



universität
wien

Diplomarbeit

Titel der Diplomarbeit

3D-2D Objektschneidetest: Neukonstruktion eines
Rasch skalierten Spatial Visualisationtests.

verfasst von

Thomas Halmetschlager

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2015

Studienkennzahl lt. Studienblatt: A298

Studienrichtung lt. Studienblatt: Psychologie

Betreut von: Ao. Univ.-Prof. Dr. Georg Gittler

Inhaltsverzeichnis

Einleitung	5
I. THEORETISCHER TEIL	7
1. Räumliches Vorstellungsvermögen	7
2. Unterschied zwischen Mental Rotation und Spatial Visualisation	13
3. Klassifizierung von bisherigen Raumvorstellungstests.....	15
4. Test „Schnitte“ von Fay & Quaiser-Pohl (1999).....	22
5. Testtheoretische Prinzipien des 3D-2D Objektschneidetests	25
6. Das dichotom logistische Modell von Rasch (nach Koller et al., 2012).....	26
6.1 Kriterien und Vorteile des Rasch Modells	26
7. Geschlechtsunterschiede bei Raumvorstellungsleistungen	29
8. Alterseffekte bei Raumvorstellungsleistungen	31
9. Fragestellung und Hypothesen.....	33
II. METHODE.....	35
10. Testkonstruktion – 3D-2D Objektschneidetest	35
10.1 Instruktion.....	37
10.2 Übungsbeispiele	40
10.3 Testitems.....	43
11. Untersuchungsablauf	45
12. Stichprobendarstellung	46
13. Analyseplan	47
III. DARSTELLUNG UND ERGEBNISSE	50
14. Hypothesenprüfung	50
14.1 Hypothese 1 - Überprüfung auf Gültigkeit des Rasch Modells.....	50
14.2 Hypothese 2 - Geschlechtsunterschiede	63
14.3 Hypothese 3 - Altersunterschiede	64
14.4 Hypothese 4 – Ausbildungszweige	66
14.5 Hypothese 5 – Höchste abgeschlossene Ausbildung.....	69
14.6 Hypothese 6 - Selbsteinschätzungsskala	70

15. Diskussion, Ausblick, Limitationen	71
Zusammenfassung	78
Abstract	79
Literaturverzeichnis.....	80
Abbildungsverzeichnis.....	85
Tabellenverzeichnis.....	87

Einleitung

Raumvorstellung ist ein Gebiet, das seit Jahrzehnten im Rahmen der Intelligenzforschung erforscht wurde. Es gibt eine Vielzahl an Testverfahren, die darauf abzielen Raumvorstellung zu messen. Diese Verfahren basieren Großteils auf der Klassischen Testtheorie und es existieren nur wenige Tests, die Rasch homogene Itemsets aufweisen. Die Subfaktoren der Raumvorstellung (Mental Rotation, Spatial Visualisation und Spatial Perception) befinden sich auf unterschiedlichen Forschungsständen. So hat Gittler (1990) bereits mit dem 3DW (Dreidimensionaler Würfeltest) ein Rasch homogenes Instrument für Spatial Perception entwickelt. Für Mental Rotation liegt ebenfalls mit dem EST (Endlosschleifentest) ein Rasch homogener Test von Gittler & Arendasy (2003) vor. Für Spatial Visualisation fehlen noch Instrumente, die die Qualitäten des Rasch Modells (Eindimensionalität, Stichprobenunabhängigkeit, usw.) aufweisen. Hiervon ausgehend wurde der Subfaktor Spatial Visualisation für diese Arbeit gewählt und ein Rasch homogener Test erstellt. Da für Spatial Visualisation noch keine Rasch homogenen Tests im deutschsprachigen Raum publiziert wurden, soll ein aktueller Test, basierend auf der Klassischen Testtheorie, als Ausgangspunkt herangezogen werden. Der neu konstruierte 3D-2D Objektschneidetest soll dem Format 3D-Stimulusmaterial und 2D-Lösungsmaterial entsprechen.

Der 3D-2D Objektschneidetest soll die Forschung im Raumvorstellungsbereich vorantreiben und das Rasch Modell in diesem Bereich weiter verankern. Zusätzlich soll auch auf bisherige Forschungsarbeiten aus dem Bereich Spatial Visualisation Bezug genommen und mit dem neuen Test repliziert werden. Darunter zählen die Studien zu Geschlechtsdifferenzen (Linn & Petersen, 1985; Harris, Hirsh-Pasek, & Newcombe, 2013) und die Studien zu Alterseffekten und Leistungsdifferenz auf Grund deren beruflichen Ausbildung (Salthouse, Babcock, Skovronek, Mitchell, & Palmon, 1990).

I. THEORETISCHER TEIL

1. Räumliches Vorstellungsvermögen

Die Beginne der Raumvorstellungsforschung starteten im frühen 20. Jahrhundert mit der Intelligenzforschung. Genauer gesagt mit den Forschungen zur zwei-faktoriellen Theorie des „g“ Faktors und die Existenz von untergeordneten „s“ Faktoren von Spearman (1904). Aufbauend auf seiner Korrelationsforschung postuliert Spearman eine mittlere Korrelation zwischen den Tests zur Messung der Intelligenz. Darauf lässt sich schlussfolgernd erklären, dass die gemeinsame Varianz in den Tests Aufschluss auf einen gemeinsamen Faktor der Intelligenz gibt, den er als *g-Faktor* bezeichnete. Der *s-Faktor* betitelt die spezifischen Unterkomponenten der Intelligenz, die durch die unterschiedlichen Facetten der Untertests erhoben werden. Diese Untertests umfassten zum Beispiel mathematische Aufgaben und sprachliche und räumliche Problemstellungen. Spearmans Studien waren der Grundbaustein für Folgeforschungen in Bezug auf Raumvorstellung, die von Thurstone (1931) als eine der sieben Primärfaktoren der Intelligenz identifiziert wurde. Auch El Koussy (1935) nahm den Begriff der Raumvorstellung auf und beschrieb ihn als einen integrierten Faktor der Intelligenz.

El Koussy erstellte 1935 eine Testbatterie zur Messung der Intelligenz und testete diese an 162 Jungen zwischen 11 und 13 Jahren. Er stützte sich dabei auf die Tetraden Technik, die Spearman in seinen Studien zur Erfassung des g-Faktors verwendete. El Koussy hoffte Gruppenfaktoren der Raumvorstellung zu finden, indem er den g-Faktor herauspartialisierte. Er identifizierte einen Gruppenfaktor, den er „k“-Faktor nannte. Präsentierte wurde der Faktor als „the ability to obtain and the facility to utilize visual, spatial imagery“ (El Koussy, 1935, S. 65)

Ein ähnlicher Ansatz vertrat den Standpunkt, dass die Leistung in einem Test nicht nur von einem Überfaktor ausgeht, sondern von verschiedenen Unterfaktoren abhängt. Bei der Überprüfung dieser These von Thurstone im Jahr 1938 stieß er durch die Vorgabe von 56 Tests an 218 Collegestudenten auf die Unterfaktoren, die er *Primary Abilities* nannte. Durch die multifaktorielle Faktorenanalyse extrahierte er 13 Faktoren. Nach der Rotation der Achsen

war es ihm möglich 9 Faktoren auf Grund der Ladungen auf den Tests zu benennen. Einen dieser Faktoren benannte er *space*. Diesen beschrieb er wie folgt:

The space factor is found in tests which require that the subject manipulate an object imaginatively in two or three dimensions. It is quite different from perceptual processes which require only the perception of detail in a flat surface, and which do not require the imaginal movement in two or three dimensions. The psychological question remains whether the spatial factor is somehow related to kinesthetic imagery. (Thurstone, 1931, zit. n. Eliot & Smith, 1983, S. 3)

Drei Jahre später führten Thurstone & Thurstone (1941) erneut eine Studie durch, um die Primal Abilities der Intelligenz zu replizieren und stießen nicht wie zuvor auf 9 Faktoren, sondern auf 7 Faktoren, worunter auch der Faktor „space“ zu finden war.

Weitere Studien zur Raumvorstellung erfolgten in den USA während des zweiten Weltkriegs, als dringender Bedarf für Flugpersonal bestand. Ausgehend von diesem Bedarf waren Selektionsmethoden notwendig. Dies wurde über Studien mittels Screeningmethoden durch Wahrnehmung-, Raumvorstellungs- und mechanische Tests erreicht, in denen Tausende von Fluganwärtern untersucht wurden. Hieraus ergab sich für die Forschung eine Gelegenheit eine breite Stichprobe für die Ausarbeitung der Subfaktoren zu erreichen.

Die Forschungen zu diesem Zeitpunkt resultierten in die Unterteilung der Raumvorstellung in die zwei Unterfaktoren *Spatial Relation* und *Visualisation*. *Spatial Relation* wurde von Guilford & Lacey (1947) definiert als „(1) ability to perceive visual-spatial arrangements, (2) ability to organize movements in spatially-determined order, or (3) ability to relate specific spatial locus or arrangement within the stimulus pattern“ (Guilford & Lacey, 1947, S. 479). *Visualisation* wurde definiert als „ability to imagine the rotation of depicted objects, the folding and unfolding of flatpatterns, and the relative changes of position of objects in space“ (Guilford & Lacey, 1947, S. 271)

Thurstone (1949) widmete sich der Exploration der Subfaktoren der Raumvorstellung. Er führte eine Studie an 350 Highschool Jungen mit 32 Gruppen und 25 Einzeltests durch. Er wollte dadurch Unterschiede in den Fähigkeiten aufzeigen, in denen man sich stationäre bzw. im Raum bewegte Objekte in Gedanken vorstellt. Dabei wurden zwei Faktoren gefunden. Der *Space Factor S1*, der auf den *Cards and Figure Tests* lädt, wurde definiert als „the ability to visualize a rigid configuration when it is moved into different positions“. (Thurstone & Thurstone, 1949, S. 13) Der *Space Factor S2*, der auf Tests aus dem Bereich *Surface Development* und *Paper Puzzle* lädt wurde definiert als „the ability to visualize a configuration in which there is movement or displacement among the parts of the configuration“. (Thurstone, 1949, S. 13) Die Unterschiede zwischen den beiden Unterfaktoren der Raumvorstellung S1 und S2 beziehen sich auf die statische oder bewegliche Eigenschaft der vorgestellten Objekte.

French widmete sich 1951 dem Thema der Unterfaktoren der Raumvorstellung. Er postulierte, dass es die Subfaktoren „Spatial“, „Orientation“ und „Visualisation“ gibt. Spatial wurde als die Fähigkeit beschrieben, Muster visuell aufzunehmen und mit anderen zu vergleichen. Orientation sah er als die Fähigkeit an, durch unterschiedliche Präsentation von visuellen Mustern nicht irritiert zu werden. Visualisation wurde als die Fähigkeit gesehen, Objekte visuell zu erfassen und imaginär manipulieren zu können. (French, 1951, zit. n. Eliot & Smith, 1983, S. 4)

Die wichtigste Unterscheidung der Unterfaktoren der Raumvorstellung zu dieser Zeit kann durch die Präsentation von rigiden und nicht-rigiden Material vorgenommen werden, wobei sich Visualisation auf nicht-rigides Material stützt. Von einer klaren Abgrenzung ohne Interkorrelationen der beiden Unterfaktoren kann aufgrund verschiedener Stichproben und Verzerrungen in der Testdurchführung nicht ausgegangen werden.

Abschließend sollen noch zwei Autorentams mit ihren Studien vorgestellt werden, die einen großen Einfluss auf die aktuelle Unterscheidung der Subfaktoren und deshalb auch auf diese Arbeit haben.

Linn & Petersen (1985) haben sich in ihrer Metaanalyse der Frage der Geschlechtsunterschiede im Raumvorstellungsbereich gewidmet. Im Rahmen dieser Studie befassen sie sich auch mit dem Thema der Subfaktoren der Raumvorstellung. Dazu kategorisierten sie Raumvorstellung in die drei Bereiche *Spatial Perception*, *Mental Rotation* und *Spatial Visualisation*. Diese Faktoren Spatial Perception und Spatial Visualisation sind ident mit denen von French (1951), wobei Orientation und Perception zusammengefasst werden.

Unter Spatial Perception wird die Fähigkeit verstanden, visuelle Relationen trotz ablenkenden Informationen zu erfassen. Beispiele hierfür sind der *Rod-Frame Test* von Witkin (1950), bei dem der Testteilnehmer einen Stab in eine senkrechte Position zum Boden bringen muss, während er sich in einer schiefen Position befindet, und der *Waterlevel-Test* von Piaget und Inhelder (1967), bei dem die Testteilnehmer Bilder von halb befüllten Wasserflaschen gezeigt bekommen und jene identifizieren müssen, bei denen der Wasserspiegel horizontal ist.

Mental Rotation bezeichnet die Fähigkeit zwei- oder dreidimensionale Objekte zu rotieren. Beispiel hierfür sind die modifizierte Version des *Mental Rotation Test (MRT)* (Shepard & Metzler, 2004) und der *Card Rotation Test* (French, Ekstrom, & Price, 1963). Bei diesem Format wird als Ausgangspunkt ein Körper vorgegeben. Im Lösungsmaterial befindet sich mehrere Antwortalternativen mit dem gleichen Körper und gespiegelten Körpern in unterschiedlicher Lage. Aufgrund von vorgestellter Rotation müssen die richtigen (nicht gespiegelten) Körper gefunden werden.

Spatial Visualisation wird als Fähigkeit bezeichnet, visuelles Material in mehrstufigen komplexen Abläufen zu verarbeiten und zu modifizieren. Prozesse die in Spatial Perception und in Mental Rotation notwendig sind, können als Lösungsschritte in Spatial Visualisation Aufgaben angewandt werden. Beispiele hier sind der *Embedded Figure Test* (Witkin, Oltman, Raskin, & Karp, 1971) und der *Paper Folding Test* (French et al., 1963). Beim Embedded Figure Test müssen einfache geometrische Figuren aus komplexen Mustern erkannt werden. Der Paper Folding Test basiert darauf, dass ein Stück Papier mehrmals gefaltet und dann ein Loch durchgestanzt wird. Es werden mehrere Möglichkeiten präsentiert, wie das entfaltete Stück

Papier aussieht. Der Teilnehmer muss das Papier in Gedanken entfalten um die Aufgabe zu lösen.

Eine neuere Perspektive die die Subfaktoren von Linn & Peterson, French und Thurstone vereinen soll ist die von Maier (1999). In seinem Modell werden die Subfaktoren nicht zusammengefasst, wodurch sich Interkorrelationen ergeben. Die 5 Subfaktoren benennt er als:

- Veranschaulichung oder Räumliche Visualisierung (Spatial Visualisation)
- Räumliche Beziehungen (Spatial Relation)
- Räumliche Orientierung (Spatial Orientation) und Faktor K¹
- dRäumliche Wahrnehmung (Spatial Perception)
- Vorstellungsfähigkeit von Rotationen (Mental Rotation)

Maier skizzierte ein Modell, in dem die Faktoren aufgrund der Position der Person und der Verwendung von dynamischen und statischen Denkvorgängen unterschieden werden können. Die Zuteilung zu dynamischen und statischen Denkvorgängen weist hohe Ähnlichkeit mit der Unterteilung in rigide und nicht-rigide Figuren von Thurstone (1949) auf. Das Modell ist in Abbildung 1 zusammengefasst.

Die hellgrauen Flächen beschreiben Faktoren in denen sich die Person nicht als Zentrum bzw. Ausgangssituation sieht, sondern die Aufgabe von „außen“ löst. Der Fokus liegt auf der Manipulation des Testmaterials aus einer Blickrichtung. Der Paper Folding Test und Water-Level Test fallen in diesen Bereich.

Bei den dunkelgrauen Flächen, sieht sich die Testperson als Zentrum und orientiert sich aufgrund der Position der eigenen Person im Raum. Dies ist zum Beispiel bei Tests der Fall, bei denen sich Personen auf einer Landkarte oder imaginären Stadt orientieren müssen. Bei dem Subfaktor Mental Rotation können beide Methoden verwendet werden. So könnte sich eine

¹ Der von Maier aufgenommene Faktor K ist hierbei nicht derselbe, wie ihn El Koussy in seinen Studien identifizierte. Es handelt sich um den Faktor *Kinesthetic Imagery* von Michael, Guilford, Fruchter & Zimmerman (1957). Kinesthetic imagery definiert rechts – links Unterscheidungen in Bezug auf die Position des eigenen Körpers und wird deshalb von Maier unter Räumliche Orientierung genannt.

Testperson ein Objekt in Gedanken vorstellen und dieses Objekt rotieren, ohne die Position des Blicks zu ändern oder das Objekt starr in Gedanken lassen und sich imaginär um das Objekt bewegen.

Standpunkt der Person	Art der Denkvorgänge	
	Dynamisch Bewegung von Objekten und Veränderung räumlicher Relation	Statisch Räumliche Relationen am Objekt bleiben unveränderlich
Außen (Zentrierung auf das Objekt)	Spatial Visualisation (<i>Paper Folding Test, Schnitte</i>)	Spatial Relation (<i>Rod-Frame Test</i>)
Innerhalb (Position der Blickrichtung)	Mental Rotation* (<i>Mental Rotation Test, EST</i>)	Spatial Perception* (<i>Embedded Figure Test, 3DW</i>)
	Spatial Orientation* (<i>Three Mountain Task, PST</i>)	
		Faktor K

Abbildung 1. Die Komponenten des räumlichen Denkens nach Maier (1999); Bei Subfaktoren, die mit einem * gekennzeichnet sind, existieren bereits Rasch homogen Verfahren.

Durch die klare Abgrenzung der Subfaktoren wird Maiers Ansatz als zentral für diese Arbeit angesehen. Das Modell fasst die Unterfaktoren nicht zusammen, wodurch eine Zuordnung von Verfahren zu den Subfaktoren erleichtert wird. Der EST von (Gittler & Arendasy, 2003) und der 3DW von Gittler (1990) sind Rasch homogene Verfahren zur Messung von Mental Rotationen bzw. Spatial Perception. Der Rasch homogene Pilot's Spatial Test (PST) von Grössenbrunner (1994) misst einerseits Mental Rotation aber auch Spatial Orientation. Für die restlichen beiden Unterfaktoren existieren im deutschen Forschungsbereich noch keine Rasch homogenen Verfahren. Durch die starke Vertretung von Spatial Visualisation in einer Vielzahl von Forschungsberichen in den letzten 100 Jahren soll in dieser Arbeit ein Rasch homogener Spatial Visualisation Test konstruiert werden.

2. Unterschied zwischen Mental Rotation und Spatial Visualisation

Wie oben in der Kategorisierung des Testmaterials und der Identifizierung der Subfaktoren der Raumvorstellung gezeigt wurde, haben Mental Rotation und Spatial Visualisation Gemeinsamkeiten und sollen aufgrund ihrer oberflächlichen Ähnlichkeit hier näher analysiert werden. Beide Subfaktoren können anhand der Dimensionen intrinsisch und dynamisch beschrieben werden und zusätzlich aufgrund des verwendeten Stimulus und Lösungsmaterials (2D oder 3D) differenziert werden.

Da zur Darstellung von Spatial Visualisation der Paper Folding Test bereits exemplarisch dargestellt wurde, soll zur besseren Erklärung eine MRT-Version zur Erfassung von Mental Rotation (Abbildung 2) dargestellt werden, um eine gute Abgrenzen der beiden Bereiche herzustellen. Beim MRT geht es darum, aus vorgegebenen Objekten, diejenigen herauszufinden, die nicht ident mit dem Stimulusitem (in Abbildung 2, links) sind.

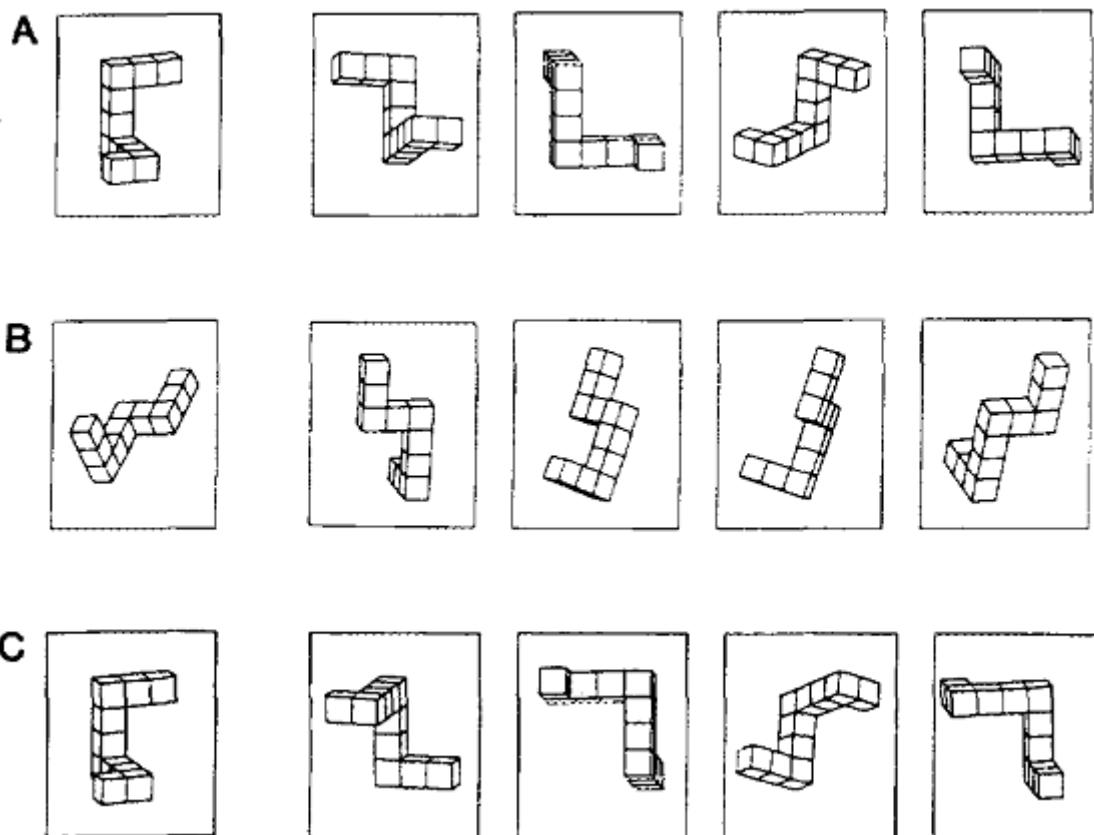


Abbildung 2. Mental Rotation Test von Vandenberg & Kuse (1978). Auszug aus Peters et al. (1995)

In diesem Zusammenhang findet auch der Terminus *Mental Folding*, dass eine Art von Spatial Visualisation darstellt, seinen Platz. Harris et al. nennen grundlegende Unterschiede in der Bearbeitung von Mental Rotation und Mental Folding (bzw. Spatial Visualisation) Aufgaben, wie zum Beispiel, dass es sich bei Mental Rotation um eine rigide Form der Bearbeitung handelt, während dies bei Mental Folding nicht der Fall ist. Schon Thurstone (1949) erkannte dies und schrieb in seinen Arbeiten, dass man Raumvorstellung in zwei große Bereiche trennen könne. In *rigide* und *nicht rigide* Transformationen.

Ein zweiter großer Unterschied findet sich darin, dass bei Mental Rotation hauptsächlich räumliche Fähigkeiten zur Lösung Verwendung finden, während bei Mental Folding auch analytische Fähigkeiten angewandt werden. Bei Mental Rotation findet aufgrund der einmaligen Rotation keine Verwendung des verbal-analytischen Lösungsweges statt, während dies bei Mental Folding, aufgrund der mehrstufigen Transformation der Fall ist. (Lohman, 1979)

Davon ausgehend darf nicht angenommen werden, dass bei Mental Rotation nur ein Lösungsweg verwendet werden kann, auch hier gibt es mehrere Möglichkeiten diese Art von Aufgaben zu lösen. Neben den bereits genannten Unterschieden zwischen den beiden Raumvorstellungsdimensionen weisen sie dennoch eine hohe Gemeinsamkeit auf. So können zur Lösung von Spatial Visualisation Aufgaben neben Transformation auch Komponenten von Rotationen verwendet werden.

Harris et al. (2013) verglichen Mental Rotation und Mental Folding mit den Leistungen von Personen die *STEM*-Kurse (Science, Technology, Engineering und Mathematics) absolvierten, um Gemeinsamkeiten und Unterschiede zwischen Mental Rotation und Mental Folding (Spatial Visualisation) aufzudecken. Die Ergebnisse ihrer Analyse zeigen, dass die Fähigkeit zu Mental Rotation vor allem bei Personen höher waren, die Kurse in Mathematik und Chemie besuchten, während Mental Folding bei Personen besser ausgebildet war, die Fächer Engineering, Chemie und Geologie besuchten. Bei den Biologiestudenten unterscheidet sich die Leistung zwischen Mental Rotation und Mental Folding nicht.

