



universität  
wien

# Diplomarbeit

Raumvorstellung aus Sicht des TARV-F 4.0 und EST.  
Eine Validierung anhand von Extremgruppen

Ana Grubescic

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.<sup>a</sup> rer. nat.)

Wien, im November 2010

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.-Prof. Dr. Mag. Klaus D. Kubinger



## **Danksagung**

Mein größter Dank gebührt Univ.-Prof. Dr. Mag. Klaus D. Kubinger für die Betreuung, ohne den die Umsetzung dieser Arbeit nicht möglich gewesen wäre. Weiters geht mein Dank an Mag. Lisbeth Weitensfelder und Ao. Univ.-Prof. Dr. Georg Gittler, die mich durch ihre Bereitstellung der angewandten Tests und ihr motivierendes und fachliches Engagement bei der Arbeit unterstützt haben.

Mein Dank richtet sich auch an die Austro Control Mitarbeiter Bernhard Söllner und Mag. Markus Winkler sowie das JA Flight Training Center für ihre organisatorische Unterstützung.

Ebenso gilt mein Dank meiner Familie, im Besonderen meiner Schwester Mag. Ivona Grubescic, und meinen Freunden, die mich immer wieder neu motiviert und vorangetrieben haben.



## Abstract

Die vorliegende Arbeit beschäftigt sich mit der Validierung der Kurzversion eines neuen Raumvorstellungstests: *Test zur Angewandten Raumvorstellung (TARV*; Weitensfelder, unpubl.). Inwiefern dieser in der Lage ist, die von ihm postulierten Fähigkeiten zu messen, wurde versucht mittels Extremgruppenvalidierung zu beantworten. Zur Bildung von Extremgruppen wurde zum einen die Fähigkeit zur Raumvorstellung als ein Kriterium herangezogen (wonach Personen aus dem Bereich der Luftfahrt eine bessere Raumvorstellungsfähigkeit aufweisen, als jene Personen die sich in einem geisteswissenschaftlichen Studium oder Psychologiestudium befinden) und zum anderen die erzielten Testkennwerte in einem anderen Raumvorstellungstest; dem Endlosschleifentest (EST; Gittler & Arendasy, unpubl.). In Anbetracht der hier vorliegenden Ergebnisse, kann die Extremgruppenvalidierung dahingehend als gelungen gesehen werden, als der TARV-F 4.0 in der Lage ist zwischen Extremgruppen auch tatsächlich zu unterscheiden. Die zugrundeliegenden Hypothesen, wonach die mittels Extremgruppenselektion (EST-Gesamtrohscore) gebildeten Gruppen, sich hinsichtlich ihrer Testwerte in der jeweils interessierenden Facette des TARV-F 4.0 (*Rotation* und *Orientierung*; wonach Personen die gute Ergebnisse im EST erzielen auch im TARV gut sind) unterscheiden, konnte nicht bestätigt werden.

The research at hand focused on the validation of a short version of a new spatial ability test: *Test zur Angewandten Raumvorstellung (TARV*; Weitensfelder, unpubl.). The intention of this research was to find out whether or not the TARV measures spatial ability by using extreme groups. The extreme groups were created by two criterions: first, the spatial ability in everyday life (people who work in aviation domains, whom we implied to have good spatial ability, were chosen on the one extreme and on the other extreme subjects studying humanities or psychology were recruited). The second criterion was the test score of another spatial ability test, namely Endlosschleifentest (EST; Gittler & Arendasy, unpubl.). The results of the current research support the hypothesis that the TARV-F 4.0 is able to differ between everyday life extreme groups. The hypothesis that the TARV-F 4.0 – especially the mental rotation and spatial orientation – is able to differ between subjects with very high scores and very low scores in the EST (a good score at the EST implies a good score at the interested domain of TARV-F 4.0 ) could not be verified.



# Inhaltsverzeichnis

1	Einleitung .....	9
<b>I Theoretischer Teil .....</b>		<b>11</b>
2	Kognitive Fähigkeit (cognitive ability) .....	13
3	Die Fähigkeit zur „Raumvorstellung“ .....	14
3.1	Definitionsversuche.....	14
3.2	Raumvorstellung als eine Komponente der Intelligenz .....	16
3.2.1	Spearman's Generalfaktor- Intelligenztheorie.....	18
3.2.2	Primärfaktoren nach Thurstone (primary mental abilities).....	19
3.2.3	Das Modell nach Lohman (1979).....	21
3.2.4	Linn und Petersen (1985) .....	22
3.2.5	Aktueller Forschungsstand .....	24
4	Raumvorstellung und das Phänomen des Geschlechtunterschiedes .....	27
4.1	Metaanalyse nach Linn und Petersen (1985).....	28
4.2	Metaanalyse nach Voyer, Voyer und Bryden (1995).....	30
4.3	Forschungsstand zum Geschlechterunterschied aus heutiger Sicht.....	32
5	Validität psychologisch- diagnostischer Verfahren .....	38
5.1	Inhaltsvalidität .....	38
5.2	Kriteriumsvalidität.....	38
5.3	Konstruktvalidität.....	39
6	Raumvorstellung und Beruf.....	40
6.1	Tätigkeitsbeschreibung eines Fluglotsen .....	41
6.2	Tätigkeitsbeschreibung eines Piloten.....	42
<b>II Empirischer Teil .....</b>		<b>45</b>
7	Fragestellung und Hypothesen .....	47
8	Statistische Methode .....	50
8.1	Optimaler Stichprobenumfang .....	50
9	Methodik und Studiendesign .....	51
9.1	Verfahren zur Erfassung der räumlichen Vorstellungsfähigkeit.....	51

9.1.1	Test zur Angewandten Raumvorstellung (TARV) .....	51
9.1.2	Endlosschleifentest (EST).....	54
9.1.3	Raumvorstellung aus Sicht des EST und TARV .....	57
9.2	Ablauf der Studie.....	58
10	Stichprobe .....	61
10.1	Stichproben zur Validierung von Extremgruppen.....	62
10.1.1	Extremgruppenbildung nach der Selbsteinschätzung .....	65
10.1.2	Extremgruppenbildung nach dem EST-Gesamtrohwert.....	67
11	Ergebnisdarstellung.....	68
11.1	<i>Luftfahrt</i> versus <i>Geisteswissenschaft</i> .....	68
11.2	Selbsteinschätzung zur räumlichen Vorstellungsfähigkeit.....	71
11.3	Facette <i>Orientierung</i> .....	73
11.4	Facette <i>Rotation</i> .....	73
11.5	Einfluss des Geschlechts auf die Leistung im EST und TARV-F 4.0 .....	74
11.5.1	Geschlechtsunterschiede im EST .....	74
11.5.2	Geschlechtsunterschiede im TARV-F 4.0 .....	76
12	Zusammenfassung und Diskussion.....	81
	<b>III Literaturverzeichnis .....</b>	<b>87</b>
	<b>IV Abbildungsverzeichnis.....</b>	<b>95</b>
	<b>V Tabellenverzeichnis .....</b>	<b>96</b>
	<b>VI SPSS-Output-Verzeichnis .....</b>	<b>97</b>
	<b>VII Anhang.....</b>	<b>99</b>
	<b>Lebenslauf .....</b>	<b>109</b>

# 1 Einleitung

Raumvorstellung etablierte sich im Rahmen der Intelligenzforschung als ein wichtiger Aspekt der allgemeinen kognitiven Fähigkeiten (vgl. Kubinger, 2009) und wird nach Gitter und Arendasy (2003) „zu den gesicherten Breichen (Primärfaktoren) menschlicher Intelligenz“ (S. 164) gezählt.

Dennoch kann trotz jahrzehntelanger Forschung keine klare Definition von Raumvorstellung gegeben werden.

In der Literatur findet sich eine Vielzahl an Definitionsversuchen, so beispielsweise beschreibt Lohman (1996) diese, als die Fähigkeit visuelle Informationen zu erzeugen, zu verarbeiten, zu behalten und zu transformieren. Einigkeit besteht jedoch darin, dass es sich um kein „Einheitskonstrukt“ handelt, sondern um einen Fähigkeitsbereich, welcher sich aus mehreren Aspekten räumlichen Vorstellungsvermögens zusammensetzt. Dennoch, „welche und wie viele Faktoren notwendig sind, um die dimensionale Binnenstruktur dieses Fähigkeitsbereichs adäquat zu beschreiben“ (Gittler & Arendasy, 2003, S. 164) scheint bis heute noch unklar zu sein.

An der Test- und Beratungsstelle des Arbeitsbereichs Psychologische Diagnostik der Universität Wien wurde ein neuer, ein sich noch in der Entwicklungsphase befindlicher, facettenorientierter Raumvorstellungstest entwickelt –Test zur Angewandten Raumvorstellung (TARV; Weitensfelder, unpubl.). Während sich viele der gängigen Raumvorstellungstests auf die menschliche Rotationsfähigkeit beziehen (z.B.: Dreidimensionaler Würfeltest, 3DW; Gittler, 1990) deutet bereits die Bezeichnung „facettenorientierter Test“ beim TARV darauf hin, dass dieser nicht nur die mentale Rotation, sondern auch andere sogenannte Facetten der Raumvorstellung – nämlich Relationen und Orientierung – zu erfassen vermag.

Inwiefern dieser jedoch in der Lage ist, die von ihm postulierten Fähigkeiten zu messen, soll im Rahmen der vorliegenden Arbeit mittels Extremgruppenvalidierung, in Anbetracht an das psychologische Gütekriterium *Validität*, geklärt werden. Diesem zufolge gilt ein Test dann als valide, wenn der besagte Test auch tatsächlich die Eigenschaft misst, welche er vorgibt zu erfassen.

Zwei Fragen sollen im Rahmen der vorliegenden Studie beantwortet werden. Zum einen die Frage, inwiefern der Test zur Angewandten Raumvorstellung fähig ist

zwischen Extremgruppen zu unterscheiden und zum anderen inwiefern die im TARV gemessene Auffassung von Raumvorstellung mit jener eines anderen Raumvorstellungstests zusammenhängt; beispielsweise dem, auf den Schlauchfiguren (Stumpf & Fay, 1983) aufbauenden Endlosschleifentest (EST; Gittler & Arendasy, unpubl.).

# I Theoretischer Teil



## 2 Kognitive Fähigkeit (cognitive ability)

“GENERAL ABILITY is concerned with all sorts of tasks, but especially those of a cognitive or intellectual sort.” (English & English, 1958; zit.n. Carroll, 1993, S. 4).

Dieser Definition zufolge befasst sich die allgemeine kognitive Fähigkeit (Intelligenz) mit unterschiedlichen, vor allem aber mit kognitiven oder intellektuellen, Aufgaben. In Bezug darauf führen die Autoren weiter an, dass es sich im Vergleich zur allgemeinen Fähigkeit bei den sogenannten „speziellen Fähigkeiten“ um Aufgabenmaterial handelt, welches sich auf bestimmte Fähigkeitsaspekte bezieht:

„Each special ability should, when possible, be so defined as not to overlap with other special abilities.“ (English & English, 1958; zit.n. Carroll, 1993, S. 4).

Hunter (1986) beschreibt die allgemeine kognitive Fähigkeit wie folgt:

General cognitive ability in the industrial psychology literature means cognitive ability as it has been developed in adult workers or job applicants. General cognitive ability is usually measured by summing across tests of several specific aptitudes, usually verbal aptitude, quantitative aptitude, and sometimes technical aptitude. A typical measure would sum across a vocabulary test, an arithmetic reasoning test, and a test of 3-dimensional spatial patterns. (S. 341)

Hülshager und Maier (2008) argumentieren:

Die menschliche Intelligenz ist hierarchisch strukturiert, wobei der g-Faktor allgemeine kognitive Fähigkeiten an der Spitze dieser Hierarchie steht. Spezifische kognitive Fähigkeiten sind die einzelnen Komponenten, die allgemeiner kognitiver Fähigkeit unterliegen und eng umgrenzte Fähigkeitsbereiche umschreiben, beispielsweise verbales Verständnis, Wortflüssigkeit, Umgang mit Zahlen, Raumvorstellung, Gedächtnis, Wahrnehmungsgeschwindigkeit und schlussfolgerndes Denken. Die

spezifischen Fähigkeiten weisen eine unterschiedliche Nähe zum g-Faktor auf. (Hülshager & Maier, 2008, S. 115-116)

### **3 Die Fähigkeit zur „Raumvorstellung“**

Die Auseinandersetzung mit dem Faktor Raumvorstellung fand ihren Ursprung in Studien zu mechanisch-praktischen Leistungen (Quaiser-Pohl, 1998; zit.n. Stepan, 2004, S. 8).

Eine Prüfung der Raumvorstellung im Zusammenhang mit Intelligenz-Testbatterien ist kaum mehr wegzudenken; unabhängig davon ob „als räumliches Vorstellungsvermögen, als „Raumvorstellung“, als Fähigkeit zur Vorstellung räumlicher Relationen, „Raum-Lage-Orientierung“ oder, gleich englisch, als *Spatial Ability* bezeichnet, die entsprechende Fähigkeit wird regelmäßig als eine wichtige Komponente von „Intelligenz“ [...] abgeprüft.“ (Kubinger, 2009, S. 199-200; Hervorhebungen aus dem Original übernommen).

Glück (1999) betont, dass allein „das Wort „Raumvorstellung“ [...] problematisch [sei], weil es eine viel engere Definition andeutet als das englische *space* bzw. *spatial ability*“ (S. 17; Hervorhebungen aus dem Original übernommen). Die Bezeichnung *spatial ability* würde sich der Autorin zufolge nur auf die Fähigkeit zum Umgang mit räumlichen Aufgaben beziehen.

Auch Voyer, Voyer und Bryden (1995) führten in ihrer Metaanalyse an, dass zunächst klar sein muss „whether spatial ability is unitary concept or involves a number of diverse components.“ (S. 251).

In diesem Sinne soll zunächst ein kurzer Überblick zu den Definitionsversuchen und in weiterer Folge zu den zugrundeliegenden Intelligenztheorien gegeben werden, um diese von Voyer et al. (1995) gestellte „Anforderung“ begründen zu können.

#### **3.1 Definitionsversuche**

Nach Thurstone und Thurstone (1941) bezieht sich Raumvorstellung auf Aufgaben, welche „räumliches Vorstellen und Orientieren sowie das Erkennen von

Objekten unter anderem Bezugswinkel erfordern“ (Amelang & Bartussek, 2006, S. 181).

Linn und Petersen (1985) definieren in ihrer Metaanalyse Raumvorstellung als Fähigkeit zur Darstellung, zur Transformation, zur Erzeugung und zum Wiederaufruf bildhafter Informationen und führen weiter an, dass diese als eine wichtige Komponente intellektueller Fähigkeit anzusehen sei:

It is generally agreed that spatial ability is an important component of intellectual ability, yet its nature remains to be clarified. Activities as disparate as perception of horizontality, mental rotation of objects, and location of simple figures within complex figures have all been referred to as measures of spatial ability. No consensus exists for categorization of measures of spatial ability, although many different schemes have been presented. The only view achieving agreement is that spatial ability involves multiple processes. (Linn & Petersen, 1985, S. 1479)

Lohman (1994) definiert diese „as the ability to generate, retain, retrieve, and transform well-structured visual images.“ (zit.n. Glück, 1999, S. 18; vgl. Lohman, 1996).

Glück (1999) zufolge handelt es sich hierbei „um die Fähigkeit, mentale Repräsentationen visuell-räumlicher Information zu erzeugen und intern zu manipulieren oder transformieren.“ (S.18).

Colom, Contreras, Botella und Santacreu (2001) fassen die diversen Definitionsversuche zusammen und sehen Raumvorstellung als eine Fähigkeit, um visuelle Bilder zu erzeugen, zu speichern, abzurufen und zu transformieren.

Eliot (2002)<sup>1</sup> bezieht sich weiters auch auf die Unterscheidung zwischen räumlicher Intelligenz und Raumvorstellung und argumentiert wie folgt:

---

<sup>1</sup> Zur räumlichen Intelligenz vgl. Eliot (2002).

Whereas spatial ability typically refers to one of several domains of human abilities within factor structures, [...] spatial intelligence refers to a very broad range of capacities which are associated with our awareness of the relational distribution of things and our ability to use that awareness to solve 'spatial' problems. Although many assume that spatial ability is the same as spatial intelligence, our existing tests of spatial ability represent only fragments of spatial intelligence (Eliot & Smith, 1983). (Eliot, 2002, S. 479-480)

### **3.2 Raumvorstellung als eine Komponente der Intelligenz**

Die Anfänge der Forschung sahen Raumvorstellung als ein eigenes Konstrukt an, welches vor allem im Rahmen der Intelligenztheorien untersucht wurde.

El-Koussy (1935) wies als erster die Existenz des Faktors „Raumvorstellung“ im Zusammenhang mit der Intelligenz auf. Ihm gelang es durch Herausparsialisierung des g-Faktors, den sogenannten k-Faktor zu extrahieren (vgl. Eliot, 1987; Stepan, 2004).

Um den Umfang der vorliegenden Arbeit nicht zu sprengen, wird im Folgenden gezielt auf Theorien eingegangen, welche für diese relevant sind.

Zuvor soll aber noch ein kurzer Einblick in die verschiedenen Forschungsperspektiven gegeben werden, welche nach Gittler (1992) als eine Orientierungshilfe zu betrachten sind.

Linn und Petersen (1985) zufolge können die meisten der bisher durchgeführten Studien aus vier verschiedenen Forschungsperspektiven betrachtet werden:

(a) the *differential* perspective, involving comparison of spatial ability for different populations (such as males and females); (b) the *psychometric* perspective, involving comparison of correlations between different spatial tasks in order to define “factors” in spatial ability; (c) the *cognitive* perspective, involving the identification of the processes used universally to solve a particular spatial ability task, albeit with quantitatively different efficiency; and (d) the *strategic* perspective, involving identification of the qualitatively different strategies used to solve a given spatial ability task by

different respondents. (Linn & Petersen, 1985, S. 1480; Hervorhebungen aus dem Original übernommen).

Eliot (1987) spricht von drei Forschungsansätzen: Zunächst führt er den differentialpsychologischen bzw. psychometrischen Ansatz an, welcher sich auf die individuellen Differenzen zugrundeliegender Intelligenzkomponente bezieht. Dabei wurde ein besonderes Augenmerk auf die Wahrnehmungsfähigkeit gerichtet. Nach Spearman's Entdeckung des g-Faktors bemühte sich die Forschung um die Aufdeckung weiterer, von diesem Faktor unabhängiger Dimensionen; das räumliche Vorstellungsvermögen stand dabei im Vordergrund.

Als zweiten Ansatz beschreibt er die entwicklungspsychologische Sicht, wonach hier die entwicklungspezifische Aufgabenverarbeitung im Vordergrund steht. Der dritte Forschungsansatz widmet sich der experimentalpsychologischen Sichtweise. Diese geht davon aus, dass unser Verständnis bzw. Wissen durch die Auseinandersetzung mit der physikalischen Welt entsteht (vgl. Gittler, 1992).

Aufgrund jahrzehntelanger Forschungstradition auf dem Gebiet der Raumvorstellung versuchten Eliot und Smith (1983) einen historischen Überblick<sup>2</sup> (zum aktuellen Forschungsstand siehe Abschnitt 3.2.5) zum Verständnis von „spatial abilities“ in drei Phasen wiederzugeben, nach welchen auch die nachfolgenden Modelle einzuordnen sind:

In the first phase (1904-1938), researchers investigated the evidence for and against the existence of a spatial factor over and above a general factor of intelligence. In the second phase (1938-1961), they attempted to ascertain the extent to which spatial factors differed from one another. And in the most recent phase (1961-1982), researchers have attempted to designate the status of spatial abilities within the complex interrelationship of other abilities, and to examine a number of sources of variance which affect performance on spatial tests. (Eliot & Smith, 1983, S. 1; vgl. auch Carroll, 1993).

---

<sup>2</sup> Zum historischen Hintergrund von Raumvorstellung vgl. auch Eliot (1987; 2002).

Im Weiteren sollen die verschiedenen Modelle und Forschungsansätze aus psychometrischer Sicht dargestellt werden, welche sich nach Eliot (2002) mit Fragen auseinandersetzen, wie beispielsweise “whether 'space' could be quantified, whether it exists as a unitary trait or as several unrelated abilities, whether it is associated with particular patterns of abilities, and whether predictions can be made from certain tests.” (S. 482).

Zugleich soll auch ein Teil des historischen Hintergrunds wiedergegeben werden.

### 3.2.1 Spearman's Generalfaktor- Intelligenztheorie

Nach Eliot (1987) befasste sich die erste Forschungsphase (1904-1938) neben dem sogenannten Generalfaktor der Intelligenz zusätzlich mit der Suche nach einem „spatial factor“ (S. 38).

Spearman gelang es 1904 mit Hilfe von Korrelationsstudien ein Intelligenzmodell aufzustellen, wonach sich jede Variable (Intelligenzmaß) auf zwei Faktoren zurückführen lässt, einem Generalfaktor sowie einem „*specific factor unique to that variable*“ (Carroll, 1993, S.39; vgl. Eliot, 1987).

Spearman ging demnach von einem Generalfaktor (g-Faktor) aus, welcher intellektuellen Leistungen zugrunde liegt; wohingegen „the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others“ (Spearman, 1904, zit.n. Rost, 1976, S. 26).

Die Existenz eines sogenannten Gruppenfaktors einer spezifischen Fähigkeit (beispielsweise Raumvorstellung), würde Spearman (1904) zufolge „disturb the pattern of intercorrelations among tests so that the requirements of the Tetrad-difference equation would not be satisfied.“ (zit.n. Eliot, 1987, S. 40).

Schließlich führte er drei mentale Prozesse an, welche auf den Generalfaktor g zurückzuführen sind: Zum einen geht Spearman von der Fähigkeit zur Erfassung von Erfahrungen aus und zum anderen von Fähigkeiten, durch welche eine Person in der Lage ist, Relationen sowie Zusammenhänge zwischen den Reizen ableiten bzw. herstellen zu können:

The first of these is „the apprehension of experience“. The other two [...] are the “eduction of relations” and the “eduction of correlates,” where the word *eduction* means the drawing out of some logical abstraction or consequence from two or more stimuli. (Carroll, 1993, S. 53; Hervorhebungen aus dem Original übernommen)

Spearman's Modell war jedoch nicht im Stande die Existenz eines Gruppenfaktors zur Raumvorstellungsfähigkeit aufzudecken.

Dies gelang 1935 El-Koussy, indem er aus Spearman's g-Faktor einen nach ihm benannten k-Faktor extrahieren konnte (vgl. Stepan, 2004; Eliot, 1987):

There is no evidence for group factor running through the whole field of spatial perception...Spatial tests are primarily tests of (G). But some spatial tests involve a group factor over and above their (G) content. This group factor, called the (K) factor, receives a ready explanation in terms of visual imagery. (El-Koussy, 1935; zit.n. Eliot, 1987, S.43)

El-Koussy zufolge repräsentiert dieser die Fähigkeit einer Person sich visuell-räumliche Bilder einzuholen und diese anzuwenden (vgl. Eliot & Smith, 1983; Eliot, 1987).

### 3.2.2 Primärfaktoren nach Thurstone (primary mental abilities)

Im Vergleich zu Spearman spricht Thurstone von gleichgestellten Faktoren der Intelligenz, welche “in verschiedenem Ausmaß für einzelne Leistungen von Bedeutung sind” (Stepan, 2004, S. 9). Er widmete sich nicht der Korrelationsanalyse wie Spearman es tat, sondern bezog sich auf die Faktorenanalyse (vgl. Eliot, 1987).

Anhand der Anwendung des Verfahrens zur Multiplen Faktorenanalyse gelang Thurstone 1938 die Unterscheidung von sieben voneinander unabhängigen Faktoren der Intelligenz, welche unter der Bezeichnung „Primary Mental Abilities“ (im Deutschen als „Primärfaktoren“) Bekanntheit erlangten (vgl. Carroll, 1993; Eliot, 1987; Stepan, 2004; Winkler, 2007):

1. verbal comprehension (v): Wortverständnis, sprachliches Verständnis

2. word fluency (w): Wortflüssigkeit
3. number (n): Rechenfertigkeiten, rechnerisches Denken
4. perceptual speed (p): Wahrnehmungsgeschwindigkeit
5. memory (m): Gedächtnis, Merkfähigkeit
6. reasoning (r): logisches Denken, schlussfolgerndes Denken
7. space (s): Raumvorstellung, räumliches Vorstellungsvermögen

Während die Entdeckung der von Thurstone postulierten sieben Primärfaktoren für das Ende der ersten Phase stand, führte die Identifikation des Faktors „space“ als eigenständiger Primärfaktor zum Beginn der, nach Eliot und Smith (1983) bezeichneten, zweiten historischen Phase.

Nach Thurstone und Thurstone (1941) umfasst der s-Faktor (space) die „Bewältigung von Aufgaben, die räumliches Vorstellen und Orientieren sowie das Erkennen von Objekten unter anderem Bezugswinkel erfordern“ (Amelang & Bartussek, 2006, S. 181).

Tests, welche für diesen Faktor charakteristisch sind, seien Thurstone und Thurstone (1941) zufolge eine „facility in holding a mental image and mentally twisting, turning, or rotating it to a different position and then matching this transformed image with a suggested solution“ (S. 21; zit.n. Eliot, 1987, S. 45).

Erfasst wird also die Fähigkeit einer Person, sich Objekte zwei- oder gar dreidimensional vorzustellen bzw. diese räumlich zu erkennen (vgl. Winkler, 2007).

