et al., 1995) mit Geschlechtsunterschieden zusammenhängt. In ihrer Auswertung mittels Latent-Class-Analyse (LCA, z.B. Rost, 1988) wählten die Autoren eine Fünf-Klassen-Lösung für beide Testteile als die am besten passende Beschreibung der Daten aus. Es zeigten sich große Unterschiede in der Häufigkeitsverteilung von Männern und Frauen über die Klassen hinweg, wobei Männer am häufigsten in den beiden Klassen mit den besten Leistungen anzutreffen

waren, während Frauen in der Klasse mit der niedrigsten Leistung und einer Klasse mit einer mittelmäßigen Leistung, aber einem sehr langsamen Arbeitstempo dominierten.

2.4.6 Die Berücksichtigung von Anlage- und Umweltaspekten bei der Erforschung von Geschlechtsunterschieden sowie Leistungsunterschieden innerhalb der Subgruppe der Frauen

Einen anderen Schwerpunkt als die Studien der bisher erwähnten Wissenschaftler wählte Casey (1996), die sich mit dem Zusammenhang zwischen Anlage- und Umweltaspekten beschäftigte, um Erklärungen für die vorhandenen Geschlechtsunterschiede bei Raumvorstellungsleistungen zwischen Männern und Frauen sowie Unterschiede innerhalb der Gruppe der Frauen zu finden.

Casey legte ihren Überlegungen die Theorie von Annett (1985) zugrunde: Annett geht davon aus, dass die meisten Menschen einen *right-shift-Faktor* erben. Der *right-shift-Faktor* ist ein einzelnes Gen, dessen dominantes Allel Rechtshändigkeit und eine linksseitig dominante Sprachverarbeitung festlegt. Das rezessive Allel führt zu einem zufälligen Lateralitätsmuster; Händigkeit und Sprachverarbeitung können links- oder rechtsseitig dominant sein.

Nach Annetts Theorie sind Rechtshänder mit ausschließlich rechtshändigen nahen Verwandten homozygot für den *right-shift-Faktor* und linksdominant bei der Sprachverarbeitung; sie bevorzugen verbale Strategien zum Problemlösen und sind schlechte Raumvorsteller. Im Gegensatz dazu sind Rechtshänder mit linkshändigen oder beidhändigen Verwandten heterozygot für das *right-shift-Gen* und besitzen das rezessive und dominante Allel für den *right-shift-Faktor*. Auf Grundlage des Konzepts des heterozygoten Vorteils, das besagt, dass Menschen mit dominantem und rezessivem Allel „besser ausgestattet sind“, schließt Annett, dass heterozygote Frauen Sprache nicht so stark linkshemisphärisch verarbeiten, beim Problemlösen weniger von verbalen Strategien abhängig sind und das Potential für gute Raumvorstellungsleistungen haben.

Nach Annett haben ca. 50 Prozent der Frauen den heterozygoten Vorteil für den *right-shift-Faktor* und daher eine Organisationsstruktur im Gehirn, die es diesen Frauen ermöglichen sollte, gute Raumvorstellungsfähigkeiten zu entwickeln. Auf Basis von Annetts Theorie formuliert Casey die Hypothese, dass Frauen mit dieser bestimmten Organisationsstruktur im Hirn und entsprechenden Raumvorstellungserfahrungen bessere Raumvorstellungsleistungen erzielen können als Frauen, die dasselbe Ausmaß an Erfahrungen besitzen, das Orga-

nisationsmuster im Gehirn aber nicht aufweisen. Casey nimmt auch an, dass Frauen, die zwar die bestimmte Form der Organisationsstruktur aufweisen, aber keine Erfahrungen mit Raumvorstellung machen konnten, keine ausgeprägten Raumvorstellungsfähigkeiten entwickeln. Indem Casey Organisationsstrukturen im Gehirn und Raumvorstellungserfahrungen einbezieht, berücksichtigt sie in ihren Forschungsarbeiten Anlage- und Umweltaspekte.

Casey beschreibt in ihrem 1996 publizierten Artikel drei unterschiedliche empirische Untersuchungen, die ihre Annahmen unterstützen:

Casey und Brabeck (1989) verglichen die Leistungen bei mentalen Rotationsaufgaben von Rechtshänderinnen mit rechtshändigen Eltern und Geschwistern mit Rechtshänderinnen mit zumindest einem links- oder beidhändigen nahen Verwandten sowie mit Nicht-Rechtshänderinnen. Als kritisches Merkmal für Raumvorstellungserfahrungen wurde das Hauptfach im Studium („Mathematik“ bzw. „Naturwissenschaften“ oder ein anderes Hauptfach) herangezogen. Die Auswertungen ergaben für Frauen eine signifikante Interaktion zwischen Hauptfach und Händigkeit in der Familie. Innerhalb der Gruppe der Frauen, die Mathematik oder Naturwissenschaften als Hauptfach gewählt hatten, erzielten die rechtshändigen Frauen mit nicht-rechtshändigen Verwandten im Vergleich zu den beiden anderen Frauengruppen (Rechtshänderinnen mit nur rechtshändigen Verwandten, nicht-rechtshändige Frauen) bessere Ergebnisse; sie schnitten bei den Testaufgaben sogar auf demselben Level wie Männer ab. In der Gruppe mit anderen Hauptfächern fanden Casey und Brabeck keine signifikanten Effekte bei der familiären Händigkeit. Die Autorinnen schlossen daraus, dass biologische (Anlage-) mit Umweltfaktoren interagieren und dabei die Fähigkeit zur mentalen Rotation bei Frauen beeinflussen.

Eine Bestätigung der Ergebnisse von Casey und Brabeck gelang Casey, Colón und Goris (1992) an einer anderen Stichprobe. Die Autoren untersuchten Schüler einer High School. Innerhalb der Gruppe der weiblichen Testpersonen zeigten wieder diejenigen die besten Raumvorstellungsleistungen, die sich freiwillig mathematische und naturwissenschaftliche Fächer ausgesucht hatten und den heterozygoten Vorteil (Rechtshänderinnen mit nicht-rechtshändigen Verwandten) aufwiesen.

In einer anderen Studie verwendeten Pezaris und Casey (1991) ein Interferenz-Paradigma, das davon ausgeht, dass zwei Aktivitäten mit denselben Prozessressourcen um eine limitierte Prozesskapazität wetteifern (Paivio, 1982), während zwei Aktivitäten, die unterschiedliche Ressourcen benötigen, weniger Interferenz produzieren. Daraus ergibt sich, dass eine Testperson mit räumlichen Bearbeitungsstrategien bei gleichzeitiger Darbietung einer visuellen

Grafik eine größere Interferenz zeigen sollte als bei Darbietung einer auditiven Ziffernfolge. Umgekehrt sollte bei Testpersonen mit verbalen Bearbeitungsstrategien die Darbietung einer auditiven Ziffernfolge mehr Interferenz produzieren. Die empirische Untersuchung ergab, dass die größte Interferenz in der männlichen Stichprobe bei gleichzeitiger Bearbeitung von mentalen Rotationsaufgaben und dem Auftrag, sich eine dargebotene Grafik zu merken, auftrat. Im Gegensatz dazu zeigten die meisten Frauen die größte Interferenz, wenn ihnen gleichzeitig zu den Rotationsaufgaben eine verbale Interferenz – der Auftrag, eine Zahlenkombination zu behalten – dargeboten wurde. Pezaris und Casey konnten bei rechtshändigen Mädchen mit nicht-rechtshändigen Verwandten, die gute Leistungen bei mathematischen und naturwissenschaftlichen Aufgaben erzielten, nachweisen, dass diese ein ähnliches Interferenzmuster wie Buben zeigten – nämlich signifikant mehr räumliche und weniger verbale Interferenz.

Kritik an manchen Teilen von Caseys (1996) Arbeiten (und Lob für andere Teile) kommt beispielsweise von Halpern (1996). Sie zitiert Hellige (1994), der Annetts (1985) Vorhersage, dass Linkshändigkeit häufig mit einer rechtsdominanten Sprachverarbeitung gekoppelt ist, in Frage stellt. Halpern nennt auch als umstrittenen Punkt in Caseys Theorien die Händigkeit in der Familie, da aus der Literatur nicht klar hervorgeht, ob und in welchem Ausmaß Händigkeit vererbt wird (zu den Kritikpunkten siehe auch Kapitel 2.5 Händigkeit).

Aus den berichteten Ergebnissen kann zusammenfassend der Schluss gezogen werden, dass Geschlechtsunterschiede durch Interaktionen zwischen biologischen, sozialen und psychologischen Faktoren entstehen; welche Wechselwirkungen sie eingehen, ist allerdings unklar.

2.5 Händigkeit

Das letzte Kapitel des Theorieteils soll der Händigkeit und dem Zusammenhang zwischen Händigkeit und Leistung gewidmet sein. Im Anschluss an eine Definition werden verschiedene Ergebnisse zur Handpräferenz und zur Hemisphärendominanz dargestellt. Den Abschluss des Kapitels bilden Untersuchungen zur Beziehung zwischen Händigkeit und Leistung.

Unter Händigkeit versteht man im Allgemeinen das Phänomen, dass Menschen für anspruchsvolle, feinmotorische Aktivitäten (z.B. Schreiben) eine Hand - die dominante - bevorzugt einsetzen. Genauso wichtig wie die dominante Hand ist bei vielen Tätigkeiten die Unterstützung der nicht-dominanten (Wilson, 2002).

Im Tierreich kann beispielsweise bei Katzen, Ratten und Affen beobachtet werden, dass sie eine Gliedmaße für anspruchsvolle Tätigkeiten bevorzugen; die Verteilung auf die Gliedmaßen ist allerdings – im Gegensatz zur menschlichen Bevölkerung – annähernd gleich groß (Stein & Stoodley, 2006). Bei Menschen (in Europa) wird nämlich von zehn bis 15 Prozent linkshändigen Personen ausgegangen. Neben diesen gibt es wenige Beidhänder, die große Mehrheit ist rechtshändig (Porac & Coren, 1981). In einer Untersuchung von über 3000 Personen der deutschen Bevölkerung zeigte sich, dass ca. 91 Prozent eine Präferenz für die rechte Hand aufweisen (Arnold-Schulz-Gahmen, Ehrenstein, Schweingruber, Selinski, Urfer & Zschesche, 1998). Diese überwiegende Rechtshändigkeit dürfte schon bei den Urmenschen und Neandertalern vorgeherrscht haben (Cashmore, Uomini & Chapelain, 2008) und kann auch in der heutigen Zeit in allen Kulturen beobachtet werden. Warum sich die dominierende Rechtshändigkeit entwickelt hat, ist unklar; der Vorteil der Präferenz für eine Hand wird darin gesehen, dass diese Hand für spezialisierte Tätigkeiten gut trainiert ist (Stein & Stoodley, 2006).

Die Handpräferenz, die für das ganze Leben festgelegt ist (Sattler, 2005), dürfte genetische Ursachen haben, wobei die Vererbung scheinbar nicht den Mendelschen Gesetzen folgt, da bei einem Viertel der eineiigen Zwillinge eine unterschiedliche Händigkeit beobachtet werden kann. Stein und Stoodley (2006) zufolge wird eine Stelle auf Chromosom 2 mit Händigkeit in Verbindung gebracht.

Für biologische Prozesse bei der Händigkeit sprechen auch Studien, die von einer sichtbaren Handpräferenz im Mutterleib berichten. Santrock (2009) beschreibt, dass in einer Untersuchung neun von zehn Föten an ihren rechten Daumen lutschten. Bei einer anderen Studie

wurde die Präferenz von Neugeborenen für die eine oder die andere Körperhälfte untersucht. 65 Prozent der Neugeborenen drehten ihren Kopf auf die rechte Seite, wenn sie am Rücken lagen, 15 Prozent auf die linke Seite und 20 Prozent zeigten keine Präferenz. Diese Präferenzen hingen mit der späteren Händigkeit zusammen (Santrock, 2009).

Interessant ist auch die Verknüpfung zwischen Händigkeit und der Sprachverarbeitung im Hirn. Sprache wird in der so genannten dominanten Hirnhälfte verarbeitet (Young, Young & Tolbert, 2008); bei 95 Prozent der Rechtshänder passiert das in der linken Hirnhemisphäre, bei zwei Prozent in der rechten und drei Prozent verarbeiten Sprache in beiden Hemisphären. Obwohl die rechte Hand mit der linken Hirnhälfte und die linke Hand mit der rechten Hirnhälfte verbunden ist und man daher annehmen könnte, dass Linkshänder Sprache rechts verarbeiten, zeigt sich ein anderes Bild: Auch 70 Prozent der Linkshänder verarbeiten Sprache links, 15 Prozent rechts und 15 Prozent in beiden Hemisphären (Stein & Stoodley, 2006). Das bedeutet, dass nicht von eindeutigen Zusammenhängen zwischen dominanter Hirnhälfte und Handpräferenz ausgegangen werden kann.

2.5.1 Händigkeit und Leistung

In der wissenschaftlichen Literatur werden häufig Händigkeit und Leistung in Verbindung gebracht. Gladue und Bailey (1995) nahmen beispielsweise eine Untersuchung zu den Zusammenhängen zwischen mentaler Raumvorstellungsleistungen, Händigkeit und der sexuellen Orientierung bei Männern und Frauen vor. Sie konnten an einer relativ großen Stichprobe signifikante Geschlechtsunterschiede bei Raumvorstellungsleistungen belegen, bei der Händigkeit zeigten sich keine Unterschiede. Die sexuelle Orientierung stand – im Gegensatz zu den Ergebnissen anderer Studien (z. B. Gladue, Beatty, Larson & Staton, 1990) – in keinem signifikanten Zusammenhang mit anderen Variablen.

Santrock (2009) fasst aus mehreren Studien zusammen, dass Linkshänder – in manchen Studien ist auch von nicht-rechtshändigen Personen bzw. Nicht-Rechtshändern die Rede – zu außergewöhnlich guten räumlichen Leistungen tendieren und dass Architekten, Mathematiker, Musiker sowie Künstler häufiger Linkshänder sind als zu erwarten wäre. Santrock beschreibt darüber hinaus, dass bei einer Studie mit mehr als 100 000 Testpersonen, die den Scholastic Aptitude Test (SAT) bearbeiteten, in der Gruppe mit den höchsten Scores 20 Prozent linkshändig waren. Das heißt, dass doppelt so viele Linkshänder in der Top-Gruppe waren wie der Verteilung in der Population nach (bei zehn Prozent) anzunehmen gewesen wäre.

Linkshändigkeit wird aber zugleich mit häufigeren Schwierigkeiten bei sprachlichen Fertigkeiten in Zusammenhang gebracht: Mehr Linkshänder haben – empirisch bestätigt – Lese-schwierigkeiten und Defizite in der Phonologie als Rechtshänder (Santrock, 2009). Einige der hier beschriebenen Ergebnisse finden sich auch, verpackt in ein theoretisches Erklärungsmodell, in der Theorie von Geschwind, Behan und Galaburda (Zitate siehe unten), die im Folgenden angerissen werden soll.

2.5.2 Die Theorie von Geschwind, Behan und Galaburda (GBG)

Geschwind, Behan und Galaburda (Geschwind & Behan, 1982; Geschwind & Galaburda 1985a, 1985b, 1985c; siehe auch McManus & Bryden, 1991) entwickelten eine umfassende Theorie zur Lateralisation des Gehirns, nachdem sie bei Linkshändern eine erhöhte Auftretensrate von Migräne, Erkrankungen des Immunsystems und Lernstörungen beobachtet hatten.

Zusammengefasst und vereinfacht dargestellt besagt ihre Theorie, dass der Einfluss von Testosteron im Mutterleib zu einem bestimmten Zeitpunkt in der Schwangerschaft die Reifung des embryonalen Gehirns beeinflussen kann, indem höhere Dosen des Hormons, das von der Mutter, von einem männlichen Embryo oder einem männlichen Zwillings produziert wird, zu Veränderungen in der Lateralisation des Hirns führen kann. Vor allem das Wachstum der linken Hirnhemisphäre, die für Störfaktoren wie Testosteron anfällig ist, wird gehemmt, während die rechte Hirnhälfte kompensatorisch wächst, die dominante Hemisphäre wird und Funktionen wie beispielsweise die Sprachverarbeitung übernimmt, die in der Regel linkshemisphärisch angesiedelt ist. Das Testosteron kann auch (über den Thymus) Auswirkungen auf die Entwicklung des Immunsystems haben, wodurch vermehrt Anfälligkeiten für Krankheiten auftreten.

Die Theorie von Geschwind, Behan und Galaburda verknüpft also eine Reihe von Phänomenen miteinander, nämlich männliches Geschlecht (und dadurch mehr Testosteron), Störungen des Immunsystems, Linkshändigkeit, gute Leistungen bei Fertigkeiten, die der rechten Hemisphäre zugeschrieben werden (Musik, Raumvorstellung, Mathematik) und schwächere linkshemisphärische Leistungen oder sogar Auffälligkeiten wie Stottern oder Dyslexie.

Das theoretische Konzept ist so umfassend, dass eine empirische Überprüfung des gesamten Inhalts kaum möglich ist (Petermann, Niebank & Scheithauer, 2004; McManus & Bryden,

1991). Teilaspekte, zum Beispiel, dass Linkshändigkeit, Autismus und Stottern bei Männern häufiger auftreten als bei Frauen oder dass Männer bei Raumvorstellungsleistungen besser abschneiden als Frauen, konnten in verschiedenen Untersuchungen bestätigt werden (Morfit & Weekes, 2001; Petermann et al., 2004; Neave, 2008; siehe dazu auch Kapitel 2.4 Geschlechtsunterschiede bei Raumvorstellung).

3. Empirischer Teil

3.1 Entstehungsgeschichte der Endlosschleifen

Die Entwicklung der *Endlosschleifen* begann 1995 in einem Forschungsseminar bei G. Gittler (detailliert beschrieben in Arendasy, 2000), in dem die psychometrischen Eigenschaften des Tests *Schlauchfiguren* von Stumpf und Fay (1983) unter die Lupe genommen wurden.

Die Items des *Schlauchfigurentest* bestehen aus jeweils zwei nebeneinander angeordneten Abbildungen von durchsichtigen Würfeln, in denen sich ein gewundener Schlauch befindet. Der links abgebildete Würfel stellt die Vorderansicht, der rechts abgebildete Würfel eine um 90 oder 180 Grad gedrehte Perspektive der Vorderansicht dar. Aufgabe der Testpersonen ist es, anzugeben, ob der rechte Würfel von rechts, links, oben, unten oder von hinten zu sehen ist.

Es zeigte sich, dass der *Schlauchfigurentest* nicht Rasch-homogen war, was Gittler und Arendasy (2003a) auf einige inhaltliche Punkte zurückführen. So vermuten die Autoren, dass die offenen Enden der Schläuche bei der Orientierung helfen, wodurch einige Items ohne Betrachtung der gesamten Figur und mit weniger räumlichen Fähigkeiten als erwünscht gelöst werden können. Ein weiterer Kritikpunkt, den Gittler und Arendasy anführen, betrifft die Testitems, die Spiegelbilder der Vorderansicht darstellen und mit der Antwortalternative „von hinten“ zu lösen sind. Bei diesen Aufgaben sind Testpersonen im Vorteil, die die Tatsache der Spiegelbilder erkennen, da sie weniger räumliche Fähigkeiten einsetzen müssen als Testpersonen, die die Spiegelbilder nicht entdecken und daher eine andere – aufwendigere – Bearbeitungsstrategie wählen müssen. Schließlich erwähnen Gittler und Arendasy, dass es aufgrund der Darbietungsweise des Testmaterials zu Verwechslungsfehlern bei der Bearbeitungsrichtung in dem Sinn kommen kann, dass trotz korrekter Lösung im Kopf (z.B. „von rechts“) eine falsche Antwortalternative markiert wird („von links“).

Um das übergeordnete Ziel bei der Entwicklung der *Endlosschleifen* – nämlich die Konstruktion von Items, die dem dichotomen logistischen Modell von Rasch (Rasch, 1960, zitiert nach Fischer, 1974) entsprechen – zu erreichen, schlugen Gittler und Arendasy (2003a) vor, die offenen Schlauchenden zu schließen, so dass die Orientierungshilfe wegfällt, auf spiegelbildliche Aufgaben mit der Lösungsalternative „von hinten“ zu verzichten sowie durch die Darbietungsweise die Gefahr der falschen Bearbeitungsrichtung zu minimieren.

In einer ersten Version entstanden 17 fototechnisch hergestellte *Endlosschleifen*, so genannte *photographische Endlosschleifen* (Gittler & Arendasy, 2003a, Beispielaufgabe siehe Kapitel 1. Einleitung), die aufgrund fehlender Rasch-Homogenität überarbeitet wurden. Nach kleinen Änderungen an den Items und der Instruktion entsprachen die Items dem Modell von Rasch.

Seit oben erwähntem Forschungsseminar wurden die *Endlosschleifen* laufend weiterentwickelt; es entstanden unter anderem computergenerierte sowie dreidimensional darzubietende Aufgaben, die in verschiedenen Versionen erprobt wurden. Die Testinstruktion wurde mehrfach angepasst und optimiert. Schließlich wurden auch verschiedene Vorgabemodi und Antwortschemata erarbeitet und getestet (siehe Arendasy, 2000).

3.2 Entwicklung der Fragestellung für die vorliegende Arbeit

Ausgangspunkt der vorliegenden Arbeit sind – wie schon in der Einleitung (Kapitel 1) angesprochen – Datensätze aus Seminaren, die 1998/99 erhoben wurden. 299 Testpersonen bearbeiteten zwei Blöcke mit jeweils 17 *photographischen Endlosschleifen*, wobei sich die Items der beiden Blöcke nur darin unterschieden, dass in einem Aufgabenset die Startansichten mit den Zielansichten vertauscht wurden. Als Auswertungsergebnis wurde erwartet, dass sich die 17 korrespondierenden Items in beiden Vorgabeversionen in ihrer Schwierigkeit nicht unterscheiden. Um diese Hypothese zu testen, waren zwei Auswertungsschritte notwendig.

Zunächst wurde ein Rasch-Modell mit 34 Items gerechnet, von denen im Vorhinein zwei Items samt ihren korrespondierenden Items (insgesamt also vier) ausgeschieden wurden. Der Modelltest niedriger versus hoher Score war für die 30 Items ($df = 29$) Rasch-homogen. Auch der Martin-Löf-Test (Martin-Löf, 1973, zitiert nach Gittler & Arendasy 2003b), der nach Unterschieden zwischen den Vorgabeversionen testet, war nicht signifikant.

Anschließend wurden Analysen mit dem linearen logistischen Testmodell (LLTM; Fischer, 1997) mit 14 freien Basisparametern berechnet, mit der Annahme, dass die jeweils korrespondierenden Items den gleichen Itemparameter besitzen, das heißt gleich schwierig sind. Der Likelihood-Quotienten-Test (Andersen, 1973) war signifikant ($\chi^2 = 175,46$; $\chi^2_{krit(1\%)} = 30,60$; $p < 0,001$). Das bedeutet, dass die 14 freien Basisparameter die Daten nicht so gut erklärten wie die 29 freien Parameter des Rasch-Modells. Die Ergebnisse legten daher nahe,

dass Schwierigkeitsunterschiede zwischen den korrespondierenden Items bei Vertauschung der Start- und Zielansichten aufgetreten waren.

Bei genauerer Durchsicht der Itemschwierigkeiten in den beiden vorliegenden Datensätzen konnte festgestellt werden, dass die Größe der Schwierigkeitsunterschiede zwischen den korrespondierenden Items erheblich variierte. Dieses Phänomen, das wahrscheinlich mit Itemmerkmalen in Zusammenhang gebracht werden kann, ist deshalb interessant, weil sich bei jeder Einzelaufgabe weder die Aufgabenstellung noch die Bilder (Start-/Zielansicht), die diese Aufgabe definieren, änderten, sondern lediglich die Position von Start- und Zielansicht und damit die durch das Itemdesign vorgeschriebene Bearbeitungsrichtung. Der kognitive Aufwand war vor und nach Vertauschung der Start- und Zielansichten als gleichwertig anzusehen; auch die Lösungen waren deterministisch ineinander überführbar: War beispielsweise eine Aufgabe mit einer Drehung nach rechts zu lösen, so lautete die korrekte Antwort bei vertauschter Bearbeitungsrichtung „Drehung nach links“. Und dennoch waren aufgabenspezifisch Unterschiede in den Itemschwierigkeiten zu beobachten.

Für die Überprüfung dieser empirischen Beobachtung wurde eine Forschungsfassung mit *computergenerierten*, frei schwebenden *Endlosschleifen* (siehe Abbildung 5) verwendet, die für die vorliegende Arbeit von G. Gittler zur Verfügung gestellt wurde.

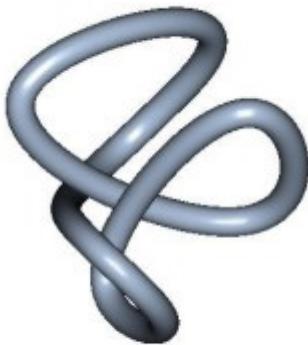


Abbildung 5: Computergenerierte Endlosschleife

Die Wahl fiel auf *computergenerierte* an Stelle *photographischer Endlosschleifen*, weil bei letzteren die Vermutung nahe liegt, dass die bei einigen Items sichtbaren Standflächen der *Endlosschleifen* zu Verzerrungen in den Itemschwierigkeiten geführt haben. Zusätzlich könn-

te die bei manchen Aufgaben nicht vollständig vorhandene Tiefeninformation, die zur Lösung der Items notwendig ist, die widersprüchlichen Ergebnisse verursacht haben.

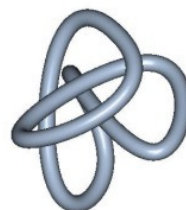
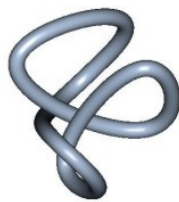
In einem ersten Auswertungsschritt sollen die psychometrischen Eigenschaften der *computergenerierten Endlosschleifenaufgaben* mit Hilfe des Rasch-Modells untersucht werden. Als Hauptfragestellung lässt sich formulieren, ob bei *computergenerierten Endlosschleifen* nach Vertauschung der Start- und Zielansicht unterschiedliche Itemparameter auftreten. Sollte dieses Phänomen zu beobachten sein, stellt sich die Frage, ob sich die Schwierigkeitsunterschiede qualitativ, das heißt anhand von Itemmerkmalen erfassen lassen.

3.3 Beschreibung des verwendeten Testmaterials

3.3.1 Endlosschleifenaufgaben

Für die vorliegende Arbeit wurde ein Paper-Pencil-Test mit 15 *computergenerierten Endlosschleifenaufgaben* erstellt. Die Items wurden nach einer Instruktion samt Beispielaufgaben mit Kontrollmöglichkeit in der unten dargestellten Form präsentiert (siehe Abbildung 6).

Startansicht



Zielansicht

Abbildung 6: Beispielitem einer *computergenerierten Endlosschleife* in der Startansicht und der um 90 Grad rotierten Zielansicht.

Die Testpersonen wurden angeleitet, die Items mit einer von zwei möglichen Strategien zu bearbeiten, einer Perspektivwechsel- oder einer Rotationsstrategie. Ihre Aufgabe war es, zu markieren, aus welcher von vier möglichen Richtungen die Startansicht betrachtet werden muss, um mit der Zielansicht identisch zu sein (Perspektivwechselstrategie) bzw. in welche von vier möglichen Richtungen die Startansicht gedreht werden muss, um mit der Zielansicht übereinzustimmen (Rotationsstrategie). Als zusätzliche Möglichkeit konnten die Testpersonen ein Fragezeichen-Symbol für Aufgaben, die sie nicht lösen konnten, markieren. Abbildung 7 zeigt das Antwortschema für eine Testaufgabe mit der beispielhaft vorgenommenen Markierung der Alternative „von rechts“ (Perspektivwechselstrategie).

Start- und Zielansicht sind gleich, wenn die Startansicht ...

Aufgabe 1	
... von oben ↓ links → rechts ↑ unten betrachtet wird	... nach gedreht wird.

Abbildung 7: Antwortschema für eine Testaufgabe mit einer beispielhaft vorgenommenen Markierung. Der linke Teil spiegelt die Perspektivwechselstrategie, der rechte Teil die Rotationsstrategie wider. Das „?“ steht als zusätzliche Auswahlmöglichkeit mit dem sinngemäßen Inhalt „Ich weiß die Lösung nicht“ zur Verfügung.

Es wurden drei verschiedene Testversionen mit jeweils 15 Items verwendet (Form a, Form b, Form c), wobei die ersten fünf Items in allen Versionen dieselben waren (so genannte Link-Items). Die Aufgaben 6 bis 15 unterschieden sich folgendermaßen: In der Form b wurden Start- und Zielansicht im Vergleich zu Form a vertauscht. In der Form c waren die Items 6 bis 15 analog der Form a abgebildet, allerdings konnten die Testpersonen bei diesen zehn Aufgaben die Bearbeitungsrichtung frei wählen, d.h. jede einzelne Aufgabe von links oben nach rechts unten oder von rechts unten nach links oben bearbeiten. Tabelle 5 beschreibt zusammenfassend die drei Testversionen.

Tabelle 5: Beschreibung der drei Testversionen

Testform	Link-Items	weitere Items
Form a	Items 1 bis 5	Items 6 bis 15
Form b		Items 6 bis 15 analog a, nur mit vertauschter Start- und Zielansicht
Form c		Items 6 bis 15 analog a mit der Möglichkeit, die Items wie bei Form a oder in umgekehrter Richtung (wie bei Form b) zu bearbeiten

Die Testpersonen, die die Form c bearbeiteten, wurden durch eine zusätzliche Instruktion samt Beispielaufgaben nach den Items 1 bis 5 auf den neuen Vorgabemodus hingewiesen; auch das Antwortschema wurde entsprechend adaptiert (siehe Abbildungen 8 und 9).

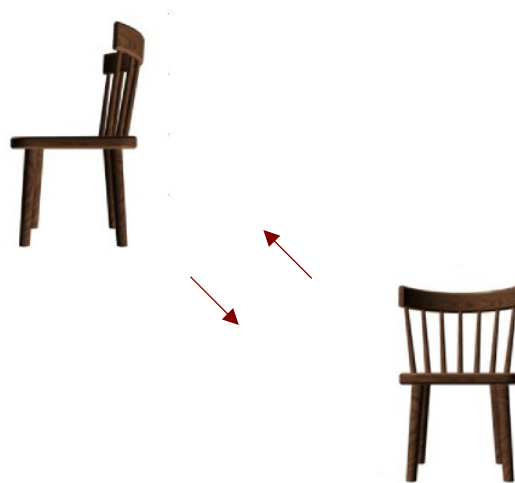


Abbildung 8: Beispielaufgabe aus der zusätzlichen Instruktion in der Form c. Die zwei Pfeile zeigen an, dass die Aufgabe von links oben nach rechts unten oder von rechts unten nach links oben bearbeitet werden kann.

Start- und Zielansicht sind gleich, wenn
bei gewählter Bearbeitungsrichtung die Startansicht ...

Aufgabe 6	
1) Bearbeitungsrichtung:	Start Ziel Ziel Start
2) Lösung:	... von oben links rechts unten betrachtet wird ... nach gedreht wird. ?

Abbildung 9: Antwortschema für die Items 6 bis 15 in der Form c, wobei für jede Aufgabe zwei Markierungen vorzunehmen sind. Die Testpersonen müssen einerseits die Bearbeitungsrichtung und andererseits auf dem Perspektivwechsel- oder dem Rotationsschema die Lösung ankreuzen.

3.3.2 Zusätzliches Testmaterial

Neben den Endlosschleifenaufgaben wurden Sozialdaten zu Geschlecht, Alter und Ausbildung erhoben sowie die Frage nach dem Besuch von Darstellendem Geometrieunterricht gestellt. Weiters wurden den Testpersonen zwei Fragebogen vorgelegt, einerseits ein mit G. Gittler erstellter Fragebogen zu den Testaufgaben sowie dem Bearbeitungsverhalten, andererseits ein Händigkeitsfragebogen (siehe Anhang 1, 2 und 3).

Ersterer beinhaltet unter anderem Aussagen zur Einschätzung der eigenen Raumvorstellungsfähigkeit, Aussagen zum Testmaterial (z.B. ob der Aufgabentypus der Testperson gefallen hat), Aussagen zum Antwortverhalten (z.B. ob die Testperson die gegebenen Antworten genau oder nicht genau überprüft hat) sowie zu Bearbeitungsstrategien (z.B. ob die Testperson versucht hat, sich die Testaufgaben mit Hilfe von Bildern oder mit Hilfe verbaler Benennungen zu verdeutlichen).

Der Händigkeitfragebogen basierte auf einem Fragebogen von Annett (1970) und wurde in übersetzter und abgeänderter Form von G. Gittler übergeben. Die Testpersonen sollten angeben, mit welcher Hand sie gewöhnlich verschiedene Tätigkeiten wie schreiben, schneiden, einen Ball werfen, ein Streichholz anzünden etc. ausüben. Als Antwortmöglichkeiten standen „rechts“, „links“, „wechselnd“ sowie „weiß nicht“ zur Verfügung.

3.4 Weitere Fragestellungen

Neben der oben formulierten Hauptfragestellung, ob bei computergenerierten Endlosschleifen nach Vertauschung der Start- und Zielansicht unterschiedliche Itemparameter auftreten, können weitere interessante Fragestellungen formuliert werden, nämlich ob

- sich unterschiedliche Bearbeitungsstrategien (Perspektivwechsel versus Rotation) auf die Testleistung auswirken,
- sich die Leistungen in den Testformen voneinander unterscheiden,
- Geschlechtsunterschiede bei *computergenerierten Endlosschleifenaufgaben* auftreten,
- ein Zusammenhang zwischen Händigkeit und Testleistung beobachtet werden kann,
- sich Persönlichkeitseigenschaften bzw. die Art der Aufgabenbearbeitung auf die Leistung auswirken, das heißt, ob beispielsweise Personen besser abschneiden,
 - denen das Testmaterial sympathisch ist,
 - die sich bezüglich ihrer Raumvorstellungsfähigkeiten besser einschätzen,
 - die ihre Antworten weniger spontan geben und genauer überprüfen.

In den folgenden Kapiteln sollen die erhobenen Sozialdaten der Stichprobe deskriptiv statistisch beschrieben sowie die Verteilung der Testpersonen auf die drei Testversionen dargestellt werden. Schließlich wird über die vorgenommene Item- und Datenselektion sowie die Verteilung des Scores berichtet.

Prozentangaben werden im gesamten empirischen Teil auf eine Nachkommastelle gerundet, bei Wahrscheinlichkeiten und Kennzahlen statistischer Tests werden drei Nachkommastellen angegeben; alle übrigen Zahlen werden mit zwei Nachkommastellen angeführt.

3.5 Stichprobenbeschreibung

Die Daten wurden im Jahr 2003 größtenteils an Wiener Schulen erhoben, einige wenige Testpersonen wurden aus dem Bekanntenkreis rekrutiert.

3.5.1 Geschlecht

Es wurden 294 Personen getestet, davon waren 130 männlich (44,2%), 163 weiblich (55,4%) und 1 Person (0,3%) ohne Angabe (siehe Abbildung 10).