Unterschiede können auch im Geschlechtervergleich gefunden werden, wie bereits im Abschnitt zu Geschlechtsunterschiede global für den Faktor Raumvorstellung beschrieben wurde. Das lässt darauf schließen, dass die beiden Konstrukte nicht exakt dasselbe messen, sondern Unterschiede in der Verarbeitung und Lösung von Aufgaben im Bereich Mental Rotation und Spatial Visualisation vorherrschen. Dies lässt darauf schließen, dass Mental Rotation und Mental Folding (Spatial Visualisation) nicht ausschließlich als separate Prozesse angesehen werden können. Beide Operationen verwenden interne Repräsentationen von Objekten (Harris et al., 2013). Diese Aussage ist übereinstimmend mit dem Modell von Uttal et al. (2013) der bereits die Einteilung nach intrinsischen und extrinsischen Dimensionen vornahm.

3. Klassifizierung von bisherigen Raumvorstellungstests

Da der Fokus dieser Arbeit eine Neukonstruktion eines Tests im Bereich Spatial Visualisation war, stellt eine Abgrenzung zu den anderen Bereichen einen wichtigen Punkt dar. Um Überschneidungen zu vermeiden, sollen hier wichtige Tests der bereits beschriebenen Subfaktoren vorgestellt werden.

Uttal et al. (2013) nehmen dazu eine ähnliche Perspektive wie Maier (1999) ein. Maier teilte die Subfaktoren aufgrund der eigenen Position im Raum und den dynamischen bzw. statischen Denkvorgängen ein. Uttal et al. bauten auf diesem Ansatz auf und unterteilten die bereits vorhandenen Tests zur Raumvorstellung aufgrund von zwei Dimensionen: *Dynamic – Static* und *Intrinsic – Extrinsic*.

Intrinsic – Extrinsic:

Als intrinsisch werden Operationen kategorisiert die eine Definierung von Objekten beschreibt. Darunter fallen die Identifizierung und Unterscheidung von Objekten. Intrinsische Aufgaben benötigen außerdem nur die Betrachtung des Objekts per se und nicht die Wahrnehmung der Umgebung des Objekts. Als extrinsisch werden Aktivitäten genannt, die visuelle Relationen zwischen Objekten oder Gruppen von Objekten beinhalten. Extrinsisch kann deshalb auch als *‘between-object’* bezeichnet werden.

Static – Dynamic:

Aufgaben in denen das Material stationär (eine inhaltsgetreue Übersetzung des Begriffs 'static') vorgegeben werden, werden als Static bezeichnet. Dabei verändert das Material weder ihre Orientierung, Platz oder Dimension (zweidimensional oder dreidimensional). Als dynamisch werden Stimuli bezeichnet, die bewegt werden, sei es physikalisch oder mental. So gibt es Aufgaben in denen man entweder das Objekt oder die eigene Person bewegen muss und dynamische Interaktionen zur Lösung voraussetzt. Eine Zusammenfassung der Dimensionen findet sich in Tabelle 1.

Tabelle 1

Kombination der Dimensionen Static – Dynamic und Intrinsic - Extrinsic nach Uttal et al. (2013)

Räumliche Fähigkeitsdimension	Beschreibung	Beispiele zur Messung
Intrinsic & Static	Objekte, Wege oder räumliche Beziehungen erkennen	• Embedded Figure Task
Intrinsic & Dynamic	Zusammenfügung von Objekten zu komplexeren Figuren, Visualisierung und Transformation von 2D zu 3D Objekten und vice versa, Rotieren von Objekten	• Form Board • Block Design • Paper Folding • Mental Rotations Test • Cube Comparison • Card Rotation Test
Extrinsic & Static	Verstehen von abstrakten räumlichen Gesetzmäßigkeiten, wie horizontale und vertikale Beständigkeit	• Water-Level Test • Rod and Frame Test
Extrinsic & Dynamic	Visualisierung der Umgebung von unterschiedlichen Ausgangspunkten	• Piaget's Three Mountains Task

Klassifikation: Intrinsic - Static

Dieser Bereich beinhaltet Verfahren, in denen die Vorgänge zur Lösung der Aufgabe hauptsächlich intrinsisch vor sich gehen und das Testmaterial zur Lösung nicht transformiert werden muss. Wahrnehmung von Relationen oder Konfigurationen von visuellen Objekten sind ein Hauptbestandteil dieses Bereichs. Ein Beispiel befindet sich in Abbildung 3. Das Stimulusmaterial besteht aus geometrischen zweidimensionalen Figuren und komplexen Zeichnungen, die aus vielen verschiedenen Mustern und Figuren besteht. Ziel ist es, die oben dargestellten Figuren in den unteren komplexen Zeichnungen wiederzuerkennen.

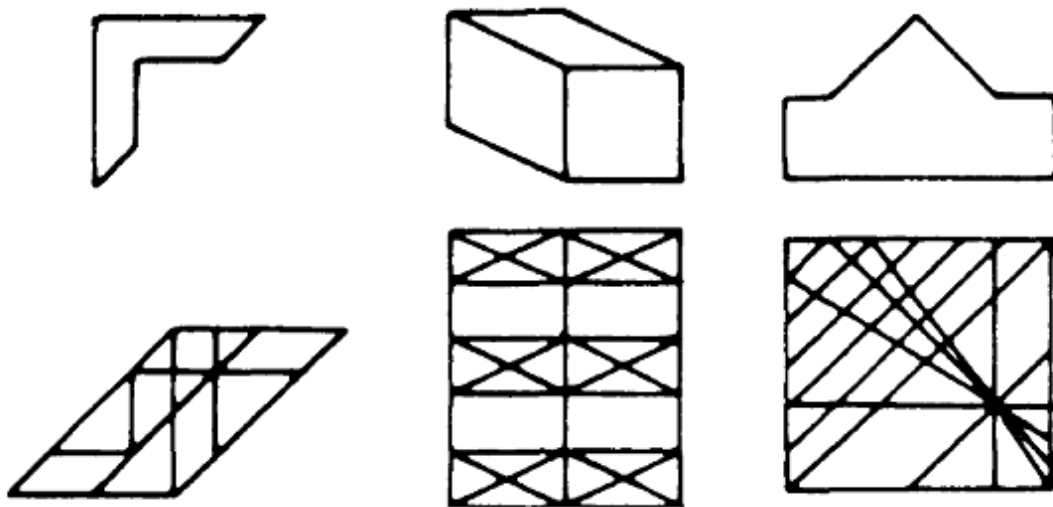


Abbildung 3. Embedded Figures Task nach Witkin, Oltman, Raskin, & Karp (1971)

Klassifikation: Intrinsic - Dynamic

Die Vorgänge zur Lösung dieser Art von Aufgaben erfolgen intrinsisch und die Verarbeitung des Testmaterials dynamisch. Dies bedeutet, dass das Testmaterial gedanklich manipuliert wird. Hierunter zählen modifizieren, schneiden und zusammensetzen von Objekten. Die Modifikation von Objekten kann hier auch die Veränderung von zweidimensionalen zu

dreidimensionalen Objekten und vice versa bedeuten. Durch die Relevanz für die Erstellung eines Tests im Bereich Spatial Visualisation mit dynamischen Inhalt sollen hier Tests für zwei- und dreidimensionale Objekte näher dargestellt werden. Die Beschreibung der Tests folgt hierbei dem Schema, dass die erste Dimension die Beschreibung des Stimulusmaterials angibt und die zweite Dimension, die Beschreibung des Lösungsmaterials. Ein Test nach dem Schema 2D-3D würde also aus zweidimensionalen Stimuli und dreidimensionalen Lösungen bestehen.

Aufgaben nach dem Schema 2D-2D:

Das Stimulus-, als auch das Lösungsmaterial sind zweidimensional. Ein Beispiel hierfür ist der Test *Formboard*. Die Aufgabe besteht darin, die karierten geometrischen Figuren so zu rotieren und zusammenzufügen, dass die weiße Fläche ausgefüllt wird. Ein Beispiel befindet sich in Abbildung 4.

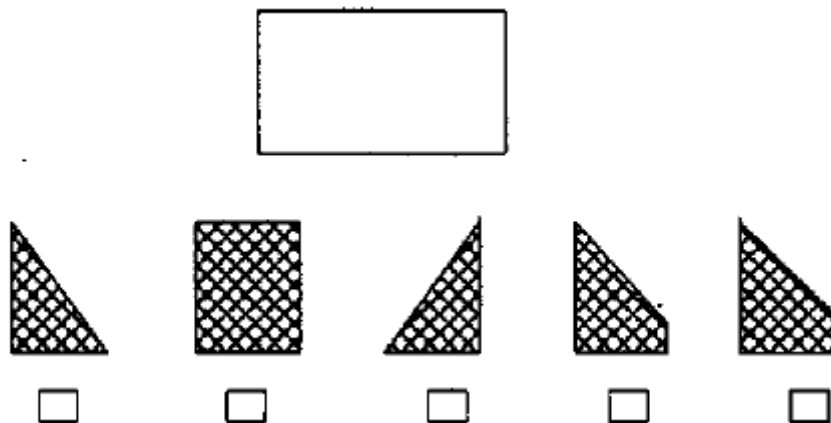


Abbildung 4. Kit of Reference Tests for Cognitive Factors – Formboard Vz-1 nach French et al. (1963)

Aufgaben nach dem Schema 2D-3D:

Die Ausgangssituation ist ein zweidimensionales Objekt, zur Lösung wird dieses Objekt aber in eine dreidimensionale Figur transformiert. Ein Beispiel hierfür ist der *Surface Development Test* in Abbildung 5. Hier müssen die Zahlen auf den Kanten des zweidimensionalen Objekts den Buchstaben auf den Kanten des dreidimensionalen Körpers zugeteilt werden. Zur Lösung gelangt man durch die Faltung der Zeichnung in Gedanken.

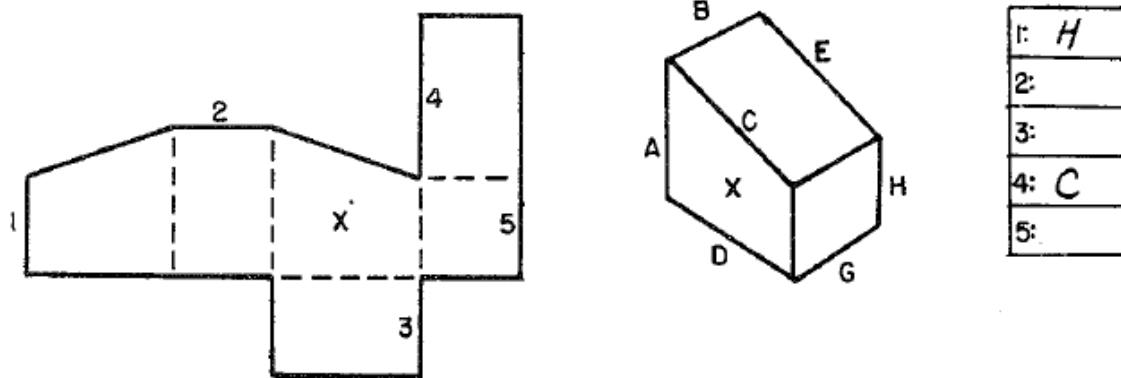


Abbildung 5. Surface Development Test; Auszug aus Manual for Kit of Factor-Referenced Cognitive Tests nach Ekstrom, French, Harman, & Dermen (1976)

Aufgaben nach dem Schema 3D-2D:

Die Ausgangssituation ist ein dreidimensionaler Körper. Um zur Lösung zu gelangen, muss dieser Körper in eine zweidimensionale Fläche transformiert werden. Hierzu zählen der bekanntere Paper Folding Test als auch der *Test „Schnitte“*. Erstgenannter soll hier dargestellt werden und Zweitgenannter wird als eigener Abschnitt weiter unter behandelt, da er ausschlaggebend für die Konstruktion des neuen Test im Rahmen dieser Arbeit war. Im Paper Folding Test ist die Ausgangssituation ein gefaltetes Stück Papier (die Darstellung wirkt zwar zweidimensional, doch aufgrund der Faltungen handelt es sich um ein dreidimensionales Objekt) und der Lösungsstimuli ist zweidimensional. Ein Stück Papier wird in Schritten gefaltet und danach ein Loch durch das Papier gestanzt. Der Lösungsweg besteht darin, das gestanzte Papier zu entfalten und zu erkennen, wo sich die gestanzten Löcher befinden, wenn das Papier komplett entfaltet ist. Ein Beispiel befindet sich in Abbildung 6.

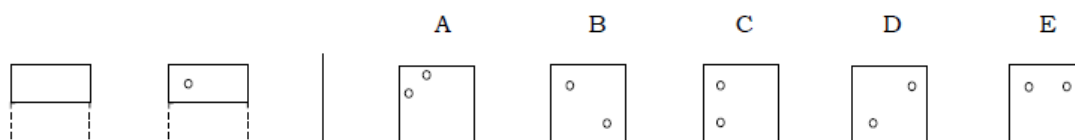


Abbildung 6. Kit of Reference Tests for Cognitive Factors – Paperfolding: Vz-2 nach French et al. (1963)

Aufgaben nach dem Schema 3D – 3D:

Ausgangs- wie auch Lösungsstimuli sind in dreidimensionalen Körpern dargestellt. Ein Beispiel hierfür ist der *Cube Comparison Test* in Abbildung 7. Hier werden Paare von Würfeln dargestellt und durch Rotation eines der Würfel soll darauf geschlossen werden, ob es sich um idente (s – same) oder verschiedene Würfel (d – different) handelt. Als Hinweise, die bei der Rotation der Würfel Aufschlüsse darauf geben sollen, ob es sich um gleiche Würfel handelt, dienen dabei Buchstaben.



Abbildung 7. Kit of Reference Tests for Cognitive Factors – Cube Comparison Test S-2 nach French et al. (1963)

Klassifikation: Extrinsic - Static

Das Testmaterial beruht auf dem Vergleich von visuellen Objekten. Das Verständnis von abstrakten visuellen Gesetzmäßigkeiten und der Vergleich von visuellen Hinweisen werden zur Bearbeitung solcher Aufgaben benötigt. Das Material ist in dieser Kategorie statischer Natur. Der *Water-Level Test* in Abbildung 8 ist ein Beispiel für diese Klassifikation. Hier wird eine Flasche mit Wasser befüllt, verschlossen und mit einem Tuch bedeckt. Danach wird die Flasche rotiert und die Testperson soll den Wasserspiegel der Flasche im Lösungsmaterial markieren.

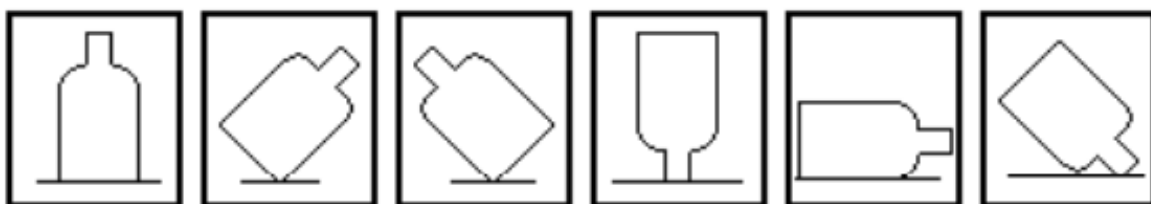


Abbildung 8. Water Level-Task nach Piaget & Inhelder (1967)

Klassifikation: Extrinsic - Dynamic

Auch hier erfolgt die Verarbeitung des Stimulusmaterials durch den Vergleich von visuellen Objekten - es handelt sich jedoch um ein dynamisches Vorgehen. Die dynamische Komponente wird definiert durch den Positionswechsel der eigenen Person oder des Materials um einen extrinsischen Vergleich zwischen den Elementen zu erreichen. Bei dem Three Mountain Task in Abbildung 9 wird ein dreidimensionales Modell von drei Bergen vorgegeben. Die Testperson muss sich aufgrund des Stimulusmaterials (Abbildung 9 – obere Grafik) festlegen auf welcher Position A, B, C oder D sie sich im Lösungsmaterial befindet.

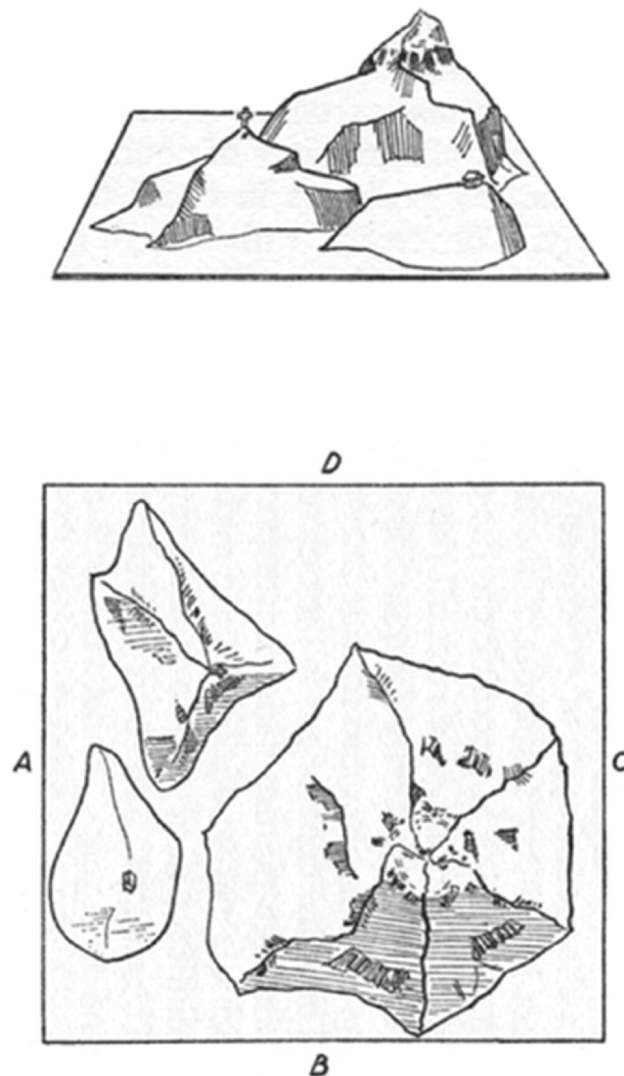


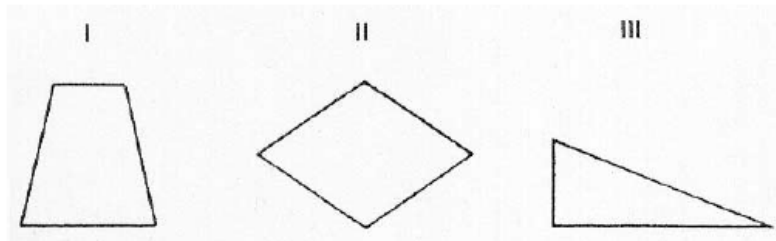
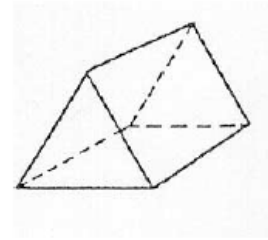
Abbildung 9. Piaget's Mountain Task nach (Piaget & Inhelder, 1967)

Die Zuteilung der Testverfahren zu den Dimensionen nach Uttal et al. (2013) zeigt, dass bereits eine Fülle von Tests in der klassischen Testtheorie existieren. Es herrscht aber eine Lücke in der Klassifikation Intrinsic – Dynamic in denen Tests zu Spatial Visualisation und Mental Rotation ihren Platz finden. Der EST und PST können hier als einziges Rasch homogenes Verfahren zur Messung von Mental Rotation im deutschen Bereich angesehen werden. Für Spatial Visualisation ist noch kein Rasch homogenes Verfahren erstellt worden. Da Spatial Visualisation die Transformation von Objekten in den Vordergrund stellt und eine Abgrenzung zu Mental Rotation (die sich in derselben Klassifizierung intrinsic – dynamisch befinden) geschehen soll, wurde das Schema 3D-2D gewählt.

4. Test „Schnitte“ von Fay & Quaiser-Pohl (1999)

Der Test „Schnitte“ von Fay & Quaiser-Pohl (1999) ist ein Raumvorstellungstest, der in die Klassifizierung: Intrinsic – Dynamic eingeordnet werden kann. Der Test behauptet von sich selbst den Faktor Spatial Visualisation als Ganzes zu erheben, indem er alternierend dreidimensionales und zweidimensionales Stimulus- als auch Lösungsmaterial verwendet. Dies erfolgt durch die Vorgabe eines Körpers der mit einem geraden Schnitt durchtrennt wird. Aufgrund dieses Schnitts entsteht eine neue Schnittfläche auf dem Körper. Diese Fläche kann aufgrund des Winkels und der Position des Schnitts alternieren. Exemplarisch soll ein derartiges Testitem in Abbildung 10 dargestellt werden. Der Körper der Grafik rechts oben wird durch einen Schnitt durchtrennt. Der Proband muss herausfinden, welche Fläche aufgrund eines Schnitts durch den Körper entstehen kann.

A prism, consisting of three rectangles and two Equilateral triangles at the ends (as seen on the right) is going to be cut by a plane. Which of the following cut surfaces can result?



- (A) only I and II
- (B) only II and III
- (C) only I and III
- (D) I, II, and III
- (E) none of those cut surfaces can emerge

Abbildung 10. Item aus dem Test „Schnitte“ (Quaiser-Pohl, 2003)

Der Test besteht zu dem nicht nur aus visuellen Materialien, sondern basiert in manchen Testitems (sowohl Lösungs- als auch Stimulusmaterial) auf reiner Textbasis. Hierzu muss die Aufgabe ohne visuelle Hinweise erfolgen, was zur Folge hat, dass die mentale Fähigkeit des Probanden ausschlaggebend ist. Dies soll das Problem beheben, dass Personen die Aufgaben aufgrund von Reasoning lösen. (Quaiser-Pohl, 2003)

Die Konstruktion des Tests „Schnitte“ besteht hierbei aus 3 Teilen, die Fay & Quaiser-Pohl (1999) als Elemente

- K - Teilmenge der zu schneidenden Körper (Kugeln, Quader, Pyramide, Zylinder),
- O - Menge der Operationen (Schnitt) und
- P - Menge der Produkte (Quader, Kreise, Dreiecke)

beschreiben.

Ausgangspunkt (Stimulusitem) der Itemkonstruktion können entweder die Elemente Körper oder Produkte oder eine Kombination von beiden sein - die Operation ist immer der Schnitt. Das Lösungsmaterial ist ausgehend von der Wahl des Stimulusitems ein Element der Körper oder der Produkte.

Der Test „Schnitte“ besteht aus einer Vielfalt an Testmaterialien, um die Facetten der Raumvorstellung abzudecken. Die Variation an verschiedenem Ausgangs- und Lösungsmaterial, erschwert jedoch die Berechnung der Objektivität. Manche Testpersonen verstehen die Instruktion nicht im ersten Durchgang und eine Interaktion zwischen Versuchsleiter und Testperson ist hier nicht zu umgehen. Deshalb ist die Durchführungsobjektivität im Test Schnitte auch nicht berechnet worden. Auswertungs- und Interpretationsobjektivität wurden durch die Verwendung von standardisierten Schablonen als gegeben angenommen. Die Split-Half-Reliabilität beträgt .81 und die zeitliche Stabilität nach ca. 15 Wochen $r_{tt} = .72$, womit zufriedenstellende Kennwerte ermittelt wurden. Bei der Validierung stützten sich Fay & Quaiser-Pohl auf Kommentare von Testteilnehmern. Leistungsstarke und leistungsschwache Personen sollten mittels der Methode des lauten Denkens Aufschluss geben, was dieser Test misst. Als Hinweis auf konvergente Validität finden sich Zusammenhänge mit dem Mechanik-Test ($r = .53$), Figuren-Denk-Test ($r = .41$) und dem Rechen-Test ($r = -.40$) aus dem Inventar zur Erfassung des mathematisch-technisch-naturwissenschaftlichen Bereichs. Die diskriminative Validität wurde aufgrund des Zusammenhangs mit Sprachfähigkeitstests und Wahrnehmungsfähigkeit berechnet. Mittels Skala zur Selbsteinschätzung der eigenen Raumvorstellungsfähigkeit wurde die ökologische Validität erhoben. Hier ergab sich ein positiver signifikanter Zusammenhang zwischen der Leistung und der Selbsteinschätzung (Fay & Quaiser-Pohl, 1999)

Der Test „Schnitte“ legt einen guten Grundbaustein für die aktuelle Arbeit dar. Die Testkennwerte sind zufriedenstellend und der Test scheint auch Spatial Visualisation zu messen. Bei der Neukonstruktion eines Rasch homogenen Tests sollen jedoch einige Veränderungen unternommen werden. So soll ein standardisierter Ablauf des kompletten Tests möglich sein, um somit die Durchführung unabhängig vom Testleiter zu ermöglichen. Eine genaue und einfache Instruktion ist ausschlaggebend dafür. Auf Aufgaben mit reiner Textbasis soll verzichtet werden, da sprachliche Defizite in die Bearbeitung der Items einfließen könnten und somit die Eindimensionalität gefährdet wäre.