Einige Jahre später fügte Thurstone 1950 den Faktor „space“ drei Subfaktoren hinzu (vgl. Guilford, 1972; Rost, 1979; Stepan, 2004; Winkler, 2007):

- a) *Spatial relations* (S1): Hierunter versteht Thurstone die Fähigkeit räumliche Konfigurationen und Beziehungen von Objekten und deren Teilen zu erfassen und diese aus unterschiedlichen Perspektiven zu erkennen: „an ability to visualize a rigid configuration when it is moved into different positions.“ (Eliot, 1987, S. 49).
- b) *Visualization* (S2): Diesen Subfaktor beschreibt Thurstone als die Fähigkeit „to visualize a configuration in which there is movement or displacement of the several parts of the configuration.“ (Guilford, 1972, S. 133); hierbei

handelt es sich um dynamische Denkvorgänge, wobei diese bei „spatial relation“ (oder kurz S1) statisch sind (vgl. Stepan, 2004).

- c) *Spatial orientation* (S3): Diese Dimension beschreibt die Fähigkeit des eigenen Zurechtfindens einer Person im Raum bzw. in einer räumlichen Situation (vgl. Rost, 1979).

Mit der Entdeckung des Generalfaktors durch El Koussy (1935) und Thurstones Modell wurden wegweisende Schritte im Bereich der Raumvorstellungsforschung gesetzt, welche bis heute immer wieder ihren Weg in Forschungsarbeiten finden; nachfolgend sollen Arbeiten vorgestellt werden, welche auf diesen beiden, vor allem aber auf Thurstones Forschung aufbauen.

### 3.2.3 Das Modell nach Lohman (1979)

Vertreter der dritten Phase (1961-1982), so auch Lohman, strebten nach der Erforschung von Ursachen, welche für die differenzierte Leistung in der Raumvorstellung vorantwortlich sein könnten (vgl. Eliot, 1987); die Geschlechtsunterschiede schienen dabei von zentralem Interesse zu sein.

Das Modell von Lohman (1979; vgl.; Eliot, 1987; Carroll, 1993; Glück, 1999; Stepan, 2004), basierend auf faktorenanalytischen und multidimensionalen Analysen, beschreibt drei Hauptfaktoren (major spatial factors), welche an die von Thurstone postulierte Unterteilung erinnern, mit diesen jedoch wenig Übereinstimmung finden. Lohman und Kyllonen (1983) definieren drei sogenannte Hauptfaktoren der Raumvorstellung wie folgt (vgl. auch Eliot, 1987):

1. Spatial Relations. This factor is defined by tests such as Cards, Flags, and Figures (Thurstone, 1983b). These tests are all parallel forms of one another, and the factor emerges only if these or highly similar tests are included in the battery. Although mental rotation is the common element, the factor probably does not represent speed of mental rotation; rather, it represents the ability to solve such problems quickly, by whatever means.
2. Spatial Orientation. This factor appears to involve the ability to imagine how a stimulus array will appear from another perspective. In the true spatial orientation test, the subjects must imagine that they are reoriented in space,

and then make some judgment about the situation. There is often a left-right discrimination component in these tasks, but this discrimination must be made from the imagined perspective. However, the factor is difficult to measure since tests designed to tap it are often solved by mentally rotating the array rather than reorienting an imagined self.

3. Visualization. The factor is represented by a wide variety of tests such as Paper Folding, Form Board, Surface Development, Hidden Figures, Copping, and so forth [...]. The tests that load on this factor, in addition to their spatial-figural content, share two important features: they are all administered under relatively unspeeded conditions, and most are much more complex than corresponding tests that load on the center of a two dimensional scaling representation, and are often quite close to tests of Spearman's *g* (such as Raven Matrices or Figure Classification) or Cattell's (1963) *Gf*.

(Lohman & Kyllonen, 1983; zit.n. Glück, 1999, S. 20; vgl. auch Carroll, 1993)

Neben Lohman und Thurstone nehmen auch Linn und Petersen (1985) in ihrer Metaanalyse die nun mehrmals beschriebenen drei Hauptfaktoren räumlicher Fähigkeit (räumliche Wahrnehmung, mentale Rotation und räumliche Visualisierung) als Bezugspunkt ihrer Arbeit.

#### 3.2.4 Linn und Petersen (1985)

Das Hauptaugenmerk der Metaanalyse nach Linn und Petersen liegt vor allem auf den Geschlechtsunterschieden im Bereich des räumlichen Vorstellungsvermögens, auf welches im Verlauf der vorliegenden Arbeit eingegangen und daher hier nicht weiter erörtert wird.

In die Analyse gingen Studien und Tagungsergebnisse ein, welche zwischen 1974 und 1982, zum Thema *spatial ability* publiziert wurden<sup>3</sup>.

Für diese Arbeit von Bedeutung ist jedoch die Sichtweise der Autoren in Bezug auf die Struktur des Faktors Raumvorstellung. Insgesamt 172 von 200 Effektgrößen gingen in die Metaanalyse ein und wurden den drei Faktoren, welche aus der

---

<sup>3</sup> Genaueres zur Studienselektion nachzulesen bei Linn und Petersen (1985).

psychometrischen Perspektive entsprungen sind, zugeteilt – räumliche Wahrnehmung (spatial perception), mentale Rotation (mental rotation) und räumliche Visualisierung (spatial visualization).

Linn und Pertersen (1985) beschrieben diese wie folgt:

„In spatial perception tests, subjects are required to determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information.“ (Linn & Petersen, 1985, S.1482).

Die mentale Rotation wird als die Fähigkeit beschrieben, zwei- oder dreidimensionale Objekte nicht nur korrekt, sondern auch rasch in die Zielansicht zu drehen; wobei die Lösungsgeschwindigkeit im Vordergrund steht.

Unter räumlicher Visualisierung verstehen die Autoren “those spatial ability tasks that involve complicated, multistep manipulations of spatially presented information. These tasks may involve the processes required for spatial perception and mental rotations but are distinguished by the possibility of multiple solution strategies” (S.1484); hierunter fallen beispielsweise die Papier-Faltaufgaben (vgl. Abbildung 3.1).

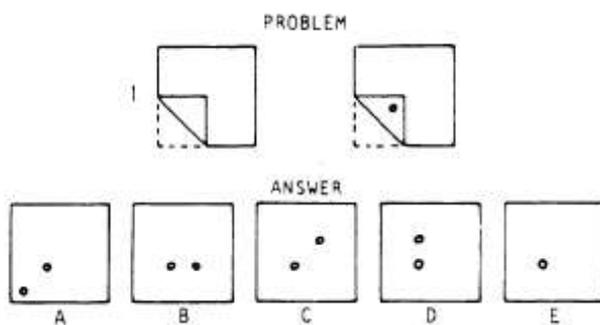


Abbildung 3.1: Papier-Falt-Aufgabe (entnommen aus Linn & Petersen, 1985, S. 1485)

Zusammenfassend kann davon ausgegangen werden, dass trotz einer Vielzahl postulierter Faktoren der Raumvorstellung – wie beispielsweise Carroll (1993), welcher sich auf fünf Hauptfaktoren bezieht – diese sich nach Colom, Contreras, Botella und Santacreu (2001) auf drei bekanntesten Faktoren reduzieren lassen:

Vz, SR, and SO are among the most prominent factors. Vz, or visualization, is defined by complex spatial tests such Paper folding or Surface development. SR, or spatial relations, is defined by the speed in manipulating simple visual patterns such as mental rotations. SO, or spatial

orientation, can sometimes be distinguished. SO tests require the subject to imagine how an array would appear from a different perspective and then to make a judgement from that imagined perspective. SO is not always recognized as a separate factor (Carroll). (Colom et al., 2001, S. 904)

### 3.2.5 Aktueller Forschungsstand

Auf die Wichtigkeit und Rolle räumlicher Intelligenz in unserem alltäglichen Tun macht Eliot (2002) aufmerksam:

Spatial intelligence is fundamental in the sense that it may operate at any given moment at several levels of human consciousness and, in combination with other cognitive functions, it may also contribute both to how we learn to process different kinds of information and how we develop strategies for solving many types of everyday problems. It is pervasive in the sense that we use it whenever we act upon information about the positional, distance, and directional relationships of people and objects in our daily lives. (S. 479)

Ihm zufolge sei räumliche Intelligenz eine Mischung aus „physical capacities, coordinated movement, learned skills, mental processes, forms of representations, and dimensions of thought at different levels of awareness for various tasks in changing surroundings.“ (Eliot, 2002, S. 485).

Die Auseinandersetzung mit dem Fähigkeitsbereich Raumvorstellung sei laut Eliot (2002) dahingehend interessant, wenn dieser anhand jener aus der Theorie und den verschiedenen Modellen generierten Hypothesen untersucht, in beobachtbaren und nachvollziehbaren Bedingungen getestet sowie in quantifizierbaren Antworten und Verhaltensmustern mit Hilfe von Computersimulationen hervorgebracht werden kann.

Dies erklärt wohl auch, warum immer wieder neue Verfahren erstellt werden, welche dieses komplexe Konstrukt möglichst, in vollem Ausmaß erfassen sollen. Auch der Einsatz des Mediums „Computer“ scheint heutzutage immer beliebter, nicht nur aufgrund des ökonomischen Nutzens, sondern auch aufgrund der immer komplexeren und vielfältigen Darstellungen von Objekten.

So gingen beispielsweise Colom, Contreras, Shih und Santacreu (2003) im Rahmen ihrer Studien der Effizienz eines neuen Verfahrens zur Messung der

Raumvorstellung nach, dem sogenannten „Spatial Orientation Dynamic Test“ (SODT-R; Santacreu, 1999; vgl. Abbildung 3.2).

Hierbei handelt es sich um einen computergenerierten Raumvorstellungstest „in which the person is required to simultaneously orient two moving points to a given destination that change from trial to trial“ (Colom et al., 2003, S. 92).

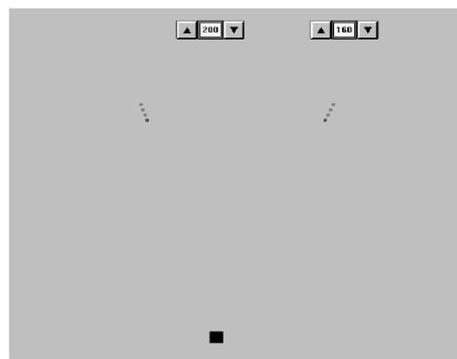


Abbildung 3.2: Spatial Orientation Dynamic Test (SODT)-  
entnommen aus Colom et al. 2003 (S. 94).

Im Konkreten gingen sie dem Nachweis von einzelnen Raumvorstellungsfaktoren, wie beispielsweise die von ihnen gewählten Faktoren wie Visualisierung (Vz), räumliche Relation (SR) und „dynamic spatial performance“ (DSP; Colom et al., 2001, S. 904) nach– beziehungsweise inwiefern „general spatial ability (Gv) can be measured reasonably well through a single test“ (Colom et al., 2003, S. 93).

Zum Einsatz kamen neben Papier-Bleistift-Verfahren auch computerisierte Testverfahren; Colom et al. (2001) zufolge kann DSP nur anhand dieser erfasst werden, mit der Begründung die Bezeichnung *Dynamic* beziehe sich auf die Vorhersage „*where* a moving object is going and *when* it will arrive at its predicted destination“ (Colom et al., 2001, S. 904; Hervorhebungen auf dem Original übernommen).

Der Stichprobe bestehend aus insgesamt 105 Psychologiestudenten ( $n_{\text{♂}} = 16$  versus  $n_{\text{♀}} = 89$ ) wurde eine Testbatterie<sup>4</sup>, welche sich aus insgesamt elf Verfahren (entsprechend den Konstrukten Vz, SR und DSP) zur Messung der Raumvorstellung zusammensetzte, vorgegeben.

Das Ergebnis der hierarchischen Faktorenanalyse deutet darauf hin, dass die Trennung der einzelnen Konstrukte, in diesem Fall SR, Vz und DSP, nicht

---

<sup>4</sup> Genauere Informationen zur Testbatterie nachzulesen bei Colom et al. (2001, 2003).

unproblematisch sei (vgl. Colom et al., 2001). Weiters kamen Colom et al. (2003) zu dem Schluss, wonach „SODT-R seems to be a test especially appropriate to assess general spatial ability [...]“ (S.97).

Dennoch, der Großteil der aktuellen Forschungsarbeiten scheint vermehrt an den Geschlechtsunterschieden in der räumlichen Fähigkeit interessiert zu sein. Eine Vielzahl von Untersuchungen widmet sich dieser Thematik (vgl. Linn & Petersen, 1985; Voyer, Voyer & Bryden, 1995). Der Fokus vieler empirischer Studien liegt dabei vor allem auf dem Aspekt der mentalen Rotation (z. B. Glück & Fabrizii, 2010; Neubauer, Bergner & Schatz, 2010). Im nachfolgenden Kapitel 4 wird auf diese Thematik näher eingegangen werden.

Zunächst soll noch kurz auf die Problematik zur Separierbarkeit von Raumvorstellung und Reasoning eingegangen werden, welche in Zusammenhang mit Papier-Bleistift-Verfahren zur Erfassung räumlicher Vorstellungsfähigkeit aufkam.

### *3.2.5.1 Abgrenzung von Raumvorstellung gegenüber Reasoning*

Raumvorstellungstests in sogenannter „printed“ Form, also jene Verfahren, welche als Papier-Bleistift-Verfahren vorgegeben werden, scheinen Kubinger (2009) zufolge dahingehend problematisch zu sein, „dass sie Gefahr laufen, mehr schlussfolgerndes Denken [...] als tatsächlich räumliches Vorstellungsvermögen zu prüfen“ (S. 200).

Reasoning wird als die „Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“ (Kubinger, 2009, S. 206) definiert.

In diesem Zusammenhang präsentierte Gittler (1999) zwei voneinander unabhängig durchgeführte Studien<sup>5</sup> bei denen „die Abgrenzung von Raumvorstellung gegenüber schlussfolgerndem Denken (Reasoning)“ (S. 70) behandelt wurde.

Gittler kam zu folgendem Ergebnis:

---

<sup>5</sup> Zum Nachweis der Separierbarkeit zwischen Raumvorstellung und Reasoning vgl. Gittler (1999).

Bei gemeinsamer Verrechnung der Rasch-homogenen Subskalen - 3DW als Prototyp für "räumliches Vorstellungsvermögen" und WMT als Prototyp für "schlußfolgerndes Denken" - konnte im Rahmen der probabilistischen Testtheorie mittels geeigneter Modelltests der Nachweis der Separierbarkeit der zugrundeliegenden latenten Fähigkeitsdimensionen erbracht werden. (Gittler, 1999, S. 79)

#### **4 Raumvorstellung und das Phänomen des Geschlechtsunterschiedes**

Das Phänomen des Geschlechtsunterschiedes wirft im Rahmen der Raumvorstellungsforschung immer wieder Fragen auf, und es wird insbesondere versucht diese Unterschiede so gut wie möglich zu unterbinden bzw. zu reduzieren. Nach Rost (1976) zeigen sich geschlechtsspezifische Leistungsunterschiede bereits ab dem 6. Lebensjahr, dies zugunsten der Jungen.

Linn und Petersen setzten sich 1985 mit dieser Thematik im Rahmen einer Metaanalyse auseinander; zehn Jahre später folgten Voyer, Voyer und Bryden (1995). Ziel einer Metaanalyse ist es, zu einer spezifischen Thematik die Ergebnisse mehrere Untersuchungen zusammenzufassen, um in weiterer Folge eine komprimierte Fassung des aktuellen Forschungsstandes wiedergeben zu können.

Auch Glück (1999) widmete sich möglichen zugrundeliegenden Ursachen individueller (geschlechtsspezifischer) Unterschiede im räumlichen Vorstellungsvermögen und bezieht sich dabei auf biologische und psychosoziale Erklärungsmodelle. Im Rahmen biologischer Ansätze wird von physiologischen, wohingegen bei psychosozialen Theorien von erlernten bzw. anerzogenen Determinanten gesprochen wird. Auf diese beiden Ansätze wird hier nicht näher eingegangen – der interessierte Leser findet eine ausführliche Darstellung dieser Ansätze bei Glück (1999) und Rost (1976).

Nachfolgend wird auf die beiden Metaanalysen nach Voyer et al. (1995) sowie Linn und Petersen (1985) eingegangen; im Rahmen neuester Untersuchungen

sollen weitere Erklärungsversuche zu geschlechtsspezifischen Leistungsunterschieden in der Raumvorstellung hervorgehoben werden.

#### 4.1 Metaanalyse nach Linn und Petersen (1985)

Linn und Petersen (1985) fokussierten sich im Rahmen ihrer Arbeit gezielt auf drei Fragestellungen in Bezug auf den Geschlechtsunterschied bei Raumvorstellungsaufgaben – Ausmaß des Unterschiedes, Aspekte auf denen dieser beruht und vor allem das Alter, ab wann die ersten Anzeichen des Unterschiedes zu registrieren sind. Insgesamt 172 Studien (seit Maccoby & Jacklin (1974) publiziert) fanden ihre Aufnahme in die Metaanalyse, wobei die Sicht der psychometrischen und kognitiven Perspektive (siehe Abschnitt 3.2) im Vordergrund stand. Nachfolgend sollen die drei Bereiche der Raumvorstellung, welche schon vorab angesprochen wurden (vgl. Abschnitt 3.2.4), in Zusammenhang mit den Geschlechtsunterschieden dargestellt und anhand von Beispielen illustriert werden:

- Spatial perception: Hierunter fällt zum Beispiel der Water Level Test (WLT; z.B.: nach Piaget & Inhelder, 1956; vgl. Abb. 4.1). Die Aufgabe der Testperson besteht darin, den Wasserspiegel bei einem schief positionierten Behälter einzuzeichnen; ohne Berücksichtigung von „interferierender Information“ (Glück, 1999, S. 111), jedoch unter Berücksichtigung der Lage des eigenen Körpers.

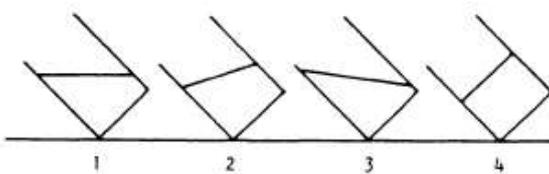


Abbildung 4.1: Itembeispiel des Water Level Tests (WLT)- entnommen aus Linn & Petersen, 1985, S. 1482.

- Mental rotation: Shepard- Metzler- Aufgaben (Vandenberg & Kuse, 1987; vgl. Abb. 4.2) und aus Thurstones Primary Mental Abilities (PMA; Thurstone & Thurstone, 1941) der Subfaktor „space“. Bei diesen Aufgaben steht jedoch die Zeit, welche der Proband benötigt um sich das Objekt mental vorstellen zu können und in weiterer Folge mental zu drehen, um dieses in die korrekte Position/Lage zu bringen, im Vordergrund (vgl. Linn & Petersen, 1985).

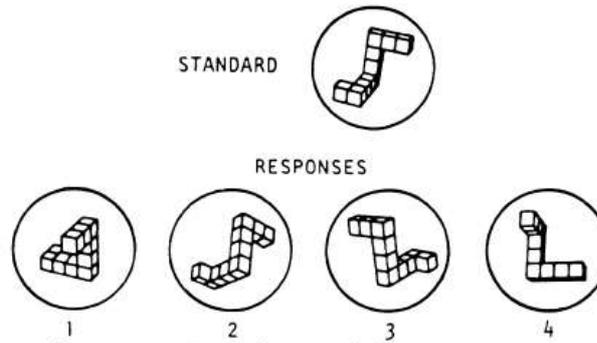


Abbildung 4.2: Aufgabe zur mentalen Rotation:

Identifiziert werden sollen die zwei möglichen Lösungen, jedoch aus einem anderen Blickwinkel aus betrachtet (aus Linn & Petersen, 1985, S. 1483).

- Spatial visualization: Linn und Petersen (1985) sprechen hier von Tests welche eine komplexere, schrittweise Verarbeitung von den Testpersonen abverlangen und „may involve the processes required for spatial perception and mental rotations but are distinguished by the possibility of multiple solution strategies“ (S. 1484); als Beispiel wären hier die bereits in Abbildung 3.1 dargestellten Papier-Faltaufgaben zu nennen.

Linn und Petersen kamen in ihrer Metaanalyse zu folgenden Ergebnissen:

Geschlechtsunterschiede konnten vor allem in den Bereichen der räumlichen Wahrnehmung und mentalen Rotation gefunden werden. Vor allem das Lösen von mentalen Rotationsaufgaben fiel den männlichen Testpersonen leichter als den weiblichen Testpersonen. Insbesondere die Version von Vandenberg und Kuse (1978) des Shepard-Metzler Tests sorgte für deutliche Unterschiede zwischen den Geschlechtern ( $\hat{\delta}_{\text{geschätzte relative Effektgröße}} = .94$ ), welche die Autoren auf zugrundeliegende Faktoren wie „differential rate of rotation, differential efficiency in strategy application, differential use of analytic processes, or differential caution“ (Linn & Petersen, 1985, S. 1489) zurückführen.

Im Bereich der räumlichen Visualisierung konnten geringe jedoch nicht signifikante Unterschiede gefunden werden. Beiden Geschlechtern schienen die Aufgaben gleich schwer oder gleich leicht zu fallen. In Bezug auf die Veränderungen der räumlichen Visualisierung im Laufe des Lebens, führen Linn und Petersen (1985) an:

Furthermore, sex differences in spatial visualization do not change across the life span. For most of the studies, no significant differences between males and females were found. Therefore sex differences in spatial visualization are not detected at any point in the life span. (S. 1490)

Die Unterschiede im Bereich der räumlichen Visualisierung führen die Autoren vielmehr auf die zugrundeliegende Lösungsstrategie zurück. Personen, die im Bereich der räumlichen Visualisierung gute Ergebnisse erzielen, hätten den Autoren zufolge „an appropriate repertoire of strategies and because they have effective meta-strategies to govern the selection of a strategy for each item“ (S.1490).

Im Bezug auf das Alter äußern sich Linn und Petersen (1985) wie folgt:

Our meta-analysis results contradict the assertion that sex differences in spatial ability are first detected in adolescence. For spatial perception, differences are detected in individual studies at about age 8 and, for grouped studies, emerge statistically only at age 18. For mental rotation, sex differences are detected whenever measurement is possible, although some versions of the test are inappropriate for those under 13. For spatial visualization, there are no differences. Thus sex differences in spatial ability are detected prior to adolescence for some categories of spatial ability and not at all for others. (S. 1492)

#### **4.2 Metaanalyse nach Voyer, Voyer und Bryden (1995)**

Im Vergleich zu Linn und Petersen (1985) nahmen Voyer, Voyer und Bryden nur jene Studien in ihre Metaanalyse auf, welche zwischen 1974 und 1993 sowie an mindestens fünf unabhängigen Stichproben durchgeführt wurden.

Die Zahl fünf begründen die Autoren damit, dass „a minimum of five studies allows for a meaningful test of the homogeneity of effect sizes for sex differences in the different tasks and for an examination of age effects through regression analysis“ (Voyer, Voyer & Bryden, 1995, S. 252).

Insgesamt 286 Effektgrößen konnten sie in ihrer Analyse bestimmen.

Um die Homogenität dieser sicherzustellen gab es kaum einen Unterschied in ihrer Vorgehensweise zu jener von Linn und Petersen (1985): Zunächst wurden die Effektgrößen auf die drei Bereiche der Raumvorstellung und weiters auch nach der Alterskategorie aufgeteilt (vgl. Tabelle 4.1).

Voyer et al. (1995) bezeichneten diese als nahezu homogen, wenn „the obtained chi-square is significant at the .05 level but not at the .01 level“ (S.252).

Nur einige der Subkategorien konnten als homogen oder als nahezu homogen bezeichnet werden: die Kategorie < 13 Jahren in allen drei Testkategorien sowie im Bereich der räumlichen Wahrnehmung bei 13 - 18 Jährigen.

Tabelle 4.1: Effektgrößen nach Voyer et al. (1995) der Testkategorie und dem Alter entsprechend (entnommen aus Voyer et al., 1995, S.258)

Category of tests	N	Weighted estimator of effect size		Test of significance for effect size (Z)	Homogeneity statistic ( $\chi^2$ )
		Present study	Linn & Petersen (1985)		
Mental rotation					
All ages	78	0.56	0.73	4.63*	560.19*
Under 13 years	13	0.33		2.00*	10.72*
13-18 years	23	0.45		4.21*	200.62*
Over 18 years	42	0.66		5.55*	263.70*
Spatial perception					
All ages	92	0.44	0.44	2.25*	158.74*
Under 13 years	21	0.33	0.37	1.73	36.31 <sup>b</sup>
13-18 years	18	0.43	0.37	2.18*	16.22*
Over 18 years	53	0.48	0.64	2.48*	97.16*
Spatial visualization					
All ages	116	0.19	0.13	1.43	241.47*
Under 13 years	40	0.02		0.12	53.01 <sup>b</sup>
13-18 years	20	0.18		1.52	41.58*
Over 18 years	56	0.23		2.00*	112.48*

<sup>a</sup> Homogeneity achieved. <sup>b</sup> Close to homogeneity.  
\*  $p < .05$ .

Den Unterschied machten jedoch die nachfolgenden Schritte aus. Da die Homogenität durch diese Vorgehensweise nicht in allen Bereichen gegeben war, wurde schließlich auch eine Gliederung nach den spezifisch eingesetzten Tests durchgeführt. Insgesamt sechs Tests, welche noch immer eine heterogene Effektgröße aufwiesen, wurden neben dem Alter der Testpersonen auch nach anderen Variablen unterteilt. Beispiele hierzu wären etwa der *Identical Blocks Test* (Stafford, 1961), welcher nach einer Individual- oder Gruppentestung gesplittert oder der *Mental Rotation Test* (Vandenberg & Kuse, 1978), der nach der Scoreberechnung (20 oder 40) aufgeteilt wurde.