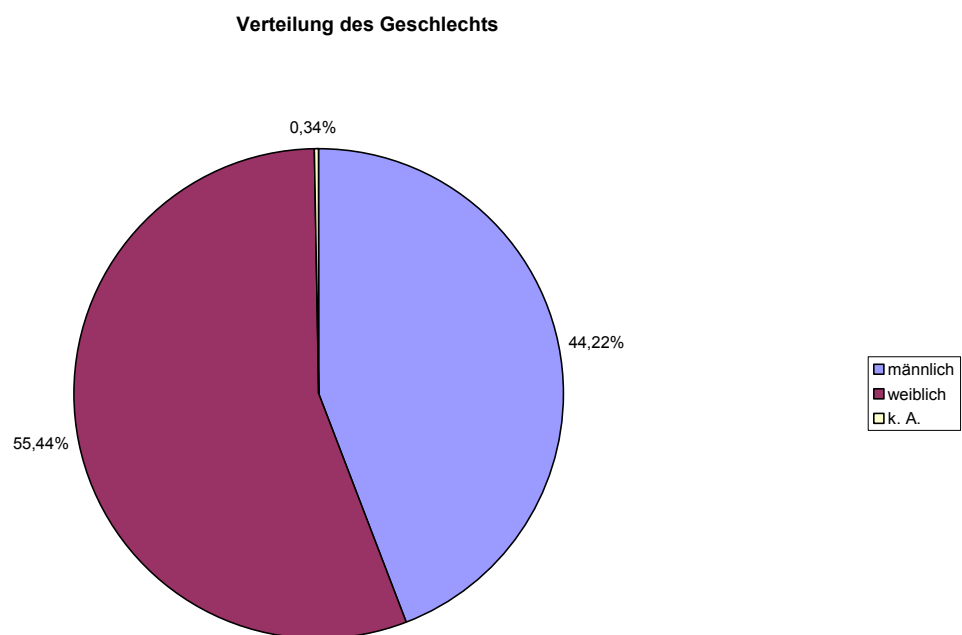


Abbildung 10: Verteilung des Geschlechts in der erhobenen Stichprobe

3.5.2 Alter

Das Alter der Testpersonen lag zwischen 14 und 33 Jahren, mit einem Schwerpunkt zwischen 14 und 18 Jahren (93,9%); eine Testperson vergaß, eine Altersangabe zu machen (0,3%). Tabelle 6 und Abbildung 11 zeigen die Altersverteilung in der erhobenen Stichprobe.

Tabelle 6: Altersverteilung in der erhobenen Stichprobe mit den Häufigkeiten in Zahlen (n) und Prozent (%).

Alter	n	%
14	7	2,4
15	44	15,0
16	101	34,4
17	103	35,0
18	20	6,8
19	2	0,7
20	1	0,3
21	1	0,3
24	9	3,1
25	2	0,7
26	1	0,3
27	1	0,3
33	1	0,3
ohne Angabe	1	0,3
Gesamt	294	100

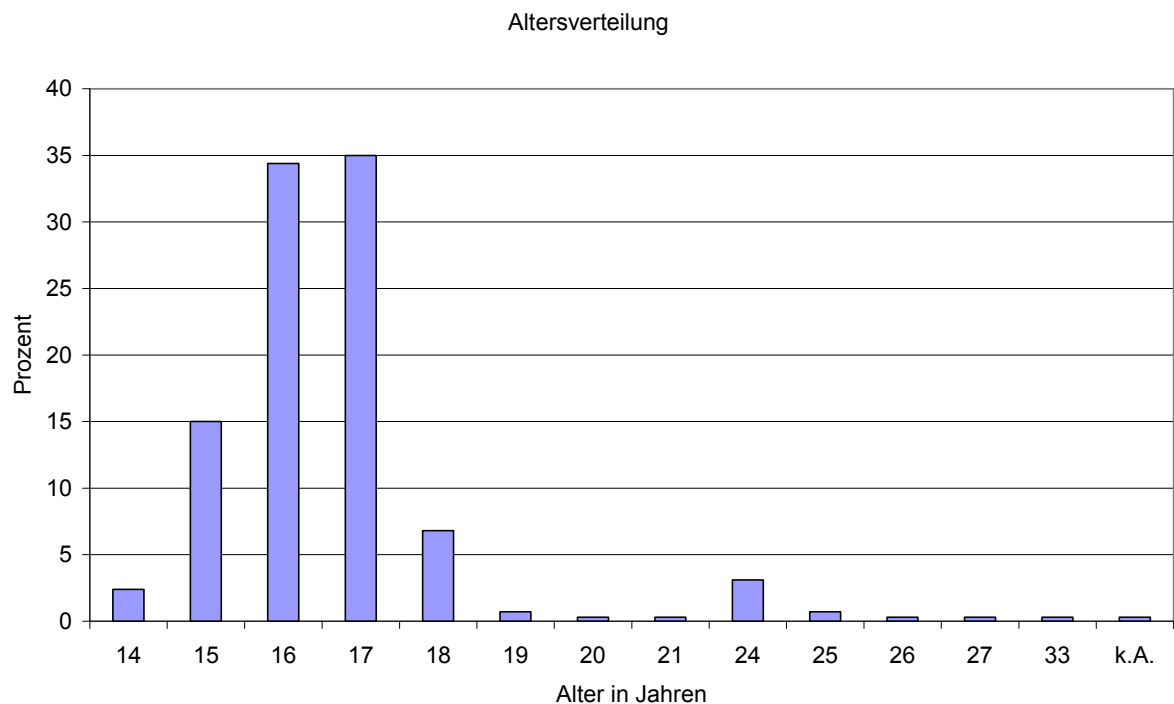


Abbildung 11: Altersverteilung in der erhobenen Stichprobe

3.5.3 Bildungsstand

Von den 294 Testpersonen hatten 7 (2,4%) ihre Ausbildung abgeschlossen, 287 (97,6%) befanden sich noch in Ausbildung. Von den 7 Personen mit abgeschlossener Ausbildung hatte 1 Person eine technische Studienrichtung absolviert, 6 Personen nicht-technische Studienrichtungen. Von den 287 Personen, die sich in Ausbildung befanden, strebten 278 Personen eine AHS-Matura an, 1 Person strebte einen Abschluss einer technischen Studienrichtung an, 8 Personen den Abschluss einer nicht-technischen Studienrichtung (siehe auch Tabelle 7).

Tabelle 7: Bildungsstand in der erhobenen Stichprobe mit den Häufigkeiten der angestrebten bzw. abgeschlossenen Ausbildungen in Zahlen (n) und Prozent (%).

		Art der angestrebten bzw. abgeschlossenen Ausbildung			n	%
		AHS-Matura	techn. Studienrichtung	nicht-techn. Studienrichtung		
noch in Ausbildung	ja	278	1	8	287	97,6
	nein	0	1	6	7	2,4
Gesamt		278	2	14	294	100

3.5.4 Besuch von Unterricht in Darstellender Geometrie

Von den 294 getesteten Personen besuchten 39 (13,3%) Unterricht in Darstellender Geometrie, 253 Personen (86,1%) hatten keinen DG-Unterricht und 2 (0,7%) Personen machten keine Angabe.

3.5.5 Bearbeitungszeit

Um die Testbearbeitungszeit für die 15 *Endlosschleifenaufgaben* zu messen, wurden die Testpersonen gebeten, die Uhrzeit nach dem Durcharbeiten der Instruktion (samt Beispielaufgaben) sowie nach Bearbeitung des letzten Items selbst in das Antwortheft einzutragen. Die Bearbeitungszeit lag zwischen 3 und 60 Minuten, wobei die Berechnung bei 12 Testpersonen (4,1%) nicht möglich war, da Zeitangaben fehlten oder nicht korrekt waren. Tabelle 8 sowie Abbildung 12 zeigen die Verteilung der Bearbeitungszeit.

Tabelle 8: Verteilung der Bearbeitungszeit mit den Häufigkeiten der berechneten Zeiten (in Minuten) in Zahlen (n) und Prozent (%).

Zeit in min	n	%	Zeit in min	n	%
3	2	0,7	23	6	2,0
6	5	1,7	24	6	2,0
7	7	2,4	25	13	4,4
8	6	2,0	26	6	2,0
9	9	3,1	27	5	1,7
10	22	7,5	28	2	0,7
11	6	2,0	29	9	3,1
12	10	3,4	30	8	2,7
13	11	3,7	31	1	0,3
14	12	4,1	32	2	0,7
15	40	13,6	33	4	1,4
16	12	4,1	35	2	0,7
17	17	5,8	38	1	0,3
18	14	4,8	41	1	0,3
19	8	2,7	47	1	0,3
20	22	7,5	60	1	0,3
21	6	2,0	ohne Angabe	12	4,1
22	5	1,7	Gesamt	294	100

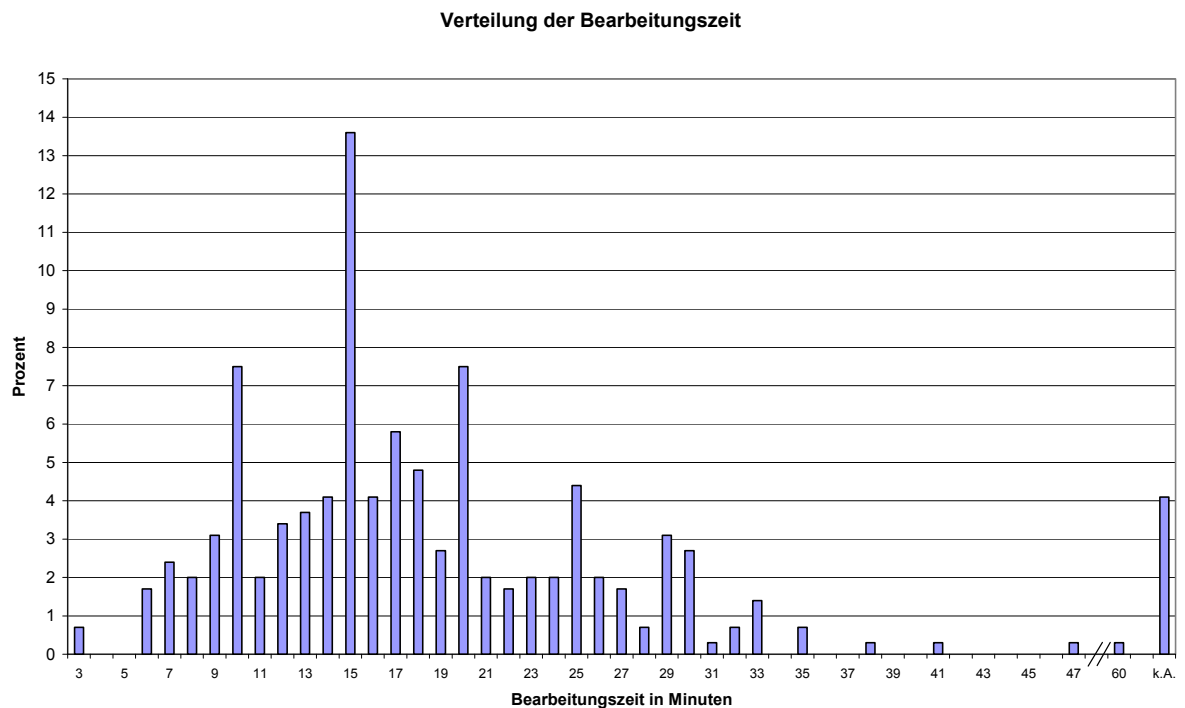


Abbildung 12: Verteilung der Bearbeitungszeit in Minuten sowie „k. A.“ (keine Angabe) für die Personen mit ausgelassenen oder fehlerhaften Zeitangaben.

3.5.6 Verteilung der Testpersonen auf die Testversionen

Die Testhefte wurden den Testpersonen randomisiert zugewiesen. 98 Personen (33,3%) bearbeiteten die Form a, 99 Personen (33,7%) die Form b und 97 Personen (33,0%) die Form c (siehe Tabelle 9).

Tabelle 9: Verteilung der Testversionen auf die Testpersonen mit den Häufigkeiten in Zahlen (n) und Prozent (%).

Testversion	n	%
Form a	98	33,3
Form b	99	33,7
Form c	97	33,0
Gesamt	294	100,0

3.5.7 Item- und Datenselektion

3.5.7.1 Itemselektion

Viele Testpersonen bezeichneten das elfte Item in mündlichen Rückmeldungen als „unlösbar“ bzw. „mehrdeutig“, was sich auch in der Wahl der Antwortalternativen widerspiegelte: 77 von 294 Personen (26,2%) kreuzten bei dieser Aufgabe das Fragezeichen-Symbol an, weitere 6 Personen (2,0%) nahmen keine Markierung vor. Die Aufgabe konnte von 40 Testpersonen (13,6%) gelöst werden, 254 Personen (86,4%) lösten das Item nicht.

Bei Betrachtung des Items (Abbildung 13) wird deutlich, dass die Lösung der Aufgabe, auch wenn die zu markierenden Alternativen bekannt sind – nämlich die Betrachtung „von unten“ (Perspektivwechsel) bzw. die Drehung „nach oben“ (Rotation) – schwer nachvollziehbar ist. Ein Grund dafür dürfte fehlende Tiefeninformation sein.

Startansicht



Zielansicht

Abbildung 13: Item 11, wie es in der Form a vorgegeben wurde. Die Lösung lautet „von unten“ (Perspektivwechsel) bzw. „Drehung nach oben“ (Rotation).

Sämtliche Auswertungen werden daher mit dem um diese Aufgabe reduzierten Satz von 14 Items durchgeführt.

3.5.7.2 Datenselektion

In einem ersten Schritt der Datenselektion wurde überprüft, wie viele Fragezeichen-Symbole jede Testperson gewählt und wie viele Items sie ausgelassen bzw. fehlerhaft markiert hatte (z.B. Markierung von zwei Lösungen oder keine Markierung der Bearbeitungsrichtung bei den Items 6 bis 15 der Form c). Eine Aufschlüsselung der Häufigkeiten in Zahlen und Prozent ist den Tabellen 10 bis 12 zu entnehmen.

Aus Tabelle 12 ist ersichtlich, dass insgesamt fünf Testpersonen bei sieben oder mehr Items (von 14) das Fragezeichen-Symbol angekreuzt bzw. die Aufgaben ausgelassen oder fehlerhaft markiert haben. Diese fünf Personen werden von allen weiteren Analysen ausgeschlossen; es resultiert daher eine Stichprobengröße von 289 Testpersonen.

Tabelle 10 Häufigkeiten der markierten Fragezeichen-Symbole in Zahlen (n) und Prozent (%)

Anzahl ?-Symbol	N	%
0	199	67,7
1	42	14,3
2	32	10,9
3	10	3,4
4	4	1,4
5	3	1,0
6	2	0,7
7	1	0,3
13	1	0,3
Gesamt	294	100,0

Tabelle 11 Häufigkeiten der Auslassungen (¥) bzw. fehlerhaften Markierungen in Zahlen (n) und Prozent (%).

Anzahl ¥ / fehlerhafte Markierung	n	%
0	261	88,8
1	21	7,1
2	5	1,7
3	2	0,7
4	2	0,7
5	1	0,3
6	1	0,3
9	1	0,3
Gesamt	294	100,0

Tabelle 12: Häufigkeiten der ausgelassenen (¥), mit dem Fragezeichen-Symbol (?) sowie fehlerhaft markierten Items in Zahlen (n) und Prozent (%).

Anzahl ¥, ? und fehlerhaften Markierungen	n	%
0	177	60,2
1	54	18,4
2	29	9,9
3	17	5,8
4	6	2,0
5	4	1,4
6	2	0,7
7	2	0,7
9	2	0,7
13	1	0,3
Gesamt	294	100,0

3.5.8 Verteilung des Scores

Die Verteilung des Scores über alle drei Testformen gemeinsam ist in Tabelle 13 bzw. Abbildung 14 dargestellt. Der maximal zu erreichende Score lag bei 14 gelösten Items.

Tabelle 13: Anzahl der gelösten Items (Score) über alle drei Testformen mit Häufigkeitsangaben in Zahlen (n) und Prozent (%).

Score	n	%
1	4	1,4
2	10	3,5
3	24	8,3
4	40	13,8
5	42	14,5
6	40	13,8
7	45	15,6
8	48	16,6
9	18	6,2
10	11	3,8
11	4	1,4
12	3	1,0
13	0	0,0
14	0	0,0
Gesamt	289	100,0

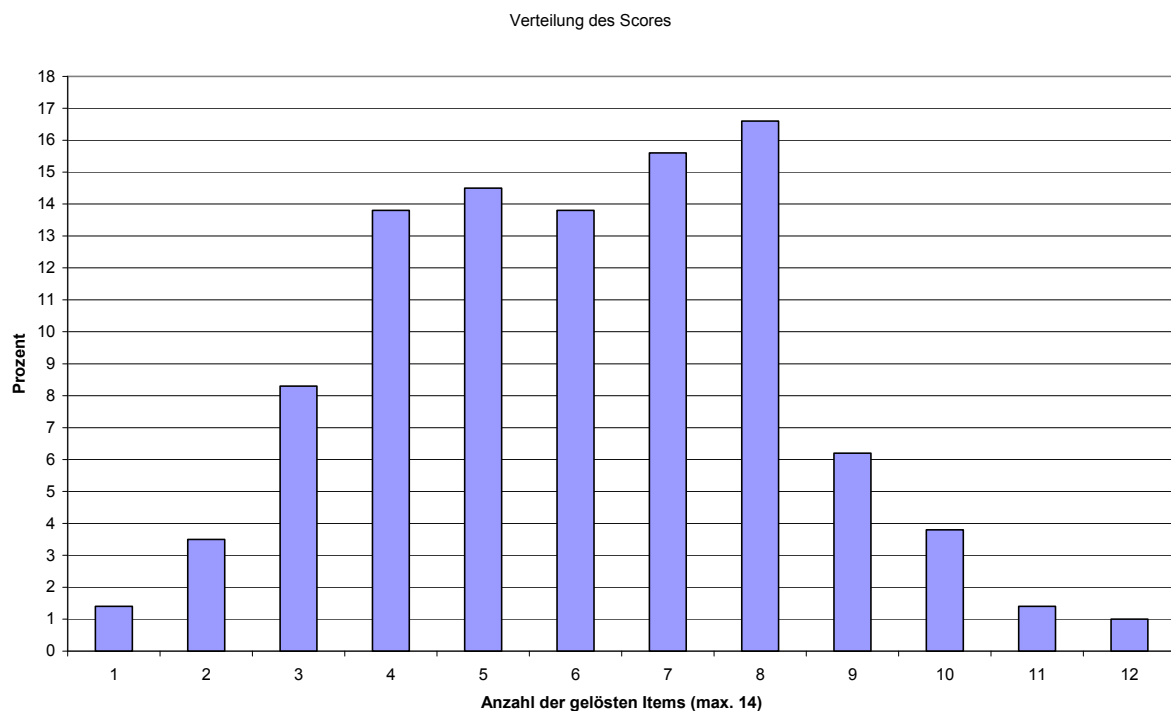


Abbildung 14: Verteilung des Scores über alle drei Testformen

Wie aus Tabelle 13 und Abbildung 14 ersichtlich lag der beobachtete Score zwischen einem gelösten Item (4 Testpersonen) und 12 gelösten Items (3 Testpersonen). Einen Überblick über die Scoreverteilung in den einzelnen Testversionen a, b und c liefern Tabelle 14 und Abbildung 15.

Tabelle 14: Verteilung des Scores in den Testformen a, b und c mit Häufigkeitsangaben in Zahlen (n_a , n_b , n_c) und Prozent ($\%_a$, $\%_b$, $\%_c$).

Score	n_a	$\%_a$	n_b	$\%_b$	n_c	$\%_c$
1	1	0,3	1	0,3	2	0,7
2	4	1,4	3	1,0	3	1,0
3	11	3,8	9	3,1	4	1,4
4	14	4,8	14	4,8	12	4,2
5	12	4,2	15	5,2	15	5,2
6	14	4,8	12	4,2	14	4,8
7	13	4,5	17	5,9	15	5,2
8	20	6,9	15	5,2	13	4,5
9	6	2,1	4	1,4	8	2,8
10	2	0,7	4	1,4	5	1,7
11	0	0,0	3	1,0	1	0,3
12	0	0,0	1	0,3	2	0,7
Gesamt	97	33,6	98	33,9	94	32,5

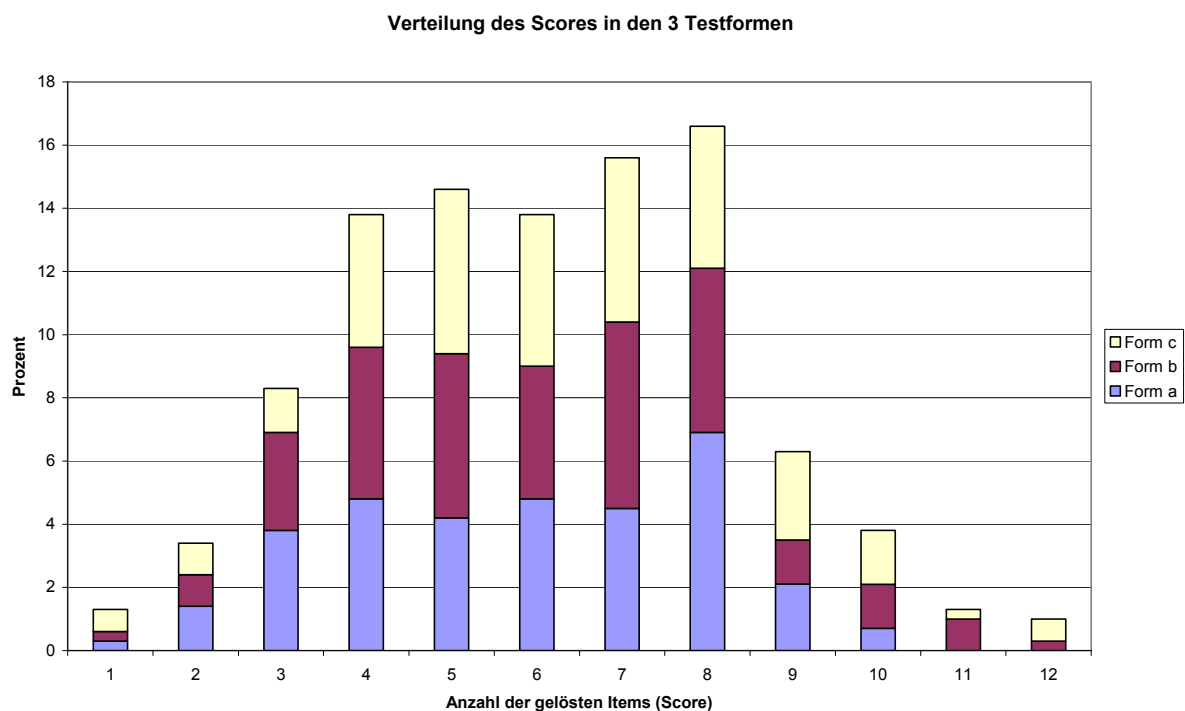


Abbildung 15: Verteilung des Scores in den drei Testformen a, b und c

Bei Betrachtung der Verteilung des Scores in den einzelnen Testformen (Tabelle 14 bzw. Abbildung 15) zeigt sich, dass in jeder Testform zumindest 1 Item gelöst wurde. Scores von 11 bzw. 12 wurden nur in den Testformen b und c erreicht, der größte beobachtete Score in der Testform a beträgt 10 gelöste Items.

Nach Darstellung der deskriptivstatistischen Kennwerte werden in den folgenden Kapiteln die Ergebnisse der Auswertungen mit Methoden der probabilistischen und klassischen Testtheorie beschrieben.

3.6 Auswertungen mit Methoden der Probabilistischen und Klassischen Testtheorie

3.6.1 Analysen nach dem Modell von RASCH

3.6.1.1 Einleitende Worte

An erster Stelle steht die Frage, ob die vorgegebenen *Endlosschleifenaufgaben* dem dichotomen logistischen Modell von Rasch (1960, zitiert nach Fischer, 1974) entsprechen.

Auf eine detaillierte Beschreibung des Modells von Rasch wird verzichtet und auf die Literatur verwiesen (z.B. Fischer, 1974, 1997; Fischer & Molenaar, 1995; Bortz & Döring, 2006; Gittler & Arendasy, 2003a), einige grundlegende Annahmen des Modells sollen dennoch dargestellt werden.

Das Rasch-Modell, das zur probabilistischen Testtheorie oder Item-Response-Theorie gezählt wird, geht vom Vorhandensein einer latenten, nicht-beobachtbaren Dimension aus, auf die durch beobachtbare (manifeste) Merkmale – die Antworten einer Testperson – zurückgeschlossen werden kann. Jede gegebene Antwort wiederum hängt von der Fähigkeit der Testperson (Personenparameter) und der Schwierigkeit der Aufgabe (Itemparameter) ab. Von zwei gegebenen Items wird jenes mit größerer Wahrscheinlichkeit gelöst, das leichter ist und von zwei Testpersonen wird jene ein Item wahrscheinlicher lösen, die höhere Fähigkeiten besitzt.

Die systematische Beziehung zwischen der beobachtbaren Reaktion und der latenten Dimension wird für jedes Item über die so genannte Itemcharakteristik erzeugt. Je nachdem, welche Ausprägung eine Testperson auf der latenten Dimension hat, besitzt sie eine bestimmte Wahrscheinlichkeit, eine Aufgabe zu lösen. Sinnvoll sind jene Aufgaben zur Messung einer latenten Dimension, die von einer Person mit höheren Fähigkeiten eher gelöst werden als von einer Person mit weniger Fähigkeiten.

Eine Grundannahme des Rasch-Modells ist die „lokale stochastische Unabhängigkeit“, die besagt, dass die Lösung eines Items nur von den Fähigkeiten einer Person und der Schwierigkeit des Items abhängt und beispielsweise nicht davon, welche Aufgaben eine Person schon bearbeitet hat.

Ob das Rasch-Modell empirisch gilt, d.h. ob die (erwünschte) Nullhypothese beibehalten werden kann, kann inferenzstatistisch mittels Likelihood-Quotienten-Tests (LQT) berechnet werden. Meistens wird hierfür der LQT nach Andersen (1973) herangezogen, der prüft, ob die Itemparameter in verschiedenen Subgruppen (männlich versus weiblich, hoher versus niedriger Rohscore, jung versus alt etc.) statistisch gleich sind.

Die Geltung des Rasch-Modells bringt einige Vorteile mit sich, nämlich

- dass die Testaufgaben eindimensional bzw. homogen in dem Sinn sind, dass sie für alle Personen dieselbe latente Dimension messen.
- dass die Anzahl der gelösten Aufgaben eine „erschöpfende Statistik“ darstellt. Das bedeutet, dass die Information, wie viele Aufgaben eine Person gelöst hat, ausreicht und nicht mehr von Interesse ist, welche Aufgaben sie gelöst hat. Daraus ergibt sich auch, dass der Testrohwer (Summe der gelösten Aufgaben) bei Geltung des Rasch-Modells ein faires Maß für die Leistung darstellt.
- die „spezifische Objektivität“, die besagt, dass der Vergleich von zwei Testpersonen möglich ist, unabhängig davon, welche Items sie bearbeitet haben und der Vergleich von zwei Items, unabhängig davon, welche Personen sie bearbeitet haben.

3.6.1.2 Auswertungsvarianten mit dem Rasch-Modell für die vorliegenden Daten

Aufgrund der Datenlage sind unterschiedliche Auswertungsvarianten mit dem Rasch-Modell möglich: In einem ersten Schritt werden die drei Testformen a, b und c getrennt voneinander auf Rasch-Homogenität überprüft, im Anschluss daran erfolgt die Überprüfung der Testformen gemeinsam. Bei der gemeinsamen Auswertung von Testformen können die Daten der drei Testformen a, b und c zusammen analysiert werden oder auch nur die Daten der beiden „Extremtestformen“ a und b.

Bei den Auswertungen der Testform c und den Auswertungen mehrerer Testformen gemeinsam können Rasch-Analysen für vollständige und unvollständige Daten berechnet werden. Beim unvollständigen Rasch-Modell werden die Items 6 bis 15 der verschiedenen Testformen so behandelt, als wären sie unterschiedliche Aufgaben (die Items 1 bis 5 sind Link-Items und daher für alle Testformen gleich). Das wird erreicht, indem die Items 6 bis 15 der Form a die Positionen 6 bis 15 zugewiesen bekommen und die Items 6 bis 15 der Form b die Positionen 16 bis 25. Die Items 6 bis 15 der Form c werden auf die Positionen 6 bis 25 umgeschrieben, in Abhängigkeit davon, ob die Testperson jedes einzelne Item von der Richtung analog der Form a oder analog der Form b bearbeitet hat.

Beim Rasch-Modell für vollständige Daten werden die Items 6 bis 15 der verschiedenen Testformen so behandelt als wären sie gleich, d.h. jedes Item bleibt – unabhängig von der Bearbeitungsrichtung – auf seiner Position. (Das Item 11 wurde - wie im Kapitel 3.5.7 „Item- und Datenselektion“ beschrieben - ausgeschieden, die Items behalten aber zwecks Übersichtlichkeit ihre ursprünglichen Nummern; d.h. Item 12 wird nach Itemselektion nicht Item 11 genannt, Item 13 nicht Item 12 etc. Daher wird auch von den Positionen 6 bis 15 und 16 bis 25 gesprochen und nicht 6 bis 14 bzw. 15 bis 23).

Schließlich ergeben sich noch zwei Auswertungsvarianten aus der Überlegung, wie mit fehlerhaft oder dem Fragezeichen-Symbol markierten bzw. ausgelassenen Aufgaben umgegangen wird. Eine Auswertungsstrategie ist, alle Items in die Analysen einzubeziehen und Aufgaben, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden mit „nicht gelöst“ gleichzusetzen. Die Itemanzahl beträgt dann für alle Testpersonen 14 (15 minus 1). Die zweite Möglichkeit ist, Auslassungen, fehlerhaft und mit dem Fragezeichen-Symbol markierte Aufgaben so zu bewerten, als wären diese Items nicht vorgegeben worden. In diesem Fall liegt die Anzahl der bearbeiteten Aufgaben in den vorliegenden Daten zwischen 8 und 14.

Aus der Kombination der eben beschriebenen Auswertungsvarianten ergeben sich für die Rasch-Analysen der einzelnen Testformen getrennt voneinander insgesamt acht verschiedene Berechnungsmöglichkeiten und für die Analyse der Testformen gemeinsam ebenso acht Auswertungsmodelle, die in den Tabellen 15 und 16 überblicksmäßig dargestellt werden.

Tabelle 15: Auswertungsmöglichkeiten mit dem Rasch-Modell für die einzelnen Testformen getrennt voneinander mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden sowie der Bezeichnung der Modelle.

Testform	A		b		C			
Itemanzahl	14 Items		14 Items		23 Items		14 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell A	Modell A	Modell B	Modell b	Modell C23	Modell c23	Modell C	Modell c

„“, ¥, ? nicht bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben so behandelt werden, als wären sie nicht vorgegeben worden; **alle Items** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben als nicht gelöst gewertet werden.

Tabelle 16: Auswertungsmöglichkeiten mit dem Rasch-Modell für die Testformen gemeinsam mit der Information, welche Testformen in die Analysen eingehen (vier Varianten mit allen Testformen und vier Varianten mit den Formen a und b), mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden und der Bezeichnung der Modelle.

Testform	alle Testformen				Testformen a und b			
Itemanzahl	23 Items		14 Items		23 Items		14 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell 1a	Modell 1b	Modell 2a	Modell 2b	Modell 3a	Modell 3b	Modell 4a	Modell 4b

„“, ¥, ? **nicht** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben so behandelt werden, als wären sie nicht vorgegeben worden; **alle Items** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben als nicht gelöst gewertet werden.

Bei allen Rasch-Analysen wird die Irrtumswahrscheinlichkeit für die Likelihood-Quotienten-Tests mit $\alpha = 1\%$ festgelegt. Als interne Teilungskriterien werden Mittelwert und Median herangezogen, als externe Teilungskriterien werden Geschlecht, Alter (Teilung nach dem Median des Alters), Bearbeitungszeit (Teilung nach dem Median der Bearbeitungszeit), der Besuch von Darstellendem Geometrieunterricht (DG) und – sofern aus der Datenlage möglich – die Testform eingesetzt.

Testpersonen, die Angaben zum Geschlecht (1 Testperson), Alter (1 Testperson), zur Bearbeitungszeit (10 Testpersonen) und zum Besuch von Darstellendem Geometrieunterricht (2 Testpersonen) nicht bzw. fehlerhaft vornahmen, wurden zufällig einer Teilungsgruppe zugewiesen.

Im Folgenden sollen die in den Tabellen 15 und 16 dargestellten Modelle ausführlicher beschrieben werden.

Rasch-Analysen für die drei Testformen getrennt voneinander

Das **Modell A** (Tabelle 17) ist ein vollständiges Rasch-Modell für die Testform a mit 97 Testpersonen, 14 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 29 verschiedenen Untergruppen führt. Die Gesamtl likelihood beträgt -592,56.

Tabelle 17: Rasch-Analysen für das Modell A mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	P
Mittelwert			13	18,14	27,72	0,152
Niedrig	61	-248,18				
Hoch	36	-235,31				
Median			13	18,89	27,72	0,127
Niedrig	64	-366,47				
Hoch	33	-216,65				
Geschlecht			13	15,49	27,72	0,278
Männlich	37	-225,97				
Weiblich	60	-358,85				
Alter			13	9,72	27,72	0,717
niedrig (14-16J.)	57	-349,06				
hoch (>16J.)	40	-238,65				
DG			13	28,53	27,72	0,008*
Nein	87	-528,02				
Ja	10	-50,28				
Zeit			13	13,50	27,72	0,410
niedrig (bis 15min)	52	320,96				
hoch (>15min)	45	264,85				

* Das Teilungskriterium „DG“ ist mit $p=0,008$ signifikant. Allerdings ist anzumerken, dass die Anzahl der Testpersonen, die Unterricht in Darstellender Geometrie besucht haben mit einer Stichprobengröße von 10 zu klein ist, um eine zuverlässige Schätzung der Parameter zu gewährleisten. Das Teilungskriterium wird trotzdem mit dem Hinweis angeführt, dass eine α -Adjustierung (zur Minimierung des Fehlers 1. Art) vorgenommen werden kann, da sechs unabhängige Tests zu einer zentralen Hypothese gerechnet werden. Die Formel für das adjustierte α lautet: $\alpha' = 1 - (1 - \alpha)^{1/m}$ (m =Anzahl der Tests; Formel aus Bortz & Schuster, 2010, S. 230), woraus sich bei einer Irrtumswahrscheinlichkeit von 1% ein α' von 0,0017 ergibt. Das bedeutet, dass das Ergebnis von 0,008 nicht signifikant ist.

Das **Modell a** (Tabelle 18) ist ein vollständiges Rasch-Modell für die Testform a mit 97 Testpersonen, 14 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht gelöst gewertet. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -621,68.

Das **Modell B** (Tabelle 19) ist ein vollständiges Rasch-Modell für die Testform b mit 98 Testpersonen, 14 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 31 verschiedenen Untergruppen führt. Die Gesamtl likelihood beträgt -614,98.

Tabelle 18: Rasch-Analysen für das Modell a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			13	21,44	27,72	0,065
Niedrig	42	-240,05				
Hoch	55	-370,92				
Median			13	9,93	27,72	0,699
Niedrig	56	-336,01				
Hoch	41	-280,71				
Geschlecht			13	13,74	27,72	0,393
männlich	37	-232,31				
weiblich	60	-382,51				
Alter			13	10,53	27,72	0,650
niedrig (14-16J.)	57	-367,47				
hoch (>16J.)	40	-248,95				
DG			13	27,27	27,72	0,011
Nein	87	-557,29				
Ja	10	-50,76				
Zeit			13	12,34	27,72	0,500
niedrig (bis 15min)	52	-342,00				
hoch (>15min)	45	-273,51				

Tabelle 19: Rasch-Analysen für das Modell B mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			13	18,50	27,72	0,139
Niedrig	63	-373,39				
Hoch	35	-232,34				
Median			13	9,01	27,72	0,772
Niedrig	74	-459,32				
Hoch	24	-151,15				
Geschlecht			13	16,23	27,72	0,237
männlich	48	-302,30				
weiblich	50	-304,56				
Alter			13	17,32	27,72	0,185
niedrig (14-16J.)	48	-296,08				
hoch (>16J.)	50	-310,24				
DG			13	21,24	27,72	0,068
Nein	83	-521,17				
Ja	15	-83,19				
Zeit			13	14,27	27,72	0,355
niedrig (bis 15min)	48	-309,62				
hoch (>15min)	50	-298,23				

Das **Modell b** (Tabelle 20) ist ein vollständiges Rasch-Modell für die Testform b mit 98 Testpersonen, 14 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht gelöst gewertet. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -655,86.

