



universität
wien

Diplomarbeit

Titel der Arbeit

Geschlechterspezifische Unterschiede in Bezug auf
schlussfolgerndes Denken im Rahmen EMS-ähnlicher
Aufgaben unter Berücksichtigung von Stereotype
Threat und Risikoverhalten

Verfasserin

Silvia Rosa Hameseder

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im Juni 2012

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.-Prof. Mag. Dr. Ulrike Willinger

Meinen Eltern ...

Danksagung

An dieser Stelle möchte ich mich bei all jenen Menschen bedanken, die mich bei meiner Diplomarbeit unterstützt, begleitet und beraten haben.

Zuallererst gilt mein Dank meiner Betreuerin Frau Univ.-Prof. Mag. Dr. Ulrike Willinger für ihre wertvollen Anregungen und ihre Unterstützung bei der Erstellung meiner Diplomarbeit.

Ein weiterer Dank gilt auch meinen Kolleginnen Helene, Anita, Gabi und Katrin für die gute Zusammenarbeit, das angenehme Arbeitsklima und vor allem die gegenseitige Stütze.

Weiters sei auch allen Schulen gedankt, die mit ihrer Teilnahme diese Untersuchung unterstützt und überhaupt erst möglich gemacht haben.

Meinen Freundinnen Silke und Gabi danke für ihr offenes Ohr und ihre motivierenden Worte, wann immer ich sie brauchte.

Ein besonderer Dank gilt meinem Freund Walter. Ohne ihn wäre ich niemals so weit gekommen. Danke, dass du immer für mich da warst und mich auch in schwierigen Zeiten unterstützt hast.

Abschließend bedanke ich mich von ganzem Herzen bei meiner Familie, besonderes bei meinen Eltern für ihre bedingungslose und liebevolle Unterstützung jeglicher Art und Weise. Danke, dass ihr mich immer bestärkt und an mich geglaubt habt.

Anmerkung

Die empirische Untersuchung, die dieser Diplomarbeit zugrunde liegt, wurde in Zusammenarbeit mit vier weiteren Diplomandinnen, Katrin Anzirk, Gabriele Hangl, Helene Lager und Anita Teufl, durchgeführt. Jede Diplomandin behandelt in ihrer Arbeit unterschiedliche Themenschwerpunkte. Alle Arbeiten basieren auf einer gemeinsam erfassten Stichprobe und der zum Teil gleichen Grundlage der Literatur.

Inhaltsverzeichnis

Danksagung	I
Anmerkung.....	II
1 Einleitung	1
2 Theoretischer Teil	3
2.1 Eignungstest für das Medizinstudium (EMS)	3
2.1.1 Einsatz des EMS in Österreich	3
2.1.2 Entwicklung und Aufbau des EMS	4
2.1.3 Kritische Auseinandersetzung mit dem EMS	7
2.1.4 Studien zur Bearbeitung von Aufgabengruppen	11
2.1.4.1 Hinweise des Heftes „TestInfo“ zu den Aufgabengruppen	12
2.2 Schlussfolgerndes Denken	14
2.2.1 Induktives Denken und deduktives Denken	14
2.2.2 Problemlösen	16
2.2.2.1 Mathematische Kompetenz und naturwissenschaftliche Kompetenz... 18	
2.2.3 Der Faktor schlussfolgerndes Denken in verschiedenen Intelligenzmodellen	19
2.2.4 Geschlechterunterschiede im schlussfolgernden Denken	26
2.3 Stereotype Threat	33
2.3.1 Definition von Stereotyp	33
2.3.2 Definition und Ursprung von Stereotype Threat	34
2.3.2.1 Einfluss von Stereotype Threat auf kognitive Leistungen und Methoden um diesen zu reduzieren	36
2.3.2.2 Mögliche verantwortliche Prozesse für Stereotype Threat	40
2.4 Risikoverhalten und Entscheidungen	44
2.4.1 Begriffsbestimmung Risiko und Entscheidung	44
2.4.2 Entscheidungsmodelle	46
2.4.3 Unterschiede im Risikoverhalten	48
2.4.4 Entscheidungen unter Risiko im Game of Dice Task (GDT)	51

3 Zielsetzung, Fragestellung und Hypothesen	53
3.1 Zielsetzung und Fragestellung	53
3.2 Hypothesen	53
3.2.1 Haupthypothesen	53
3.2.2 Nebenhypothesen	53
4 Empirischer Teil	54
4.1 Methode	54
4.1.1 Untersuchungsplan und intendierte Stichprobe	54
4.1.1.1 Voruntersuchung	56
4.1.2 Erhebungsinstrumente	56
4.2 Untersuchung	70
4.2.1 Untersuchungsdurchführung	70
4.2.2 Auswertungsverfahren	71
4.2.3 Stichprobenbeschreibung	73
4.2.3.1 Stichprobenumfang und Geschlechterverteilung	74
4.2.3.2 Alter in der Gesamtstichprobe und in den Versuchsgruppen	75
4.2.3.3 Staatsbürgerschaft	76
4.2.3.4 Schulzweige	78
4.2.3.5 Wiederholen eines Schuljahres	79
4.2.3.5.1 Wiederholen eines Schuljahres nach Geschlecht	80
4.2.3.5.2 Wiederholen eines Schuljahres im naturwissenschaftlichen Schulzweig	81
4.2.3.6 Absicht ein Medizinstudium zu beginnen	82
4.2.3.6.1 Absicht ein Medizinstudium zu beginnen nach Geschlecht	83
4.2.3.6.2 Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS allgemein	83
4.2.3.7 Auseinandersetzung mit EMS allgemein nach Geschlecht	85
4.2.3.7.1 Auseinandersetzung mit EMS allgemein im naturwissenschaftlichen Schulzweig	86
4.2.3.8 Versuchsbedingungen	87
4.2.3.8.1 Stereotype Threat-Bedingungen	87
4.2.3.8.1.1 Stereotype Threat-Bedingungen nach Geschlecht	88
4.2.3.8.2 Risikobedingungen	89

4.2.3.8.2.1 Risikobedingungen nach Geschlecht	90
4.2.3.8.3 Stereotype Threat-Bedingungen und Risikobedingungen	91
4.3 Deskriptivstatistik und relative Lösungshäufigkeiten der EMS-ähnlichen Aufgaben	92
4.3.1 Relative Lösungswahrscheinlichkeiten bzw. Itemschwierigkeiten	93
4.3.1.1 Relative Lösungshäufigkeiten im Untertest Quantitative und formale Probleme	93
4.3.1.2 Relative Lösungshäufigkeiten im Untertest Medizinisch-naturwissen- schaftliches Grundverständnis	94
4.3.1.3 Relative Lösungshäufigkeiten im Untertest Diagramme und Tabellen	96
4.4 Ergebnisse	97
4.4.1 Überprüfung der experimentellen Stereotype Threat-Versuchsbedingungen mittels Wissenstest	97
4.4.2 Reliabilitätsanalysen	105
4.4.3 Interkorrelationen zwischen den Untertests	109
4.4.4 Uni- und multivariate Ergebnisse	110
4.4.4.1 Haupthypothesen	110
4.4.4.1.1 Hypothesenblock zum Untertest Quantitative und formale Probleme	110
4.4.4.1.2 Hypothesenprüfung zum Untertest Quantitative und formale Probleme	111
4.4.4.1.3 Hypothesenblock zum Untertest Medizinisch-naturwissen- schaftliches Grundverständnis	114
4.4.4.1.4 Hypothesenprüfung zum Untertest Medizinisch-naturwissen- schaftliches Grundverständnis	115
4.4.4.1.5 Hypothesenblock zum Untertest Diagramme und Tabellen	119
4.4.4.1.6 Hypothesenprüfung zum Untertest Diagramme und Tabellen	120
4.4.4.2 Nebenhypothesen zu den EMS-ähnlichen Aufgaben	125
4.4.4.2.1 Nebenhypothesen zum Untertest Quantitative und formale Probleme	126
4.4.4.2.2 Nebenhypothesenprüfung zum Untertest Quantitative und formale Probleme	126
4.4.4.2.3 Nebenhypothesen zum Untertest Medizinisch-naturwissen- schaftliches Grundverständnis	128

4.4.4.2.4 Nebenypothesenprüfung zum Untertest Medizinisch-naturwissenschaftliches Grundverständnis	128
4.4.4.2.5 Nebenypothesen zum Untertest Diagramme und Tabellen	129
4.4.4.2.6 Nebenypothesenprüfung zum Untertest Diagramme und Tabellen	130
4.4.4.2.7 Nebenypothese zu den EMS-ähnlichen Leistungen in Abhängigkeit des Schulzweiges sowie Nebenypothesenprüfung	131
4.4.4.3 Nebenypothesen zum GDT	134
4.4.4.3.1 Nebenypothesenprüfung zum GDT	134
5 Diskussion und Ausblick	139
6 Zusammenfassung	148
7 Abstract	151
8 Literaturverzeichnis	153
9 Abbildungsverzeichnis	165
10 Tabellenverzeichnis	166
11 Anhang	169
Anhang A: Tabellen	169
Anhang B: Lebenslauf	173

1 Einleitung

Unterschiede zwischen den Geschlechtern in kognitiven Aufgaben sind eine vieldiskutierte und aktuelle Thematik, mit der sich zahlreiche Untersuchungen auseinandersetzen (u. a. Hyde, Fennema & Lamon, 1990; Kuhn & Holling, 2009). Auch im *Eignungstest für das Medizinstudium* (EMS) sind geschlechterspezifische Unterschiede zu beobachten. Männer erzielen bessere Leistungen im EMS als Frauen (Mallinger et al., 2009). Leistungsunterschiede zwischen den Geschlechtern sind jedoch nicht auf mangelnde kognitive Fähigkeiten zurückzuführen und verschwinden im Laufe des Medizinstudiums (Mitterauer, Frischenschlager & Haidinger, 2007). Es gibt keine allgemeingültigen Antworten auf die komplexen Fragen zu derartigen kognitiven Geschlechterunterschieden (Halpern, 2000).