5. Testtheoretische Prinzipien des 3D-2D Objektschneidetests

Der 3D-2D Objektschneidetest soll auf 4 Bereichen aufgebaut sein. *Instruktion, Erhebung der soziodemografischen Daten, Übungsbeispiele* und *Testitems*. Die Zeitdauer soll dabei aus Gründen der Ökonomie maximal 40 Minuten betragen, wobei zur Bearbeitung des Tests kein Zeitlimit vorgegeben werden soll, um einen reinen Powertest zu erhalten. Insgesamt wird der Test aus 19 Items bestehen, davon sind 4 Übungsbeispiele, die zum Verständnis vorgegeben werden und nicht in den Rohscore eingerechnet werden. Die Schwierigkeit des Tests soll so angelegt werden, dass eine Normalverteilung des Rohscores in der Normalpopulation erreicht wird.

Das Antwortniveau war eine wichtige Überlegung. In den Arbeiten zum Paper Folding Test wird das Antwortformat 1 aus 5 verwendet. Das ist eine Ratewahrscheinlichkeit von 20 %. Da bei einer Überprüfung mittels Rasch Modells die Ratewahrscheinlichkeit sehr klein gehalten werden muss, war die Wahl eines Antwortformats, ausschlaggebend. Der Gedanke war, dass nicht nur eine Antwortalternative richtig sein soll, sondern mehrere. Dies würde verhindern, dass Testpersonen beim sofortigen Erkennen der richtigen Antwortalternative, die anderen Alternativen nicht mehr in Betracht ziehen. Zur Auswahl standen die Formate 2 aus 5, 3 aus 5, eine Kombination aus 2 oder 3 aus 5 bzw. n aus 5. Die Ratewahrscheinlichkeit bei 2 aus 5 und 3 aus 5, beträgt rund 10 %, bei der Kombination der beiden 5 % und bei n aus 5 ist die Wahrscheinlichkeit 3.125 %. Jedoch wurde entschieden, dass nie alle Antwortmöglichkeiten richtig sein können. Da es sich bei dem Test um eine Neukonstruktion handelt, wurde n aus 5 gewählt, um einerseits eine sehr geringe Ratewahrscheinlichkeit zu erhalten, als auch den Test nicht aufgrund eines zu einfachen Antwortformats zu verfälschen. Das Testmaterial wurde am PC vorgegeben, die Eintragung der Antworten erfolgte aber mittels Pen and Paper.

Die geplante Stichprobe soll sich aus 200 Personen zusammensetzen, wobei Heterogenität angestrebt wird. Eine heterogene Stichprobe ist bei der Erstüberprüfung auf Geltung des Rasch Modells notwendig, wie es auch Koller, Alexandrowicz, & Hatzinger (2012) in ihrem Buch „Das Rasch Modell in der Praxis“ beschreiben. Deshalb werden nicht nur Studenten aus ausgewählten Studienrichtungen getestet, sondern auch berufstätige Personen zwischen 16 und 65 Jahren.

6. Das dichotom logistische Modell von Rasch (nach Koller et al., 2012)

Zur testtheoretischen Überprüfung der konstruierten Items soll im Folgenden auf das gewählte Modell eingegangen werden, das zur Überprüfung des Tests angewendet worden ist. Hierbei handelt es sich um das dichotome logistische Modell von Rasch.

Bei der Konstruktion eines Tests geht es darum, ein Verfahren zu erstellen, das es ermöglicht nicht direkt erfassbare Konstrukte zu erheben. Dies hat seinen Ursprung darin, dass es nicht möglich ist Raumvorstellung, in diesem Fall Spatial Visualisation, direkt zu erfassen, wie zum Beispiel das Körpergewicht über eine Waage. Deshalb muss theoriegeleitet die Konstruktion eines Tests erfolgen, mittels dessen vorhandenen Items es möglich ist, diese Konstrukte zu erfassen. Diese Items dienen als manifeste Indikatoren, die Aufschluss auf die zugrundeliegenden latenten Konstrukte (Spatial Visualisation) geben.

Die Ausprägung des latenten Konstrukts hängt von der Höhe der Personeneigenschaft, bzw. Personenfähigkeitsparameter θ ab. Die Schwierigkeit eines Items wird Itemschwierigkeitsparameter bzw. umgepolt Itemleichtigkeitsparameter genannt. Sind Personen- und Itemschwierigkeitsparameter bekannt, können weiterführend Rückschlüsse auf das Antwortverhalten bei den vorgelegten Items gezogen werden und der Test als Ganzes analysiert werden. So können über das Rasch Modell Wahrscheinlichkeiten beschrieben werden in denen eine Person v das Item i löst, wobei diese Wahrscheinlichkeit von der Fähigkeit der Person θ und von der Itemschwierigkeit β abhängt.

6.1 Kriterien und Vorteile des Rasch Modells

Wie bei der Messung von manifesten Konstrukten (zB. bei Körpergewicht eine Waage geeicht werden muss, um immer wieder exakte Werte zu liefern) muss auch ein Test mit manifesten Indikatoren (den Items) „geecht“ werden. Dies wird durch die Beantwortung der Items durch die Zielpopulation erreicht, indem Rückschlüsse auf die Testitems geleistet werden können. Ein Test mit hoher Qualität muss deshalb nicht nur aus Items bestehen, die theoriegeleitet

aufgebaut sind sondern auch den Qualitätskriterien des Rasch Modells entsprechen, die eine Art Eichung des Test übernehmen. Die Qualitäten lassen sich auflisten als:

- Eindimensionalität
- Spezifische Objektivität/Stichprobenunabhängigkeit
- Streng monoton steigende Itemcharakteristikkurven
- Trennschärfe
- Lokale stochastische Unabhängigkeit
- Suffizienz

Dichotome Items

Dichotome Items sind binärer oder zweikategorieller Natur. Es handelt sich hierbei entweder um die Ausprägung 0 = falsch oder 1 = richtig oder in einem Einstellungstest 0 = stimme nicht zu oder 1 = stimme zu. Die folgenden Qualitätskriterien lassen sich besonders gut mit dichotomen Items überprüfen.

Eindimensionalität

Wie oben beschrieben sollen die Items als manifeste Indikatoren dienen, um die latente Fähigkeit von Spatial Visualisation zu erheben. Dies kann aber nur geschehen, wenn die Items auch wirklich nur eine Fähigkeitsdimension erheben und nicht mit zusätzlichen Fähigkeiten konfundiert sind. Außerdem müssen die Items in den Subgruppen gleich schwierig zu lösen sein (Differential Item Functioning). Liegt kein DIF vor, so funktionieren alle Items in den verschiedenen Subgruppen gleich und der Test stellt ein faires Instrument dar.

Spezifische Objektivität/Stichprobenunabhängigkeit

Bei einem Rasch konformen Modell ist es irrelevant, welche Items verwendet werden, um Personen zu vergleichen. Eine Methode zur Überprüfung der spezifischen Objektivität auf Seite der Personen ist die Subgruppeninvarianz. Personen mit derselben Fähigkeit in den unterschiedlichen Subgruppen sollten den gleichen Personenscore erlangen. Eine Verletzung

liegt vor, wenn Personen aufgrund ihrer Zuteilung zu den Subgruppen verschiedene Scores erlangen.

Streng monoton steigende Item Characteristic Curve (ICC)

Gilt das Rasch Modell und sind alle Item homogen, dann kann auch davon ausgegangen werden, dass alle ICC streng monoton steigend sind. Mit steigendem Fähigkeitsparameter steigt auch die Wahrscheinlichkeit das Item zu lösen. Weisen Personenfähigkeit und Itemschwierigkeit einen gleich hohen Wert auf, so liegt man bei einer Lösungswahrscheinlichkeit von 50 %.

Trennschärfe

Eine zusätzliche Qualität in der Überprüfung des Verfahrens ist die Trennschärfe der einzelnen Items. Ein trennscharfes Item kann zwischen fähigen und weniger fähigen Personen gut diskriminieren.

Lokale stochastische Unabhängigkeit

Ob eine Aufgabe gelöst wird oder nicht, soll nur von der Fähigkeit der Testperson und der Schwierigkeit der Aufgabe abhängen und nicht davon welche Aufgaben bisher gelöst worden sind. Unabhängige Items stellen eine Voraussetzung für die Gültigkeit des Rasch Modells dar.

Suffizienz

Sind alle oben beschriebenen Eigenschaften zugetroffen, so kann davon ausgegangen werden, dass das Rasch Modell gilt und eine suffiziente Statistik darstellt. Werden alle Eigenschaften erfüllt, dann ist es irrelevant, welche Person welches Items gelöst hat. Alle Informationen sind dann in den Randsummen (richtige Antworten aller Personen in einem Item, bzw. Rohscore für jede Person) enthalten.

7. Geschlechtsunterschiede bei Raumvorstellungsleistungen

In den Jahren von 1960 bis 1980 waren differentialpsychologische Studien ein zentrales Forschungsthema. Ein Vertreter, der sich der geschlechtsspezifischen Forschung widmete war Werdelin (1961). In seiner Studie gab er 23 Tests an 453 High School Studenten vor. Dabei fand er in den Gruppen der Jungen unter den 6 faktorenanalytisch gewonnenen Faktoren zwei *Spatial Factors*, die Ähnlichkeiten mit den Faktoren von Thurstone aufwiesen. Den Faktor *S1* beschrieb Werdelin als Fähigkeit eine Gestalt als Ganzes zu erkennen und zu manipulieren und den Faktor *S2* als Fähigkeit eine Gestalt in verschiedene Teile zu zerteilen, zu manipulieren und wieder zusammenzufügen.

In der Gruppe der Mädchen fand er nur den Spatial Factor *S1*. Während keine Geschlechtsunterschiede in der Fähigkeit zur Manipulation von Gestalten (*S1*) zu finden waren, konnte Werdelin nach Analyse der Testitems und der Ergebnisse schlussfolgern, dass Geschlechtsunterschiede in der Fähigkeit bestehen, visuelle Strukturen zu verstehen, zu organisieren und reorganisieren (*S2*).

Neben den besseren Leistungen der Jungen in dem Faktor *S2*, lassen sich auch Geschlechtsunterschiede in anderen Unterfaktoren finden. In Mental Rotation und Spatial Perception zeigen Männer eine bessere Leistung als Frauen. In Spatial Visualisation sind die Unterschiede zugunsten der Männer sehr gering bis zu nicht existent. Geschlechtsunterschiede scheinen vor allem dann aufzutreten, wenn effiziente Lösungen aufgrund schneller Rotation notwendig sind, wie es beim Mental Rotation Test (MRT) der Fall ist. In Spatial Visualisation Aufgaben wird eine effiziente Lösungsmöglichkeit gesucht, in denen die Manipulation von Objekten eine Rolle spielt. Aufgrund der fehlenden Geschwindigkeitskomponente scheinen keine Geschlechtsunterschiede auf. (Linn & Petersen 1985)

Die Geschlechtsunterschiede lassen sich auch aufgrund Anzahl der zur Verwendung stehenden Lösungswege erklären. Es gibt nur wenig unterschiedliche Lösungswege mit denen Mental Rotation und Spatial Perception Aufgaben gelöst werden können. Bei Spatial Visualisation gibt es eine Vielzahl von Lösungswegen. So ist es möglich den MRT mittels

analytischem² Lösungsweg zu lösen, anstatt das Stimulusmaterial zu rotieren. Personen, die einen für sich persönlich weniger effektiven Lösungsweg verwenden, erzielen folgendermaßen einen niedrigeren Score, als Personen, die einen effektiveren Lösungsweg verwenden. (Linn & Peterson, 1985) Personen, die nur eine Strategie zur Lösung von Aufgaben im Bereich Mental Rotationen verwenden, weichen von dieser nachträglich auch nicht mehr ab. So verhält es sich auch in Spatial Perception. Vor allem Frauen sollen sich als Erstes für ineffektivere Lösungswege entscheiden und nicht mehr davon abweichen, wodurch die genannten Geschlechtsdifferenzen entstehen. (Shepard & Cooper, 1986)

Aufgrund der Komplexität der Aufgaben im Subfaktor Spatial Visualisation entscheiden sich Personen zu Beginn des Tests nicht sofort für einen Lösungsweg, sondern eruieren auch andere Lösungswege und verwenden nach der Erprobung den für sie persönlich Effektivsten. Da mehrere Lösungswege ausprobiert werden und sich Frauen und Männer für ihren persönlich favorisierten Lösungsweg entscheiden, entstehen hier keine Geschlechtsunterschiede. Zusätzlich postulieren die Autoren, dass Metastrategien zur Lösung von Spatial Visualisation Aufgaben verwendet werden, wohingegen bei Mental Rotation und Spatial Perception spezifische Strategien angewandt werden und deshalb Geschlechtsunterschiede entstehen.

Neben der Wahl der Strategie können auch gesellschaftliche Aspekte die Leistung von Frauen verschlechtern. So wird Frauen eine schlechtere Raumvorstellungsfähigkeit zugeschrieben. Studien zeigen, dass dieses Vorurteil die Leistung von Frauen beeinflussen kann. Wird Frauen vor der Bearbeitung eines Mental Rotation Tests mitgeteilt, dass Frauen eine schlechtere Leistung erzielen als Männer, so spiegelt sich dies auch in der tatsächlich erbrachten Leistung im Test wieder. Wird Frauen im Vorhinein allerdings die Information gegeben, dass Frauen im Allgemeinen eine bessere Raumvorstellungsfähigkeit aufweisen als Männer, sind keine Geschlechtsunterschiede zu finden. (Moè, 2009)

Einfluss auf die Leistung der Raumvorstellungsfähigkeit haben auch Aktivitäten, mit denen man sich täglich beschäftigt. Ist Raumvorstellung in diesen Aktivitäten integriert, so scheint es

² Der analytische Lösungsweg beinhaltet Strategien, die schlussfolgerndes Denken voraussetzen, wodurch nicht mehr Rotation zur Lösungsfindung verwendet wird.

nicht verwunderlich, dass deshalb auch die Leistung in Raumvorstellungstest bei diesen Personen besser ist. Da Frauen und Männer verschiedenen Tätigkeiten nachgehen, kann hier eine Ursache für Geschlechtsunterschiede gesehen werden. So weisen Frauen und Männer, die in ihrer Kindheit mit raumvorstellungsnahem Spielzeug (Murmeln, Knetmasse und Videospiele) gespielt haben, bessere Leistungen als Kinder, die mit raumvorstellungsfernen Spielzeug (Karten, Malen, Basteln) gespielt haben. Da Jungen öfter mit raumvorstellungsnahem Spielzeug interagieren, können so auch höhere Leistungen im späteren Alter erklärt werden. Neben der frühkindlichen Erfahrung mit Raumvorstellung beeinflussen auch die Freizeitaktivitäten die Raumvorstellungsleistung. So zeigen sich bei Personen, die raumvorstellungsnahen Sportrichtungen (Bowling, Fußball, Lacrosse, Broomball, Himmel und Hölle) nachgehen, bessere Raumvorstellungsfähigkeiten, als bei Personen, die raumvorstellungsferne Sportrichtungen (Eiskunstlauf, Gymnastik, Fahrradfahren, Reiten, Kampfsport, Laufen, Wrestling) betreiben. Da Frauen vor allem in den raumvorstellungsfernen Sportrichtungen zu finden sind, zeigen sie auch geringere Leistungen in der Raumvorstellung. Die Effekte wurden im MRT und im Water Level Test nachgewiesen, wobei im MRT stärkere Unterschiede zu Gunsten der Männer gefunden wurden. (Voyer, Nolan, & Voyer, 2000)

Neben dem biologischen Geschlecht wurde auch die sexuelle Orientierung mit der Raumvorstellungsleistung in Verbindung gebracht. Hier ergab sich bei heterosexuellen Männer eine signifikant höhere Leistung im MRT als bei homosexuellen Männer. Bei Frauen hingegen wurden keine signifikanten Unterschiede bezüglich der sexuellen Einstellung gefunden. (Gladue, Beatty, Larson & Staton, 1990)

8. Alterseffekte bei Raumvorstellungsleistungen

Dass mit steigendem Alter altersbedingte Veränderungen auftreten, ist durch eine Fülle an Studien, wie zum Beispiel durch Reischies & Lindenberger (2010) bekannt. Salthouse et al. (1990) erforschten anhand von 4 Tests, die Spatial Visualisation messen, wie sich die Leistungsfähigkeit in diesem Subfaktor mit dem Alter verändert. Dazu testeten sie Universitätsprofessoren an der technischen Universität im Alter zwischen 24 und 67 Jahren. Den Testpersonen wurde in einer Querschnittsstudie die Tests *Paper Folding Test*, *Form Board*, *Surface Development* und *Cube Comparison* vorgegeben.

Die Ergebnisse der Studie sind in Abbildung 11 dargestellt. Die Regressionslinien zeigen dabei für jeden Raumvorstellungstest einen negativen Verlauf. Die Leistungsunterschiede betragen im Alter von 60 Jahren 1-2 Standardabweichungen weniger als im 20. Lebensjahr. Im Verlauf kann durchschnittlich ein Absinken von 0.3 Standardabweichungen pro Dekade gezeigt werden.

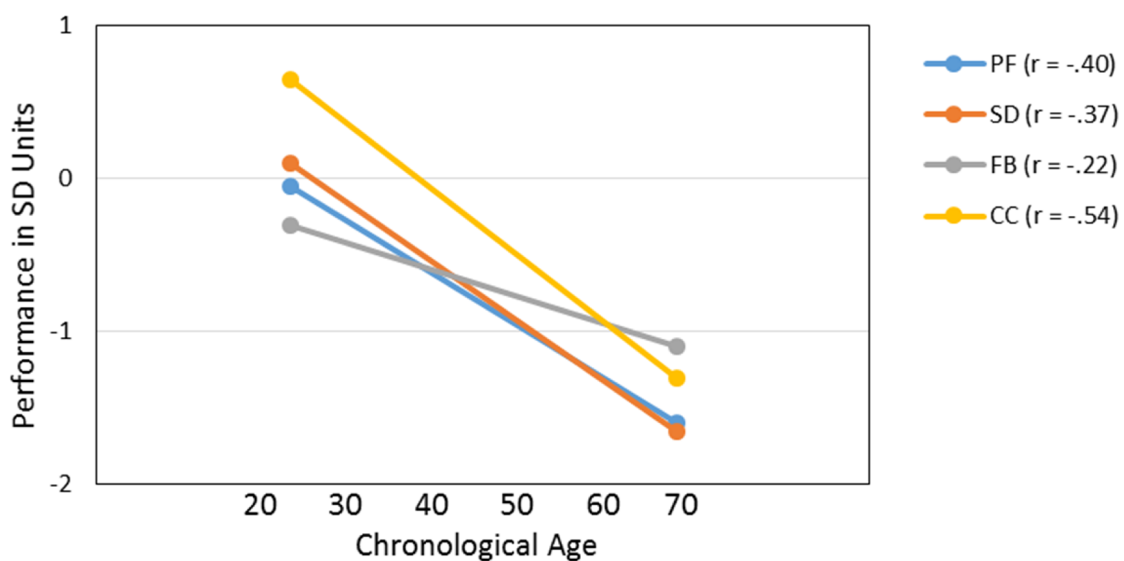


Abbildung 11. Alterseffekte nach Salthouse et al. (1990) samt Regressionskoeffizienten pro Test; PF = Paper Folding Test; SD = Surface Development Test; FB = Form Board; CC = Cube Comparison

Auch das berufliche Umfeld hängt mit der Raumvorstellungsleistung im Alter zusammen. In ihrer Analyse zeigten die Autoren, dass Menschen, die länger in einem raumvorstellungsnahem Berufsfeld (Architekten) arbeiten, auch im höheren Alter durchschnittlich höhere Werte in Raumvorstellungstests erreichen.

Zusammenfassend kann aufgrund der angeführten Theorien gezeigt werden, dass Raumvorstellung ein fundiertes Thema darstellt. Die Subfaktoren Spatial Visualisation und Mental Rotation sind in den meisten Modellen vertreten. Die Subfaktoren Spatial Perception, Spatial Relation und Spatial Orientation werden entweder kombiniert als Faktor angesehen (Linn & Peterson, 1985) oder einzeln als selbständige Faktoren beschrieben (Maier, 1999). Aufgrund der Vorteile des Rasch Modells finden in den letzten Jahren vermehrt Forschungen in Richtung der Konstruktion von Rasch homogenen Raumvorstellungsverfahren statt. Bisher

gibt es nur in den Subfaktoren Spatial Perception, Mental Rotation und Spatial Orientation deutschsprachige Rasch homogene Verfahren. Für den Faktor Spatial Visualisation, der in der Forschung durchgehend seinen Platz findet und einen fundierten Faktor der Raumvorstellung darstellt, fehlen Rasch homogene Verfahren. Deshalb soll sich diese Arbeit der Entwicklung eines Rasch konformen Spatial Visualisationstests widmen.

9. Fragestellung und Hypothesen

Abgeleitet aus der Theorie ist der wichtigste Punkt dieser Arbeit die Neukonstruktion des 3D-2D Objektschneidetests und dessen Überprüfung mittels Rasch Modell. Zusätzlich sollen einige Hypothesen aus der Theorie repliziert werden.

Hypothese 1

Das Rasch Modell kann für das aktuelle Itemset angenommen werden. Die Überprüfung der folgenden Hypothesen ist nur mit Rasch homogenen Items möglich.

Hypothese 2

Wie schon erwähnt, sind Geschlechterdifferenzen in der Raumvorstellung breit erforscht. Sieht man sich aber die Subbereiche Spatial Visualisation genauer an, so wurden keine bis wenige Geschlechtsdifferenzen postuliert. (Linn & Petersen, 1985) Aufbauend auf diesen Forschungen sollen mit Hilfe der erhobenen Stichprobe Geschlechtsdifferenzen überprüft werden. Die Hypothesen sollen neben Vertiefung in dem Gebiet der Raumvorstellung auch eine Erfassung der Validität ermöglichen. Eine Übereinstimmung der Ergebnisse mit schon in der Literatur erhobenen Hypothesen kann als Validitätsansatz interpretiert werden.

Hypothese 3

Wie im Bereich zu Altersunterschieden beschrieben, gibt es auch schon früh zu erkennende Altersdefizite in Spatial Visualisation. (Salthouse et al., 1990) Dies wurde anhand des Paper Folding Test überprüft, der eine hohe theoretische Ähnlichkeit zum 3D-2D

Objektschneidetest, aufgrund seiner Kategorisierung in der Dimension der 3D zu 2D Testverfahren, aufweist. Deshalb wurde der Alterseffekt als Hypothese formuliert.

Hypothese 4

Wie schon in der Theorie erwähnt, gibt es aufgrund des beruflichen Tätigkeitsbereichs oder der Ausbildung in Berufen mit Ähnlichkeit zur Raumvorstellung bessere Leistung als Personen die in raumvorstellungsfremden Bereichen arbeiten. (Salthouse et al., 1990) Diese Hypothese soll ebenfalls vor allem zur Validierung dienen. Misst der Test Spatial Visualisation, so sollten Personen im raumvorstellungsnahen Tätigkeitsbereich auch höhere Leistung erzielen.

Hypothese 4.1

Vor allem Techniker und Studenten technischer Berufe fallen in die Gruppe von Personen, die aufgrund ihrer Profession hohen Bezug zur Raumvorstellung haben. Deshalb sollen diese mit einer Gruppe von Personen verglichen werden, die aufgrund ihrer Profession wenig mit Raumvorstellung zu tun haben, in diesem Fall die Testpersonen im Tätigkeitsbereich Psychologie.

Hypothese 5

Zusätzlich zu der Art des Berufs soll auch überprüft werden, ob aufgrund der bisher abgeschlossenen Ausbildung signifikante Unterschiede zwischen Personen mit keinem Schulabschluss, Pflichtschule, Matura und universitären Abschluss gibt.

Hypothese 6

Die Erhebung der Selbsteinschätzung im Bereich Spatial Visualisation soll in Bezug zu der tatsächlichen Leistung im Test gesetzt werden, um dann darauf schließen zu können, ob der Test auch misst, was er vorgibt zu messen (Validität).