Voyer et al. (1995) kommen zu folgendem Schluss:

...sex differences, favoring males, are clearly significant and homogeneous on the Cards Rotation Test, the generic mental rotation tasks, the Spatial Relations subtest of the PMA, and the Paper Form Board. Sex differences on the Spatial Relation subtest of the DAT and Paper Folding are homogeneous but not significant. The rod-and frame test and the Block Design subtest of the various Wechsler intelligence scales show sex differences in some age groups but not others. Finally, scoring and testing procedures proved to have an important influence on the magnitude of sex differences on the Mental Rotations Test, the Water Level Test, the Identical Blocks Test, and the Embedded Figures Test. [...] Most particularly, the Mental Rotations Test appears to produce the most robust sex differences among all tests included in the present analysis. This is especially obvious when this test is scored out of 20, in which case the average effect size represents a mean difference of 0.94 standard deviation units between the means of men and women. (S. 265)

Den Unterschied zur Studie von Linn und Petersen (1985) machen die im Bereich der räumlichen Visualisierung gefundenen Geschlechtsunterschiede aus; diese waren nur bei jenen Verfahren sichtbar, welche eine Rotationskomponente zur Bearbeitung erforderten.

#### **4.3 Forschungsstand zum Geschlechterunterschied aus heutiger Sicht**

Wie bereits bei Linn und Petersen (1985) und Voyer et al. (1995) konnten auch bei Dabbs, Chang, Strong und Milun (1998) im Rahmen ihrer Untersuchung, zum geografischen Verständnis und den Navigationsstrategien von Männern und Frauen ( $n = 194$ ), Geschlechtsunterschiede in der Fähigkeit zur mentalen Rotation zugunsten von Männern bestätigt werden ( $t = 4.04$ ,  $p < .001$ ); im Landkartenwissen ( $t = 3.10$ ,  $p < .01$ ) erzielten die männlichen Teilnehmer ebenfalls bessere Ergebnisse.

Weiss, Kemmler, Deisenhammer, Fleischhacker und Delazer. (2003) sehen dabei das zeitlich gesetzte Limit als einen möglichen Erklärungsansatz für den Geschlechtsunterschied.

In Bezug auf die im MRT gefundenen Geschlechtsunterschiede argumentieren die Autoren wie folgt:

It is worth mentioning, that the Mental Rotation Test was performed with a time limit of 7 min in our study, which may be relevant for the outcome of the task. [...] Sex differences in spatial abilities appear to be due, at least in part, to performance or situational factors. [...] Males, who may have more exposure to spatial tasks, and on the other hand, women who may have more encouragement in the verbal domain, work more quickly and perhaps more confidently than the respective other sex. However, this advantage disappears when the premium for rapid performance is taken away.

(Weiss et al., 2003, S. 871/872)

Auch Glück und Fabrizii (2010) konnten einen Geschlechtsunterschied im MRT nachweisen und kamen ebenso zu dem Schluss, wonach Geschlechtsunterschiede nur dann auftraten, wenn ein Zeitlimit vorhanden war. Zusätzlich sei auch das Antwortformat von Bedeutung; der Unterschied beim MRT 0-4 – hier variiert die Anzahl von korrekten Antwortalternativen zwischen den Items von 0 bis 4) – ( $\hat{\delta} = .67$ ) verringerte sich im Vergleich zur Standardversion (MRT, Vandenberg & Kuse, 1978; vgl. Abb. 4.3); ( $\hat{\delta} = .93$ ).

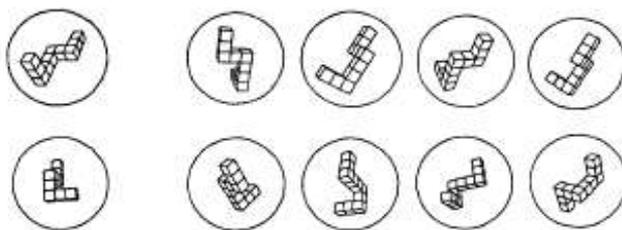


Abbildung 4.3: Beispielitems des Mental Rotation Tests: In der ersten Zeile stimmen Figur 1 und 4 mit der Vergleichsfigur überein. In der zweiten sind Figur 2 und drei richtig (entnommen aus Vandenberg & Kuse, 1978, S. 600).

Die Unterschiede werden auch auf einen möglichen Trainingseffekt zurückgeführt (vgl. Glück, 1999; Weiss et al., 2003). Bereits Sherman (1967; vgl. Weiss et al., 2003) führte diesbezüglich an, dass Frauen weniger die Möglichkeit hätten Aufgaben betreffend räumlicher Fähigkeiten zu üben.

Saccuzzo, Craig, Johnson und Larson (1996) widmeten sich bei ihrer Analyse Raumvorstellungstests mit dynamisch dargebotenen Aufgaben; sowohl männliche als auch weibliche Testpersonen wurden unter vier Bedingungen zu zwei

Testzeitpunkten getestet – computerisierte Aufgaben sowie Papier-Bleistift-Aufgaben mit jeweils einer *speed*- als auch *power*-Komponente.

In allen vier Bedingungen, zu beiden Messzeitpunkten, erzielten die männlichen Testpersonen höhere Werte. Sowohl bei männlichen als auch bei weiblichen Testpersonen konnte ein sog. Trainingseffekt, mit der Tendenz zur Verbesserung, festgestellt werden.

Vor allem aber bei den Computertests mit einer *speed*-Komponente wiesen die weiblichen Testpersonen die höchste Verbesserung auf, sodass sie (zum zweiten Messzeitpunkt; vgl. Abbildung 4.4) annähernd gleiche Werte erzielten wie die männlichen Testpersonen.

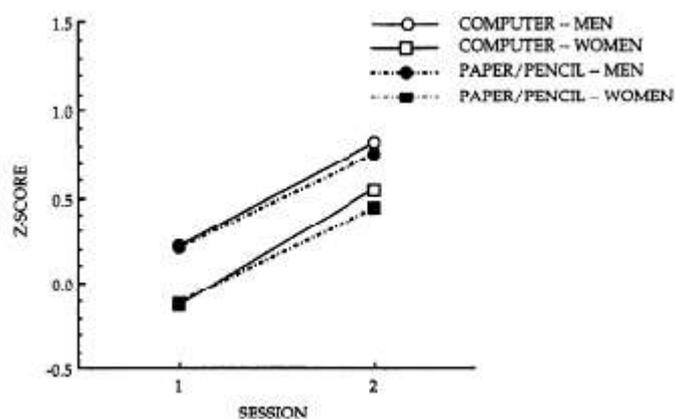


Abbildung 4.4: Interaktiondiagramm –Geschlecht x Papier/ Bleistift oder Computer x Testzeitpunkt (entnommen aus Saccuzzo et al., 1996)

Saccuzzo et al. (1996) konnten aus ihrer Untersuchung zwei Schlüsse ziehen:

The present findings add to our knowledge in two ways. First, they show a clear cut practice effect, not only on paper and pencil tasks but also on the new computerized tasks. Second, they show that women catch up to men on the speeded computerized test. Thus the results support the promise of computerized tests in reducing gender differences on speeded tests of spatial abilities. (S. 606)

Erklärungen für Geschlechtsunterschiede wurden weiters auch im Zusammenhang mit der Darbietungsform des Aufgabenmaterials gefunden.

So konnten Stepan (2004) sowie Neubauer, Bergner und Schatz (2010) im Rahmen ihrer Untersuchungen aufzeigen, dass die Geschlechtsunterschiede bei mentalen Rotationsaufgaben, welche dreidimensional dargestellt wurden (unter Einsatz von Anaglyphenbrillen; vgl. Stepan, 2004) geringer ausfielen, als es bei der herkömmlichen zweidimensionalen Vorgabe der Fall war.

Die schlechteren Leistungen weiblicher Testpersonen in Raumvorstellungsaufgaben spiegeln sich wiederum in ihrer Selbsteinschätzung wieder. Aufgrund mangelnden Trainings bzw. Kontaktes mit räumlichen Aufgabenmaterial scheinen sich, nach Goldstein und Chance (1965; vgl. Weiss et al., 2003), weibliche Testpersonen unwohl und benachteiligt zu fühlen und somit „respond with more caution and therefore perform at a slower rate than men“ (Weiss et al., 2003, S. 872).

In zahlreichen Studien wird auch die Selbsteinschätzung in Bezug auf räumliche Fähigkeiten (räumliche Intelligenz) behandelt.

So stellten beispielsweise Rammstedt und Rammsayer (2002) ein Messinstrument vor, das sogenannte *Inventar zur selbsteingeschätzten Intelligenz* (kurz ISI), wonach dieses in der Lage sei die selbsteingeschätzte Intelligenz von Personen zu erfassen.

Das ISI bezieht sich auf folgende elf Bereiche der Intelligenz (nach dem Modell von Gardner und Thurstone): „Verbales Verständnis, Wortflüssigkeit, mathematische Intelligenz, räumliche Intelligenz, Gedächtnisfähigkeit, Wahrnehmungsgeschwindigkeit, logisches Denken, musikalische Intelligenz, körperlich-kinästhetische Intelligenz, interpersonale Intelligenz und intrapersonale Intelligenz.“ (Rammstedt und Rammsayer, 2002, S. 436).

Ihre Untersuchung wurde an einer Stichprobe von insgesamt 852 Personen, im Alter zwischen acht und 64 Jahren ( $M = 20.8$  Jahre), durchgeführt, welche in weiterer Folge auf vier Teilstichproben aufgeteilt wurde – 130 GrundschülerInnen und 247 GymnasiastInnen ( $M = 11.8$ ,  $SD = 2.1$  Jahre), 139 BerufsschülerInnen ( $M = 19.6$ ,  $SD = 3.0$  Jahre), 240 StudentInnen ( $M = 23.8$ ,  $SD = 4.4$  Jahre) sowie 97 ältere Erwachsene ( $M = 53.2$ ,  $SD = 5.1$  Jahre).

Ein signifikanter Geschlechtsunterschied für die Gesamtstichprobe – wonach sich männliche Personen, im Vergleich zu den weiblichen, besser einschätzen – findet

sich in den Bereichen verbales Verständnis, mathematische Intelligenz, räumliche Intelligenz, Wahrnehmungsgeschwindigkeit und logisches Denken.

Auch in den Alters- und Bildungsgruppen können Geschlechtsunterschiede, zugunsten der männlichen Testpersonen, in den Bereichen mathematische Intelligenz, räumliche Intelligenz sowie logische Intelligenz, festgestellt werden (vgl. Rammstedt & Rammsayer, 2002). Die weiblichen Testpersonen schätzten sich beispielsweise in der musikalischen, inter- und intrapersonalen Intelligenz besser ein (vgl. Furnham & Petrides, 2004).

Holling und Preckel (2005) kritisierten die bisherigen Studien, da die Validitätsprüfung mit Hilfe von Korrelationsstudien erfolgte und somit durch systematische Unter- und Überschätzung verzerrt sein könnte.

Die Autoren argumentieren diesbezüglich:

The size of these correlations does not seem to justify the use of self-estimates as a replacement for intelligence tests in vocational counseling. On the other hand, these studies usually contain no information about the standard deviations of the independent and dependent variables. It is possible that restriction in range of one or both variables may have led to small observed correlations. In addition, the standard error of estimate may be small despite a small correlation, and thus the criterion may be predicted quite well. It is also possible that correlations might be biased by systematic over- or underestimates of subsamples. (Holling & Preckel, 2005, S. 504)

Im Gegensatz zu den vorhergehenden Studien bezog sich ihre Studie auf die Validitätsprüfung selbsteingeschätzter Intelligenz durch Abstandsmessungen zwischen selbsteingeschätzter und getesteter Intelligenz (vgl. Holling & Preckel, 2005).

In Zusammenhang mit dem Geschlecht konnten keine Unterschiede in der selbsteingeschätzten Intelligenz gefunden werden, jedoch in spezifischen Fähigkeitsbereichen: „Male participants rated themselves higher than female participants in the domains of abstract reasoning (FM, NS) and spatial ability (SO). Furthermore, male participants estimated their memory (ME) better than female participants” (Holling & Preckel, 2005, S. 511).

Die Ergebnisse stimmen mit jenen der vorhergehenden Studien überein, so konnte auch hier die Präferenz männlicher Personen zur höheren Einschätzung in den Fähigkeiten zur Raumvorstellung und abstrakten Denkens nachgewiesen werden. Auch Furnham und Buchanan (2005) kamen zu diesem Ergebnis, auch hier schätzten sich die männlichen Testpersonen in ihrer räumlichen Intelligenz – „the ability to find your way around the environment and from mental images“ (S. 547) – signifikant höher ein.

Zahlreiche Studien führten diese Unterschiede auf Persönlichkeitsfaktoren zurück (z.B.: Furnham & Thomas, 2004; Rammstedt & Rammsayer, 2002). Furnham und Buchanan (2005) zufolge scheint, dass „high neuroticism may well be associated with lower subjective estimates of various cognitive abilities.“ (S. 544).

Auch hier konnten Unterschiede zwischen den Geschlechtern zugunsten von männlichen Personen gefunden werden.

Zusammenfassend geben die Autoren wieder:

Gender differences in self-estimated intelligence must be explained in some other way than personality differences between men and women. This paper has shown that personality and gender factors relate systematically to self-perceived conventional intelligence. They also show that test experience (and to a lesser extent belief in test validity) are strongly predictive of these scores. (Furnham & Buchanan, 2005, S. 553)

## 5 Validität psychologisch- diagnostischer Verfahren

Ein Test ist dann valide, wenn „er tatsächlich jenes psychische Merkmal misst, welches er zu messen behauptet“ (Kubinger, 2009, S. 55).

Die *Validität* (Gültigkeit) eines Tests stellt eines der wichtigsten Testgütekriterien dar; d.h. ein Raumvorstellungstest soll auch tatsächlich räumliches Vorstellungsvermögen testen können und nicht etwa Reasoning (schlussfolgerndes Denken).

Insgesamt werden drei Validitätsarten unterschieden: Inhalts-, Konstrukt- und Kriteriumsvalidität.

### 5.1 Inhaltsvalidität

Inhaltliche Gültigkeit wird am besten durch ein sogenanntes „Experten-*Rating*“ (Kubinger, 2009, S. 56) gewonnen. Die herangezogenen Experten prüfen das Testmaterial, im Konkreten jedes Item, dahingehend „ob es in Bezug auf die gegebene operationale Definition dessen, was der Test messen soll, passt.“ (Kubinger, 2009, S. 56).

### 5.2 Kriteriumsvalidität

Steht ein zu validierender Test mit einem Außenkriterium in Übereinstimmung, so wird von Kriteriumsvalidität gesprochen. Im Vergleich zu den beiden anderen Ansätzen ist die Kriteriumsvalidität dadurch im Vorteil, dass diese eine statistische Kennzahl liefert.

Das Außenkriterium, mit dem der Test korreliert werden soll, kann einerseits ein anderer Test (Übereinstimmungsvalidität), der vorgibt dasselbe Konstrukt zu messen, oder andererseits ein sich in Zukunft befindliches Ereignis sein (Prognostische Validität).

In Bezug auf die prognostische Validität ist es fraglich, wie weit dieses Ereignis in der Zukunft liegen darf (z.B. beruflicher Erfolg), um dieses auch korrekt vorhersagen zu können.

Kubinger (2009) zufolge, ist eine „valide Prognose in Bezug auf ein einmaliges Verhalten, welches zwar schwer wiegt, aber allgemein selten auftritt, [...] mit Hilfe psychologisch-diagnostischer Verfahren völlig unrealistisch“ (S. 65).

Warum dennoch auch weniger valide Verfahren in der Praxis zum Einsatz kommen (beispielsweise bei der Bewerberauswahl), wird mit Hilfe von *Taylor-Russel-Tafeln* erklärt, wonach diese die Berechnung des Prozentsatzes geeigneter Bewerber (bei bekannter Grund- und Selektionsrate) ermöglichen (vgl. Kubinger, 2009).

### **5.3 Konstruktvalidität**

Bei der Konstruktvalidität steht der Test selbst im Vordergrund. Dieser wird danach geprüft, ob er neben pragmatischen vor allem theoriegeleitete Ansprüche, auf ein Konstrukt<sup>6</sup> bezogen, erfüllt (vgl. Lienert & Raatz, 1994).

Nach Bortz und Döring (2009) weist ein Test Konstruktvalidität dann auf, wenn die aus dem Konstrukt abgeleiteten Hypothesen durch die Testergebnisse bestätigt werden können. Eine Vielzahl an Hypothesen, das bestimmte Konstrukt betreffend sowie seine Verbindungen und Beziehungen zu anderen latenten und manifesten Variablen, wird formuliert. Wenn die Ergebnisse der Testung mit den empirischen und theoriegeleiteten Hypothesen übereinstimmen, kann dies als ein Hinweis auf Konstruktvalidität eines Tests betrachtet werden. Wenn die Hypothesen mit den Ergebnissen der Testung nicht konform sind, dann ist zu bezweifeln, dass das verwendete Instrument valide ist.

---

<sup>6</sup> „Merkmale, Zustände oder Instanzen, die nicht direkt beobachtbar sind, sondern aufgrund von (Verhaltens-) Beobachtungen erschlossen werden“ (Tewes & Wildgrube, 1999; zit.n. Kubinger, 2009, S. 57).

## 6 Raumvorstellung und Beruf

Die Fähigkeit zur Raumvorstellung ist nicht nur in unserem alltäglichen Tun von Bedeutung, sie ist bereits in so manchen Berufen unerlässlich und somit auch zu einer wichtigen Voraussetzung im Rahmen der berufsspezifischen Eignungsdiagnostik geworden.

Dass eine allgemeine kognitive Fähigkeit (Intelligenz) einen validen Prädiktor für den beruflichen „Erfolg“ darstellt konnte bereits in zahlreichen Metaanalysen zur Validität allgemeiner kognitiver Fähigkeit für berufsspezifische Leistungskriterien und Ausbildungserfolg verdeutlicht werden (vgl. Hunter, 1986; Hülshager & Maier, 2008), wobei „die Höhe der Validität in Abhängigkeit des Komplexitätsgrades des Berufes schwankt [...]“ (Hülshager & Maier, 2008, S. 114).

Nach Schmidt-Atzert, Deter und Jaeckel (2004) kann unter allgemeiner Intelligenz „die allen kognitiven Leistungstests gemeinsame Varianz verstanden“ (S. 148) werden, sodass durch Leistungstests neben spezifischen Fähigkeiten (wie etwa die zur Raumvorstellung) auch Aspekte der allgemeinen Intelligenz erfasst werden können. Den Autoren zufolge sei die „allgemeine Intelligenz der beste Prädiktor für den theoretischen Teil der Prüfung, während spezifische Intelligenzkomponenten und zusätzliche Leistungstests den praktischen Teil der Prüfung am besten vorhersagen“ (Schmidt-Atzert et al., 2004, S.156). Eine Kombination aus diesen sei, den Autoren zufolge, am besten geeignet um einen Ausbildungserfolg vorhersagen zu können.

Hülshager und Maier (2008) nach sei die Validität dieser spezifischen Intelligenzkomponenten, im Vergleich zur allgemeinen kognitiven Fähigkeit zwar niedrig, dennoch scheint sie essentiell zu sein und ihre Erfassung sei nicht als ungeeignet zu betrachten.

Inwiefern ein Zusammenhang zwischen den gemessenen Leistungen eines Raumvorstellungstests mit den zukünftigen ausbildungs- und berufsbezogenen räumlichen Leistungen bestehe, wurde bereits in Zusammenhang mit der prognostischen Validität (vgl. Kapitel 5) erörtert. So ist es auch nicht verwunderlich, dass die Berufanforderungsprofile immer präziser beschrieben

werden. Mit Hilfe spezifischer Verfahren sollen die dem Berufsbild nach erforderlichen Fähigkeiten erfasst werden.

Da sich ein Teil der Stichprobe der hier vorliegenden Arbeit aus Personen aus dem Bereich der Luftfahrt, genauer aus Piloten und Flugverkehrsleitern (anstelle der Bezeichnung „Flugverkehrsleiter“ wird in weiterer Folge der Begriff „Fluglotse“ verwendet) zusammensetzt, sollen diese beiden Tätigkeitsbereiche, in denen das räumliche Vorstellungsvermögen eine wichtige Voraussetzung darstellt, als „grob“ dargestellte Beispiele fungieren.

## **6.1 Tätigkeitsbeschreibung eines Fluglotsen**

Die Austro Control Österreichische Gesellschaft für Zivilluftfahrt mbH ist neben der sicheren Abwicklung des Flugverkehrs für ganz Österreich auch für die Ausbildung künftiger Fluglotsen<sup>7</sup> verantwortlich.

Der Aufgabenbereich eines Fluglotsen umfasst einerseits die Sicherheit von Passagieren andererseits die der Flugzeuge. Während ihrer Dienstzeit sind sie für den reibungslosen Ablauf des Flugverkehrs verantwortlich. Der Zuständigkeitsbereich eines Fluglotsen liegt neben der Bekanntgabe von Flugroute und Flughöhe, auch beim Erteilen der Start- bzw. der Landeerlaubnis.

Die Arbeit eines Fluglotsen erfordert unter anderem eine rasche Auffassungs- und Kombinationsgabe und hohe Konzentrationsfähigkeit. Essentiell dabei ist das räumliche Vorstellungsvermögen, um die erforderlichen Mindestabstände zwischen den Flugzeugen einhalten zu können. Die Abwicklung aller Befehle eines Flugverkehrsleiters erfolgt vom Tower aus (vgl. Aschenbrenner, 2007; Gruber, 2008). Informationen zum Flugverkehr erhält der Fluglotse über einen Radarbildschirm (visuell) bzw. Sprechfunk (auditiv), welche er in weiterer Folge in einem dreidimensionalen Bild umsetzen muss, um aus den Flugbewegungen die richtigen Befehle ableiten zu können.

Wie aus der Beschreibung hervorgeht, so ist das räumliche Vorstellungsvermögen eine der wichtigsten Voraussetzungen eines Fluglotsen. Um jedoch diese auch als gegeben „bescheinigen“ zu können, werden die Anwärter im Rahmen des von der Austro Control durchgeführten Aufnahmeverfahrens in der Hauptsektion (um in diese zu gelangen müssen die vorgegebenen Grundvoraussetzungen sowie die

---

<sup>7</sup> Nähere Informationen zum Auswahlverfahren und zur Ausbildung von Fluglotsen sind der Homepage der Austro Control zu entnehmen- [www.austrocontrol.at](http://www.austrocontrol.at) [14.11.2010].

Vorselektion erfolgreich absolviert werden) in Form von Eignungstestungen in den Bereichen Leistung und Persönlichkeit, am Kuratorium für Verkehrsicherheit, auf diese hin getestet.

Personen die das Aufnahmeverfahren erfolgreich abgeschlossen haben werden als „Trainees“ zur Ausbildung zugelassen.

## **6.2 Tätigkeitsbeschreibung eines Piloten**

Im Vergleich zur der Tätigkeit als FluglotsIn, kann die Ausbildung zum PilotIn<sup>8</sup> auch an privaten Flugschulen erfolgen (vgl. Gruber, 2008).

So wurden die für die vorliegende Arbeit erforderlichen Personen vom Jet Alliance Flight Training Center zur Verfügung gestellt.

Hierbei handelt es sich um eine seit 2006 von der Austro Control GmbH zugelassene Ausbildungseinrichtung, in südlicher Richtung von Wien.

Im Rahmen des Qualifying werden die Grundvoraussetzungen geprüft. Das räumliche Vorstellungsvermögen wird hier, im Vergleich zu den FluglotsInnen, nicht durch einen Leistungstest erfasst, sondern mit Hilfe des Simulators.

In der PPL-Phase (Privatpilotschein) werden den Anwärtern Basisfertigkeiten näher gebracht. Im Rahmen des Theorieunterrichts (ca. 125 Stunden) wird mitunter der Gebrauch der Checklisten sämtlicher Flugplanungsdetails als auch der Sprechfunkverkehr erlernt. Insgesamt 47 Flugstunden müssen in dieser Phase geleistet werden.

Nach erfolgreichem Abschluss der PPL-Phase folgt die nächste Stufe CPL (Berufspilotenlizenz). Im Laufe dieser Phase müssen ca. 500 Theoriestunden und mindestens 150 Flugstunden absolviert werden.

Die sogenannte Instrumentenflugberechtigung (IR) berechtigt den Privatpilotscheinbesitzer „Flugzeuge bis zu einer Entscheidungsmindesthöhe von 200 Fuß (60 m) nach Instrumentenflugregeln zu führen“ (Gruber, 2008, S. 31). Die letzte Stufe nimmt die ATPL-Ausbildung (Verkehrspilotenlizenz) ein.

Der Leistungsumfang dieser Phase umfasst ca. 800 Theoriestunden, weiters müssen „1500 Flugstunden nachgewiesen werden, diese müssen mindestens 75

---

<sup>8</sup> Zusätzliche Informationen zur Ausbildung finden sich auf der Homepage [www.jetallianz.at](http://www.jetallianz.at) [14.11.2010].

Stunden Instrumentenflugzeit und 100 Stunden Nachtflug beinhalten“ (Gruber, 2008, S. 31).

Gruber (2008) führt berufliche oder ausbildungsbezogene Situationen an, in welchen die Fähigkeit zur Raumvorstellung unerlässlich erscheint. Das Steuern des Flugzeuges im Raum erfolgt mit Hilfe der Informationen aus der Umgebung oder den Cockpitinstrumenten, wobei die Umgebung beim Instrumentenflug bewegt wahrgenommen wird. Der Pilot muss die Lage des Flugzeuges so lange ändern, bis die angestrebte Position am Horizont eingenommen wurde – diese mentale Verarbeitung wird als eine Rotationsstrategie beschrieben. Wird die Lage des Flugzeuges jedoch falsch interpretiert, so wird dies als räumliche Desorientierung bezeichnet. Im Gegensatz dazu kann es zu einem Orientierungsfehler kommen, wenn in Relation zur Umgebung, die Position des Flugzeuges falsch eingeschätzt wird (vgl. Gruber, 2008).