Die vorliegende Diplomarbeit soll dazu beitragen, geschlechterspezifische Leistungsunterschiede im EMS, insbesondere in jenen Aufgabenbereichen die das schlussfolgernde Denken betreffen, zu erklären. Dabei werden auch die sozialen Einflussfaktoren Stereotype Threat und Risikoverhalten berücksichtigt. In der Literatur stützen zahlreiche Belege das Phänomen Stereotype Threat. Demnach wird die kognitive Leistung von Frauen, aufgrund der Angst, ihre Leistung könnte mit negativen Stereotypen ihrer Gruppe übereinstimmen (z.B. „Frauen sind schlecht in Mathematik“), negativ beeinträchtigt (u. a. Davies, Spencer, Quinn & Gerhardstein, 2002; Keller & Dauenheimer, 2003; Spencer, Steele & Quinn, 1999). Auch im Risikoverhalten können Geschlechterunterschiede belegt werden. Männer weisen ein risikofreudigeres Verhalten auf als Frauen (Ben-Shakhar & Sinai, 1991).

Zu Beginn des theoretischen Teils dieser Arbeit wird der EMS näher vorgestellt. Im Anschluss daran wird die Fähigkeit des schlussfolgernden Denkens erläutert, zudem wird neben dem Faktor schlussfolgerndes Denken in verschiedenen Intelligenzmodellen, auch auf Forschungsarbeiten bezüglich Geschlechterunterschiede im schlussfolgernden Denken eingegangen. Des Weiteren wird an das Phänomen Stereotype Threat herangeführt und dessen

Einfluss auf geschlechterspezifische Leistungsunterschiede behandelt. Der letzte Abschnitt des theoretischen Teils erläutert den Einflussfaktor Risikoverhalten. Dabei wird auf Entscheidungsmodelle sowie auf Unterschiede im Risikoverhalten, insbesondere zwischen Frauen und Männern, eingegangen.

Im empirischen Teil werden die Erhebungsinstrumente, die Untersuchungsdurchführung sowie die Stichprobe beschrieben. Im Anschluss an die Darstellung der Hypothesenprüfungen zu den EMS-ähnlichen Aufgabenbereichen des schlussfolgernden Denkens (Untertest *Quantitative und formale Probleme*, Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* und Untertest *Diagramme und Tabellen*) werden die Ergebnisse diskutiert.

Während sich diese Arbeit explizit den geschlechterspezifischen Leistungsunterschieden in EMS-ähnlichen Aufgaben zum schlussfolgernden Denken widmet, werden geschlechterspezifische Unterschiede in der Merkfähigkeit im Rahmen EMS-ähnlicher Aufgaben in der Arbeit von Helene Lager bearbeitet. Ergebnisse zur Leistungsmotivation, Selbstwirksamkeitserwartung und Theory of Mind im Rahmen EMS-ähnlicher Aufgaben sind in den Diplomarbeiten von Gabriele Hangl, Anita Teufl und Katrin Anzirk nachzulesen.

2 Theoretischer Teil

Der nachfolgende theoretische Teil dieser Arbeit soll dem Leser einen Überblick über Konstrukte und wissenschaftlichen Begriffe geben, auf welchen in weiterer Folge im empirischen Teil eingegangen wird.

2.1 Eignungstest für das Medizinstudium (EMS)

Im folgenden Teil wird der *Eignungstest für das Medizinstudium* (EMS) vorgestellt. Es wird u. a. auf den Einsatz des EMS in Österreich sowie auf die Entwicklung und den Aufbau des EMS eingegangen.

2.1.1 Einsatz des EMS in Österreich

Mit dem Beschluss des Europäischen Gerichtshofs im Jahre 1995, musste Österreich den Bewerbern aus allen EU-Ländern die Chance auf einen Studienplatz, ermöglichen. Der damit verbundene hohe Anstieg der Bewerberanzahl im Medizinstudium in Österreich, macht es deshalb seit 2006 notwendig, dass das Auswahlverfahren *Eignungstest für das Medizinstudium* (EMS) an der Medizinischen Universität Wien und der Medizinischen Universität Innsbruck, zum Einsatz kommt. Auch in der Schweiz wird dieses Auswahlverfahren eingesetzt (Mallinger et al., 2008). Der EMS ist eine abgewandelte Form des *Test für Medizinische Studiengänge* (TMS), der in Deutschland von 1986 bis 1996 bei der Studienbewerberauswahl im Fach Medizin Einsatz fand, und seit 2007 in Teilen Deutschlands wieder dafür eingesetzt wird (Hänsgen & Spicher, 2007; Trost et al., 1998). Das Zentrum für Testentwicklung und Diagnostik (ZTD) in Freiburg, Schweiz, stellt den Test zur Verfügung und übernimmt die Auswertung für den EMS in Österreich (Mallinger et al., 2009).

Für die Studienrichtungen Humanmedizin und Zahnmedizin sind in Wien und Innsbruck insgesamt 1140 Studienplätze verfügbar (Mallinger et al., 2009). Gemäß § 124b Universitätsgesetz 2002, sind 75% der Studienplätze für

Österreicher, 20% für EU-Bürger und 5% für Nicht-EU-Bürger vorbehalten. Die Zulassung zum Medizinstudium erfolgt aufgrund des Testwertes, welcher zwischen 70 und 130 liegt (mit einem Mittelwert von 100) (Mallinger et al., 2009). Es wird ein Grenzwert festgelegt und Bewerber deren Testwert besser oder gleich dem Grenzwert ist, haben, sofern alle gesetzlichen Voraussetzungen erfüllt werden, Anspruch auf einen Studienplatz (Zentrum für Testentwicklung und Diagnostik [ZTD], 2009). Erreichen Personen den gleichen Testwert, so ist der mittlere Rangplatz der Testleistungen in den Untertests entscheidend (Mallinger et al., 2009). Aufgabe von Auswahlverfahren ist die Erhebung der Studieneignung und damit die Ermöglichung einer Prognose des künftigen Studienerfolgs (Troost et al., 1998). Die begrenzten Kapazitäten im Medizinstudium sind Ursache und nicht Folge des EMS-Einsatzes (Hänsgen, 2008).

2.1.2 Entwicklung und Aufbau des EMS

Zu Beginn der Entwicklung des TMS, auf welchem der EMS beruht, wurden in Deutschland aufwändige Anforderungsanalysen für ein Medizinstudium durchgeführt. Basierend auf diesen Analysen wurden in zahlreichen Voruntersuchungen verschiedene Aufgabengruppen erprobt, aus denen 13 als am besten geeignet für ein Übergangsverfahren für die Zulassung zu den medizinischen Studiengängen ausgewählt wurden (Troost et al., 1998). Aufgrund zahlreicher Untersuchungen während des Übergangsverfahrens, wurden vier der ursprünglich 13 Aufgabengruppen, welche sich als weniger prognosekräftig zeigten, eliminiert. Die Aufgabengruppe *Planen und Organisieren* wurde dann im EMS hinzugefügt (Hänsgen, 2008).

Seit 2010 wird zusätzlich online die Möglichkeit geboten, freiwillig und anonym an einem Self-Assessment teilzunehmen. Ziel dabei ist es, sich mit der eigenen Eignung für ein Medizinstudium auseinanderzusetzen und sich den eigenen Stärken und Risiken bewusst zu werden. Es werden Persönlichkeitsmerkmale wie Sozialverhalten, Arbeitsverhalten und Belastbarkeit rückgemeldet (Mallinger et al., 2010).

Der EMS setzt sich aus zehn Untertests zusammen. Tabelle 1 gibt einen Überblick über den Aufbau und die zeitliche Abfolge des EMS. Die Tabelle zeigt des Weiteren je Untertest die Anzahl der Aufgaben, die zu erreichende Punktezahl und die Bearbeitungszeit, sowie die im Untertest erfassten Fähigkeiten.

Tabelle 1: Aufbau, Ablauf und erfassende Fähigkeiten des EMS (modifiziert nach ZTD, 2009)

Untertests	Fähigkeiten	Aufgaben	max. Punktezahl	Bearbeitungszeit
Quantitative und formale Probleme	Umgang mit Zahlen, Größen, Einheiten und Formeln in medizinischen und naturwissenschaftlichen Fragestellungen	20	20	50 min
Schlauchfiguren	räumliches Vorstellungsvermögen	20	20	12 min
Textverständnis	Verarbeitung und Aufnahme von umfangreichen und komplexem Textmaterial	18	18	45 min
Planen und Organisieren	effiziente Selbstorganisation	20	20	60 min
Konzentriertes und sorgfältiges Arbeiten	rasches, sorgfältiges und konzentriertes Arbeiten	Blatt mit 1600 Zeichen	20	8 min
1 Stunde Pause				
<i>Einprägephase</i> Figuren lernen	Einzelheiten von Gegenständen einprägen und merken	Vorlagen zum Einprägen werden gezeigt		4 min
<i>Einprägephase</i> Fakten lernen	Fakten einprägen und merken			6 min
Medizinisch-naturwissenschaftliches Grundverständnis	Verständnis für Fragen der Medizin und Naturwissenschaften	20	20	50 min
<i>Abrufphase</i> Figuren reproduzieren	Abrufen eingepprägter Einzelheiten von Gegenständen	20	20	5 min
<i>Abrufphase</i> Fakten reproduzieren	Abrufen eingepprägter Fakten	20	20	7 min
Muster zuordnen	Musterausschnitte in einem komplexen Bild wiedererkennen	20	20	18 min
Diagramme und Tabellen	Diagramme und Tabellen richtig	20	20	50 min

	analysieren und interpretieren		
Gesamttest		198	198
Gesamtdauer (inkl. Pause) 9.00 bis ca. 17.00 Uhr			

Als „Speed- und Power-Test“ stellt der EMS unter Zeitdruck hohe Leistungsanforderungen (Kubinger, 2006).