II. METHODE

10. Testkonstruktion – 3D-2D Objektschneidetest

Wie bereits erwähnt soll sich die Neukonstruktion des Tests im Bereich Spatial Visualisation, bzw. genauer gesagt im 3D-2D (dreidimensionales Stimulusmaterial und zweidimensionales Lösungsmaterial) Gebiet stattfinden. Diese Entscheidung wurde getroffen, um eine Abtrennung zum Bereich Mental Rotation zu erreichen. Der Paper Folding Test wird als eines der besten Testverfahren im 3D-2D Gebiet angesehen und soll mitsamt des Tests „Schnitte“ als Ausgangspunkt zur Neukonstruktion fungieren. Allerdings soll sich hier auch vom Test „Schnitte“ distanziert werden, da lediglich Ähnlichkeiten in der Vorgabe der Testitems und der gewählten Operationen erfolgen. Die Wahl der Objekte, Flächen, visuelle Darstellung der Körper und Testkonstruktion wurden eigenhändig erarbeitet. So erfolgt im 3D-2D Objektschneidetest auch keine textbasierte Darstellung der Items wie dies im Test „Schnitte“ der Fall ist. Dies wurde entschieden, da nicht ohne weitere Forschung davon ausgegangen werden kann, dass die sprachliche Kompetenz nicht in die Art der Bearbeitung dieser Raumvorstellungsaufgaben einfließt. Das hat in dieser Arbeit großen Einfluss, da die Berechnung des Tests über das Rasch Modell erfolgt und die Eindimensionalität nicht verletzt werden darf. Auch die Art und die Menge der Präsentation der Antwortalternativen und die Instruktion sollen anders gestaltet werden als im Test „Schnitte“.

Um neben der Überprüfung der Konformität des Rasch Modells auch einen Einblick zu bekommen, ob denn dieser Test auch misst, was er zu messen vorgibt, nämlich Spatial Visualisation, wurde nach den Übungsbeispielen anhand einer 6-stufigen Likert Skala die Frage gestellt: „Schätzen Sie bitte ein, wie gut Sie in diesem Test voraussichtlich abschneiden werden“. Die Überlegung dahinter war die, dass sich Personen nach der Instruktion und den Übungsbeispielen in ihrer Fähigkeit zu Spatial Visualisation einschätzen und diese mit dem erzielten Rohscore übereinstimmt. Daraus soll die Validität des Tests abgeleitet werden.

Verwendete Programme:

Zur Erstellung der 3 dimensionalen Körper wurde das Programm Sketch up 2013 (<http://www.sketchup.com/de>, Abbildung 12) verwendet. Mit diesem Programm ist es möglich Körper zu zeichnen und dann zu durchtrennen, um daraus die resultierenden Flächen zu zeichnen. Dies hilft zusätzlich der Fehlervermeidung bei der Testkonstruktion, da die Lösungen direkt aus dem bereits erstellten Körper entnommen werden können.

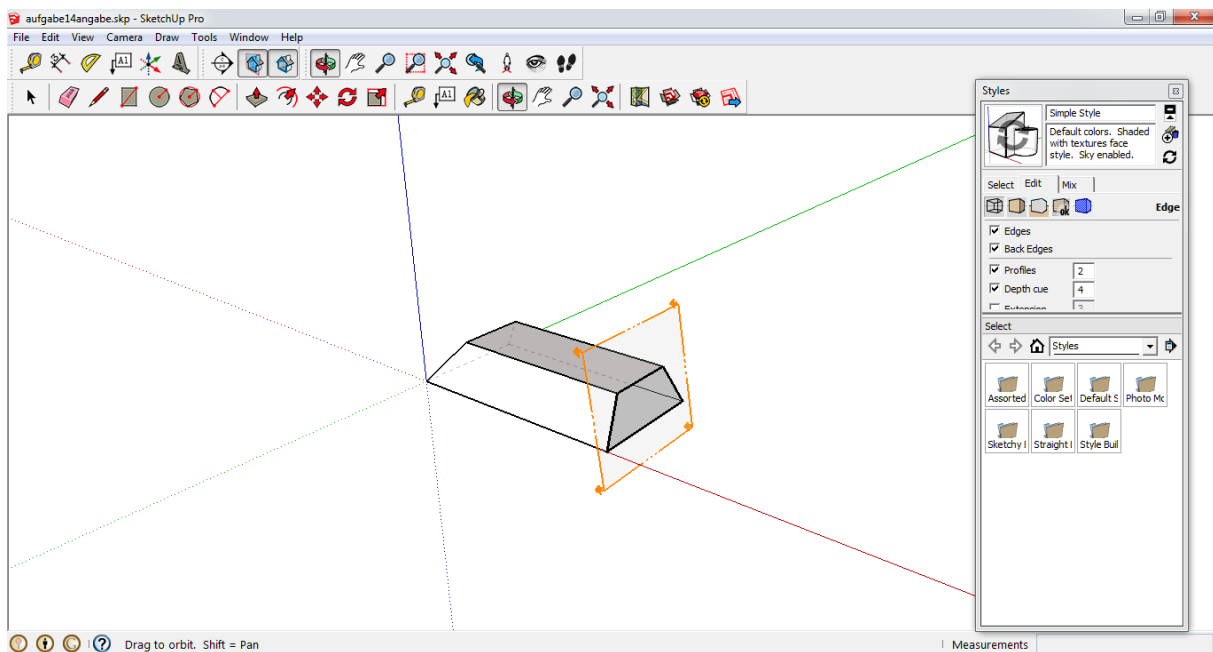


Abbildung 12. Screenshot aus dem Programm Sketch Up 2013

Zur grafischen Aufbereitung wurde das Bildbearbeitungsprogramm Gimp 2.8 verwendet (<http://www.gimp.org>). Hier erfolgte die Speicherung und Verarbeitung der gezeichneten Körper aus Sketch Up, um diese in ein passendes Format für die Testoberfläche zu konvertieren.

Zur Programmierung des Tests als PC-Verfahren wurde Microsoft Access 2013 verwendet (<http://products.office.com/de-de/access>). Die Funktion zur Formularerstellung dient dazu, um einen Ablauf der Instruktion, Beispielitems und Testitems zu erreichen.

10.1 Instruktion

Die Instruktion soll einen kompakten aber genauen Einblick in die notwendigen Lösungsschritte geben. Begonnen wird wie in Abbildung 13 gezeigt wird, mit der Darstellung des Lösungswegs in grafischer, sowieso in textbasierter Form. Das soll gewährleisten, dass Personen, die eine schlechte Kompetenz im verbalen Bereich aufweisen bzw. Personen mit schlechteren Deutschkenntnissen sich den Lösungsweg grafisch ansehen können.

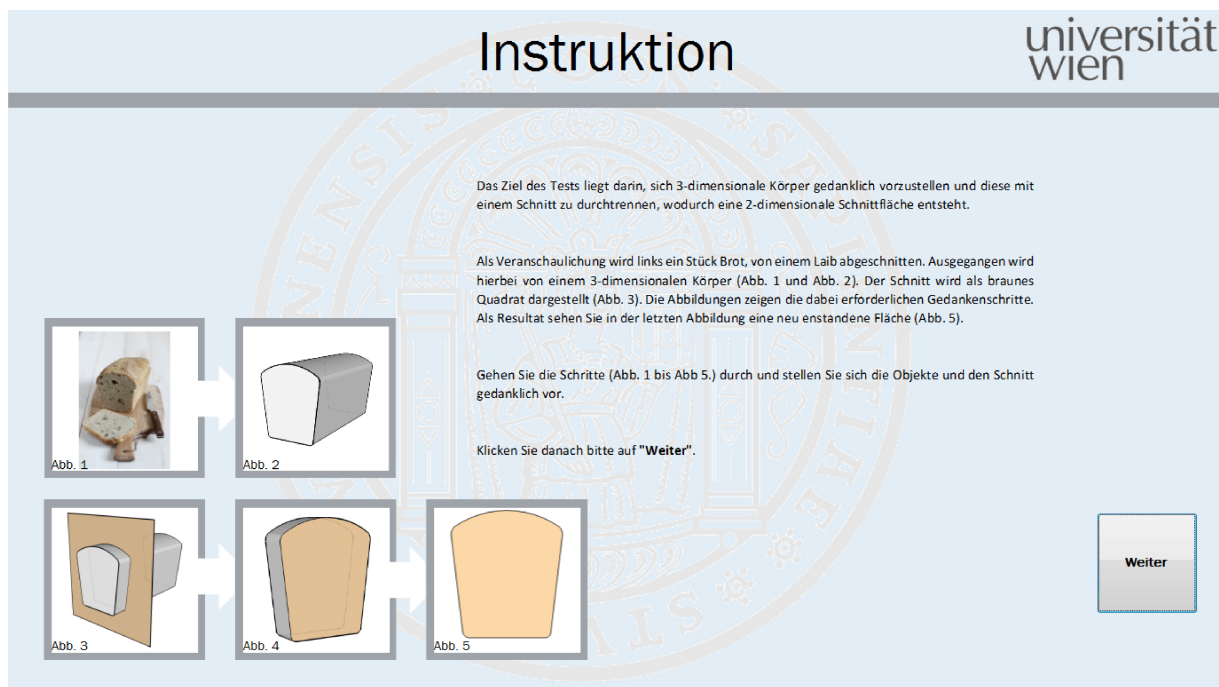


Abbildung 13. Instruktionsseite 1 des 3D-2D Objektschneidetests.

Auf der 2. Seite (Abbildung 14) soll gezeigt werden, dass es verschiedene Möglichkeiten gibt, einen Schnitt durch den Körper durchzuführen und dass es dadurch verschiedene Möglichkeiten gibt, neue Flächen entstehen zu lassen.

Instruktion

Wie Sie links sehen, gibt es nicht nur eine Möglichkeit ein Stück abzuschneiden, sondern unendlich viele. Schnitte an unterschiedlichen Stellen, mit unterschiedlichen Winkeln, resultieren in verschiedenen Flächen. Die Abbildungen 1a bis 1c sind mit der zuvor gezeigten Seite ident. Abbildungen 2a bis 2c und 3a bis 3c sind mögliche alternative Schnitte.

Es ist wichtig, dass ein durchgehender Schnitt, wie in den Abbildungen gezeigt, durchgeführt wird. Dabei muss nicht durch den Mittelpunkt des Körpers geschnitten werden.

Beachten Sie, dass die Flächen in unterschiedlicher Größe im Körper vorkommen können, die Proportionen aber erhalten bleiben.

Klicken Sie bitte auf "**Weiter**".

Zurück

Weiter

Abbildung 14. Instruktionssseite 2 des 3D-2D Objektschneidetests.

Hier werden wichtige Komponenten genannt, die bei der Durchführung des Tests wichtig sind. So muss zum Beispiel ein durchgehender Schnitt durch den Körper stattfinden und der Schnitt muss dabei nicht durch den Mittelpunkt des Körpers reichen. Außerdem können aufgrund der Variabilität des Schnitts Flächen entstehen, die gleich aussehen, aber verschiedene Größe aufweisen, die Proportionen der Außenlinien aber identisch sind. Es ist aber nicht von Relevanz welche Größe die Flächen haben, sondern nur, ob die Proportionen der Schnittfläche ident mit den Antwortalternativen sind.

Seite 3 (Abbildung 15) erklärt, wie das in der Instruktion gezeigte Beispiel als Testitem aussehen würde. Hier sieht man, dass der 3 dimensionale Körper links oben dargestellt wird. Die Testperson soll nun erkennen, welche unten gezeigten Flächen entstehen können, wenn ein Schnitt durch den Körper gezogen wird. In diesem Fall sind Antworten a, b und e korrekt.

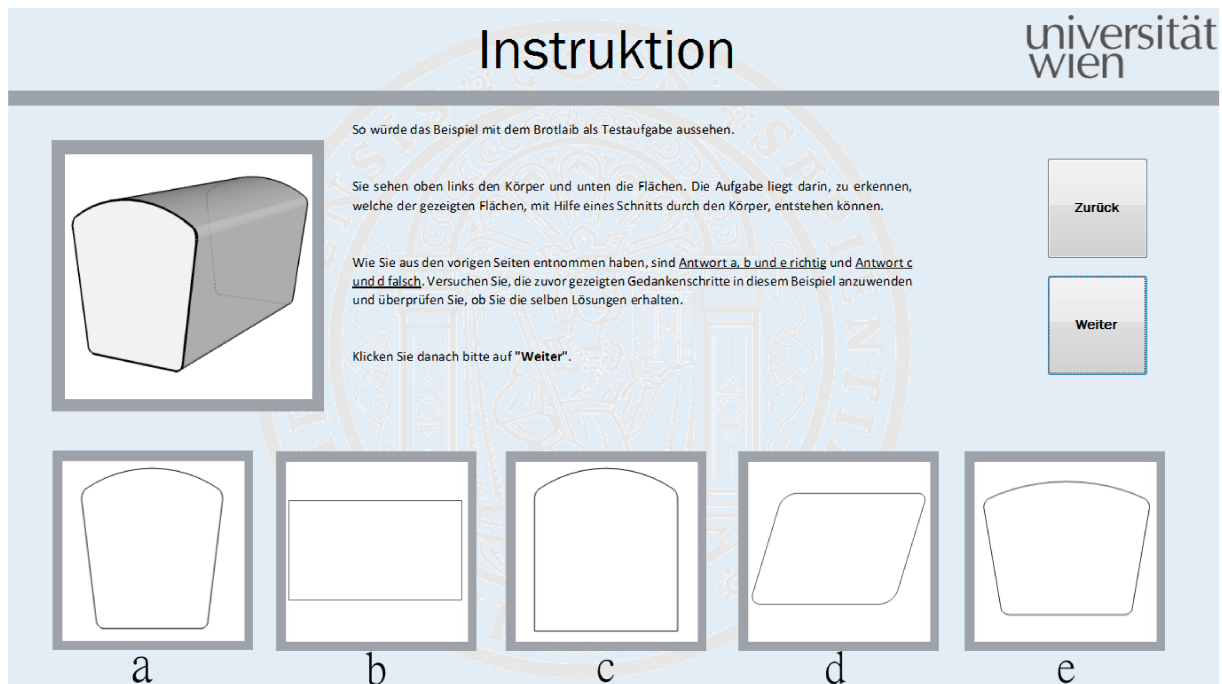


Abbildung 15. Instruktionssseite 3 des 3D-2D Objektschneidetests.

Auf Seite 4 (Abbildung 16) wird dargestellt, wie das Eintragen der gewählten Antwortalternativen erfolgen soll. Die Kreuze werden im Lösungsheft zu den korrespondierenden Antwortalternativen gesetzt. Wird ein Kreuz gesetzt und wird dann festgestellt die Antwortalternative ist doch inkorrekt, dann wird das Feld komplett ausgefüllt, um als falsch gewertet zu werden.

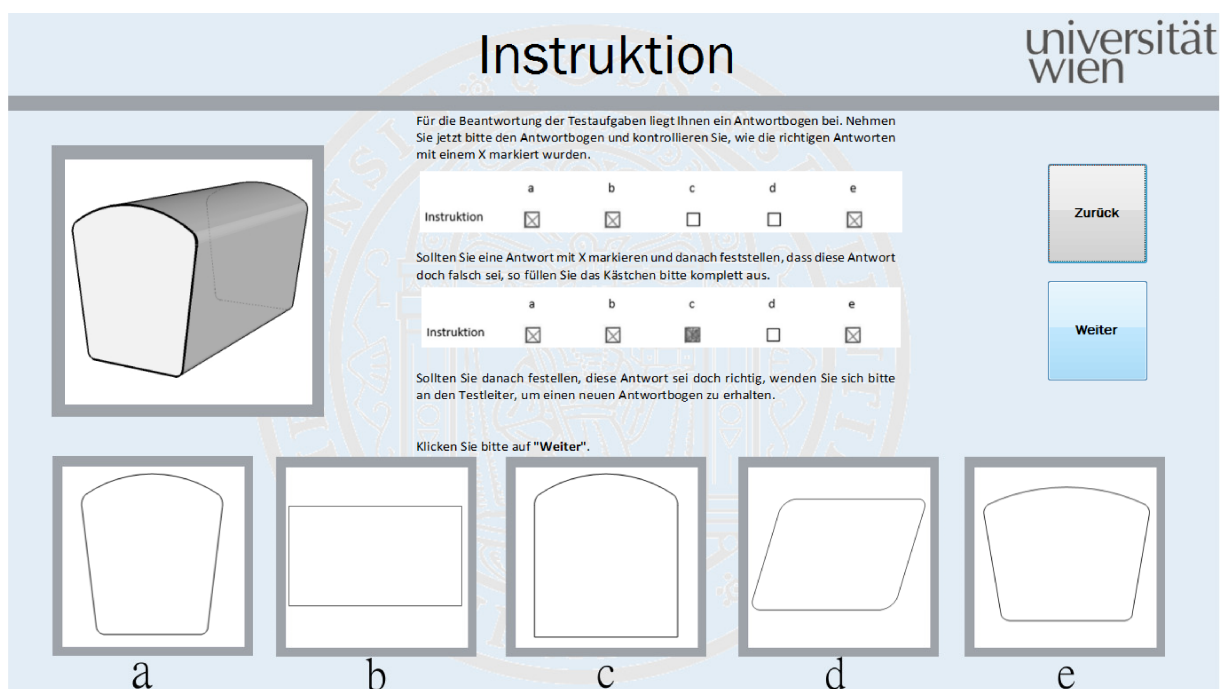


Abbildung 16. Instruktionssseite 4 des 3D-2D Objektschneidetests.

Auf der letzten Seite der Instruktion (Abbildung 17) werden nochmal wichtige Informationen zum Test zusammengefasst, die das Antwortformat und Zeitlimit betreffen.

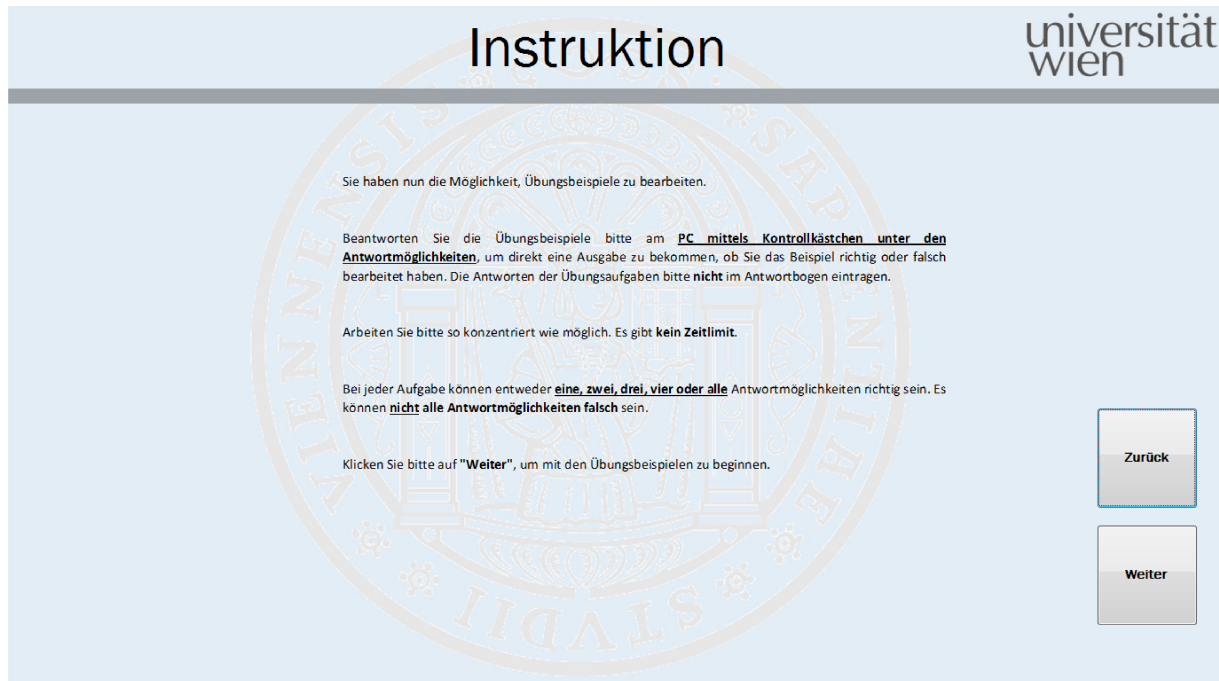


Abbildung 17. Instruktionsseite 5 des 3D-2D Objektschneidetests.

10.2 Übungsbeispiele

Die Übungsbeispiele sollen einen ersten Einblick geben, wie die Testitems gelöst werden sollen. Die Lösung wird anders als bei den Testitems direkt am Computer eingegeben, um den Testpersonen eine Rückmeldung über ihre Leistung zu ermöglichen. Zur Veranschaulichung soll exemplarisch das Übungsbeispiel Nummer 1 (Abbildung 18) gezeigt werden.

Übungsbeispiel 1

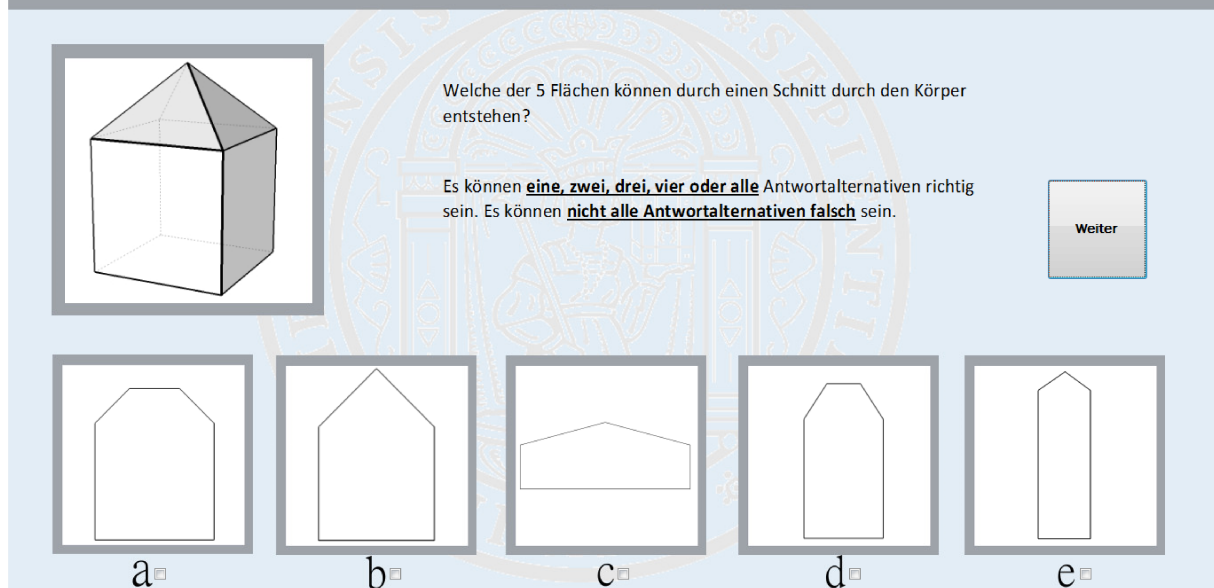


Abbildung 18. Übungsbeispiel 1 des 3D-2D Objektschneidetests.

Anders als bei den Testitems erfolgt bei den 4 Übungsbeispielen eine Rückmeldung darüber, ob das Beispiel richtig oder falsch beantwortet worden ist. Beantwortet die Testperson das Übungsbeispiel falsch, dann wird einerseits angezeigt, dass die Antwort inkorrekt war und andererseits bekommt man einen Hinweis, worauf man bei diesem Beispiel achten muss. (Abbildung 19)

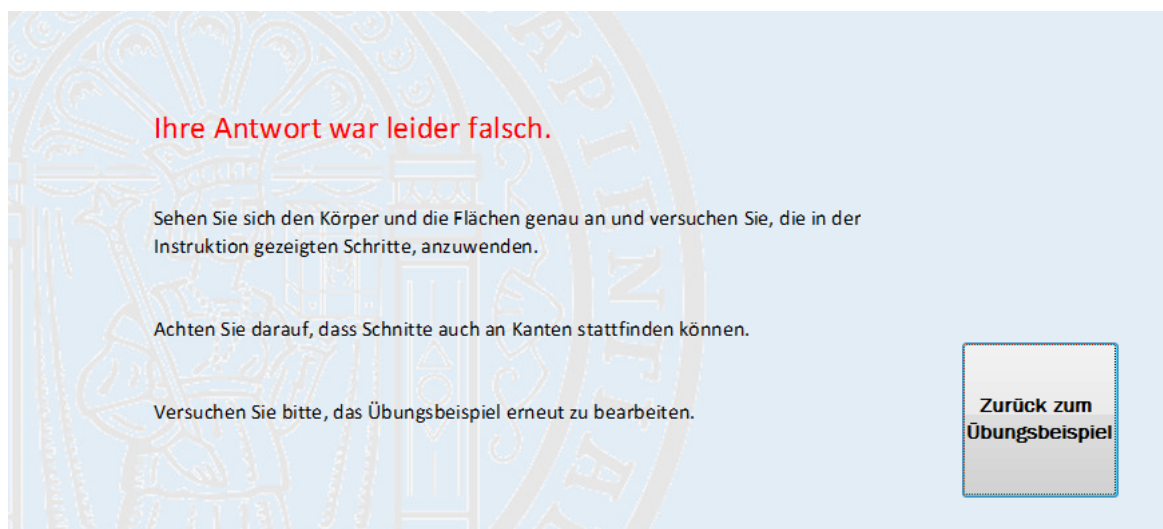


Abbildung 19. Verkleinerter Ausschnitt der Rückmeldung des 3D-2D Objektschneidetests, wenn das Übungsbeispiel 1 falsch beantwortet wurde, mit einem Hinweis zur richtigen Beantwortung des Beispiels.