## **II Empirischer Teil**



## 7 Fragestellung und Hypothesen

Im Rahmen der vorliegenden Arbeit soll der Frage nachgegangen werden, inwiefern der Test zur Angewandten Raumvorstellung (TARV, Weitensfelder, unpubl.) fähig ist, zwischen Extremgruppen zu unterscheiden, und inwiefern die hier gemessene Auffassung von Raumvorstellung mit jener des Endlosschleifentests (EST; Gittler & Arendasy, unpubl.) zusammenhängt.

Um den ersten Teil der Fragestellung beantworten zu können – inwiefern ist der TARV fähig zwischen Extremgruppen zu unterscheiden, um zumindest als Screeningverfahren zum Einsatz zu kommen – ergibt sich die entsprechende Haupthypothese; vorab sei angemerkt, dass im Rahmen der vorliegenden Untersuchung eine Kurzversion des TARV zum Einsatz kam (vgl. Abschnitt 9.1.1.2) – Version F (für FluglotsInnen und PilotInnen) 4.0 – und die nachfolgenden Hypothesen sich auf diese (vorgegebene Version) beziehen:

### 1. Hypothese:

*H0: Personen aus dem Bereich der Luftfahrt weisen gleich hohe oder geringere Werte im TARV-F 4.0 auf, als Personen mit geisteswissenschaftlichem Schwerpunkt.*

*H1: Personen aus dem Bereich der Luftfahrt weisen im TARV-F 4.0 eine generalisiert höhere Raumvorstellungsfähigkeit auf, als Personen mit geisteswissenschaftlichem Schwerpunkt.*

Wie aus bisherigen Studien hervorgeht, scheinen auch Unterschiede im Zusammenhang mit der selbsteingeschätzten räumlichen Vorstellungsfähigkeit, vor allem zwischen Männern und Frauen zu bestehen (vgl. Abschnitt 4.3; Forschungsstand zum Geschlechterunterschied aus heutiger Sicht).

Inwiefern nun TARV-F 4.0 in der Lage ist zwischen Personen, die sich als gut in ihrer räumlichen Vorstellung und Personen, die sich als weniger gut hinsichtlich dieser Fähigkeit einschätzen, zu unterscheiden, soll mit der nachfolgenden Hypothese geprüft werden; wobei nur 25% der Personen, die ihre Fähigkeiten als gut einschätzen und 25% der Personen, die diese als weniger gut einschätzen

berücksichtigt werden, die mittleren 50% entfallen. Diese Extremgruppen werden im Weiteren mit den im TARV-F 4.0 erbrachten Leistungen verglichen.

## **2. Hypothese:**

*H0: Testpersonen, die sich als gut in ihrer räumlichen Vorstellungsfähigkeit einschätzen, erzielen im TARV-F 4.0 keine besseren Ergebnisse, als jene Testpersonen die sich als schlecht einschätzen.*

*H1: Die Testpersonen, die sich als gut in ihrer räumlichen Vorstellungsfähigkeit einschätzen, erzielen bessere Testwerte im TARV-F 4.0, als jene Testpersonen, die ihre räumliche Vorstellungsfähigkeit schlecht einschätzen.*

Dem Endlosschleifentest (EST; Gittler & Arendasy, unpubl.) werden zwei Lösungsmöglichkeiten zugesprochen; einerseits die Methode der mentalen Rotation und andererseits die des mentalen Perspektivenwechsels des eigenen Körpers (vgl. Stepan, 2004).

In Bezug auf den zweiten Teil der Fragestellung, inwiefern die im TARV-F 4.0 gemessene Auffassung von Raumvorstellung mit jener des EST zusammenhängt ergeben sich folgende Hypothesen, mit der Annahme, dass Personen, die im EST gute Ergebnisse erzielen, auch auf der entsprechenden Facette des TARV-F 4.0 gut sind; wobei nur Personen berücksichtigt werden, deren Testergebnisse des EST sich im oberen (25%) und unteren (25%) Bereich wiederfinden, die mittleren 50% werden auch hier ausgelassen.

## **3. Hypothese:**

*H0: Die Testpersonen, die im EST höhere Testkennwerte aufweisen, erzielen im TARV-F 4.0 keine besseren Werte in der Facette „Orientierung“, als jene Testpersonen, die im EST niedrigere Testkennwerte erreichen.*

*H1: Die Testpersonen, die im EST hohe Testkennwerte erreichen, erzielen in der Facette „Orientierung“ des TARV-F 4.0 bessere Ergebnisse in Richtung höherer*

*Orientierungsfähigkeit, als jene Testpersonen die im EST niedrigere Werte aufweisen.*

#### **4. Hypothese:**

*H0: Die Testpersonen, die im EST höhere Testkennwerte aufweisen, erzielen im TARV-F 4.0 keine besseren Ergebnisse in der Facette „mentale Rotation“, als jene Testpersonen, die im EST niedrigere Testkennwerte erreichen.*

*H1: Die Testpersonen, die im EST hohe Testkennwerte erreichen, erzielen in der Facette „mentale Rotation“ des TARV-F 4.0 bessere Testkennwerte in Richtung höherer Fähigkeit zur mentalen Rotation, als jene Testpersonen, die im EST niedrigere Testkennwerte aufweisen.*

## 8 Statistische Methode

Zur Prüfung der einzelnen Hypothesen wird je ein *Welch-Test*<sup>9</sup> bei einseitiger Fragestellung gerechnet. Als abhängige Variablen dienen jeweils die Rohwerte der interessierenden Skala des TARV-F 4.0, unabhängige Variable stellt die Zuteilung zur jeweiligen Extremgruppe dar.

Die einseitige Fragestellung begründet sich aus der theoriegeleiteten Gerichtetheit der Hypothesen – Beispielsweise: *Die Testpersonen, die im EST hohe Testkennwerte erreichen, erzielen in der Facette „Orientierung“ des TARV-F 4.0 bessere Ergebnisse in Richtung höherer Orientierungsfähigkeit, als jene Testpersonen die im EST niedrigere Werte aufweisen.*

### 8.1 Optimaler Stichprobenumfang

Unter Berücksichtigung der relevanten Effekte wurde die Stichprobengröße, für den Vergleich zweier unabhängiger Stichproben, im Vorhinein mit Hilfe des Statistikprogramms *CADEMO Light* (Version 3.27) berechnet (vgl. auch Rasch und Kubinger, 2006).

Vorgegebene Parameter werden in der nachfolgenden Abbildung 8.1 dargestellt:

$d = 1.3333$ $s^2(1) = s^2(2) = 1.0000$ $\text{Risiko 1. Art } (\alpha) = 0.0500 \text{ und Risiko 2. Art } (\beta) = 0.0500$
---

Abbildung 8.1: Vorgegebene Parameter zur Berechnung des optimalen Stichprobenumfangs

Die Bestimmung des optimalen Stichprobenumfangs ergibt für einen einseitigen Test zweier unabhängiger Stichproben, einen Stichprobenumfang je Gruppe  $n_1 = n_2 = 14$ ,  $n_\Sigma = 28$ . Im Hinblick auf die Hypothesen 2 - 4 und aufgrund der Tatsache, dass die mittleren 50% aus der Analyse ausgeschlossen werden, muss von einer doppelt so großen Stichprobengröße ausgegangen werden, womit ein Umfang von  $n_\Sigma = 56$  als optimal anzusehen ist.

<sup>9</sup> Zum Vorzug des *Welch-Tests* gegenüber dem *T-Test* vgl. Kubinger, Rasch und Moder (2009).

## **9 Methodik und Studiendesign**

Um die hier gestellte Fragestellung beantworten zu können wurden den Personen zwei Verfahren zur Erfassung des räumlichen Vorstellungsvermögens vorgegeben. Zum einen der Endlosschleifentest (EST; Gittler & Arendasy, unpubl.) und zum anderen eine Kurzversion des Tests zur Angewandten Raumvorstellung; Version F (für FluglotsInnen und PilotInnen) 4.0 (TARV-F 4.0; Weitensfelder, unpubl.).

Zusätzlich wurde zwischen den beiden Raumvorstellungstests ein Androgynie-Fragebogen (ANDRO; Gittler, unpubl.) von den Testpersonen bearbeitet. Hierbei handelt es sich um einen Fragebogen zur Geschlechtsrollenidentität. Dieser beinhaltet 63 Eigenschaftswörter, wobei 17 davon für die männliche und 25 für die weibliche Geschlechtsrolle charakteristisch sind. Aufgabe der Personen ist es, sich auf einer sechsstufigen Antwortskala hinsichtlich jedes dieser Eigenschaftsbegriffe einzuschätzen (von „trifft nicht zu“ bis „trifft zu“; vgl. Anhang A). Die Ergebnisse des ANDRO stehen nicht im Zusammenhang mit der hier vorliegenden Fragestellung, weshalb auch auf diesen nicht näher eingegangen wird. Da jedoch im Rahmen dieser Untersuchung ein „Puffer“ zwischen den beiden Raumvorstellungstests benötigt wurde (vgl. Testvorgabe, Tab. 9.3) und die Daten für eventuelle spätere Studien herangezogen werden, erwies sich ANDRO im Hinblick auf die hier vorliegende Untersuchung als passend.

Nachfolgend sollen der Endlosschleifentest als auch der Test zur Angewandten Raumvorstellung im Hinblick auf den jeweiligen theoretischen Hintergrund sowie deren Einsatz in der vorliegenden Arbeit dargestellt werden.

### **9.1 Verfahren zur Erfassung der räumlichen Vorstellungsfähigkeit**

#### **9.1.1 Test zur Angewandten Raumvorstellung (TARV)**

Hierbei handelt es sich um ein an der Test- und Beratungsstelle des Arbeitsbereichs Psychologische Diagnostik der Universität Wien neu entwickeltes Verfahren zur Erfassung der räumlichen Vorstellungsfähigkeit.

### 9.1.1.1 Theoretischer Hintergrund

Inspiziert durch die Arbeiten von Lohman (1979), Thurstone (1938) sowie Linn und Petersen (1985) – vgl. Kapitel 3.2.3 und 3.2.4 – zielt das Verfahren auf drei Facetten der Raumvorstellung ab: *Relationen* (Erkennen von Größen- und Abstands-verhältnissen), *Rotation* und *Orientierung*.

Die Facette *Relationen* erfasst die Fähigkeit der Person Relationen, besonders in Bezug auf Abstands- und Größenverhältnisse, bei Objekten zu erkennen.

Mit der zweiten Facette *Rotation* wird die mentale Rotationsfähigkeit einer Person – gedankliche Vorstellung der Drehbewegung eines Gebildes – beschrieben.

Der Aspekt der *Orientierung* bezieht sich auf das Einschätzen von Positionen verschiedener Teile eines Objektes.

Das Verfahren baut auf zwei Phasen auf. Neben den oben beschriebenen Facetten soll weiters die Fähigkeit des *Schließens von 3D auf 2D* (1.Phase) und umgekehrt, des *Schließens von 2D auf 3D* (2.Phase), erfasst werden.

Die erste Testphase erfordert das Schließen von dreidimensional vorgegebenen Figuren auf die zugehörigen (richtig zu beurteilenden) zweidimensionalen Planansichten, wobei diese aus verschiedenen Perspektiven (rechte Seite, linke Seite, Top (von oben)) dargestellt werden.

In der zweiten Phase wird entsprechend zur ersten vorgegangen – mit dem Unterschied, dass hier das Schließen von zweidimensional vorgegebenen Figuren auf die zugehörigen (richtig zu beurteilenden) dreidimensionalen Planansichten geprüft werden soll.

### 9.1.1.2 Vorgabe

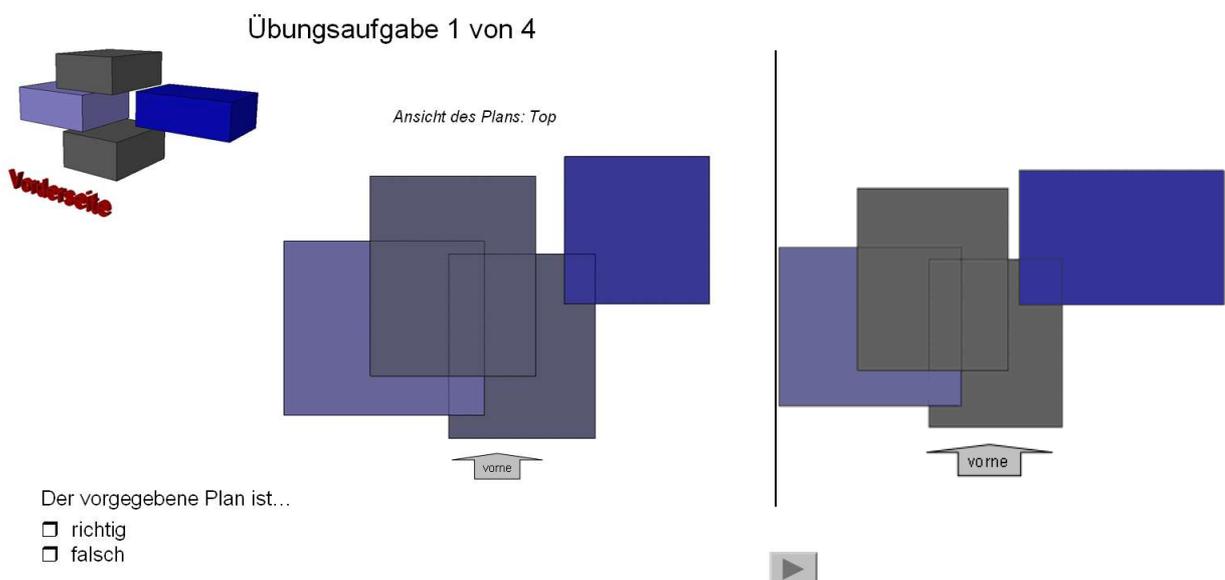
Die Präsentation der einzelnen Aufgaben erfolgt über den Computerbildschirm. Da sich der Test jedoch noch in der Entwicklungsphase befindet, werden hier die Items in Form einer Power-Point-Präsentation dargeboten. Die Antworten der Testpersonen werden nicht automatisch über den Computer registriert, sondern müssen händisch (von der Testperson selbst) auf einem Antwortbogen vermerkt werden (Antwortbogen siehe Anhang B). TARV stellt ein sequentiell aufgebautes Verfahren dar, bestehend aus insgesamt zehn Aufgaben (vgl. Tabelle 9.1).

TARV-F 4.0 Facette	2D→ 3D	3D→ 2D	Aufgabenanzahl
S1 -Relationen	3	1	4
S2- Rotation	2	1	3
S3- Orientierung	1	2	3
Gesamt	6	4	<b>10</b>

Tabelle 9.1: Übersicht zur Aufgabenzusammenstellung im TARV-F 4.0

Jede Aufgabe besteht aus drei Teilen. Die Testpersonen müssen das ihnen dargebotene Objekt (welches zuvor als alleinstehendes „Ausgangsobjekt“ präsentiert wird) aus drei verschiedenen Planansichten nach seiner Richtigkeit beurteilen, wobei ein oder zwei Perspektivendarstellungen fehlerhaft sein können. Die Pläne werden als transparent/durchscheinend dargestellt (d.h. es sind sowohl Kanten als auch Teile in den Plänen eingezeichnet, welche sonst nicht sichtbar wären; vgl. Weitensfelder, Grubestic, Kubinger & Gittler, submitted) – vgl. Abbildung 9.1: Der Testperson wird eine Perspektivenmöglichkeit des Gebildes dargeboten (hier: Ansicht des Plans: Top), welche nach ihrer Richtigkeit in Bezug auf das Zielobjekt (links außen) als richtig oder falsch zu bewerten ist. In diesem Fall handelt es sich um eine falsche Darstellung des Ausgangsobjektes (einige Teile des Gebildes standen nicht im richtigen Größenverhältnis zueinander); rechts außen wird die richtige Lösung dargestellt.

Abbildung 9.1: Übungsaufgabe 1 des TARV-F 4.0:  
Links die fehlerbehaftete Perspektivendarstellung; rechts die richtige Lösung der Aufgabe.



### 9.1.1.3 Verrechnung

Das Item wird letztlich als „gelöst“ gewertet, wenn alle drei Ansichten richtig beurteilt wurden. Somit gilt der Plan als richtig, wenn alle Teile im Plan richtig zueinander liegen, sich am richtigen Ort befinden und in der richtigen Größe zueinander stehen. In jeder Aufgabe überwiegt nur eine Fehlerart, um die Facetten voneinander zu trennen und nicht miteinander zu vermischen. In Bezug auf die Facette Relationen können sich die Größen- und Abstandsverhältnisse von und zwischen Objektteilen verändern. Die Aufgabe der Testpersonen hierbei besteht darin, fehlerbehaftete Winkel von Objektteilen (falsch rotierte Teile) zu erkennen. Bei der dritten Facette, der Orientierung, kann es im Vergleich zu anderen zur Veränderung von Richtungen einzelner Objektteile kommen (vgl. Weitensfelder et al., submitted).

### 9.1.2 Endlosschleifentest (EST)

#### 9.1.2.1 Theoretischer Hintergrund

##### 9.1.2.1.1 Schlauchfiguren (Stumpf & Fay, 1983)

Stumpf und Fay (1983) konstruierten Schläuche, welche in einem durchsichtigen, hohlen Plastikwürfel in eine feste Position gebracht und aus sechs verschiedenen Perspektiven fotografiert wurden. Eine Schlauchfigur und eine zweite Abbildung, werden dargeboten und die Aufgabe der Testperson ist es nun zu beurteilen, aus welcher Sicht (von rechts, von links, von hinten, von oben oder von unten) die zweite Abbildung dargestellt ist (vgl. Stumpf & Fay, 1983).

Im Zusammenhang mit den im Kapitel 3.2 beschriebenen Modellen, scheint die zugrundeliegende Fähigkeit der Schlauchfiguren insbesondere die Fähigkeit zur mentalen Orientierung zu sein (vgl. Gruber, 2008). Wobei diese nicht immer als ein eigenständiger Faktor anzusehen ist (vgl. Colom et al. 2001) und somit eine Abgrenzung zur mentalen Rotation nicht einfach erscheint (vgl. Spatial Orientation nach Lohman).

Gittler und Arendasy (2003) kritisierten die fehlende Rasch- Homogenität der Schlauchfiguren und konnten diese im Rahmen empirischer Studien belegen.

Ihre Annahmen stützen sich auf drei mögliche Begründungen, wonach „diese Aufgaben [der Schlauchfiguren] den Homogenitätsanforderungen des RM mit hoher Wahrscheinlichkeit nicht standhalten“ (Gittler & Arendasy 2003, S. 167).

Zunächst führen die Autoren an, dass Items vorhanden wären, bei denen weniger das räumliche Vorstellungsvermögen im Vordergrund stehe, sondern Items welche „aufgrund korrekter Bearbeitung der sichtbaren Enden der gewundenen Schläuche [...] gelöst werden“ (Gittler & Arendasy 2003, S. 167).

Weiters bilden Items, die mit „von hinten“ zu lösen sind, lediglich eine spiegelbildliche Abbildung der „Startansicht“. Das Lösen dieser Aufgabe sei den Autoren zufolge weniger auf das räumliche Vorstellungsvermögen zurückzuführen. Zuletzt führen Gittler und Arendasy (2003) an, dass „infolge der Verwechslung von Ausgangs- und Zielansichten eine nominell inkorrekte Antwort gegeben wird“ (S. 167).

Durch die Darstellung von in sich geschlossenen Figuren sowie das Ausschließen von Items mit der Lösung „von hinten“ – stattdessen wurde die Antwortmöglichkeit „Ich weiß nicht“ eingeführt – und die Präzisierung der Instruktionen konnte die Gültigkeit des Rasch-Modells nicht nachgewiesen werden. Dies führte zu weiteren Studien, in welchen zusätzliche Verbesserungen des Testmaterials angestrebt wurden.

So wurden in weiterer Folge Schattierungen beseitigt und die Figuren als frei stehend (ohne einem Rahmen) abgebildet.

Abbildung 9.2 bildet eine Beispielaufgabe der Endversion des EST ab, wie sie in der vorliegenden Arbeit verwendet wurde. Ziel ist es neben dem Drehwinkel auch die Drehrichtung, ausgehend von der Startansicht (links), für die daraus resultierende Zielansicht (rechts) anzugeben.

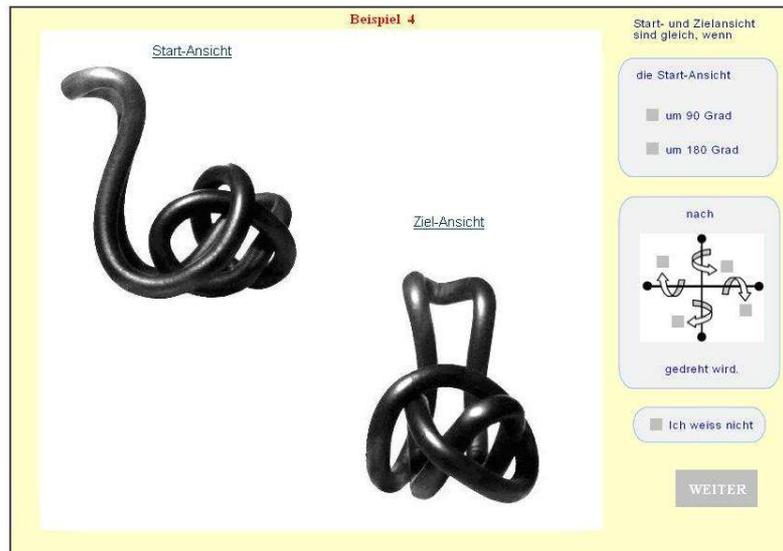


Abbildung 9.2: EST- Beispielitem 4-  
(Screenshot aus der Version Grubesic)

Gittler und Arendasy (2003) kreierten auf diese Weise einen Test zur Messung des räumlichen Vorstellungsvermögens, welcher eindimensionales Aufgabenmaterial beinhaltet.

#### 9.1.2.2 Vorgabe

Im Rahmen dieser Studie wurde der Endlosschleifentest als ein Computerverfahren, bestehend aus 16 zweidimensional dargestellten Items (inklusive eines Warming-Up-Items), zur Bearbeitung vorgegeben<sup>10</sup>.

Vor der eigentlichen Testung mussten die Teilnehmer einige demografische Daten (u.A. Geschlecht, Alter, Ausbildung) angeben. In diesem Zusammenhang wurden auch die schulischen Leistungen der Personen mit Hilfe einer Analogskala – in den Fächern Mathematik, Physik, Darstellende Geometrie, Geographie und Englisch – erfasst; diese sind für die vorliegende Fragestellung nicht relevant, sondern Teil einer anderen Untersuchung und werden daher auch nicht weiter berücksichtigt. Für die Fragestellung essentiell, war jedoch die Selbsteinschätzung der Teilnehmer (welche ebenfalls zuvor erfasst wurde), hinsichtlich ihrer räumlichen Vorstellungsfähigkeit; diese wurde ebenfalls auf einer Analogskala vermerkt (von 0%- 100%).

<sup>10</sup> Hierbei handelt es sich um eine, für die vorliegende Untersuchung, angepasste Version des EST- vgl. Anhang C.

Nachdem die Testpersonen erneut über den Bildschirm nach ihrem Einverständnis zur Testung gefragt wurden und sie diese bestätigten, konnte die Testphase beginnen. Zuvor wurden Übungsbeispiele vorgeben mit anschließender Rückmeldung, ob die Items richtig oder falsch gelöst wurden – eine Rückmeldung wird in der eigentlichen Testphase nicht gegeben.

Um die mögliche Verwechslungsgefahr zwischen der Start- und Zielansicht zu reduzieren, wurde diesbezüglich zwischen den Beispielaufgaben eine entsprechende Meldung präsentiert.

Von den insgesamt 16 zu bearbeitenden Items werden 7 Items der *power*-Bedingung und 8 weitere Items der *worklimit*-Bedingung zugeschrieben. Während bei der *power*-Bedingung die Gründlichkeit und Konzentration im Vordergrund steht, bezieht sich die *worklimit*-Bedingung auf die Lösungsgeschwindigkeit; die Testpersonen werden aufgefordert die (nachfolgenden) Aufgaben so rasch wie möglich zu bearbeiten.

#### 9.1.2.3 Verrechnung

Wie vorab beschrieben, besteht die Aufgabe der Testpersonen darin durch Bestimmung des Drehwinkels sowie der Drehrichtung die Figur aus der Startansicht in die dargestellte Zielansicht zu überführen.

Die Verrechnung erfolgt dichotom, wodurch ein Item als „richtig gelöst“ („1“) gewertet wird, wenn beide Angaben (Drehwinkel und –richtung) richtig angegeben worden sind und als „nicht gelöst“ („0“) wenn auch nur eine der beiden Aufgaben falsch beantwortet oder die Antwortmöglichkeit „Ich weiß nicht“ („0“) gewählt wurde.

#### 9.1.3 Raumvorstellung aus Sicht des EST und TARV

Während sich der TARV auf die vorab genannten Facetten (vgl. Kapitel 9.1.1.1) der Raumvorstellung bezieht – somit auch als ein facettenorientiertes Verfahren beschrieben wird (vgl. Weitensfelder et al., submitted) – und die Lösung der einzelnen Aufgaben jeweils auf einen dieser drei Fähigkeitsbereiche (Identifikation von Größen- und Abstandsverhältnissen, mentales Rotieren und Orientierungsfähigkeit) zurückzuführen ist, stehen beim EST zwei Lösungsalternativen zur Verfügung. Nach Gruber (2008) bezieht sich der *Schlauchfiguren*

Test (Stumpf & Fay, 1983) auf den Faktor „räumliche Orientierung“ (S. 11). Da EST auf diesem aufbaut, könnte der Schluss gezogen werden, dass der EST ebenfalls auf diesem Faktor beschrieben werden kann. Durch den Wegfall der offenen Enden (wie es bei den Schlauchfiguren jedoch der Fall ist) im EST, strebten die Autoren (vgl. Kapitel 9.1.2) gleichzeitig auch die Elimination der „markanten Orientierungshilfen“ (Gittler & Arendasy, 2003, S. 168) an.