Mit Ausnahme des Untertests *Konzentriertes und sorgfältiges Arbeiten*, sind die Aufgaben des EMS nach dem Multiple-Choice-Prinzip aufgebaut, dabei werden bei jedem Untertest fünf Lösungsvorschläge gegeben, wovon einer korrekt ist bzw. beim Untertest *Planen und Organisieren* werden vier Lösungsvorschläge vorgegeben (ZTD, 2009). Für jede richtige Antwort wird ein Punkt vergeben, in der Broschüre „TestInfo’09“ wird den Bewerbern empfohlen, bei jeder Aufgabe eine Antwort zu geben, da es bei falschen Antworten keinen Punkteabzug gibt, außer im Untertest *Konzentriertes und sorgfältiges Arbeiten* (ZTD, 2009).

Die EMS-Aufgaben sind Problemstellungen wie sie in einem Lehrbuch zum Grundstudium Medizin zu finden sein könnten, dabei muss das Problem erkannt und analysiert werden und eine Lösung gefunden werden. Alle Informationen die zum Lösen der Aufgaben nötig sind, sind in den Aufgabenstellungen enthalten, spezielles medizinisches Vorwissen ist hierfür nicht erforderlich, weshalb auch eine geringe Trainierbarkeit angegeben wird (Mallinger et al., 2009). Es wird jedoch empfohlen sich mit dem Originaltestmaterial vertraut zu machen und unter Ernstfallbedingungen zu üben (ZDT, 2009). Hierfür werden ca. 20 Stunden zur Vorbereitung/Übung befürwortet um bei der tatsächlichen Testbearbeitung keine Zeit zu verlieren. Zwischen den Begriffen Vorbereitung und Training ist hier zu unterscheiden, da mit dem Begriff Training häufig suggeriert wird, dass jede Leistung von jeder Person erreicht werden kann (Hänsgen & Spicher, 2009).

Mit Hilfe der Faktorenanalyse, welche das „klassische“ Verfahren einer psychometrischen Strukturanalyse darstellt, wurden die gemeinsamen Komponenten welche sich in den 10 Aufgabenbereichen finden, ermittelt. Etwa 30% der Varianz der Testergebnisse werden durch den Faktor

schlussfolgerndes Denken aufgeklärt, repräsentiert durch die Untertests *Medizinisch-naturwissenschaftliches Grundverständnis*, *Diagramme und Tabellen*, *Textverständnis*, *Quantitative und formale Probleme* sowie *Planen und Organisieren*. Etwa 14% der Varianz der Testergebnisse werden durch den Faktor Merkfähigkeit aufgeklärt, repräsentiert durch die Untertests *Fakten Lernen* und *Figuren Lernen*, und ca. 18% der Varianz der Testergebnisse werden durch den Faktor visuelle Fähigkeiten aufgeklärt, repräsentiert durch die Bereiche *Schlauchfiguren*, *Konzentriertes und sorgfältiges Arbeiten* und *Muster zuordnen* (Mallinger et al., 2008).

Die Leistungsunterschiede der Geschlechter in Österreich, sind seit dem Einsatz des EMS zu beobachten und werden nach wie vor diskutiert. Unterschiede zugunsten der Männer zeigten sich in jedem Jahr, am größten waren diese Differenzen jedoch 2007 (Mallinger et al., 2010).

Die Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* sind die Untertests zum schlussfolgernden Denken, auf welche in dieser Arbeit eingegangen wird.

Die Aufgaben im Untertest *Quantitative und formale Probleme* zeigen Ähnlichkeit zum Mathematikunterricht. Die Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* erheben naturwissenschaftlich-medizinische Fähigkeiten (Spiel, Schober & Litzenberger, 2008). Mathematik und Naturwissenschaften geben insbesondere die Inhalte des ersten Studienabschnittes im Medizinstudium an und sind deshalb wichtige EMS-Inhalte (Mallinger et al., 2010).

2.1.3 Kritische Auseinandersetzung mit dem EMS

Der Testwert des EMS, korreliert nachweislich hoch mit der Studieneignung (Mallinger et al., 2009). Nach Hell, Trapmann, Weigand, Hirn und Schuler (2005, zitiert nach Spiel et al., 2008, S. 22) zeigen allgemeine und fachspezifische Studierfähigkeitstests, neben Schulnoten, welche aber

aufgrund fehlender Objektivität problematisch sind, die höchste prognostische Validität. Der EMS hat als Studierfähigkeitstest, verglichen mit Verfahren wie Interviews oder schulischen Fachnoten, eine hohe Durchführungs- und Auswertungsobjektivität (Heine, Briedis, Didi, Haase & Trost, 2006).

Spiel et al. (2008) führten eine Evaluationsstudie zur Analyse der Chancengleichheit des EMS bezüglich Geschlecht, sozialer Herkunft, Schultyp und Nationalität, mittels fünfminütigem Fragebogen durch. Dieser umfasste 36 Fragen und wurde nach der EMS-Testung 2007 den Bewerbern vorgegeben. Wie Spiel et al. (2008) anführten, wurden schlechte Leistungen der Frauen im EMS oft als nicht überraschend argumentiert, was mit den schlechteren Prüfungsleistungen und höheren Drop-Out-Quoten, die Frauen auch vor der EMS-Einführung aufwiesen, begründet wurde. Mitterauer et al. (2007) zeigten in ihrer Studie zum Studienverlauf jedoch, dass Geschlecht kein valider Prädiktor für den Studienerfolg ist. Zwar war der Erfolg beim Bestehen der ersten summativen integrativen Prüfung (SIP1) zu Studienanfang, bei den weiblichen Studierenden geringer, sie holten diese Unterschiede jedoch mit einem Jahr Verzögerung auf. Zu Studienbeginn zeigten sich aber tatsächlich die Geschlechterunterschiede im Studienerfolg, womit nach Hänsgen (2008) die Studienerfolgsprognose mittels SIP1 wiederum insgesamt richtig sei, denn der im EMS festgestellte Geschlechterunterschied entspricht genau dem in der SIP1 festgestellten Unterschied. Mitterauer et al. (2007) sowie Haidinger, Mitterauer, Rimroth und Frischenschlager (2008) schlugen vor, statt der umfangreichen SIP1 mehrere Teilprüfungen vorzugeben, um die Benachteiligung der Frauen zu beheben. Sie begründeten dies mit der Schlussfolgerung, dass der Zeitverlust der weiblichen Medizinstudenten, aufgrund des Zusammenwirkens der SIP-spezifischen Erfordernisse und dem geschlechterspezifischen Lernverhalten erklärt werden können und nicht mangels kognitiver Fähigkeiten. Bereits 2005 zeigten Frischenschlager, Mitterauer und Haidinger in ihrer Untersuchung, dass die von ihnen erhobenen Leistungsfaktoren Schulnoten, fachspezifische Vorkenntnisse und allgemeine und fachspezifische Lernleistung nicht für ein Auswahlverfahren geeignet sind.

In den letzten Jahren hat sich der relative Männeranteil der Bewerbungen innerhalb der Österreich-Quote zwar erhöht, die Bewerberanzahl der Frauen ist jedoch höher, die Zulassungsanzahl ist hingegen bei den Männern höher (Mallinger et al., 2010). Bedeutsam für die Geschlechterverteilung im Medizinstudium ist die Testleistung, nicht die Quote. Der Mittelwertsunterschied zuungunsten der Frauen betrug 2009, wie auch 2010 etwa vier Testwertpunkte (Mallinger et al., 2010). Aufgrund der Quotenregelung werden deutlich weniger deutsche Bewerber aufgenommen, als dies der Fall wäre wenn nur die Testleistung Zulassungskriterium wäre (Spiel et al., 2008). Bei den Nicht-EU-Bürgern werden aufgrund der geringen Bewerberanzahl auch Personen zugelassen, die schlechte Testleistungen erzielen, ihnen kommt die Quotenregelung zugute (Spiel et al., 2008). Spiel et al. (2008) empfahlen um dem gegen zu wirken die Einführung eines Mindesttestwerts. Mallinger et al. (2010) ließen die Frage offen, ob die Chance der Nicht-EU-Bürger auf einen Studienplatz mit einer 5% Quote, im Vergleich zur Österreich-Quote und EU-Bürger-Quote, hier zu hoch sei.

Deutsche BewerberInnen erbringen in den Untertests *Textverständnis*, *Medizinisch-naturwissenschaftliches Grundverständnis* sowie *Diagramme und Tabellen* bessere Leistungen als österreichische BewerberInnen. Auch sind deutsche BewerberInnen schon mehrfach zu Zulassungstests für das Medizinstudium angetreten (Spiel et al., 2008).

Nach Spiel et al. (2008) zeigten sich bei österreichischen BewerberInnen mit und ohne naturwissenschaftlichen Schwerpunkt in der Schule, keine Leistungsunterschiede, wohingegen deutsche BewerberInnen aus Schulen mit naturwissenschaftlichem Schwerpunkt bessere Leistungen erbrachten.