Nach dem Hinweis bekommt die Testperson einen erneuten Versuch das Übungsbeispiel zu lösen. Wird hier erneut die falsche Antwort geben, wird der Testperson der richtige Lösungsweg grafisch dargestellt. Beantwortet die Testperson das Übungsbeispiel beim 1. oder 2. Versuch richtig, erfolgt ebenfalls eine Darstellung des grafischen Lösungsweges, um Testpersonen, die das Item richtig oder falsch gelöst haben nicht zu bevorzugen. Die Testpersonen haben zu diesen Zeitpunkt immer noch die Möglichkeit sich das Übungsbeispiel anzusehen, um die gezeigten Lösungsschritte selbst zu überprüfen. Abbildung 20 zeigt die Seite nach Beendigung des Übungsbeispiel 1. Gezeigt werden die Lösungsschritte und die richtigen Antworten. Wird das Item richtig gelöst erscheint dieselbe Seite, nur mit dem Text „Ihre Antwort war richtig“ in grüner Schrift.

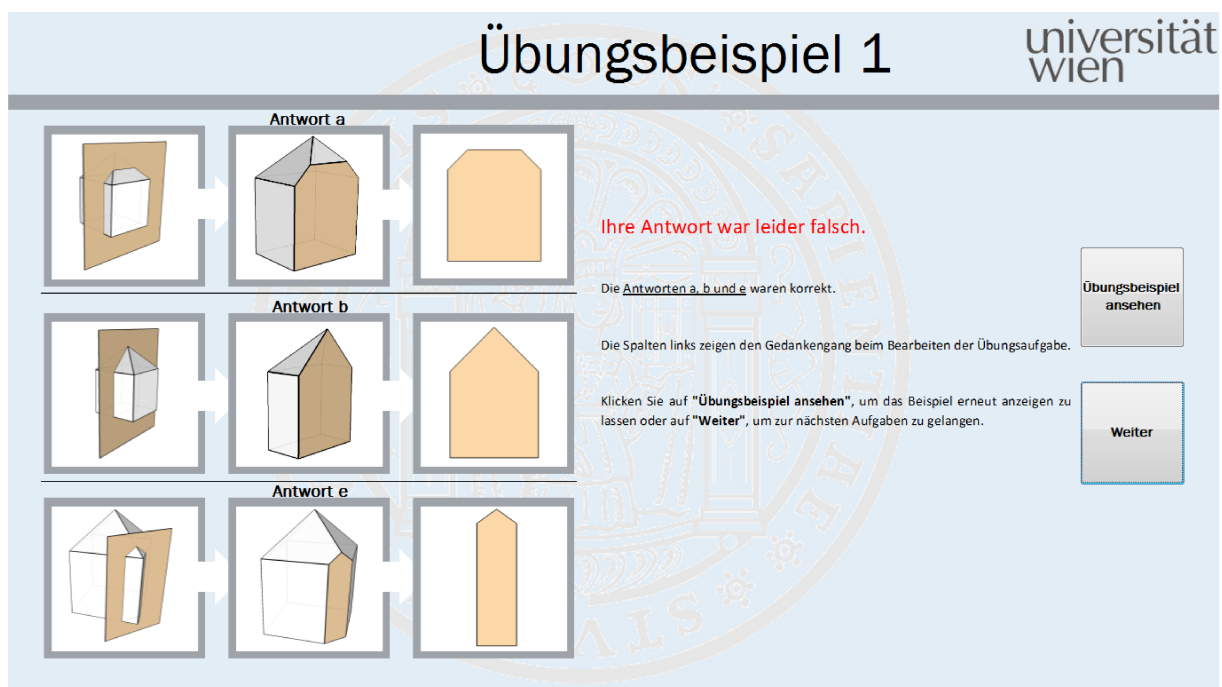


Abbildung 20. Übungsbeispiel 1 des 3D-2D Objektschneidetests nach falscher Beantwortung nach dem 2. Durchlauf.

Nach den 4 Übungsbeispielen wird eine Zwischenseite (Abbildung 21) mit Informationen eingeschaltet, dass es kein Zeitlimit gibt, wie das Antwortformat erfolgt, dass die Antworten ins Antwortheft eingetragen werden sollen und dass man sich keine Notizen machen darf.

Sie haben die Übungsbeispiele beendet. Füllen Sie bitte die Frage zur Selbsteinschätzung aus, die dem Testheft beiliegt, bevor Sie mit dem Test beginnen.

Sollten Sie Fragen zum Test haben, wenden Sie sich bitte an den Testleiter.

Die Aufgaben sind so konstruiert, dass nicht jede Person alle Aufgaben lösen kann. Arbeiten Sie trotzdem **so konzentriert wie möglich** weiter. Es gibt **kein Zeitlimit**.

Bei jeder Aufgabe können entweder **eine, zwei, drei, vier oder alle** Antwortmöglichkeiten richtig sein. Es können **nicht alle Antwortmöglichkeiten falsch** sein. Tragen Sie Ihre Lösungen bitte im Antwortbogen ein.

Machen Sie sich bitte keine Notizen, die Ihnen bei der Lösung der Beispiele helfen könnten.

Wenn Sie bereit sind, klicken Sie bitte auf **"Test starten"**.

Test starten

Abbildung 21. Zwischenseite vor der Vorgabe der Testitems des 3D-2D Objektschneidetests.

10.3 Testitems

Bei der Konstruktion der Testitems war die systematische Festlegung der Schwierigkeit ein großes Thema. Im Endeffekt wurden 4 Arten zur Schwierigkeitskonfiguration verwendet.

1. Schnitt parallel zur den Außenflächen des Körpers: Wird ein Schnitt durchgeführt, der sich parallel zu den Außenflächen befindet, so ist dieser leichter durchzuführen, als ein Schnitt der in einem Winkel zu den Außenflächen erfolgt.
2. Tiefe des Schnitts: Wird ein Schnitt nahe an den Außenflächen durchgeführt, so ist dies leichter. Vor allem in komplexen Körpern ist ein Schnitt in der Nähe der Mitte des Objekts mit mehr Vorstellungsleistung verbunden, als am Rand.
3. Die Komplexität des Körpers: Körper mit einer einfachen geometrischen Form, wie Kugel, Quader, Würfel, usw. sind einfacher zu manipulieren als komplexe Figuren, wie zum Beispiel eine 6 seitige Pyramide mit Aussparungen und einem angehefteten Quader als Grundfläche.

4. Vorgabe der Lösungsmöglichkeiten bzw. Einfluss von Distraktoren: Die Wahl der Lösungsmöglichkeiten hat einen großen Einfluss auf die Schwierigkeit des Items. Es kann ein sehr komplexer Körper verwendet werden, sind jedoch die Schnitte und somit die Lösungsmöglichkeiten einfach, kann auch davon ausgegangen werden, dass das Item leicht werden wird. Die Wahl der Distraktoren ist hier ein zusätzlicher Punkt. Werden Distraktoren gewählt, die hohe Ähnlichkeiten mit den richtigen Schnitten haben, wird das Item zugleich schwerer als mit Distraktoren, die dem Körper und den Schnitten fremd sind.

Die Schwierigkeitskonfigurationen können hierbei isoliert betrachtet werden, jedoch beeinflussen sie sich auch wechselseitig.

6.4.1 Distraktoren und falsche Antwortmöglichkeiten

Unter Distraktoren werden Antwortalternativen verstanden, die von den richtigen Alternativen ablenken sollen und die Testperson dazu verleiten sollen sich für die falsche Alternative zu entscheiden. Kubinger (2009) definiert einen Distraktor wie folgt:

Aus dem Englischen kommend (*distractor*) meint diese Bezeichnung aber „Ablenker“; die Aufgabe eines Distraktors ist es also, eine nicht lösungskompetente Tp von der Lösung abzulenken, insbesondere dadurch, dass er aus teilweise richtigen Komponenten besteht, so dass er der Lösung möglichst nahe kommt. So gesehen ist ein Distraktor dann brauchbar, wenn er die Tp „verwirrt“, wodurch nur ausreichend befähigte Tpn zur Lösung gelangen. (Kubinger, 2009, S. 135)

Die Konstruktion der Distraktoren und der falschen Antwortmöglichkeiten erfolgte anhand von 3 Methoden.

1. Hohe Ähnlichkeit zu einem Schnitt der richtig ist, in dem aber die Proportionen des Distraktors so geändert wurden, dass er im Stimulusobjekt nicht vorkommen kann. Hierbei ist es wichtig darauf zu achten, dass sich Schnitt und Distraktor nicht zu ähnlich

sind. Wäre dies der Fall, so würden differentielle Unterschiede von Personen aufgrund der Fähigkeit der Wahrnehmung entstehen.

2. Eine Fläche, die mit dem Objekt und den Schnitten sehr wenig Zusammenhang hat. Dies wäre der Fall, wenn bei der Vorgabe eines Würfels eine symmetrische 6-seitige Fläche vorgegeben wird. Dies ist bei einfachen Körpern eine relativ schnell erkennbarere falsche Antwortmöglichkeit, jedoch bei komplexen Körpern ist dies eine gute Wahl der Konstruktion eines Distraktors.
3. Eine weitere Möglichkeit der Distraktorkonstruktion ist die Ausnutzung von Denkfehlern. Dies zeigt sich jedoch als schwierigste Methode, da es sehr aufwendig bzw. komplex herausstellt, solche Denkfehler zu identifizieren. Eine Möglichkeit dafür besteht darin, Personen in einer Vorstudie Items vorzugeben und diese zu bitten die Aufgaben mittels lautem Denken zu lösen. Hiermit können Lösungswege und Denkfehler identifiziert und zur Distraktorkonstruktion verwendet werden. Exemplarisch soll die Konstruktion eines Distraktors in Abbildung 22 grafisch dargestellt werden. Anhand eines Denkfehlers kann es passieren, dass man sich aufgrund eines Schnittes, von der linken unteren Ecke bis zur Mitte des Kreisbogens, vorstelle, dass die rechte Ecke der resultierenden Fläche rund werde. Das ist jedoch nicht der Fall. Durch so einen Schnitt würde immer die Schnittfläche rechts resultieren.

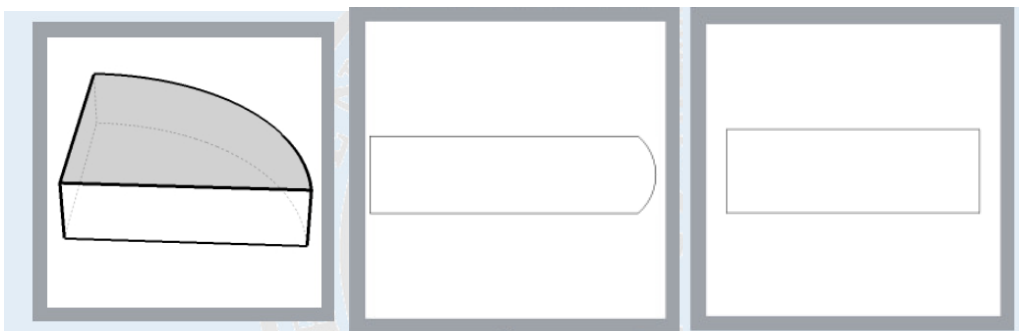


Abbildung 22. Auszug aus dem Item 5 mit dem Distraktor in der Mitte und der richtigen Antwort rechts.

11. Untersuchungsablauf

Der Ablauf ist standardisiert und erfolgt durch die Vorgabe des Tests auf dem Laptop bei den Testpersonen zu Hause oder in vorher festgelegten ruhigen Plätzen. Es fanden 3 Termine mit Gruppentestungen á 15 Leuten statt. Die restlichen Daten wurden mittels Einzeltestungen erhoben. Alle Teilnehmer nahmen freiwillig an der Testung teil. Begonnen wird mit der

Begrüßung der Versuchsperson, sowie die Versicherung der Anonymisierung der Daten. Die Erfassung der soziodemografischen Daten erfolgte mittels Pen and Paper zu Beginn des Tests. Erhoben wurden Alter, Geschlecht, Nationalität, bisher höchste abgeschlossene Ausbildung und aktuell ausgeführter Beruf bzw. Studium. Nach der Instruktion und den Übungsbeispielen erfolgt die Einschätzung der Leistung, gefolgt von der Bearbeitung der Testitems.

12. Stichprobendarstellung

Die Stichprobe umfasst nach der Datenerhebung 174 Personen. 4 Personen mussten nach standardisierter Auswertung aus dem Stichprobenumfang exkludiert werden, da diese Teilnehmer kein Item richtig gelöst hatten. Die Modellüberprüfung erfordert Datensätze mit mindestens einem richtig und einem falsch gelösten Item. In Tabelle 2 sind die soziodemografischen Verteilungen dargestellt.

Tabelle 2

Soziodemografische Verteilung der Stichprobe

Demografische Variable	Kennwert
Teilnehmeranzahl	174
- Männer	89
- Frauen	85
Alter	29.9 Jahre
- Spannweite	16-56 Jahre
Nationalität	
- Österreich	92.5 %
- Deutschland	4.0 %
- Sonstige	2.9 %
Höchste bisher abgeschlossene Ausbildung	
- Pflichtschule	30
- Matura	86
- Universitärer Abschluss	31
- Sonstige abgeschlossene Ausbildung	27

Anmerkung. Unter sonstige abgeschlossene Ausbildung fallen Weiterbildungen und Lehrgänge. Fehlende Daten wurden in der Tabelle nicht angeführt.

Die erfragten Variablen „zur Zeit ausgeübter Beruf“ und „aktuelle Studienrichtung“ können als zusammenhängende Themenblöcke aufgefasst werden. So können Techniker als Beruf bzw. eine Studienrichtung im technischen Bereich zum Themenblock „technischer Bereich“ subsummiert werden. Daraus lassen sich aus den 2 Variablen schließlich 10 Themenblöcke ableiten. Die Themenblöcke sind in Tabelle 3 zusammengefasst.

Tabelle 3

*Die erstellten Themenblöcke aus den Variablen
Aktueller Beruf und aktuelle Studienrichtung*

Themenblock	N
Psychologie	23
Kaufm. Tätigkeit	37
Sozial / Gesundheit	8
Betriebswirtschaft	13
Technischer Bereich	23
Lehramt	7
Rechtswissenschaften	8
IT-Bereich	12
Sonstige Studienrichtungen	26
In Ausbildung	6
Fehlend	11
Gesamt	174

Da es sich um einen reinen Powertest hielt, wurden die Testzeiten auch nicht systematisch protokolliert, wobei die kürzeste Bearbeitung des Test bei ungefähr 15 Minuten lag und die längste Bearbeitung bei ungefähr 60 Minuten. Durchschnittlich brauchten die Testpersonen 30-40 Minuten zur Durchführung des Tests.

13. Analyseplan

Der Likelihoodquotiententest nach Andersen (1973)

Likelihood ist im Rahmen der Wahrscheinlichkeitsberechnung ein Verfahren, mit dem es möglich ist, aufgrund von unabhängigen Ergebnissen das Outcome zu schätzen. Wenn das

Rasch Modell gilt, kann aufgrund der stochastischen Unabhängigkeit der Items auch hier die Likelihoodberechnung durchgeführt werden.

Anhand des LRT kann überprüft werden, ob die geschätzten Parameter in den gewählten Subgruppen die gleichen Schätzwerte aufweisen. Ein signifikantes Ergebnis geht mit der Verwerfung der H_0 (H_0 = Das Modell gilt) einher und die Gültigkeit des Rasch Modells wird nicht angenommen.

Die Testgröße ist asymptotisch χ^2 , mit den Freiheitsgraden $df = (k-1)(G-1)$ verteilt. K sind hierbei die Anzahl der Items und G die Anzahl der Teilgruppen. Wird die Testgröße mit der χ^2 -Verteilung verglichen, so erhält man einen p -Wert mit dem man entscheiden kann, ob die H_0 verworfen werden muss und somit das Rasch Modell keine Gültigkeit finden kann.

Der Wald Test nach Lord (1980)

Der Wald Test basiert auf dem gleichen Ansatz wie der LRT, der Datensatz wird aber auf Itemebene geteilt. So erhält man die Hypothese $H_0: \beta_i^{(1)} = \beta_i^{(2)} = \beta_i$. Signifikante Unterschiede zwischen den Teildatensätzen weisen auf eine Verletzung des Rasch Modells hin.

Martin-Löf-Test (1973, zit. n. Koller et al., 2012)

Der Martin-Löf-Test dient der Überprüfung der Itemhomogenität. Auch er beruht auf der Methode der Likelihoodquotienten, aber anders als beim LRT, werden hier die Items in Gruppen unterteilt. Die Idee dahinter ist, dass bei unabhängigen eindimensionalen Testaufgaben die Fähigkeitsparameter der Personen in den beiden Itemgruppen ähnlich sein müssen. Personen mit einer hohen Fähigkeit in der einen Gruppe sollten auch hohe Fähigkeit in der anderen Gruppe aufweisen.

Die Testgröße ist χ^2 verteilt sich mit den Freiheitsgraden $k_1 \cdot k_2 - 1$. Wird das Ergebnis signifikant, so muss davon ausgegangen werden, dass die Personenscores nicht eindimensional sind und somit die Items nicht als homogen angenommen werden können.

Die Teilung der Kriterien erfolgt standardmäßig mittels dem internen Teilungskriterium Median oder Mittelwert bzw. theoriegeleitet anhand von externen Teilungskriterien, wie bei der Überprüfung der lokal stochastischen Unabhängigkeit.

III. DARSTELLUNG UND ERGEBNISSE

14. Hypothesenprüfung

Die Hypothesenprüfung erfolgt, wenn nicht anders beschrieben 2-seitig mit $\alpha = .05$.

14.1 Hypothese 1 - Überprüfung auf Gültigkeit des Rasch Modells

Die Überprüfung auf Gültigkeit des Rasch Modells lehnt sich an die Arbeiten von Koller et al. (2012) an und soll demnach laut ihrem Vorschlag zur Durchführung des Rasch Modells erfolgen. Abbildung 23 soll das gewählte Vorgehen beschreiben.

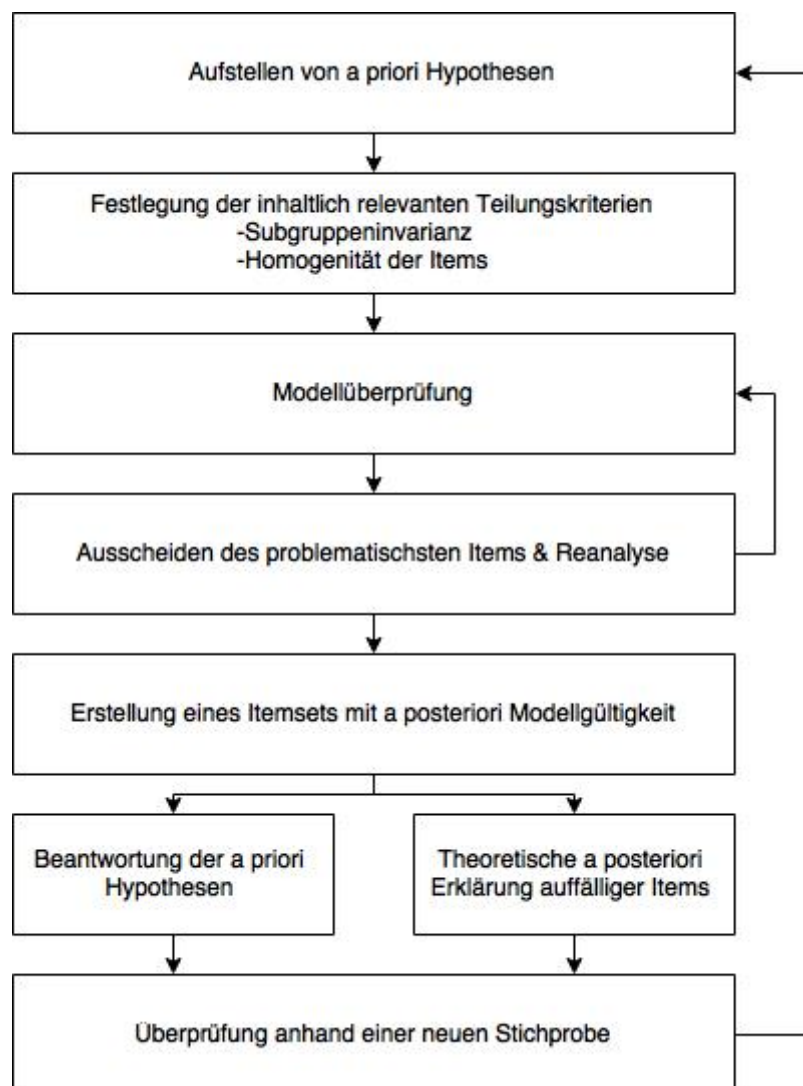


Abbildung 23. Vorgehen für eine Überprüfung auf das Rasch Modell nach Koller et al. (2012)

Das Aufstellen der a-priori Hypothesen stellt einen wichtigen Aspekt bei der Modellprüfung dar. Um einer artifiziellen Modellanpassung entgegen zu wirken, muss eine theoriegeleitete Selektionsstrategie erstellt werden. Diese soll einer Modellkontrolle anhand rein „technischer“ Kriterien entgegen wirken. Mit einer Selektionsstrategie können externe Teilungskriterium für den LRT erarbeitet und somit systematische Unterschiede zwischen Item und Personengruppen reflektiert werden. Inhomogenitäten können zu einem späteren Zeitpunkt durch diese Vorüberlegungen besser erklärt werden. (Gittler, 1986)

Die Teilungskriterien für den LRT orientieren sich in dieser Modellkontrolle an den internen Teilungskriterien Median und den theoriegeleiteten externen Teilungskriterien Geschlecht, Alter und bisherige abgeschlossene Ausbildung. In den drei externen Teilungskriterien können Unterschiede im Antwortverhalten oder unterschiedliche Lösungsstrategien angenommen werden. Die folgenden Berechnungen fanden im R-Studio mit der Libraryerweiterung eRm statt.

Modellprüfung

Subgruppeninvarianz

Die Subgruppeninvarianz dient der Überprüfung der spezifischen Objektivität bzw. Stichprobenunabhängigkeit. Als erstes Teilungskriterium wird der Median gewählt. Der Mittelwert wird nicht verwendet, da aufgrund der Häufigkeitsverteilung des Rohscores in der Stichprobe Ausreißer im oberen Fähigkeitsbereich festzustellen sind und eine Linksschiefe angenommen werden kann. Abbildung 24 zeigt die Verteilung des Rohscores.

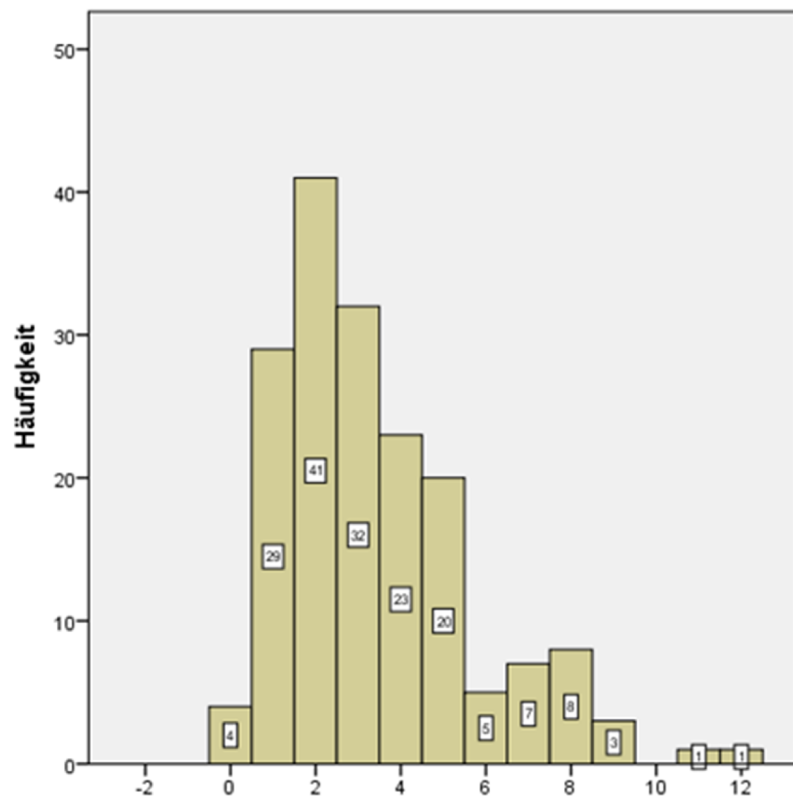


Abbildung 24. Häufigkeitsverteilung des Rohscores in der Gesamtstichprobe

In den Berechnungen zur Subgruppeninvarianz werden 4 Teilungskriterien verwendet. Durch die wiederholte Testung der Hypothesen muss deshalb eine Alphakorrektur nach Bonferroni durchgeführt werden. Koller et al. (2012) gehen davon aus, dass aufgrund des Zusammenspiels von Alpha- und Betafehlern ein Signifikanzniveau von $\alpha = .10$ als optimal anzunehmen ist. Die Alpha-Korrektur für die Subgruppeninvarianzberechnung für den Andersen LRT berechnet sich wie folgt:

$$\alpha^* = \frac{\alpha}{q}$$

q ... Anzahl Teilungskriterien

Für die Alphakorrektur im Waldtest (Überprüfung der Subgruppen auf Itemebene) gilt die Alphakorrektur mit

$$\alpha^* = \frac{\alpha}{q * k}$$

q ... Anzahl Teilungskriterien

k ... Anzahl der Items im Test

Die Modellprüfung mittels LRT zeigt die Ergebnisse in Tabelle 4.