Nach Gittler (persönliche Mitteilung, 14.09.2010) stehe beim EST die Fähigkeit zur mentalen Rotation des Aufgabenmaterials im Vordergrund und sei somit dem Faktor „mentale Rotation“ zuzuordnen. Durch die gedankliche Drehbewegung des Objektes (nach links, rechts, oben oder unten) soll die Person das Objekt in seine Zielansicht überführen.

Weiters wird die Möglichkeit eines mentalen Perspektivenwechsels angeführt (vgl. Stepan, 2004), wobei „die eigene Position um die Endlosschleifen gedanklich geändert“ (S. 45) wird, um zur Lösung der Aufgabe zu gelangen.

Auch eine geringe Korrelation zwischen dem EST und dem Dreidimensionalen Würfelttest (3DW; Gittler, 1990) lässt darauf schließen, dass eine andere Art oder anderer Aspekt der Orientierung für das Lösen der Aufgaben erforderlich erscheint (vgl. Gitter & Arendasy, 2003).

## **9.2 Ablauf der Studie**

Die Testungen wurden an der jeweils entsprechenden Einrichtungsstätte und in deren Schulungsräumlichkeiten, in Form von Gruppentestungen, durchgeführt.

Die Fluglotsen wurden einerseits in den Büroräumlichkeiten der Austro Control Zentrale und andererseits im Rahmen ihrer Kurseinheiten (zu zwei Terminen), auch nach Absprache mit den Kursleitern, in den Schulungsräumen der Austro Control Akademie getestet.

Auch bei der Gruppe der Piloten wurde vom Jet Alliance Flight Training Center der Schulungs- bzw. Testraum, ebenfalls für zwei Termine, zur Verfügung gestellt.

Für die Personen aus verschiedenen Studienrichtungen wurden die Räume der Computerdiagnostik der Test- und Beratungsstelle der Fakultät für Psychologie, ebenfalls für zwei Testzeitpunkte, bereitgestellt. Zu Beginn der am 18. Mai 2010 (ca. 14 Uhr) durchgeführten Testung, wurden in den Nebenräumen der Computerdiagnostik Bohrarbeiten durchgeführt, da diese jedoch (nach

Rückmeldung der Testpersonen) in einem nicht beeinträchtigend lautem Ausmaß vorhanden und gegen 15 Uhr beendet waren, konnte die Testung durchgeführt werden.

Alle drei Gruppen konnten somit unter sehr ähnlichen Bedingungen erfasst werden und erfüllten auch die Voraussetzung der geforderten Mindestausbildung; in allen drei Stichproben handelt es sich dabei um die Berufsreifeprüfung/Matura (vgl. Kapitel 10).

Tabelle 9.2 gibt einen Überblick über die Testtermine und die zum Datum (entsprechend) getestete Gruppe wieder.

Tabelle 9.2: Testtermine im Überblick

16.04.2010	PilotInnen-Trainees	Jet Alliance Flight Training Center
18.05.2010	StudentInnen	Computerdiagnostik der Fakultät für Psychologie
21.05.2010	PilotInnen-Trainees	Jet Alliance Flight Training Center
25.05.2010	StudentInnen	Computerdiagnostik der Fakultät für Psychologie
07.06.2010	LotsInnen-Trainees	Austro Control Zentrale
08.06.2010	LotsInnen-Trainees	Austro Control Zentrale
	StudentInnen	Computerdiagnostik der Fakultät für Psychologie
11.06.2010	LotsInnen-Trainees	Austro Control Akademie
14.06.2010	LotsInnen-Trainees	Austro Control Akademie

Vor dem Testbeginn musste jede Person nach dem Zufallsprinzip einen Code ziehen. Aufgrund der Bearbeitung am Computer sollten dadurch Gruppenbildungen soweit wie möglich unterbunden werden, weiters dienten die Codes dazu die Anonymität der einzelnen Testpersonen zu wahren.

Die Codes setzten sich aus Buchstaben und Zahlen zusammen. Die Buchstaben („TPL“ und „GW“) legten schon die entsprechende Stichprobe (bzw. Extremgruppe) fest, die anhängenden Zahlen standen für die laufende Nummer der Testpersonen (z.B.: TPL001 oder GW035). Zusätzlich wurden sowohl die Codes, als auch die TARV-F 4.0-Antwortbögen in zwei verschiedenen Farben gedruckt. Diese gaben die Bearbeitungsreihenfolge zwischen den beiden Raumvorstellungstests vor, wobei der Androgynie Fragebogen immer dazwischen bearbeitet wurde (vgl. Tab. 9.3). Da die Testungen immer unabhängig voneinander durchgeführt wurden (vgl. Tab. 9.2), wurden der jeweiligen Stichprobe auch nur die entsprechenden Codes vorgegeben. So erhielten die

Personen aus der Stichprobe der Geisteswissenschaften nur Codes mit GW in den oben angeführten Farben.

Die weitere – je nach Gruppenzugehörigkeit (PilotIn oder FluglotsIn) – vorgenommene Trennung, wurde von mir persönlich am Antwortbogen des TARV-F 4.0 zusätzlich notiert – wobei die Testzeitpunkte (nach dem Datum der Testung, welches im EST erfasst wird) eine weitere Sicherstellung der Einteilung darstellten. Eine weibliche Person aus der Gruppe der Geisteswissenschaften konnte jedoch nicht eindeutig einen bestimmten Tag der Testung zugewiesen werden, da diese statt des aktuellen Datums ihr Geburtsdatum angegeben hatte.

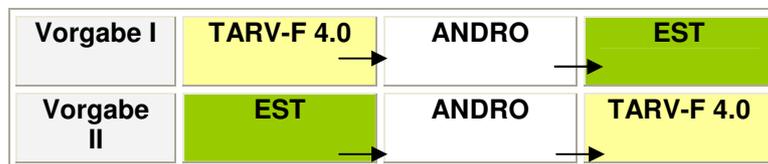


Tabelle 9.3: Reihenfolge der Testbearbeitung

Am Ende jeder Testung hatte jede einzelne Person insgesamt zwei Tests zur Messung der Raumvorstellung und einen Androgynie-Fragebogen bearbeitet.

Da kein zeitliches Limit vorgegeben wurde – wobei beim EST ein Worklimit vorhanden war, jedoch nur mit der Anweisung sie mögen so rasch wie möglich weiterarbeiten, aber ohne weiterer Konsequenzen (wie beispielsweise eines Testabbruchs nach verstrichener Zeit) – erstreckte sich die Testdauer (nach subjektiver Wertung) von ca. 1 Stunde bis maximal 2 Stunden. Im Durchschnitt jedoch betrug die Bearbeitungszeit ungefähr 1,5 Stunden.

## 10 Stichprobe

Insgesamt 60 Personen wurden im Zeitraum vom April bis Juni 2010 (siehe Tabelle 9.2) für die vorliegende Arbeit getestet, wobei eine Person (FluglotsIn) aus der späteren Auswertung ausgeschlossen wurde, da diese angab einen Teil der Instruktion nicht sorgfältig gelesen zu haben.

Somit gingen die Testergebnisse von 31 männlichen und 28 weiblichen Testpersonen ( $n = 59$ ) in die Analyse zur Extremgruppenvalidierung ein, bei einem durchschnittlichen Alter von 23.76 Jahren ( $SD = 2.794$ ; vgl. Tabelle 10.1). Von den insgesamt 59 Personen gaben 86,4% die Matura einer Allgemeinbildenden Höheren Schule (AHS) oder Berufsbildenden Höheren Schule (BHS), als die höchst abgeschlossene Ausbildung an. Gefolgt von einem universitären Abschluss (6.8%), einer Fachhochschule oder Akademie (5.1%) und zuletzt von einer berufsbildenden mittleren Schule (Fachschule oder Lehre, mit 1.7%; entspricht einer Person).

Alter in Jahren					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	19	1	1,7	1,7	1,7
	20	5	8,5	8,5	10,2
	21	9	15,3	15,3	25,4
	22	3	5,1	5,1	30,5
	23	12	20,3	20,3	50,8
	24	8	13,6	13,6	64,4
	25	5	8,5	8,5	72,9
	26	9	15,3	15,3	88,1
	27	3	5,1	5,1	93,2
	28	2	3,4	3,4	96,6
	29	1	1,7	1,7	98,3
	35	1	1,7	1,7	100,0
	Total	59	100,0	100,0	

Tabelle 10.1: Altersverteilung über die Gesamtstichprobe ( $n=59$ )

In Bezug auf die Testreihenfolge ergab eine Häufigkeitsanalyse (vgl. Tab. 10.2) folgendes Bild: Von den  $n = 59$  Testpersonen bearbeiteten  $n = 30$  Personen die Tests in der TAE-Reihenfolge und  $n = 29$  Personen in umgekehrter Darbietung (vgl. Tab. 9.3).

**Reihenfolge der Testvorgabe**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	TARV-EST	30	50,8	50,8	50,8
	EST-TARV	29	49,2	49,2	100,0
	Total	59	100,0	100,0	

Tabelle 10.2: Häufigkeitsverteilung zur Testvorgabe (n=59)

Aus der Tabelle ist eine annähernd gleichverteilte Testvorgabe erkennbar, welche auch im entsprechenden SPSS-Output des Chi-Quadrat-Tests (bei  $\alpha = 5\%$ ) bestätigt wird (vgl. SPSS-Output 10.1).

**Test Statistics**

	Reihenfolge der Testvorgabe
Chi-Square	,017 <sup>a</sup>
df	1
Asymp. Sig.	,896

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 29,5.

SPSS-Output 10.1: Chi-Quadrat-Test zur Vorgabeverteilung

### 10.1 Stichproben zur Validierung von Extremgruppen

Für die Extremgruppenvalidierung wurden die Extremgruppen je nach Ausbildung entweder der Stichprobe *Luftfahrt* oder *Geisteswissenschaften* gebildet (somit kann auch von natürlich bestehenden Extremgruppen gesprochen werden).

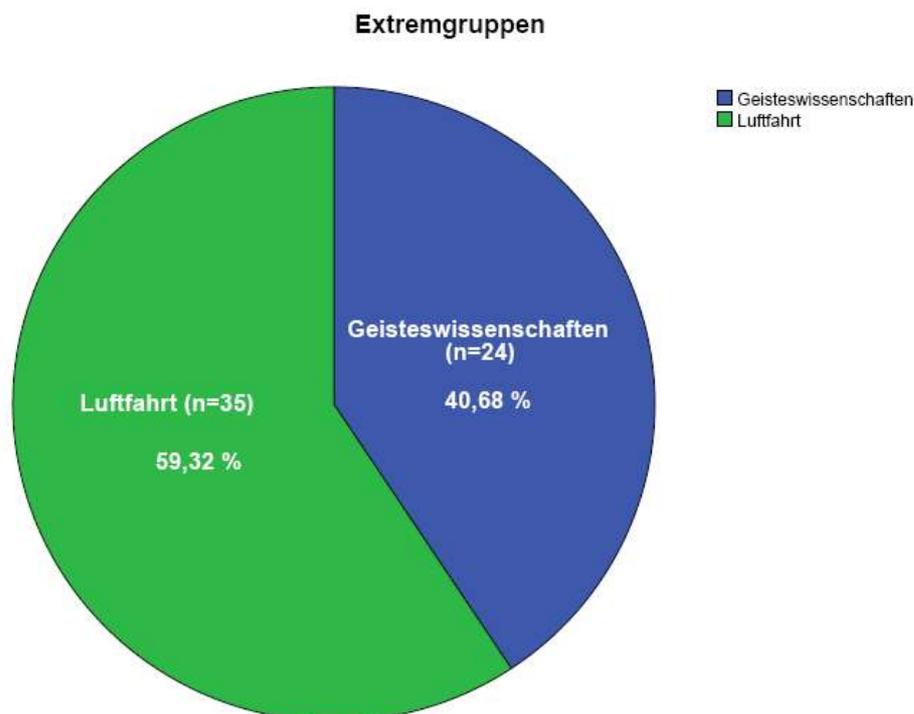
Im Falle der hier vorliegenden Arbeit bilden die PilotInnen und FluglotsInnen die Gruppe der *Luftfahrt*. Die Personen dieser Extremgruppe befinden sich derzeit in der Ausbildung zu PilotIn bzw. FluglotsIn (Trainees).

Das räumliche Vorstellungsvermögen stellt in Rahmen dieser eine wichtige Voraussetzung dar, welche mit Hilfe der Vorauswahl (Selektion), durch die entsprechende Einrichtung, als gegeben eingestuft wurde.

Studenten verschiedener „geisteswissenschaftlicher“ Studienrichtungen gingen in die Gruppe der *Geisteswissenschaften*<sup>11</sup> ein. Diese Extremgruppe setzte sich ausschließlich aus Studenten zusammen, deren Studienrichtungen kein räumliches Vorstellungsvermögen abverlangten (wie beispielsweise Theologie oder Psychologie).

Aus der Häufigkeitsanalyse geht hervor, dass insgesamt 24 Personen (40,7%) der Gruppe der *Geisteswissenschaften* und 35 Personen (59,3%) der Gruppe der *Luftfahrt* zugewiesen werden konnten. Im Folgenden soll die Häufigkeitsverteilung anhand des Kreisdiagramms verdeutlicht werden (Abb. 10.1); der entsprechende Geschlechteranteil der Gesamtstichprobe (n = 59) wird in der Tabelle 10.3 dargestellt.

Abbildung 10.1: Verteilung der Gesamtstichprobe nach Extremgruppen: *Luftfahrt* und *Geisteswissenschaften*



<sup>11</sup> Um eine klare Abgrenzung zur Luftfahrt zu verdeutlichen, bei welcher die Raumvorstellungsfähigkeit als Kriterium gehandhabt wird, werden alle Studienrichtungen, in welchen die Raumvorstellung (Psychologie, Publizistik, etc.) nicht von Bedeutung ist, unter dem Sammelbegriff „Geisteswissenschaft“ zusammengefasst.

### Geschlecht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	männlich	31	52,5	52,5	52,5
	weiblich	28	47,5	47,5	100,0
Total		59	100,0	100,0	

Tabelle 10.3: Geschlechteranteil der Gesamtstichprobe (n = 59)

In Bezug auf die Geschlechtsverteilung konnten in beiden Extremgruppen signifikante Unterschiede gefunden werden. Während in der Gruppe der Geisteswissenschaften der Anteil weiblicher Teilnehmerinnen ( $p < .05$ ;  $n_{\text{♀}} = 17$  versus  $n_{\text{♂}} = 7$ ) höher ist, ist dies in der Gruppe der Luftfahrt umgekehrt der Fall; hier dominiert der Männeranteil signifikant ( $p < .05$ ;  $n_{\text{♀}} = 11$  versus  $n_{\text{♂}} = 24$ ). Die entsprechenden SPSS-Outputs 10.2. und 10.3 geben die Ergebnisse der Chi-Quadrat-Tests (bei  $\alpha = 5\%$ ) beider Extremgruppen wieder.

SPSS-Output 10.2: Chi-Quadrat-Test zur Geschlechterverteilung der Extremgruppe Geisteswissenschaften (n = 24)

SPSS-Output 10.3: Chi-Quadrat-Test zur Geschlechterverteilung der Extremgruppe Luftfahrt (n = 35)

#### Test Statistics<sup>b</sup>

	Geschlecht
Chi-Square	4,167 <sup>a</sup>
df	1
Asymp. Sig.	,041

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 12,0.

b. Extremgruppen = Geisteswissenschaften

#### Test Statistics<sup>b</sup>

	Geschlecht
Chi-Square	4,829 <sup>a</sup>
df	1
Asymp. Sig.	,028

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 17,5.

b. Extremgruppen = Luftfahrt

Im Hinblick auf die Gesamtstichprobe n = 59 konnte kein signifikanter Unterschied ( $p = .696$ ) in der Geschlechterverteilung gefunden werden.

Wie bereits im Abschnitt 9.1 angeführt, wurden je nach formulierter Hypothese weitere Extremgruppen aus der Gesamtstichprobe (n = 59) gebildet, wobei die mittleren 50% nicht berücksichtigt werden.

Somit ergeben sich drei weitere Extremgruppen entsprechend der Hypothesen 2 - 4 (vgl. Kapitel 7) wie sie in der nachfolgenden Grafik zur Veranschaulichung

(Abbildung 10.2) dargestellt werden und auf welche im Weiteren auch eingegangen wird.

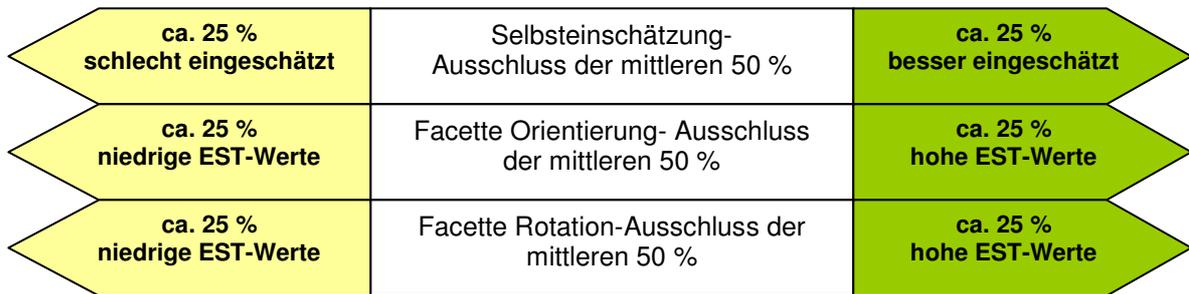


Abbildung 10.2: Bildung weiterer Extremgruppen (aus n = 59) durch Ausschluss der mittleren 50%.

### 10.1.1 Extremgruppenbildung nach der Selbsteinschätzung

Das Extrahieren der zwei Extremgruppen, bestehend aus Personen die sich auf einer Analogskala hinsichtlich ihrer räumlichen Vorstellungsfähigkeit als gut bzw. schlecht einschätzen, erfolgte mit Hilfe von Quartilen:

Im ersten Quartil befinden sich diejenigen Testpersonen, die sich auf der Analogskala vom schlechten Ende bis maximal 51% der Skala zugeschrieben haben, im zweite jene von 51% bis 73%, im dritten von 73% bis 83% und zuletzt befinden sich jene Testpersonen im vierten Quartil, die sich minimal 83% und aufwärts (bis 100%) der Analogskala zugeschrieben haben– vgl. SPSS-Output 10.4.

**Statistics**

Selbsteinschätzung

N	Valid	59
	Missing	0
Percentiles	25	51,00
	50	73,00
	75	83,00

SPSS-Output 10.4: Überblick über die Perzentile zur Extrahierung von Extremgruppen nach der selbsteingeschätzten räumlichen Fähigkeit

Wie bereits oben angeführt, bilden nur Personen die sich als gut in ihrer räumlichen Vorstellungsfähigkeit einschätzen eine Extremgruppe; in die zweite Extremgruppe gehen Personen ein, die sich als schlecht in ihrer räumlichen

Vorstellungsfähigkeit einschätzen – dies entspricht dem unteren (bis max. 51%) und dem oberen (ab 83%) Quartil.

In die Analyse können somit insgesamt 30 Personen aufgenommen werden; davon sind  $n = 16$  Personen (wovon  $n_{\text{♀}} = 11$  weiblichen und  $n_{\text{♂}} = 5$  männlichen Geschlechts sind), die sich hinsichtlich ihrer räumlichen Vorstellungsfähigkeit als weniger begabt einschätzten und weitere  $n = 14$  Personen ( $n_{\text{♀}} = 6$  und  $n_{\text{♂}} = 8$ ), die sich als gut in dieser eingeschätzt hatten (vgl. Abb. 10.3).

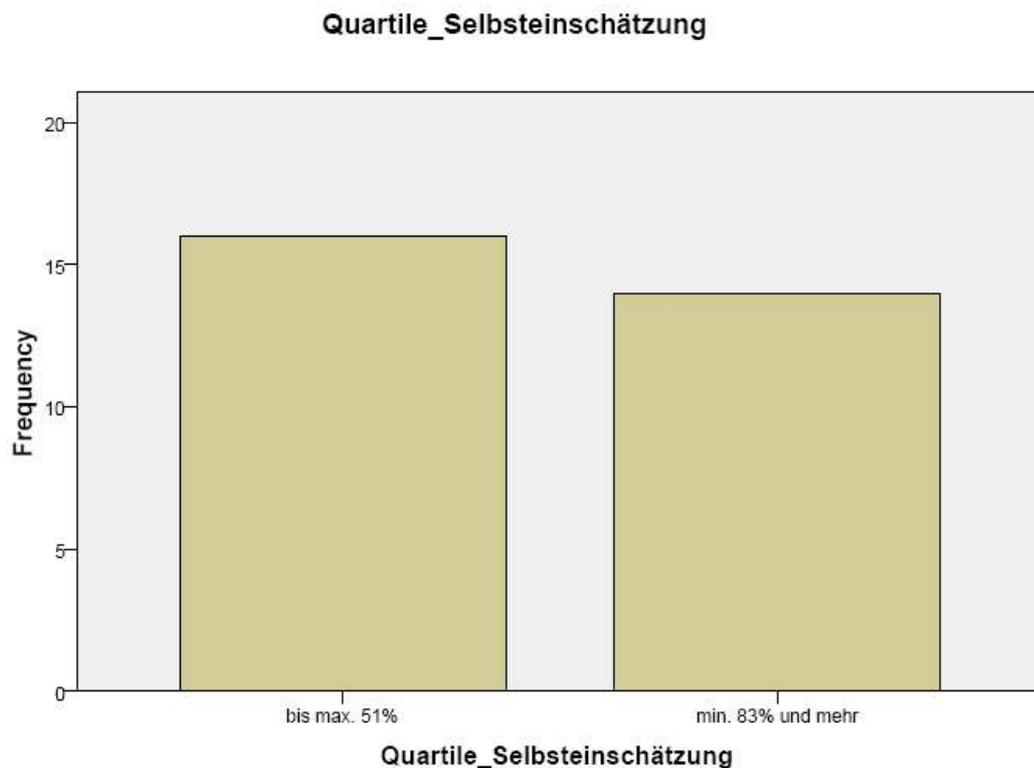


Abbildung 10.3: Extremgruppenbildung nach der Selbsteinschätzung ( $n = 30$ )

In weiterer Folge werden diese beiden Extremgruppen im Hinblick auf ihre Ergebnisse im TARV-F 4.0 miteinander verglichen. Um einen möglichen Einfluss der Testvorgabe ausschließen zu können – mit der Annahme, dass sich diese aufgrund der Testreihenfolge auf die Selbsteinschätzung (vgl. Abschnitt 9.1.2.2-diese wird nur vor der Bearbeitung des EST erfasst) auswirken könnte – soll dieser bei der Analyse berücksichtigt und anhand einer partiellen Korrelation kontrolliert werden.

### 10.1.2 Extremgruppenbildung nach dem EST-Gesamtrahwert

Hierbei wurden, wie auch schon in Abschnitt 10.1.1 beschrieben, die Extremgruppen ebenfalls auf Basis von Quartilen gebildet.

Bei den zur Hypothesenprüfung (3. und 4. Hypothese) notwendigen Testpersonen handelt es sich um jene, deren EST-Gesamtrahwert sich im untersten beziehungsweise obersten Quartil wiederfand, die restlichen 50% wurden aus der Analyse ausgeschlossen. Wie aus der Berechnung hervorgeht, erzielten  $n = 15$  (davon sind  $n_{\text{♀}} = 9$  weiblichen und  $n_{\text{♂}} = 6$  männlichen Geschlechts) Testpersonen niedrige und  $n = 14$  (davon sind  $n_{\text{♀}}=5$  und  $n_{\text{♂}}=9$ ) Personen hohe Testkennwerte im EST (vgl. Abbildung 10.4). Aus den insgesamt  $n = 29$  Personen stammen  $n = 12$  aus der Gruppe der Geisteswissenschaften und  $n = 17$  Personen aus der Gruppe der Luftfahrt.