Spiel et al. (2008) stellten fest, dass die EMS-Aufgaben das Kriterium der Verrechnungsfairness (Skalierung gemäß des Rasch-Modells) nicht erfüllt, da nicht nur die gemessene Kompetenz bzw. das untersuchte Wissen, sondern auch andere Faktoren für die Lösung eine Rolle spielen. Des Weiteren sind nach Spiel et al. (2008) die Trennschärfen vieler Aufgaben zu gering um zwischen „guten“ und „schlechten“ Testpersonen zu differenzieren. Spiel et al.

(2008) führte für die Verrechnungsfairness die Definition an, dass Testaufgaben ausschließlich die interessierende Kompetenz (bzw. den interessierenden Wissensbereich) erfassen sollen. Hierzu betonte Hänsgen (2008) den Bezug der Definition auf die probabilistische Testtheorie (PTT) bzw. Rasch-Skalierung. Wie viele Studierfähigkeitstests wurde der EMS nach der klassischen Testtheorie konstruiert. Bei der Konstruktion des TMS entschied sich der damalige „Beirat für psychologische und allgemeine Fragen der Testentwicklung und des Testeinsatzes bei der Hochschulzulassung“ bewusst für die klassische Testtheorie. Um der Realität eines Studiums, in welchem unterschiedliche kognitive Fähigkeiten gleichzeitig notwendig sind, nahe zu kommen, sollte der TMS möglichst komplexe Aufgaben beinhalten. Auch beim EMS würden höhere Trennschärfen, wie sie von Spiel et al. (2008) gefordert wurden, die erwünschte Aufgabenvielfalt verringern, so Hänsgen (2008).

Spiel et al. (2008) wiesen überdies in ihrer Untersuchung auf geringe Reliabilitäten der Untertests hin, Hänsgen (2008) betonte hierzu die ausreichende Zuverlässigkeit des EMS-Testwerts und entgegnete, dass die etwas niedrigeren Reliabilitäten in den einzelnen Untertests Preis für die größere Vielfalt und größere Heterogenität in den Aufgabengruppen seien.

Um die Ratewahrscheinlichkeit zu beseitigen empfahlen Spiel et al. (2008), dass beim Multiple-Choice-Antwortformat nicht nur eine von fünf Alternativen richtig sein sollte, sondern unterschiedlich eine bis vier Antwortmöglichkeiten richtig sein könnten, wobei nur das Ankreuzen aller zutreffenden Alternativen als richtig bewertet werden soll. Nach Hänsgen (2008) gleiche die mehrfach gegebene Instruktion, alle Fragen zu beantworten - also am Ende immer bei jenen zu raten, die unbeantwortet sind - den Einfluss der Ratewahrscheinlichkeit aus. Die Bedingungen seien somit, laut Hänsgen (2008), für alle BewerberInnen gleich und echte Leistung werde als Differenz zur Ratewahrscheinlichkeit verstanden. Ein Antwortformat mit mehreren richtigen Lösungen je Aufgabe sei nach Hänsgen (2008) denkbar. Eine nachträgliche Änderung der erprobten Aufgaben sei jedoch nicht möglich, weil sich damit die Schwierigkeit drastisch ändern würde.

2.1.4 Studien zur Bearbeitung von Aufgabengruppen

Der Bereich *Quantitative und formale Probleme* setzt sich aus „mathematischen Sachaufgaben“ zusammen, für deren Lösung neben schlussfolgerndem Denken, formallogische, kombinatorische, arithmetische oder algebraische Operationen notwendig sind. Studien zur Bearbeitung von Aufgaben des Bereiches *Quantitative und formale Probleme* zeigten, dass diese Aufgabengruppe kein Konstrukt enthielt, für welches ein „standardisiertes allgemeingültiges Prozessmodell“ herstellbar war (Klieme, 1989, zitiert nach Trost et al., 1998, S. 51 f.). Nach Klieme (1989, zitiert nach Trost et al., 1998, S. 52) besaßen leistungsstarke Bearbeiter in dieser Aufgabengruppe die komplexe Fähigkeit, aus dem Aufgabentexten die zentralen naturwissenschaftlichen Zusammenhänge zu filtern und in adäquate Situationsmodelle zu transformieren sowie mathematische Strukturen richtig zu erfassen und abstrakt zu repräsentieren. Hingegen leistungsschwächere Personen erfassten hier mathematische Strukturen nicht oder nur fehlerhaft und bildeten die Problemstellungen in konkreten Situationsmodellen ab. Die Analyse der Prozessverläufe zeigte, dass bei leistungsstarken Personen die Kontroll- und Steuerungsaktivitäten weitgehend automatisiert verliefen und sie wenige metakognitive Äußerungen verbalisierten. Leistungsschwächere verbalisierten häufig Zwischenergebnisse, Kontrollabsichten und Ziele, was daraus schließen lässt, dass die betreffenden Denkprozesse nicht automatisiert waren.

Der Bereich *Medizinisch-naturwissenschaftliches Grundverständnis* erfasst die Fähigkeit, aus Geschriebenem rasch das Wesentliche zu extrahieren und richtige Schlüsse daraus zu ziehen. Die Fragestellungen können allein durch schlussfolgerndes Denken im Rahmen von qualitativen mentalen Modellen beantwortet werden. Es werden zu einem medizinischen oder naturwissenschaftlichen Sachverhalt Behauptungen aufgestellt, welche zumeist die Form einer Konditionalaussage haben, diese setzt sich zusammen aus Bedingungsteil (wenn...) und Konsequenzteil (dann...). Die aufgestellten Behauptungen müssen auf ihre Gültigkeit überprüft werden (Trost et al., 1998).

Trost et al. (1998) beschäftigten sich in ihrer Studie zur Bearbeitung von Aufgaben des Bereiches *Medizinisch-naturwissenschaftliches Grundverständnis* mit zwei Arten von Aufgaben, nämlich Aufgaben die „Wirkungszusammenhänge“ beschrieben und Aufgaben die „räumliche Konstellationen“ beschrieben. Bei ersteren werden die Textaussagen als Abfolge von „Fact-Nodes“ repräsentiert, die durch Kausalrelationen miteinander verknüpft sind, diese Aufgaben benötigen geringen kognitiven Aufwand. Bei Aufgaben die „räumliche Konstellationen“ beschreiben, werden die Textaussagen in ein Vorstellungsbild transformiert, diese Aufgaben erfordern mehr kognitiven Aufwand. Nach Trost et al. (1998) differenzierten Leistungsschwache hier nicht oder unzulänglich zwischen den beiden Aufgabentypen, wohingegen Leistungsstarke automatisch unterschieden. Leistungsschwache Personen hatten Schwierigkeiten beim Verständnis der Konditionalaussagen und bei der Unterscheidung zwischen Bedingungs- und Konsequenzteil. Überdies bereitete es leistungsschwachen Personen Schwierigkeiten den Text präsent zu halten, weshalb sie doppelt so lange für die Behauptungsüberprüfungen brauchten, als leistungsstarke Personen.

Derartige prozessorientierte Studien waren Basis für die Gestaltung des „TestInfos“, welches Testteilnehmern zur Vorbereitung auf den Test überlassen wird und sie mit geeigneten Bearbeitungstechniken vertraut machen soll.

2.1.4.1 Hinweise des Heftes „TestInfo“ zu den Aufgabengruppen

In der Broschüre „TestInfo'09“ (ZDT, 2009) wird beim Untertest *Quantitative und formale Probleme* auf die wichtigsten vier Grundmuster hingewiesen: Aufteilen und Untergliedern einer bestimmte Grundmenge in verschiedene Teile; Proportionale Beziehungen; Formeln aufstellen, umformen und interpretieren; Rechnen mit Einheiten und Dimensionen.

Naturwissenschaftlicher Sachverhalt soll verstanden und der richtige Lösungsansatz gefunden werden. Dies wird durch das Vertrautmachen mit den Grundmustern erleichtert (ZDT, 2009).

Beim Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* wird im Heft „TestInfo’09“ (ZDT, 2009) darauf hingewiesen, dass um den Sachverhalt zu strukturieren, Wesentliches bereits beim Lesen unterstrichen und markiert werden soll. Bei komplexeren Aufgaben sollen Skizzen angefertigt werden um zentrale Elemente in ihrer Beziehung zueinander schematisch darzustellen. Das Nachvollziehen bzw. Schlussfolgern aus den im Text enthaltenen Informationen, ermöglicht die Beurteilung der einzelnen Aussagen, spezielle Sachkenntnisse sind dafür nicht erforderlich.

Beim Untertest *Tabellen und Diagramme* wird im Heft „TestInfo’09“ (ZDT, 2009) darauf hingewiesen, dass keine speziellen naturwissenschaftlichen, medizinischen oder statistischen Kenntnisse zum Lösen der Aufgaben benötigt werden. Allein durch die graphisch oder tabellarisch dargebotenen Informationen und dem jeweiligen Aufgabentext können die richtigen Antworten abgeleitet werden. In Tabellen sollen Werte veränderlicher Größen verglichen oder in Relation zu anderen Werten gesetzt werden. Es sollen Maxima, Minima, und Schwankungen ermittelt werden und Gesetzmäßigkeiten bzw. verallgemeinernde Aussagen analysiert werden. In Säulendiagrammen oder Histogrammen werden Häufigkeits- und Mengenangaben graphisch dargestellt, auch hier ist es notwendig Vergleiche anzustellen und Gesetzmäßigkeiten zu analysieren. Es werden Säulendiagramme mit absoluten sowie mit relativen Einheiten vorgegeben, die Skalenbeschriftung ist zu beachten. Auch bei Kurvendiagrammen sind Aussagen über Maxima und Minima zu beurteilen, des Weiteren kommt ihnen eine große Bedeutung bei der Analyse von Gesetzmäßigkeiten zu. Es können gesetzmäßige Zusammenhänge wie eine positiv monotone Beziehung, ein linearer Zusammenhang, ein proportionaler Zusammenhang, eine negativ monotone Beziehungen oder ein negativ linearer Zusammenhang, anhand des Kurvenverlaufes abgelesen werden, so die Angaben und Hinweise der Broschüre „TestInfo’09“ (ZDT, 2009).