Tabelle 4

LRT für die 4 Teilungskriterien

Teilungskriterien für LRT - Untergruppenbeschreibung	k	N	$LRT(\chi^2)$	p	α^*
Median	15	170	26.385	.023*	.025
- <= Median		106			
- > Median		64			
Geschlecht	15	170	22.217	.076	.025
- Männer		88			
- Frauen		82			
Alter	15	168	16.801	.267	.025
- < 30 Jahre		114			
- >= 30 Jahre		54			
Ausbildung	15	170	40.146	.553	.025
- Pflichtschule		30			
- Matura		86			
- Universität		31			
- Andere		27			

Wie in Tabelle 4 zu sehen ist, gibt es nur im internen Teilungskriterium Median signifikante Werte ($LRT(\chi^2) = 26.385$, $p = .023$). Somit wird die H_0 verworfen und der Test kann nicht als Rasch konform angenommen werden. Aufgrund der Signifikanz im LRT muss entschieden werden, welches Item nicht Rasch homogen ist und aus der Modellberechnung ausgeschlossen wird. Zur Klarifizierung wurde der Wald-Test durchgeführt. Tabelle 5 zeigt den

Waldtest und Abbildung 25 die grafische Veranschaulichung der Subgruppeninvarianz mit Teilungskriterium Median.

Tabelle 5

Waldtest für das Teilungskriterium Median

Beta Item #	z-Statistik	p	α^*
Beta Item 1	0.870	.384	.0022
Beta Item 2	-1.201	.230	.0022
Beta Item 3	2.305	.021	.0022
Beta Item 4	2.104	.035	.0022
Beta Item 5	-1.143	.253	.0022
Beta Item 6	0.189	.850	.0022
Beta Item 7	-2.779	.005	.0022
Beta Item 8	0.916	.360	.0022
Beta Item 9	-0.687	.492	.0022
Beta Item 10	0.624	.533	.0022
Beta Item 11	-0.565	.572	.0022
Beta Item 12	-0.915	.360	.0022
Beta Item 13	0.189	.850	.0022
Beta Item 14	0.820	.412	.0022
Beta Item 15	1.474	.140	.0022

Anmerkung. der z-Wert beschreibt, ob das Item in einer Gruppe häufiger gelöst wurde als in der Anderen. (Ein Wert > 0 bedeutet, dass das Item von Personen mit einem Rohscore über dem Median besser gelöst wurde und vice versa) Die Alpha-Korrektur erfolgt anhand von 3 Teilungskriterien x 15 Testitems; Es werden nur 3 Teilungskriterien herangezogen, da das Teilungskriterium „bisher abgeschlossene Ausbildung“ nicht im Waldtest berechnet werden kann und somit auch nicht in die Alpha-Korrektur einfließt.

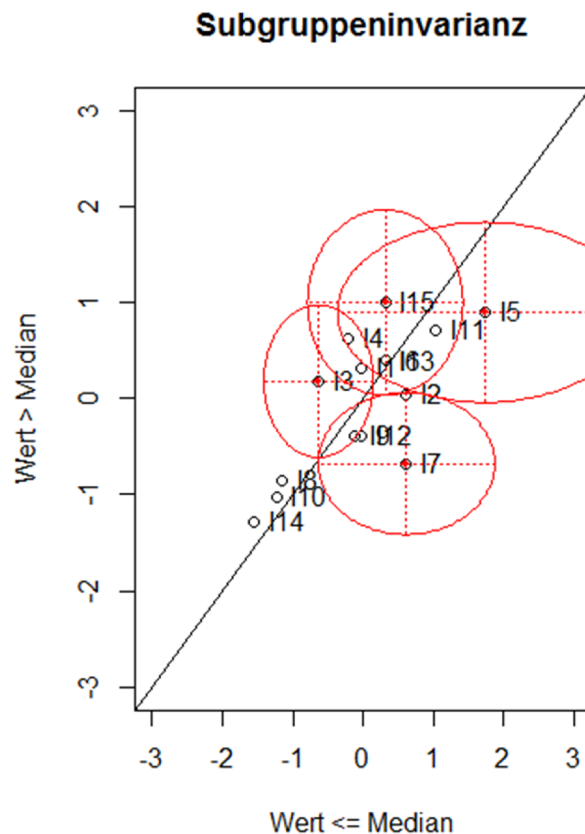


Abbildung 25. Grafische Darstellung des LRT mit Teilungskriterium Median. Items mit den Ellipsen werden in den Subgruppen unterschiedlich oft gelöst. Liegen die Items auf der Diagonale werden sie von beiden Gruppen gleich gut gelöst.

Zu sehen ist, dass es einige Items gibt, die Auffälligkeiten zeigen und somit zum Ausschluss bereit stehen. Dabei handelt es sich um Item 7 ($z = -2.779, p = .005$), Item 3 ($z = 2.305, p = .021$) und Item 4 ($z = 2.104, p = .035$). Die p-Werte sind nach der Alpha-Korrektur zwar nicht signifikant, jedoch haben sie trotzdem Einfluss auf die schlechten Werte im LRT. In Abbildung 25 zeigen diese drei Items knapp keine Modellverletzung, Item 7 scheint aber von Personen mit niedriger Fähigkeit leichter lösbar zu sein, als von Personen mit hoher Fähigkeit. Bei Item 3 und 4 verhält es sich gegensätzlich. Da Item 7 nahe an einer Modellverletzung liegt, soll es deshalb für eine Reanalyse vom Test ausgeschlossen werden.

Reanalyse

Subgruppeninvarianz

Die Reanalyse des Andersen LRT ist in Tabelle 6 abgebildet.

Tabelle 6

LRT für die 4 Teilungskriterien ohne Item 7

Teilungskriterien für LRT -Untergruppenbeschreibung	<i>k</i>	<i>N</i>	<i>LRT</i> (χ^2)	<i>p</i>	α^*
Median	14	170	15.572	.273	.025
- <= Median		79			
- > Median		91			
Geschlecht	14	170	20.360	.087	.025
- Männer		88			
- Frauen		82			
Alter	14	168	16.291	.234	.025
- < 30 Jahre		114			
- >= 30 Jahre		54			
Ausbildung	14	170	36.547	.582	.025
- Pflichtschule		30			
- Matura		82			
- Universität		31			
- Andere		27			

Ohne Item 7 zeigen sich im LRT keine signifikanten Werte und deshalb kann davon ausgegangen werden, dass die Items Stichprobenunabhängigkeit aufweisen.

Lokale stochastische Unabhängigkeit (LSU)

Die Ergebnisse der LSU sind in Tabelle 7 zusammengefasst.

Tabelle 7

Lokal stochastische Unabhängigkeit (LSU) für alle Items

LSU Item #	k	LRT(χ^2)	p	α^*
Item 1	13	8.61	.736	.10
Item 2	13	8.30	.761	.10
Item 3	13	15.62	.209	.10
Item 4	13	15.88	.197	.10
Item 5	13	10.14	.604	.10
Item 6	13	12.41	.413	.10
Item 8	13	7.55	.819	.10
Item 9	13	13.66	.323	.10
Item 10	13	14.87	.249	.10
Item 11	13	4.50	.958	.10
Item 12	13	13.73	.318	.10
Item 13	13	17.88	.119	.10
Item 14	13	12.14	.434	.10
Item 15	13	11.29	.505	.10

Anmerkung. Das Item in der Spalte „LSU Item #“ gibt an, dass dieses Item zur Splittung des Datensatzes verwendet wurde.

In den Berechnungen der LSU mittels Andersen LRT erfolgen keine Alpha-Korrekturen, weshalb das $\alpha = .10$ verwendet wird. Anzumerken ist, dass vor dem Ausschluss des Items 7, Item 9 und 13 noch signifikante Werte aufwiesen.

Item-Homogenität

Der Martin-Löf Test zur Überprüfung der Item-Homogenität zeigt mit $\chi^2 (48, n = 170) = 49.04$, $p = .431$ kein signifikantes Ergebnis. Deshalb kann davon ausgegangen werden, dass die Items homogen sind.

Itemparameter

Mithilfe der Schätzung der Itemparameter ist ein Einblick möglich, wie sich die Schwierigkeitsparameter der Items über den Test verteilen. In Tabelle 8 befindet sich die Liste mit den Rasch konformen Itemleichtigkeitsparameter (β). Ein negativer Wert gibt Aufschluss darüber, dass es sich um eine schwierige Aufgabe handelt.

Tabelle 8

Itemleichtigkeitsparameter (β) der Rasch homogenen Items

Item #	Geschätzte Std	Unteres KI	Oberes KI
Item 1	-0.20	-0.59	0.19
Item 2	-0.24	-0.63	0.15
Item 3	0.18	-0.17	0.54
Item 4	-0.28	-0.68	0.11
Item 5	-1.08	-1.60	-0.57
Item 6	-0.38	-0.78	0.03
Item 8	0.97	0.66	1.28
Item 9	0.25	-0.10	0.60
Item 10	1.10	0.79	1.41
Item 11	-0.81	-1.27	-0.34
Item 12	0.22	-0.14	0.57
Item 13	-0.38	-0.78	0.03
Item 14	1.39	1.09	1.70
Item 15	-0.75	-1.20	-0.29
Bedingte log-likelihood:		-834.57	
Anzahl der Wiederholungen:		13	

Anmerkung. Std = Standardabweichung; KI = Konfidenzintervall

Die grafische Darstellung der Itemparameter wird durch die ICC dargestellt.

Item Characteristic Curves (ICC)

Die ICC sollen hier von allen Items aufgelistet werden. Anzumerken ist, dass die Ausprägungen der latenten Dimension auf den Kurven liegen, bzw. die Konfidenzintervalle mit der ICC kreuzen sollen. Die Konfidenzintervalle im hohen Fähigkeitsbereich sind in diesem Test durch hohe Varianz gekennzeichnet, da nur wenige Versuchspersonen im mittleren oder hohen Fähigkeitsbereich liegen. Abbildungen 26-29 zeigen die Auflistung der ICC.

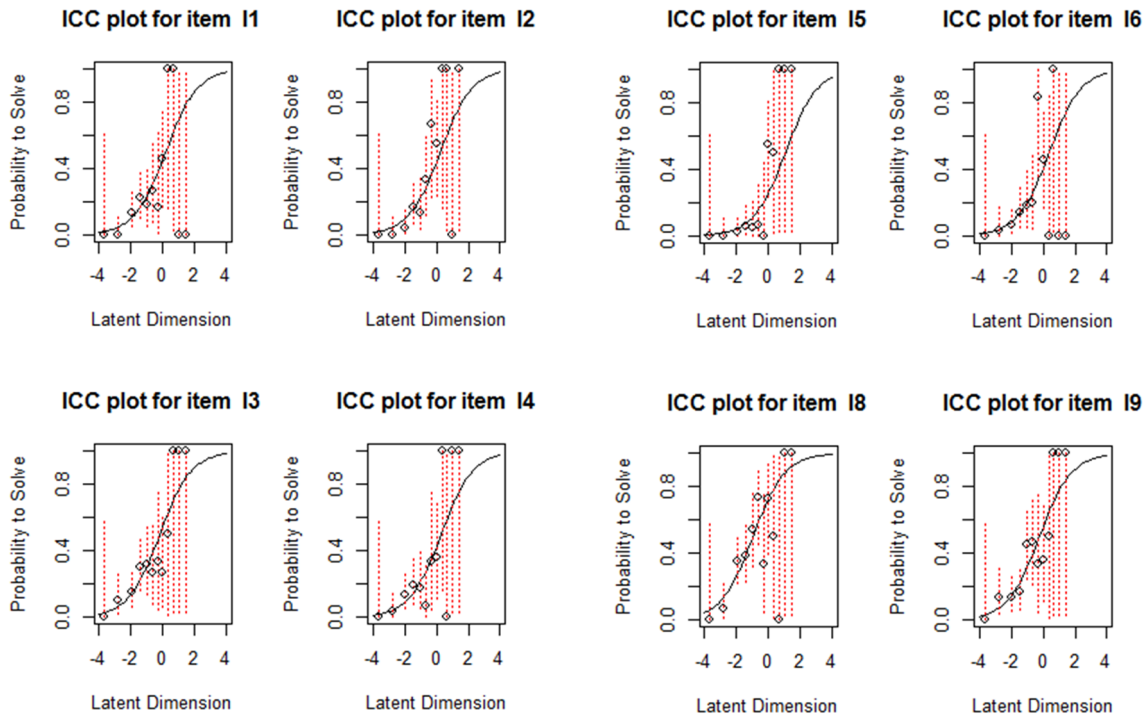


Abbildung 26. ICC für Item 1-4

Abbildung 27. ICC für Item 5,6, 8 und 9

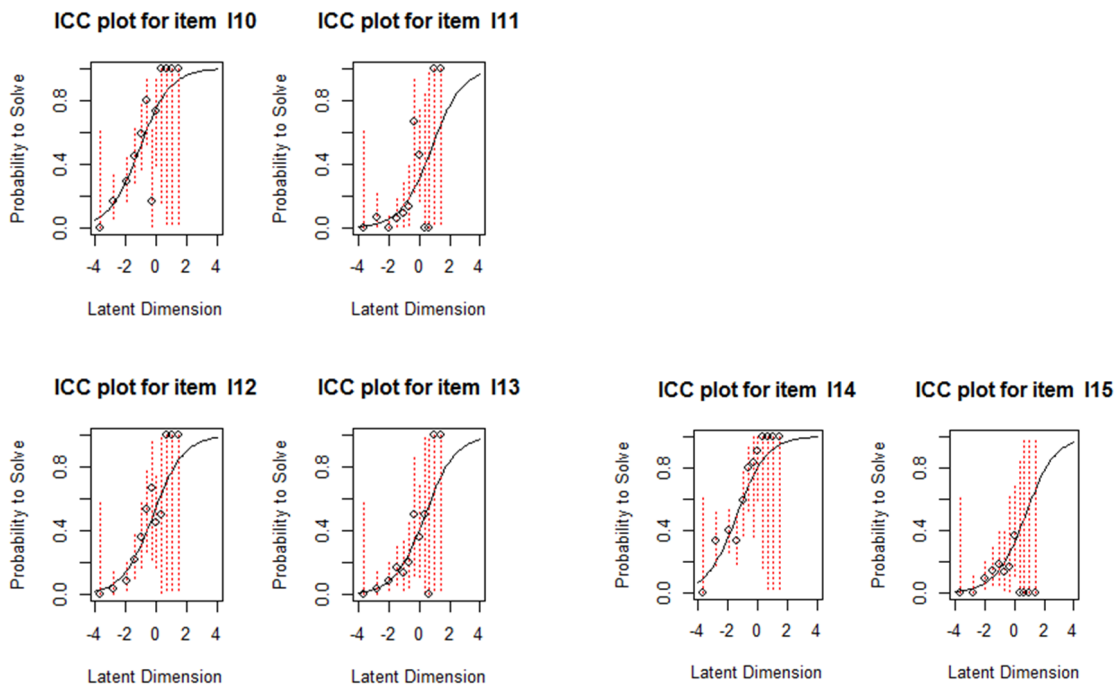


Abbildung 28. ICC für Item 10-13

Abbildung 29. ICC für Item 14 und 15

Alle Ausprägungen der latenten Fähigkeit mit den Konfidenzintervallen liegen auf den 14 Kurven, obwohl die Konfidenzintervalle enorme Maße angenommen haben. Dies ist der Fall, da nur wenige Personen einen höheren Rohscore aufweisen und somit im höheren latenten Bereich weniger Daten vorhanden sind als im unteren Bereich. Wie man erkennen kann, liegt

die Hauptinformationen der Daten nicht um den Wendepunkt, sondern knapp darunter, dies lässt darauf schließen, dass die beste Diskrimination zwischen Fähigkeitsparametern im unteren Fähigkeitsbereich zu leisten ist. In allen ICC lassen sich hohe Steigerungen im unteren Fähigkeitsbereich und flachere Steigerung im oberen Fähigkeitsbereich verzeichnen. Die Abweichungen zu einer perfekten ICC sind destotrotz nicht schwerwiegend genug, um eine Inhomogenität des Rasch Modells anzunehmen.

Personenparameter

Die Personenparameter werden mithilfe der Itemparameter geschätzt. Jeder Ausprägung des Rohscores wird hier ein eigener Parameter Score zugewiesen. Ein Rohscore von 0 hat einen besonders niedrigen (und negativen) Personenparameter (θ), ein hoher Rohscore hat einen hohen positiven Personenparameter. In Tabelle 9 zeigen sich die Personenparameter im Rasch homogenen Itemset.

Tabelle 9

*Die geschätzten Personenparameter
für die Rasch homogenen Items*

Rohscore	Schätzwert
0	-3.68
1	-2.79
2	-1.97
3	-1.44
4	-1.01
5	-0.65
6	-0.31
7	0.01
8	0.34
9	0.67
10	1.02
11	1.44

Es ergibt sich ein ähnliches Bild, wie es auch schon bei den ICC zu sehen war. Die Testitems weisen eine zu hohe Schwierigkeit auf und deshalb gibt es keine Personen die mehr als 11 Items korrekt beantwortet haben und die Schätzparameter weisen besonders niedrige Werte

auf. Dies spiegelte sich schon aufgrund der großen Konfidenzintervalle über dem Wert 0 in den ICC wieder.

Zur grafischen Veranschaulichung von Personen und Itemparameter soll die Person-Item-Map in Abbildung 30 dienen. Die PI-Map stellt die Personenparameter mit den Itemparameter gegenüber und lässt gut überblicken, wie sich die Verteilung der Personen über die Items verhält.

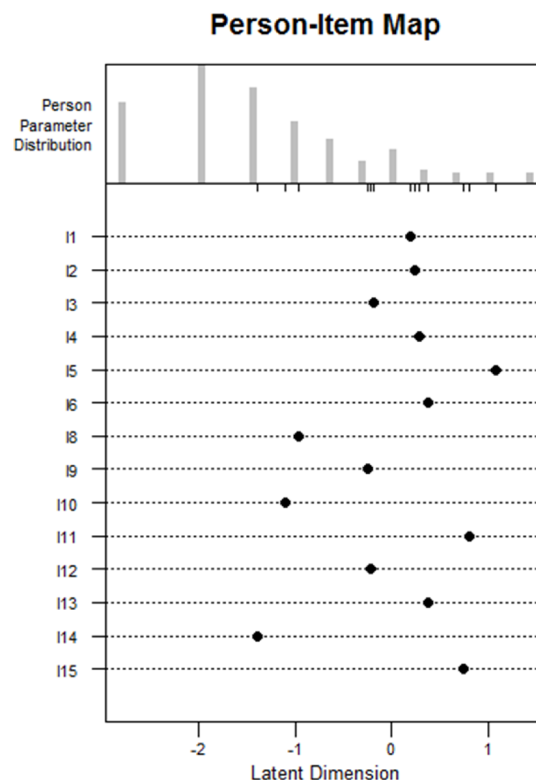


Abbildung 30. Person-Item-Map; Oben = Die Verteilung der Personenparameter; Unten = Die Verteilung der Itemparameter

Wie zu sehen ist, positionieren sich die Personenparameter (auf der Grafik oben) eher links, auf der Seite der niedrigen latenten Dimension, während sich die Itemparameter auf der rechten Seite positionieren, in dem eine hohe latente Fähigkeit zur Lösung der Items benötigt wird. Dies bedeutet, dass die Items eine zu hohe Schwierigkeit haben und die Testpersonen deshalb eine niedrige Leistung erreichen.

Itemüberprüfung: Item 7

Da sich das Item 7 als nicht Rasch konform herausgestellt hat, soll es in Abbildung 31 näher dargestellt werden, um eine mögliche Fehlkonstruktion feststellen zu können.

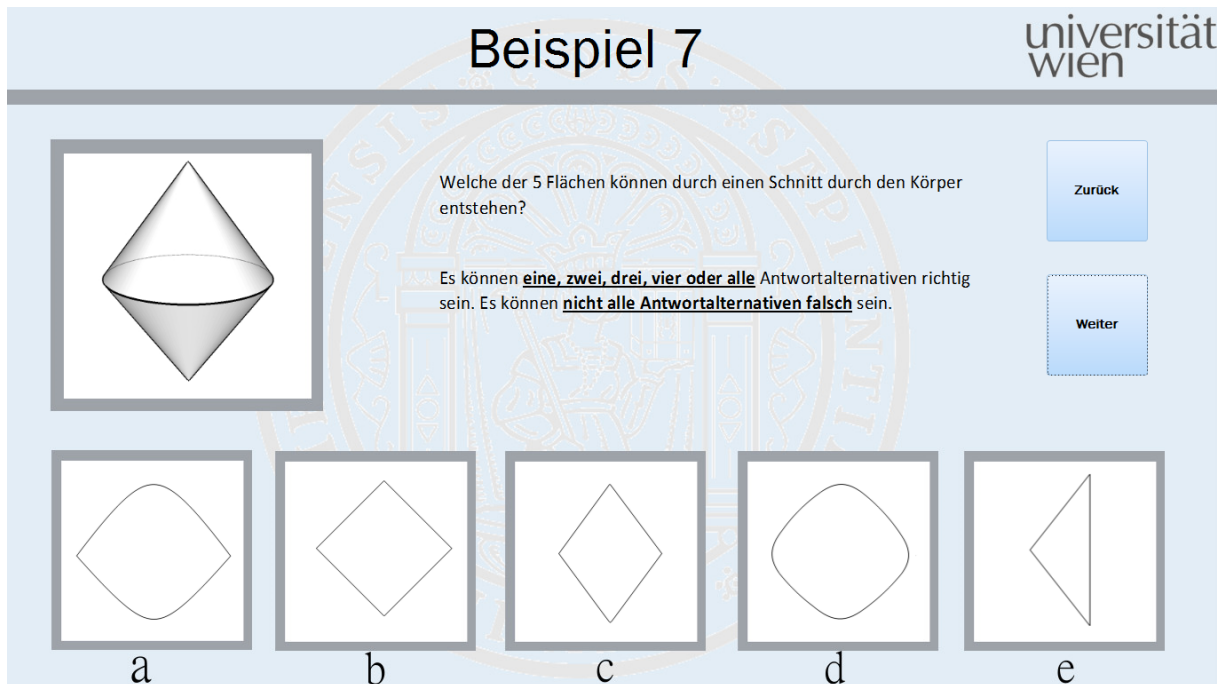


Abbildung 31. Item 7 des 3D-2D Objektschneidetests.

Schon in der Berechnung der Subgruppeninvarianz mittels Median, lässt sich erkennen, dass Item 7 von Personen mit Leistung unterhalb des Median leichter gelöst worden ist, als von Personen oberhalb des Median. Personen mit schlechterer Fähigkeit scheinen also dieses Item besser zu lösen, als Personen mit hoher Fähigkeit. Nach näherer Betrachtung der Antwortalternativen lässt sich erkennen, dass es zwischen Antwortalternative b und c Unstimmigkeiten gibt. Nach Überlegungen könnte dies der Fall sein, da Alternative b und c sehr ähnlich sind und durch die Wahrnehmung des 3D Körpers visuell nicht erkannt werden kann, ob es sich um einen Doppelkegel mit gleicher Höhe und Durchmesser handelt, oder ob der Doppelkegel höher, als breiter ist. Durch diese Wahrnehmungsschwierigkeit könnten Personen eine andere Lösungsstrategie gewählt haben, nämlich die des Ratens, während bessere Personen versucht haben, das Item aufgrund ihrer Fähigkeit zu lösen.

14.2 Hypothese 2 - Geschlechtsunterschiede

H_0 : Es gibt keine Leistungsunterschiede zwischen Frauen und Männern.

H_1 : Es gibt Leistungsunterschiede zwischen Frauen und Männern.