Tabelle 20 kann entnommen werden, dass die Teilungskriterien Mittelwert und Median signifikant sind. Bei der Kontrolle der z-Werte zeigt die fünfte Aufgabe mit $z = 5,36$ den auffälligsten Wert (siehe Tabelle 21). Diese Aufgabe wurde in allen Testversionen als Link-Item vorgegeben und führte in den bislang berechneten Modellen A, a und B zu keinen signifikanten Likelihood-Quotienten-Tests.

Tabelle 20: Rasch-Analysen für das Modell b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			13	29,76	27,72	0,005
Niedrig	54	-352,16				
Hoch	44	-288,83				
Median			13	29,76	27,72	0,005
Niedrig	54	-352,16				
Hoch	44	-288,83				
Geschlecht			13	19,86	27,72	0,099
männlich	48	-317,09				
weiblich	50	-328,84				
Alter			13	19,80	27,72	0,100
niedrig (14-16J.)	48	-314,00				
hoch (>16J.)	50	-331,96				
DG			13	19,85	27,72	0,099
Nein	83	-560,00				
Ja	15	-85,93				
Zeit			13	16,22	27,72	0,237
niedrig (bis 15min)	48	-324,16				
hoch (>15min)	50	-323,60				

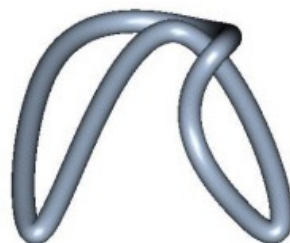
Bei genauer Betrachtung, wie viele und welche Testpersonen mit der Testform b das Item 5 korrekt bearbeiten konnten, präsentiert sich folgendes Bild: Insgesamt markierten 33 Personen (33,7%) die korrekte Alternative, 65 Testpersonen (66,3%) lösten die Aufgabe nicht. Wird die Stichprobe anhand des Medians in die zwei Gruppen „gute Raumvorsteller“ und „schlechte Raumvorsteller“ geteilt, zeigt sich, dass die 54 schlechten Raumvorsteller (55,1%) mit einem Score zwischen 1 und 6 bei dem Item 5 mit 35,2% eine höhere Lösungswahrscheinlichkeit haben als die 44 guten Raumvorsteller (44,9%) mit einer Lösungswahrscheinlichkeit von 28,2%.

lichkeit von 31,8%. Mangelnde Tiefeninformation dürfte die Ursache für diese widersprüchlichen Ergebnisse sein (siehe Abbildung 16).

Tabelle 21: z-Werte der Items 1 bis 15 (ohne Item 11) aus dem Modell b mit der Itemnummer (Itemnr.), den z-Werten und den Signifikanz-Angaben.

Itemnr.	z-Werte	Signifikanz
1	-1,5295	n.s.
2	2,2664	sig. (5%)
3	-1,6710	n.s.
4	-0,8706	n.s.
5	5,3600	sig. (5% & 1%)
6	0,3337	n.s.
7	-1,0387	n.s.
8	0,1305	n.s.
9	-0,9436	n.s.
10	-1,2925	n.s.
12	-1,3037	n.s.
13	0,4952	n.s.
14	1,1638	n.s.
15	3,9896	sig. (5% & 1%)

Startansicht



Zielansicht

Abbildung 16: Item 5

Dieses Item kann also nicht zu einer zuverlässigen Messung der Raumvorstellungsfähigkeit beitragen und muss für alle weiteren Analysen ausgeschieden werden. Ebenso müssen die Rasch-Analysen für die Modelle A, a, B und b ohne Item 5 wiederholt werden.

Wiederholung der Rasch-Analysen für die Modelle A, a, B und b ohne Item 5

Aufgrund der notwendigen Eliminierung von Item 5 ändert sich die Anzahl der Testaufgaben, die in die Rasch-Analysen eingehen (siehe Tabelle 22).

Tabelle 22: Neuerliche Darstellung der Auswertungsmöglichkeiten mit dem Rasch-Modell für die einzelnen Testformen getrennt voneinander mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden sowie der Bezeichnung der Modelle.

Testform	A		B		c			
Itemanzahl	13 Items		13 Items		22 Items		13 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell A	Modell A	Modell B	Modell B	Modell C22	Modell c22	Modell C	Modell c

„“, ¥, ? nicht bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben so behandelt werden, als wären sie nicht vorgegeben worden; **alle Items** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben als nicht gelöst gewertet werden.

Das **Modell A** (Tabelle 23) ist ein vollständiges Rasch-Modell für die Testform a mit 97 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 25 verschiedenen Untergruppen führt. Die Gesamtl likelihood beträgt -531,64.

Das **Modell a** (Tabelle 24) ist ein vollständiges Rasch-Modell für die Testform a mit 97 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht gelöst gewertet. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -553,03.

Tabelle 23: Neuerliche Rasch-Analysen für das Modell A mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	10,36	26,25	0,584
niedrig	50	-247,73				
hoch	47	-278,73				
Median			12	13,30	26,25	0,347
niedrig	60	-310,96				
hoch	37	-214,04				
Geschlecht			12	11,95	26,25	0,450
männlich	37	-202,68				
weiblich	60	-322,99				
Alter			12	9,60	26,25	0,651
niedrig (14-16J.)	57	-313,68				
hoch (>16J.)	40	-213,17				
DG			12	13,75	26,25	0,317
nein	87	-474,50				
ja	10	-50,28				
Zeit			12	11,83	26,25	0,460
niedrig (bis 15min)	52	-287,67				
hoch (>15min)	45	-238,06				

Tabelle 24: Neuerliche Rasch-Analysen für das Modell a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	6,34	26,25	0,898
niedrig	44	-225,37				
hoch	53	-324,50				
Median			12	7,62	26,25	0,814
niedrig	61	-332,34				
hoch	36	-216,87				
Geschlecht			12	10,49	26,25	0,573
männlich	37	-206,38				
weiblich	60	-341,40				
Alter			12	10,69	26,25	0,555
niedrig (14-16J.)	57	-327,75				
hoch (>16J.)	40	-219,93				
DG			12	12,05	26,25	0,442
nein	87	-496,24				
ja	10	-50,76				
Zeit			12	10,67	26,25	0,557
niedrig (bis 15min)	52	-303,48				
hoch (>15min)	45	-244,22				

Das **Modell B** (Tabelle 25) ist ein vollständiges Rasch-Modell für die Testform b mit 98 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 29 verschiedenen Untergruppen führt. Die Gesamtl likelihood beträgt -557,27.

Tabelle 25: Neuerliche Rasch-Analysen für das Modell B mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	7,75	26,25	0,805
niedrig	61	-325,95				
hoch	37	-227,45				
Median			12	6,57	26,25	0,885
niedrig	72	-404,32				
hoch	26	-149,67				
Geschlecht			12	15,34	26,25	0,224
männlich	48	-270,79				
weiblich	50	-278,82				
Alter			12	15,28	26,25	0,226
niedrig (14-16J.)	48	-267,77				
hoch (>16J.)	50	-281,86				
DG			12	20,43	26,25	0,059
nein	83	-471,82				
ja	15	-75,24				
Zeit			12	7,69	26,25	0,809
niedrig (bis 15min)	48	-279,45				
hoch (>15min)	50	-273,98				

Das **Modell b** (Tabelle 26) ist ein vollständiges Rasch-Modell für die Testform b mit 98 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht gelöst gewertet. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -594,00.

Das **Modell C22** (Tabelle 27) ist ein unvollständiges Rasch-Modell für die Testform c mit 94 Testpersonen, 22 Items und 4 externen Teilungskriterien. Die Items 6 bis 15 werden (wie oben beschrieben) auf die Positionen 6 bis 15 umgeschrieben, wenn sie analog der Testform a bearbeitet wurden und auf die Positionen 16 bis 25, wenn sie analog der Testform b bearbeitet wurden. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet. Es gibt 72 verschiedene Untergruppen. Die Gesamtl likelihood beträgt -524,52.

Tabelle 26: Neuerliche Rasch-Analysen für das Modell b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	14,98	26,25	0,242
Niedrig	48	-278,21				
Hoch	50	-308,30				
Median			12	14,03	26,25	0,299
Niedrig	57	-339,14				
Hoch	41	-247,84				
Geschlecht			12	18,04	26,25	0,115
Männlich	48	-284,37				
Weiblich	50	-300,60				
Alter			12	19,22	26,25	0,083
niedrig (14-16J.)	48	-281,43				
hoch (>16J.)	50	-302,96				
DG			12	19,64	26,25	0,074
Nein	83	-507,04				
Ja	15	-77,14				
Zeit			12	10,58	26,25	0,565
niedrig (bis 15min)	48	-291,35				
hoch (>15min)	50	-297,36				

Tabelle 27: Rasch-Analysen für das Modell C22 mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			21	23,04	38,96	0,342
Niedrig	84	-464,32				
Hoch	10	-48,67				
Median			21	23,04	38,96	0,342
Niedrig	84	-464,32				
Hoch	10	-48,67				
Geschlecht			21	16,04	38,96	0,767
Männlich	44	-242,54				
Weiblich	50	-273,95				
Alter			21	17,74	38,96	0,666
niedrig (14-16J.)	47	-248,39				
hoch (>16J.)	47	-267,26				
DG			21	24,63	38,96	0,263
Nein	80	-441,92				
Ja	14	-70,28				
Zeit			21	28,03	38,96	0,139
niedrig (bis 18min)	50	-293,70				
hoch (>18min)	44	-216,80				

Das **Modell c22** ist ein unvollständiges Rasch-Modell für die Testform c. Es werden alle Items als vorgegeben betrachtet, d.h. dass fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items als „nicht gelöst“ gewertet werden. Das führt dazu, dass nur jene 63 von 94 Testpersonen in die Analysen einbezogen werden können, die alle Items bearbeitet haben und deren Antworten daher der Bearbeitungsrichtung analog der Form a bzw. analog der Form b zugeordnet werden können. Auf die Berechnung dieses Rasch-Modells muss aufgrund der zu geringen Fallzahlen verzichtet werden.

Das **Modell C** (Tabelle 28) ist ein vollständiges Rasch-Modell für die Testform c mit 94 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 24 verschiedenen Untergruppen führt. Die Gesamtl likelihood beträgt -532,89.

Tabelle 28: Rasch-Analysen für das Modell C mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	8,55	26,25	0,741
Niedrig	67	-368,52				
Hoch	2	-160,09				
Median			12	8,55	26,25	0,741
Niedrig	67	-368,52				
Hoch	27	-160,09				
Geschlecht			12	11,41	26,25	0,494
Männlich	44	-248,39				
Weiblich	50	-278,80				
Alter			12	9,06	26,25	0,697
niedrig (14-16J.)	47	-257,40				
hoch (>16J.)	47	-270,96				
DG			12	11,90	26,25	0,454
Nein	80	-451,20				
Ja	14	-75,74				
Zeit			12	18,73	26,25	0,095
niedrig (bis 18min)	50	-297,91				
hoch (>18min)	44	-225,61				

Das **Modell c** (Tabelle 29) ist ein vollständiges Rasch-Modell mit der Testform c, 94 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht gelöst gewertet. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -563,73.

Tabelle 29: Rasch-Analysen für das Modell c mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	7,83	26,25	0,798
Niedrig	45	-252,88				
Hoch	49	-306,94				
Median			12	5,87	26,25	0,922
Niedrig	62	-365,94				
Hoch	32	-194,85				
Geschlecht			12	12,49	26,25	0,407
Männlich	44	-262,48				
Weiblich	50	-295,01				
Alter			12	9,16	26,25	0,690
niedrig (14-16J.)	47	-276,13				
hoch (>16J.)	47	-283,02				
DG			12	11,32	26,25	0,502
Nein	80	-479,38				
Ja	14	-78,69				
Zeit			12	22,45	26,25	0,033
niedrig (bis 18min)	50	-314,16				
hoch (>18min)	44	-238,34				

Alle in den Tabellen 23 bis 29 beschriebenen Modelltests sind nicht signifikant. Die Itemparameter in den Subgruppen niedriger versus hoher Median, männliches versus weibliches Geschlecht, niedriges versus hohes Alter, Besuch versus kein Besuch von Unterricht in Darstellender Geometrie sowie niedrige versus hohe Bearbeitungszeit unterscheiden sich also in den einzelnen Testformen a, b und c statistisch nicht voneinander.

Die Bedeutung dieser Aussage geht für die Testform c, bei der Wahlfreiheit der Bearbeitungsrichtung für die Items 6 bis 15 vorliegt, noch weiter als die Bedeutung der Aussagen für die Testformen a und b. Dass nämlich die Rasch-Analysen für die Version c mit 13 Items allen Modelltests standhalten, heißt, dass sich die Itemparameter in den verschiedenen Subgruppen nicht unterscheiden, obwohl die Items 6 bis 15 in zwei unterschiedliche Richtungen bearbeitet werden konnten und daher nicht exakt gleich sind.

Für die Items der Testform a, der Form b sowie der Version c kann also angenommen werden, dass sie im Sinne des Rasch-Modells eindimensional sind und für alle Personen dieselbe latente Dimension messen.

Rasch-Analysen für die Testformen gemeinsam

Nach der für jede Testform getrennten Überprüfung folgen die Rasch-Analysen mehrerer Testformen gemeinsam; die verschiedenen Auswertungsvarianten können der (wegen der notwendigen Eliminierung von Item 5) überarbeiteten Übersichtstabelle (Tabelle 30) entnommen werden.

Tabelle 30: Neuerliche Darstellung der Auswertungsmöglichkeiten mit dem Rasch-Modell für die Testformen gemeinsam mit der Information, welche Testformen in die Analysen eingehen (vier Varianten mit allen Testformen und vier Varianten mit den Formen a und b), mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden und der Bezeichnung der Modelle.

Testform	alle Testformen				Testformen a und b			
Itemanzahl	22 Items		13 Items		22 Items		13 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell 1a	Modell 1b	Modell 2a	Modell 2b	Modell 3a	Modell 3b	Modell 4a	Modell 4b

„“, ¥, ? nicht bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben so behandelt werden, als wären sie nicht vorgegeben worden; **alle Items** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben als nicht gelöst gewertet werden.

Das **Modell 1a** (Tabelle 31) ist ein unvollständiges Rasch-Modell für alle Testformen mit 289 Testpersonen, 22 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet. Es wurden 121 verschiedene Untergruppen beobachtet. Die Gesamtlikelihood beträgt -1636,31.

Das **Modell 1b** (Tabelle 32) ist ein unvollständiges Rasch-Modell für alle Testformen. Es werden alle Items als vorgegeben betrachtet, d.h. dass fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items als „nicht gelöst“ gewertet werden. Das führt dazu, dass bei der Testform c nur jene Testpersonen in die Analysen einbezogen werden können, die alle Items bearbeitet haben und deren Antworten daher der Bearbeitung analog der Form a bzw. analog der Form b zugeordnet werden können. Zu 195 Testpersonen, die Form a und Form b bearbeitet haben, kommen 63 Testpersonen mit der Testform c, auf die die beschriebenen Kriterien zutreffen; das sind insgesamt 258 Testpersonen. Das Modell beinhaltet 22 Items und 4 externe Teilungskriterien, es wurden 42 Untergruppen beobachtet. Die Gesamtlikelihood beträgt -1537,90.

Tabelle 31: Rasch-Analysen für das Modell 1a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			21	21,54	38,96	0,427
niedrig	204	-1117,92				
hoch	85	-507,62				
Median			21	21,54	38,96	0,427
niedrig	204	-1117,92				
hoch	85	-507,62				
Geschlecht			21	19,69	38,96	0,541
männlich	129	-731,49				
weiblich	160	-894,97				
Alter			21	16,11	38,96	0,763
niedrig (14-16J.)	152	-854,75				
hoch (>16J.)	137	-773,50				
DG			21	23,21	38,96	0,333
nein	250	-1410,16				
ja	39	-214,55				
Zeit			21	27,25	38,96	0,163
niedrig (bis 16min)	147	-853,52				
hoch (>16min)	142	-769,16				

Tabelle 32: Rasch-Analysen für das Modell 1b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			21	26,36	38,96	0,193
niedrig	139	-791,60				
hoch	119	-733,11				
Median			21	22,43	38,96	0,375
niedrig	171	-1002,78				
hoch	87	-523,90				
Geschlecht			21	28,78	38,96	0,120
männlich	112	-660,43				
weiblich	146	-863,08				
Alter			21	17,50	38,96	0,681
niedrig (14-16J.)	134	-795,01				
hoch (>16J.)	124	-734,14				
DG			21	25,06	38,96	0,245
nein	223	-1337,81				
ja	35	-187,56				
Zeit			21	28,28	38,96	0,132
niedrig (bis 16min)	134	-815,15				
hoch (>16min)	124	-708,61				

Das **Modell 2a** (Tabelle 33) ist ein vollständiges Rasch-Modell für alle Testformen mit 289 Testpersonen, 13 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 59 Untergruppen führt. Die Gesamtl likelihood beträgt -1642,02.

Tabelle 33: Rasch-Analysen für das Modell 2a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	6,60	26,25	0,883
niedrig	175	-946,59				
hoch	114	-692,14				
Median			12	6,03	26,25	0,914
niedrig	180	-975,90				
hoch	109	-663,11				
Geschlecht			12	14,48	26,25	0,271
männlich	129	-736,91				
weiblich	160	-897,87				
Alter			12	9,45	26,25	0,664
niedrig (14-16J.)	152	-858,31				
hoch (>16J.)	137	-778,99				
DG			12	8,72	26,25	0,727
nein	250	-1414,80				
ja	39	-222,86				
Zeit			12	23,00	26,25	0,028
niedrig (bis 16min)	147	-857,29				
hoch (>16min)	142	-773,22				

Das **Modell 2b** (Tabelle 34) ist ein vollständiges Rasch-Modell für alle Testformen mit 289 Testpersonen, 13 Items und 4 externen Teilungskriterien. Es werden alle Items als vorgegeben betrachtet, d.h. dass fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items als „nicht gelöst“ gewertet werden. Es gibt nur 1 Untergruppe. Die Gesamtl likelihood beträgt -1731,56.

Das **Modell 3a** (Tabelle 35) ist ein unvollständiges Rasch-Modell für die Testformen a und b mit 195 Testpersonen, 22 Items und 4 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet. Es wurden 54 verschiedene Untergruppen beobachtet. Die Gesamtl likelihood beträgt -1089,51.

Tabelle 34: Rasch-Analysen für das Modell 2b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	12,43	26,25	0,412
niedrig	137	-767,07				
hoch	152	-958,28				
Median			12	9,39	26,25	0,670
niedrig	180	-1053,18				
hoch	109	-673,68				
Geschlecht			12	16,11	26,25	0,186
männlich	129	-770,17				
weiblich	160	-953,33				
Alter			12	10,14	26,25	0,603
niedrig (14-16J.)	152	-904,93				
hoch (>16J.)	137	-821,56				
DG			12	7,70	26,25	0,808
nein	250	-1499,64				
ja	39	-228,07				
Zeit			12	25,52	26,25	0,013
niedrig (bis 16min)	147	-899,20				
hoch (>16min)	142	-819,59				

Tabelle 35: Rasch-Analysen für das Modell 3a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			21	14,68	38,96	0,839
niedrig	111	-575,52				
hoch	84	-506,65				
Median			21	18,69	38,96	0,605
niedrig	132	-716,39				
hoch	63	-363,77				
Geschlecht			21	22,82	38,96	0,354
männlich	85	-473,95				
weiblich	110	-604,15				
Alter			21	18,72	38,96	0,603
niedrig (14-16)	105	-584,36				
hoch (>16)	90	-495,79				
DG			21	32,73	38,96	0,049
nein	170	-946,59				
ja	25	-126,55				
Zeit			21	18,85	38,96	0,595
niedrig (bis 15min)	100	-567,30				
hoch (>15min)	95	-512,79				

Das **Modell 3b** (Tabelle 36) ist ein unvollständiges Rasch-Modell für die Testformen a und b mit 195 Testpersonen, 22 Items und 4 externen Teilungskriterien. Es werden alle Items als vorgegeben betrachtet, d.h. dass fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items als „nicht gelöst“ gewertet werden. Es gibt 2 Untergruppen. Die Gesamtl likelihood beträgt -1147,40.

Tabelle 36: Rasch-Analysen für das Modell 3b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			21	18,04	38,96	0,647
Niedrig	92	-504,97				
Hoch	103	-633,41				
Median			21	17,80	38,96	0,662
Niedrig	118	-673,28				
Hoch	77	-465,22				
Geschlecht			21	23,72	38,96	0,307
Männlich	85	-491,35				
Weiblich	110	-644,19				
Alter			21	21,89	38,96	0,406
niedrig (14-16J.)	105	-612,06				
hoch (>16J.)	90	-524,40				
DG			21	30,21	38,96	0,088
Nein	170	-1003,36				
Ja	25	-128,93				
Zeit			21	20,38	38,96	0,497
niedrig (bis 15min)	100	-595,06				
hoch (>15min)	95	-542,15				

Das **Modell 4a** (Tabelle 37) ist ein vollständiges Rasch-Modell für die Testformen a und b mit 195 Testpersonen, 13 Items und 5 externen Teilungskriterien. Items, die fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen wurden, werden als nicht vorgegebene Items betrachtet, was zu 45 Untergruppen führt. Die Gesamtl likelihood beträgt -1094,62.

Tabelle 37: Rasch-Analysen für das Modell 4a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	6,79	26,25	0,871
niedrig	115	-611,37				
hoch	80	-479,85				
Median			12	5,97	26,25	0,917
niedrig	119	-635,52				
hoch	76	-456,11				
Geschlecht			12	12,26	26,25	0,425
männlich	85	-482,15				
weiblich	110	-606,34				
Alter			12	13,63	26,25	0,325
niedrig (14-16J.)	105	-587,13				
hoch (>16J.)	90	-500,67				
DG			12	11,09	26,25	0,521
nein	170	-951,12				
ja	25	-137,96				
Zeit			12	11,79	26,25	0,462
niedrig (bis 15min)	100	-570,75				
hoch (>15min)	95	-517,98				
Testform			12	11,40	26,25	0,495
a	97	-531,64				
b	98	-557,27				

Das **Modell 4b** (Tabelle 38) ist ein vollständiges Rasch-Modell mit den Testformen a und b, 195 Testpersonen, 14 Items und 5 externen Teilungskriterien. Es werden alle Items als vorgegeben betrachtet, d.h. dass fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items als „nicht gelöst“ gewertet werden. Es gibt 1 Untergruppe. Die Gesamtl likelihood beträgt -1153,66.

Tabelle 38: Rasch-Analysen für das Modell 4b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilungskriterium	n	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert		-507,99	12	13,87	26,25	0,309
niedrig	92	-638,74				
hoch	103					
Median			12	8,91	26,25	0,711
niedrig	118	-678,72				
hoch	77	-470,49				
Geschlecht			12	12,83	26,25	0,381
männlich	85	-500,84				
weiblich	110	-646,40				
Alter			12	16,53	26,25	0,168
niedrig (14-16J.)	105	-614,90				
hoch (>16J.)	90	-530,49				
DG			12	9,85	26,25	0,629
nein	170	-1008,62				
ja	25	-140,12				
Zeit			12	13,71	26,25	0,319
niedrig (bis 15min)	100	-598,77				
hoch (>15min)	95	-548,03				
Testform			12	13,27	26,25	0,350
a	97	-553,03				
b	98	-594,00				

Auch die verschiedenen Auswertungsvarianten mit zwei bzw. drei Testformen gemeinsam ergeben, dass die Nullhypothese – das Rasch-Modell gilt – nicht verworfen werden muss.

Für die Modelle mit 22 Items (1a, 1b, 3a, 3b) bedeuten die Ergebnisse, dass sich die Itemparameter in den Subgruppen niedriger versus hoher Median, männliches versus weibliches Geschlecht, niedriges versus hohes Alter, Besuch versus kein Besuch von Unterricht in Darstellender Geometrie sowie niedrige versus hohe Bearbeitungszeit statistisch nicht voneinander unterscheiden. Dieselben Aussagen gelten für die Modelle mit 13 Items und allen drei Testformen (2a, 2b), was allerdings als „höherwertiger“ einzustufen ist, weil Items mit unterschiedlichen Bearbeitungsrichtungen nicht als zwei unterschiedliche Items verrechnet werden, sondern als „ein Item“ in die Analysen eingehen. Die Itemparameter unterscheiden

sich also nach Vertauschung der Start- und Zielansicht in den verschiedenen Subgruppen nicht voneinander.

Noch wichtiger erscheinen die Ergebnisse der Rasch-Analysen für die Modelle mit 13 Items und den Testformen a und b (4a, 4b): Auch hier kann die Nullhypothese der Modellgeltung beibehalten werden, da alle Modelltests unauffällig verlaufen. Besonders weitreichend ist die Aussage, dass der Modelltest, der die Formen a und b direkt miteinander vergleicht, nicht signifikant ist. Das bedeutet, dass sich die Itemparameter in den beiden Testformen nicht voneinander unterscheiden, obwohl die Items nicht gleich sind, da sie in zwei unterschiedliche Richtungen bearbeitet wurden.

Zusammenfassend bedeutet die gemeinsame Rasch-Homogenität der Testformen a und b bzw. a, b und c, dass sich die Testformen in ihrer Dimensionalität gleichen und kein qualitativer Unterschied zwischen den Versionen besteht.

Die Hauptfragestellung, ob sich die Itemparameter bei *computergenerierten Endlosschleifen* nach Vertauschung der Start- und Zielansicht voneinander unterscheiden, kann in Anbetracht der berichteten Ergebnisse verneint werden.

3.6.1.3 Likelihood-Quotienten-Tests zwischen den verschiedenen Rasch-Modellen

Mit Hilfe von Likelihood-Quotienten-Tests (LQT) soll untersucht werden, ob Modelle mit weniger Parametern (13 Items) die Daten genauso gut erklären wie Modelle mit mehr Parametern (22 Items). Für die Überprüfung werden die Likelihoods der Gesamtstichprobe sowie die Freiheitsgrade der zwei zu vergleichenden Modelle benötigt; zu beachten ist, dass nur Modelle mit gleicher Testpersonenanzahl verglichen werden können.

Die Berechnungsformel für die Likelihood-Quotienten-Tests lautet:

$$\chi^2 = 2 \cdot (\ln L_1 - \ln L_2) \text{ bei } df = df_1 - df_2$$

$\ln L_1$... Gesamtlikelihood der Stichprobe 1

$\ln L_2$... Gesamtlikelihood der Stichprobe 2

df_1 ... Freiheitsgrade der Stichprobe 1

df_2 ... Freiheitsgrade der Stichprobe 2

Alle für die Likelihood-Quotienten-Tests notwendigen Informationen sind Tabelle 39 zu entnehmen.

Tabelle 39: Notwendige Informationen für die Berechnung der Likelihood-Quotienten-Tests mit den involvierten Testformen, der Itemanzahl, der Angabe, welche Items verrechnet wurden, der Bezeichnung der Modelle, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) sowie der Likelihood (ln L).

Testform	alle Testformen				Testformen a und b			
Itemanzahl	22 Items		13 Items		22 Items		13 Items	
„“, ¥, ?	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell 1a	Modell 1b	Modell 2a	Modell 2b	Modell 3a	Modell 3b	Modell 4a	Modell 4b
n	289	258	289	289	195	195	195	195
df	21	21	12	12	21	21	12	12
ln L	-1636,31	-1537,90	-1642,02	-1731,56	-1089,51	-1147,40	-1094,62	-1153,66

„“, ¥, ? nicht bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben so behandelt werden, als wären sie nicht vorgegeben worden; **alle Items** bedeutet, dass fehlerhaft und mit dem Fragezeichen-Symbol markierte sowie ausgelassene Aufgaben als nicht gelöst gewertet werden.

Bei Betrachtung der Tabelle 39 ist erkennbar, dass die Modelle 1a und 2a, 3a und 4a sowie 3b und 4b miteinander verglichen werden können. Auf die Modelle 1b und 2b, die in der bislang beschriebenen Form aufgrund der unterschiedlichen Testpersonenanzahl nicht gegenübergestellt werden können, wird weiter unten eingegangen.

Likelihood-Quotienten-Test für die Modelle 1a und 2a

Für den Vergleich der Gesamtl likelihoods der beiden Modelle (Tabelle 39) ergibt das χ^2 11,42 bei 9 (21-12) Freiheitsgraden. Der kritische Wert für 1% beträgt 21,67; der χ^2 -Wert liegt darunter ($p=0,248$). Das bedeutet, dass das Modell 2a mit 13 Parametern die Daten genauso gut erklärt wie das Modell 1a mit 22 Parametern.

Likelihood-Quotienten-Test für die Modelle 3a und 4a

Das χ^2 ergibt 10,22 und liegt mit 9 Freiheitsgraden unter dem kritischen Wert von 21,67 für 1% ($p=0,333$). Auch hier erklärt das Modell (4a) mit weniger Parametern die Daten genauso gut wie das Modell (3a) mit mehr Parametern.

Likelihood-Quotienten-Test für die Modelle 3b und 4b

Das χ^2 von 12,52 liegt mit 9 Freiheitsgraden unter dem kritischen Wert (1%) von 21,67 ($p=0,186$). Das Modell mit weniger Parametern (4b) kann für weitere Berechnungen herangezogen werden.

Likelihood-Quotienten-Test für die Modelle 1b und 2b

Um die Modelle 1b mit 258 Testpersonen und 2b mit 289 Testpersonen miteinander vergleichen zu können, werden weitere Rasch-Analysen analog dem Modell 2b unter dem Namen Modell 2b2 durchgeführt – allerdings nur mit jenen 258 Testpersonen, die in Modell 1b involviert sind.

Das **Modell 2b2**, das in Tabelle 40 zusammengefasst dargestellt wird, ist ebenso wie Modell 2b ein vollständiges Rasch-Modell für alle Testformen mit 13 Items und 4 externen Teilkriterien. Alle Items werden als vorgegeben betrachtet. Es werden die Daten aller Testpersonen der Formen a und b in die Analysen einbezogen sowie die Daten der Testpersonen mit Form c, die alle Aufgaben bearbeitet haben (und keine Items fehlerhaft oder mit dem Fragezeichen-Symbol markiert bzw. ausgelassen haben), das sind insgesamt 258. Es gibt eine Untergruppe, die Gesamtl likelihood beträgt -1544,99.

Tabelle 40: Rasch-Analysen für das Modell 2b2 mit der Anzahl der Testpersonen pro Teilkriterium (n), der Likelihood in den Teilkriterien (ln L), den Freiheitsgraden (df), der Prüfgröße des LQs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).

Teilkriterium	N	ln L	df	χ^2	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert			12	11,87	26,25	0,456
Niedrig	118	-659,21				
Hoch	140	-879,84				
Median			12	5,69	26,25	0,931
Niedrig	158	-923,02				
Hoch	100	-619,12				
Geschlecht			12	19,99	26,25	0,067
Männlich	112	-667,55				
Weiblich	146	-867,44				
Alter			12	12,25	26,25	0,426
niedrig (14-16J.)	134	-798,99				
hoch (>16J.)	124	-739,87				
DG			12	9,59	26,25	0,652
Nein	223	-1342,87				
Ja	35	-197,32				
Zeit			12	23,93	26,25	0,021
niedrig (bis 16min)	134	-819,66				
hoch (>16min)	124	-713,37				

Wie aus Tabelle 40 ersichtlich ist, kann auch für das Modell 2b2 Rasch-Homogenität angenommen werden. Die Modelle 1b und 2b2 können jetzt mittels Likelihood-Quotienten-Test verglichen werden. Das χ^2 ergibt 14,18, der kritische Wert für 1% liegt mit 9 Freiheitsgraden bei 21,67 (p= 0,116). Das bedeutet, dass das Modell 2b2 mit weniger Parametern die Daten genauso gut erklärt wie das Modell 1b.

3.6.1.4 Likelihood-Quotienten-Tests auf Ebene der Teilungskriterien

Zusätzlich können nicht nur die Gesamtlikelhoods zweier Modelle verglichen werden, sondern auch die Likelihoods in den einzelnen Teilungsgruppen (männlich versus weiblich, niedriges versus hohes Alter usw.). Diese Vergleiche werden nur für die Modelle vorgenommen, bei denen fehlerhaft oder mit dem Fragezeichen-Symbol markierte bzw. ausgelassene Items verrechnet wurden; das trifft auf die Modelle 1b und 2b2 sowie 3b und 4b zu. Tabelle 41 können die Ergebnisse des Vergleichs zwischen den Modellen 1b und 2b2 entnommen werden.

Tabelle 41: Likelihood-Quotienten-Tests für die einzelnen Teilungskriterien der Modelle 1b und 2b2 mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen ($\ln L_1$, $\ln L_2$), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie zur Wahrscheinlichkeit (p).

	Modell 1b			Modell 2b2						
Teilungskriterium	n_1	$\ln L_1$	df_1	n_2	$\ln L_2$	df_2	χ^2	$df_1 - df_2$	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert										
Niedrig	139	-791,60	21	118	-659,21	12	aufgrund unterschiedlicher n kein Vergleich möglich			
Hoch	119	-733,11	21	140	-879,84	12				
Median										
Niedrig	171	-1002,78	21	158	-923,02	12	aufgrund unterschiedlicher n kein Vergleich möglich			
Hoch	87	-523,90	21	100	-619,12	12				
Geschlecht										
Männlich	112	-660,43	21	112	-667,55	12	14,24	9	21,67	0,114
Weiblich	146	-863,08	21	146	-867,44	12	8,72	9	21,67	0,464
Alter										
niedrig (14-16J.)	134	-795,01	21	134	-798,99	12	7,96	9	21,67	0,538
hoch (>16J.)	124	-734,14	21	124	-739,87	12	11,46	9	21,67	0,245
DG										
Nein	223	-1337,81	21	223	-1342,87	12	10,12	9	21,67	0,341
Ja	35	-187,56	21	35	-197,32	12	19,52	9	21,67	0,021
Zeit										
niedrig (≤ 16 min)	134	-815,15	21	134	-819,66	12	9,02	9	21,67	0,435
hoch (>16min)	124	-708,61	21	124	-713,37	12	9,52	9	21,67	0,391

Um die Modelle 1b und 2b2 auch anhand des Rohscores mittels Likelihood-Quotienten-Test gegenüberstellen zu können, muss sichergestellt werden, dass in beiden Modellen dieselben Testpersonen in den Gruppen „niedrig“ und „hoch“ enthalten sind. Das gelingt, indem für Modell 1b mit 42 Untergruppen ein weiteres externes Teilungskriterium gebildet wird, das die Stichprobe anhand des Medians des Gesamtscores teilt. Für das Modell 2b2 mit nur einer Untergruppe muss dieses externe Kriterium nicht gebildet werden; es kann die Teilung nach dem Median (Tabelle 41) herangezogen werden. Tabelle 42 zeigt die Ergebnisse der Like-

likelihood-Quotienten-Tests in den Gruppen hoher und niedriger Median des Gesamtscores für die Modelle 1b und 2b2.

Tabelle 42: Likelihood-Quotienten-Tests in den nach dem Median geteilten Gruppen hoher und niedriger Score für die Modelle 1b und 2b2 mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen ($\ln L_1$, $\ln L_2$), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQTs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{krit(1\%)}$) sowie zur Wahrscheinlichkeit (p).

Teilungskriterium	Modell 1b			Modell 2b2			χ^2	$df_1 - df_2$	$\chi^2_{krit(1\%)}$	p
	n_1	$\ln L_1$	df_1	n_2	$\ln L_2$	df_2				
Score ≤ 6	158	-917,48	21	158	-923,02	12	11,08	9	21,67	0,270
Score > 6	100	-614,92	21	100	-619,12	12	8,4	9	21,67	0,494

Da alle Likelihood-Quotienten-Tests auf Ebene der einzelnen Teilungskriterien nicht signifikant sind, kann geschlossen werden, dass das Modell 2b2 mit weniger Parametern die Daten genauso gut erklärt wie das Modell 1b mit mehr Parametern.

Neben dem Vergleich der Modelle 1b und 2b2 auf Ebene der Teilungskriterien werden auch die Likelihoods in den Teilungsgruppen der Modelle 3b und 4b gegenübergestellt. Die Ergebnisse sind in Tabelle 43 dargestellt.