2.2 Schlussfolgerndes Denken

Im folgenden Kapitel soll an die Fähigkeit des schlussfolgernden Denkens näher herangeführt werden. Neben der begrifflichen Bestimmung wird der Faktor schlussfolgerndes Denken in verschiedenen Intelligenzmodellen behandelt. Weiters werden Befunde der Intelligenzforschung in Bezug auf Geschlechterunterschiede im schlussfolgernden Denken angeführt.

Schlussfolgerndes Denken bedeutet aus gegebenen Informationen neues Wissen abzuleiten. Es basiert auf logischen Überlegungen, da logisch korrekte Ableitungen und Generalisierungen gegebener Informationen erforderlich sind (Lohaus, Vierhaus & Maass, 2010).

Aufgaben wie die Bearbeitung mathematischer Probleme und Sachaufgaben, das Erkennen von Gesetzmäßigkeiten in Zahlenfolgen oder Figurenfolgen und das analoge Schließen gelten als Indikatoren für den Kernbereich der Intelligenz, der schlussfolgerndes Denken, Reasoning oder Verarbeitungskapazität genannt wird (Klieme, Funke, Leutner, Reimann & Wirth, 2001).

2.2.1 Induktives Denken und deduktives Denken

Das systematische Überprüfen von Hypothesen erfordert schlussfolgernde Denkleistungen, dabei handelt es sich im Wesentlichen um zwei Formen des Schlussfolgerns, um das induktive und das deduktive Denken/Schlussfolgern (Lohaus et al., 2010; Maderthaner, 2008).

Induktives Denken ermöglicht es, aus spezifischen Erfahrungen allgemeine Regeln, Muster, Konzepte oder Gesetzmäßigkeiten abzuleiten und auf neue Ereignisse anzuwenden (Waldmann & Weinert, 1990). In der Literatur gibt es unzählige Einzeltests bzw. Subtests, die es sich zum Ziel gesetzt haben, schlussfolgerndes Denken zu messen. Dem induktiven Denken werden vor allem die Aufgabentypen Analogien (z.B. mit Wörtern, Zahlen oder

geometrischen Figuren), Reihen ergänzen (z.B. mit Buchstaben oder Zahlen), Matrizenaufgaben (z.B. vom Raven-Test-Typus) und Klassifikationsaufgaben zugerechnet (Beckmann & Guthke, 1999).

Deduktives Denken ist ein systematischer Denkprozess, der von einer bestimmten Menge an Prämissen unter Verwendung logischer Regeln zu einer Menge von Schlussfolgerungen führt. Allein die Wahrheit der Prämissen garantiert die Wahrheit der Schlussfolgerungen, unabhängig davon welche konkreten Inhalte die Prämissen haben (Waldmann & Weinert, 1990). Das deduktive Denken betreffend, beziehen sich kognitionspsychologische Untersuchungen insbesondere auf konditionale und kategoriale Schlüsse (Maderthaner, 2008). Konditionalschlüsse bestehen aus dem Wenn-Teil (Antecedens) und dem Dann-Teil (Konsequens) (Anderson, 2007). Eine wichtige logische Regel (Ableitungsoperation) ist dabei der Modus ponens. Er erlaubt die Ableitung eines gültigen Konsequens aus einem gegebenen Antecedens, so zum Beispiel, dass man aus der Proposition „Wenn A, dann B“ bei der gegebenen Proposition A, auf B schließen kann. Eine weitere Ableitungsoperation ist der Modus tollens, bei dem von einem nicht gegebenen Konsequens auf das Nichtvorliegen des Antecedens geschlossen wird (Anderson, 2007; Maderthaner, 2008). Auch Syllogismen sind dem deduktiven Denken zuzuordnen, sie enthalten in den Prämissen Quantoren wie alle, einige, keine, usw. und basieren auf Operationen der Quantorenlogik (Maderthaner, 2008; Waldmann & Weinert, 1990). Der überwiegende Anteil der psychologischen Quantorenforschung konzentriert sich den kategorialen Syllogismus, welcher aus zwei Prämissen und einer Schlussfolgerung besteht (Anderson, 2007; Waldmann & Weinert, 1990).

Ursachen für Fehler beim deduktiven Schlussfolgern, wie z.B. Fehler bei der Beurteilung von Syllogismen, lassen sich unter anderem auf das Vergessen von Modellen, gemäß der Theorie mentaler Modelle von Johnson-Laird, zurückführen (Maderthaner, 2008). In der Theorie mentaler Modelle (Johnson-Laird, 1999) wird davon ausgegangen, dass Prämissen (Aussagen) zu Einzelvorstellungen (Modellen) umgewandelt werden und gemäß ihrer Bedeutungen, im Arbeitsgedächtnis zu einem oder mehreren komplexen

mentalen Modellen zusammengefügt werden. Aus diesen können Schlüsse abgeleitet werden, die innerhalb der einzelnen Prämissen nicht enthalten sind. Insbesondere bei mehreren oder komplexen Prämissen kann es zum Vergessen von Annahmen und somit zu Fehlern im deduktiven Schlussfolgern kommen (Maderthaner, 2008).

2.2.2 Problemlösen

Es liegt nahe, eine enge Beziehung zwischen den Konstrukten Intelligenz und Problemlösekompetenz zu sehen (Leutner, Funke, Klieme, & Wirth, 2005). Sie hängt sicherlich mit Maßen der Intelligenz im Sinne des Reasoning zusammen und ist von dieser kognitiven Grundfähigkeit nicht abgrenzbar (Klieme et al., 2001). Auch Problemlöseaufgaben sind typische für Tests die das schlussfolgernde Denken erheben (Anderson, 2007). Bei den involvierten Denkprozessen des Problemlösens handelt es sich im Wesentlichen um schlussfolgerndes Denken (Maderthaner, 2008).

Unter einem Problem versteht man eine Situation, für deren gewünschte Überführung von der Ist- in eine Soll-Situation, ein Aufwand zu erbringen ist. Ein Problem in der Mathematik ist beispielsweise eine lineare Gleichung, bei der die Zustandsvariablen so verändert werden müssen, dass die Gleichsetzung zwischen beiden Seiten der Formeln korrekt ist. Eine Krankheit ist beispielsweise ein Problem in der Medizin, dass durch gezielten Einfluss auf den Zustand des Kranken (mittels Medikamenten, Anweisungen, Operationen etc.) und durch das Herbeiführen eines weniger belastenden Zustandes gelöst werden soll. Gut definierte Probleme, d. h. wenn der Anfangszustand, der Endzustand und die zur Verfügung stehende Maßnahmen (Operationen, Prozeduren etc.) bekannt sind, sind leichter lösbar als schlecht definierte Probleme (Maderthaner, 2008).

Als Problemlösen wird in der Forschung häufig das Absuchen eines Problemraumes beschrieben, dieser besteht aus unterschiedlichen Problemzuständen. Die möglichen Zustände eines Problems vom Ausgangszustand bis zum Zielzustand, können mittels erlaubter Operatoren,

welche man sich als Wege von einem Zustand zum anderen vorstellen kann, erreicht werden (Anderson, 2007).

Die Bezeichnung „Komplexes Problemlösen“ bezieht sich eigentlich auf die Komplexität der Anforderung, die verglichen mit den Anforderungen beim „einfachen“ Problemlösen erheblich größer ausfällt (Funke, 2003). Beim Lösen komplexer Probleme sind nach Dörner (1976) folgende Aspekte zu berücksichtigen, die Komplexität der Problemsituation, die Vernetztheit der beteiligten Variablen, die Dynamik inklusive zeitlicher Entwicklung, die Intransparenz der nicht zugänglichen Teile des Systems, die Polytelie im Sinne mehrerer konkurrierender Ziele sowie die Wirkungssicherheit, die Wirkungsbreite, die Reversibilität und die Anwendungsvoraussetzungen von Problemlösemaßnahmen.

Die Übergänge vom Ausgangszustand zum Zielzustand einer Problemsituation können mittels mentalen Modells geistig simuliert werden (Maderthaler, 2008).

Ist für ein Problem kein Algorithmus d. h. keine eindeutige Lösungsstrategie verfügbar, so können Heuristiken eine schrittweise Annäherung an Zielzustände ermöglichen (Maderthaler, 2008).

Ein heuristisches Verfahren zur Optimierung von Zuständen welches auch in der Mathematik eingesetzt wird ist die Unterschiedsreduktion. Dabei werden Operatoren ausgewählt welche den Unterschied zwischen dem aktuellen Zustand und dem Zielzustand am stärksten reduzieren. Man bezeichnet dieses Verfahren auch als Bergsteigen (hill climbing), welches den Nachteil birgt, dass nur betrachtet wird, ob der direkt folgende Schritt eine Verbesserung darstellt, nicht aber ob der Zielzustand erreicht wird (Anderson, 2007; Maderthaler, 2008).

Die Mittel-Ziel-Analyse bringt eine globalere Sichtweise in das Problemlösen, sie wurde von Newell und Simon mit Hilfe des Computerprogramms General Problem Solver (GPS) untersucht. Bei der Mittel-Ziel-Analyse wird um den Unterschied, der zwischen dem aktuellem Zustand und der Anwendungsvoraussetzung eines Operators besteht, zu eliminieren, der Zielzustand in Teilziele zerlegt. Sie ist eine wichtige Methode des

Problemlösens (Anderson, 2007). Nach Ernst und Newell (1969, zitiert nach Anderson, 2007, S. 310) erörtert die Mittel-Ziel-Analyse die Lösung für verschiedene Probleme, darunter für Probleme aus der Algebra und der Differentialrechnung sowie für Logikprobleme.