Die Voraussetzungen für die Hypothesenprüfung können angenommen werden. Die Messung weist Intervallskalenniveau auf und die Normalverteilung des untersuchten Merkmals kann in beiden Gruppen angenommen werden. (siehe Abbildung 32)

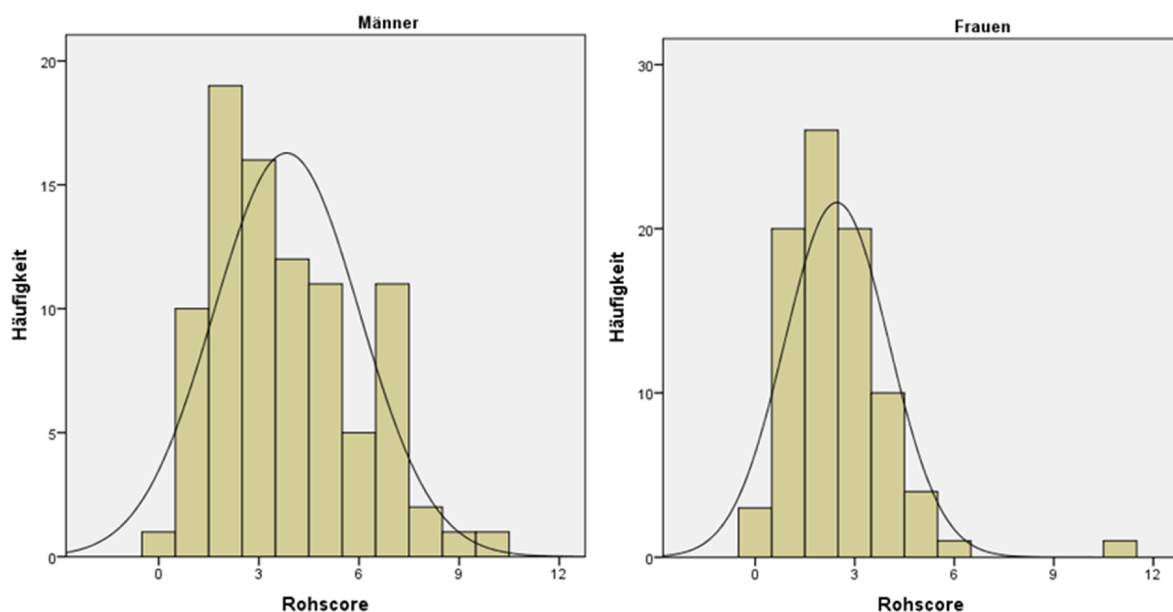


Abbildung 32. Häufigkeitsverteilung des Rohscore für Männer und Frauen.

Männer ($n = 89$) weisen einen durchschnittlichen Rohscore von $M = 3.83$ ($SD = 2.18$) auf und Frauen ($n = 85$) einen durchschnittlichen Rohscore von $M = 2.46$ ($SD = 1.57$). Aufgrund des Antwortformats von n aus 5 (es können nicht alle Antwortalternativen falsch sein) liegt die naive Ratewahrscheinlichkeit bei rund 3.30 %.

Aufgrund der Verletzung der Homogenität der Varianzen, mittels Levene-Test ($F(1, 173) = 14.55, p < .001$), wird der Welch-Test für unabhängige Stichproben mit ungleichen Varianzen berechnet. Das Ergebnis des Welch-Test weist mit $t(89, 85) = 4.78, p < .001$ ein signifikantes Ergebnis auf.

Ausgehend davon, dass verschiedene Berufe, unterschiedlich viel mit Spatial Visualisation gemeinsam haben, und dies Auswirkung auf die Berechnung dieser Hypothese haben kann, soll der Beruf, bzw. Ausbildungszweig konstant gehalten werden und das Geschlecht mit der Leistung korreliert werden. Hierfür wurde die partielle Korrelation gewählt. Wie schon oben gezeigt ist die Normalverteilung für beide Geschlechtsvariablen gegeben und intervallskaliert. Deshalb kann eine partielle Korrelation durchgeführt werden.

Wird der Berufs- und Ausbildungszweig konstant gehalten so zeigt die partielle Korrelation mit $r(168) = -.318, p < .001$ einen mittleren negativen Zusammenhang zwischen dem Geschlecht und der erbrachten Leistung im Test.

14.3 Hypothese 3 - Altersunterschiede

H_0 : Es gibt keinen Altersunterschied zwischen jüngeren und älteren Personen aufgrund ihrer Testleistung.

H_1 : Es gibt Altersunterschiede zwischen jüngeren und älteren Personen aufgrund ihrer Testleistung.

Hierfür wurden die Testpersonen in 3 Gruppen unterteilt, wie in Tabelle 10 zu sehen ist. Da die ersten Defizite in Spatial Visualisation bereits mit circa 25 Jahren beginnen, wurde die erste Gruppe bis zu dieser Grenze definiert. Die ersten beiden Gruppen haben eine Spannweite von 10 Jahren. Da es ab 36 Jahren nur mehr wenige Personen in der Stichprobe gibt, wurden die restlichen Testpersonen in dieser Gruppe zusammengefasst.

Tabelle 10

Altersaufteilung der Stichprobe aufgrund von 3 Altersgruppen

Gruppe	Altersbereich	<i>N</i>	<i>M</i>	<i>SD</i>
1	16-25	58	3.43	2.33
2	26-35	79	2.78	1.64
3	36 +	37	3.54	2.16

Die Voraussetzungen für die Berechnung einer einfachen Varianzanalyse für unabhängigen Stichproben sind erfüllt. In allen 3 Gruppen liegt Intervallskalierung und Normalverteilung (siehe Abbildung 33) vor.

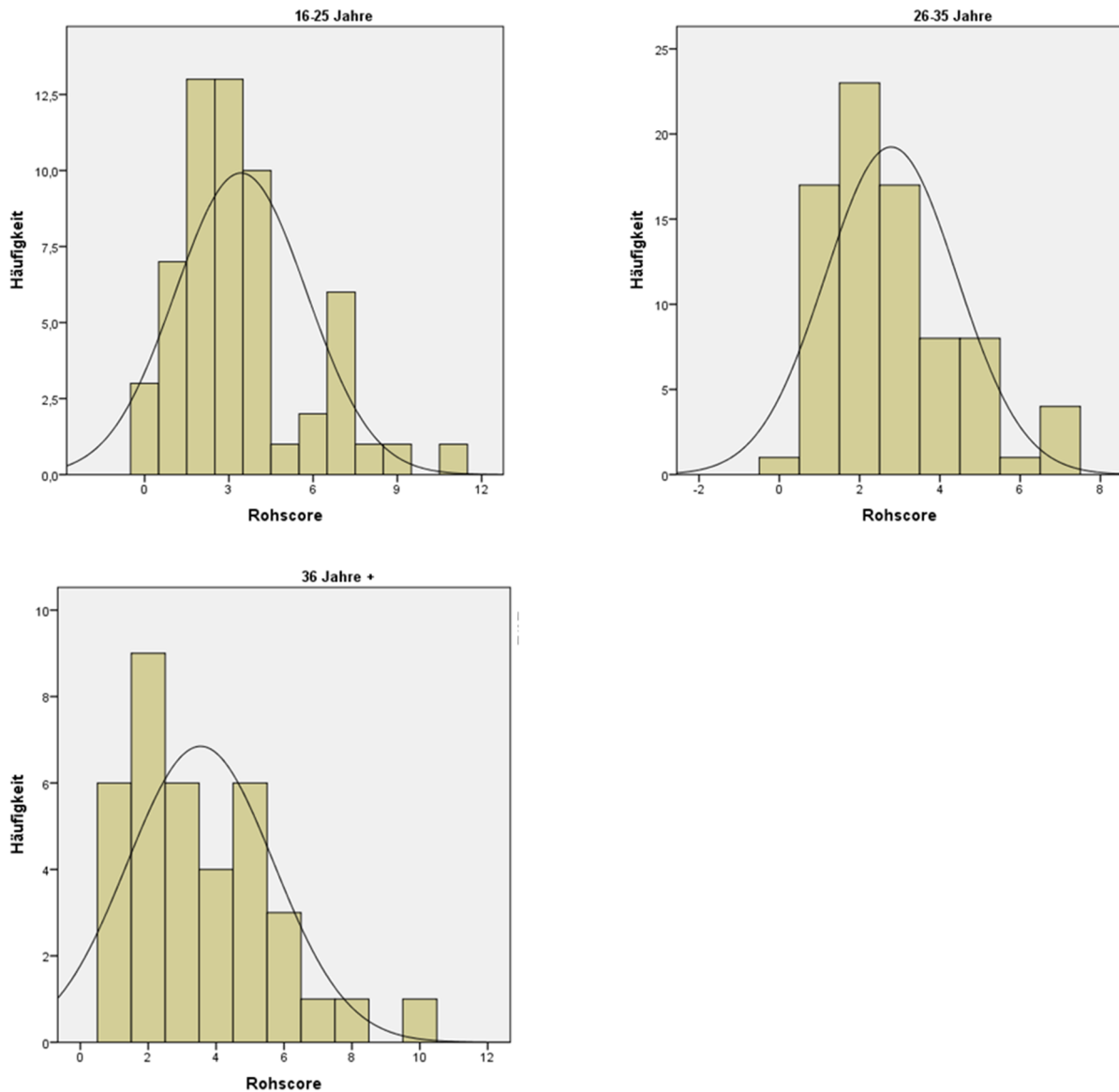


Abbildung 33. Häufigkeitsverteilung des Rohscores in den 3 Altersgruppen.

Nach dem Levene-Test ist die Inhomogenitätsprüfung der Varianzen mit $F(2, 171) = 3.16$, $p = .045$ signifikant. Aufgrund dieser Verletzung soll im Rahmen der einfaktoriellen ANOVA der Welch-Test gerechnet werden. Es lassen sich mit $t(2, 83.64) = 2.75$, $p = .07$ keine signifikanten Mittelwertsunterschiede in den Gruppen finden.

Aufgrund der Unterschiede im Rohscore (Tabelle 10) zwischen der Gruppe 1 (16-25) und Gruppe 2 (25-36) soll noch rein explorativ ein t-Test für unabhängige Stichproben berechnet werden. Angesichts der explorativen Natur der Berechnung soll auch keine Alpha-Korrektur durchgeführt werden. Die Normalverteilung der Daten konnte bereits weiter oben angenommen werden. Der Levene-Test zeigt inhomogene Varianzen ($F(1, 136) = 5.12, p = .025$) und auch der Welch-Test zeigt keine signifikanten Mittelwertsunterschiede zwischen den beiden Gruppen ($t(58, 79) = 1.91, p = .074$).

14.4 Hypothese 4 – Ausbildungszweige

H_0 : Es gibt keine Unterschiede in der Leistung zwischen verschiedenen Berufs- und Ausbildungszweigen.

H_1 : Es gibt Unterschiede in der Leistung zwischen verschiedenen Berufs- und Ausbildungszweigen.

Die Variable zur Prüfung dieser Hypothese wurde aus den Variablen „zur Zeit ausgeführter Beruf“ und „aktuelle Studienrichtung“ zusammengefasst. Aufgrund der geringen Gruppengröße in manchen Untergruppen und der folglich Verletzung der Voraussetzungen zur Verwendung eines parametrischen Verfahrens in der Normalverteilung diesen Untergruppen, wurde der Kruskal-Wallis-Test durchgeführt.

Die Ergebnisse zeigen, dass die Teststatistik mit $\chi^2(9, n = 163) = 8.92, p = .445$ kein signifikantes Ergebnis liefert. Sieht man sich die Mittleren Ränge in Tabelle 11 an, sind die Bereiche IT und Technischer Bereich mit den höchsten Rängen vertreten.

Tabelle 11

Ausbildungszweige mit den mittleren Rängen laut Kruskal-Wallis-Test

Themenblock	N	Mittlerer Rang
Psychologie	23	64.61
Kaufm. Tätigkeit	37	81.24
Sozial / Gesundheit	8	78.38
Betriebswirtschaft	13	90.92
Technischer Bereich	23	97.22
Lehramt	7	65.00
Rechtswissenschaften	8	67.38
IT-Bereich	12	91.25
Sonstige Studienrichtungen	26	87.58
In Ausbildung	6	77.17
Gesamt	163	

14.4.1 Hypothese 4.1 – Vergleich Psychologie - Technik

H₀: Personen im technischen Bereich unterscheiden sich nicht signifikant von Personen im psychologischen Bereich aufgrund ihrer Leistung.

H₁: Personen im technischen Bereich unterscheiden sich signifikant von Personen im psychologischen Bereich aufgrund ihrer Leistung.

Die Voraussetzungen für den t-Test für unabhängige Stichproben sind nicht gegeben. Die Variablen sind zwar intervallskaliert aber die Normalverteilung kann, wie in Abbildung 34 zu sehen ist, im technischen Bereich nicht angenommen werden.

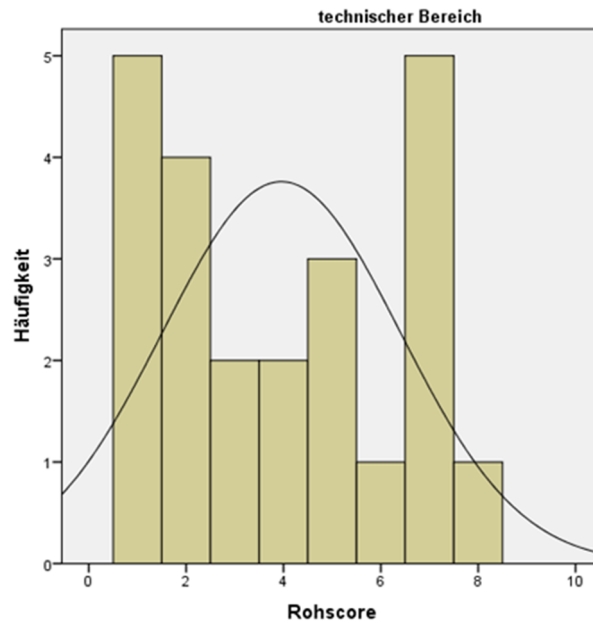


Abbildung 34. Häufigkeitsverteilung des Rohscore im technischen Bereich

Aufgrund der Verletzung der Normalverteilung muss ein nicht parametrischer Test verwendet werden, in diesem Fall der Mann-Whitney-U-Test. Die Voraussetzungen für den U-Test können als erfüllt angenommen werden. Die untersuchten Merkmale sind stetig, die Daten sind mindestens rangskaliert und die Stichproben sind unabhängig voneinander. Tabelle 12 zeigt die Ränge der Themenblöcke Psychologie und technischer Bereich.

Tabelle 12

Ränge für die Themenblöcke Psychologie und technischer Bereich

Themenblock	N	Mittlerer Rang	Rangsumme
Psychologie	23	19.35	445.00
Technischer Bereich	23	27.65	636.00
Gesamt	46		

Der Mann-Whitney-U-Test liefert mit $U(23, 23) = 169$, $p = .033$ ein signifikantes Ergebnis. Es kann also davon ausgegangen werden, dass sich Personen im technischen Bereich von Personen im psychologischen Bereich unterscheiden.

14.5 Hypothese 5 – Höchste abgeschlossene Ausbildung

H₀: Es gibt keine signifikanten Unterschiede in der Leistung zwischen den bisher abgeschlossenen Ausbildungsgruppen.

H₁: Es gibt signifikante Unterschiede in der Leistung zwischen den bisher abgeschlossenen Ausbildungsgruppen.

Auch hier sind die Voraussetzungen nicht erfüllt, da die Normalverteilung in der Gruppe mit universitärem Abschluss nicht angenommen werden kann (Abbildung 35).

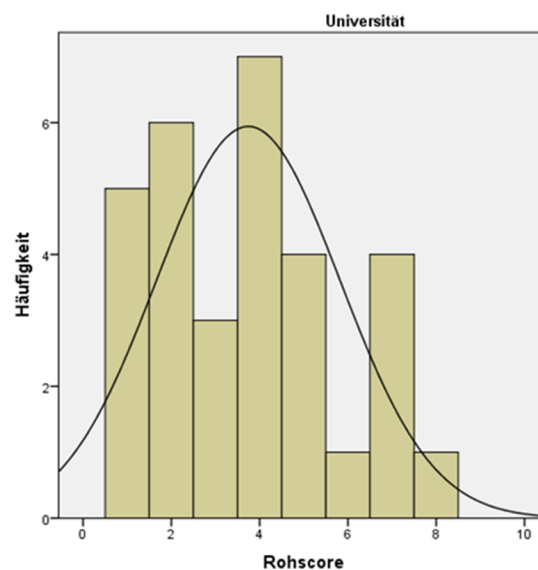


Abbildung 35. Häufigkeitsverteilung des Rohscores der Personen mit universitärem Abschluss.

Deshalb wurde hier der Kruskal-Wallis-Test für die Ausbildungsgruppen durchgeführt. Tabelle 13 zeigt die Verteilung der Testpersonen über die Gruppen, sowie die mittleren Ränge. Nach dem Kruskal-Wallis-Test lassen sich keine signifikanten Unterschiede zwischen den Gruppen annehmen ($\chi^2(3, n=174) = 7.38, p = .061$).

Tabelle 13

Ränge für die höchste abgeschlossene Ausbildung laut Kruskal-Wallis-Test

Höchste abgeschlossene Ausbildung	N	Mittlerer Rang
Pflichtschule	30	99.68
Matura	86	81.27
Universität	31	102.56
Sonstige	27	76.52
Gesamt	174	

Aufgrund des mittleren Rangs der Gruppe Matura und Universität soll explorativ noch ein U-Test für diese beiden Gruppen erfolgen. Es wird auch hier keine Alpha-Korrektur durchgeführt, da sich diese Berechnungen nicht auf Hypothesen stützen und rein explorativ angesehen werden. Der U-Test liefert mit $U(86, 31) = 1001$, $p = .037$ einen signifikanten Unterschied zwischen den beiden Gruppen.

14.6 Hypothese 6 - Selbsteinschätzungsskala

H_0 : Es gibt keine Korrelation zwischen der Einschätzung der eigenen Leistung und der tatsächlichen Leistung.

H_1 : Die Einschätzung der eigenen Leistung korreliert positiv mit der tatsächlichen Leistung.

Die Variable „Einschätzung der eigenen Leistung“ wurde auf einer 6-stufigen Likert Skala kodiert (1 = schlechte Leistung, 6 = gute Leistung). Niedrige Werte bedeuten eine geringe antizipierte Leistung im Test und vice versa.

Da es sich bei der Selbsteinschätzungsskala um eine Ordinalskalierung handelt, wurde eine Rangkorrelation nach Spearman durchgeführt. Die Ergebnisse zeigen, dass sich eine mäßige Korrelation von $r_s(169) = .285$, $p < .001$ ergibt. Dies bedeutet, dass Personen, die sich besser einschätzen, auch eine bessere Leistung im Test erreichen.

15. Diskussion, Ausblick, Limitationen

Wie zu sehen ist, entspricht der 3D-2D Objektschneidetest (nach Ausschluss des Items 7) dem Rasch Modell, auch wenn die Itemschwierigkeiten, bzw. die ICC nicht optimal verlaufen. Die Gültigkeit des Rasch Modells kann aber aufgrund des Falsifikationsprinzips nur a posteriori angenommen werden. Das Falsifikationsprinzip besagt, dass das Rasch Modell nie bewiesen werden kann. Ein Test kann sich dementsprechend nur sensu Popper qualifizieren. Anhand einer Überprüfung des a posteriori gültigen Itemsets an einer neuen unabhängigen Stichprobe kann überprüft werden, ob eine erneute Modellverträglichkeit gegeben ist. (Kubinger & Draxler, 2007)

Aus den hohen Itemschwierigkeitsparametern und den stark negativen Personenparameter lässt sich schließen, dass ein Großteil des Tests eine zu hohe Schwierigkeit aufweist. Auch bei der Datenerhebung hat sich aufgrund der hohen Abbruchquote gezeigt, dass der Test viele Teilnehmer überforderte. Es wurde eine durchschnittliche Abbruchrate von 30 % verzeichnet, wobei mehr Frauen als Männer den Test abbrachen. Viele Testteilnehmer brachen nach der Instruktion bzw. vor den Übungsbeispielen ab. Daraus kann geschlossen werden, dass die Instruktion zu schwer gestaltet wurde oder zu viele Informationen auf einmal präsentiert wurden. Eine Befragung nach Abbruch des Tests fand nicht statt. Für zukünftige Testungen mit dem 3D-2D Objektschneidetest wäre eine standardisierte Befragung der Beweggründe des Abbruchs vom Vorteil. Dies würde helfen den Test ökonomischer und zumutbarer zu gestalten und somit die Abbruchrate zu verringern. Ausgehend davon sollte bei einer erneuten Erhebung, bzw. Umkonstruktion des Tests in Betracht gezogen werden, dass einige Items geändert werden, bzw. eine Stichprobe gewählt wird, die höhere Leistung im Bereich Raumvorstellung vorzeigt. Eine solche Stichprobe könnte aus Personen bestehen, die im beruflichen Umfeld mit Raumvorstellung zu tun haben oder hochbegabten Personen.

Eine andere Alternative wäre die Umkonstruktion der am wenigsten gelösten Items zu leichteren Versionen, um so die globale Testschwierigkeit zu verringern. Dies würde erlauben, genauere Itemparameterschätzwerte zu erreichen und somit bessere Kennwerte in der Berechnung des Rasch Modells zu erzielen.

In der Theorie postulieren Linn & Petersen (1985), dass in Spatial Visualisation keine bis geringe Geschlechtsunterschiede zu finden sind. In dieser Stichprobe finden sich Gruppenunterschiede bzw. eine mäßige Korrelation von $r = -.318, p < .001$. Dies widerspricht damit den Annahmen von Linn & Petersen. Ein Grund für die höheren Geschlechtsunterschiede könnte die Verwendung von verschiedenen Lösungswegen sein, wie es bereits Shepard & Cooper (1986) beschrieben. Da aber angenommen werden muss, dass Personen aufgrund der Komplexität der Aufgaben im Bereich Spatial Visualisation verschiedene Lösungsstrategien ausprobieren und die effektivste verwendet wird, muss ein weiterer Faktor die Geschlechtsunterschiede auslösen. Die Schwierigkeit der Items könnte mit der Wahl der Lösungswege interagieren. Aufgrund der hohen Schwierigkeit und der Undurchschaubarkeit, ob ein Item richtig oder falsch gelöst worden ist, ist es den Testteilnehmern nicht möglich herauszufinden, ob ein Lösungsweg effektiv ist oder nicht. Dadurch bleibt den Testpersonen nichts weiter über, als eine beliebige Strategie zu wählen. Wie Shepard & Cooper (1986) schon erklärten, wählen Personen eine Strategie und bleiben dann während des Tests bei dieser. So könnte es passieren, dass Männer eine effektivere Strategie wählen als Frauen. Eine weitere Möglichkeit besteht darin, dass die Geschlechtsunterschiede aufgrund der Zufallsstichprobe entstanden sind, bzw. dass der Test nicht vollständig Spatial Visualisation misst, sondern auch andere Facetten der Raumvorstellung, was nicht als verwunderlich zu erachten ist, da Spatial Visualisation eine Vielzahl von Operationen verwendet, die auch in den Subfaktoren Mental Rotation und Spatial Perception verwendet werden.

Ob der 3D-2D Objektschneidetest wie bei der Datenerhebung auch im Gruppensetting zu verwenden ist bleibt fraglich. Der Einfluss vom Verhalten von Männern und Frauen während der Testung kann nicht kontrolliert werden. So sind während der Gruppentestung Männer schneller fertig geworden und haben ihren Sitzplatz verlassen, während Frauen noch den Test bearbeiteten. Solche Einflüsse könnten Frauen suggerieren, dass Männer in diesem Test besser sind als Frauen. Solche Suggestionen beeinflussen die Leistung von Frauen, wie (Moè, 2009) in seinen Studien zum gender belief zeigte. Bisher wurde der gender belief nur in Mental Rotation untersucht, doch könnte dieser Effekt auch in Spatial Visualisation auftreten und somit die Geschlechtsdifferenzen im 3D-2D Objektschneidetest erklären. Bevor nicht ausgeschlossen werden kann, dass der gender belief in dieser Studie einen Einfluss auf die

Leistung hat, sind Gruppensettings für zukünftige Testungen mit dem 3D-2D Objektschneidetest nicht zu empfehlen.

Die Theorien zur sexuellen Orientierung (Gladue et al., 1990) und zur Freizeitaktivität (Voyer et al., 2000) konnten im Rahmen der Hypothesenprüfung aufgrund der Testlänge keinen Platz finden. So könnte es aber für zukünftige Testungen mit dem 3D-2D Objektschneidetest interessant sein, ob auch in Spatial Visualisation Leistungsunterschiede zwischen Freizeitaktivitäten zu erkennen sind.