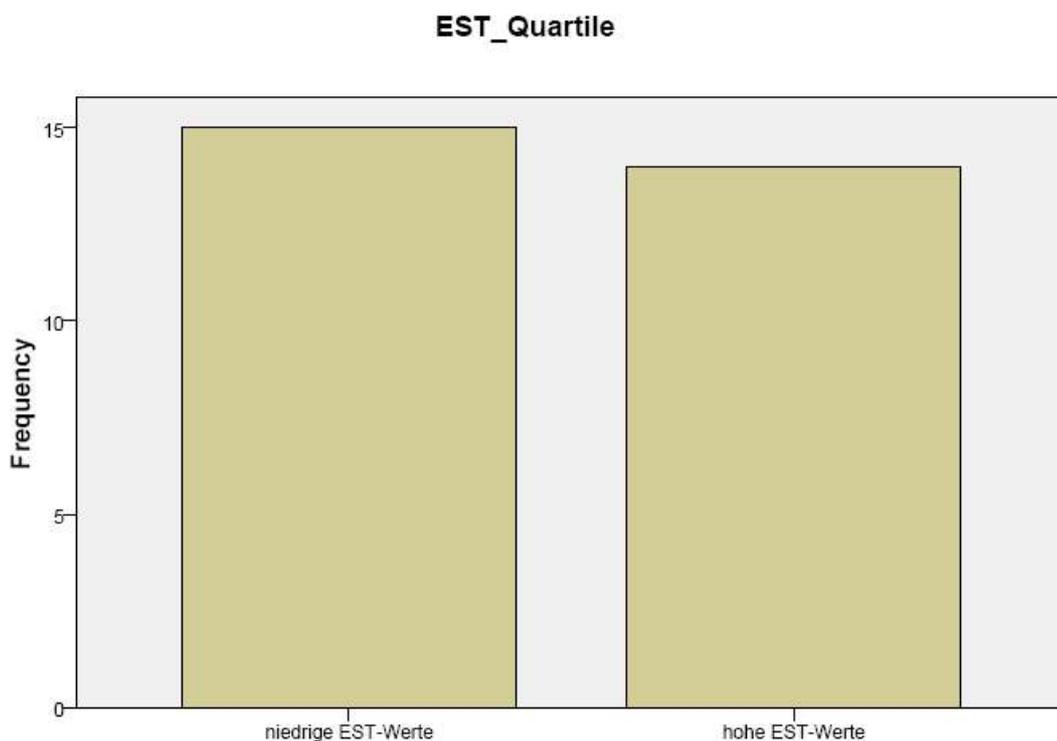


Abbildung 10.4: Extremgruppenbildung nach dem EST- Gesamtrahwert

In weiterer Folge sollen diese neu gebildeten Extremgruppen hinsichtlich ihrer Ergebnisse im TARV-F 4.0 auf der Facette Rotation und Orientierung miteinander verglichen werden.

# 11 Ergebnisdarstellung

Nachfolgend sollen die Ergebnisse der einzelnen Hypothesen, entsprechend der zu vergleichenden Extremgruppen, welche zuvor gebildet wurden (vgl. Abschnitt 10.1), dargestellt werden.

Die Analyse erfolgte mit Hilfe des Statistikprogramms SPSS (Version 17.0). Die Haupthypothese und alle weiteren Nebenhypothesen werden mittels *Welch*-Tests berechnet; dabei wird immer die Ausgabezeile *Varianzen sind nicht gleich* betrachtet (vgl. Rasch & Kubinger, 2006). Die abhängige Variable stellt jeweils der Gesamtrohwert des Tests bzw. der Facette, auf welchem/welcher der Vergleich beruht, dar.

## 11.1 Luftfahrt versus Geisteswissenschaft

Vom Gesamtstichprobenumfang ( $n = 59$  Personen) können für die Extremgruppenvalidierung insgesamt  $n_L = 35$  Personen der Gruppe *Luftfahrt* und  $n_G = 24$  Personen der Gruppe *Geisteswissenschaften* (vgl. Abschnitt 10.1) zugewiesen werden.

Die Berechnung des *Welch*-Tests, mit dem erzielten Gesamtrohwert im TARV-F 4.0 als abhängige Variable, wird im SPSS-Output 11.1 dargestellt.

Durchschnittlich erzielten Personen aus der Extremgruppe *Luftfahrt* ( $M = 2.857$ ;  $SD = 1.375$ ), im Vergleich zu Personen aus dem Bereich der *Geisteswissenschaften* ( $M = 2.042$ ;  $SD = 1.3345$ ), signifikant höhere Werte (hohe Werte sprechen für ein gutes räumliches Vorstellungsvermögen) im Gesamtrohwert des TARV-F 4.0;  $p = .0135$ ,  $p < .05$ .

		Independent Samples Test									
		Levene's Test for Equality of Variances		t-Test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
TARV Gesamtrohwert	Equal variances assumed	.507	.480	2,264	57	.027	,81548	,36012	,09435	1,53660	
	Equal variances not assumed			2,277	50,555	.027	,81548	,35808	,09645	1,53451	

SPSS-Output 11.1: Welch-Test des Extremgruppenvergleichs: Luftfahrt vs. Geisteswissenschaften

Aufgrund des signifikanten Ergebnisses kann die Alternativhypothese angenommen werden, wonach Personen aus dem Bereich der *Luftfahrt* eine

generalisiert höhere Raumvorstellungsfähigkeit aufweisen, als Personen mit geisteswissenschaftlichem Schwerpunkt.

Wie jedoch aus der deskriptiven Statistik und die vorab berechneten Chi-Quadrat-Tests (vgl. Abschnitt 10.1) hervorgeht, besteht ein signifikanter Unterschied in der Geschlechterverteilung innerhalb beider Extremgruppen.

Aufgrund dieser Tatsache wurde zusätzlich eine punktbiserale Korrelation berechnet mit den Variablen Geschlecht und Gesamtrohscore; es besteht ein signifikanter Zusammenhang (das Vorzeichen ergibt sich aufgrund der Codierung des Geschlechts) zwischen dem Geschlecht und der Leistung im TARV-F 4.0 ( $r_{pb} = -.285$ ,  $p < .05$ ;  $r^2 = .081$ ). In diesem Fall macht das Geschlecht 8% der Varianz im Gesamtrohscore des TARV-F 4.0 aus.

Bei zusätzlicher Betrachtung der Verteilung der Ergebnisse ist erkennbar, dass in beiden Extremgruppen der Bereich der gelösten Items von 0 bis 5 reicht (vgl. Abb. 11.1); mit Ausnahme einer Person der Extremgruppe *Luftfahrt*, welche insgesamt sechs Items richtig bearbeiten konnte.

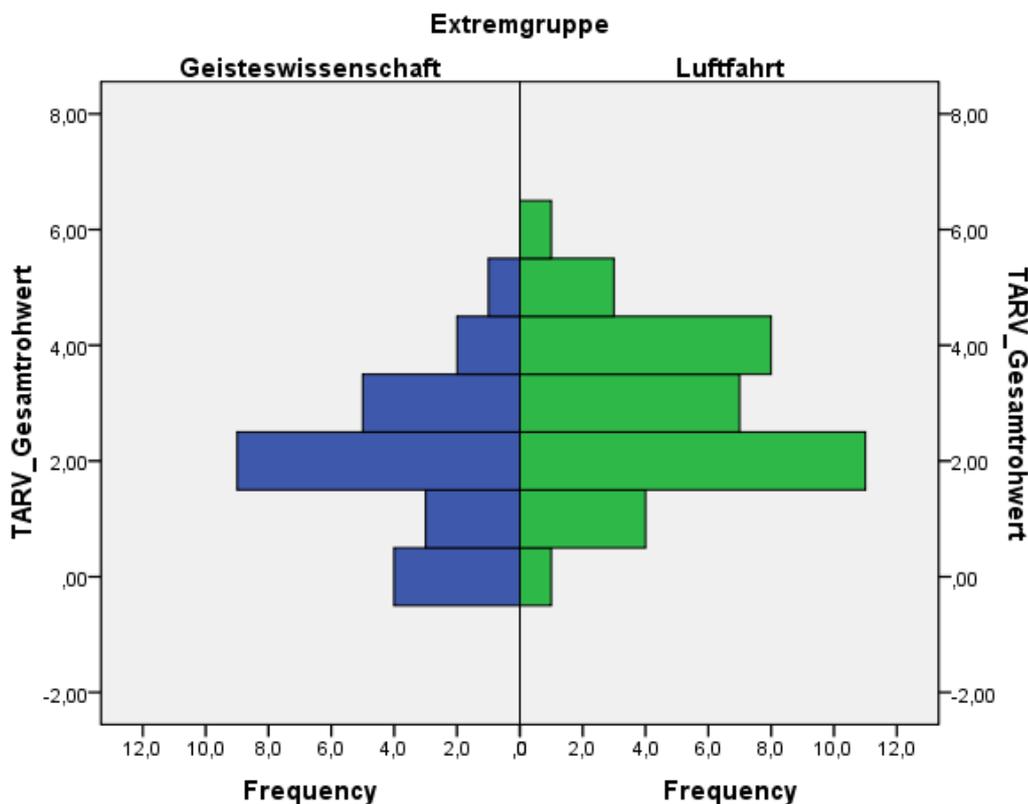


Abbildung 11.1: Häufigkeitsverteilung des Gesamtrohwertes im TARV-F 4.0 (Geisteswissenschaft vs. Luftfahrt)

Anhand der Abbildung ist weiters ersichtlich, dass Personen aus der Extremgruppe *Luftfahrt*, im Vergleich zu jenen aus der Extremgruppe *Geisteswissenschaften*, öfters mindestens zwei Aufgaben richtig gelöst haben (vgl. Tab. 11.1); aber auch im Bereich höherer Lösungshäufigkeiten (3 oder 4 richtige Aufgaben) ist diese Guppe im Vorteil.

**TARV\_Gesamtrohwert<sup>a</sup>**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	4	16,7	16,7	16,7
1,00	3	12,5	12,5	29,2
2,00	9	37,5	37,5	66,7
3,00	5	20,8	20,8	87,5
4,00	2	8,3	8,3	95,8
5,00	1	4,2	4,2	100,0
Total	24	100,0	100,0	

a. Extremgruppen = Geisteswissenschaften

**TARV\_Gesamtrohwert<sup>a</sup>**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	1	2,9	2,9	2,9
1,00	4	11,4	11,4	14,3
2,00	11	31,4	31,4	45,7
3,00	7	20,0	20,0	65,7
4,00	8	22,9	22,9	88,6
5,00	3	8,6	8,6	97,1
6,00	1	2,9	2,9	100,0
Total	35	100,0	100,0	

a. Extremgruppen = Luftfahrt

Tabelle 11.1: Übersicht über die Anzahl richtig gelöster Items im TARV-F 4.0 pro Extremgruppe

### **Nachtragshypothese: Facettenorientierter Zugang aus Sicht der Luftfahrt**

Wie aus dem in Kapitel 6 bereits erörterten Anforderungen zukünftiger PilotInnen und LotsInnen hervorgeht, so scheint eine facettenorientierte Sichtweise dahingehend nötig zu sein, als das diese Personen im Rahmen des Qualifyings und Auswahlverfahrens auch tatsächlich nach jenen Fähigkeitsaspekten ausgewählt werden, welche auch tatsächlich im Zusammenhang mit ihrem zukünftigen Arbeitsplatz benötigt werden.

Aufgrund des signifikanten Unterschiedes zwischen Personen aus der Extremgruppe der *Luftfahrt* und jener aus der Gruppe der *Geisteswissenschaft* im Gesamtrohwert des TARV-F 4.0, sollen nachfolgend die einzelnen Bereiche des

TARV-F 4.0 darauf geprüft werden, in welchen Aspekten des TARV-F 4.0 Personen aus dem Bereich der *Luftfahrt* besser sind.

Wird nun für jede einzelne Facette bzw. Subdimension des TARV-F 4.0 ein einseitiger *Welch*-Test (die Einseitigkeit ergibt sich aufgrund der berufsrelevanten Anforderung und der Vorannahme, dass Personen aus dem Bereich der Luftfahrt bessere Leistungen erzielen) berechnet, ergibt sich ein signifikanter Unterschied einzig in der Subdimension *Schließen von 2D auf 3D* zwischen den beiden Extremgruppen; vgl. SPSS Output 11.2. Durchschnittlich erzielten Personen aus der Gruppe der *Luftfahrt*, im Vergleich zu Personen aus der Gruppe der Geisteswissenschaften, einen signifikant höheren Rohscore im Schließen von 2D auf 3D;  $p = .0135$ ,  $p < .05$ .

In den übrigen Aspekten des TARV-F 4.0 zeigen sich keine Unterschiede zwischen den beiden Extremgruppen.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
S1_Rohscore	Equal variances assumed	,257	,614	-1,724	57	,090	-,437	,253	-,944	,071
	Equal variances not assumed			-1,712	48,370	,093	-,437	,255	-,950	,076
S2_Rohscore	Equal variances assumed	,025	,874	-1,349	57	,183	-,245	,182	-,609	,119
	Equal variances not assumed			-1,369	52,016	,177	-,245	,179	-,605	,114
S3_Rohscore	Equal variances assumed	,170	,682	-,732	57	,467	-,133	,182	-,498	,231
	Equal variances not assumed			-,749	53,265	,457	-,133	,178	-,490	,224
3D auf 2D	Equal variances assumed	,010	,921	-,837	57	,406	-,164	,196	-,557	,229
	Equal variances not assumed			-,860	53,867	,393	-,164	,191	-,547	,219
2D auf 3D	Equal variances assumed	,564	,456	-2,340	57	,023	-,651	,278	-1,208	-,094
	Equal variances not assumed			-2,289	45,647	,027	-,651	,284	-1,224	-,078

SPSS-Output 11.2: Welch-Test des Extremgruppenvergleichs für die Facetten und Subdimensionen des TARV-F 4.0

## 11.2 Selbsteinschätzung zur räumlichen Vorstellungsfähigkeit

Zur Prüfung der hier zugrundeliegenden Hypothese wurde von der Gesamtstichprobe ( $n = 59$ ) – unter Ausschluss der mittleren 50% und unter Berücksichtigung der oberen und unteren 25% (Personen die sich als „sehr schlecht“ oder „sehr gut“ einschätzen) – ein neuer Stichprobenumfang von insgesamt  $n_{\text{Selbsteinschätzung}} = 30$  Personen errechnet. Davon schätzten sich  $n = 14$  Personen hoch und  $n = 16$  Personen niedrig in ihrer räumlichen Vorstellungsfähigkeit ein (vgl. Abschnitt 10.1.1).

Es zeigt sich kein signifikanter Unterschied zwischen diesen beiden, anhand der Selbsteinschätzung gebildeten, Extremgruppen im Gesamtrahwert des TARV-F 4.0 (vgl. SPSS Output 11.3). Die durchschnittliche Leistung im Gesamtrahwert des TARV-F 4.0 jener Testpersonen, die sich in ihrer Fähigkeit zur Raumvorstellung hoch einschätzten liegt bei  $M = 2.786$  Punkten ( $SD = 1.578$ ) und jene von Personen, die sich eher schlecht (unteren 25%) einschätzten bei  $M = 2.00$  ( $SD = 1.155$ ) erzielten Punkten.

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
TARV_Gesamtrahwert	Equal variances assumed	2,755	,108	1,570	28	,128	,78571	,50043	-,23938	1,81080
	Equal variances not assumed			1,538	23,557	,137	,78571	,51100	-,26998	1,84141

SPSS-Output 11.3: Welch-Test des Extremgruppenvergleichs in der Selbsteinschätzung

Zur Prüfung der Annahme, wonach die Testreihenfolge (EAT oder TAE) einen möglichen Einfluss auf die Selbsteinschätzung ( $n_{\text{Selbsteinschätzung}} = 30$ ) und die Leistung im TARV-F 4.0 haben könnte, wurde weiters eine einseitige partielle Korrelation berechnet, unter Kontrolle der Testreihenfolge, bei welcher kein signifikanter Zusammenhang zwischen diesen beiden Variablen ersichtlich wurde (vgl. SPSS-Output 11.4).

Aufgrund des nicht signifikanten Ergebnisses muss die Nullhypothese beibehalten werden: Die anhand der Selbsteinschätzung gebildeten Extremgruppen unterscheiden sich im Hinblick auf ihr erzieltes Gesamtergebnis im TARV-F 4.0 nicht voneinander.

Correlations			Selbsteinschätzung	TARV_Gesamtrahwert
<b>Control Variables</b>				
Reihenfolge der Testvorgabe	Selbsteinschätzung	Correlation	1,000	,278
		Significance (1-tailed)	.	,072
		df	0	27
TARV_Gesamtrahwert	TARV_Gesamtrahwert	Correlation	,278	1,000
		Significance (1-tailed)	,072	.
		df	27	0

SPSS-Output 11.4: Partielle Korrelation

### 11.3 Facette *Orientierung*

Zwei weitere Extremgruppen konnten mit einem Umfang von  $n = 15$  Personen mit niedrigen EST-Testwerten ( $M = .40$ ;  $SD = .632$ ) und  $n = 14$  Testpersonen ( $M = .79$ ;  $SD = .699$ ) mit hohen EST-Werten identifiziert werden (gefragt waren die ca. 25% der schlechtesten und die ca. 25% besten, also jene Personen die im EST die schlechtesten und die besten Werte erzielt haben). Wird nun für diese beiden Gruppen ein einseitiger *Welch*-Test berechnet, mit der Testleistung der Testpersonen in der Facette *Orientierung* als abhängige Variable, so zeigt sich ein nicht signifikanter Unterschied zwischen diesen beiden Extremgruppen ( $p > .05$ ). SPSS-Output 11.5 veranschaulicht die Ergebnisse des *Welch*-Tests.

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
S3_Rohscore	Equal variances assumed	.046	.831	1,560	27	.130	.386	.247	-.122	.893
	Equal variances not assumed			1,554	26,230	.132	.386	.248	-.124	.896

SPSS-Output 11.5: Welch-Test des Extremgruppenvergleichs für die Facette Orientierung des TARV-F 4.0

Personen mit hohen Testwerten im EST erzielen im Gesamtrahwert (hoher Wert bedeutet bessere Orientierungsfähigkeit) der Facette *Orientierung* des TARF-F 4.0, keine besseren Ergebnisse, als Personen die niedrigere EST-Werte erzielten. Die Nullhypothese wird daher beibehalten.

### 11.4 Facette *Rotation*

Auch auf der Facette *Rotation* lässt sich kein signifikanter Unterschied zwischen den beiden Extremgruppen feststellen. Wie auch in der vorhergehenden Hypothesenprüfung wurden auch hier die Extremgruppen auf Grundlage der im EST erreichten Testwerte ( $n = 14$  Testpersonen mit hohen Werten und  $n = 15$  Personen mit niedrigen Werten) gebildet (vgl. Abschnitt 11.3). Der Mittelwertunterschied zwischen den beiden Extremgruppen erweist sich als recht gering. Der einseitige *Welch*-Test mit dem Gesamtrahwert der Facette *Rotation* des TAFV-F 4.0 als abhängige Variable zeigt einen nicht signifikanten Unterschied ( $p = .3325$ ;  $p > .05$ ).

Die Ergebnisse werden im SPSS-Output 11.6 dargestellt.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
S2_Rohscore	Equal variances assumed	,014	,908	,438	27	,665	,105	,239	-,385	,595
	Equal variances not assumed			,438	26,823	,665	,105	,239	-,386	,595

SPSS-Output 11.6: Welch-Test des Extremgruppenvergleichs für die Facette Rotation des TARV-F 4.0

Aufgrund des vorliegenden Ergebnisses und des daraus ableitbaren nicht signifikanten Unterschiedes zwischen den vorliegenden Extremgruppen, muss die Nullhypothese beibehalten werden; Personen mit hohen Testwerten im EST erzielen im Gesamtrohwert (hoher Wert bedeutet bessere Fähigkeit zur mentalen Rotation) der Facette *Rotation* des TARF-F 4.0, keine besseren Ergebnisse, als Personen, die niedrigere EST-Werte erzielten.

### 11.5 Einfluss des Geschlechts auf die Leistung im EST und TARV-F 4.0

Wie aus zahlreichen Vorstudien hervorgeht, so scheint das Geschlecht im Rahmen der Raumvorstellung einen besonderen Stellenwert einzunehmen. Dabei fallen vor allem die Unterschiede hinsichtlich der Fähigkeit zur mentalen Rotation ins Auge, welche beispielsweise auf mögliche Trainingseffkte und Vorgabemodi zurückzuführen wären (vgl. Kapitel 4).

Diesen sollen sich die nachfolgenden Abschnitte widmen.

#### 11.5.1 Geschlechtsunterschiede im EST

Ziel des EST ist die Erfassung der zugrundeliegenden Fähigkeit zur mentalen Rotation, wobei der Aspekt der Orientierung (aufgrund der Tatsache, dass dieser auf den Schlauchfiguren aufbaut) ebenfalls eine Rolle spielen und eine totale Abgrenzung dennoch schwierig erscheinen kann.

Wie bereits in Vorstudien berichtet (vgl. Glück et al, 2010; Weiss et al, 2003), scheint das zeitlich vorgegebene Limit bei der Bearbeitung von Raumvorstellungstests, vor allem in Bezug auf die mentale Rotation, für einen Geschlechtsunterschied verantwortlich zu sein.

Nachfolgend soll dieser Aspekt auch im Rahmen der vorliegenden Arbeit zusätzlich überprüft werden. Dabei soll erfasst werden, inwiefern sich die angesprochenen Geschlechtsunterschiede in der hier dargebotenen Version des EST zugunsten von Männern zeigen.

Von den insgesamt  $k = 15$  Items (ohne Warmup) wurden  $k = 7$  Items unter der *Power*-Bedingung und  $k = 8$  Items unter der *Worklimit*-Bedingung vorgegeben. Die Personen wurden in der *Worklimit*-Bedingung – mit der entsprechenden Instruktion – darauf aufmerksam gemacht, die nachfolgenden Aufgaben so rasch wie möglich zu bearbeiten, ohne jedoch eine genaue Zeitvorgabe erhalten zu haben.

Aufgrund einer unvollständigen Datenmatrix konnte für die abhängige Variable *Gesamtscore unter Power-Bedingung* kein Box-M-Test berechnet werden. Aus diesem Grund erfolgt die Analyse mit einem *Welch*-Test, mit dem Geschlecht als unabhängige Variable. Für die beiden anderen abhängigen Variablen *Gesamtscore im EST* und *Gesamtscore unter Worklimit-Bedingung* konnte ein nicht signifikanter Box-M-Test und somit eine multivariate Varianzanalyse (MANOVA) mit den festen Faktoren Geschlecht (Faktor A) und Ausbildungsgruppe (Faktor B; Luftfahrt und Geisteswissenschaften) berechnet werden.

Für beide Variablen konnte keine Signifikanz auf dem Faktor B gefunden werden. Obwohl für den Gesamtrahwert des EST kein signifikanter Geschlechtsunterschied erkennbar ist, gilt dies nicht für die *Worklimit-Bedingung*: durchschnittlich erzielten die männlichen Teilnehmer ( $n = 30$ ;  $M = 3.84$ ;  $SD = 1.614$ ), im Vergleich zu den weiblichen Testpersonen ( $n = 28$ ;  $M = 2.71$ ;  $SD = 1.512$ ), signifikant bessere Leistungen; im nachfolgenden SPSS-Output 11.7 werden die Ergebnisse der multivariaten Varianzanalyse wiedergegeben.

SPSS-Output 11.7: Zweifaktorielle Varianzanalyse mit den Faktoren EST und Worklimit  
(bei  $\alpha = .05$ )

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Gesamtscore EST (ohne Warmup, k=15)	30,641 <sup>a</sup>	3	10,214	1,626	,194
	Worklimitscore k=8 Items	18,744 <sup>b</sup>	3	6,248	2,459	,072
Intercept	Gesamtscore EST (ohne Warmup, k=15)	1602,382	1	1602,382	255,087	,000
	Worklimitscore k=8 Items	507,347	1	507,347	199,650	,000
Sex_gr	Gesamtscore EST (ohne Warmup, k=15)	24,011	1	24,011	3,822	,056
	Worklimitscore k=8 Items	13,908	1	13,908	5,473	,023
Extremgruppen	Gesamtscore EST (ohne Warmup, k=15)	,166	1	,166	,026	,871
	Worklimitscore k=8 Items	,099	1	,099	,039	,844
Sex_gr * Extremgruppen	Gesamtscore EST (ohne Warmup, k=15)	,889	1	,889	,142	,708
	Worklimitscore k=8 Items	,058	1	,058	,023	,880
Error	Gesamtscore EST (ohne Warmup, k=15)	345,494	55	6,282		
	Worklimitscore k=8 Items	139,765	55	2,541		
Total	Gesamtscore EST (ohne Warmup, k=15)	2347,000	59			
	Worklimitscore k=8 Items	803,000	59			
Corrected Total	Gesamtscore EST (ohne Warmup, k=15)	376,136	58			
	Worklimitscore k=8 Items	158,508	58			

a. R Squared = ,081 (Adjusted R Squared = ,031)

b. R Squared = ,118 (Adjusted R Squared = ,070)

Das Ergebnis des *Welch*-Tests für die verbleibene Variable (*Gesamtscore unter der Power-Bedingung*) weist keinen signifikanten Unterschied zwischen den Geschlechtern auf ( $p = .196$ ; vgl. SPSS-Output 11.8).

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Powerscore k=7 Items	Equal variances assumed	,206	,651	-,864	57	,391	-,291	,337	-,967	,384
	Equal variances not assumed			-,861	55,510	,393	-,291	,338	-,970	,387

SPSS-Output 11.8: Welch-Test für die Variable *Gesamtscore unter der Power-Bedingung*

11.5.2 Geschlechtsunterschiede im TARV-F 4.0

Um mögliche Geschlechtsunterschiede auch im Raumvorstellungstest TARV-F 4.0 aufzeigen zu können wurde auch hier eine MANOVA berechnet; auch im TARV-F 4.0 konnte kein Box-M-Test, aufgrund einer unvollständigen Datenmatrix, für die abhängigen Variablen Rohwert der Facette *Rotation* sowie die Subdimension

*Schließen von 2D auf 3D* berechnet werden. Aus diesem Grund erfolgt für diese beiden Variablen die Analyse mit einem *Welch*-Test, mit dem Geschlecht als unabhängige Variable.

Die übrigen abhängigen Variablen, *Gesamtrohwert* im TARV-F 4.0, Rohwerte der Facetten *Relationen* (S1) und *Orientierung* (S3) sowie die Subdimension *Schließen von 3D auf 2D* konnten aufgrund eines nicht signifikanten Box-M-Tests mittels einer MANOVA berechnet werden. Als feste Faktoren (mit je zwei Stufen) wurden die Gruppenzugehörigkeit (Faktor A: Luftfahrt versus Geisteswissenschaft), die Testreihenfolge (Faktor B: EAT versus TAE) sowie das Geschlecht (Faktor C: männlich versus weiblich) festgesetzt.