2.2.2.1 Mathematische Kompetenz und naturwissenschaftliche Kompetenz

Problemlösefähigkeit ist eine fächerübergreifende Kompetenz und ist nicht als eindimensionales Konstrukt zu sehen (Klieme et al., 2001). Das „Programme for International Student Assessment“ (PISA) definiert die Problemlösekompetenz als die Fähigkeit einer Person, kognitive Prozesse zu nutzen, um sich mit realen, fächerübergreifenden Problemstellungen auseinanderzusetzen und sie zu lösen. Bei diesen Problemstellungen ist der Lösungsweg nicht sofort ersichtlich und die zur Lösung notwendigen Wissensbereiche entstammen nicht einem einzelnen Fachbereich der Mathematik, der Naturwissenschaften oder des Lesens (Organisation for Economic Co-operation and Development [OECD], 2003).

Als mathematische Kompetenzen werden von Leiß und Blum (2006) das mathematische Argumentieren, das mathematische Modellieren, die Verwendung von mathematischen Darstellungen, der Umgang mit symbolischen, formalen und technischen Elementen der Mathematik, das mathematische Kommunizieren und das mathematisches Problemlösen angeführt. Beim mathematischen Problemlösen können auch heuristische Prozesse oder Hilfsmittel angewendet werden, welche sich bei Problemlöseprozessen allgemein als zielführend erweisen. Typische Lösungsstrategien dafür sind das Zerlegungsprinzip („In welche Teilprobleme kann man ein Problem zerlegen?“) und das Analogieprinzip („Wurden bereits ähnliche Probleme gelöst?“). Außerdem das Vorwärtsarbeiten („Was lässt sich aus den gegebenen Daten folgern?“), das Rückwärtsarbeiten („Was wird benötigt, um das Gesuchte zu erhalten?“), das systematische Probieren und die Veranschaulichung durch eine mathematische Figur, Tabelle oder Skizze.

Naturwissenschaften befassen sich mit der systematischen Erforschung der Natur und ihren gesetzmäßigen Zusammenhängen. Sie setzen sich aus verschiedenen Disziplinen zusammen, den exakten Naturwissenschaften (Physik, Chemie, Astronomie) und den biologischen Naturwissenschaften (Biologie, Medizin u. ä.) (Beerman, Heller & Menacher, 1992). Die naturwissenschaftliche Kompetenz beschreibt PISA als die Fähigkeit naturwissenschaftliche Fragen zu erkennen und naturwissenschaftliches Wissen anzuwenden. Außerdem sollen aus Belegen Schlussfolgerungen gezogen werden, um Entscheidungen zu verstehen und zu treffen, die die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen (OECD, 2003).

2.2.3 Der Faktor schlussfolgerndes Denken in verschiedenen Intelligenzmodellen

Es gibt weder eine einheitliche Definition des Begriffes Intelligenz noch eine einheitliche Forschungsperspektive. Möchte man einen kleinsten gemeinsamen Nenner finden, so versteht man unter Intelligenz die Fähigkeit, aus Erfahrungen zu lernen und sich an die Erfordernisse der Umgebung anzupassen. Der Frage, wie Intelligenz gemessen werden kann, haben sich verschiedene Forscher verschrieben (Funke & Vaterrodt-Plünnecke, 2004).

Die Fähigkeit des schlussfolgernden Denkens wird als zentraler Aspekt der Intelligenz verstanden. In diesem Kapitel soll die Rolle des schlussfolgernden Denkens in den unterschiedlichen Intelligenzmodellen hervorgehoben werden.

Binet, welchem die Urheberschaft des allerersten Intelligenztests, welcher 1905 erschien, zugesprochen wird, nahm an, dass es sich bei Intelligenz um eine ganzheitliche homogene Fähigkeit handle. Er legte Aufgaben vor welche eher logisches Denken als Auswendiglernen erforderten (Funke & Vaterrodt-Plünnecke, 2004). Weitere Forschungen haben verdeutlicht, dass Intelligenz nicht eine einzige Dimension ist, sondern die individuellen Unterschiede der

Kognition komplexer sind (Anderson, 2007). Im Laufe der Zeit wurden unterschiedliche Modelle der Intelligenz entwickelt.

Spearman setzte wie viele Intelligenzforscher nach ihm, die Faktorenanalyse als methodisches Instrument ein, um festzulegen welche Aufgabenreihen miteinander korrelieren (Funke & Vaterrodt-Plünnecke, 2004). Bei seiner Zwei-Faktoren-Theorie der Intelligenz stellte Spearman (1927) die Annahme auf, dass Intelligenzleistungen auf einem allgemeinen Faktor (g-Faktor) und einem anderen, für jeden einzelnen Test spezifischen Faktor (s) beruhen. Aufgrund hoher Korrelationen von $r = .68$ bis $.84$ zwischen dem induktiven Denken und dem g-Faktor belegte er die große Bedeutung dieser Fähigkeit für die Intelligenz.

Thurstone stellte der Zwei-Faktoren-Theorie ein Primärfaktorenmodell gegenüber (Funke & Vaterrodt-Plünnecke, 2004). Er nahm anstelle eines einzigen g-Faktors mehrere voneinander unabhängige Faktoren an, die die Testleistung bestimmen.

Mittels multipler Faktorenanalyse postulierte Thurstone sieben Primärfähigkeiten, welche einzeln eine höchst unterschiedliche Breite aufweisen (Amelang, Bartussek, Stemmler & Hagemann, 2006).

Die sieben Primärfähigkeiten nach Thurstone (Jäger, 1967) sind:

1. verbales Verständnis (Verbal Comprehension – V)
2. Wortflüssigkeit (Word Fluency – W)
3. schlussfolgerndes Denken (Reasoning – R)
4. räumliches Vorstellungsvermögen (Space – S)
5. Merkfähigkeit (Memory – M)
6. Rechenfähigkeit (Number – N)
7. Wahrnehmungsgeschwindigkeit (Perceptual Speed – P)

Thurstone definierte zunächst den Faktor Induktion, der durch das Erkennen einer Regel im Testmaterial geprüft wird (Jäger, 1967). Ist die Regel auch anzuwenden und nicht nur zu erkennen, so könnte man nach Jäger (1967)

auch von Deduktion sprechen. Solche Überlegungen haben wohl auch dazu geführt, dass Thurstone später den Faktor Reasoning definierte, welcher Induktion und Deduktion umfasst, nicht an eine bestimmte Art von Material gebunden ist und heute allgemein als schlussfolgerndes Denken übersetzt wird (Beckmann & Guthke, 1999; Jäger, 1967). Die Primärfähigkeit schlussfolgerndes Denken lädt in Aufgaben verschiedener Formate und Anforderungen (Amelang et al., 2006).

Eine weitere Modellvorstellung von Intelligenz, die Theorie der fluiden und kristallinen Intelligenz von Cattell, kann als Synthese aus der Zwei-Faktoren-Theorie und dem Primärfaktorenmodell angesehen werden (Holling, Preckel & Vock, 2004). Cattell, ein Schüler Spearman, ging 1957 davon aus, dass sich der g-Faktor aus zwei unabhängigen Komponenten, zwei Faktoren zweiter Ordnung zusammensetzt (Funke & Vaterrodt-Plünnecke, 2004). Er elaborierte und verbesserte seine Theorie mehrmals, u. a. in Zusammenarbeit mit Horn.

Cattell unterscheidet die fluide Intelligenz (Gf) und die kristalline Intelligenz (Gc) (Funke & Vaterrodt-Plünnecke, 2004). Die fluide Intelligenz bezeichnet die Fähigkeit, neue Probleme und Situationen zu bewältigen, sie ist von gelerntem Wissen unabhängig (Holling et al., 2004). Cattell nimmt an, dass fluide Intelligenz vor allem mittels Reasoning-Tests gemessen wird. (Beckmann & Guthke, 1999). Die kristalline Intelligenz sind kognitive Fertigkeiten, die durch die Kumulierung von Lernerfahrungen von Geburt an entwickelt wurden (Holling et al., 2004). Während die kristalline Intelligenz bis ins hohe Alter stabil bleibt oder sogar noch ansteigt, nimmt die fluide Intelligenz im Alter ab (Süß, 2003).

Fluide Intelligenz und kristalline Intelligenz werden durch mehrere untergeordnete Primärfaktoren determiniert (Horn & Cattell, 1966). Die fluide Intelligenz wird nach Horn und Cattell (1966) vor allem durch Aufgaben des schlussfolgernden Denkens, meist bei mehr figural-abstraktem Material, gemessen. Spätere Untersuchungen zeigten jedoch, dass Induktion nicht nur im Sekundärfaktor fluide Intelligenz, sondern auch in der kristallinen Intelligenz hoch lädt (Beckmann & Guthke, 1999). Nach Ansicht von Beckmann und

Guthke (1999) werden Testaufgaben die z.B. sprachlich-semantisches Material mit hohem Bildungsanteil enthalten, auch wenn sie fluide Intelligenz im Sinne des logischen Schlussfolgerns erfassen, immer auch Ladungen im Faktor kristalline Intelligenz aufweisen. Schlussfolgerndes Denken ist jedoch der wesentlichste Bestandteil der fluiden Intelligenz.

Sehr stark wird fluide Intelligenz durch mathematische Tests angesprochen, dennoch bezeichnet man sie besser als Fähigkeit zum schlussfolgernden Denken denn als mathematische Fähigkeit per se (Anderson, 2007).