Tabelle 43: Likelihood-Quotienten-Tests für die einzelnen Teilungskriterien der Modelle 3b und 4b mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen ($\ln L_1$, $\ln L_2$), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie zur Wahrscheinlichkeit (p).

	Modell 3b			Modell 4b						
Teilungskriterium	n_1	$\ln L_1$	df_1	n_2	$\ln L_2$	df_2	χ^2	$df_1 - df_2$	$\chi^2_{\text{krit}(1\%)}$	p
Mittelwert										
Niedrig	92	-504,97	21	92	-507,99	12	6,04	9	21,67	0,736
Hoch	103	-633,41	21	103	-638,74	12	10,66	9	21,67	0,300
Median										
Niedrig	118	-673,28	21	118	-678,72	12	10,88	9	21,67	0,284
Hoch	77	-465,22	21	77	-470,49	12	10,54	9	21,67	0,309
Geschlecht										
Männlich	85	-491,35	21	85	-500,84	12	18,98	9	21,67	0,025
Weiblich	110	-644,19	21	110	-646,40	12	4,42	9	21,67	0,882
Alter										
niedrig (14-16J.)	105	-612,06	21	105	-614,90	12	5,68	9	21,67	0,771
hoch (>16J.)	90	-524,40	21	90	-530,49	12	12,18	9	21,67	0,203
DG										
Nein	170	-1003,36	21	170	-1008,62	12	10,52	9	21,67	0,310
Ja	25	-128,93	21	25	-140,12	12	22,38	9	21,67	0,008*
Zeit										
niedrig (≤ 15 min)	100	-595,06	21	100	-598,77	12	7,42	9	21,67	0,593
hoch (>15min)	95	-542,15	21	95	-548,03	12	11,76	9	21,67	0,227

* Der Likelihood-Quotienten-Test für die Teilstichproben „DG ja“ aus den Modellen 3b und 4b ist signifikant. Da aber 12 Tests zu einer zentralen Hypothese berechnet werden, kann α adjustiert werden (siehe oben, S. 82). Das adjustierte α für 1% beträgt 0,0008; das bedeutet, dass das Ergebnis von 0,008 nicht signifikant ist. Kritisch anzumerken ist aber auch die kleine Stichprobengröße von 25 Testpersonen mit Unterricht in Darstellender Geometrie, was zu fehlerhaften Schätzungen geführt haben könnte.

Auch die Likelihood-Quotienten-Tests zwischen den einzelnen Teilungsgruppen der Modelle 3b und 4b ergeben, dass das Modell 3b mit mehr Parametern die Daten nicht besser erklärt als das Modell 4b mit weniger Parametern.

3.6.1.5 Weitere Likelihood-Quotienten-Tests

Zusätzlich zu den bislang berechneten Likelihood-Quotienten-Tests können die Gesamt-Likelihoods aus den Rasch-Analysen der Testformen einzeln mit den Gesamt-Likelihoods aus den Rasch-Analysen der Testformen gemeinsam verglichen werden; möglich sind diese Berechnungen für alle Modelle mit 13 Items. In den Tabellen 44 und 45 sind alle für die Likelihood-Quotienten-Tests notwendigen Informationen zusammengefasst.

Tabelle 44: Für die Berechnung der Likelihood-Quotienten-Tests notwendige Informationen aus den Rasch-Analysen der Testformen einzeln mit Angaben über Testform und Itemanzahl sowie darüber, welche Items verrechnet wurden, der Bezeichnung des Modells, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) und der Likelihood (ln L).

Testform	A		B		c	
Itemanzahl	13 Items		13 Items		13 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell A	Modell A	Modell B	Modell B	Modell C	Modell c
n	97	97	98	98	94	94
df	12	12	12	12	12	12
ln L	-531,64	-553,03	-557,27	-594,00	-532,89	-563,73

Tabelle 45: Für die Berechnung der Likelihood-Quotienten-Tests notwendige Informationen aus den Rasch-Analysen der Testformen gemeinsam mit Angaben über die involvierten Testformen, der Itemanzahl sowie darüber, welche Items verrechnet wurden, der Bezeichnung des Modells, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) und der Likelihood (ln L).

Testform	a, b und c		a und b	
Itemanzahl	13 Items		13 Items	
Welche Items	„“, ¥, ? nicht	alle Items	„“, ¥, ? nicht	alle Items
Bezeichnung des Modells	Modell 2a	Modell 2b	Modell 4a	Modell 4b
N	289	289	195	195
Df	12	12	12	12
ln L	-1642,02	-1731,56	-1094,62	-1153,66

Likelihood-Quotienten-Test für die Modelle A, B und C versus Modell 2a

Die auf S.99 angegebene Berechnungsformel kann für den Vergleich von drei Teilstichproben mit einer Gesamtstichprobe adaptiert werden. Die Berechnungsformel lautet dann:

$$\chi^2 = 2 \cdot ((\ln L_1 + \ln L_2 + \ln L_3) - \ln L_{\text{ges}}) \text{ bei } df = (df_1 + df_2 + df_3) - df_{\text{ges}}$$

ln L₁ ... Gesamtl likelihood der Teilstichprobe 1

ln L₂ ... Gesamtl likelihood der Teilstichprobe 2

ln L₃ ... Gesamtl likelihood der Teilstichprobe 3

ln L_{ges} ... Gesamtl likelihood der Gesamtstichprobe

df₁ ... Freiheitsgrade der Teilstichprobe 1

df₂ ... Freiheitsgrade der Teilstichprobe 2

df₃ ... Freiheitsgrade der Teilstichprobe 3

df_{ges} ... Freiheitsgrade der Gesamtstichprobe

Werden die entsprechenden Zahlen aus den Tabellen 44 und 45 in die Formel eingesetzt, ergibt sich:

$$\chi^2 = 2 \cdot (((-531,64) + (-557,27) + (-532,89)) - (-1642,02))$$

$$\chi^2 = 2 \cdot 20,22$$

$$\chi^2 = 40,44 \text{ bei } df = (12 + 12 + 12) - 12 = 24$$

Das χ^2 von 40,44 liegt mit 24 Freiheitsgraden unter dem kritischen Wert von 42,98 für 1% ($p=0,019$). Das bedeutet, dass die Modelle A, B und C einzeln nicht mehr Informationen beinhalten als das Modell 2a mit den drei Testformen gemeinsam.

Likelihood-Quotienten-Test für die Modelle a, b und c versus Modell 2b

Für den Vergleich der Likelihoods zwischen den Modellen a, b und c versus Modell 2b ergibt sich:

$$\chi^2 = 2 \cdot (((-553,03) + (-594,00) + (-563,73)) - (-1731,56))$$

$$\chi^2 = 2 \cdot 20,80$$

$$\chi^2 = 41,6 \text{ bei } df = (12 + 12 + 12) - 12 = 24$$

Der kritische Wert bei $df=24$ und 1% (42,98) wird mit 41,6 nicht überschritten ($p=0,014$); Die Modelle a, b und c beinhalten einzeln nicht mehr Informationen als Modell 2b mit allen drei Testformen gemeinsam.

Likelihood-Quotienten-Test für die Modelle A und B versus Modell 4a

Weiters können die Ergebnisse der Rasch-Analysen der Testformen a und b einzeln mit der Gesamtlikelihood der Rasch-Analysen der Formen a und b gemeinsam verglichen werden.

Die Formel für den Vergleich von zwei Stichproben lautet:

$$\chi^2 = 2 \cdot ((\ln L_1 + \ln L_2) - \ln L_{\text{ges}}) \text{ bei } df = (df_1 + df_2) - df_{\text{ges}}$$

$\ln L_1$... Gesamtlikelihood der Teilstichprobe 1

$\ln L_2$... Gesamtlikelihood der Teilstichprobe 2

$\ln L_{\text{ges}}$... Gesamtlikelihood der Gesamtstichprobe

df_1 ... Freiheitsgrade der Teilstichprobe 1

df_2 ... Freiheitsgrade der Teilstichprobe 2

df_{ges} ... Freiheitsgrade der Gesamtstichprobe 2

Nach Einsetzen der entsprechenden Werte aus den Tabellen 44 und 45 ergibt sich:

$$\chi^2 = 2 \cdot (((-531,64) + (-557,27)) - (-1094,62))$$

$$\chi^2 = 2 \cdot 5,71$$

$$\chi^2 = 11,42 \text{ bei } df = (12 + 12) - 12 = 12$$

Der kritische Wert für 1% und 12 Freiheitsgraden beträgt 26,25; das χ^2 liegt mit 11,42 deutlich darunter ($p=0,493$). Die Modelle A und B beinhalten nicht mehr Informationen als das Modell 4a.

Likelihood-Quotienten-Test für die Modelle a und b versus Modell 4b

Durch Einsetzen der Werte aus den Tabellen 44 und 45 erhält man:

$$\chi^2 = 2 \cdot (((-553,03) + (-594,00)) - (-1153,66))$$

$$\chi^2 = 2 \cdot 6,63$$

$$\chi^2 = 13,26 \text{ bei } df = (12 + 12) - 12 = 12$$

Das χ^2 entspricht bei 1% und 12 Freiheitsgraden einem p von 0,350 ($\chi^2_{\text{krit}1\%} = 26,25$). Die Modelle a und b beinhalten einzeln nicht mehr Informationen als das Modell 4b.

3.6.1.6 Schlussfolgerungen aus den Ergebnissen der Likelihood-Quotienten-Tests

Aus den im Kapitel „Likelihood-Quotienten-Tests zwischen den verschiedenen Rasch-Modellen“ (Subkapitel unter 3.6.1) berichteten Ergebnissen lässt sich schließen, dass die Modelle mit weniger Parametern die Daten genauso gut erklären wie die Modelle mit mehr Parametern. Die weiteren klassisch-testtheoretischen Auswertungen werden daher mit den Daten aus Modellen mit 13 Items durchgeführt. Aus den Modellen mit 13 Items werden jene herangezogen, bei denen alle Items als vorgegeben betrachtet wurden. Im Folgenden werden also Berechnungen für die Modelle 2b, 2b2 und 4b (Übersicht über alle Modelle in Tabelle 30, bzw. Modell 2b2) vorgenommen.

Das Modell 2b (13 Items, 289 Testpersonen) hat zwar kein „Gegenstück-Modell“, mit dem es mittels Likelihood-Quotienten-Tests verglichen werden kann; dennoch werden die Daten aus diesem Modell für weitere Auswertungen genutzt, da das Modell alle 289 Testpersonen und alle drei Testformen beinhaltet. Im Vergleich dazu sind in Modell 2b2 zwar alle Testformen, aber nur 258 Testpersonen involviert, weil nur Personen mit der Testversion c in die Analysen eingehen konnten, die alle Items bearbeitet haben. Die Ergebnisse dieses Modells könnten allerdings verzerrt sein, da voraussichtlich eher jene Testpersonen wegfielen, die den

Test weniger sorgfältig bearbeitet haben oder schwächer begabt waren. Von Modell 4b sind insofern interessante Ergebnisse zu erwarten, als nur die beiden „Extremtestformen“ a und b involviert sind; der Nachteil ist allerdings, dass die erhobenen Daten der Testform c bei den Auswertungen nicht berücksichtigt werden.

3.7 Klassisch-testtheoretische Auswertungen

Nach der abgeschlossenen Prüfung auf Gültigkeit des Rasch-Modells werden in diesem Kapitel klassisch-testtheoretische Berechnungen für die Modelle 2b, 2b2 und 4b beschrieben. Für die Auswertungen wird nicht der Score, sondern der nach Warm (1989) geschätzte Personenparameter (WLE-Parameter) herangezogen; zwecks besserer Verständlichkeit werden neben den WLE-Parametern die Scores angeführt. Das α -Niveau wird für alle Auswertungen mit 5% festgesetzt.

3.7.1 Schätzung der Personenparameter nach Warm

Die Ergebnisse der Schätzung der Personenparameter nach Warm (1989) sowie die dazugehörigen Scores für die Modelle 2b, 2b2 und 4b können Tabelle 46 entnommen werden. Tabelle 47 zeigt die Häufigkeiten der beobachteten WLE-Parameter bzw. der Scores.

Tabelle 46: Den Scores entsprechende WLE-Parameter für die Modelle 2b, 2b2 und 4b mit Angaben über die involvierten Testformen und die Stichprobengrößen (n).

Modell	Modell 2b	Modell 2b2	Modell 4b
Testformen	Formen a, b, c	Formen a, b, c	Formen a, b
n	289	258	195
Score	WLE-Parameter	WLE-Parameter	WLE-Parameter
0	-3,57	-3,58	-3,58
1	-2,34	-2,35	-2,36
2	-1,70	-1,71	-1,72
3	-1,23	-1,24	-1,24
4	-0,83	-0,84	-0,85
5	-0,48	-0,49	-0,49
6	-0,15	-0,15	-0,16
7	0,17	0,17	0,17
8	0,50	0,50	0,50
9	0,85	0,85	0,86
10	1,23	1,24	1,25
11	1,69	1,70	1,71
12	2,32	2,33	2,34
13	3,52	3,54	3,56

Tabelle 47: Häufigkeiten der Scores und der WLE-Parameter in den Modellen 2b, 2b2 und 4b in Zahlen (n) und Prozent (%).

	Modell 2b			Modell 2b2			Modell 4b		
Score	WLE	N	%	WLE	n	%	WLE	n	%
1	-2,34	4	1,4	-2,35	2	0,8	-2,36	2	1,0
2	-1,70	20	6,9	-1,71	19	7,4	-1,72	15	7,7
3	-1,23	31	10,7	-1,24	24	9,3	-1,24	23	11,8
4	-0,83	43	14,9	-0,84	40	15,5	-0,85	31	15,9
5	-0,48	39	13,5	-0,49	33	12,8	-0,49	21	10,8
6	-0,15	43	14,9	-0,15	40	15,5	-0,16	26	13,3
7	0,17	47	16,3	0,17	44	17,1	0,17	36	18,5
8	0,50	37	12,8	0,50	33	12,8	0,50	25	12,8
9	0,85	15	5,2	0,85	14	5,4	0,86	11	5,6
10	1,23	7	2,4	1,24	6	2,3	1,25	4	2,1
11	1,69	3	1,0	1,70	3	1,2	1,71	1	0,5
Gesamt		289	100,0		258	100,0		195	100,0

Anhand Tabelle 47 ist ersichtlich, dass die Anzahl der gelösten Aufgaben zwischen 1 und 11 liegt. Das bedeutet, dass alle Testpersonen zumindest 1 Item gelöst haben, aber keine Person beobachtet wurde, die 12 bzw. alle 13 verrechneten Items lösen konnte.

Die Mittelwerte der WLE-Parameter liegen in Abhängigkeit des jeweiligen Modells zwischen -0,33 und -0,29. Detaillierte Angaben über Mittelwerte und Standardabweichungen der WLE-Parameter und der Scores sowie über die Mediane können Tabelle 48 entnommen werden.

Tabelle 48: Mittelwerte (Mw WLE, Mw Score), Standardabweichungen (s WLE, s Score) und Mediane (Med WLE, Med Score) für die WLE-Parameter und Scores in den Modellen 2b, 2b2 und 4b samt Stichprobengröße (n).

	Modell 2b	Modell 2b2	Modell 4b
Mw WLE	-0,31	-0,29	-0,33
s WLE	0,79	0,78	0,79
Mw Score	5,62	5,69	5,58
s Score	2,19	2,17	2,19
Med WLE	-0,15	-0,15	-0,16
Med Score	6	6	6
n	289	258	195

Beim Vergleich der Kennwerte der drei Modelle fällt auf, dass der mittlere WLE-Parameter in Modell 4b, das nur die Testformen a und b involviert, am niedrigsten ist. Bei Modell 2b2 mit dem höchsten Mittelwert kommt wahrscheinlich die bereits weiter oben angesprochene Vermutung zum Tragen, dass die Ergebnisse wegen des notwendigen Ausschlusses einiger Personen mit der Testform c verzerrt sind.

3.7.2 Überprüfung von Unterschieden zwischen Testpersonengruppen

Nach Darstellung der Häufigkeiten der WLE-Parameter und Scores soll im Folgenden mittels t-Tests für unabhängige Stichproben und Varianzanalysen geprüft werden, ob sich Testpersonengruppen mit unterschiedlichen Merkmalsausprägungen in ihren Raumvorstellungsleistungen unterscheiden. Konkret geht es um Unterschiede zwischen den Geschlechtern, um Altersunterschiede, Unterschiede in Abhängigkeit der Bearbeitungszeit, Unterschiede zwischen Personen, die Unterricht bzw. keinen Unterricht in Darstellender Geometrie besucht haben sowie um Leistungsunterschiede je nach bearbeiteter Testform.

In einem ersten Schritt wird eine Voraussetzung für t-Tests und Varianzanalysen – die Annahme der Normalverteilung der WLE-Parameter – in allen Teilstichproben der Modelle 2b, 2b2 und 4b überprüft. Bei Teilstichprobengrößen ab 90 Testpersonen wird die Normalverteilungsannahme geprüft, indem Schiefe (Skewness) und Wölbung (Kurtosis) der WLE-Parameter-Verteilung untersucht werden. Die Interpretation erfolgt wie bei z-Werten: Ist $z \geq \alpha$ ($z = 1,96$ bei einem α von 5%), kann keine Normalverteilung angenommen werden. Bei kleineren Teilstichproben wird der Kolmogoroff-Smirnov-Test eingesetzt. Testpersonen, die Angaben zu einer der gefragten Merkmale nicht bzw. fehlerhaft vornahmen, werden bei den folgenden Analysen nicht berücksichtigt. Die Ergebnisse der Normalverteilungsüberprüfung für die drei Modelle können den Tabellen 49, 50 und 51 entnommen werden.

Tabelle 49: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 2b in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤ 16 min, > 16 min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).

	Geschlecht		Alter		DG		Bearbeitungszeit		Testform		
	M	w	14-16	>16	ja	nein	≤ 16 min	> 16 min	a	b	c
n	128	160	151	137	39*	248	141	138	97	98	94
Schiefe	-0,41	0,03	-0,07	-0,19		-0,12	-0,14	-0,22	-0,36	-0,05	-0,08
s Schiefe	0,21	0,19	0,20	0,21		0,16	0,20	0,21	0,25	0,24	0,25
Wölbung	0,19	-0,45	-0,46	0,03		-0,41	-0,55	-0,13	-0,82	-0,40	0,28
s Wölbung	0,43	0,38	0,39	0,41		0,31	0,41	0,41	0,49	0,48	0,49

* Für die Teilstichprobe „Besuch von Unterricht in Darstellender Geometrie“ wird aufgrund der geringen Testpersonenanzahl ein Kolmogoroff-Smirnov-Test gerechnet, der mit einem Kolmogoroff-Smirnov-Z von 0,861 nicht signifikant ist ($p=0,450$).

Tabelle 50: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 2b2 in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤16min, >16min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).

	Geschlecht		Alter		DG		Bearbeitungszeit		Testform		
Teilstichproben	M	w	14-16	>16	ja	nein	≤16min	>16min	A	b	c
N	111	146	133	124	35*	222	129	122	97	98	63*
Schiefe	-0,36	0,03	-0,02	-0,22		-0,12	-0,10	-0,22	-0,36	-0,05	
s Schiefe	0,23	0,20	0,21	0,22		0,16	0,21	0,22	0,25	0,24	
Wölbung	0,10	-0,34	-0,50	0,05		-0,40	-0,62	-0,10	-0,82	-0,42	
s Wölbung	0,46	0,40	0,42	0,43		0,33	0,42	0,44	0,49	0,48	

* Die Kolmogoroff-Smirnov-Tests für die Teilstichproben „DG ja“ (Kolmogoroff-Smirnov-Z=0,723; p=0,673) und „Testform c“ (Kolmogoroff-Smirnov-Z=1,062; p=0,210) sind nicht signifikant.

Tabelle 51: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 4b in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤15min, >15min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).

	Geschlecht		Alter		DG		Bearbeitungszeit		Testform	
Teilstichproben	M	w	14-16	>16	ja	nein	≤15min	>15min	A	b
n	84*	110	104	90	25*	169	96	93	97	98
Schiefe		-0,05	-0,07	-0,32		-0,16	0,01	-0,40	-0,35	-0,04
s Schiefe		0,23	0,24	0,25		0,19	0,25	0,25	0,25	0,24
Wölbung		-0,65	-0,89	-0,03		-0,66	-0,74	-0,17	-0,83	-0,41
s Wölbung		0,46	0,47	0,50		0,37	0,49	0,50	0,49	0,48

* Der Kolmogoroff-Smirnov-Tests für die Teilstichprobe „männliches Geschlecht“ (Kolmogoroff-Smirnov-Z = 1,414; p = 0,037) ist signifikant; für die Teilstichprobe „Besuch von Unterricht in Darstellender Geometrie“ (Kolmogoroff-Smirnov-Z = 0,516; p = 0,953) zeigt sich ein nicht signifikantes Ergebnis.

Aufgrund der in den Tabellen 49 bis 51 berichteten Ergebnisse kann für alle Teilstichproben – bis auf „männliches Geschlecht“ in Modell 4b – Normalverteilung angenommen werden.

3.7.2.1 Überprüfung der Frage nach Leistungsunterschieden zwischen den Geschlechtern

Die Frage nach Leistungsunterschieden zwischen Männern und Frauen wird mittels t-Tests für unabhängige Stichproben (Modelle 2b und 2b2) bzw. mittels U-Test (Modell 4b) untersucht. Die wichtigsten statistischen Kennwerte für die Modelle 2b, 2b2 und 4b können Tabelle 52 entnommen werden.

Tabelle 52: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „männliches Geschlecht“ (m) und „weibliches Geschlecht“ (w) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.

	Modell 2b					Modell 2b2					Modell 4b				
	n	Mw WLE	s WLE	Mw Sco	s Sco	N	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco
m	128	-0,16	0,81	6,06	0,20	111	-0,09	0,77	6,25	2,16	84	-0,15	0,82	6,10	2,25
w	160	-0,43	0,76	5,28	0,17	146	-0,44	0,76	5,27	2,09	110	-0,47	0,75	5,19	2,08
n_{ges}	288					257					194				

Für die Modelle 2b und 2b2 sind die t-Tests mit $t=2,928$ und $p=0,004$ bzw. $t=3,598$ und $p<0,01$ signifikant, die Varianzen sind homogen (Modell 2b: $F=0,001$; $p=0,969$; Modell 2b2: $F=0,075$; $p=0,784$). Das Ergebnis des U-Tests für Modell 4b ist mit $U=3531,500$ und $p=0,005$ ebenso signifikant; der mittlere Rang für die Männer beträgt 110,46, für die Frauen 87,60. Männer lösen in allen drei Modellen durchschnittlich nicht ganz ein Item mehr als Frauen. Damit entsprechen die beobachteten Resultate den erwarteten und im Theorieteil (Kapitel 2.4 Geschlechtsunterschiede bei Raumvorstellung) berichteten Ergebnissen.

3.7.2.2 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit des Alters

Auch die Frage nach Leistungsunterschieden in Abhängigkeit des Alters wird mit t-Tests für unabhängige Stichproben untersucht. Die Gesamtstichprobe wird anhand des Medians des Alters in die Gruppen „14 bis 16 Jahre“ und „älter als 16 Jahre“ geteilt. Die statistischen Kennwerte sind in Tabelle 53 angeführt.

Tabelle 53: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „14 bis 16 Jahre“ (≤ 16) und „älter als 16 Jahre“ (> 16) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.

	Modell 2b					Modell 2b2					Modell 4b				
	n	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco
≤ 16	151	-0,46	0,06	5,22	0,18	133	-0,41	0,78	5,34	2,16	104	-0,49	0,78	5,14	2,13
> 16	137	-0,15	0,06	6,07	0,18	124	-0,15	0,77	6,07	2,14	90	-0,15	0,78	6,09	2,17
n_{ges}	288					257					194				

Das Ergebnis des t-Tests für Modell 2b ist mit $t=-3,381$ und $p=0,001$ bei homogenen Varianzen ($F=1,099$; $p=0,295$) signifikant. Für Modell 2b2 ergibt der t-Test mit $t=-2,727$ und $p=0,007$ ebenso ein signifikantes Ergebnis wie für Modell 4b mit $t=-3,031$ und $p=0,003$; die Varianzen sind homogen (Modell 2b2: $F=0,420$; $p=0,517$; Modell 4b: $F=0,126$; $p=0,723$). Ältere Testpersonen lösen im Schnitt nicht ganz ein Item mehr als jüngere. Die Erwartung, dass ältere Teilnehmer bessere Raumvorstellungsleistungen erbringen als jüngere, wurde bestätigt und entspricht den im Theorieteil berichteten Ergebnissen von Geiser, Lehmann und Eid (2008; siehe Kapitel 2.4.4 Veränderungen bei Geschlechtsunterschieden).

3.7.2.3 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit des Besuchs von Unterricht in Darstellender Geometrie

Weitere t-Tests werden für die Beantwortung der Frage nach Leistungsunterschieden in Abhängigkeit davon, ob Unterricht in Darstellender Geometrie besucht wurde oder nicht, gerechnet. Tabelle 54 können die statistischen Kennwerte entnommen werden.

Tabelle 54: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „Besuch von Unterricht in Darstellender Geometrie“ (DG) und „kein Unterricht in Darstellender Geometrie“ (k.DG) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.

	Modell 2b					Modell 2b2					Modell 4b				
	n	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco
DG	39	-0,39	0,73	5,36	2,01	35	-0,39	0,76	5,40	2,08	25	-0,41	0,77	5,36	2,10
k.DG	248	-0,29	0,79	5,68	2,21	222	-0,27	0,79	5,73	2,19	169	-0,32	0,80	5,60	2,21
n_{ges}	287					257					194				

Das Ergebnis des t-Tests für Modell 2b ist bei homogenen Varianzen ($F=1,895$; $p=0,170$) mit $t=0,779$ und $p=0,437$ nicht signifikant; dasselbe gilt für die Modelle 2b2 mit $t=0,794$ und $p=0,428$ sowie 4b mit $t=0,491$ und $p=0,624$ bei homogenen Varianzen (Modell 2b2: $F=0,759$; $p=0,385$; Modell 4b: $F=0,747$; $p=0,388$).

Überraschend ist die empirische Beobachtung, dass Testpersonen ohne Unterricht in Darstellender Geometrie besser abschneiden als Personen, die DG-Unterricht besucht haben (siehe Tabelle 54). Das widerspricht einerseits der intuitiven Erwartung, andererseits auch publizierten Ergebnissen wie beispielsweise jenen von Gittler (1994), der einen positiven Einfluss von DG-Unterricht auf Raumvorstellungsleistungen beobachten konnte. Kritisch anzumerken ist allerdings in der vorliegenden Untersuchung die geringe Anzahl an Testpersonen, die DG-Unterricht besuchte. χ^2 -Tests für alle drei Modelle ergeben, dass die Teilstichproben „DG-Unterricht“ und „kein DG-Unterricht“ mit $p<0,01$ nicht gleich verteilt sind (siehe Tabelle 55).

Tabelle 55: Verteilung der Testpersonen aus den Teilstichproben DG-Unterricht (DG) und kein DG-Unterricht (kein DG) mit den beobachteten sowie erwarteten Stichprobengrößen, den χ^2 -Werten (χ^2), Freiheitsgraden (df) und der Wahrscheinlichkeit (p) für die Modelle 2b, 2b2 und 4b.

	Modell 2b		Modell 2b2		Modell 4b	
	Beobachtet	erwartet	beobachtet	erwartet	beobachtet	erwartet
DG	39	143,5	35	128,5	25	97
kein DG	248	143,5	222	128,5	169	97
χ^2	152,20		136,07		106,89	
df	1		1		1	
P	<0,01		<0,01		<0,01	

3.7.2.4 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der Bearbeitungsdauer

Die Frage, ob sich Testpersonen in ihren Leistungen unterscheiden, die weniger bzw. mehr Zeit zur Testbearbeitung in Anspruch nahmen, lässt sich ebenfalls mit t-Tests beantworten. Die Gesamtstichprobe wird anhand des Medians der Bearbeitungszeit in zwei Gruppen geteilt. Alle wichtigen statistischen Kennwerte sind in Tabelle 56 zusammengefasst.

Tabelle 56: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „niedrige Bearbeitungszeit“ (≤ 16 min bzw. ≤ 15 min) und „hohe Bearbeitungszeit“ (> 16 min bzw. > 15 min) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.

Modell 2b						Modell 2b2					
	n	Mw WLE	s WLE	Mw Sco	s Sco		n	Mw WLE	s WLE	Mw Sco	s Sco
≤ 16 min*	141	-0,46	0,75	5,21	2,07	≤ 16 min*	129	-0,43	0,74	5,27	2,06
> 16 min*	138	-0,17	0,82	6,02	2,29	> 16 min*	122	-0,13	0,81	6,12	2,26
n_{ges}	279					n_{ges}	251				

Modell 4b					
	n	Mw WLE	s WLE	Mw Sco	s Sco
≤ 15 min*	96	-0,53	0,76	5,03	2,09
> 15 min*	93	-0,14	0,80	6,13	2,21
n_{ges}	189				

* In den Modellen 2b und 2b2 liegt das Teilungskriterium bei ≤ 16 Minuten im Gegensatz zu Modell 4b, bei dem das Teilungskriterium ≤ 15 min beträgt. Die unterschiedlichen Teilungskriterien kommen dadurch zustande, dass in Modell 4b die Testform c, bei der die durchschnittliche Bearbeitungsdauer durch die zusätzliche Instruktion nach den fünf Link-Items größer ist, nicht involviert ist. Die durchschnittliche Bearbeitungszeit über alle Testpersonen (mit gültigen Angaben, $n=279$), alle Formen und alle 15 Items beträgt 17,8 Minuten; die Zeiten für die einzelnen Testformen liegen bei 16,4 Minuten für Form a ($n=93$), 17,1 Minuten für Form b ($n=96$) und 20,0 Minuten für Form c ($n=90$).

Die t-Tests sind mit $t=-3,043$ und $p=0,003$ (Modell 2b) bzw. $t=-3,059$ und $p=0,002$ (Modell 2b2) sowie $t=-3,433$ und $p=0,001$ (Modell 4b) signifikant, die Varianzen homogen (Modell 2b: $F=0,807$; $p=0,370$; Modell 2b2: $F=0,329$; $p=0,567$; Modell 4b: $F=0,002$; $p=0,961$). Das Ergebnis, dass Testpersonen, die sich für die Bearbeitung der Endlosschleifen mehr Zeit nehmen, ca. ein Item mehr lösen als Personen, die schneller arbeiten, erscheint einleuchtend. Auch Motivation könnte bei der Bearbeitungszeit eine Rolle spielen – in dem Sinn, dass weniger motivierte Personen die Bearbeitung des Tests schneller hinter sich bringen wollen als motivierte Personen.

Bezugnehmend auf den Theorieteil wird noch eine Hypothese zur Bearbeitungszeit überprüft: Goldstein, Haldane und Mitchell (1990) sowie Linn und Petersen (1985) berichten in ihren Untersuchungen, dass Frauen Raumvorstellungsaufgaben vorsichtiger und daher langsamer bearbeiten. Sollte diese Hypothese für die vorliegenden Daten zutreffen, müssten sich deutlich mehr Frauen in der Gruppe mit der höheren Bearbeitungszeit im Vergleich zur Gruppe mit der niedrigeren Bearbeitungszeit befinden. Mittels χ^2 -Tests kann die Verteilung der Frauen auf die beiden Gruppen überprüft werden. Die Ergebnisse sind für alle drei Mo-

delle nicht signifikant und daher keine Bestätigung der im Theorieteil berichteten Resultate. Es zeigt sich sogar das gegensätzliche Bild, dass sich mehr Frauen in der Gruppe mit niedrigeren Bearbeitungszeiten finden. Die genauen Ergebnisse können Tabelle 57 entnommen werden.

Tabelle 57: Verteilung der Frauen auf die Gruppen „niedrige“ und „hohe Bearbeitungszeit“ mit den beobachteten sowie erwarteten Stichprobengrößen, den χ^2 -Werten (χ^2), Freiheitsgraden (df) und der Wahrscheinlichkeit (p) für die Modelle 2b, 2b2 und 4b.

	Modell 2b		Modell 2b2		Modell 4b	
	beobachtet	erwartet	beobachtet	erwartet	beobachtet	erwartet
niedrige Zeit	79	78	74	71,5	56	53,5
hohe Zeit	77	78	69	71,5	51	53,5
χ^2	0,026		0,175		0,234	
df	1		1		1	
p	0,873		0,676		0,629	

3.7.2.5 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der Testform

Die Frage, ob sich die Raumvorstellungsleistungen in Abhängigkeit der Testform unterscheiden, lässt sich für die Modelle 2b und 2b2 (mit jeweils drei Testformen) mittels Varianzanalysen, für das Modell 4b (mit zwei Testformen) mittels t-Test beantworten.

Analysen für die Modelle 2b und 2b2

Die Mittelwerte und Standardabweichungen der WLE-Parameter und Scores für die Testformen a, b und c können Tabelle 58 entnommen werden, die wichtigsten statistischen Kennwerte der Varianzanalysen sind in Tabelle 59 dargestellt.

Tabelle 58: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) in den Teilstichproben „Form a“, „Form b“, „Form c“ und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b und 2b2.

	Modell 2b					Modell 2b2				
	n	Mw WLE	s WLE	Mw Sco	s Sco	n	Mw WLE	s WLE	Mw Sco	s Sco
Form a	97	-0,39	0,76	5,40	2,12	97	-0,39	0,77	5,40	2,12
Form b	98	-0,26	0,80	5,76	2,26	98	-0,26	0,81	5,76	2,26
Form c	94	-0,27	0,80	5,71	2,21	63	-0,15	0,75	6,03	2,09
n_{ges}	289					258				

Tabelle 59: Statistische Kennwerte der Varianzanalysen zwischen den Testformen mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und Gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q) für die Modelle 2b und 2b2.

	Modell 2b			Modell 2b2		
	QS	df	mittl. Q	QS	df	mittl. Q
Zwischen	1,01	2	0,51	2,30	2	1,15
Innerhalb	178,09	286	0,62	154,28	255	0,61
Gesamt	179,11	288		156,58	257	

Das Ergebnis der Varianzanalyse für Modell 2b ist mit einem F-Wert von 0,812 und einem p von 0,445 nicht signifikant, die Varianzen sind homogen (Levene Statistik=0,275; p=0,760). Für Modell 2b2 ergibt sich mit F=1,899 und p=0,152 ebenso ein nicht signifikantes Ergebnis bei homogenen Varianzen (Levene Statistik=1,625; p=0,199). Das bedeutet, dass sich die Leistungen in den drei Testformen a, b und c nicht voneinander unterscheiden.

Bei Betrachtung der Mittelwerte fällt auf, dass bei Modell 2b2 der durchschnittliche Score bei der Testform c mit ca. 6 gelösten Items deutlich über dem Wert von 5,4 gelösten Items bei der Testform a und ein wenig über dem Wert von 5,8 gelösten Items bei der Form b liegt (siehe Tabelle 58). Beim Vergleich der Scores in der Testform c zwischen den Modellen 2b und 2b2 ist erkennbar, dass der Wert bei Modell 2b2 um ca. 1/3 Item größer ist als bei Modell 2b. Das bedeutet, dass die weiter oben angesprochene Vermutung von verzerrten Daten in Modell 2b2 bestätigt werden kann: Durch das für die Rasch-Analysen notwendige Herauslösen von 31 Testpersonen mit Form c, die nicht alle Aufgaben (korrekt) bearbeitet bzw. Items mit dem Fragezeichen-Symbol markiert haben, fallen scheinbar systematisch Personen weg, die weniger begabt oder vielleicht weniger motiviert an die Aufgabenbearbeitung herangegangen sind.

Für Modell 2b zeigen die vorliegenden Daten einen Trend in die Richtung, dass die Leistungen in den Testformen b und c mit 5,76 bzw. 5,71 gelösten Items besser sind als bei der Testform a mit 5,4 gelösten Items (siehe Tabelle 58).

Analysen für das Modell 4b

Für das Modell 4b mit zwei unabhängigen Variablen (Testformen a und b) wird ein t-Test gerechnet; die Mittelwerte und Standardabweichungen der WLE-Parameter und Scores für die Formen a und b können Tabelle 60 entnommen werden.