Es zeigt sich ein signifikanter Haupteffekt A mit dem Gesamtrohwert des TARV-F 4.0 als abhängige Variable (vgl. Abschnitt 11.1).

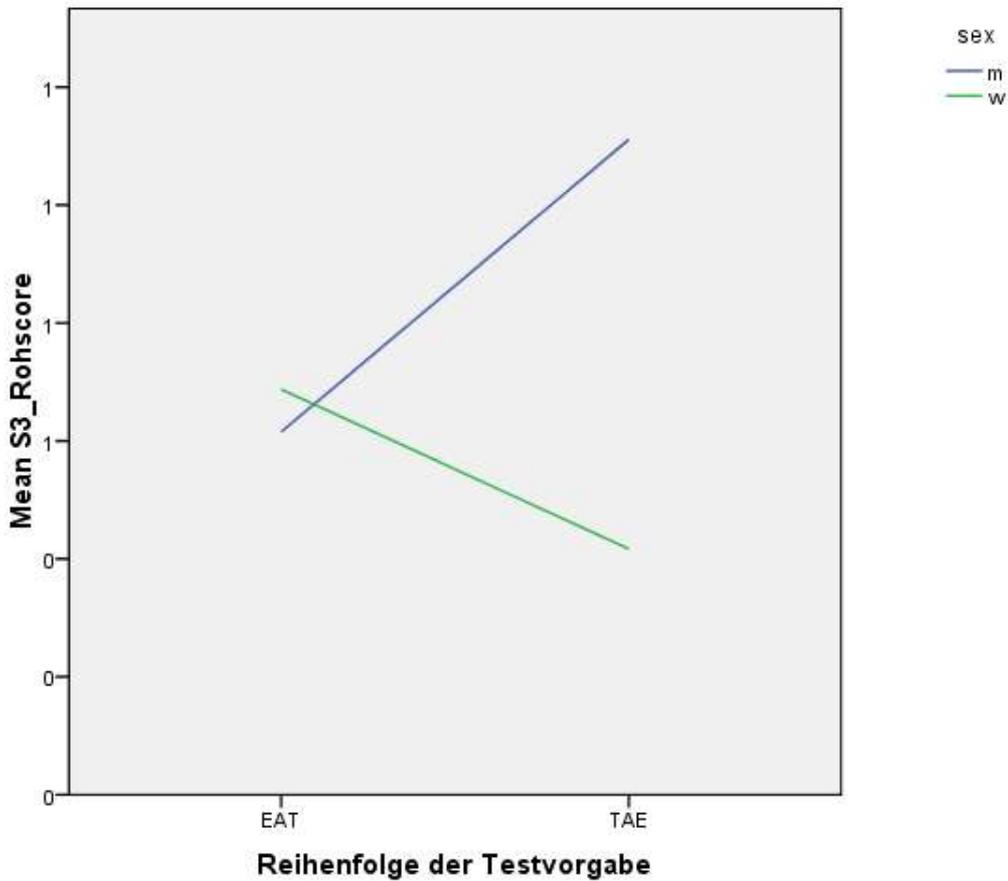
In Zusammenhang mit den übrigen abhängigen Variablen konnte kein signifikanter Haupteffekt gefunden werden.

In Bezug auf die Interaktionen zwischen den Faktoren sind weiters zwei signifikante Interaktionen 1. Ordnung – Interaktion BxC, mit der entsprechenden abhängigen Variable *Gesamtrohwert* im TARV-F 4.0 ( $p = .016$ ) beziehungsweise Rohwert auf der Facette *Orientierung* (S3;  $p = .010$ ) – und eine Interaktion 2. Ordnung (Gruppenzugehörigkeit x Testreihenfolge x Geschlecht- kurz: AxBxC) für die Facette S3 ( $p = .009$ ) erkennbar.

Dabei fällt vor allem die Wechselwirkung auf der Facette S3 ins Auge. Es zeigt sich, dass die Leistungen von den männlichen Testpersonen ( $M = .62$ ;  $SD = .768$ ) sowie weiblichen Testpersonen ( $M = .69$ ;  $SD = .479$ ) bei der EAT-Vorgabe annähernd gleich gut sind; während bei der TAE-Vorgabe die Leistung der männlichen Testpersonen höher ( $M = 1.11$ ;  $SD = .676$ ) ist, ist diese bei den weiblichen Testpersonen ( $M = .42$ ;  $SD = .669$ ) niedriger. Demnach scheint die EAT-Vorgabe weniger Unterschiede zwischen den Geschlechtern hervorzurufen, als die diesbezüglich weniger günstigere TAE-Vorgabe– vgl. Abbildung 11.2.

Ein Ausschnitt des Ergebnisses der mehrfaktoriellen Varianzanalyse wird im SPSS-Output 11.9 veranschaulicht.

Abbildung 11.2: Interaktionsdiagramm BxC; mit der abhängige Variable Rohwert der Facette Orientierung



Die Berechnung eines einseitigen *Welch*-Tests, jeweils für die beiden verbleibenden abhängigen Variablen dem Rohwert der Facette *Rotation* und dem der Subdimension *Schließen von 2D auf 3D* (als unabhängige Variable wurde das Geschlecht festgesetzt – mit der Annahme, dass Männer bessere Leistungen erbringen; vgl. Weitensfelder, et al. (submitted)) ergab folgendes Ergebnis:

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
S2_Rohscore	Equal variances assumed	,098	,756	-2,980	57	,004	-,503	,169	-,842	-,165
	Equal variances not assumed			-3,011	56,446	,004	-,503	,167	-,838	-,169
2D auf 3D	Equal variances assumed	1,222	,274	-1,697	57	,095	-,475	,280	-1,035	,085
	Equal variances not assumed			-1,687	54,407	,097	-,475	,281	-1,039	,089

SPSS-Output 11.9: Welch-Test für die Facette Rotation und die Subdimension "Schließen von 2D auf 3D"

Beide Variablen zeigen einen signifikanten Unterschied (vgl. SPSS-Output 11.10), wonach die durchschnittliche Testleistung der männlichen Testpersonen, im Vergleich zu den Leistungen weiblicher Testpersonen, signifikant besser ist – die entsprechenden deskriptiven Statistiken werden in nachfolgender Tabelle 11.2 dargestellt.

**Group Statistics**

Geschlecht		N	Mean	Std. Deviation	Std. Error Mean
S2_Rohscore	weiblich	28	,46	,576	,109
	männlich	31	,97	,706	,127
2D auf 3D	weiblich	28	1,43	1,136	,215
	männlich	31	1,90	1,012	,182

Tabelle 11.2: Gruppenstatistiken für die Variablen Rotation und Schließen von 2D auf 3D

SPSS-Output 11.10: Mehrfaktorielle Varianzanalyse des TARV-F 4.0 ( $\alpha = .05$ )

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	TARV_Gesamtrohwert	28,647 <sup>a</sup>	7	4,092	2,425	,032
	S1_Rohscore	4,698 <sup>b</sup>	7	,671	,683	,686
	S3_Rohscore	7,047 <sup>c</sup>	7	1,007	2,550	,025
	3D auf 2D	5,337 <sup>d</sup>	7	,762	1,479	,196
Intercept	TARV_Gesamtrohwert	241,551	1	241,551	143,138	,000
	S1_Rohscore	50,190	1	50,190	51,042	,000
	S3_Rohscore	19,520	1	19,520	49,433	,000
	3D auf 2D	26,947	1	26,947	52,273	,000
Extremgruppen	TARV_Gesamtrohwert	7,847	1	7,847	4,650	,036
	S1_Rohscore	2,435	1	2,435	2,476	,122
	S3_Rohscore	,317	1	,317	,803	,374
	3D auf 2D	,263	1	,263	,511	,478
Testreihenfolge	TARV_Gesamtrohwert	1,963	1	1,963	1,163	,286
	S1_Rohscore	,030	1	,030	,031	,862
	S3_Rohscore	1,172	1	1,172	2,967	,091
	3D auf 2D	,000	1	,000	,001	,978
Sex_gr	TARV_Gesamtrohwert	1,891	1	1,891	1,121	,295
	S1_Rohscore	,335	1	,335	,340	,562
	S3_Rohscore	,529	1	,529	1,341	,252
	3D auf 2D	,805	1	,805	1,561	,217
Extremgruppen * Testreihenfolge	TARV_Gesamtrohwert	,654	1	,654	,388	,536
	S1_Rohscore	,060	1	,060	,061	,807
	S3_Rohscore	,020	1	,020	,050	,825
	3D auf 2D	,057	1	,057	,111	,741
Extremgruppen * Sex_gr	TARV_Gesamtrohwert	,133	1	,133	,079	,780
	S1_Rohscore	,957	1	,957	,973	,329
	S3_Rohscore	,001	1	,001	,002	,969
	3D auf 2D	,000	1	,000	,001	,978
Testreihenfolge * Sex_gr	TARV_Gesamtrohwert	10,563	1	10,563	6,259	,016
	S1_Rohscore	,062	1	,062	,063	,802
	S3_Rohscore	2,789	1	2,789	7,063	,010
	3D auf 2D	1,372	1	1,372	2,662	,109
Extremgruppen * Testreihenfolge * Sex_gr	TARV_Gesamtrohwert	5,946	1	5,946	3,524	,066
	S1_Rohscore	,114	1	,114	,116	,735
	S3_Rohscore	2,941	1	2,941	7,448	,009
	3D auf 2D	1,934	1	1,934	3,752	,058
Error	TARV_Gesamtrohwert	86,065	51	1,688		
	S1_Rohscore	50,149	51	,983		
	S3_Rohscore	20,139	51	,395		
	3D auf 2D	26,290	51	,515		
Total	TARV_Gesamtrohwert	491,000	59			
	S1_Rohscore	120,000	59			
	S3_Rohscore	60,000	59			
	3D auf 2D	74,000	59			
Corrected Total	TARV_Gesamtrohwert	114,712	58			
	S1_Rohscore	54,847	58			
	S3_Rohscore	27,186	58			
	3D auf 2D	31,627	58			

a. R Squared = ,250 (Adjusted R Squared = ,147)

b. R Squared = ,086 (Adjusted R Squared = -,040)

c. R Squared = ,259 (Adjusted R Squared = ,158)

d. R Squared = ,169 (Adjusted R Squared = ,055)

## 12 Zusammenfassung und Diskussion

Im Mittelpunkt der vorliegenden Arbeit stand die Validierung der Kurzversion des Raumvorstellungstests *Test zur Angewandten Raumvorstellung – Version F* (für FluglotsInnen und PilotInnen) 4.0 (kurz TARV-F 4.0; Weitensfelder, unpubl.) – mittels Extremgruppen.

Die eingesetzte Testbatterie setzte sich neben dem oben genannten TARV-F 4.0 aus einem weiteren Test zur Messung von Raumvorstellung, dem Rasch-Modell konformen computergenerierten *Endlosschleifentest* (EST; Gittler & Arendasy, unpubl.) und einem *Androgyniefragebogen* (ANDRO; Gittler, unpubl.) zusammen. Beide Raumvorstellungstests wurden an die hier vorliegende Untersuchung adaptiert, indem die Aufgabenanzahl reduziert wurde, sodass eine Bearbeitungsdauer von maximal zwei Stunden nicht überschritten werden sollte. Weiters wurden die Aufgaben des EST nur in der herkömmlichen zweidimensionalen (EST-2D) und nicht in der ebenfalls verfügbaren stereoskopisch-dreidimensionalen (EST-3D) Version vorgegeben.

Die Leistungen von insgesamt 60 Personen konnten erfasst werden, wobei eine Person aus der weiteren Analyse ausgeschlossen werden musste. Von den restlichen 59 Personen, im Alter zwischen 19 und 35 Jahren, befanden sich 35 Testpersonen als sogenannte Trainees in der Ausbildung zum PilotInnen oder FluglotsInnen (n = 11 PilotInnen und n = 24 LotsInnen) und 24 Personen in einer Studienrichtung, in welcher die Fähigkeit zur Raumvorstellung keine besondere Bedeutung zukam; hingegen stellt diese bei der Gruppe von PilotInnen und FluglotsInnen ein wichtiges Kriterium dar, um zur Ausbildung zugelassen zu werden.

Die Bearbeitung der Tests erfolgte in unterschiedlicher Reihenfolge – EAT- oder der TAE-Vorgabe (die Großbuchstaben stehen für den Anfangsbuchstaben des entsprechenden Tests bzw. Verfahrens) – wobei diese zufällig durch das Ziehen von sogenannten Probandencodes erfolgte.

Zwei Fragen sollten mit Hilfe der vorliegenden Untersuchung beantwortet werden: Erstens, *inwiefern ist der Raumvorstellungstest TARV- 4.0 in der Lage, zwischen Extremgruppen zu unterscheiden und zweitens, inwiefern hängt die hier gemessene Auffassung von Raumvorstellung mit jener des EST zusammen?*

Die Extremgruppenvalidierung zur Beantwortung des ersten Teils der Fragestellung wurde vorgenommen, indem die Fähigkeit zur Raumvorstellung als Kriterium für die Bildung der Extremgruppen herangezogen wurde; wonach sich Personen aus dem Bereich der Luftfahrt im TARV-F 4.0 von jenen Personen unterscheiden, welche sich in einem geisteswissenschaftlichen Studium oder Psychologiestudium befinden.

Das Ergebnis des einseitigen *Welch*-Tests zeigt einen signifikanten Unterschied zwischen den beiden Extremgruppen;  $p = .0135$  ( $p < .05$ ). Unabhängig vom Geschlecht erzielten Personen aus dem Bereich der *Luftfahrt* signifikant höhere Werte im Gesamtrohwert des TARV-F 4.0.

Die Extremgruppenvalidierung für den ersten Teil der Fragestellung kann als gelungen betrachtet werden.

Aufgrund einer signifikanten Geschlechterverteilung in beiden Extremgruppen wurde zusätzlich eine punktbiserale Korrelation berechnet; es konnte ein signifikanter Zusammenhang zwischen dem Geschlecht und der Leistung im TARV-F 4.0 ( $r_{pb} = -.285$ ,  $p < .05$ ;  $r^2 = .081$ ) gefunden werden.

Ein nachträglich berechneter *Welch*-Test je Bereich des TARV-F 4.0, weist einzig für die Subdimension *Schließen von 2D auf 3D* einen signifikanten Unterschied zwischen den beiden Extremgruppen auf; die durchschnittliche Leistung von Personen aus dem Bereich der *Luftfahrt* betrug  $M = 1.94$  ( $SD = .998$ ) und jene von Personen ohne Raumvorstellung als berufsrelevantes Kriterium  $M = 1.29$ ; ( $SD = 1.122$ ).

Weiters konnte kein signifikanter Unterschied im einseitigen *Welch*-Test zwischen Personen, die sich in ihrer räumlichen Vorstellungsfähigkeit als gut einschätzten von jenen die sich als schlecht einschätzten in den von ihnen erzielten Leistungen im Gesamtrohwert des TARV-F 4.0 gefunden werden ( $p = .137$ ).

Ein nachträglich berechneter einseitiger *Welch*-Test für die Variable Gesamtscore im EST (die Einseitigkeit ergibt sich aufgrund der Annahme, dass Personen, die sich als gut in ihrer räumlichen Vorstellungsfähigkeit einschätzen, auch gute Testkennwerte im EST erzielen) ergab ebenfalls keinen signifikanten Unterschied ( $p = .053$ ) zwischen den mittels Selbsteinschätzung gebildeten Extremgruppen.

Im Hinblick auf den zweiten Teil der Fragestellung, inwiefern die im TARV-F 4.0 gemessene Auffassung von Raumvorstellung mit jener des EST zusammenhängt, waren die zugrundeliegenden Hypothesen, dass die mittels Extremgruppenselektion (nach dem EST-Gesamtscore unter Ausschluss der mittleren 50%) gebildeten Extremgruppen sich hinsichtlich ihrer Testwerte in der jeweils interessierenden Facette des TARV-F 4.0 (*Rotation* und *Orientierung*) unterscheiden.

Von insgesamt  $n = 29$  Personen erreichten  $n = 15$  Testpersonen hohe Testkennwerte im EST. Die Extremgruppenvalidierung beider Facetten kann jedoch aufgrund nicht signifikanter Ergebnisse im einseitigen *Welch*-Test, als nicht gelungen betrachtet werden.

Dem theoretischen Hintergrund des EST zufolge und dem damit verbundenen und empirisch bestätigten Anspruchs auf Eindimensionalität (vgl. Gittler & Arendasy, 2003) sowie den ihm postulierten Lösungsmöglichkeiten der mentalen Rotation oder des mentalen Perspektivenwechsels (vgl. Arendasy, 2000; Stepan, 2008), können diese beiden Lösungsansätze im Rahmen der hier vorliegenden Ergebnisse nicht bestätigt werden. Dies lässt die Annahme zu, wonach die Konstruktion der Items im TARV-F 4.0 eher suboptimal sei und eine Überarbeitung dieser als notwendig erscheint.

Zahlreiche Studien widmeten sich der Frage nach dem Geschlechtsunterschied im Rahmen von Raumvorstellungstests, dabei zeigten sich die größten Unterschiede vor allem in der Fähigkeit zur mentalen Rotation, zum Nachteil von weiblichen Testpersonen. Inwiefern sich dies auch in der vorliegenden Studie aufzeigen lässt, wurde ebenfalls für beide hier eingesetzte Raumvorstellungstests zusätzlich erfasst.

Wie bereits von Weiss et al. (2003) berichtet, sahen auch Glück et al. (2010) das zeitlich vorgegebene Limit als einen möglichen Erklärungsansatz für die Geschlechtsunterschiede bei der Bearbeitung von mentalen Rotationsaufgaben. Dies konnte auch in der vorliegenden Untersuchung in der Subdimension *Worklimit* ( $k = 8$  Items) des EST im Rahmen einer multivariaten Varianzanalyse (MANOVA) mit den Zwischensubjektfaktoren Geschlecht und Ausbildungsgruppe (Extremgruppe; mit vs. ohne Raumvorstellung als Kriterium) bestätigt werden ( $p = .023$ ). Für den Gesamtrohwert des EST ( $p = .056$ ), sowie im für die Power-Bedingung berechneten *Welch*-Test ( $k = 7$  Items;  $p = .393$ ) konnte keine Signifikanz gefunden werden.

Ein varianzanalytischer Vergleich der Leistungen im *Gesamtrohwert des TARV-F 4.0* sowie in den Facetten *Relationen* und *Orientierung* und der Subdimension *Schließen von 3D auf 2D* weist keinerlei Signifikanz auf dem Faktor Geschlecht auf; dies gilt jedoch nicht für die Variablen *Schließen von 2D auf 3D* und *Rotation*, hier zeigt ein je Variable berechneter einseitiger *Welch*-Test Signifikanz auf.

Der Unterschied zwischen männlichen und weiblichen Testpersonen im Zusammenhang mit den im TARV-F 4.0 erfassten Subdimensionen werden (mit Vorbehalt) auf den gefundenen Zusammenhang zwischen der Subdimension *Schließen von 2D auf 3D* und der Facette *Rotation* zurückgeführt (vgl. Weitensfelder et al. submitted). Neubauer et al. (2010) sowie Stepan (2004) führen die Geschlechtsunterschiede auf die Vorgabe des Aufgabenmaterials zurück; wonach im Vergleich zur herkömmlichen zweidimensionalen Darbietung (2D), die dreidimensionale (stereoskopische; 3D) Darbietung von mentalen Rotationsaufgaben weniger Geschlechtsunterschiede hervorrufen würde.

Weiters konnte eine zweifache Wechselwirkung zwischen der Testreihenfolge und dem Geschlecht für die Variablen *Gesamtrohscore des TARV-F 4.0* ( $p = .016$ ) und der Facette *Orientierung* ( $p = .010$ ) sowie eine deutliche Dreifachwechselwirkung zwischen der Ausbildungsgruppe, der Testreihenfolge und dem Geschlecht ebenso für die Variable *Orientierung* ( $p = .009$ ) gefunden werden.

Interessant erscheint hier, dass sich ein deutlicher Unterschied zwischen den Geschlechtern in der Testbearbeitungsreihenfolge abzeichnet. Durchschnittlich ist kaum ein Unterschied zwischen männlichen und weiblichen Testpersonen in der EAT-Testbearbeitung vorhanden, jedoch ist dieser bei der durchschnittlichen

Leistung in der TAE-Vorgabe ausgeprägter und zwar zu Ungunsten der weiblichen Testpersonen. Die hier gefundenen Ergebnisse weisen auf einen möglichen Trainingseffekt der Orientierungsleistung im TARV-F 4.0 hin. Weiters scheinen die weiblichen Testpersonen vom Training mehr zu profitieren, als die männlichen Testpersonen; dies entspricht den Ergebnissen von Neubauer et al. (2010). Warum jedoch die männlichen Testpersonen in der EAT-Vorgabe schlechter werden ist verwunderlich und bedarf weiterer Forschungsarbeit.

In Anbetracht der hier vorliegenden Ergebnisse kann die Extremgruppenvalidierung dahingehend als gelungen betrachtet werden als der TARV-F 4.0 in der Lage ist zwischen Extremgruppen auch tatsächlich zu unterscheiden; somit konnten Personen, welche sich als Trainees in der Ausbildung zu PilotIn und FluglotsIn befinden und bei welchen das räumliche Vorstellungsvermögen dahingehend wichtig erscheint, um die von ihnen abverlangten Berufsanforderungen im Rahmen dieser auch tatsächlich zu erfüllen, eine bessere Leistung im TARV-F 4.0, im Vergleich zu Personen einer Studienrichtung, in welcher das räumliche Vorstellungsvermögen nicht von Bedeutung ist vorweisen. Auch in Zusammenhang mit der Subdimension Schließen von 2D auf 3D, welche für die Berufsgruppe der Fluglotsen dahingehend als wichtig erscheint, dass die Personen in der Lage sind Informationen zum Flugverkehr auch richtig einschätzen zu können, beispielsweise um die erforderlichen Mindestabstände zwischen den Flugzeugen auch einhalten zu können und dadurch für einen reibungslosen Ablauf des Flugverkehrs sorgen zu können, konnten signifikante Unterschiede im (nachträglich berechneten) einseitigen *Welch*-Test (die Einseitigkeit ergibt sich aufgrund der berufsrelevanten Anforderung;  $p = .0135$ ) zwischen den beiden Extremgruppen gefunden werden.

In Zusammenhang mit der von Gruber (2008) postulierten Anforderung der mentalen Rotationsfähigkeit im Pilotenberuf konnten jedoch keinerlei Unterschiede zwischen den Ausbildungsgruppen gefunden werden. Weitensfelder et al. (submitted) führen dies auf die mangelnde Flugerfahrung der Testpersonen zurück.

Um jedoch eine eindeutige Eigenständigkeit den einzelnen Facetten des TARV-F 4.0 zuschreiben zu können, bedarf es weiterer Forschungsarbeit und Überarbeitung des Itemmaterials (vgl. Weitensfelder et al, submitted).



### III Literaturverzeichnis

- Amelang, M. & Bartussek, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6. Aufl.). Stuttgart: Kohlhammer.
- Aschenbrenner, V. (2007). *Belastung und Beanspruchung in organisationalen Veränderungen der FlugverkehrsleiterInnen*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Carroll, J. B. (1993). *Human cognitive abilities: a survey of the factor-analytic studies*. Cambridge [u.a.]: Cambridge Univ. Press.
- Colom, R., Contreras, M.J., Botella, J. & Santacreu, J. (2001). Vehicles of spatial ability. *Personality and Individual Differences*, 32, 903–912.
- Colom, R., Contreras, M. J., Shin, P. C. & Santacreu, J. (2003). The Assessment of Spatial Ability with a Single Computerized Test. *European Journal of Psychological Assessment*, 19/2, 92-100.
- Dabbs, Jr., J. M., Chang, E-L., Strong, R. A. & Milun, R. (1998). Spatial Ability, Navigation Strategy, and Geographic Knowledge Among Men and Women. *Evaluation and Human Behavior*, 19/2, 89-98.
- Eliot, J. (1987). *Models of psychological space: psychometric, developmental and experimental approaches*. New York: Springer.
- Eliot, J., & Smith, I. M. (1983). *An international directory of spatial tests*. Windsor, England: NFER/ Nelson; and Atlantic Highlands, NJ: Humanities Press.

- Eliot, J. (2002). About spatial Intelligence: I. *Perceptual and Motor Skills*, 94, 479-486.
- El-Koussy, A. A. H. (1935). The visual perception of space. *British Journal of Psychology Monograph Supplement*, 7, Nr. 20.
- English, H. B. & English, A. C. (1958). *A comprehensive dictionary of psychological and psychoanalytical terms: A guide to usage*. New York: Mc Kay.
- Furnham, A. & Buchanan, T. (2005): Personality, gender and self-perceived intelligence. *Personality and Individual Differences*, 39, 543-555.
- Furnham, A. & Thomas, C. (2004). Parents gender and personality and estimates of their own and their children's intelligence. *Personality and Individual Differences*, 37, 887–903.
- Furnham, A. & Petrides, K. V. (2004). Parental estimates of five intelligences. *Australian Journal of Psychology*, 56, 10-17.
- Gittler, G. (1990). *Dreidimensionaler Würfelttest (3DW). Ein rasch-skaliertes Test zur Messung des räumlichen Vorstellungsvermögens*. Weinheim: Beltz.
- Gittler, G. (1992). *Testpsychologische Aspekte der Raumvorstellungsforschung- Kritik, Lösungsansätze und empirische Befunde*. Wien: Habilitationsschrift der Universität Wien.
- Gittler, G. (1999). Sind Raumvorstellung und Reasoning sperierbare Fähigkeitsdimensionen? Dimensionalitätsanalysen zweier Raschskalierter Tests: 3DW und WMT. *Diagnostica*, 45/2, 69-81.
- Gittler, G. & Arendasy, M. (2003). Endlosschleifen. Psychometrische Grundlagen des Aufgabentyps EP. *Diagnostica*, 49, 164–175.