Bei Hypothese 3 zeigt sich auf den ersten Blick, dass im Gegensatz zu Salthouse et al. (1990) keine Alterseffekte zu herrschen scheinen. Hierbei muss angemerkt werden, dass Salthouse et al. eine größere Spannweite des Alters verwendeten. Die hier angeführte Stichprobe verfügt einerseits über keine Gleichverteilung des Alters und andererseits über sehr wenige Testpersonen über 35 Jahren. Wie in den Ergebnissen zu sehen war, zeigten Gruppe 1 (16-25 Jahren) und Gruppe 3 (36+ Jahren) gleich gute Leistungswerte mit durchschnittlich 3.5 Lösungen und Gruppe 2 (26-35 Jahre) geringere Werte mit durchschnittlich 2.8 Lösungen. Dies würde darauf schließen lassen, dass im Alter um 25 Jahren ein Einbruch der Leistung in Spatial Visualisation erfolgt, wobei die Leistung ab 35 Jahren wieder ansteigt. Dies ist nicht nur ein Gegensatz zu der Forschung von Salthouse et al. sondern auch ein Gegensatz zum Defizitmodell der Intelligenz. Eine mögliche Erklärung für dieses Resultat ist, dass die Stichprobe Verzerrungen aufweist. Die beste Leistung wird von der Gruppe der Techniker erreicht (siehe Hypothese 4 & 4.1). Die Verteilung der Techniker ist aber im Alter von 36+ überrepräsentativ. Dies hat zur Auswirkung, dass die Gruppe 3 (36+ Jahre) aufgrund von Stichprobenverzerrungen eine bessere Leistung aufweist, als die anderen Gruppen.

Vergleicht man jedoch Gruppe 1 und 2, in denen diese Stichprobenverzerrung nicht auftritt, lassen sich Tendenzen zu den Alterseffekten finden. Gruppe 1 unterscheidet sich dabei zwar nicht signifikant von Gruppe 2 mit $t(57, 78) = 1.91, p = .074$, dies ist aber der Fall, da die Hypothesen zu Beginn dieser Studie 2-seitig formuliert wurde. Geht man davon aus, dass es nur möglich ist, dass die Leistung über das Alter nachlässt und somit nur eine 1-seitige Formulierung in Frage kommt, wären die gefundenen Ergebnisse signifikant und Alterseffekte in Gruppe 1 und 2 wären vorhanden.

Bei der Überprüfung der Leistungsunterschiede in den Ausbildungszweigen, in Bezug auf die Forschung von Salthouse et al. zeigt sich, dass es global über alle 10 Ausbildungszweige hinweg keine Unterschiede gibt ($\chi^2(9, n = 163) = 8.92, p = .445$). Dies ist bei der großen Anzahl von Ausbildungszweigen nicht verwunderlich, deswegen wurden in Hypothese 3.1 a priori 2 Berufsfelder gewählt, die hohen (Techniker) bzw. wenig (Psychologen) Bezug zum Thema Spatial Visualisation haben. Der Vergleich der beiden Gruppen zeigt, dass die Testpersonen in den Gruppen signifikante Leistungsunterschiede zeigen ($U(23, 23) = 169, p = .033$). Dieses Ergebnis geht mit den Forschungen von Salthouse et al. einher.

In der bisherigen abgeschlossenen Ausbildung lassen sich keine signifikante Unterschiede erkennen, jedoch sind die Mittleren Ränge im Kruskal-Wallis-Test für Personen mit Maturaabschluss und universitärem Abschluss sehr unterschiedlich, weshalb explorativ der Mann-Whitney-U-Test durchgeführt wurde. Dieser zeigt mit $U(86, 31) = 1001, p = .037$ signifikante Unterschiede zwischen den beiden Gruppen. Eine universitäre Ausbildung könnte also mit einer höheren Leistung in Spatial Visualisation einhergehen. Dies könnte verschiedene Gründe haben. Erstens könnten wieder die Funde von Shepard & Cooper (1986) herangezogen werden. Da bei Spatial Visualisation aufgrund der Komplexität verschiedene Lösungswege ausprobiert werden und schließlich der effektivste gewählt wird, könnte aufgrund der besseren Ausbildung ein breiteres Repertoire an Lösungswegen bereitstehen, was in der möglichen Verwendung von effizienteren Strategien resultiert. Ein zweiter Ausgangspunkt ist die Forschung von Workman & Lee, (2004). Sie postulieren neben einem kulturellen Vergleich im Paper Folding Test auch, dass Spatial Visualisation einem Lerneffekt unterliegt, bzw. dass Lösungsstrategien erlernt werden können. So könnten Personen mit universitärem Abschluss Lösungsstrategien erlernt haben, bzw. den Umgang mit bereits gelernten Lösungsstrategien in ihrer Ausbildung vertieft haben, wodurch sie einen Vorteil gegenüber Maturanten haben. Eine weitere Möglichkeit besteht darin, dass Absolventen der Universität gelernt haben ihre Lösungsstrategien rasch zu ändern und zu adaptieren und so während des Tests schneller zum richtigen Lösungsweg finden als Maturanten.

Die Selbsteinschätzung zeigt eine mittlere Korrelation $r_s(169) = .285, p < .001$ mit der Leistung. Personen, die sich schlecht einschätzen, schneiden auch schlecht im Test ab. Jedoch sollte die

Erhebung der Selbsteinschätzung bei erneuter Erhebung nicht nach der Bearbeitung der Übungsbeispiele erfolgen, da hier die Personen schon auf ihre bereits geleistete Fähigkeit Bezug nehmen. Da auch die Übungsbeispiele nicht leicht sind und viele Personen hier schon falsche Antworten geben, schätzen sich viele Personen nach den Übungsbeispielen möglicherweise schlechter ein, als vor der Bearbeitung der Übungsbeispiele.

Die Selbsteinschätzungsskala wurde erstellt, um einen Bezug zur Validität des Tests herzustellen. Geplant war, dass Personen ihre Fähigkeit in Spatial Visualisation einschätzen und diese mit der Leistung verglichen wird. Jedoch kann nicht davon ausgegangen werden, dass die Testpersonen ihre Fähigkeit in Bezug auf Spatial Visualisation einschätzen, sondern vielmehr ihre Fähigkeit diesen Test zu lösen. Wobei noch unklar ist, ob dieser Test auch Spatial Visualisation misst. Dieses Problem entstand dadurch, da die Skala nach den Übungsbeispielen vorgegeben wurde und somit die Selbsteinschätzung mit der bisherigen Leistung in den Übungsbeispielen konfundiert wurde. Deshalb ist es nicht möglich, die Validität aus dieser Skala herzuleiten.

Bei der Selbsteinschätzungsskala muss in zukünftlichen Untersuchungen mit dem 3D-2D Objektschneidetest neben der Position im Test auch Überlegungen getätigt werden, ob die Skala nicht den gender belief (Moè, 2009) auslöst. Es könnte möglich sein, dass neben der Reflektion der eigenen Leistung in Spatial Visualisation auch ein Geschlechtervergleich stattfindet. So könnte der gender belief ausgelöst werden und Frauen schätzen sich schlechter ein als Männer. Das könnte Einfluss auf die erreichte Leistung im Test haben. Da dieser Effekt nicht ausgeschlossen werden kann, muss der Einsatz der Selbsteinschätzungsskala mit Bedacht stattfinden. Es kann nicht davon ausgegangen werden, dass es keine Leistungsunterschiede im Geschlecht aufgrund der Vorgabe der Selbsteinschätzungsskala (infolge des gender beliefs) gibt.

Ein weiterer Ansatz für die Feststellung, dass dieser Test auch wirklich Spatial Visualisation misst, war die Replizierung der Hypothesen aus der Literatur. Dafür wurden die Hypothesen 2, 3 und 4.1 und 5 herangezogen. Aufgrund der Stichprobenverzerrungen (viele raumvorstellungsnahe Berufe im Altersbereich 36+) im oberen Altersbereich konnten die Ergebnisse leider nicht vollständig repliziert werden, jedoch stimmen einige post-hoc

Ergebnisse mit der bisherigen Forschung überein. So konnten Alterseffekte (explorative Berechnung) zwischen den Gruppen 16-25 Jahre und 36+ Jahren gefunden werden, sowie Leistungsunterschiede zwischen Ausbildungszweigen. Aufgrund der Ähnlichkeit des Tests zu anderen 3D-2D Verfahren und Anlehnung an den Test Schnitte, der von sich behauptet Spatial Visualisation zu messen und der teilweisen Replizierbarkeit der Hypothesen, kann davon ausgegangen werden, dass dieser Test auch Spatial Visualisation misst, wobei es unsicher ist, ob nicht Facetten der anderen Subbereiche der Raumvorstellung Einzug finden.

Zusätzlich besteht noch die Möglichkeit, anhand der Methode des lauten Denkens zu erheben, welche Denkvorgänge während der Lösung der Testaufgaben Verwendung finden. Diese Methode fand auch schon bei der Validitätsprüfung im Test „Schnitte“ (Fay & Quaiser-Pohl, 1999) seinen Platz. So könnte anhand dieser Methode festgestellt werden, ob Denkprozesse verwendet werden, die für Spatial Visualisation charakterisierend sind. Die Methode des lauten Denkens könnte auch dazu verwendet werden, um die verwendeten Lösungsstrategien zu identifizieren. Dies könnte Aufschluss darüber geben, ob die Anwendung von spezifischen Strategien im Zusammenhang mit dem Geschlecht oder von Berufsgruppen steht.

In der aktuellen Auflage des 3D-2D Objektschneidetests ist keine Zeitkomponente vorgesehen. Die Testzeiten wurden auch nicht protokolliert. Es könnte aber von Vorteil sein, die individuellen Zeiten pro Testperson zu notieren um diese bei der Hypothesenprüfung konstant zu halten. Das könnte Aufschluss darüber geben ob bestimmte Ausbildungs- oder Berufsgruppen den Test schneller beantworten als andere. Es könnte auch ein Zusammenhang zwischen der erbrachten Leistung und Dauer erschlossen werden. Es ist durchaus plausibel, dass Personengruppen, die sich selbst schlecht in Raumvorstellung einschätzen, den Test schnell und ohne viel Sorgfalt bearbeiten und deshalb einen niedrigeren Rohscore erzielen. Aktuell konnte diese Komponente nicht erfasst werden. Für zukünftige Studien mit dem 3D-2D Objektschneidetest ist deshalb eine Erfassung der Testzeiten zu empfehlen, wobei darauf zu achten ist, dass das Verfahren auf keine maximale Testdauer beschränkt wird. Es soll sich beim 3D-2D Objektschneidetest um einen reinen Power-Test halten. Ein Einfluss einer Speed-Komponente könnte die Rasch homogenität des Tests beeinflussen.

Bei zukünftigen Itemkonstruktionen zur Erweiterung des Itemsets ist darauf zu achten, dass Objekte und Körper visuell klar zu unterscheiden sind. So hat die Analyse von Item 7 gezeigt, dass ähnliche Figuren die Lösbarkeit der Aufgabe beeinflusst. So ist es auch wichtig, dass die Lösungsalternativen klar unterscheidbar sind. Es sollte darauf verzichtet werden, ähnliche Schnittflächen (zB. Variationen von Rechtecken) zu verwenden. In Kombination mit komplexen Körpern kann die Präsentation von gleichartigen Schnittflächen die Aufgabe um ein vielfaches erschweren. Eine klare Unterscheidung der Schnittflächen ist deshalb grundlegend für die Konstruktion zukünftiger Aufgaben.

Abschließend kann gesagt werden, dass die Rasch konforme Konstruktion des 3D-2D Objektschneidetests gelungen ist und so ein Erstentwurf für einen a posteriori Rasch homogenen Test im Bereich Spatial Visualisation entstanden ist. Die vorgeschlagenen Verbesserungen zur Verringerung der Abbruchraten, Schwierigkeit des Tests und zum Testsetting könnten den 3D-2D Objektschneidetest zu einem wertvollen Beitrag für die aktuelle Forschung der Raumvorstellung machen. So gibt es nun neben den Subfaktoren Mental Rotation, Spatial Relation und Spatial Orientation nun auch einen Rasch homogenen Test zur Messung von Spatial Visualisation, mithilfe dessen die Forschung in diesem Bereich weiter ausgeweitet werden kann.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Neukonstruktion und testtheoretischen Untersuchung des Raumvorstellungstests „3D-2D Objektschneidetest“, mit dem Ziel den Subfaktor Spatial Visualisation zu messen. In Anlehnung an den Test „Schnitte“ von Fay & Quaiser-Pohl (1999) wurden 15 neue Items computerbasiert generiert, die anhand einer Selbsteinschätzungsskala und theoriegeleiteten Hypothesen validiert wurden. Der 3D-2D Objektschneidetest wurde anhand einer repräsentativen Stichprobe von 174 Personen, die aus 88 Männern und 86 Frauen, im Alter von 16-65 Jahren bestand, mittels Rasch Modell überprüft. Dabei zeigte sich nach Ausschluss des Items 7 und der Berechnung des Andersen-LRT in allen Teilungskriterien ein Rasch homogenes Bild. Die Hypothesen zu den Geschlechtsdifferenzen zeigten einen positiven signifikanten Effekt in Richtung der Männer. Bei den Alterseffekten zeigten sich keine signifikanten Unterschiede zwischen den Altersgruppen.

Personen in einem Ausbildungszweig, der mit Raumvorstellung in Bezug steht (Technik, Informationstechnologie, ...), erzielten eine bessere Leistung, als Personen in einem raumvorstellungsfremden Ausbildungszweig (Betriebswirtschaftslehre, Psychologie, ...). Personen mit universitärem Abschluss zeigten im Vergleich zu Maturanten bessere Leistungen. Die Selbsteinschätzungsskala der eigenen Fähigkeit zur Raumvorstellung zeigt einen mäßigen Zusammenhang mit der tatsächlich erreichten Leistung im Test. Die Validierung des Rasch konformen Tests kann angenommen werden, wobei nicht alle Hypothesen repliziert werden konnte und nachträglich Limitierungen im Validierungsmodell gefunden wurden.

Abstract

This study deals with the construction and test theoretic research of the „3D-2D Objektschneidetest“, with the focus on the sub factor Spatial Visualisation. 15 new items got created on the basis of the Test „Schnitte“ by Fay & Quaiser-Pohl (1999). These items got validated by a self-assessment scale of anticipated performance and theory based hypothesizes. The Rasch Model was used for the analysis of the data of the „3D-2D Objektschneidetest“. The data includes 174 person, consisting 88 men and 86 women, between 16 and 65 years. After excluding Item 7 from the item set, the Andersen-LRT showed homogeneity of the remaining data in all split-criteria. The analysis of the gender differences presents a positive significant effect towards men. There were no significant differences in age.

Comparing groups with jobs with high affinity to spatial abilities (engineering, information technology) and groups with jobs with low affinity to spatial abilities (psychology, business economics) showed a better performance towards the high affinity group. Individuals with university degree displayed better performance than high-school graduates. The self-assessment scale of the anticipated Spatial Visualisation performance presents a medium correlation with the actual achieved performance. It can be concluded that the „3D-2D Objektschneidetest“ measures Spatial Visualisation, although not all hypothesis could be replicated, which limits the power of the used validation model.

Literaturverzeichnis

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123–140.
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor referenced cognitive tests*. Princeton: Educational Testing Service.
- El Koussy, A. A. H. (1935). *An investigation into the factors in tests involving the visual perception of space* (Vol. 20). California: The University Press.
- Fay, E., & Quaiser-Pohl, C. (1999). *“Schnitte” - Ein Test zur Erfassung des räumlichen Vorstellungsvermögens*. Frankfurt/M: Swets Test Services.
- French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors (revised 1963)*. Princeton, N. J.: Educational Testing Service.
- Gittler, G. (1986). Inhaltliche Aspekte bei der Itemselektion nach dem Modell von Rasch. *Zeitschrift Für Experimentelle Und Angewandte Psychologie*, 33(3), 386–412.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest (3DW) – Ein Rasch-skaliertes Test zur Messung des räumlichen Vorstellungsvermögens*. Weinheim: Beltz Test GmbH.
- Gittler, G., & Arendasy, M. (2003). Endlosschleifen: Psychometrische Grundlagen des Aufgabentyps EP. *Diagnostica*, 49(4), 164–175.
- Gladue, B. A., Beatty, W. W., Larson, J., & Staton, R. D. (1990). Sexual orientation and spatial ability in men and women. *Psychobiology*, 18(1), 101–108.
- Grössenbrunner, P. (1994). *Pilot's Spatial Test (PST). Software und Manual*. Mödling: Dr. G. Schuhfried GmbH.
- Guilford, J. P., & Lacey, J. I. (1947). *Printed classification tests* (pp. 269–295, 477–509). DTIC Document. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=AD0651781>

- Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: similarities and differences between mental rotation and mental folding. *Cognitive Processing*, 14(2), 105–115.
- Koller, I., Alexandrowicz, R., & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: eine Einführung in eRm*. UTB.
- Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Hogrefe Verlag.
- Kubinger, K. D., & Draxler, C. (2007). Probleme bei der Testkonstruktion nach dem Rasch-Modell. *Diagnostica*, 53(3), 131–143.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479–1498.
- Lohman, D. F. (1979). *Spatial Ability: A Review and Reanalysis of the Correlational Literature*. DTIC Document. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA075972>
- Lord, F. M. (1980). In *Applications of item response theory to practical testing problems* (pp. 44–63). Routledge.
- Maier, P. H. (1999). *Räumliches Vorstellungsvermögen: Ein theoretischer Abriss des Phänomens räumliches Vorstellungsvermögen. Mit didaktischen Hinweisen f. d. Unterricht*. Auer Verlag in Der Aap, Lehrfachverlage GmbH.
- Michael, W. B., Guilford, J., Fruchter, B., & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement*, 17, 185–199.
- Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, 19(1), 21–27.

- 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(1), 39–58.
- Piaget, J., & Inhelder, B. (1967). *The child's conception of space*. New York: W. W. Norton and Company, Inc.
- Quaiser-Pohl, C. (2003). The Mental Cutting Test “Schnitte” and the Picture Rotation Test-Two New Measures to Assess Spatial Ability. *International Journal of Testing*, 3(3), 219–231.
- Reischies, F. M., & Lindenberger, U. (2010). Grenzen und Potentiale kognitiver Leistungsfähigkeit im Alter. In *Die Berliner Altersstudie* (pp. 375–401). Akademie Verlag.
- Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R., & Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, 26(1), 128–136.
- Shepard, R. N., & Cooper, L. A. (1986). *Mental images and their transformations*. Cambridge, MA, US: The MIT Press.
- Shepard, R. N., & Metzler, J. (2004). Mental rotation of three-dimensional objects. In *Cognitive Psychology: Key Readings* (pp. 254–257). New York: Psychology Press.
- Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201–292.
- Thurstone, L. L. (1949). *MECHANICAL APTITUDE III ANALYSIS OF GROUP TESTS*. Chicago: DTIC Document. Retrieved from <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=AD0490040>
- Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, 2, 94.

- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., & Warren, C. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402.
- Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles*, 43(11-12), 891–915.
- Werdelin, I. (1961). *Geometrical ability and the space factors in boys and girls*. Oxford, England: U. Lund.
- Witkin, H. A. (1950). Perception of the upright when the direction of the force acting on the body is changed. *Journal of Experimental Psychology*, 40(1), 93.
- Witkin, H., Oltman, P., Raskin, E., & Karp, S. (1971). *Manual for embedded figures test, children's embedded figures test, and group embedded figures test*. Palo Alto, CA: Consulting Psychologists Press.
- Workman, J. E., & Lee, S.-H. (2004). A cross-cultural comparison of the apparel spatial visualization test and paper folding test. *Clothing and Textiles Research Journal*, 22(2), 22–30.

Abbildungsverzeichnis

Abbildung 1. Die Komponenten des räumlichen Denkens nach Maier (1999)	10
Abbildung 2. Mental Rotation Test von Vandenberg & Kuse (1978). Auszug aus Peters et al. (1995)	11
Abbildung 3. Embedded Figures Task nach Witkin, Oltman, Raskin, & Karp (1971)	15
Abbildung 4. Kit of Reference Tests for Cognitive Factors – Formboard Vz-1 nach French et al. (1963)	16
Abbildung 5. Surface Development Test; Auszug aus Manual for Kit of Factor- Referenced Cognitive Tests nach Ekstrom, French, Harman, & Dermen (1976)	17
Abbildung 6. Kit of Reference Tests for Cognitive Factors – Paperfolding: Vz-2 nach French et al. (1963)	17
Abbildung 7. Kit of Reference Tests for Cognitive Factors – Cube Comparison Test S-2 nach French et al. (1963)	18
Abbildung 8. Water Level-Task nach Piaget & Inhelder (1967)	18
Abbildung 9. Piaget’s Mountain Task nach (Piaget & Inhelder, 1967)	19
Abbildung 10. Item aus dem Test „Schnitte“ (Quaiser-Pohl, 2003)	21
Abbildung 11. Alterseffekte nach Salthause et al. (1990) samt Regressionskoeffizienten pro Test	30
Abbildung 12. Screenshot aus dem Programm Sketch Up 2013	34
Abbildung 13. Instruktionsseite 1 des 3D-2D Objektschneidetests	35
Abbildung 14. Instruktionsseite 2 des 3D-2D Objektschneidetests	36
Abbildung 15. Instruktionsseite 3 des 3D-2D Objektschneidetests	37
Abbildung 16. Instruktionsseite 4 des 3D-2D Objektschneidetests	37
Abbildung 17. Instruktionsseite 5 des 3D-2D Objektschneidetests	38
Abbildung 18. Übungsbeispiel 1 des 3D-2D Objektschneidetests	39
Abbildung 19. Verkleinerter Ausschnitt der Rückmeldung des 3D-2D Objektschneidetests	39
Abbildung 20. Übungsbeispiel 1 des 3D-2D Objektschneidetests nach falscher Beantwortung nach dem 2. Durchlauf	40

Abbildung 21. Zwischenseite vor der Vorgabe der Testitems des 3D-2D Objektschneidetests	41
Abbildung 22. Auszug aus dem Item 5 mit dem Distraktor in der Mitte und der richtigen Antwort rechts	43
Abbildung 23. Vorgehen für eine Überprüfung auf das Rasch Modell nach Koller et al. (2012)	48
Abbildung 24. Häufigkeitsverteilung des Rohscores in der Gesamtstichprobe	50
Abbildung 25. Grafische Darstellung des LRT mit Teilungskriterium Median	53
Abbildung 26. ICC für Item 1-4	57
Abbildung 27. ICC für Item 5,6, 8 und 9	57
Abbildung 28. ICC für Item 10-13	57
Abbildung 29. ICC für Item 14 und 15	57
Abbildung 30. Person-Item-Map	59
Abbildung 31. Item 7 des 3D-2D Objektschneidetests	60
Abbildung 32. Häufigkeitsverteilung des Rohscore für Männer und Frauen	61
Abbildung 33. Häufigkeitsverteilung des Rohscores in den 3 Altersgruppen	63
Abbildung 34. Häufigkeitsverteilung des Rohscore im technischen Bereich	66
Abbildung 35. Häufigkeitsverteilung des Rohscores der Personen mit universitärem Abschluss	67

Tabellenverzeichnis

Tabelle 1. Kombination der Dimensionen Static – Dynamic und Intrinsic - Extrinsic nach Uttal et al. (2013)	14
Tabelle 2. Soziodemografische Verteilung der Stichprobe	44
Tabelle 3. Die erstellten Themenblöcke aus den Variablen aktueller Beruf und aktuelle Studienrichtung	45
Tabelle 4. LRT für die 4 Teilungskriterien	51
Tabelle 5. Waldtest für das Teilungskriterium Median	52
Tabelle 6. LRT für die 4 Teilungskriterien ohne Item 7	54
Tabelle 7. Lokal stochastische Unabhängigkeit (LSU) für alle Items	55
Tabelle 8. Itemleichtigkeitsparameter (β) der Rasch homogenen Items	56
Tabelle 9. Die geschätzten Personenparameter für die Rasch homogenen Items	58
Tabelle 10. Altersaufteilung der Stichprobe aufgrund von 3 Altersgruppen	62
Tabelle 11. Ausbildungszweige mit den mittleren Rängen laut Kruskal-Wallis-Test	65
Tabelle 12. Ränge für die Themenblöcke Psychologie und technischer Bereich	66
Tabelle 13. Ränge für die höchste abgeschlossene Ausbildung laut Kruskal-Wallis-Test	68

Hiermit bestätige ich, dass ich die vorliegende Arbeit in allen relevanten Teilen selbstständig durchgeführt habe.

Wien, am 13.11.2015

Unterschrift:

Thomas Halmetschlager

Lebenslauf

Persönliche Daten

Name:	Thomas HALMETSCHLAGER
Geburtsdaten:	11.08.1986 in Gmünd
Familienstand:	Ledig
Staatsbürgerschaft:	Österreich

Bildungsweg

1992 – 1996	Volksschule in Gmünd
1996 – 2001	Bundesrealgymnasium in Gmünd
2001 – 2006	Bundeshandelsakademie in Gmünd
Seit 2008	Psychologiestudium an der Universität Wien

Zivildienst

Juli 2006 – März 2007	Rotes Kreuz Gmünd
-----------------------	-------------------

Berufstätigkeit

Seit 2011	Büroangestellter in Gmünd
-----------	---------------------------

Psychologisches Praktikum

Jänner – März 2015	Praktikum bei datenwerk in Wien
--------------------	---------------------------------