Tabelle 60: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) in den Teilstichproben „Form a“ und „Form b“ und Gesamtpersonenanzahl (n_{ges}) für das Modell 4b.

	Modell 4b				
	N	Mw WLE	s WLE	Mw Sco	s Sco
Form a	97	-0,40	0,77	5,40	2,12
Form b	98	-0,27	0,81	5,76	2,26
n_{ges}	195				

Der t-Test ergibt mit $t=-1174$ und $p=0,242$ bei homogenen Varianzen ($F=0,071$; $p=0,791$) ein nicht signifikantes Ergebnis; das heißt, dass keine Leistungsunterschiede zwischen Personen, die Testform a und Personen, die Testform b bearbeitet haben, beobachtet werden konnten. Anhand der Mittelwerte der WLE-Parameter bzw. der Scores ist erkennbar, dass Personen mit der Testform b besser abschnitten als Personen mit Form a (ca. 5,8 gelöste Items im Vergleich zu 5,4 gelösten Aufgaben; siehe Tabelle 60).

Zusammenfassend kann die Frage nach Leistungsunterschieden in Abhängigkeit der Testform verneint werden.

3.7.2.6 Überprüfung der Frage nach der Auswirkung von Wechselwirkungen zwischen Geschlecht, Alter und Zeit auf die Leistung

Dieses Kapitel widmet sich der Frage, ob Wechselwirkungen zwischen den Variablen Geschlecht, Alter und Zeit in Zusammenhang mit der Leistung bestehen. Die bisher berichteten Ergebnisse haben gezeigt, dass die Variablen männliches Geschlecht, höheres Alter und höhere Bearbeitungszeit mit signifikant besseren Leistungen in Verbindung stehen. Um zu klären, ob Wechselwirkungen zwischen den oben genannten Faktoren auftreten, werden für die Modelle 2b und 4b mehrfache Varianzanalysen gerechnet; auf die Berechnung für Modell 2b2 wird verzichtet. Mittelwerte und Standardabweichungen sind in Tabelle 61, die Ergebnisse der Varianzanalysen in Tabelle 62 dargestellt.

Tabelle 61: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Gruppen „männliches Geschlecht“ (m) versus „weibliches Geschlecht“ (w) mit den Altersgruppen „14 bis 16 Jahre“ ($\leq 16J.$) versus „älter als 16 Jahre“ ($> 16J.$) und den Bearbeitungszeitgruppen „bis 16 Minuten“ ($\leq 16min$) bzw. „bis 15 Minuten“ ($\leq 15min$) versus „mehr als 16 Minuten“ ($> 16min$) bzw. „mehr als 15 Minuten“ ($> 15min$) inklusive Testpersonenanzahl pro Untergruppe (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b und 4b.

Ge- schlecht	Alter	Modell 2b						Modell 4b					
		Zeit	n	Mw WLE	s WLE	Mw Sco	s Sco	Zeit	n	Mw WLE	s WLE	Mw Sco	s Sco
m	$\leq 16J.$	$\leq 16min$	34	-0,48	0,92	5,21	2,52	$\leq 15min$	23	-0,69	0,81	4,61	2,21
		$> 16min$	21	-0,01	0,82	6,52	2,16	$> 15min$	16	0,21	0,53	7,13	1,54
	$> 16J.$	$\leq 16min$	28	-0,38	0,64	5,39	1,73	$\leq 15min$	17	-0,32	0,78	5,65	2,06
		$> 16min$	39	0,23	0,69	7,15	2,00	$> 15min$	25	0,26	0,72	7,24	2,05
w	$\leq 16J.$	$\leq 16min$	50	-0,60	0,71	4,80	1,97	$\leq 15min$	33	-0,68	0,74	4,61	2,05
		$> 16min$	38	-0,57	0,75	4,89	2,06	$> 15min$	27	-0,57	0,74	4,93	1,98
	$> 16J.$	$\leq 16min$	29	-0,26	0,67	5,72	1,93	$\leq 15min$	23	-0,31	0,69	5,61	1,95
		$> 16min$	39	-0,27	0,85	5,74	2,36	$> 15min$	24	-0,29	0,83	5,71	2,29
n_{ges}			278						188				

Tabelle 62: Ergebnisse der Varianzanalysen mit den Quadratsummen (QS) für die Wechselwirkungen „Geschlecht*Alter“, „Geschlecht*Zeit“, „Alter*Zeit“ und „Geschlecht*Alter*Zeit“ samt Freiheitsgraden (df), mittleren Quadraten (mittl. Q), F-Werten und den Wahrscheinlichkeiten (p) für die Modelle 2b und 4b.

	Modell 2b					Modell 4b				
	QS	df	mittl.Q	F	Sig.	QS	df	mittl.Q	F	Sig.
Geschlecht*Alter	0,37	1	0,37	0,63	0,427	0,15	1	0,15	0,27	0,606
Geschlecht*Zeit	4,62	1	4,62	8,00	0,005	5,17	1	5,17	9,39	0,003
Alter*Zeit	0,05	1	0,05	0,09	0,767	0,46	1	0,46	0,83	0,363
Geschlecht*Alter*Zeit	0,14	1	0,14	0,23	0,629	0,14	1	0,14	0,26	0,614

Die Varianzanalyse für Modell 2b zeigt bei homogenen Varianzen ($F=1,343$; $df_1=7$; $df_2=270$; $p=0,230$) ein signifikantes Ergebnis für die Wechselwirkung Geschlecht*Zeit ($p=0,005$). Für Modell 4b beträgt die Wahrscheinlichkeit $p=0,003$ für die Wechselwirkung Geschlecht*Zeit, die Varianzen sind homogen ($F=0,916$; $df_1=7$; $df_2=180$; $p=0,495$). Die Frage nach signifikanten Wechselwirkungen kann also bejaht werden, es gibt Interaktionen zwischen den Variablen Geschlecht und Zeit in Zusammenhang mit der Leistung, nicht aber für die Variablen Geschlecht und Alter, Alter und Zeit sowie Geschlecht, Alter und Zeit.

Bei Betrachtung der Mittelwerte der Scores in den verschiedenen Gruppierungen, die sich aus Kombination der Variablen Geschlecht, Alter und Bearbeitungszeit ergeben (siehe Tabelle 61), fällt auf, dass in den Modellen 2b und 4b ältere Männer mit einer höheren Bearbei-

tungszeit die besten Leistungen zeigen (ca. 7,2 gelöste Items), gefolgt von jüngeren Männern mit einer höheren Bearbeitungszeit (ca. 6,5 gelöste Items in Modell 2b und 7,1 gelöste Items in Modell 4b). An dritter Stelle ist in beiden Modellen die erste Frauengruppe zu finden, nämlich ältere Frauen mit einer höheren Bearbeitungszeit mit ca. 5,7 gelösten Items. Am schlechtesten schneiden in Modell 2b jüngere Frauen mit einer niedrigeren Bearbeitungszeit ab (ca. 4,8 gelöste Items) bzw. in Modell 4b gleichermaßen jüngere Frauen und jüngere Männer mit einer niedrigeren Bearbeitungszeit mit ca. 4,6 gelösten Aufgaben.

In Bezug auf die signifikante Wechselwirkung zwischen Geschlecht und Zeit zeigt sich – dargestellt für die Daten aus Modell 2b –, dass Männer und Frauen in ihrer jeweiligen Gruppe mit der niedrigeren Bearbeitungszeit annähernd gleich abschneiden (5,29 versus 5,14 gelöste Items); in der jeweiligen Gruppe mit der höheren Bearbeitungszeit ist ein deutlicher Unterschied beobachtbar: Männer lösen durchschnittlich 6,9 Items, Frauen 5,3. Tabelle 63 können die Mittelwerte und Standardabweichungen der WLE-Parameter sowie Mittelwerte und Standardabweichungen der Scores für die Variablen Geschlecht und Bearbeitungszeit entnommen werden; Abbildung 17 zeigt die graphische Darstellung der Leistungen in den vier Teilstichproben.

Tabelle 63: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) für die Variablen „Geschlecht“ („männlich“, „weiblich“) und „Bearbeitungszeit“ („≤ 16min“, „>16min“) für das Modell 2b.

Geschlecht	Zeit	n	Mw WLE	s WLE	Mw Sco	s Sco
männlich	≤16min	62	-0,43	0,80	5,29	2,18
	>16min	60	0,15	0,74	6,93	2,06
weiblich	≤16min	79	-0,47	0,71	5,14	1,99
	>16min	77	-0,41	0,81	5,32	2,24
n _{ges}		278				

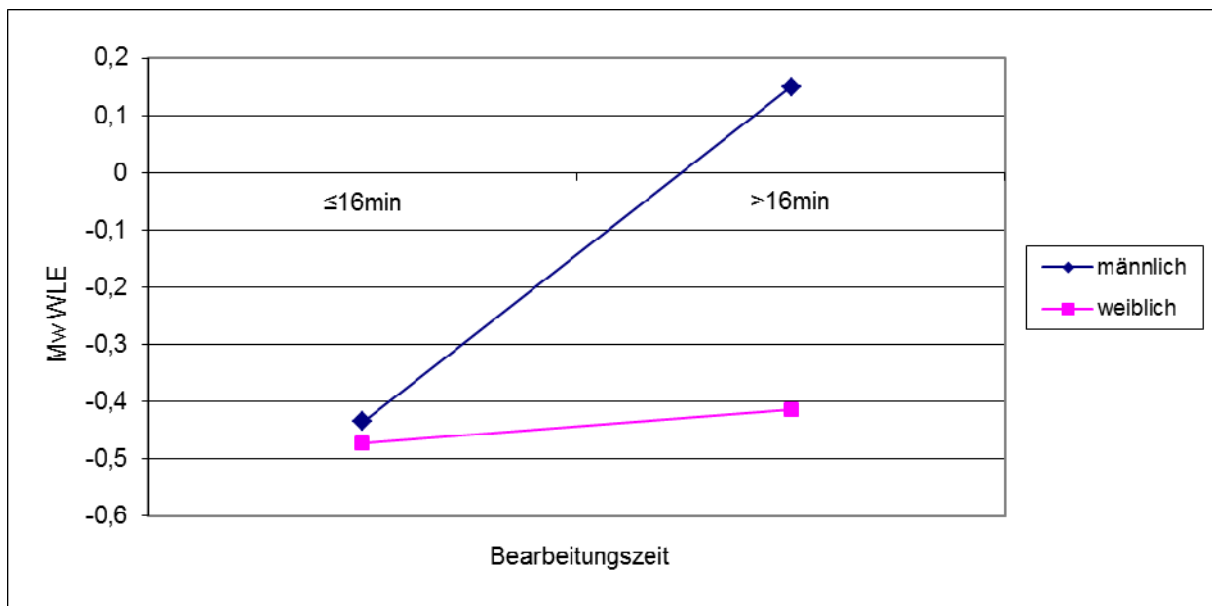


Abbildung 17: Graphische Veranschaulichung der signifikanten Wechselwirkung zwischen Geschlecht und Bearbeitungszeit mit den Leistungen (WLE-Parameter) in den Teilstichproben „weiblich, niedrigere Bearbeitungszeit“, „weiblich, höhere Bearbeitungszeit“, „männlich, niedrigere Bearbeitungszeit“ sowie „männlich, höhere Bearbeitungszeit“ für das Modell 2b.

3.7.2.7 Überprüfung der Frage nach Leistungsunterschieden in Abhängigkeit der gewählten Strategie

Die letzte Auswertung im Kapitel „Überprüfung von Unterschieden zwischen Testpersonen-gruppen“ widmet sich der Bearbeitungsstrategie. Die Testpersonen konnten – wie im Kapitel 3.3.1 beschrieben – ihre Lösungen im Answerheft in einem Rotations- oder Perspektivwechselschema markieren. Durch das von der Testperson gewählte Antwortformat kann für jede Testperson eine Aussage darüber gemacht werden, welches Item sie mit welchem Lösungsweg bearbeitet hat bzw. kann für alle Items angegeben werden, wie häufig sie mittels Rotations- und Perspektivwechselschema bearbeitet wurden. Die Häufigkeiten der beiden Bearbeitungswege können für Modell 2b Tabelle 64 entnommen werden.

Tabelle 64: Häufigkeiten in Zahlen (n) und Prozent (%) für Modell 2b, mit denen die Items 1 bis 15 (ohne Items 5 und 11) mittels Rotations- und Perspektivwechselstrategie (Rot, PPW) bearbeitet wurden.

Item	1		2		3		4		6	
	n	%	n	%	n	%	n	%	n	%
PPW	94	33,3	111	40,2	99	35,0	106	38,7	85	31,4
Rot	188	66,7	165	59,8	184	65,0	168	61,3	186	68,6
n _{ges}	282	100,0	276	100,0	283	100,0	274	100,0	271	100,0

Item	7		8		9		10		12	
	n	%	n	%	n	%	n	%	n	%
PPW	97	34,5	103	38,1	86	31,3	104	36,6	112	40,7
Rot	184	65,5	167	61,9	189	68,7	180	63,4	163	59,3
n _{ges}	281	100,0	270	100,0	275	100,0	284	100,0	275	100,0

Item	13		14		15	
	n	%	n	%	n	%
PPW	108	38,6	99	39,6	92	35,0
Rot	172	61,4	151	60,4	171	65,0
n _{ges}	280	100,0	250	100,0	263	100,0

Die Endlosschleifenaufgaben wurden – wie aus Tabelle 64 ersichtlich – häufiger mit einer Rotations- als mit einer Perspektivwechselstrategie bearbeitet. Das Rotationsschema wurde über alle Items betrachtet von 59,3% (Item 12) bis 68,7% (Item 9) der Testpersonen gewählt, das Perspektivwechselschema markierten zwischen 31,3% (Item 9) und 40,7% (Item 12) der Testpersonen.

Bezüglich der verwendeten Strategie stellt sich die Frage, ob sich die Wahl eines Bearbeitungsschemas auf die Leistung auswirkt. Zur Beantwortung muss für jede Testperson festgelegt werden, welchem von drei Strategietypen sie entspricht – dem Rotierer, dem Perspektivwechsler oder dem Mischtyp. Für die Zuteilung zu einem der drei Typen wurde ein 2/3-Kriterium herangezogen: Zu den Rotierern bzw. Perspektivwechslern zählen jene Testpersonen, die mindestens 2/3 aller bearbeiteten Aufgaben auf dem Rotations- bzw. dem Perspektivwechselschema markiert haben; Mischtypen sind jene Testpersonen, die weniger als 2/3 aller Aufgaben auf einem der beiden Schemata markiert haben. Die auf diese Art vorgenommene Zuordnung ergibt, dass von den 289 Testpersonen 78 Personen (27,0%) als Perspektivwechsler und 157 Personen (54,3%) als Rotierer deklariert werden können. Die restlichen 54 Testpersonen (18,7%) sind Mischtypen. In Tabelle 65 sind die Mittelwerte und Standardabweichungen der WLE-Parameter bzw. der Scores für die drei Strategietypen für Modell 2b dargestellt.

Tabelle 65: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der 3 Strategietypen inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%) für Modell 2b.

Strategietyp	n	%	Mw WLE	s WLE	Mw Sco	s Sco
Perspektivwechsler	78	27,0	-0,21	0,83	5,91	2,30
Rotierer	157	54,3	-0,31	0,72	5,60	2,03
Mischtyp	54	18,7	-0,45	0,91	5,28	2,46
Gesamt	289	100,0				

Die Perspektivwechsler schneiden mit ca. 5,9 gelösten Items am besten ab, gefolgt von Rotierern mit ca. 5,6 gelösten Aufgaben. Die schlechtesten Leistungen erbringen Mischtypen, sie lösen durchschnittlich ca. 5,3 Items.

Mittels Varianzanalyse kann untersucht werden, ob sich die Leistungen der drei Strategietypen signifikant voneinander unterscheiden. Die Prüfung auf Normalverteilung ergibt in allen Gruppen, dass die Leistungen normalverteilt sind (Tabelle 66). Informationen zu den Quadratsummen, mittleren Quadraten und Freiheitsgraden sind in Tabelle 67 dargestellt.

Tabelle 66: Prüfung auf Normalverteilung der WLE-Parameter in den Teilstichproben Perspektivwechsler, Rotierer und Mischtyp samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p) für Modell 2b.

Strategietyp	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
Perspektivwechsler	78	27,0					0,915	0,373
Rotierer	157	54,3	-0,08	0,19	-0,34	0,39		
Mischtyp	54	18,7					0,917	0,370
Gesamt	289	100,0						

Tabelle 67: Statistische Kennwerte der Varianzanalyse für die drei Strategietypen mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q) für Modell 2b.

	QS	df	mittl. Q.
Zwischen	1,99	2	1,00
Innerhalb	177,12	286	0,62
Gesamt	179,11	288	

Die Varianzanalyse zeigt mit $F=1,606$ und $p=0,202$ ein nicht signifikantes Ergebnis bei homogenen Varianzen (Levene Statistic=2,587; $p=0,077$). Die Leistungen unterscheiden sich also nicht in Abhängigkeit der gewählten Strategie.

3.7.3 Zusammenfassung der Ergebnisse der Überprüfung von Unterschieden zwischen Testpersonengruppen

Die Überprüfung der Unterschiedshypothesen ergibt für alle drei Modelle signifikante Leistungsunterschiede in Abhängigkeit des Geschlechts, des Alters und der Bearbeitungszeit. Männliches Geschlecht, höheres Alter und höhere Bearbeitungszeit sind mit besseren Leistungen assoziiert. Bei den Variablen „Besuch von Unterricht in Darstellender Geometrie“, „Testform“ und „Bearbeitungsstrategie“ konnten keine signifikanten Unterschiede beobachtet werden. Bezüglich Wechselwirkungen zwischen Variablen zeigt sich (bei Modell 2b) eine Interaktion zwischen Geschlecht und Bearbeitungszeit in die Richtung, dass Männer mit einer höheren Bearbeitungszeit deutlich besser abschneiden als Personen anderer Teilstichproben.

3.7.4 Fragebogenanalysen

3.7.4.1 Fragebogen zur Einschätzung der Raumvorstellungsleistung, den Testaufgaben und dem Bearbeitungsverhalten

In diesem Kapitel sollen die Ergebnisse des Fragebogens mit Aussagen zur Einschätzung der eigenen Raumvorstellungsleistung, zur Einschätzung der Testaufgaben und des eigenen Bearbeitungsverhaltens (siehe Anhang 1 und 2) für das Modell 2b (289 Testpersonen, 3 Testformen, alle Items als vorgegeben betrachtet) dargestellt werden. Der Fragebogen liegt in zwei annähernd gleichen Versionen mit 13 Aussagen für die Testformen a und b sowie 14 Aussagen für die Testform c vor. Der geringfügige Unterschied findet sich darin, dass spezifische Aussagen zur frei wählbaren (Form c) bzw. nicht frei wählbaren Bearbeitungsrichtung (Formen a und b) formuliert wurden. Eine Aussage war auf einer 3-stufigen Skala, die übrigen auf vierstufigen Skalen, bei denen nur die Extremantworten verbal benannt waren, zu markieren. Der Fragebogen wurde den Testpersonen im Anschluss an die Endlosschleifenaufgaben vorgegeben.

Zwischen 282 und 286 Testpersonen haben die einzelnen Fragebogensaussagen, die für die drei Testformen gleich waren, bearbeitet. Die zwei spezifischen Aussagen für die Form c wurden von 92 bzw. 93 Testpersonen, die spezifische Aussage für die Formen a und b von 189 Testpersonen beantwortet.

Im Folgenden werden die einzelnen Fragebogensaussagen mittels t-Tests, Varianzanalysen und Kreuztabellen analysiert sowie deren Voraussetzungen, die Normalverteilung der WLE-

Parameter in allen Teilstichproben und die Homogenität der Varianzen, untersucht. Zur Normalverteilungsprüfung werden analog oben (Kapitel 3.7.2) bei Stichprobengrößen ab 90 Testpersonen Schiefe (Skewness) und Wölbung (Kurtosis) der WLE-Parameter herangezogen, bei kleineren Teilstichproben wird der Kolmogoroff-Smirnov-Test eingesetzt.

Aussage 1

Ich schätze mein räumliches Vorstellungsvermögen als

niedrig ausgeprägt ein ○ ○ ○ ○ hoch ausgeprägt ein.

Mittels Varianzanalyse wird untersucht, ob die sich die Raumvorstellungsleistungen der Testpersonen je nach Einschätzung ihrer Fähigkeit unterscheiden. Als Ergebnis wird erwartet, dass jene Personen im Test besser abschneiden, die ihr Raumvorstellungsvermögen als höher ausgeprägt einschätzen. Die Prüfung auf Normalverteilung, die Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores sowie die Ergebnisse der Varianzanalyse können den Tabellen 68 bis 70 entnommen werden.

Tabelle 68: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „räumliches Vorstellungsvermögen niedrig ausgeprägt“ bis „räumliches Vorstellungsvermögen hoch ausgeprägt“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

räumliches Vorstellungsvermögen	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 niedrig ausgeprägt	60	21,0					1,262	0,083
2	105	36,7	-0,34	0,24	-0,06	0,47		
3	109	38,1	0,26	0,23	-0,05	0,46		
4 hoch ausgeprägt	12	4,2					0,586	0,882
Gesamt	286	100,0						

Tabelle 69: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

räumliches Vorstellungsvermögen	n	%	Mw WLE	s WLE	Mw Sco	s Sco
1 niedrig ausgeprägt	60	21,0	-0,64	0,81	4,72	2,21
2	105	36,7	-0,42	0,75	5,32	2,06
3	109	38,1	-0,06	0,73	6,33	2,08
4 hoch ausgeprägt	12	4,2	-0,01	0,84	6,42	2,39
Gesamt	286	100,0				

Tabelle 70: Statistische Kennwerte der Varianzanalyse für die Aussage 1 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	15,67	3	5,22
Innerhalb	162,90	282	0,58
Gesamt	178,57	285	

Die Varianzanalyse zeigt mit einem F-Wert von 9,043 bei homogenen Varianzen (Levene Statistik = 0,512; $p=0,674$) ein signifikantes Ergebnis ($p<0,01$). Signifikante Unterschiede im Post-Hoc-Test (nach Scheffé, siehe Tabelle 71) zeigen sich zwischen Stufe 1 (niedrig ausgeprägt) und 3 sowie 2 und 3. Die Leistungen steigen erwartungsgemäß von Stufe 1 bis 4 kontinuierlich an (wegen der geringen Anzahl von nur 12 Testpersonen, die Stufe 4 („hoch ausgeprägt“) markierten, gibt es keine signifikanten Unterschiede zu anderen Stufen). Das bedeutet, dass die Testpersonen ihr Raumvorstellungsvermögen sehr realistisch in Bezug auf ihre Leistungen bei den Endlosschleifenaufgaben einschätzen.

Tabelle 71: Scheffé-Test zur Einschätzung des Raumvorstellungsvermögens mit den mittleren Differenzen (mittl. Diff.) und der Wahrscheinlichkeit (p).

Einschätzung der RV		mittl. Diff.	p
1 niedrig ausgeprägt	2	-0,22	0,367
	3	-0,58	<0,01
	4	-0,62	0,083
2	1	0,22	0,367
	3	-0,36	0,008
	4	-0,41	0,384
3	1	0,58	<0,01
	2	0,36	0,008
	4	-0,04	0,998
4 hoch ausgeprägt	1	0,62	0,083
	2	0,41	0,384
	3	0,04	0,998

Neben Leistungsunterschieden in Abhängigkeit der Einschätzung der Raumvorstellungsfähigkeit interessieren die Zusammenhänge zwischen der Einschätzung und dem Geschlecht. Hier wird – wie im Kapitel 2.4.2.3 (Soziale und Sozokulturelle Ursachen) berichtet – angenommen, dass sich innerhalb der Frauen mehr Personen in jener Gruppe befinden, die ihre Fähigkeiten schlechter einschätzen. Um eine sinnvolle Anzahl von Testpersonen in den vier Feldern einer Kreuztabelle (Einschätzung und Geschlecht) zu erhalten, werden die Stufen 1 und 2 („eher niedrig ausgeprägt“) sowie 3 und 4 („eher hoch ausgeprägt“) zusammengefasst. Tabelle 72 zeigt die Verteilung der Frauen und Männer auf die beiden Stufen ($n=285$).

Tabelle 72: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Einschätzung der Raumvorstellungsfähigkeit („eher niedrig ausgeprägt“, „eher hoch ausgeprägt“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 1.

			Einschätzung der RV	
			eher niedrig ausgeprägt	eher hoch ausgeprägt
Geschlecht	m	beobachtetes n	49	77
		% innerhalb Geschlecht	38,9	61,1
		% innerhalb Aussage 1	29,7	64,2
	w	beobachtetes n	116	43
		% innerhalb Geschlecht	73,0	27,0
		% innerhalb Aussage 1	70,3	35,8

Ein χ^2 -Test zeigt mit $p < 0,01$ (1-seitig) signifikante Verteilungsunterschiede zwischen Geschlecht und Einschätzung der Raumvorstellungsfähigkeit: Unter den Frauen schätzen 73% ihre räumlichen Fähigkeiten eher niedrig ein, 27% eher hoch. Bei den Männern ist das Verhältnis umgekehrt: 61,1% sprechen von eher guten Fähigkeiten, 38,9% von einer eher niedrigen Ausprägung.

Aussage 2

Die Bearbeitung der Testaufgaben fiel mir

sehr leicht ○ ○ ○ ○ sehr schwer.

Bei Aussage 2 wird untersucht, ob sich die Leistungen der Testpersonen je nach vorgenommener Markierung unterscheiden. Da nur 3 Testpersonen die Markierung „sehr leicht“ wählten, werden die Testpersonen in zwei Gruppen zusammengefasst (sinngemäß „eher leicht“ versus „eher schwer“) und ein t-Test berechnet. Eine a-priori-Hypothese lautet, dass Personen, denen die Bearbeitung der Testaufgaben leicht fiel (Stufe 1 und 2), besser abschneiden als Personen, die Stufe 3 oder 4 markierten. Tabelle 73 können die Ergebnisse der Prüfung auf Normalverteilung, Tabelle 74 Informationen über Mittelwert und Standardabweichungen der WLE-Parameter und der Scores entnommen werden.

Tabelle 73: Prüfung auf Normalverteilung der WLE-Parameter in den 2 Teilstichproben „Bearbeitung der Testaufgaben fiel eher leicht“ und „Bearbeitung der Aufgaben fiel eher schwer“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Bearbeitung der Aufgaben fiel ...	N	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 sehr leicht	3	1,1	-0,14	0,17	-0,35	0,33	0,980	0,292
2	66	23,2						
3	169	59,3						
4 sehr schwer	47	16,5						
Gesamt	285	100,0						

Tabelle 74: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage 2 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Bearbeitung der Aufgaben fiel ...	N	%	Mw WLE	s WLE	Mw Sco	s Sco
eher leicht	69	24,2	-0,22	0,71	5,86	2,02
eher schwer	216	57,8	-0,34	0,82	5,55	2,26
Gesamt	285	100,0				

Der t-Test ist bei homogenen Varianzen ($F=2,799$; $p=0,095$) mit $t=1,110$ und $p=0,268$ nicht signifikant. Die Leistungen bei den Endlosschleifenaufgaben unterscheiden sich nicht in Abhängigkeit der subjektiv empfundenen Leichtigkeit bzw. Schwierigkeit der Testaufgaben; als Trend ist jedoch erkennbar, dass die Mittelwerte des Scores bzw. des WLE-Parameters bei Testpersonen, die Stufe 1 oder 2 („eher leicht“) markierten, höher sind als bei Testpersonen, die Stufe 3 oder 4 wählten („eher schwer“).

Zusätzlich soll bei Aussage 2 die Verteilung der Geschlechter auf die beiden zusammengefassten Stufen der Fragebogensaussage untersucht werden, wobei erwartet wird, dass sich mehr Frauen als Männer unter jenen Personen finden, denen die Bearbeitung schwer fiel. Tabelle 75 zeigt die Aufteilung der Testpersonen auf die beiden Stufen der Aussage ($n=284$).

Tabelle 75: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogen-aussage 2 („Bearbeitung der Aufgaben fiel eher leicht“ versus „eher schwer“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 2.

			Bearbeitung der Aufgaben fiel	
			eher leicht	eher schwer
Geschlecht	m	beobachtetes n	40	85
		% innerhalb Geschlecht	32,0	68,0
		% innerhalb Aussage 2	58,8	39,4
	w	beobachtetes n	28	131
		% innerhalb Geschlecht	17,6	82,4
		% innerhalb Aussage 2	41,2	60,6

Die mittels χ^2 -Test überprüfte Verteilung von Männern und Frauen auf die beiden Felder „eher leicht“ versus eher „schwer“ ist mit $p=0,004$ (1-seitig) signifikant. Sowohl innerhalb der Frauen als auch innerhalb der Männer finden sich deutlich mehr Personen in der Gruppe jener, denen die Aufgabenbearbeitung eher schwer fiel (82,4% bzw. 68,0%). Nur 17,6% der Frauen und 32,0% der Männer gaben an, dass ihnen die Bearbeitung eher leicht fiel. Das Verhältnis der Männer auf die beiden Fragebogensgruppen beträgt ca. 1/3 zu 2/3, bei Frauen ca. 1/6 zu 5/6.

Aussage 3

Die Bearbeitung der Testaufgaben war für mich

anstrengend ○ ○ ○ ○ nicht anstrengend.

Mittels Varianzanalyse soll geprüft werden, ob sich Testpersonen je nach Markierung auf der vierstufigen Skala in ihren Leistungen unterscheiden. Bezüglich des Ergebnisses wird erwartet, dass Personen, die die Testbearbeitung weniger anstrengend empfanden, besser abschneiden als Personen, die die Bearbeitung mehr anstregte. Die Überprüfung auf Normalverteilung (Tabelle 76) zeigt, dass die WLE-Parameter in einer Teilstichprobe (Stufe 3) nicht normalverteilt sind. Auf die Berechnung der Varianzanalyse muss verzichtet werden; stattdessen wird eine Rangvarianzanalyse gerechnet. Tabelle 77 können die Mittelwerte und Standardabweichungen der WLE-Parameter und Scores entnommen werden, Tabelle 78 die Kennwerte der Rangvarianzanalyse.

Tabelle 76: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Bearbeitung der Testaufgaben war anstrengend“ bis „Bearbeitung der Testaufgaben war nicht anstrengend“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Bearbeitung der Aufgaben war...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 anstrengend	50	17,5					1,036	0,233
2	117	40,9	-0,26	0,22	-0,30	0,44		
3	89	31,1					1,382	0,044
4 nicht anstrengend	30	10,5					0,774	0,587
Gesamt	286	100,0						

Tabelle 77: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 3 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Bearbeitung der Aufgaben war ...	n	%	Mw WLE	s WLE	Mw Sco	s Sco
1 anstrengend	50	17,5	-0,30	0,81	5,62	2,28
2	117	40,9	-0,31	0,77	5,62	2,14
3	89	31,1	-0,28	0,82	5,73	2,27
4 nicht anstrengend	30	10,5	-0,40	0,79	5,37	2,16
Gesamt	286	100,0				

Tabelle 78: Statistische Kennwerte der Rangvarianzanalysen mit den mittleren Rängen der WLE-Parameter für die 4 Antwortstufen der Aussage 3 samt Testpersonenanzahl (n).

Bearbeitung der Aufgaben war ...	N	mittl. Ränge
1 anstrengend	50	140,43
2	117	143,62
3	89	148,32
4 nicht anstrengend	30	133,83

Die Rangvarianzanalyse zeigt mit $p=0,851$ ein nicht signifikantes Ergebnis ($\chi^2=0,795$; $df=3$). Anhand der Mittelwerte der WLE-Parameter bzw. der Scores ist entgegen der oben geäußerten Vermutung erkennbar, dass jene Testpersonen am schlechtesten abschneiden, die die Bearbeitung der Aufgaben subjektiv als nicht anstrengend empfanden (Stufe 4).

Aussage 4

Das Testmaterial hat mir

gut gefallen ○ ○ ○ ○ nicht gut gefallen.

Der Frage nach den Auswirkungen auf die Testleistung in Abhängigkeit davon, ob Personen das Testmaterial gefiel oder nicht, kann ebenfalls mit einer Varianzanalyse nachgegangen werden. Ähnlich wie bei den bisher berichteten Aussagen besteht die Annahme, dass Personen besser abschneiden, denen der Test besser gefiel. Die Prüfung auf Normalverteilung ist in Tabelle 79 dargestellt, Mittelwerte und Standardabweichungen in Tabelle 80 sowie die Kennwerte der Varianzanalyse in Tabelle 81.

Tabelle 79: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testmaterial hat mir gut gefallen“ bis „Testmaterial hat mir nicht gut gefallen“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Testmaterial hat mir ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 gut gefallen	67	23,5					0,984	0,287
2	100	35,1	-0,53	0,24	0,27	0,48		
3	69	24,2					1,131	0,155
4 nicht gut gefallen	49	17,2					0,834	0,489
Gesamt	285	100,0						

Tabelle 80: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 4 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Das Testmaterial hat mir ...	N	%	Mw WLE	s WLE	Mw Sco	s Sco
1 gut gefallen	67	23,5	-0,25	0,71	5,76	2,02
2	100	35,1	-0,24	0,83	5,84	2,25
3	69	24,2	-0,36	0,76	5,46	2,16
4 nicht gut gefallen	49	17,2	-0,45	0,87	5,22	2,42
Gesamt	285	100,0				

Tabelle 81: Statistische Kennwerte der Varianzanalyse für die Aussage 4 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	1,86	3	0,62
Innerhalb	176,68	281	0,63
Gesamt	178,54	284	

Das Ergebnis ist mit einem F-Wert von 0,987 nicht signifikant ($p=0,399$), die Varianzen sind homogen (Levene Statistik=1,010; $p=0,389$). Aus den Mittelwerten der WLE-Parameter bzw. der Scores ist aber entsprechend der vorab geäußerten Vermutung erkennbar, dass Testpersonen besser abschneiden, denen der Test besser gefiel (Stufe 1 und 2) im Vergleich zu jenen, denen das Testmaterial weniger gut gefiel (Stufe 3 und 4).

Aussage 5

Ich habe mich bei der Bearbeitung der *Endlosschleifen* im Allgemeinen

- an der ganzen Schleife orientiert.
- an einem einzelnen markanten Teil der Schleife orientiert.
- an mehreren markanten Teilen der Schleife orientiert.

Bei dieser Aussage, die auf unterschiedliche Strategien abzielt, mussten die Testpersonen angeben, ob sie die Aufgaben eher holistisch (= Orientierung an der ganzen Schleife) oder analytisch (Orientierung an einem einzelnen / mehreren markanten Schleifenteilen) bearbeiteten. Mittels Varianzanalyse kann untersucht werden, ob es je nach vorgenommener Markierung Unterschiede in der Leistung gibt. In Tabelle 82 ist die Überprüfung auf Normalverteilung dargestellt; Tabelle 83 zeigt Mittelwerte und Standardabweichung der WLE-Parameter und Scores, Tabelle 84 die Kennwerte der Varianzanalyse.

Tabelle 82: Prüfung auf Normalverteilung der WLE-Parameter in den 3 Teilstichproben „Orientierung an der ganzen Schleife“, „Orientierung an einem markanten Schleifenteil“ und „Markierung an mehreren markanten Schleifenteilen“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Orientierung an ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 der ganzen Schleife	40	14,2					0,861	0,449
2 einem markanten Teil	78	27,7					1,104	0,174
3 mehrten Teilen	164	58,2	-0,23	0,19	-0,23	0,38		
Gesamt	282	100,0						

Tabelle 83: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der 3-stufigen Antwortskala der Aussage 5 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Orientierung an ...	n	%	Mw WLE	s WLE	Mw Sco	s Sco
der ganzen Schleife	40	14,2	-0,39	0,91	5,40	2,50
einem Teil der Schleife	78	27,7	-0,40	0,79	5,37	2,17
mehreren Teilen	164	58,2	-0,27	0,76	5,74	2,12
Gesamt	282	100,0				

Tabelle 84: Statistische Kennwerte der Varianzanalyse für die Aussage 5 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	1,16	2	0,58
Innerhalb	173,20	279	0,62
Gesamt	174,35	281	

Das Ergebnis der Varianzanalyse ist nicht signifikant ($F=0,934$; $p=0,394$) bei homogenen Varianzen (Levene Statistik=1,704; $p=0,184$); das bedeutet, dass sich die Leistungen in Abhängigkeit der angegebenen Bearbeitungsstrategie nicht unterscheiden.