- Glück, J. (1999). *Spatial Strategies. Kognitive Strategien bei Raumvorstellungsleistungen*. Unveröffentlichte Dissertation, Universität Wien.
- Glück, J. & Fabrizii, C. (2010). Gender Differences in the Mental Rotations Test Are Partly Explained by Response Format. *Journal of Individual Differences*, 31(2), 106–109.
- Goldstein, A. G. & Chance, J. E. (1965). Effects of practice on sex-related different performance on embedded figures. *Psychonomic Science*, 3, 361-362.
- Gruber, M. (2008). *Einsatz- und Anwendungsmöglichkeiten des 3DW und des EST bei der Selektion und Leistungsüberprüfung von Fluglotsen und Linienspiloten*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Guilford, J. P. (1972). Thurstone's Primary Mental Abilities and Structure-of-Intellect Abilities. *Psychological Bulletin*, 77(2), 129–143.
- Holling, H. & Preckel, F. (2005): Self-estimates of intelligence – methodological approaches and gender differences. *Personality and Individual Differences*, 38, 503-517.
- Hunter, J. E. (1986). Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance. *Journal of Vocational Behavior*, 29, 340- 362.
- Hülshager, U. R. & Maier, G. W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf. Eine Bestandaufnahme internationaler und nationaler Forschung. *Psychologische Rundschau*, 59(2), 108-122.
- Kubinger, K. D. (2009). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. (2. Aufl.). Göttingen: Hogrefe.

- Kubinger, K. D., Rasch, D. & Moder, K. (2009). Zur Legende der Voraussetzungen des t-Tests für unabhängige Stichproben. *Psychologische Rundschau*, 60, 26–27.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. völlig neubearbeitete und erweiterte Auflage). Weinheim: Beltz.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterisation of gender differences in spatial abilities: A meta-analysis. *Child Development*, 56, 1479–1498.
- Lohman, D. F (1979). Spatial ability: A review and reanalysis of the correlational literature (Tech. Rep. No. 8). Stanford, CA: Stanford University School of Education, Aptitude Research Project.
- Lohman, D. F (1994). Spatial ability. In R.J. Sternberg, (Hg.), *Encyclopedia of human intelligence* (Vol. 2, S. 1000- 1007). New York: Macmillan.
- Lohman, D. F. (1996). Spatial Ability and G. In Ian Dennis & Patrick Tapsfield (Hg.), *Human Abilities: Their Nature and Measurement* (97–116). Hillsdale, NJ: Erlbaum.
- Lohman, D. F. & Kyllonen, P. C. (1983). Individual differences in solution strategy on spatial tasks. In D.F. Dillon & R.R. Schmeck (Hg.), *Individual differences in cognition* (S. 105-135). New York: Academic Press.
- Maccoby, E. E. & Jacklin, C. N. (1979). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Neubauer, A., Bergner, S. & Schatz, M. (2010). Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence*, 38(5), 529–539.

- Piaget, J. & Inhelder, B. (1956). *The child's conception of space*. London: Routledge & Keagan Paul.
- Quaiser-Pohl, C. (1998). Die Fähigkeit zur räumlichen Vorstellung: zur Bedeutung von kognitiven und motivationalen Faktoren für geschlechtsspezifische Unterschiede. Münster; New York; München; Berlin: Waxmann.
- Rammstedt, B. & Rammsayer, T. (2002). Die Erfassung von selbsteingeschätzter Intelligenz. Konstruktion, teststatistische Überprüfung und erste Ergebnisse des ISI. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23(4), 435-446.
- Rasch, D. & Kubinger, K. D. (2006). *Statistik für das Psychologiestudium. Mit Softwareunterstützung zur Planung und Auswertung von Untersuchungen sowie zu sequentiellen Verfahren*. München: Spektrum.
- Rost, D. (1976). *Der Begabungsfaktor „Raumvorstellung“*. Theorie und Training. Dissertation, Universität Hamburg.
- Saccuzzo, D. P., Craig, A. S., Johnson, N. E. & Larson, G. E. (1996). Gender differences in dynamic spatial abilities. *Personality and Individual Differences*, 21(4), S. 599-607.
- Santacreu, J. (1999). *SODT-R and SVDT-R: Dynamic computerized tests for the assessment of spatial ability (revised versions)*. Technical Report. Madrid: Universidad Autónoma de Madrid.
- Sherman, J. A. (1967). Problems of sex differences in space perception and aspects of intellectual functioning. *Psychological Review*, 74, 290–299.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

- Stafford, R. E. (1961). Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and Motor Skills*, 13, 428.
- Stepan, A. (2004). *Differentielle Aspekte bei zwei- versus dreidimensionaler Darbietung von Raumvorstellungsaufgaben des Typs EST*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Stumpf, Heinrich & Fay, Ernst (1983). *Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Göttingen: Hogrefe.
- Schmidt-Atzert, L., Deter, B., & Jaeckel, S. (2004). Prädiktion von Ausbildungserfolg: Allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten. *Zeitschrift für Personalpsychologie*, 3(4), 147-158.
- Tewes, U. & Wildgrube, K. (Hrsg.). (1999). *Psychologie-Lexikon (2. Aufl.)*. München: Oldenbourg.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Thurstone, L. L. (1950). Some primary abilities in visual thinking. Psychometric Laboratory Research Report No. 59. Chicago: University of Chicago Press.
- Thurstone, L. L. & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, 2.
- Vandenberg, S. & Kuse, A. R (1978). Mental Rotations, A Group Test of Three Dimensional Spatial Visualization. *Perceptual & Motor Skills*, 47, 599-604.
- Voyer, D., Voyer, S. & Bryden, M. P. (1995). Magnitude of Sex Differences in Spatial Abilities: A Meta-Analysis and Consideration of Critical Variables. *Psychological Bulletin*, 117(2), 250–270.

- Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W. & Delazer, M. (2003): Sex differences in cognitive functions. *Personality and Individual Differences*, 35, 863-875.
- Weitensfelder, L., Grubestic, A., Kubinger, K. D. & Gittler, G. (submitted). Zur Notwendigkeit einer facettenorientierten Raumvorstellungsmessung in der Eignungsbeurteilung aus Gründen der Genderfairness.
- Winkler, P. (2007). *Die praktische Relevanz von Raumvorstellung*. Unveröffentlichte Diplomarbeit, Universität Wien



## IV Abbildungsverzeichnis

Abbildung 3.1: Papier-Falt-Aufgabe .....	23
Abbildung 3.2: Spatial Orientation Dynamic Test (SODT) .....	25
Abbildung 4.1: Itembeispiel des Water Level Tests (WLT).....	28
Abbildung 4.2: Aufgabe zur mentalen Rotation: .....	29
Abbildung 4.3: Beispielitems des Mental Rotation Tests.....	33
Abbildung 4.4: Interaktionsdiagramm.....	34
Abbildung 8.1: Vorgegebene Parameter zur Berechnung des optimalen Stichprobenumfangs .....	50
Abbildung 9.1: Übungsaufgabe 1 des TARV-F 4.0:.....	53
Abbildung 9.2: EST- Beispielitem 4 .....	56
Abbildung 10.1: Verteilung der Gesamtstichprobe nach Extremgruppen: <i>Luftfahrt</i> und <i>Geisteswissenschaften</i> .....	63
Abbildung 10.2: Bildung weiterer Extremgruppen (aus n = 59) durch Ausschluss der mittleren 50%. .....	65
Abbildung 10.3: Extremgruppenbildung nach der Selbsteinschätzung (n = 30) ...	66
Abbildung 10.4: Extremgruppenbildung nach dem EST- Gesamtrohwert .....	67
Abbildung 11.1: Häufigkeitsverteilung des Gesamtrohwertes im TARV-F 4.0.....	69
Abbildung 11.2: Interaktionsdiagramm BxC .....	78

## V Tabellenverzeichnis

Tabelle 4.1: Effektgrößen nach Voyer et al. (1995) .....	31
Tabelle 9.1: Übersicht zur Aufgabenzusammenstellung im TARV-F 4.0.....	53
Tabelle 9.2: Testtermine im Überblick .....	59
Tabelle 9.3: Reihenfolge der Testbearbeitung.....	60
Tabelle 10.1: Altersverteilung über die Gesamtstichprobe (n=59).....	61
Tabelle 10.2: Häufigkeitsverteilung zur Testvorgabe (n=59) .....	62
Tabelle 10.3: Geschlechteranteil der Gesamtstichprobe (n = 59) .....	64
Tabelle 11.1: Übersicht über die Anzahl richtig gelöster Items im TARV-F 4.0 pro Extremgruppe.....	70
Tabelle 11.2: Gruppenstatistiken für die Variablen Rotation und Schließen von 2D auf 3D.....	79

## VI SPSS-Output-Verzeichnis

SPSS-Output 10.1: Chi-Quadrat-Test zur Vorgabeverteilung .....	62
SPSS-Output 10.2: Chi-Quadrat-Test zur Geschlechterverteilung der Extremgruppe Geisteswissenschaften (n = 24) .....	64
SPSS-Output 10.3: Chi-Quadrat-Test zur Geschlechterverteilung der Extremgruppe Luftfahrt (n = 35) .....	64
SPSS-Output 10.4: Überblick über die Perzentile zur Extrahierung von Extrem- gruppen nach der selbsteingeschätzten räumlichen Fähigkeit .....	65
SPSS-Output 11.1: Welch-Test des Extremgruppenvergleichs: Luftfahrt vs. Geisteswissenschaften .....	68
SPSS-Output 11.2: Welch-Test des Extremgruppenvergleichs für die Facetten und Subdimensionen des TARV-F 4.0 .....	71
SPSS-Output 11.3: Welch-Test des Extremgruppenvergleichs in der Selbsteinschätzung .....	72
SPSS-Output 11.4: Partielle Korrelation .....	72
SPSS-Output 11.5: Welch-Test des Extremgruppenvergleichs für die Facette Orientierung des TARV-F 4.0 .....	73
SPSS-Output 11.6: Welch-Test des Extremgruppenvergleichs für die Facette Rotation des TARV-F 4.0 .....	74
SPSS-Output 11.7: Zweifaktorielle Varianzanalyse mit den Faktoren EST und Worklimit .....	76
SPSS-Output 11.8: Welch-Test für die Variable <i>Gesamtscore unter der Power- Bedingung</i> .....	76
SPSS-Output 11.9: Mehrfaktorielle Varianzanalyse des TARV-F 4.0 ( $\alpha = .05$ ) ....	80
SPSS-Output 11.10: Welch-Test für die Facette Rotation und die Subdimension „Schließen von 2D auf 3D“ .....	78



# VII Anhang



## Anhang A: Instruktion des Androgynie-Fragebogens (ANDRO; Gittler, unpubl.)

Fakultät für Psychologie der Universität Wien

Persönlichkeits-Einschätzung Selbstbild

Forschungsgruppe Differentielle Psychologie

**Instruktion**

**Geben Sie bitte an, wie Sie sich selbst - ganz privat - einschätzen.**

**Klicken** Sie dazu auf eines der **grauen Kästchen** unter dem Pfeil. **Je weiter rechts** - in Richtung **Pfeilspitze** - Sie klicken, **desto größer** oder **stärker** ist Ihre **Zustimmung**.

Es gibt weder richtige noch falsche Antworten; arbeiten Sie daher **zügig**.

Bitte bearbeiten Sie jetzt das Beispiel. Korrigieren ist möglich, indem Sie auf ein anderes Kästchen klicken.

	Ich persönlich denke von mir, ...
	
... ich bin motiviert, den Fragebogen ernsthaft zu bearbeiten	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

**WEITER**

Fakultät für Psychologie der Universität Wien

Persönlichkeits-Einschätzung Selbstbild

Forschungsgruppe Differentielle Psychologie

**Zusätzliche Informationen**

Der Fragebogen enthält **63 Aussagen** (Eigenschaften) und dauert ca. 10 Minuten.

Auf jeder der folgenden Bildschirmseiten werden Ihnen sieben Aussagen präsentiert. Klicken Sie erst auf **WEITER**, wenn Sie **alle sieben** Einschätzungen vorgenommen haben. Korrigieren ist innerhalb der gesamten Bildschirmseite möglich.

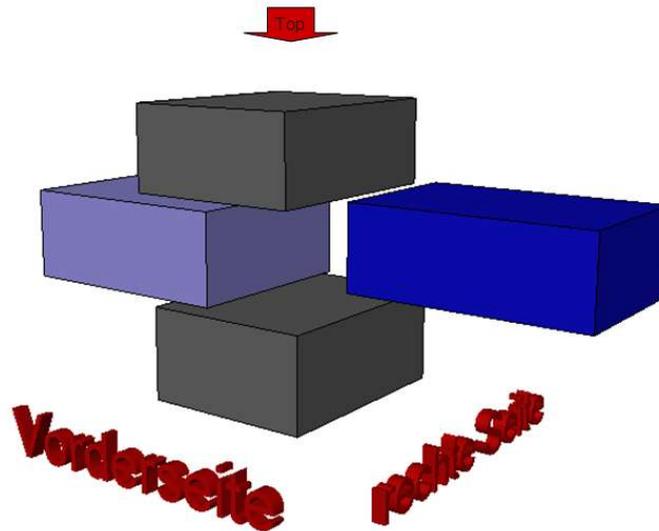
Noch einmal sei darauf hingewiesen, dass es im Folgenden um Ihre **ganz private Einschätzung** geht. Wie sehr Sie einer Aussage **zustimmen**, soll ausschließlich von Ihrer **persönlichen Sichtweise** abhängen.

**WEITER**

## Anhang B– Instruktion des TARV-F 4.0 (inkl. Übungsbeispiel) sowie der Antwortbogen

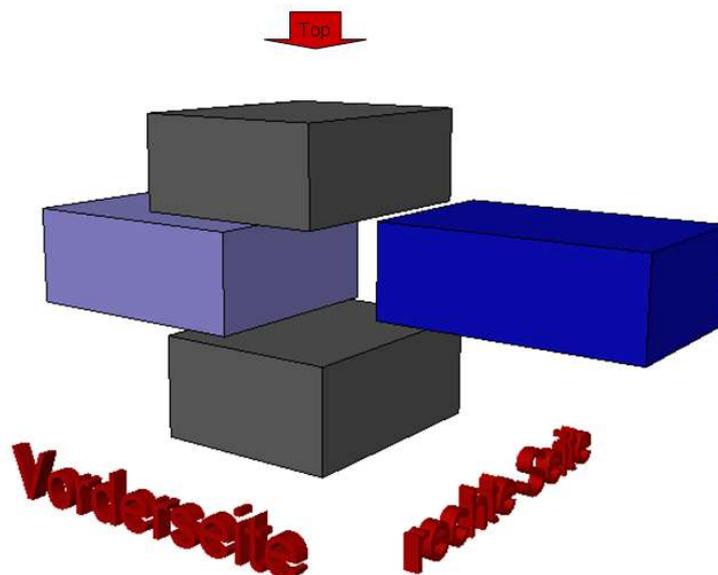
1

Hier sehen Sie eine Figur, die aus mehreren Teilen besteht.  
Zu dieser Figur gibt es verschiedene Pläne von verschiedenen Seiten.  
Ihre Aufgabe auf den nächsten Seiten wird sein, jeweils herauszufinden,  
ob die Pläne richtig oder falsch sind.



2

Die Seiten sind immer aus Ihrer Sicht angegeben:  
Wenn Sie sich vor die Figur stellen (an die Vorderseite), ist die Seite,  
die Sie rechts sehen, die rechte Seite.  
Die Seite, die Sie links sehen würden, wäre die linke Seite.

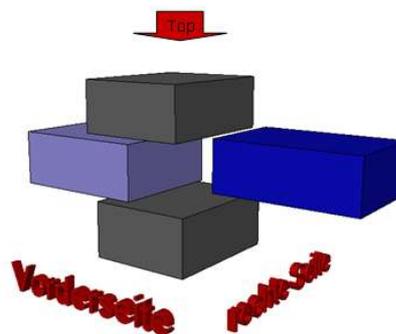


3

Der Plan einer Seite gilt als „richtig“, wenn

- alle Teile im Plan richtig zueinander liegen,
- alle Teile im Plan sich am richtigen Ort befinden und
- alle Teile im Plan zueinander in der richtigen Größe stehen.

3

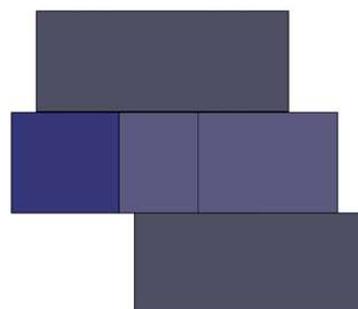


Die zu jeder Figur gehörenden Pläne sind dabei durchscheinend (transparent) dargestellt.

Es sind dadurch auch die Teile und Kanten eingezeichnet, die sonst verdeckt sind.

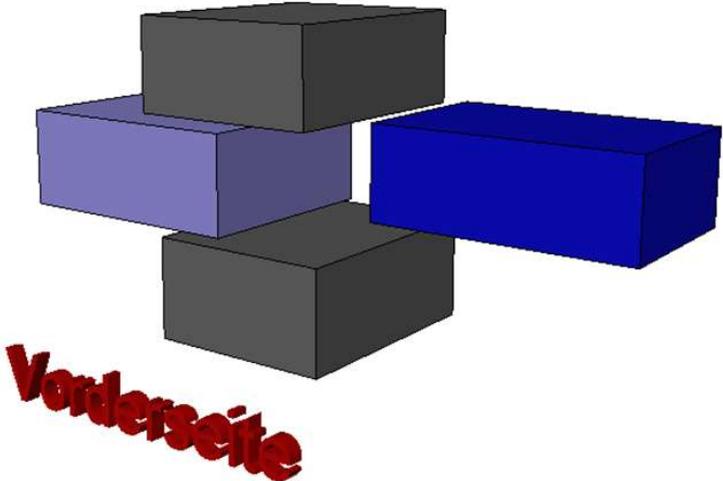
Hier ein Beispiel:

*Ansicht des Plans: linke Seite*



4

Probieren Sie das gleich einmal!

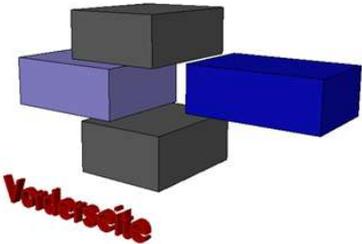


Sehen Sie sich die Figur oben gut an; in Folge werden Sie zu dieser Figur einige Pläne gezeigt bekommen.

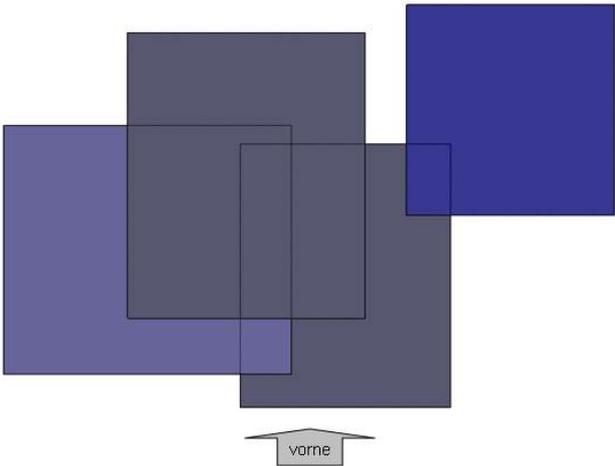


5

Übungsaufgabe 1 von 4



Ansicht des Plans: Top

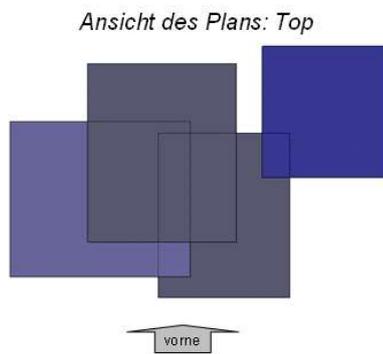
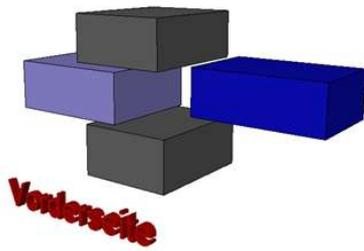


Der vorgegebene Plan ist...

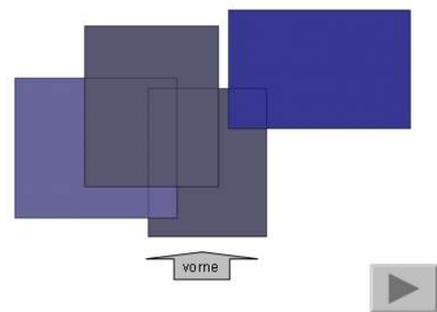
- richtig
- falsch



6



Der vorige Plan war falsch (hier: oben dargestellt). Einige Teile standen zueinander nicht in der richtigen Größe. Der richtige Plan von der angegebenen Seite würde so aussehen:



# TARV — TEST ZUR ANGEWANDTEN RAUMVORSTELLUNG (VERSION 4.0)

<b>Teil 1</b>						
Übungsaufgabe 1		<input type="checkbox"/> richtig <input type="checkbox"/> falsch	Aufgabe 6F	6Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
Übungsaufgabe 2		<input type="checkbox"/> richtig <input type="checkbox"/> falsch		6Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
Übungsaufgabe 3		<input type="checkbox"/> richtig <input type="checkbox"/> falsch		6Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
Übungsaufgabe 4		<input type="checkbox"/> richtig <input type="checkbox"/> falsch		Aufgabe 7F	7Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
Aufgabe 1F	1Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	7Fb		<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
	1Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	7Fc		<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
	1Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	Aufgabe 8F		8Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
Aufgabe 2F	2Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch			8Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
	2Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch			8Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
	2Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	Aufgabe 9F		9Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
Aufgabe 3F	3Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch			9Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
	3Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch			9Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
	3Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	Aufgabe 10F		10Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch
Aufgabe 4F	4Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch		10Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
	4Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch		10Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	
	4Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<p><i>Von TestleiterIn auszufüllen:</i></p> <p><i>Rohscore: _____</i></p>			
<b>Teil 2</b>						
Übungsaufgabe 1		<input type="checkbox"/> richtig <input type="checkbox"/> falsch				
Übungsaufgabe 2		<input type="checkbox"/> richtig <input type="checkbox"/> falsch				
Aufgabe 5F	5Fa	<input type="checkbox"/> richtig <input type="checkbox"/> falsch				
	5Fb	<input type="checkbox"/> richtig <input type="checkbox"/> falsch				
	5Fc	<input type="checkbox"/> richtig <input type="checkbox"/> falsch				

## Anhang C – Übersicht der Itemzuordnung im EST

	Stepan Nr.	Test Nr. Grubestic	qRM Nr.	Itemparameter (msig)	Testnr. Form A
Power (k=7)	Warmup 3	1		warmup	1
	90	2	29	0.6663	20
	122	3	22	-0.6283	14
	55	4	18	-1.6092	12
	61	5	28	0.3968	19
	150	6	24	-0.3613	15
	133	7	27	1.4734	18
	132	8	5	-0.0125	6
Worklimit (k=8)	111	9	11	1.2123	9
	8	10	26	0.6663	17
	2	11	20	-1.7376	13
	151	12	2	-0.0336	3
	96	13	16	-0.4494	11
	126	14	6	0.3968	5
	17	15	9	-0.8973	8
	7	16	1	0.0944	2



# Lebenslauf

<b>Name</b>	<b>Ana Grubestic</b>	
<b>Angaben zur Person</b>	Geboren: 17. November 1984 Geburtsort: Zenica (BiH) Staatsbürgerschaft: Österreich, Bosnien und Herzegowina	
<b>Ausbildung</b>	Seit 10/2004	<b>Diplomstudium Psychologie an der Universität Wien</b> 04/2007 1. Diplomprüfung
	09/1999-06/2004	<b>Bildungsanstalt für Kindergartenpädagogik Ettenreichgasse</b>
<b>Berufserfahrung</b>	Seit 09/2007	<b>Austro Control GmbH</b> <i>Sachbearbeitung/Assistenz</i>
	05-06/2010	<b>LOS- Lehrinstitut für Orthographie und Schreibtechnik</b> <i>Pädagogische Fachkraft (geringfügige Beschäftigung)</i>
	2003-2009	<b>SMZ- Ost Betriebskindergarten</b> <i>Kindergartenpädagogin (Sommer-Urlaubsvertretung)</i>
	10/2005- 06/2007	<b>Oasis- Fitness and Health</b> <i>Pädagogin in der Kinderbetreuung (geringfügige Beschäftigung)</i>
	08/2004-10/2005	<b>Firma Telemark Marketing Gebhard Zuber GmbH</b> <i>Marketing Assistentin (geringfügige Beschäftigung)</i>
<b>Publikationen</b>	Weitensfelder, L., Grubestic, A. & Kubinger, K.D. (2010). Ein facettenorientierter Raumvorstellungstest: Gendereffekte. 47. Kongress der Deutschen Gesellschaft für Psychologie, Bremen, 26.-30. September 2010.	
<b>Praktika</b>	02/2009-03/2009	<b>Kuratorium für Verkehrssicherheit</b> <i>6-Wochen-Pflichtpraktikum</i>
<b>Sonstige Aktivitäten</b>	SS 2008-WS 2008/09	<b>Teilnahme am zweisemestrigen Ausbildungsprogramm der Fakultät für Psychologie zum <i>student mentor</i> für StudienanfängerInnen</b>
	06/2009	<b>LPLUS – Training und Examination Management. Lernen und Prüfen am PC.</b>
<b>Sprachen</b>	Kroatisch/ Serbisch/ Bosnisch	Muttersprache
	Deutsch	Fließend in Wort und Schrift
	Englisch	Fließend in Wort und Schrift