Mitte der 1980er Jahre kam Horn zu dem Schluss, dass die Forschungsergebnisse neben Gf und Gc auf sieben weitere übergeordnete g-Fähigkeiten hindeuten. Das Konzept eines generellen g-Faktors gab Horn auf. Er behielt für seinen Ansatz die Bezeichnung „Gf-Gc Theorie“ bei, diese umfasst in der aktuellen Fassung neun Dimension (Maltby, Day & Macaskill, 2011).

Guilfords Ansatz, das Strukturmodell der Intelligenz, basiert nicht auf einem Generalfaktor. Er geht von einer großen Vielfalt unabhängiger Einzelfaktoren aus. Durch die Kombination von fünf Operationsklassen (Kognition, Gedächtnis, Evaluation, konvergente Produktion, divergente Produktion), vier Inhaltsklassen (figural, symbolisch, semantisch, Verhalten) und sechs Produktklassen (Einheit, Klasse, Relation, System, Transformation, Implikation) entstehen 120 verschiedene Fähigkeiten (Guilford & Hoepfner, 1976). In einer neueren Version wurde sogar auf 150 verschiedene Fähigkeiten erweitert (Funke & Vaterrodt-Plünnecke, 2004). Guilfords Analysen zum Denken führen zu Fähigkeiten, die in die Kategorien Kognition und konvergente Produktion einzuordnen, und sowohl induktiv als auch deduktiv sind (Guilford & Hoepfner, 1976). Es liegen zahlreiche kritische Einwände am Strukturmodell der Intelligenz vor, unbestritten ist jedoch, dass Guilford ein breites Spektrum an Tests entwickelte und damit viele Erkenntnisse über Einzelfunktionen u. a. zur Denkfähigkeit lieferte (Funke & Vaterrodt-Plünnecke, 2004; Guilford & Hoepfner, 1976).

Das Berliner Intelligenzstrukturmodell (BIS) von Jäger vereinigt Elemente aus den Kategoriensystemen von Spearman, Thurstone und Guildford und berücksichtigt allgemeine Erkenntnisse und gezielt durchgeführte Erhebungen der Intelligenzforschung (Amelang et al., 2006).

Eine Grundannahme des BIS-Modells ist, dass an jeder Intelligenzleistung alle intellektuellen Fähigkeiten, allerdings mit deutlich unterschiedlichen Gewichten, beteiligt sind. An der Spitze dieses hierarchischen Modells steht die allgemeine Intelligenz, welche sich aus sieben Fähigkeitskonstrukten zusammensetzt (Süß, 2003).

Die sieben Fähigkeiten lassen sich in die vier operativen Fähigkeiten Bearbeitungsgeschwindigkeit (B), Merkfähigkeit (M), Einfallsreichtum (E) und Verarbeitungskapazität (K) und in die drei inhaltlichen Fähigkeiten figuralbildhaftes Denken (F), sprachgebundenes Denken (V) und zahlengebundenes Denken (N), unterscheiden. Aus der Kombination der vier operativen mit den drei inhaltsgebundenen Fähigkeiten ergeben sich zwölf Arten von kognitiven Leistungen (Holling et al., 2004).

Der Berliner Intelligenzstruktur-Test von Jäger, Süß und Beauducel (1997) baut auf dem BIS auf. Betreffend schlussfolgerndes Denken ist vor allem die Fähigkeit Verarbeitungskapazität (K) bedeutsam. Die Skala Verarbeitungskapazität (K) erfasst Leistungsaspekte die, bei der Bearbeitung von mehr oder weniger komplexen Aufgabenstellungen, vielfältiges Beziehungsstiften, Erkennen von Regelmäßigkeiten und formallogisches Schlussfolgern erforderlich machen (z.B. in Zahlenfolgen-, Figurenfolgentests und Analogien) (Beckmann & Guthke, 1999). Für das Lösen des im BIS-Test vorgegebenen Aufgabentyps „Zahlenreihen fortsetzen“ muss beispielsweise sowohl das System der Zahlen beherrscht werden (inhaltsgebundene Komponente, N) wie auch die Fähigkeit zum logisch-schlussfolgernden Denken (operative Komponente, K) vorhanden sein (Funke & Vaterrodt-Plünnecke, 2004).

Ein weiteres Intelligenzmodell entwickelte Carroll (1993) mit der Three-Stratum-Theorie, welches Konzepte wie Spearman's g und seine spezifischen Faktoren,

Cattells Gc und Gf und Thurstones Primärfaktoren vereint. Die auf einer Reanalyse aller in der Literatur berichteten Korrelationsmatrizen basierende Three-Stratum-Theorie, weist drei Hierarchieebenen auf. Das Konstrukt der allgemeinen Intelligenz, ähnlich Spearmans g, befindet sich auf der dritten Ebene (Stratum III). Sie ist den folgenden acht Fähigkeiten der zweiten Ebene (Stratum II) übergeordnet: fluide Intelligenz (ähnlich Cattells Modell), kristalline Intelligenz, allgemeine Gedächtnisfähigkeit, visuelle Wahrnehmung, auditive Wahrnehmung, Abruffähigkeit, kognitive Geschwindigkeit, Verarbeitungsgeschwindigkeit. Diesen acht Fähigkeiten werden 69 spezifische Fähigkeiten der ersten Ebene (Stratum I) zugeordnet (Holling et al., 2004; Maltby et al., 2011). Mit der Fähigkeit fluide Intelligenz sind hier basale Prozesse des schlussfolgernden, logischen Denkens und andere mentale Aktivitäten, welche nur minimal durch Lernen und Akkulturation beeinflusst werden, gemeint (Holling et al., 2004). Carroll (1993) ordnet von den 69 spezifischen Fähigkeiten, die Konstrukte sequentielles Schlussfolgern, Induktion, quantitatives Schlussfolgern und Geschwindigkeit des Schlussfolgerns der fluiden Intelligenz zu. Schlussfolgern wird aber auch durch andere breite Fähigkeiten, wie dem Arbeitsgedächtnis, mitbestimmt (Beckmann & Guthke, 1999).

Moderne Intelligenzkonzeptionen wie die Triarchische Theorie von Sternberg und die Theorie der multiplen Intelligenzen von Gardner beziehen sich nicht auf die Ergebnisse von Faktorenanalysen (Funke & Vaterrodt-Plünnecke, 2004). Sternberg betrachtet vorwiegend mentale Prozesse, die bei Aufgabenlösungen bedeutsam sind (Holling et al., 2004). Im Fokus mehrerer kognitionspsychologischer Analysen von Sternberg ist das induktive Denken, wobei vor allem selektive Informationsverarbeitungs-, Kombinations- und Vergleichsprozesse experimentell untersucht werden (Heller, 1991).

Die Theorie von Sternberg beinhaltet die drei Subtheorien Komponenten-, Erfahrungs- und Kontextsubtheorie. Die Komponentensubtheorie betrachtet die Komponenten die zur Informationsverarbeitung erforderlich sind, sie betrifft Fragen wie „Welche geistigen Operationen sind für die Lösung von Problemen notwendig?“. Sternberg unterscheidet dabei drei wesentliche Komponenten.

Metakomponenten, sind mentale Prozesse höherer Ordnung, die bei der Planung, Überwachung und Bewertung von Problemlösungen erforderlich sind. Performanzkomponenten, sind mentale Prozesse niedrigerer Ordnung, die die Anweisungen der Metakomponenten ausführen. Wissenserwerbkomponenten, sind mentale Prozesse niedrigerer Ordnung, die Lernen und Wissenserwerb steuern (z.B. Vergleich von Informationen oder selektive Enkodierung) Die Erfahrungssubtheorie beschäftigt sich mit der Fähigkeit neuartige Probleme zu lösen sowie dafür erforderliche Informationsverarbeitung zu automatisieren. Je schneller Personen beim Lösen neuartiger Probleme sind, desto mehr Kapazitäten sind für die Automatisierung frei. Intelligenter Personen arbeiten dabei schneller und effizienter. Neuartigkeit und Automatisierung stehen demzufolge in einem wechselseitigen Zusammenhang. Die Kontextsubtheorie beschäftigt sich mit der praktischen Anwendung der Komponentensubtheorie im Umfeld einer Person. Sie soll klären, wie Personen ihre Umwelt beeinflussen, wie sie sich verschiedenen Umgebungen anpassen und wie sie sich neue Umwelten schaffen. Intelligente Menschen wissen wie und wann sie sich an eine bestimmte Umgebung anpassen können (Funke & Vaterrodt-Plünnecke, 2004; Holling et al., 2004).

Entgegen Sternberg postuliert Gardner (1991) in seiner Theorie der multiplen Intelligenzen, die Eigenständigkeit der verschiedenen Intelligenzformen. Auch hier spielt das schlussfolgernde Denken eine Rolle. Gardner (1991) diskutiert die linguistische, die logisch-mathematische und die räumliche Intelligenz sowie die musikalische, die körperlich-kinästhetische, die inter- und die intrapersonale Intelligenz.

Die logisch-mathematische Intelligenz wird bei der Lösung mathematischer Probleme, beim logischen Schließen oder beim Führen mathematischer Beweise benötigt. Nach Meinung Gardners ist die logisch-naturwissenschaftliche Intelligenz eine Erfindung der westlichen Welt. In vielen nichtwestlichen Gesellschaften sind anderen Arten der Intelligenz, wie der musikalischen, eine größere Bedeutung beizumessen (Funke & Vaterrodt-Plünnecke, 2004; Gardner, 1991).

Gardners Theorie bezieht Fähigkeiten ein, die in Intelligenztests bisher kaum erfasst werden. Es liegt jedoch keine empirische Bestätigung vor, dass die verschiedenen Intelligenzformen tatsächlich voneinander unabhängig sind, auch fehlen für die meisten dieser Intelligenzen hinreichende diagnostische Verfahren (Rost, 2008).