Neben den Auswirkungen der Strategie auf die Leistung interessieren Zusammenhänge zwischen Geschlecht und Strategie. Wie im Kapitel 2.4.5 Geschlechtsunterschiede und Bearbeitungsstrategien berichtet, zeigte sich in mehreren Studien, dass Frauen eher zu analytischen Strategien tendieren, während Männer Raumvorstellungsaufgaben eher holistisch bearbeiten. In Tabelle 85 ist die Aufschlüsselung der Testpersonen nach Geschlecht auf die drei Antwortmöglichkeiten „Orientierung an der ganzen Schleife“, „an einem Teil der Schleife“ sowie „an mehreren Teilen der Schleife“ dargestellt ($n=281$).

Tabelle 85: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogen-aussage 5 („Orientierung an der ganzen Schleife“, „an einem Teil der Schleife“, „an mehreren Teilen der Schleife“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts, innerhalb der Aussage 5 sowie Gesamt innerhalb des Geschlechts.

			Orientierung an ...		
			der ganzen Schleife	einem Schleifenteil	mehreren Schleifenteilen
Geschlecht	m	beobachtetes n	18	38	66
		% innerhalb Geschlecht	14,8	31,1	54,1
		% innerhalb Aussage 5	45,0	48,7	40,5
	w	beobachtetes n	22	40	97
		% innerhalb Geschlecht	13,8	25,2	61,0
		% innerhalb Aussage 5	55,0	51,3	59,5
Gesamt		% innerhalb Geschlecht	14,2	27,8	58,0

Ein χ^2 -Test (2-seitig; gibt kein 1-seitig) zeigt ein nicht signifikantes Ergebnis ($p=0,472$). Die Aufteilung auf die drei Antwortalternativen innerhalb der Geschlechter sieht ähnlich aus. Die meisten Testpersonen gaben an, sich an mehreren Schleifenteilen orientiert zu haben (54,1% der Männer und 61,0% der Frauen). 31,1% der Männer sowie 25,2% der Frauen orientierten sich an einem markanten Schleifenteil, während die wenigsten anführten, holistisch zu arbeiten, d.h. sich an der ganzen Schleife zu orientieren (14,8% der Männer, 13,8% der Frauen). Über beide Geschlechter gesehen sprachen 58% von der Orientierung an mehreren Schleifenteilen, 27,8% von der Orientierung an einem Schleifenteil und 14,2% von einer holistischen Arbeitsweise. Diese beobachteten Ergebnisse unterstützen nicht die im Theorie-teil berichteten Resultate. Kritisch ist allerdings anzumerken, dass es in der durchgeführten Studie nur um eine globale und im Nachhinein erhobene Einschätzung geht (siehe Kritik-punkte von Schultz, 1991, Kapitel 2.3.2).

Aussage 6

Ich habe meine Testantworten im Allgemeinen

spontan gegeben ○ ○ ○ ○ nicht spontan gegeben.

Mittels Varianzanalyse kann geprüft werden, ob sich Testpersonen in ihren Leistungen unterscheiden, je nachdem, ob sie ihre Antworten spontan oder weniger spontan gegeben haben. Die a-priori-Hypothese lautet, dass die Leistungen steigen, je weniger spontan die Antworten gegeben wurden. In den Tabellen 86 bis 88 sind die Ergebnisse der Überprüfung auf Normalverteilung, die Mittelwerte und Standardabweichungen der WLE-Parameter bzw. Scores sowie die Kennwerte der Varianzanalyse dargestellt.

Tabelle 86: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testantworten spontan gegeben“ bis „Testantworten nicht spontan gegeben“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Testantworten ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 spontan gegeben	36	12,6					0,751	0,625
2	92	32,3	-0,001	0,25	-0,70	0,50		
3	94	33,0	-0,30	0,25	-0,33	0,49		
4 nicht spontan gegeb.	63	22,1					1,117	0,165
Gesamt	285	100,0						

Tabelle 87: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 6 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Ich habe die Testantworten	n	%	Mw WLE	s WLE	Mw Sco	s Sco
1 spontan gegeben	36	12,6	-0,62	0,6	4,72	1,80
2	92	32,3	-0,38	0,79	5,42	2,22
3	94	33,0	-0,23	0,89	5,87	2,45
4 nicht spontan gegeben	63	22,1	-0,16	0,63	6,02	1,81
Gesamt	289	100,0				

Tabelle 88: Statistische Kennwerte der Varianzanalyse für die Aussage 6 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	5,92	3	1,97
Innerhalb	171,31	281	0,61
Gesamt	177,23	284	

Da die Varianzen nicht homogen sind (Levene Statistic=4,608; $p=0,004$), muss eine Rangvarianzanalyse berechnet werden, deren Ergebnisse Tabelle 89 entnommen werden können.

Tabelle 89: Statistische Kennwerte der Rangvarianzanalysen mit den mittleren Rängen der WLE-Parameter für die 4 Antwortstufen der Aussage 6 samt Testpersonenanzahl (n).

Stufe	n	mittl. Ränge
1	36	108,76
2	92	136,15
3	94	152,35
4	63	158,62

Die Rangvarianzanalyse zeigt mit $\chi^2=10,499$ (df=3) und $p=0,015$ ein signifikantes Ergebnis. Anhand der Mittelwerte der WLE-Parameter bzw. der Scores (Tabelle 87) ist erkennbar, dass – wie angenommen – die Leistungen kontinuierlich steigen, je weniger spontan die Antworten gegeben wurden. Testpersonen, die spontan geantwortet haben (Stufe 1), lösen mit ca. 4,7 korrekten Items mehr als 1 Item weniger als Personen, die nicht spontan geantwortet haben (Stufe 4, ca. 6 gelöste Items).

Weiters interessiert die Verteilung der beiden Geschlechter auf die verschiedenen Antwortstufen. Um größere Fallzahlen zu erhalten, werden die Stufen 1 und 2 („Antworten eher spontan gegeben“) sowie Stufen 3 und 4 („Antworten eher nicht spontan gegeben“) zusammengefasst. In Anlehnung an die Literatur – Frauen bearbeiten Raumvorstellungsaufgaben vorsichtiger und daher langsamer (Goldstein, Haldane & Mitchell, 1990; Linn & Petersen, 1985; siehe Kapitel 2.4.3) – sollten sich bei den Frauen mehr in jener Gruppe befinden, die nicht spontan geantwortet haben. Die Ergebnisse der Verteilung der Geschlechter auf die beiden Antwortstufen sind in Tabelle 90 dargestellt.

Tabelle 90: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogensaussage 6 („Testantworten eher spontan gegeben“ versus „Testantworten eher nicht spontan gegeben“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 6.

			Testantworten ...	
			spontan gegeben	nicht spontan gegeben
Geschlecht	m	beobachtetes n	47	78
		% innerhalb Geschlecht	37,6	62,4
		% innerhalb Aussage 6	36,7	50,0
	w	beobachtetes n	81	78
		% innerhalb Geschlecht	50,9	49,1
		% innerhalb Aussage 6	63,3	50,0

Ein χ^2 -Test ergibt mit $p=0,017$ (1-seitig) ein signifikantes Ergebnis. Entgegen der Erwartung befinden sich ungefähr gleich viele Frauen in den Gruppen der spontan und nicht spontan Antwortenden (50,9% versus 49,1%). Bei den Männern zeigt sich ein anderes Bild: 37,6% gaben ihre Antworten spontan, 62,4% nicht spontan.

Aussage 7

Ich habe meine Testantworten im Allgemeinen

genau überprüft ○ ○ ○ ○ nicht genau überprüft.

Analog der Aussage 6 wird vermutet, dass jene Testpersonen bessere Leistungen erzielen, die ihre Antworten genauer überprüfen. Mittels Varianzanalyse kann untersucht werden, ob

sich die Leistungen je nach vorgenommener Markierung auf der vierstufigen Skala unterscheiden. Die Ergebnisse der Prüfung auf Normalverteilung können Tabelle 91 entnommen werden, die Informationen über Mittelwerte der WLE-Parameter und Scores Tabelle 92 sowie die Ergebnisse der Varianzanalyse Tabelle 93.

Tabelle 91: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testantworten genau überprüft“ bis „Testantworten nicht genau überprüft“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Testantworten ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 genau überprüft	22	7,7					0,616	0,842
2	84	29,5					1,284	0,074
3	107	37,5	-0,10	0,23	-0,11	0,46		
4 nicht genau überprüft	72	25,3					1,086	0,189
Gesamt	285	100,0						

Tabelle 92: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 7 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Ich habe die Testantworten	n	%	MW WLE	s WLE	MW Sco	s Sco
1 genau überprüft	22	7,7	0,14	0,87	6,86	2,49
2	84	29,5	-0,22	0,80	5,90	2,25
3	107	37,5	-0,37	0,74	5,44	2,05
4 nicht genau überprüft	72	25,3	-0,47	0,79	5,19	2,15
Gesamt	285	100,0				

Tabelle 93: Statistische Kennwerte der Varianzanalyse für die Aussage 6 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	7,38	3	2,46
Innerhalb	171,16	281	0,61
Gesamt	178,54	284	

Die Varianzanalyse zeigt mit $F=4,038$ und $p=0,008$ ein signifikantes Ergebnis bei homogenen Varianzen (Levene Statistic=0,699, $p=0,553$). Anhand der Mittelwerte der WLE-Parameter bzw. der Scores ist erkennbar, dass die Leistungen mit der Genauigkeit der Überprüfung kontinuierlich ansteigen. Jene Personen, die ihre Antworten genau überprüften (Stufe 1), lösten ca. 1,6 Items mehr als Personen, die nicht genau überprüften (Stufe 4). Der

nach Scheffé durchgeführte Post-Hoc-Test (Tabelle 94) zeigt einen signifikanten Unterschied zwischen den Stufen 1 und 4 ($p=0,018$) sowie einen knapp nicht signifikanten Unterschied zwischen den Stufen 1 und 3 ($p=0,051$).

Tabelle 94: Scheffé-Test zur Aussage 7 mit den mittleren Differenzen (mittl. Diff.) und der Wahrscheinlichkeit (p).

		mittl. Diff.	p
1 genau überprüft	2	0,36	0,297
	3	0,51	0,051
	4	0,61	0,018
2	1	-0,36	0,297
	3	0,15	0,615
	4	0,25	0,269
3	1	-0,51	0,051
	2	-0,15	0,615
	4	0,10	0,884
4 nicht genau überprüft	1	-0,61	0,018
	2	-0,25	0,269
	3	-0,10	0,884

Zusätzlich interessiert die Verteilung zwischen Geschlecht und den in zwei Gruppen zusammengefassten Antwortstufen 1 und 2 („Antworten eher genau überprüft“) sowie 3 und 4 („Antworten eher nicht genau überprüft“). Sie kann Tabelle 95 entnommen werden ($n=284$).

Tabelle 95: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogen-aussage 7 („Testantworten eher genau überprüft“ versus „Testaufgaben eher nicht genau überprüft“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 7.

			Testantworten ...	
			genau überprüft	nicht genau überprüft
Geschlecht	m	beobachtetes n	55	71
		% innerhalb Geschlecht	43,7	56,3
		% innerhalb Aussage 7	51,9	39,9
	w	beobachtetes n	51	107
		% innerhalb Geschlecht	32,3	67,7
		% innerhalb Aussage 7	48,1	60,1

Der χ^2 -Test ist mit $p=0,033$ (1-seitig) signifikant. Die Verteilung innerhalb der Geschlechter zeigt, dass mehr Frauen ihre Antworten nicht genau überprüften (67,7%) im Vergleich zu Männern (56,3%). 32,3% der Frauen und 43,7% der Männer überprüften ihre Antworten genau.

Aussage 8

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben **bildlich** (mittels „Bilder im Kopf“) zu verdeutlichen.

trifft zu ○ ○ ○ ○ trifft nicht zu

Mittels Varianzanalyse wird untersucht, ob sich Leistungsunterschiede in Abhängigkeit davon zeigen, ob Testpersonen die Aufgaben bildlich verdeutlichten oder nicht. Die Überprüfung auf Normalverteilung ist Tabelle 96 zu entnehmen, Mittelwerte und Standardabweichungen der WLE-Parameter und Scores Tabelle 97 sowie die Ergebnisse der Varianzanalysen Tabelle 98.

Tabelle 96: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testaufgaben bildlich verdeutlicht – trifft zu“ bis „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Testaufgaben bildlich verdeutlicht	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 trifft zu	161	56,7	-0,05	0,19	-0,06	0,38		
2	82	28,9					1,292	0,071
3	27	9,5					0,701	0,710
4 trifft nicht zu	14	4,9					0,618	0,840
Gesamt	284	100,0						

Tabelle 97: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 8 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Testaufgaben bildlich Verdeutlicht	n	%	MW WLE	s WLE	MW Sco	s Sco
1 trifft zu	161	56,7	-0,34	0,78	5,54	2,17
2	82	28,9	-0,21	0,77	5,89	2,21
3	27	9,5	-0,46	0,79	5,22	2,12
4 trifft nicht zu	14	4,9	-0,47	0,86	5,21	2,29
Gesamt	284	100,0				

Tabelle 98: Statistische Kennwerte der Varianzanalyse für die Aussage 8 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	1,81	3	0,602
Innerhalb	171,99	280	0,614
Gesamt	173,80	283	

Das Ergebnis der Varianzanalyse ist mit $p=0,403$ nicht signifikant ($F\text{-Wert}=0,979$), die Varianzen sind homogen (Levene Statistik=0,109; $p=0,955$). Das bedeutet, dass es keine Leistungsunterschiede in Abhängigkeit davon gibt, ob Testpersonen die Aufgaben bildlich verdeutlicht haben oder nicht.

Aussage 9

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben **sprachlich** (mittels verbaler Formulierungen) zu verdeutlichen.

trifft zu ○ ○ ○ ○ trifft nicht zu

Da bei dieser Aussage nur sechs Testpersonen die Stufe 1 „trifft zu“ markierten, werden zwei Gruppen gebildet und ein t-Test berechnet. Die Überprüfung auf Normalverteilung ist in Tabelle 99 dargestellt, die Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores in Tabelle 100.

Tabelle 99: Prüfung auf Normalverteilung der WLE-Parameter in den 2 Teilstichproben „Testaufgaben sprachlich verdeutlicht – trifft zu“ und „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Testaufgaben sprachlich verdeutlicht	N	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 trifft zu	6	2,1	-0,14	0,15	-0,24	0,30	0,603	0,860
2	21	7,3						
3	49	17,1						
4 trifft nicht zu	210	73,4						
Gesamt	286	100,0						

Tabelle 100: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage 9 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Testaufgaben sprachlich Verdeutlicht	n	%	Mw WLE	s WLE	Mw Sco	s Sco
trifft eher zu	27	9,4	-0,27	0,80	5,74	2,26
trifft eher nicht zu	259	90,6	-0,31	0,79	5,61	2,20
Gesamt	286	100,0				

Der t-Test ist mit $t=0,277$ und $p=0,782$ nicht signifikant, die Varianzen sind homogen ($F=0,017$; $p=0,897$). Das bedeutet, dass keine Leistungsunterschiede zwischen der Gruppe, die die Aufgaben eher verbal verdeutlicht und jener, die eher nicht verbal verdeutlicht, auftreten.

Bei der Aussage 9 interessiert auch die Verteilung der Geschlechter auf die beiden zusammengefassten Antwortstufen mit der Annahme, dass Frauen eher zu verbalen Verdeutlichungen neigen (siehe z.B. Casey 1996, Kapitel 2.4.6). Tabelle 101 zeigt die Verteilung zwischen Geschlecht und der sprachlichen versus nicht-sprachlichen Verdeutlichung (n=285).

Tabelle 101: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogen-aussage 9 („Testantworten sprachlich verdeutlicht – trifft eher zu“ und „trifft eher nicht zu“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 9.

			Antworten verbal verdeutlicht	
			trifft eher zu	trifft eher nicht zu
Geschlecht	m	beobachtetes n	10	116
		% innerhalb Geschlecht	7,9	92,1
		% innerhalb Aussage 9	38,5	44,8
	w	beobachtetes n	16	143
		% innerhalb Geschlecht	10,1	89,9
		% innerhalb Aussage 9	61,5	55,2

Der χ^2 -Test ergibt mit $p=0,343$ (1-seitig) keine signifikanten Verteilungsunterschiede. Die oben geäußerte Vermutung kann nicht bestätigt werden: innerhalb der Frauen markierten nur 10,1%, dass sie die Antworten verbal verdeutlicht haben, während 89,9% eine nicht-verbale Verdeutlichung angaben. Bei den Männern sieht das Verhältnis ähnlich aus: 7,9% der Männer gaben eine verbale Verdeutlichung an, für 92,1% traf die verbale Verdeutlichung nicht zu.

Aussage ab1

Die Bearbeitungsrichtung im Test war fix vorgegeben – von links oben (Startansicht) nach rechts unten (Zielansicht). Bei manchen Aufgaben hätte ich lieber von rechts unten nach links oben (also in umgekehrter Richtung) gearbeitet.

trifft zu ○ ○ ○ ○ trifft nicht zu

Bei dieser Aussage, die nur die 195 Personen mit den Testformen a und b bearbeiteten, wird mittels Varianzanalyse untersucht, ob sich die Leistungen in Abhängigkeit der vorgenommenen Markierung unterscheiden. Es wird angenommen, dass jene Personen besser abschneiden, die mit der vorgegebenen Bearbeitungsrichtung zufrieden waren (d.h. nicht lieber in umgekehrter Richtung gearbeitet hätten). Die Ergebnisse der Überprüfung der Normalverteilung sind Tabelle 102, Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores Tabelle 103 sowie die Ergebnisse der Varianzanalyse Tabelle 104 zu entnehmen.

Tabelle 102: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Ich hätte lieber in umgekehrter Richtung gearbeitet – trifft zu“ bis „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Ich hätte lieber in umgekehrter Richtung gearbeitet ...	n	%	Kolm. Z	p
1 trifft zu	32	16,9	0,926	0,358
2	46	24,3	0,886	0,412
3	47	24,9	1,095	0,182
4 trifft nicht zu	64	33,9	0,990	0,281
Gesamt	189	100,0		

Tabelle 103: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage ab1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Ich hätte lieber in umgekehrter Richtung gearbeitet ...	n	%	MW WLE	s WLE	MW Sco	s Sco
1 trifft zu	32	16,9	-0,43	0,87	5,31	2,46
2	46	24,3	-0,40	0,80	5,39	2,24
3	47	24,9	-0,35	0,84	5,53	2,30
4 trifft nicht zu	64	33,9	-0,21	0,71	5,89	2,03
Gesamt	189	100,0				

Tabelle 104: Statistische Kennwerte der Varianzanalyse für die Aussage ab1 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	1,52	3	0,51
Innerhalb	117,13	185	0,63
Gesamt	118,65	188	

Das Ergebnis der Varianzanalyse ist mit einem F-Wert von 0,801 bei homogenen Varianzen (Levene Statistik=1,853; $p=0,139$) nicht signifikant ($p=0,495$). Anhand der Mittelwerte der WLE-Parameter bzw. der Scores ist aber – wie vermutet – erkennbar, dass jene Personen am besten abschneiden, die mit der vorgegebenen Bearbeitungsrichtung zufrieden waren. Sie lösen um ca. ein halbes Item mehr als Personen, die (einige) Aufgaben lieber in umgekehrter Richtung bearbeitet hätten.

Aussage c1

Die freie Wahl der Bearbeitungsrichtung war für mich

sehr angenehm ○ ○ ○ ○ sehr unangenehm.

Diese Frage wurde nur jenen 94 Testpersonen vorgegeben, die die Testform c bearbeiteten. Aufgrund der geringen Fallzahlen werden die vier Antwortstufen zu zwei Gruppen zusammengefasst. Mittels t-Test wird überprüft, ob sich die Leistungen zwischen Testpersonen, denen die freie Wahl der Bearbeitungsrichtung eher angenehm und jenen, denen die freie Wahl eher unangenehm war, unterscheiden. In Tabelle 105 ist die Prüfung auf Normalverteilung dargestellt, in Tabelle 106 die Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und der Scores.

Tabelle 105: Prüfung auf Normalverteilung der WLE-Parameter in den 2 zusammengefassten Teilstichproben „Die freie Wahl der Bearbeitungsrichtung war eher angenehm“ versus „eher unangenehm“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Die freie Wahl der Bearbeitungsrichtung war ...	n	%	Kolm. Z	p
1 sehr angenehm	21	22,8	1,053	0,217
2	52	56,6		
3	10	10,9	0,660	0,776
4 sehr unangenehm	9	9,8		
Gesamt	92	100,0		

Tabelle 106: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage c1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Die freie Wahl der Bearbeitungsrichtung war	N	%	Mw WLE	s WLE	Mw Sco	s Sco
eher angenehm	73	79,3	-0,34	0,76	5,53	2,10
eher unangenehm	19	20,7	-0,10	0,95	6,21	2,55
Gesamt	92	100,0				

Der t-Test ist mit $t=-1,135$ und $p=0,259$ nicht signifikant, die Varianzen sind homogen ($F=0,502$; $p=0,481$). Bei näherer Betrachtung der Mittelwerte der WLE-Parameter bzw. der Scores fällt auf, dass jene Personen um ca. 0,7 gelöste Items besser abschneiden, denen die freie Wahl der Bearbeitungsrichtung eher unangenehm war, was ein nicht erwartetes Ergebnis ist.

Aussage c2

Die freie Wahl der Bearbeitungsrichtung fiel mir im Vergleich zur fix vorgegebenen Bearbeitungsrichtung (Aufgaben 1 bis 5)

sehr leicht ○ ○ ○ ○ sehr schwer.

Auch für die zweite Aussage, die nur die Testpersonen mit Form c bearbeiteten, werden zwei Gruppen gebildet und die Mittelwerte verglichen. Erwartet wird, dass jene Personen besser abschneiden, denen die freie Wahl der Bearbeitungsrichtung eher leicht fiel. Tabelle 107 kann die Prüfung auf Normalverteilung entnommen werden. Tabelle 108 liefert Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores.

Tabelle 107: Prüfung auf Normalverteilung der WLE-Parameter in den 2 zusammengefassten Teilstichproben „Die freie Wahl der Bearbeitungsrichtung fiel mir eher leicht“ versus „eher schwer“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Die freie Wahl der Bearbeitungsrichtung fiel mir ...	n	%	Kolm. Z	p
1 sehr leicht	16	17,2	0,838	0,484
2	53	57,0		
3	20	21,5	1,002	0,268
4 sehr schwer	4	4,3		
Gesamt	93	100,0		

Tabelle 108: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage c2 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Die freie Wahl der Bearbeitungsrichtung war ...	N	%	Mw WLE	s WLE	Mw Sco	s Sco
eher leicht	69	74,2	-0,34	0,83	5,54	2,26
eher schwer	24	25,8	-0,08	0,72	6,21	2,06
Gesamt	93	100,0				

Der t-Test ist mit $t=-1,379$ und $p=0,171$ nicht signifikant, die Varianzen sind homogen ($F=0,839$; $p=0,362$). Ähnlich der Aussage c1 fällt auf, dass jene Personen besser abschneiden, denen die freie Wahl der Bearbeitungsrichtung eher schwer fiel.

Aussage 10

Ich finde, dass der Endlosschleifentest ...

... räumliches Vorstellungsvermögen

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

Mittels Varianzanalyse kann untersucht werden, ob sich unterschiedliche Markierungen auf der vierstufigen Skala auf die Leistung auswirken. Die Prüfung auf Normalverteilung ist Tabelle 109 zu entnehmen, die Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores Tabelle 110 sowie die Kennwerte der Varianzanalyse Tabelle 111.

Tabelle 109: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst räumliches Vorstellungsvermögen in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Test erfasst RV in ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 geringem Ausmaß	14	4,9					0,649	0,794
2	65	22,8					0,873	0,432
3	118	41,4	-0,36	0,22	-0,22	0,44		
4 hohem Ausmaß	88	30,9					0,978	0,294
Gesamt	285	100,0						

Tabelle 110: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 10 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Test erfasst RV ...	n	%	MW WLE	s WLE	MW Sco	s Sco
1 in geringem Ausmaß	14	4,9	-0,19	0,66	5,93	1,94
2	65	22,8	-0,43	0,75	5,26	2,12
3	118	41,4	-0,38	0,80	5,44	2,18
4 in hohem Ausmaß	88	30,9	-0,14	0,81	6,09	2,29
Gesamt	285	100,0				

Tabelle 111: Statistische Kennwerte der Varianzanalyse für die Aussage 10 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	4,35	3	1,45
Innerhalb	174,19	281	0,62
Gesamt	178,54	281	

Das Ergebnis der Varianzanalyse ist bei homogenen Varianzen (Levene Statistik=0,310; $p=0,818$) mit $F=2,339$ knapp nicht signifikant ($p=0,074$). Am besten schneiden jene Personen ab, die meinen, dass die Endlosschleifenaufgaben Raumvorstellung in hohem Ausmaß (Stufe 4) erfassen; an zweiter Stelle jene, die meinen, dass die Aufgaben räumliches Vorstellungsvermögen in geringem Ausmaß erfassen (Stufe 1).

Aussage 11

Ich finde, dass der Endlosschleifentest ...

... logisches, schlussfolgerndes Denken

in geringem Ausmaß erfasst ○ ○ ○ ○ in hohem Ausmaß erfasst.

Auch bei dieser Aussage wird eine Varianzanalyse berechnet. Die Prüfung auf Normalverteilung ist in Tabelle 112 dargestellt, die Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores in Tabelle 113 sowie die Kennwerte der Varianzanalyse in Tabelle 114.

Tabelle 112: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst logisches, schlussfolgerndes Denken in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Test erfasst logisches Denken ...	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 in geringem Ausmaß	32	11,2					0,752	0,624
2	84	29,4					1,344	0,054
3	125	43,7	-0,14	0,22	-0,30	0,43		
4 in hohem Ausmaß	45	15,7					0,875	0,429
Gesamt	286	100,0						

Tabelle 113: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 11 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Test erfasst logisches Denken ...	n	%	MW WLE	s WLE	MW Sco	s Sco
1 in geringem Ausmaß	32	11,2	-0,41	0,72	5,31	2,05
2	84	29,4	-0,35	0,80	5,51	2,18
3	125	43,7	-0,23	0,79	5,86	2,22
4 in hohem Ausmaß	45	15,7	-0,39	0,84	5,42	2,30
Gesamt	286	100,0				

Tabelle 114: Statistische Kennwerte der Varianzanalyse für die Aussage 11 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	1,64	3	0,55
Innerhalb	176,93	282	0,63
Gesamt	178,57	285	

Das Ergebnis der Varianzanalyse ist mit $p=0,455$ ($F=0,873$) nicht signifikant, die Varianzen sind homogen (Levene Statistik=0,363; $p=0,780$). Am besten schneiden jene Personen ab, die Stufe 3, am schlechtesten jene, die Stufe 1 (in geringem Ausmaß) markierten.

Aussage 12

Ich finde, dass der Endlosschleifentest ...

... Konzentrationsfähigkeit

in geringem Ausmaß erfasst ○ ○ ○ ○ in hohem Ausmaß erfasst.

Bei der letzten Aussage des Fragebogens kann ebenfalls mittels Varianzanalyse überprüft werden, ob sich die Leistungen in Abhängigkeit der vorgenommenen Markierung unterscheiden. Die Prüfung auf Normalverteilung kann Tabelle 115 entnommen werden, Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores Tabelle 116 sowie die Kennwerte der Varianzanalyse Tabelle 117.

Tabelle 115: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst Konzentrationsfähigkeit in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Test erfasst Konzentrationsfähigkeit ...	N	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
1 in geringem Ausmaß	22	7,7					0,734	0,654
2	49	17,1					0,730	0,661
3	103	36,0	-0,17	0,24	-0,63	0,47		
4 in hohem Ausmaß	112	39,2	0,11	0,23	-0,12	0,45		
Gesamt	286	100,0						

Tabelle 116: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 12 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Test erfasst Konzentrationsfähigkeit ...	N	%	MW WLE	s WLE	MW Sco	s Sco
1 in geringem Ausmaß	22	7,6	-0,56	0,77	4,95	2,03
2	49	17,0	-0,33	0,75	5,57	2,08
3	103	35,6	-0,36	0,84	5,50	2,33
4 in hohem Ausmaß	112	38,8	-0,21	0,76	5,89	2,15
Gesamt	286	100,0				

Tabelle 117: Statistische Kennwerte der Varianzanalyse für die Aussage 12 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).

	QS	df	mittl. Q
Zwischen	2,85	3	0,95
Innerhalb	175,72	282	0,62
Gesamt	178,57	285	

Das Ergebnis der Varianzanalyse ist bei homogenen Varianzen (Levene Statistik=0,809; $p=0,490$) nicht signifikant ($F=1,522$; $p=0,209$). Am wenigsten Items lösen Personen, die markierten, dass der Test Konzentrationsfähigkeit in geringem Ausmaß misst (ca. 5 Items), am meisten jene, die „in hohem Ausmaß“ markierten (ca. 5,9 Items).

Zusammenfassung der Fragebogenergebnisse

Zusammenfassend ergaben die Fragebogenanalysen, dass jene Personen bei den Endlos-schleifenaufgaben signifikant besser abschnitten, die ihr räumliches Vorstellungsvermögen höher einschätzen. Signifikante Leistungsunterschiede konnten auch in Abhängigkeit davon beobachtet werden, ob die Antworten spontan oder weniger spontan gegeben sowie genau oder weniger genau überprüft wurden. Die Leistungen stiegen kontinuierlich mit der Genauigkeit der Überprüfung und der Abnahme an Spontaneität. Zwischen den Geschlechtern zeigt sich erwartungsgemäß, dass die Testbearbeitung mehr Frauen als Männern schwer fiel und – entgegen der Erwartung –, dass mehr Frauen ihre Antworten spontaner gaben und weniger genau überprüften als Männer.

Keine signifikanten Leistungsunterschiede zeigten sich in Zusammenhang mit der subjektiv empfundenen Leichtigkeit bzw. Schwierigkeit der Testaufgaben sowie der subjektiv empfundenen Anstrengung. Ebenso konnten keine Unterschiede in Abhängigkeit davon beobachtet werden, ob den Testpersonen das Aufgabenmaterial gefiel oder nicht. Auch die Aussagen zu verschiedenen Bearbeitungsstrategien (Orientierung an der ganzen Schleife/einem Schleif-

enteil/mehreren Teilen; bildliche versus nicht-bildliche Verdeutlichung; verbale versus nicht-verbale Verdeutlichung) ergaben keine signifikanten Leistungsunterschiede. Schließlich konnten keine signifikanten Leistungsunterschiede bei den Aussagen, ob die Endlosschleifen räumliches Vorstellungsvermögen, schlussfolgerndes Denken oder Konzentrationsfähigkeit messen, beobachtet werden.

Die spezifische Frage nach der Zufriedenheit mit der vorgegebenen Bearbeitungsrichtung für die Testpersonen mit den Formen a oder b ergab keine signifikanten Leistungsunterschiede, aber einen Trend in die Richtung, dass Personen schlechter abschnitten, die lieber in die andere Richtung gearbeitet hätten. Nicht signifikante Ergebnisse aber Trends in eine nicht erwartete Richtung erbrachten die zwei spezifischen Fragen für die Personen mit der Form c: Es schnitten jene Testpersonen am besten ab, denen die freie Wahl der Bearbeitungsrichtung eher unangenehm war und schwer fiel.

3.7.4.2 Fragebogen zur Händigkeit

Der Händigkeitsfragebogen (in Anlehnung an Annett, 1970; siehe Anhang 3) umfasste insgesamt 15 Fragen. Bei zwölf Fragen mit der Formulierung „Welche Hand verwenden Sie gewöhnlich, wenn Sie ...“ ging es um die Handpräferenz bei verschiedensten Tätigkeiten wie Schreiben oder Schneiden, bei sportlichen Aktivitäten sowie diversen anderen Verrichtungen im täglichen Leben. Die Testpersonen wurden angehalten, sich die Bewegungsabläufe der gefragten Tätigkeiten vorzustellen. Es standen die vier Antwortalternativen „links“, „rechts“, „wechselnd“ (d.h. beide Hände) sowie „weiß nicht“ zur Verfügung. Tabelle 118 können die zwölf Tätigkeiten sowie die Häufigkeiten der markierten Antworten in Zahlen und Prozent entnommen werden.

Tabelle 118: Häufigkeiten in Zahlen (n) und Prozent (%) für die 12 Tätigkeiten im Händigkeitsfragebogen.

Frage	Brief Schreiben		Ball werfen		Tennis, Badminton, Squash		Streichholz anzünden		mit Schere schneiden	
	n	%	n	%	n	%	n	%	n	n
links	26	9,1	15	5,2	17	5,9	22	7,7	4	4,9
rechts	260	90,6	260	90,6	255	89,2	248	86,4	263	92,0
beide	1	0,3	11	3,8	13	4,5	16	5,6	9	3,1
weiß n.	0	0,0	1	0,3	1	0,3	1	0,3	0	0,0
Gesamt	287	100,0	287	100,0	286	100,0	287	100,0	286	100,0

Frage	Faden in Nadel fädeln		mit Besen aufkehren*		mit Schaufel schaufeln*		Spielkarten austeilen		Hammer verwenden	
	n	%	n	%	n	%	n	%	n	%
links	42	14,7	65	23,1	37	12,9	21	7,3	19	6,6
rechts	207	72,6	149	53,0	187	65,4	255	88,9	265	92,7
beide	29	10,2	56	19,9	45	15,7	9	3,1	2	0,7
weiß n.	7	2,5	11	3,9	17	5,9	2	0,7	0	0,0
Gesamt	285	100,0	281	100,0	286	100,0	287	100,0		100,0

Frage	Zähne Putzen		Glas aufschrauben	
	n	%	n	%
links	19	6,6	54	18,8
rechts	232	80,8	183	63,8
beide	35	12,2	50	17,4
weiß n.	1	0,3	0	0,0
Gesamt	287	100,0	287	100,0

* Welche Hand ist weiter oben?

Die übrigen drei Fragen bezogen sich auf das Umlernen der Schreibhand sowie das verletzungsbedingt notwendige Umlernen; die letzte offen formulierte Frage 15 („Wenn Sie für alle diese 12 Tätigkeiten die rechte Hand benutzen: Gibt es irgendwelche einhändigen Tätigkeiten, für die Sie die linke Hand benutzen?“) brachte keine Informationen.

Die Frage, ob die Schreibhand umgelernt wurde, beantworteten von insgesamt 287 Testpersonen 5 (1,7%) mit „ja“ und 282 Personen (98,3%) mit „nein“. 26 Testpersonen (9,1%) gaben an, Tätigkeiten (aus Verletzungsgründen) von einer auf die andere Hand umgelernt haben zu müssen, 261 Personen (90,9%) mussten das nicht tun (n=287).

Mit Hilfe des Händigkeitsfragebogens kann die Frage nach Leistungsunterschieden in Abhängigkeit der Schreibhand (siehe Frage 1 in Tabelle 118) untersucht werden. Von den 287 Testpersonen, die eine Antwortalternative markierten, gab eine Person an, mit beiden Händen zu schreiben, alle übrigen gaben die linke oder rechte Hand an. Aufgrund der gegebenen Datenlage muss die eine Person mit wechselnder Schreibhand ausgeschieden werden. Die Daten der 286 übrigen Testpersonen können mittels t-Test untersucht werden. Tabelle 119 zeigt die Prüfung auf Normalverteilung, Tabelle 120 die Mittelwerte und Standardabweichungen der WLE-Parameter und der Scores.

Tabelle 119: Prüfung auf Normalverteilung der WLE-Parameter in den beiden Teilstichproben „Schreibhand rechts“ bzw. „Schreibhand links“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Schreibhand	n	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
rechts	260	90,9	-0,14	0,15	-0,28	0,30		
links	26	9,1					0,763	0,605
Gesamt	286	100,0						

Tabelle 120: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „Schreibhand rechts“ und „Schreibhand links“ inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Schreibhand	n	%	Mw WLE	s WLE	Mw Sco	s Sco
Rechts	260	90,9	-0,33	0,80	5,59	2,21
Links	26	9,1	-0,14	0,75	6,08	2,15
n _{ges}	286	100,0				

Der t-Test ist bei homogenen Varianzen ($F=0,579$; $p=0,447$) mit $t=-1,132$ und $p=0,259$ nicht signifikant. Anhand der Mittelwerte der WLE-Parameter bzw. der Scores ist aber erkennbar, dass Personen, die mit der linken Hand schreiben ca. ein halbes Item mehr lösen als Personen, die rechts schreiben.

Weiters soll nicht nur die Schreibhand mit Leistung in Verbindung gesetzt werden, sondern anhand aller zwölf Tätigkeiten für jede Person eine Handpräferenz (linkshändig, rechtshändig oder beidhändig) gebildet werden, die auf Zusammenhänge mit Raumvorstellungsfähigkeit überprüft wird. Nach intensiven Überlegungen wurde für die Festlegung der Handpräferenz folgende Vorgehensweise gewählt: Unter den zwölf Tätigkeiten in dem Fragebogen (siehe Tabelle 118 bzw. Anhang 3) sind fünf, die nur mit einer Hand ausgeführt werden, und zwar sind das Schreiben, einen Ball werfen, einen Tennis-, Badminton- oder Squash-Schläger halten, mit einer Schere schneiden sowie Zähne putzen. Diese fünf einhändigen Tätigkeiten werden für die Festlegung der Handpräferenz vorrangig herangezogen. Die übrigen sieben Tätigkeiten, bei denen wie beispielsweise beim Aufkehren mit einem Besen beide Hände benötigt werden, kommen dann zum Tragen, wenn anhand der einhändigen Tätigkeiten keine eindeutige Zuordnung möglich ist.

Von den 287 Testpersonen, die den Händigkeitsfragebogen bearbeiteten, waren 220 Personen (76,7%) anhand der fünf einhändigen Tätigkeiten eindeutig zuordenbar. Das bedeutet, dass sie alle fünf einhändigen Aktivitäten mit „rechts“ bzw. „links“ markierten oder vier Aus-

sagen mit „rechts“ bzw. „links“ markiert wurden und jeweils eine Aussage ausgelassen wurde. 67 Testpersonen (23,3%) waren anhand der einhändigen Tätigkeiten nicht eindeutig zuordenbar. Tabelle 121 können die Antwortmuster samt Häufigkeiten und Prozentangaben der 287 Testpersonen für die fünf einhändigen Tätigkeiten entnommen werden.

Tabelle 121: Antwortmuster der 287 Testpersonen bei den fünf einhändigen Tätigkeiten mit den Häufigkeiten in Zahlen (n) und Prozent (%).

Antwortmuster	Kommentar	n	%
5 links	anhand der 5 einhändigen Tätigkeiten eindeutig zuordenbar (n=220)	5	1,7
4 links, 1 Auslassung		1	0,3
5 rechts		211	73,5
4 rechts, 1 Auslassung		3	1,0
4 links 1 rechts	anhand der 5 einhändigen Tätigkeiten nicht eindeutig zuordenbar (n=67)	4	1,4
4 links, 1 wechselnd		4	1,4
3 links, 2 rechts		1	0,3
3 links, 2 wechselnd		2	0,7
3 links, 1 rechts, 1 wechselnd		1	0,3
2 links, 2 rechts, 1 Auslassung		1	0,3
2 links, 2 rechts, 1 wechselnd		1	0,3
4 rechts, 1 links		4	1,4
4 rechts, 1 wechselnd		32	11,1
3 rechts, 2 links		2	0,7
3 rechts, 1 links, 1 wechselnd		4	1,4
3 rechts, 2 wechselnd		9	3,1
2 rechts, 1 links, 2 wechselnd		1	0,3
3 wechselnd, 1 rechts, 1 links		1	0,3

Die Bildung einer Handpräferenz für die 67 Testpersonen mit einer nicht-eindeutigen Präferenz erfolgte anhand aller Aussagen, indem für jede Person alle „Rechts-“ sowie „Links-Markierungen“ zusammengezählt wurden. Die Markierung „wechselnd“ wurde zur Hälfte zu „rechts“, zur Hälfte zu „links“ dazugezählt. So entstand für jede Testperson ein individuelles Antwortmuster (z. B. 8 rechts, 4 links oder 7,5 rechts, 2,5 links, 2 Auslassungen). Zur abschließenden Beurteilung der Handpräferenz wurde ein 2/3-Kriterium herangezogen: Wurden zumindest zwei Drittel aller Aussagen mit rechts bzw. links markiert, erfolgte eine Zuordnung zu „Handpräferenz rechts“ bzw. „Handpräferenz links“. Waren weniger als zwei Drittel der Aussagen mit rechts bzw. links markiert, wurde die Person als beidhändig eingestuft. Anhand der beschriebenen Vorgehensweise konnten für die 287 Testpersonen die Handpräferenzen bestimmt werden: 263 Personen (91,6%) wurden als „rechtshändig“ eingestuft, 18 Personen (6,3%) als „linkshändig“ sowie 6 Personen (2,1%) als „beidhändig“. Tabelle 122 zeigt Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter und Scores in den drei Teilstichproben.

Tabelle 122: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „linkshändig“, „rechtshändig“ und „beidhändig“ inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).

Handpräferenz	N	%	Mw WLE	s WLE	Mw Sco	s Sco
rechtshändig	263	91,6	-0,31	0,79	5,61	2,21
linkshändig	18	6,3	-0,09	0,66	6,22	1,93
beidhändig	6	2,1	-0,79	0,86	4,33	2,34
Gesamt	287	100,0				

Die besten Leistungen erbrachten Personen mit einer linkshändigen Präferenz mit ca. 6,2 gelösten Items ab, gefolgt von Rechtshändern mit durchschnittlich 5,6 korrekt bearbeiteten Items. Die schlechtesten Leistungen erbrachten Beidhänder mit ca. 4,3 gelösten Items, wobei die geringe Fallzahl von 6 Testpersonen als kritisch zu betrachten ist.

Um der Frage nach Leistungsunterschieden in Abhängigkeit der Handpräferenz nachzugehen, wird ein t-Test zwischen den Teilstichproben „linkshändig“ und „rechtshändig“ berechnet (für die Beidhänder kann wegen der geringen Stichprobengröße kein Test durchgeführt werden). Die Prüfung auf Normalverteilung kann Tabelle 123 entnommen werden.

Tabelle 123: Prüfung auf Normalverteilung der WLE-Parameter in den beiden Handpräferenzgruppen „rechtshändig“ bzw. „linkshändig“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).

Handpräferenz	N	%	Schiefe	s Schiefe	Wölbung	s Wölbung	Kolm. Z	p
rechtshändig	263	93,6	-0,13	0,15	-0,26	0,30		
linkshändig	18	6,4					0,784	0,570
Gesamt	281	100,0						

Der t-Test ist mit $t=1,177$ und $p=0,240$ bei homogenen Varianzen ($F=1,980$; $p=0,160$) nicht signifikant. Linkshänder lösen aber um ca. 0,6 Items mehr als Rechtshänder. Dieser Trend entspricht einigen im Kapitel 2.5 Händigkeit berichteten Ergebnissen.

4. Abschließende Zusammenfassung und Bewertung der Arbeit

Die vorliegende Arbeit beschäftigt sich mit Endlosschleifen, einem Material zur Messung des räumlichen Vorstellungsvermögens. Konkret geht es um computergenerierte (im Vergleich zu photographischen) Endlosschleifen (Gittler & Arendasy, 2003a) – das sind gewundene, in sich geschlossene Schläuche, die in verschiedenen, jeweils um 90 Grad gedrehten Ansichten vorliegen. Ausgangspunkt der Arbeit sind nicht erwartete Ergebnisse nach der Vorgabe von photographischen Endlosschleifenaufgaben, die zeigen, dass die Aufgabenschwierigkeiten nach Veränderung der Bearbeitungsrichtung stark variieren. Es drängt sich die Frage auf, ob dieses Phänomen mit den photographischen Endlosschleifen und deren teilweise sichtbaren Standflächen in Zusammenhang stehen könnte oder ob auch bei computergenerierten Testaufgaben starke Schwankungen bei den Itemparametern auftreten.

Im theoretischen Teil der Arbeit wird Raumvorstellung im Rahmen verschiedener Intelligenztheorien beleuchtet. In manchen Theorien wird Raumvorstellung als unabhängiger Faktor gesehen (z.B. Thurstone, 1938, zitiert nach Amelang & Bartussek, 2001), in anderen als untergeordneter Faktor (z.B. Spearman, 1904, zit. nach Asendorpf, 1999). Auch bei der Struktur des Faktors Raumvorstellung gibt es in der Literatur verschiedene, meist mit Hilfe faktoranalytischer Methoden erforschte Konzepte, die von zwei oder mehr Subfaktoren sprechen. Die Subfaktoren unterscheiden sich dabei sowohl in der Benennung als auch den inhaltlichen Komponenten. Einen kognitiven Zugang zur Erforschung der Struktur des Faktors Raumvorstellung wählten Linn und Petersen (1985), die sich an den bei der Aufgabenbearbeitung ablaufenden Prozessen orientierten und auf diesem Weg die drei Kategorien Spatial Perception (Erkennen räumlicher Beziehungen zwischen Reizen und der Position des eigenen Körpers), Spatial Visualization (= Veranschaulichung; komplexeres Aufgabenmaterial, für dessen Bearbeitung ein Repertoire verschiedener Strategien günstig ist) und Mental Rotation (Rotation zwei- oder dreidimensionaler Figuren) definierten.

Nach Darstellung einiger Intelligenztheorien und Abhandlungen zur Struktur des Faktors Raumvorstellung widmet sich ein Kapitel des theoretischen Teils den Bearbeitungsstrategien bei Raumvorstellungsaufgaben. Übereinstimmung herrscht dahingehend, dass unterschiedliche Strategien angewendet werden, die je nach Aufgabenart, Schwierigkeit und Fähigkeit der Person variieren. Ebenso können die Begriffe der analytischen und holistischen Bearbeitungsweise bei vielen Forschern (z.B. Barratt, 1953; Cooper, 1976, 1980) nachgelesen werden. Die Untersuchungen von Schultz (1991) beschäftigen sich mit den Auswirkungen ver-

schiedener Strategien auf die Leistung und belegen, dass bei bestimmten Aufgabentypen die Verwendung bzw. die Vermeidung einer Strategie signifikant mit Leistung in Zusammenhang stehen. Eine andere wichtige Erkenntnis aus verschiedenen Studien (z.B. Geiser, Lehmann & Eid, 2006) ist, dass Testpersonen Aufgaben zum Teil mit anderen Strategien bearbeiten als von der Instruktion vorgesehen wäre, was wiederum bedeutet, dass die Aufgaben etwas anderes messen als sie zu messen vorgeben. Das Kapitel über Bearbeitungsstrategien schließt die Darstellung der Forschungsergebnisse von Just und Carpenter (1976, 1985), deren zentraler Stellenwert den interindividuell unterschiedlichen kognitiven Koordinatensystemen, in die räumliche Informationen eingebettet werden, zukommt.

Das Kapitel zu Geschlechtsunterschieden bei Raumvorstellung beginnt mit der Beschreibung der Metaanalysen von Linn und Petersen (1985) sowie Voyer, Voyer und Bryden (1995), die von empirisch nachweisbaren Geschlechtsunterschieden zugunsten der Männer bei mentalen Rotationsaufgaben berichten. Keine eindeutigen Ergebnisse finden sich in den Kategorien Spatial Perception und Spatial Visualization, in denen zum Teil signifikante Geschlechtsunterschiede resultieren, zum Teil nicht, auch in Abhängigkeit davon, welches Testmaterial verwendet, wie es vorgegeben und ausgewertet wurde. Einige Unterkapitel beschäftigen sich mit Ursachen für Geschlechtsunterschiede bei Raumvorstellungsleistungen. Als mögliche ausschlaggebende Faktoren werden körperliche, biologische Merkmale wie Hormone (McGee, 1979) oder die Hirnstruktur (Hausmann, 2007), evolutionäre Theorien (Männer mit einem größeren Aktionsradius können sich mit mehr Frauen paaren; siehe Jones, Braithwaite & Healy, 2003) sowie soziale/soziokulturelle Merkmale wie die Geschlechterrollenidentifikation (z.B. Hausmann, 2007) genannt. Weitere Inhalte des Kapitels beziehen sich auf Geschlechtsunterschiede bei Umgestaltung der Testinstruktion (z.B. Vorgabe ohne Zeitlimit), sowie Untersuchungen zur Veränderung der Größe der Geschlechtsunterschiede über die Zeit hinweg (mit uneinheitlichen Ergebnissen) oder durch Training (Verbesserung der räumlichen Fertigkeiten). Schließlich werden im Kapitel zu Geschlechtsunterschieden die Arbeiten von Casey (1996) vorgestellt, die sich mit Zusammenhängen zwischen Anlage- und Umweltaspekten und deren Auswirkung auf Raumvorstellungsleistungen, vor allem bei Frauen, beschäftigen. Basis für Caseys Studien ist die Theorie des right-shift-Faktors von Annett (1985), die Verbindungen zwischen der Händigkeit in der Familie und der Sprachverarbeitung im Hirn annimmt und Vorteile bei räumlichen Fertigkeiten für Rechtshänder mit linkshändigen Verwandten sieht. Casey (1996) kann in drei empirischen Studien ihre Annahmen von guten Raumvorstellungsleistungen bei Frauen mit einer bestimmten Organisationsstruktur im Hirn (Anlageaspekt) und entsprechenden Raumvorstellungserfahrungen (Umweltaspekt) bestätigen.

Im letzten Kapitel des Theorieteils, der „Händigkeit“, werden nach einer Begriffsdefinition und Angaben über die Verteilung der Händigkeit in der Bevölkerung Informationen über die Festlegung der Handpräferenz gegeben. Den Abschluss des theoretischen Teils bilden Untersuchungsergebnisse zum Zusammenhang zwischen Händigkeit und Raumvorstellungsleistungen. Einige Resultate gehen in die Richtung, dass Linkshändigkeit mit besseren räumlichen Leistungen, aber zugleich auch mit häufigeren Schwierigkeiten bei sprachlichen Fertigkeiten in Verbindung gebracht wird (z.B. die Theorie von Geschwind, Behan und Galaburda; Geschwind & Behan, 1982; Geschwind & Galaburda 1985a-c; McManus & Bryden, 1991).

Der empirische Teil beginnt mit der Entstehungsgeschichte des Testmaterials. Die Endlosschleifen entstanden in Anlehnung an die nicht-rasch-homogenen Schlauchfiguren (Stumpf & Fay, 1983) mit dem übergeordneten Ziel, Geltung des Rasch-Modells (Rasch, 1960, zitiert nach Fischer, 1974) zu erlangen. In der vorliegenden Arbeit kamen 15 computergenerierte Endlosschleifenaufgaben im Paper-Pencil-Format in drei Testversionen a, b und c zum Einsatz. Die ersten fünf Items der drei Testversionen waren gleich (Link-Items); die Items 6 bis 15 unterschieden sich in den Formen a und b nur in der Bearbeitungsrichtung, bei Form c war die Bearbeitungsrichtung frei wählbar. Die Testpersonen bearbeiteten die Aufgaben mit einer Perspektivwechsel- oder Rotationsstrategie. Neben den Endlosschleifenaufgaben wurden ein Fragebogen mit Aussagen zum Testmaterial, zur Einschätzung der eigenen Raumvorstellungsleistung sowie Bearbeitungsstrategien und ein Fragebogen zur Händigkeit vorgelegt.

Neben der Hauptfragestellung – ob bei computergenerierten Endlosschleifenaufgaben (analog einiger Ergebnisse aus Forschungsseminaren bei Prof. Gittler mit photographischen Endlosschleifen) nach Vertauschung der Start- und Zielansicht unterschiedliche Itemparameter auftreten – werden einige Nebenfragestellungen formuliert; unter anderem, ob Leistungsunterschiede in verschiedenen Subgruppen (Geschlecht, Alter, Bearbeitungszeit, ...) auftreten, ein Zusammenhang zwischen Raumvorstellungsleistungen und Händigkeit sowie Auswirkungen verschiedener persönlicher Eigenschaften und Einschätzungen auf die Leistung beobachtet werden können.

Nach der Stichprobenbeschreibung widmen sich zwei Unterkapitel der Daten- und Itemselektion: ein Item wurde wegen mangelnder Tiefeninformation eliminiert, fünf Testpersonen wurden wegen nicht ernsthafter Testbearbeitung ausgeschieden. 289 Datensätze und 14 Items gehen in die Analysen ein.

Den Schwerpunkt der Auswertungen bilden Analysen mit dem dichotomen logistischen Modell von Rasch (1960, zitiert nach Fischer, 1974). Aufgrund der Datenlage können 16 verschiedene Auswertungsmodelle aufgesetzt werden, für die jeweils 6 bis 7 Modelltests durchgeführt werden. Zuerst werden die drei Testformen getrennt voneinander mit vollständigen und unvollständigen Daten untersucht. Zusätzlich können bei der Testform c, bei der die Items in unterschiedliche Richtungen bearbeitet wurden, die Testaufgaben 6 bis 15 so behandelt werden, als wären sie ein (gleiches) Item oder als wären sie zwei (verschiedene) Items. Anschließend werden alle drei Testversionen gemeinsam und die Formen a und b gemeinsam analysiert. Hier wird analog oben mit vollständigen und unvollständigen Daten gerechnet sowie in einer Modellart davon ausgegangen, dass die Items 6 bis 15 gleich sind, in der anderen Modellart, dass sie verschieden sind. Im Zuge der Rasch-Analysen stellt sich heraus, dass ein weiteres Item ausgeschieden werden muss. Nach Durchführung der Modelltests niedriger versus hoher Score, niedriger versus hoher Median, männliches versus weibliches Geschlecht, niedriges versus hohes Alter, Unterricht versus kein Unterricht in Darstellender Geometrie, niedrige versus hohe Bearbeitungszeit sowie in zwei Modellen Testform a versus Testform b kann für alle 16 Modelle Rasch-Homogenität angenommen werden. Die Itemparameter unterscheiden sich also nicht nach Vertauschung der Start- und Zielansicht.

Nach Darstellung der Rasch-Analysen wird die Berechnung einiger Likelihood-Quotienten-Tests zwischen verschiedenen Rasch-Modellen beschrieben. Als Ergebnis zeigt sich, dass Modelle mit weniger Parametern die Daten genauso gut erklären wie Modelle mit mehr Parametern und daher für weitere Analysen bevorzugt herangezogen werden.

Die klassisch-testtheoretischen Auswertungen, für deren Berechnung die WLE-Parameter herangezogen werden, beginnen mit der Darstellung der Überprüfung von Unterschiedshypothesen zwischen Testpersonengruppen. Signifikante Leistungsunterschiede können in Abhängigkeit des Geschlechts, des Alters und der Bearbeitungszeit beobachtet werden: männliche Testpersonen, Personen mit einem höheren Alter sowie einer höheren Bearbeitungszeit erbringen bessere Leistungen. Keine Unterschiede zeigen sich in Zusammenhang mit dem Besuch von Unterricht in Darstellender Geometrie, in Abhängigkeit der Testform und der Bearbeitungsstrategie.

Bezüglich des Fragebogens zum Testmaterial, zur Einschätzung der eigenen Raumvorstellungsleistung sowie Bearbeitungsstrategien zeigt sich, dass Personen signifikant besser abschneiden, die ihre räumlichen Fertigkeiten höher einschätzen, die ihre Antworten weniger spontan geben und genauer überprüfen. Keine signifikanten Leistungsunterschiede werden

in Abhängigkeit der empfundenen Leichtigkeit/Schwierigkeit der Testaufgaben, der empfundenen Anstrengung sowie in Abhängigkeit davon, ob das Testmaterial gefiel oder nicht beobachtet. Ebenso zeigen sich keine signifikanten Leistungsunterschiede in Zusammenhang mit verschiedenen Bearbeitungsstrategien und den Aussagen, ob Endlosschleifen räumliches Vorstellungsvermögen, schlussfolgerndes Denken oder Konzentrationsfähigkeit messen. Spezifische Fragen für die Testformen a und b mit der fix vorgegeben Bearbeitungsrichtung und der Form c mit der wahlfreien Bearbeitungsrichtung bringen keine signifikanten Leistungsunterschiede.

Die Auswertungen des Händigkeitsfragebogens ergeben, dass keine signifikanten Leistungsunterschiede sowohl in Abhängigkeit der Schreibhand (rechte oder linke Hand) als auch in Abhängigkeit einer allgemeinen Handpräferenz (rechts oder links) beobachtet werden; es besteht aber ein Trend in die Richtung, dass Personen, die mit der linken Hand schreiben bzw. eine linksseitige Handpräferenz haben, besser abschneiden.

Abschließend sollen im Zuge einer kritischen Auseinandersetzung mit der vorliegenden Arbeit einige verbesserungswürdige Punkte erwähnt werden:

- Die größtenteils an Schulen erhobenen Daten stellen keine repräsentative Stichprobe dar.
- Die Testvorgabe über den Computer hätte einige Vorteile mit sich gebracht, beispielsweise, dass Testpersonen bei entsprechender Programmierung keine Items hätten auslassen können oder dass fehlerhafte Markierungen (vor allem bei der Testform c) nicht möglich gewesen wären.
- Das Ausscheiden des uneindeutigen Items 11 hätte vielleicht verhindert werden können, indem mehr Personen im Rahmen einer Vortestung die Aufgaben bearbeitet hätten.
- Die im Fragebogen zum Test vorgegebenen Aussagen zu den Strategien können nur als vager Hinweis auf Bearbeitungsweisen genommen werden, da sie retrospektiv und auch global über alle Endlosschleifenaufgaben hinweg erfragt wurden.

Literaturverzeichnis

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123-140.

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, 61, 303-321.

Annett, M. (1985). *Left, right, hand, and brain: The right shift theory*. London: Erlbaum.

Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5., aktualisierte und erweiterte Aufl.). Stuttgart: Kohlhammer.

Amthauer, R. (1953). *Intelligenz-Struktur-Test I-S-T*. Göttingen: Hogrefe.

Amthauer, R. (1970). *Intelligenz-Struktur-Test I-S-T 70*. Göttingen: Hogrefe.

Arnold-Schulz-Gahmen, B. E., Ehrenstein, W. H., Schweingruber, T., Selinski, S., Urfer, W. & Zschiesche, E. (1998). Laterality of eye, ear, hand, and foot: Distribution and implications for sensory-motor performance. *Pfügers Archiv - European Journal of Physiology, Supplement to Vol. 435*, R 228.

Arendasy, M. (2000). Psychometrischer Vergleich computergestützter Vorgabeformen bei Raumvorstellungsaufgaben: stereoskopisch-dreidimensionale und herkömmlich-zweidimensionale Darbietung. Dissertation, Universität Wien.

Asendorpf, J. (1999). *Psychologie der Persönlichkeit* (2., überarbeitete und aktualisierte Aufl.). Berlin: Springer.

Baenninger, M. & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, 20, 327- 344.

Baenninger, M. & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, 7, 363-379.

Barratt, E. S. (1953). An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors. *Journal of Psychology*, 36, 17-25.

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Auflage). Heidelberg: Springer.

Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.

Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Cashmore, L., Uomini, N. & Chapelain, A. (2008). The evolution of handedness in humans and great apes: a review and current issues. *Journal of Anthropological Sciences*, 86, 7-35.

Casey, M. B. (1996). Understanding individual differences in spatial ability within females: a nature/nurture interactionist framework. *Developmental Review*, 16, 241-260.

Casey, M. B. & Brabeck, M. M. (1989). Exceptions to the male advantage on a spatial task: Family handedness and college major as a factor identifying women who excel. *Neuropsychologia*, 27, 689-696.

Casey, M. B., Colón, D. & Goris, Y. (1992). Family handedness as a predictor of spatial ability among minority girls in a math-science training program. *Brain and Cognition*, 18, 88-96.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2. Aufl.). New York: Academic Press.

Cooper, L. A. (1976). Individual differences in visual comparison processes. *Perception & Psychophysics*, 19, 433-444.

Cooper, L. A. (1980). Spatial information processing: Strategies for research. In R. E. Snow, P. A. Federico & W. E. Montague (Hrsg.), *Aptitude, learning and instruction: Cognitive process analyses* (S. 149-176). Hillsdale, NJ: Erlbaum.

- Cooper, L. A. & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Hrsg.), *Visual information processing* (S. 75-176). New York: Academic Press.
- Ekstrom, R. B., French, J. W. & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- Eliot, J. & Smith, I. M. (1983). *An international directory of spatial tests*. Windsor: Nfer-Nelson.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95-103.
- Feng, J., Spence, I. & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18, 850-855.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.
- Fischer, G. H. (1997). Unidimensional linear logistic Rasch models. In W. J. Van der Linden & R. K. Hambleton (Hrsg.), *Handbook of modern item response theory* (S. 225-244). New York: Springer.
- Fischer, G. H. & Molenaar, I. W. (1995). *Rasch Models. Foundations, Recent Developments, and Applications*. New York: Springer.
- French, J. W. (1965). The relationship of problem solving styles to the factor composition of tests. *Educational and Psychological Measurement*, 25, 9-28.
- Gardner, H. (1994). *Abschied vom IQ. Die Rahmen-Theorie der vielfachen Intelligenzen*. Stuttgart: Klett-Cotta.
- Gardner, H. (2002). *Intelligenzen. Die Vielfalt des menschlichen Geistes*. Stuttgart: Klett-Cotta.
- Geiser, C., Lehmann, W. & Eid, M. (2006). Separating „Rotators“ from „Nonrotators“ in the Mental Rotations Test: A multigroup Latent Class Analysis. *Multivariate Behavioral Research*, 41, 261-293.

Geiser, C., Lehmann, W. & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, 36, 556-563.

Geschwind, N. & Behan, P. (1982). Left-handedness: Association with immune disease, migraine, and developmental learning disorder. *PNAS*, 79, 5097-5100.

Geschwind, N. & Galaburda, A. M. (1985a). Cerebral Lateralization: Biological Mechanisms, Associations, and Pathology: I. A Hypothesis and a Program for Research. *Archives of Neurology*, 42, 428-459.

Geschwind, N. & Galaburda, A. M. (1985b). Cerebral Lateralization: Biological Mechanisms, Associations, and Pathology: II. A Hypothesis and a Program for Research. *Archives of Neurology*, 42, 521-552.

Geschwind, N. & Galaburda, A. M. (1985c). Cerebral Lateralization: Biological Mechanisms, Associations, and Pathology: III. A Hypothesis and a Program for Research. *Archives of Neurology*, 42, 634-654.

Gittler, G. (1994). Intelligenzförderung durch Schulunterricht: Darstellende Geometrie und räumliches Vorstellungsvermögen. In G. Gittler, M. Jirasko, U. Kastner-Koller, C. Korunka & A. Al-Roubaie (Hrsg.), *Die Seele ist ein weites Land. Aktuelle Forschung am Wiener Institut für Psychologie* (S. 105-122). Wien: WUV.

Gittler, G. & Arendasy, M. (2003a). Endlosschleifen: Psychometrische Grundlagen des Aufgabentyps E^P. *Diagnostica*, 49, 164-175.

Gittler, G. & Arendasy, M. (2003b). Psychometrische Effekte von Farbgebung. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24, 23-31.

Gladue, B. A. & Bailey, J. M. (1995). Spatial ability, handedness, and human sexual orientation. *Psychoneuroendocrinology*, 20, 487-497.

Gladue, B. A., Beatty, W. W., Larson, J. & Staton, R. D. (1990). Sexual orientation and spatial ability in men and women. *Psychobiology*, 18, 101-108.

Glück, J. (1999). *Spatial Strategies. Kognitive Strategien bei Raumvorstellungsleistungen*. Universität Wien: Unveröffentlichte Dissertation.

Goldstein, D., Haldane, D. & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition*, 18, 546-550.

Grüßing, M. (2002). Wieviel Raumvorstellung braucht man für Raumvorstellungsaufgaben? Strategien bei der Bewältigung räumlich-geometrischer Anforderungen. *ZDM - The International Journal on Mathematics Education*, 34, 37-45.

Halpern, D. F. (1996). Sex, brains, hands, and spatial cognition. *Developmental Review*, 16, 261-270.

Hausmann, M. (2007). Kognitive Geschlechtsunterschiede. In S. Lautenbacher, O. Güntürkün und M. Hausmann (Hrsg.), *Gehirn und Geschlecht. Neurowissenschaft des kleinen Unterschieds zwischen Frau und Mann* (S. 105-123). Heidelberg: Springer.

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H. M., Cohen-Kettenis, P. Z. & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, 114, 1245-1250.

Hellige, J. B. (1994). Handedness. In *Encyclopedia of human behavior* (S. 491-500). New York: Academic Press.

Horan, P.F. & Rossler, R. A. (1984). A multivariate analysis of spatial abilities by sex. *Developmental Review*, 4, 387-411.

Hosenfeld, I., Strauss, B. & Köller, O. (1997). Geschlechtsdifferenzen bei Raumvorstellungsaufgaben – eine Frage der Strategie? *Zeitschrift für Pädagogische Psychologie/German Journal of Educational Psychology*, 11, 85-94.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w and d. *American Psychologist*, 36, 892-901.

Janowsky, J. S., Oviatt, S. K. & Orwoll, E. S. (1994). Testosterone influences spatial cognition in older men. *Behavioral Neuroscience*, 108, 325-332.

Jäger, A. O., Süß, H. M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test*. Göttingen: Hogrefe.

Jones, C. M., Braithwaite, V. A. & Healy, S. D. (2003). The evolution of sex differences in spatial ability. *Behavioral Neuroscience*, 117, 403-411.

Just, M. A. & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441-480.

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.

Kail, R., Carter, P. & Pellegrino, J. (1979). The locus of sex differences in spatial ability. *Perception & Psychophysics*, 26, 182-186.

Köller, O., Rost, J. & Köller, M. (1994). Individuelle Unterschiede beim Lösen von Raumvorstellungsaufgaben aus dem IST- bzw. IST-70-Untertest „Würfelaufgaben“. *Zeitschrift für Psychologie*, 202, 65-85.

Kubinger, K. D. & Jäger, R. S. (Hrsg.). (2003). *Schlüsselbegriffe der Psychologischen Diagnostik*. Weinheim: Beltz.

Kyllonen, P. C., Lohman, D. F. & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, 76, 1325-1345.

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56, 1479-1498.

Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Technical Report No. 8, Aptitude Research Project). Stanford: University, School of Education.

Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: University Press.

Maier, P. H. (1994). *Räumliches Vorstellungsvermögen. Komponenten, geschlechtsspezifische Differenzen, Relevanz, Entwicklung und Realisierung in der Realschule*. Frankfurt am Main: Peter Lang.

McBurney, D. H., Gaulin, S. J. C., Devineni, T. & Adams, C. (1997). Superior spatial memory of women: Stronger evidence for the gathering hypothesis. *Evolution and Human Behavior*, 18, 165-174.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86, 889-918.

McGee, M. G. (1982). Spatial abilities: The influence of genetic factors. In M. Potegal (Hrsg.), *Spatial abilities: Development and physiological foundations* (S. 199-222). New York: Academic Press.

McKeever, W. F., Rich, D. A., Deyo, R. A. & Conner, R. L. (1987). Androgens and spatial ability: Failure to find a relationship between testosterone and ability measures. *Bulletin of the Psychonomic Society*, 25, 438-440.

McManus, I. C. & Bryden, M. P. (1991). Geschwind's theory of cerebral lateralization: Developing a formal, causal model. *Psychological Bulletin*, 110, 237-253.

McWilliams, W., Hamilton, C. J. & Muncer, S. J. (1997). On mental rotation in three dimensions. *Perceptual and Motor Skills*, 85, 297-298.

Morfit, N. S. & Weekes, N. Y. (2001). Handedness and Immune Function. *Brain and Cognition*, 46, 209-213.

Neave, N. (2008). *Hormones and Behaviour. A Psychological Approach*. Cambridge: Cambridge University Press.

Paivio, A. (1982). The empirical case for dual coding. In J. C. Yuille (Hrsg.), *Imagery, memory, and cognition* (S. 307-332). Hillsdale, NJ: Erlbaum.

Petermann, F., Niebank, K. & Scheithauer, H. (2004). *Entwicklungswissenschaft. Entwicklungspsychologie - Genetik - Neuropsychologie*. Berlin: Springer.

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., 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.

Pezaris, E. & Casey, M. B. (1991). Girls who use „masculine“ problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition*, 17, 1-22.

Porac, C. & Coren, S. (1981). *Lateral Preferences and Human Behavior*. New York: Springer.

Putz-Osterloh, W. (1977). Über Problemlöseprozesse bei dem Test Würfelaufgaben aus dem Intelligenztest IST und IST-70 von Amthauer. *Diagnostica*, 23, 252-265.

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. (2002). Girl's spatial abilities: Charting the contributions of experience and attitudes in different academic groups. *British Journal of Educational Psychology*, 72, 245-260.

Rosenthal, R. (1979). The „file drawer problem“ and tolerance for null results. *Psychological Bulletin*, 86, 638-641.

Rosenthal, R. (1980). Summarizing significance levels. In R. Rosenthal (Ed.), *Quantitative assessment of research domains* (S. 33-46). San Francisco: Jossey-Bass.

Rost, J. (1988). *Quantitative und qualitative probabilistische Testtheorie*. Bern: Huber.

Rost, J. (1990). Rasch Models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.

Sanders, B. & Soares, M. P. (1986). Sexual maturation and spatial ability in college students. *Developmental Psychology*, 22, 199-203.

- Santrock, J. W. (2009). *A Topical Approach to Life-Span Development*. Boston: McGraw-Hill.
- Sattler, J. B. (2005). *Der umgeschulte Linkshänder oder Der Knoten im Gehirn* (9. Aufl.). Donauwörth: Auer.
- Schultz, K. (1991). The contribution of solution strategy to spatial performance. *Canadian Journal of Experimental Psychology*, 45, 474-491.
- Shepard, R. N. & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, 171, 701-703.
- Shih, M., Pittinsky, T. L. & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10, 80-83.
- Shute, V. J., Pellegrino, J. W., Hubert, L. & Reynolds, R. W. (1983). The relationship between androgen levels and human spatial abilities. *Bulletin of the Psychonomic Society*, 21, 465-468.
- Signorella, M. L. & Jamison, W. (1978). Sex differences in the correlations among field dependence, spatial ability, sex role orientation, and performance on Piaget's Water-Level-Task. *Developmental Psychology*, 14, 689-690.
- Silverman, I. & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, L. Cosmides & J. Tooby (Hrsg.), *The adapted mind: Evolutionary psychology and the generation of culture* (S. 487-503). New York: Oxford University Press.
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Stein, J. F & Stoodley, C. J. (2006). *Neuroscience: An Introduction*. London: Wiley.
- Stumpf, H. & Fay, E. (1983). *Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Göttingen: Hogrefe.
- Stumpf, H. & Klieme, E. (1989). Sex-related differences in spatial ability: More evidence for convergence. *Perceptual and Motor Skills*, 69, 915-921.

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 559-604.

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.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale - Revised*. New York: Psychological Corporation.

Williams, R. H., Zimmerman, D. W., Zumbo, B. D. & Ross, D. (2003). Charles Spearman: British Behavioral Scientist. *Human Nature Review*, 3, 114-118.

Wilson, F. R. (2002). *Die Hand – Geniestreich der Evolution. Ihr Einfluß auf Gehirn, Sprache und Kultur des Menschen*. Reinbek bei Hamburg: Rowohlt.

Witelson, S. F. (1987). Neurobiological aspects of language in children. *Child Development*, 58, 653-688.

Witelson, S. F. (1991). Neural sexual mosaicism: Sexual differentiation of the human temporo-parietal region for functional asymmetry. *Psychoneuroendocrinology*, 16, 131-153.

Witkin, H. A., Dyk, R. B. & Faterson, H. F. (1962). *Psychological differentiation*. New York: Wiley.

Young, P. A., Young, P. H. & Tolbert, D. L (2008). *Basic Clinical Neuroscience* (Second Edition). Philadelphia: Lippincott Williams & Wilkins.

d differences in cerebral organization. In M. A. Wittig & A. C. Petersen (Eds.), *Sex-related differences in cognitive functioning* (S. 121-143). New York: Academic Press.

Tabellenverzeichnis

Tabelle 1: Zusammenfassende Darstellung einiger Faktoren aus verschiedenen Strukturkonzepten von Raumvorstellung mit den Faktorennamen, den jeweiligen Forschern sowie dem Faktorinhalt.	22
Tabelle 2: Sechs Tests mit homogenen Effektgrößen samt Autoren, einer kurzen Testbeschreibung, der Kategoriezugehörigkeit (Kat.; MR = <i>Mental Rotation</i> , SV = <i>Spatial Visualization</i>) sowie den Ergebnissen bei der Prüfung auf Geschlechtsunterschiede (GU).....	44
Tabelle 3: Sechs Tests mit heterogenen Effektgrößen samt Autoren, einer kurzen Testbeschreibung, der Kategoriezugehörigkeit (Kat.; MR = <i>Mental Rotation</i> , SP = <i>Spatial Perception</i> , SV = <i>Spatial Visualization</i>), den Teilungskriterien samt Ergebnis sowie den Ergebnissen bei Prüfung auf Geschlechtsunterschiede (GU).....	45
Tabelle 4: Zusammenfassung einiger Evolutionsthesen samt Spezies, für die sie gelten, der Vorhersage, die sie treffen sowie dem Selektionsdruck.	48
Tabelle 5: Beschreibung der drei Testversionen.....	66
Tabelle 6: Altersverteilung in der erhobenen Stichprobe mit den Häufigkeiten in Zahlen (n) und Prozent (%).	70
Tabelle 7: Bildungsstand in der erhobenen Stichprobe mit den Häufigkeiten der angestrebten bzw. abgeschlossenen Ausbildungen in Zahlen (n) und Prozent (%).	71
Tabelle 8: Verteilung der Bearbeitungszeit mit den Häufigkeiten der berechneten Zeiten (in Minuten) in Zahlen (n) und Prozent (%).	72
Tabelle 9: Verteilung der Testversionen auf die Testpersonen mit den Häufigkeiten in Zahlen (n) und Prozent (%).	73
Tabelle 10 Häufigkeiten der markierten Fragezeichen-Symbole in Zahlen (n) und Prozent (%).....	75
Tabelle 11 Häufigkeiten der Auslassungen (¥) bzw. fehlerhaften Markierungen in Zahlen (n) und Prozent (%).	75
Tabelle 12: Häufigkeiten der ausgelassenen (¥), mit dem Fragezeichen-Symbol (?) sowie fehlerhaft markierten Items in Zahlen (n) und Prozent (%).	75
Tabelle 13: Anzahl der gelösten Items (Score) über alle drei Testformen mit Häufigkeitsangaben in Zahlen (n) und Prozent (%).	76
Tabelle 14: Verteilung des Scores in den Testformen a, b und c mit Häufigkeitsangaben in Zahlen (n_a , n_b , n_c) und Prozent ($\%_a$, $\%_b$, $\%_c$).	77
Tabelle 15: Auswertungsmöglichkeiten mit dem Rasch-Modell für die einzelnen Testformen getrennt voneinander mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden sowie der Bezeichnung der Modelle.....	81
Tabelle 16: Auswertungsmöglichkeiten mit dem Rasch-Modell für die Testformen gemeinsam mit der Information, welche Testformen in die Analysen eingehen (vier Varianten mit allen Testformen und vier Varianten mit den Formen a und b), mit der	

Itemanzahl, den Angaben darüber, welche Items verrechnet werden und der Bezeichnung der Modelle.....	82
Tabelle 17: Rasch-Analysen für das Modell A mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	83
Tabelle 18: Rasch-Analysen für das Modell a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	84
Tabelle 19: Rasch-Analysen für das Modell B mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	84
Tabelle 20: Rasch-Analysen für das Modell b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	85
Tabelle 21: z-Werte der Items 1 bis 15 (ohne Item 11) aus dem Modell b mit der Itemnummer (Itemnr.), den z-Werten und den Signifikanz-Angaben.....	86
Tabelle 22: Neuerliche Darstellung der Auswertungsmöglichkeiten mit dem Rasch-Modell für die einzelnen Testformen getrennt voneinander mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden sowie der Bezeichnung der Modelle.	87
Tabelle 23: Neuerliche Rasch-Analysen für das Modell A mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	88
Tabelle 24: Neuerliche Rasch-Analysen für das Modell a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	88
Tabelle 25: Neuerliche Rasch-Analysen für das Modell B mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	89
Tabelle 26: Neuerliche Rasch-Analysen für das Modell b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	90
Tabelle 27: Rasch-Analysen für das Modell C22 mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	90
Tabelle 28: Rasch-Analysen für das Modell C mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden	

(df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	91
Tabelle 29: Rasch-Analysen für das Modell c mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	92
Tabelle 30: Neuerliche Darstellung der Auswertungsmöglichkeiten mit dem Rasch-Modell für die Testformen gemeinsam mit der Information, welche Testformen in die Analysen eingehen (vier Varianten mit allen Testformen und vier Varianten mit den Formen a und b), mit der Itemanzahl, den Angaben darüber, welche Items verrechnet werden und der Bezeichnung der Modelle.....	93
Tabelle 31: Rasch-Analysen für das Modell 1a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	94
Tabelle 32: Rasch-Analysen für das Modell 1b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	94
Tabelle 33: Rasch-Analysen für das Modell 2a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	95
Tabelle 34: Rasch-Analysen für das Modell 2b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	96
Tabelle 35: Rasch-Analysen für das Modell 3a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	96
Tabelle 36: Rasch-Analysen für das Modell 3b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	97
Tabelle 37: Rasch-Analysen für das Modell 4a mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	98
Tabelle 38: Rasch-Analysen für das Modell 4b mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).....	99
Tabelle 39: Notwendige Informationen für die Berechnung der Likelihood-Quotienten-Tests mit den involvierten Testformen, der Itemanzahl, der Angabe, welche Items	

verrechnet wurden, der Bezeichnung der Modelle, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) sowie der Likelihood (ln L).	101
Tabelle 40: Rasch-Analysen für das Modell 2b2 mit der Anzahl der Testpersonen pro Teilungskriterium (n), der Likelihood in den Teilungsgruppen (ln L), den Freiheitsgraden (df), der Prüfgröße des LQTs (χ^2), dem kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie der Wahrscheinlichkeit (p).	102
Tabelle 41: Likelihood-Quotienten-Tests für die einzelnen Teilungskriterien der Modelle 1b und 2b2 mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen (ln L_1 , ln L_2), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQTs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie zur Wahrscheinlichkeit (p).	103
Tabelle 42: Likelihood-Quotienten-Tests in den nach dem Median geteilten Gruppen hoher und niedriger Score für die Modelle 1b und 2b2 mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen (ln L_1 , ln L_2), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQTs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie zur Wahrscheinlichkeit (p).	104
Tabelle 43: Likelihood-Quotienten-Tests für die einzelnen Teilungskriterien der Modelle 3b und 4b mit Angaben zur Stichprobengröße (n_1 , n_2), zur Likelihood in den Teilungsgruppen (ln L_1 , ln L_2), zur Anzahl der Freiheitsgrade in den Gruppen (df_1 , df_2), zur Prüfgröße des LQTs (χ^2), zur Differenz der Freiheitsgrade ($df_1 - df_2$), zum kritischen Fehlerwert für 1% ($\chi^2_{\text{krit}(1\%)}$) sowie zur Wahrscheinlichkeit (p).	105
Tabelle 44: Für die Berechnung der Likelihood-Quotienten-Tests notwendige Informationen aus den Rasch-Analysen der Testformen einzeln mit Angaben über Testform und Itemanzahl sowie darüber, welche Items verrechnet wurden, der Bezeichnung des Modells, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) und der Likelihood (ln L).	106
Tabelle 45: Für die Berechnung der Likelihood-Quotienten-Tests notwendige Informationen aus den Rasch-Analysen der Testformen gemeinsam mit Angaben über die involvierten Testformen, der Itemanzahl sowie darüber, welche Items verrechnet wurden, der Bezeichnung des Modells, der Stichprobengröße (n), der Anzahl der Freiheitsgrade (df) und der Likelihood (ln L).	106
Tabelle 46: Den Scores entsprechende WLE-Parameter für die Modelle 2b, 2b2 und 4b mit Angaben über die involvierten Testformen und die Stichprobengrößen (n).	110
Tabelle 47: Häufigkeiten der Scores und der WLE-Parameter in den Modellen 2b, 2b2 und 4b in Zahlen (n) und Prozent (%).	111
Tabelle 48: Mittelwerte (Mw WLE, Mw Score), Standardabweichungen (s WLE, s Score) und Mediane (Med WLE, Med Score) für die WLE-Parameter und Scores in den Modellen 2b, 2b2 und 4b samt Stichprobengröße (n).	111
Tabelle 49: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 2b in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤ 16 min, >16min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).	112

Tabelle 50: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 2b2 in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤ 16 min, >16min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).....	113
Tabelle 51: Prüfung auf Normalverteilung der WLE-Parameter für das Modell 4b in den Teilstichproben Geschlecht (m ... männlich, w ... weiblich), Alter in Jahren (14-16, >16), Unterricht in Darstellender Geometrie (DG ja, DG nein), Bearbeitungszeit (≤ 15 min, >15min) und Testform (a, b, c) samt jeweiliger Testpersonenanzahl (n) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung).....	113
Tabelle 52: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „männliches Geschlecht“ (m) und „weibliches Geschlecht“ (w) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.....	114
Tabelle 53: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „14 bis 16 Jahre“ (≤ 16) und „älter als 16 Jahre“ (>16) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.....	115
Tabelle 54: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „Besuch von Unterricht in Darstellender Geometrie“ (DG) und „kein Unterricht in Darstellender Geometrie“ (k.DG) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.....	115
Tabelle 55: Verteilung der Testpersonen aus den Teilstichproben DG-Unterricht (DG) und kein DG-Unterricht (kein DG) mit den beobachteten sowie erwarteten Stichprobengrößen, den χ^2 -Werten (χ^2), Freiheitsgraden (df) und der Wahrscheinlichkeit (p) für die Modelle 2b, 2b2 und 4b.	116
Tabelle 56: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „niedrige Bearbeitungszeit“ (≤ 16 min bzw. ≤ 15 min) und „hohe Bearbeitungszeit“ (>16min bzw. >15min) inklusive Testpersonenanzahl in den Teilstichproben (n) und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b, 2b2 und 4b.....	117
Tabelle 57: Verteilung der Frauen auf die Gruppen „niedrige“ und „hohe Bearbeitungszeit“ mit den beobachteten sowie erwarteten Stichprobengrößen, den χ^2 -Werten (χ^2), Freiheitsgraden (df) und der Wahrscheinlichkeit (p) für die Modelle 2b, 2b2 und 4b.	118
Tabelle 58: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) in den Teilstichproben „Form a“, „Form b“, „Form c“ und Gesamtpersonenanzahl (n_{ges}) für die Modelle 2b und 2b2.	118

Tabelle 59: Statistische Kennwerte der Varianzanalysen zwischen den Testformen mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und Gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q) für die Modelle 2b und 2b2.	119
Tabelle 60: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) in den Teilstichproben „Form a“ und „Form b“ und Gesamtpersonenanzahl (n _{ges}) für das Modell 4b.....	120
Tabelle 61: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Gruppen „männliches Geschlecht“ (m) versus „weibliches Geschlecht“ (w) mit den Altersgruppen „14 bis 16 Jahre“ (≤16J.) versus „älter als 16 Jahre“ (>16J.) und den Bearbeitungszeitgruppen „bis 16 Minuten“ (≤16min) bzw. „bis 15 Minuten“ (≤15min) versus „mehr als 16 Minuten“ (>16min) bzw. „mehr als 15 Minuten“ (>15min) inklusive Testpersonenanzahl pro Untergruppe (n) und Gesamtpersonenanzahl (n _{ges}) für die Modelle 2b und 4b.....	121
Tabelle 62: Ergebnisse der Varianzanalysen mit den Quadratsummen (QS) für die Wechselwirkungen „Geschlecht*Alter“, „Geschlecht*Zeit“, „Alter*Zeit“ und „Geschlecht*Alter*Zeit“ samt Freiheitsgraden (df), mittleren Quadraten (mittl. Q), F-Werten und den Wahrscheinlichkeiten (p) für die Modelle 2b und 4b.....	121
Tabelle 63: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) inklusive Testpersonenanzahl (n) für die Variablen „Geschlecht“ („männlich“, „weiblich“) und „Bearbeitungszeit“ („≤ 16min“, „>16min“) für das Modell 2b. ...	122
Tabelle 64: Häufigkeiten in Zahlen (n) und Prozent (%) für Modell 2b, mit denen die Items 1 bis 15 (ohne Items 5 und 11) mittels Rotations- und Perspektivwechselstrategie (Rot, PPW) bearbeitet wurden.	124
Tabelle 65: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der 3 Strategietypen inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%) für Modell 2b.....	125
Tabelle 66: Prüfung auf Normalverteilung der WLE-Parameter in den Teilstichproben Perspektivwechsler, Rotierer und Mischtyp samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p) für Modell 2b.....	125
Tabelle 67: Statistische Kennwerte der Varianzanalyse für die drei Strategietypen mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q) für Modell 2b.....	125
Tabelle 68: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „räumliches Vorstellungsvermögen niedrig ausgeprägt“ bis „räumliches Vorstellungsvermögen hoch ausgeprägt“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	127

Tabelle 69: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	127
Tabelle 70: Statistische Kennwerte der Varianzanalyse für die Aussage 1 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	128
Tabelle 71: Scheffé-Test zur Einschätzung des Raumvorstellungsvermögens mit den mittleren Differenzen (mittl. Diff.) und der Wahrscheinlichkeit (p).	128
Tabelle 72: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Einschätzung der Raumvorstellungsfähigkeit („eher niedrig ausgeprägt“, „eher hoch ausgeprägt“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 1.	129
Tabelle 73: Prüfung auf Normalverteilung der WLE-Parameter in den 2 Teilstichproben „Bearbeitung der Testaufgaben fiel eher leicht“ und „Bearbeitung der Aufgaben fiel eher schwer“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	130
Tabelle 74: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage 2 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	130
Tabelle 75: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogensausage 2 („Bearbeitung der Aufgaben fiel eher leicht“ versus „eher schwer“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 2.	131
Tabelle 76: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Bearbeitung der Testaufgaben war anstrengend“ bis „Bearbeitung der Testaufgaben war nicht anstrengend“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	132
Tabelle 77: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 3 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	132
Tabelle 78: Statistische Kennwerte der Rangvarianzanalysen mit den mittleren Rängen der WLE-Parameter für die 4 Antwortstufen der Aussage 3 samt Testpersonenanzahl (n).	132
Tabelle 79: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testmaterial hat mir gut gefallen“ bis „Testmaterial hat mir nicht gut gefallen“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	133

Tabelle 80: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 4 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	133
Tabelle 81: Statistische Kennwerte der Varianzanalyse für die Aussage 4 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	134
Tabelle 82: Prüfung auf Normalverteilung der WLE-Parameter in den 3 Teilstichproben „Orientierung an der ganzen Schleife“, „Orientierung an einem markanten Schleifenteil“ und „Markierung an mehreren markanten Schleifenteilen“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	134
Tabelle 83: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der 3-stufigen Antwortskala der Aussage 5 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	135
Tabelle 84: Statistische Kennwerte der Varianzanalyse für die Aussage 5 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	135
Tabelle 85: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogenaussage 5 („Orientierung an der ganzen Schleife“, „an einem Teil der Schleife“, „an mehreren Teilen der Schleife“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts, innerhalb der Aussage 5 sowie Gesamt innerhalb des Geschlechts.	136
Tabelle 86: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testantworten spontan gegeben“ bis „Testantworten nicht spontan gegeben“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	137
Tabelle 87: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 6 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	137
Tabelle 88: Statistische Kennwerte der Varianzanalyse für die Aussage 6 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	137
Tabelle 89: Statistische Kennwerte der Rangvarianzanalysen mit den mittleren Rängen der WLE-Parameter für die 4 Antwortstufen der Aussage 6 samt Testpersonenanzahl (n).	137
Tabelle 90: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogenaussage 6 („Testantworten eher spontan gegeben“ versus „Testantworten eher nicht spontan gegeben“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 6.	138

Tabelle 91: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testantworten genau überprüft“ bis „Testantworten nicht genau überprüft“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	139
Tabelle 92: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 7 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	139
Tabelle 93: Statistische Kennwerte der Varianzanalyse für die Aussage 6 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	139
Tabelle 94: Scheffé-Test zur Aussage 7 mit den mittleren Differenzen (mittl. Diff.) und der Wahrscheinlichkeit (p).	140
Tabelle 95: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogenaussage 7 („Testantworten eher genau überprüft“ versus „Testaufgaben eher nicht genau überprüft“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 7.	140
Tabelle 96: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Testaufgaben bildlich verdeutlicht – trifft zu“ bis „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	141
Tabelle 97: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 8 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	141
Tabelle 98: Statistische Kennwerte der Varianzanalyse für die Aussage 8 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	141
Tabelle 99: Prüfung auf Normalverteilung der WLE-Parameter in den 2 Teilstichproben „Testaufgaben sprachlich verdeutlicht – trifft zu“ und „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	142
Tabelle 100: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage 9 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	142
Tabelle 101: Verteilung zwischen Geschlecht (m ... männlich, w ... weiblich) und Fragebogenaussage 9 („Testantworten sprachlich verdeutlicht – trifft eher zu“ und „trifft eher nicht zu“) mit den beobachteten Teilstichprobengrößen („beobachtetes n“) sowie Prozentangaben (%) innerhalb des Geschlechts und innerhalb der Aussage 9.	143

Tabelle 102: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Ich hätte lieber in umgekehrter Richtung gearbeitet – trifft zu“ bis „trifft nicht zu“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	144
Tabelle 103: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage ab1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	144
Tabelle 104: Statistische Kennwerte der Varianzanalyse für die Aussage ab1 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	144
Tabelle 105: Prüfung auf Normalverteilung der WLE-Parameter in den 2 zusammengefassten Teilstichproben „Die freie Wahl der Bearbeitungsrichtung war eher angenehm“ versus „eher unangenehm“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).....	145
Tabelle 106: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage c1 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	145
Tabelle 107: Prüfung auf Normalverteilung der WLE-Parameter in den 2 zusammengefassten Teilstichproben „Die freie Wahl der Bearbeitungsrichtung fiel mir eher leicht“ versus „eher schwer“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).....	146
Tabelle 108: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den beiden Teilstichproben der Aussage c2 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	146
Tabelle 109: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst räumliches Vorstellungsvermögen in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	147
Tabelle 110: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 10 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).....	147
Tabelle 111: Statistische Kennwerte der Varianzanalyse für die Aussage 10 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	147
Tabelle 112: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst logisches, schlussfolgerndes Denken in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und	

Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	148
Tabelle 113: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 11 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).	148
Tabelle 114: Statistische Kennwerte der Varianzanalyse für die Aussage 11 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	149
Tabelle 115: Prüfung auf Normalverteilung der WLE-Parameter in den 4 Teilstichproben „Der Endlosschleifentest erfasst Konzentrationsfähigkeit in geringem Ausmaß“ bis „in hohem Ausmaß“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	149
Tabelle 116: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) der jeweiligen Antwortstufe (1 bis 4) der Aussage 12 inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).	150
Tabelle 117: Statistische Kennwerte der Varianzanalyse für die Aussage 12 mit den Quadratsummen (QS) zwischen den Gruppen, innerhalb der Gruppen und gesamt samt Freiheitsgraden (df) und den mittleren Quadraten (mittl. Q).	150
Tabelle 118: Häufigkeiten in Zahlen (n) und Prozent (%) für die 12 Tätigkeiten im Händigkeitsfragebogen.	152
Tabelle 119: Prüfung auf Normalverteilung der WLE-Parameter in den beiden Teilstichproben „Schreibhand rechts“ bzw. „Schreibhand links“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	153
Tabelle 120: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „Schreibhand rechts“ und „Schreibhand links“ inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).	153
Tabelle 121: Antwortmuster der 287 Testpersonen bei den fünf einhändigen Tätigkeiten mit den Häufigkeiten in Zahlen (n) und Prozent (%).	154
Tabelle 122: Informationen über Mittelwerte und Standardabweichungen der WLE-Parameter (Mw WLE, s WLE) sowie Mittelwerte und Standardabweichungen der Scores (Mw Sco, s Sco) in den Teilstichproben „linkshändig“, „rechtshändig“ und „beidhändig“ inklusive Testpersonenanzahl (n) und Häufigkeiten in Prozent (%).	155
Tabelle 123: Prüfung auf Normalverteilung der WLE-Parameter in den beiden Handpräferenzgruppen „rechtshändig“ bzw. „linkshändig“ samt Häufigkeiten in Zahlen (n) und Prozent (%) sowie Angaben über Schiefe, Standardabweichung der Schiefe (s Schiefe), Wölbung und Standardabweichung der Wölbung (s Wölbung) bzw. Ergebnisse des Kolmogorov-Smirnoff-Tests mit Angaben zur Prüfgröße (Kolm. Z) und der Wahrscheinlichkeit (p).	155

Abbildungsverzeichnis

Abbildung 1: <i>Photographische Endlosschleife</i>	7
Abbildung 2: Schematische Darstellung der Zwei-Faktoren-Theorie von Spearman.....	9
Abbildung 3: „Structure of Intellect“-Modell	14
Abbildung 4: Berliner Intelligenzstrukturmodell	15
Abbildung 5: Computergenerierte Endlosschleife	63
Abbildung 6: Beispielitem einer <i>computergenerierten Endlosschleife</i> in der Startansicht und der um 90 Grad rotierten Zielansicht.....	64
Abbildung 7: Antwortschema für eine Testaufgabe mit einer beispielhaft vorgenommenen Markierung. Der linke Teil spiegelt die Perspektivwechselstrategie, der rechte Teil die Rotationsstrategie wider. Das „?“ steht als zusätzliche Auswahlmöglichkeit mit dem sinngemäßen Inhalt „Ich weiß die Lösung nicht“ zur Verfügung.	65
Abbildung 8: Beispielaufgabe aus der zusätzlichen Instruktion in der Form c. Die zwei Pfeile zeigen an, dass die Aufgabe von links oben nach rechts unten oder von rechts unten nach links oben bearbeitet werden kann.....	66
Abbildung 9: Antwortschema für die Items 6 bis 15 in der Form c, wobei für jede Aufgabe zwei Markierungen vorzunehmen sind. Die Testpersonen müssen einerseits die Bearbeitungsrichtung und andererseits auf dem Perspektivwechsel- oder dem Rotationsschema die Lösung ankreuzen.	67
Abbildung 10: Verteilung des Geschlechts in der erhobenen Stichprobe	69
Abbildung 11: Altersverteilung in der erhobenen Stichprobe	70
Abbildung 12: Verteilung der Bearbeitungszeit in Minuten sowie „k. A.“ (keine Angabe) für die Personen mit ausgelassenen oder fehlerhaften Zeitangaben.	72
Abbildung 13: Item 11, wie es in der Form a vorgegeben wurde. Die Lösung lautet „von unten“ (Perspektivwechsel) bzw. „Drehung nach oben“ (Rotation).....	74
Abbildung 14: Verteilung des Scores über alle drei Testformen	76
Abbildung 15: Verteilung des Scores in den drei Testformen a, b und c	77
Abbildung 16: Item 5	86
Abbildung 17: Graphische Veranschaulichung der signifikanten Wechselwirkung zwischen Geschlecht und Bearbeitungszeit mit den Leistungen (WLE-Parameter) in den Teilstichproben „weiblich, niedrigere Bearbeitungszeit“, „weiblich, höhere Bearbeitungszeit“, „männlich, niedrigere Bearbeitungszeit“ sowie „männlich, höhere Bearbeitungszeit“ für das Modell 2b.....	123

Anhänge

Anhang 1: Spezifischer Fragebogen zum Test für die Testformen a und b

Kreuzen Sie an, was am besten auf Sie zutrifft.

Ich schätze mein räumliches Vorstellungsvermögen als

niedrig ausgeprägt ein ☐ ☐ ☐ ☐ hoch ausgeprägt ein.

Die Bearbeitung der Testaufgaben fiel mir

sehr leicht ☐ ☐ ☐ ☐ sehr schwer.

Die Bearbeitung der Testaufgaben war für mich

anstrengend ☐ ☐ ☐ ☐ nicht anstrengend.

Das Testmaterial hat mir

gut gefallen ☐ ☐ ☐ ☐ nicht gut gefallen.

Ich habe mich bei der Bearbeitung der *Endlosschleifen* im Allgemeinen

- ☐ an der ganzen Schleife orientiert.
- ☐ an einem einzelnen markanten Teil der Schleife orientiert.
- ☐ an mehreren markanten Teilen der Schleife orientiert.

Ich habe meine Testantworten im Allgemeinen

spontan gegeben ☐ ☐ ☐ ☐ nicht spontan gegeben.

Ich habe meine Testantworten im Allgemeinen

genau überprüft ☐ ☐ ☐ ☐ nicht genau überprüft.

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben bildlich (mittels „Bilder im Kopf“) zu verdeutlichen.

trifft zu ☐ ☐ ☐ ☐ trifft nicht zu

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben sprachlich (mittels verbaler Formulierungen) zu verdeutlichen.

trifft zu ☐ ☐ ☐ ☐ trifft nicht zu

Die Bearbeitungsrichtung im Test war fix vorgegeben – von links oben (Startansicht) nach rechts unten (Zielansicht). Bei manchen Aufgaben hätte ich lieber von rechts unten nach links oben (also in umgekehrter Richtung) gearbeitet.

trifft zu ☐ ☐ ☐ ☐ trifft nicht zu

Ich finde, dass der Test ...

... *räumliches Vorstellungsvermögen*

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

... *logisches, schlussfolgerndes Denken*

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

... *Konzentrationsfähigkeit*

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

Anhang 2: Spezifischer Fragebogen zum Test für die Testform c

Kreuzen Sie an, was am besten auf Sie zutrifft.

Ich schätze mein räumliches Vorstellungsvermögen als

niedrig ausgeprägt ein ☐ ☐ ☐ ☐ hoch ausgeprägt ein.

Die Bearbeitung der Testaufgaben fiel mir

Sehr leicht ☐ ☐ ☐ ☐ sehr schwer.

Die Bearbeitung der Testaufgaben war für mich

anstrengend ☐ ☐ ☐ ☐ nicht anstrengend.

Das Testmaterial hat mir

gut gefallen ☐ ☐ ☐ ☐ nicht gut gefallen.

Ich habe mich bei der Bearbeitung der *Endlosschleifen* im Allgemeinen

- ☐ an der ganzen Schleife orientiert.
- ☐ an einem einzelnen markanten Teil der Schleife orientiert.
- ☐ an mehreren markanten Teilen der Schleife orientiert.

Ich habe meine Testantworten im Allgemeinen

spontan gegeben ☐ ☐ ☐ ☐ nicht spontan gegeben.

Ich habe meine Testantworten im Allgemeinen

genau überprüft ☐ ☐ ☐ ☐ nicht genau überprüft.

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben **bildlich** (mittels „Bilder im Kopf“) zu verdeutlichen.

trifft zu ☐ ☐ ☐ ☐ trifft nicht zu

Als Lösungsstrategie habe ich im Allgemeinen versucht, mir die Testaufgaben **sprachlich** (mittels verbaler Formulierungen) zu verdeutlichen.

trifft zu ☐ ☐ ☐ ☐ trifft nicht zu

Die freie Wahl der Bearbeitungsrichtung war für mich

sehr angenehm ☐ ☐ ☐ ☐ sehr unangenehm.

Die freie Wahl der Bearbeitungsrichtung fiel mir im Vergleich zur fix vorgegebenen Bearbeitungsrichtung (Aufgaben 1 bis 5)

sehr leicht ☐ ☐ ☐ ☐ sehr schwer.

Ich finde, dass der Endlosschleifentest ...

... **räumliches Vorstellungsvermögen**

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

... **logisches, schlussfolgerndes Denken**

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

... **Konzentrationsfähigkeit**

in geringem Ausmaß erfasst ☐ ☐ ☐ ☐ in hohem Ausmaß erfasst.

Anhang 3: Händigkeitfragebogen

Bitte geben Sie im Folgenden an, welche Hand Sie gewöhnlich verwenden, um Tätigkeiten auszuführen. (Stellen Sie sich dabei die Bewegungsabläufe vor).

Welche Hand verwenden Sie gewöhnlich, wenn Sie ...

... einen Brief leserlich schreiben wollen?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... einen Ball werfen, um ein Ziel zu treffen?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... einen Tennis-, Squash- oder Badmintonschläger halten?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... ein Streichholz anzünden (welche Hand hält das Streichholz)?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... mit einer Schere schneiden?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... einen Faden durch ein Nadelöhr fädeln (welche Hand hält den Faden)?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... mit einem Besen aufkehren (welche Hand ist am Besenstiel weiter oben)?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... mit einer Schaufel Sand schaufeln (welche Hand ist weiter oben)?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... Spielkarten austeilen?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... mit einem Hammer einen Nagel einschlagen (welche Hand hält den Hammer)?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... eine Zahnbürste benutzen?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

... den Deckel eines Marmeladeglases aufschrauben?

☐ linke ☐ rechte ☐ wechselnd ☐ weiß nicht

Wurde Ihre Schreibhand (z.B. in der Schule) von links auf rechts umgelernt?

☐ ja ☐ nein

Mussten Sie aus anderen Gründen (z.B. Verletzung) irgendwelche Tätigkeiten von einer auf die andere Hand umlernen?

☐ ja ☐ nein

Wenn Sie für alle diese 12 Tätigkeiten die rechte Hand benutzen: Gibt es irgendwelche einhändigen Tätigkeiten, für die Sie die linke Hand benutzen?

☐ ja ☐ nein

Wenn ja, welche: _____

Anhang 4: Abstract Deutsch

Die vorliegende Arbeit befasst sich mit der Frage, ob die Änderung der Bearbeitungsrichtung bei Raumvorstellungsaufgaben zu unterschiedlichen Itemschwierigkeiten führt. Die Idee entstand aus Daten mit photographischen Endlosschleifen, bei denen eine Umdrehung des Start- und Zielbildes zum Teil erhebliche Unterschiede in den Itemparametern mit sich brachte. Da als mögliche Ursache ungünstige Eigenschaften der photographischen Endlosschleifen vermutet werden, kommen in der vorliegenden Studie 15 computergenerierte Endlosschleifenaufgaben in drei Testversionen, die über fünf Link-Items verbunden sind, zum Einsatz. Nach der Eliminierung eines auffälligen Items werden die Daten der drei Testversionen getrennt voneinander sowie gemeinsam mit vollständigen und unvollständigen Designs mit dem dichotomen logistischen Modell von Rasch analysiert ($n=289$ Testpersonen, 16 verschiedene Modelle). Nach dem Ausscheiden eines weiteren Items und Durchführung der Modelltests niedriger versus hoher Score, niedriger versus hoher Median, niedriges versus hohes Alter, Unterricht versus kein Unterricht in Darstellender Geometrie sowie niedrige versus hohe Bearbeitungszeit kann Rasch-Homogenität für alle berechneten Modelle angenommen werden. Die Itemparameter bei computergenerierten Endlosschleifen unterscheiden sich nicht nach Vertauschung des Start- und Zielbildes. Klassisch-testtheoretische Auswertungen ergeben signifikante Leistungsunterschiede zugunsten der Männer, zugunsten Testpersonen mit einem höheren Alter (> 16 Jahre) sowie Personen mit einer höheren Bearbeitungszeit. Keine signifikanten Unterschiede treten in Abhängigkeit der Testform, der Bearbeitungsstrategie sowie Darstellendem Geometrieunterricht auf. Ein Fragebogen zum Testmaterial sowie Testverhalten zeigt signifikant bessere Leistungen bei Personen, die ihre räumlichen Fähigkeiten hoch einschätzen, ihre Antworten überprüfen und nicht spontan antworten. Die Auswertungen eines Händigkeitsfragebogens ergeben keine signifikanten Leistungsunterschiede in Abhängigkeit der Schreibhand sowie einer globalen Handpräferenz.

Anhang 5: Abstract Englisch

In the present thesis, the hypothesis is investigated whether or not a change in the processing direction of spatial tasks influences their difficulty. The idea behind this hypothesis roots in data obtained from photographic endless loops, where the inversion of the start and end images caused partly dramatic changes in the item parameters. Since adverse properties of the photographic endless loops are presumed as a reason for this observation, the present study uses 15 tasks of computer-generated endless loops in three different test versions, which are connected via five link-items. After elimination of one ambiguous item, the data from the three test versions are analyzed both independently as well as together (complete and incomplete designs) with the Rasch model for dichotomous data ($n=289$ probands, 16 different models). Rasch-homogeneity can be assumed for all calculated models after testing of high versus low score and median, high versus low age, education versus no education in descriptive geometry, short versus long processing time, and after another problematic item had been eliminated from the test. The item parameters of computer-generated endless loops do not differ upon inversion of the start and end images. The evaluation with methods of classical test theory results in significantly better performance of male probands, older probands (> 16 years) as well as probands with longer processing times. No significant differences crop up as a function of the test version, processing strategy or an educational background in descriptive geometry. A questionnaire concerning the test material as well as the test behavior proves significantly better performance of probands who estimate their spatial abilities as high, who recheck their results, and do not answer spontaneously. The evaluation of a handedness questionnaire shows no significant differences in performance as a function of the probands' writing hand or their general hand preference.

Lebenslauf

Mag.rer.nat. Veronika Steinhauser (geb. Reich-Rohrwig)

Geb. 16. März 1979 in Wien

Verheiratet, 2 Töchter (geb. 2009, 2011)

BERUFLICHER WERDEGANG

Ab Oktober 2003	Angestellte bzw. freie Mitarbeiterin Niederösterreichisches Hilfswerk, Korneuburg
März 2003 - Jän. 2006	Studienassistentin Universität Wien, Fakultät für Psychologie Institut für Klinische, Biologische und Differentielle Psychologie (Ao. Univ.-Prof. Dr. Georg Gittler)
März 2000 - Jän. 2003	Tutorin für die Lehrveranstaltung Intelligenz- und Persönlichkeits-theorien I und II

STUDIUM UND AUSBILDUNG

Mai 2006	Eintragung in die Liste der Klinischen und GesundheitspsychologInnen
Sept. 2005 - Okt. 2006	Ausbildung zur Legasthenietherapeutin Qualitätszirkel Legasthenie
April 2005 - März 2006	Ausbildung zur Klinischen und Gesundheitspsychologin Berufsverband Österreichischer PsychologInnen
Seit März 2003	Doktoratsstudium der Naturwissenschaften (Psychologie) Universität Wien
Okt. 2002 - Juni 2004	Ausbildung zur Mediatorin Institut für Mediation und Konfliktmanagement
Okt. 1997 - Jän. 2003	Diplomstudium Psychologie Universität Wien

BERUFSRELEVANTE PRAKTIKA

Sept. 2002	OMV AG , Personalabteilung
Jän. 2001 - Mai 2001	AKH Wien , Abteilung für Transplantationschirurgie
Aug. 1999 - Sept. 1999	Heilpädagogisches Zentrum Hinterbrühl
Aug. 1999 - Sept. 1998	Haus der Barmherzigkeit , 1180 Wien

SCHULDBILDUNG

Sept. 1989 - Juni 1997	BG XIX, Gymnasiumstraße , 1190 Wien Reifeprüfung Juni 1997
Sept. 1985 - Juni 1989	Volksschule Scheibenbergstraße , 1180 Wien