Trotz unterschiedlicher Definitionen und Klassifikationen, zeigt sich das schlussfolgernde Denken beim Großteil der Intelligenzmodelle, sowohl in klassischen Modellen wie in jenem von Spearman, Thurstone oder Cattell, als auch in modernen Modellen wie in jenem von Sternberg oder Gardner, als wesentlicher Aspekt bzw. Kernfaktor der Intelligenz. Schlussfolgerndes Denken wird in vielen Intelligenzmodellen als zentrale Anforderung in allen mit dem g-Faktor hochladenden Tests bezeichnet.

2.2.4 Geschlechterunterschiede im schlussfolgernden Denken

Geschlechterunterschiede in der intellektuellen Leistung haben schon viele hitzige Debatten ausgelöst. Obwohl sich Frauen und Männer nicht in ihrer allgemeinen Intelligenz unterscheiden, zeigen sie Unterschiede in spezifischen kognitiven Fähigkeiten (Halpern, 2000). Frauen erbringen bessere Leistungen bei Aufgaben, die sprachliche Intelligenz, Wahrnehmungsgeschwindigkeit und Gedächtnis erfordern. Männer verfügen über ein besseres räumliches Vorstellungsvermögen und zeigen bessere Leistungen bei Aufgaben die mathematisches Schlussfolgern erfassen, so die allgemeinen Forschungsergebnisse (Funke & Vaterrodt-Plünnecke, 2004; Halpern & LaMay, 2000). Im folgenden Teil soll vor allem auf die Geschlechterunterschiede welche schlussfolgerndes Denken betreffen eingegangen werden.

Bei Aufgaben zum proportionalen Denken und Schlussfolgern (Meehan, 1984; zitiert nach Halpern, 2000, S. 126) und Aufgaben zum naturwissenschaftlichen Schlussfolgern (Hedges & Nowell, 1995; zitiert nach Halpern, 2000, S. 126) erreichen Männer im Allgemeinen höherer Durchschnittswerte und somit Geschlechterunterschiede zu ihren Gunsten.

In einer Metaanalyse von Hyde et al. (1990), welche 100 Studien zusammenfasst, zeigte sich ein sehr kleiner Effekt ($d = 0.15$) bezüglich dem männlichen Vorteil bei mathematischen Fähigkeiten. Es ließ sich sogar für numerische Kalkulationen ein weiblicher Vorteil ($d = 0.14$) aufzeigen. Weiters ließ sich ein Alterstrend beobachten, bis zum 14. Lebensjahr waren keine bedeutsamen Unterschiede bzw. sogar ein sehr geringer Vorteil der Mädchen festzustellen, bei 15- bis 18-Jährigen war ein leichter Vorteil der Jungen erkennbar ($d = 0.29$). Bei College-Studenten ($d = 0.41$) und Erwachsenen ($d = 0.59$) verstärkte sich dieser Vorteil.

Strand, Deary und Smith (2006) zeigten unter Einsatz des Cognitive Abilities Test (CAT), dass 11 bis 12-jährige Mädchen signifikant bessere Leistungen im verbalen und non-verbalen schlussfolgernden Denken erbrachten und die Burschen hingegen im quantitativen schlussfolgernden Denken signifikant höhere Ergebnisse erzielten.

Kuhn und Holling (2009) untersuchten in ihrer Studie Geschlechterunterschiede in den schulischen Leistungen naturwissenschaftlicher Fächer und Sprachen, welche durch die Fähigkeit zum schlussfolgernden Denken herbeigeführt werden. Bei den untersuchten SchülerInnen aus den siebten bis zehnten Schulstufen, zeigte sich bezüglich Sprachen ein leistungsmäßiger Vorteil zugunsten der Mädchen, wohingegen in Naturwissenschaften und im schlussfolgernden Denken die Leistungen der männlichen Schüler überragten. Geschlechterunterschiede in der Leistung in naturwissenschaftlichen Fächern erklärten sich größtenteils aufgrund der Fähigkeit des schlussfolgernden Denkens. Die Fähigkeit zum schlussfolgernden Denken erscheint bedeutsam für die Prognose von schulischen Leistungen in naturwissenschaftlichen Fächern.

Die Geschlechterunterschiede in naturwissenschaftlichen und mathematischen Leistungen sind bei mittlerer Befähigung kleiner als bei höchstem Befähigungs- und Leistungsniveau (Halpern et al., 2007).

Vor allem in den Extrembereichen der mathematischen Fähigkeiten zeigten sich sehr starke Geschlechterunterschiede. Am höchsten waren die

Leistungsunterschiede für mathematisch Hochbegabte, hier waren beachtlich mehr Männer als Frauen zu finden. Aber auch im unteren Bereich des Fähigkeitsspektrums waren mehr Buben als Mädchen (Holling et al., 2004).

Anhand von Mathematik-Schulnoten zeigen sich oft keine Leistungsunterschiede zwischen den Geschlechtern, oder es können, wie in vielen Schulfächern, sogar bessere Noten bei den Mädchen beobachtet werden (Kimura, 2002).

Zum Vergleich von Schülerleistungen wurden die internationalen Studien TIMSS (Third International Mathematics and Science Study) und PISA durchgeführt.

Die TIMSS untersuchte Mathematik- und Naturwissenschaftsleistungen von SchülerInnen der Grundschule, der Sekundarstufe I und der Sekundarstufe II (Baumert, Bos & Watermann, 1998). Je nach Untersuchung nahmen bis zu 37 Länder an diesem internationalen Vergleich teil (Suchań, Wallner-Paschon, Bergmüller & Schreiner, 2008). Österreich beteiligte sich 1995 und 2007 an diesem Schülerleistungsvergleich, wobei 2007 nur die GrundschülerInnen an der Erhebung teilnahmen. Im Grundschulalter zeigten sich in der TIMSS 1995 keine bzw. unbedeutende Geschlechterunterschiede in den Mathematikleistungen in Österreich (Kaiser & Steisel, 2000). In der TIMSS 2007 verschlechterte sich die durchschnittliche Mathematikkompetenz wie auch die Naturwissenschaftsleistung der österreichischen VolksschülerInnen im Vergleich zum Ergebnis aus der TIMSS 1995 signifikant. Außerdem zeigte die TIMSS 2007 in beiden Bereichen signifikante Geschlechterunterschiede zugunsten der Buben auf (Suchań et al., 2008). Im Sekundarbereich I bzw. in der Mittelstufe zeigten sich in den Mathematikleistungen in einigen Ländern wie z.B. Spanien und Israel signifikante Geschlechterunterschiede nur zugunsten der Jungen, in Österreich war dieser Unterschied nicht signifikant (Kaiser & Steisel, 2000). In den naturwissenschaftlichen Leistungen erzielten die österreichischen Buben in der vierten Klasse der Mittelstufe signifikant höhere Ergebnisse als die Mädchen (Jungwirth, 1998). Weiters gaben österreichische Jungen in der Mittelstufe eine positivere Einstellung zur Mathematik an als

Mädchen (Kaiser & Steisel, 2000; Jungwirth, 1998). In der Sekundarstufe II bzw. Oberstufe schnitten die männlichen Schüler Österreichs sowohl im Bereich Mathematik als auch im Bereich Naturwissenschaften signifikant besser ab als sie weiblichen Schülerinnen (Jungwirth, 1998). Die Mathematikleistungen österreichischer OberstufenschülerInnen in der TIMSS/III lagen im internationalen Vergleich in einem schmalen Mittelfeld, das sie mit deutschen AbiturientInnen teilten. Auch beim Vergleich der Leistungsstärksten eines Altersjahrgangs lagen die österreichischen SchülerInnen im unteren Mittelfeld. Der Abstand zur internationalen Leistungsspitze war groß (Baumert et al., 1998).

PISA liefert seit dem Jahr 2000 alle drei Jahre Qualitätsindikatoren für die Grundkompetenzen in Lesen, Mathematik und Naturwissenschaft von 15- und 16-jährigen SchülerInnen in 65 Teilnehmerländern. In Österreich wiesen die letzten PISA-Erhebungen unterschiedlich große Leistungsunterschiede in Mathematik zwischen Mädchen und Buben auf. Die Erhebungen in den Jahren 2000 und 2006 zeigten in der Mathematikkompetenz relativ große Geschlechterunterschiede zugunsten der Burschen. Hingegen waren bei PISA 2003 diesbezüglich keine nennenswerten Geschlechterdifferenzen zu erkennen. 2009 stimmten die Ergebnisse für Österreich mit jenen aus den Jahren 2000 und 2006 überein. Die österreichischen Burschen lagen, wie die Burschen in den Nachbarländern Deutschland und Schweiz, im Jahr 2009 mit ihrer Mathematikleistung signifikant über jener der Mädchen. Leistungsunterschiede zwischen Österreichs weiblichen und männlichen Schülern im Bereich Naturwissenschaft, waren bei den PISA-Erhebungen mit etwas höheren Mittelwerten der Burschen, dennoch nicht signifikant (Schwantner & Schreiner, 2010).

Auch in der bei PISA 2003 erfassten fächerübergreifenden Problemlösekompetenz, unterschieden sich die Geschlechter nicht signifikant (Haider & Reiter, 2004).

Die Tests SAT (Scholastic Assessment Tests) und GRE (Graduate Record Examination) werden von vielen Colleges zur Auswahl ihrer Studierenden

eingesetzt. Diese Tests, messen nicht angeborene Eigenschaften sondern entwickelte Fähigkeiten. Sie werden häufig wie traditionelle Intelligenztests genutzt (Halpern & LaMay, 2000). Auch beim SAT (Halpern, 2000) und beim GRE-Quantitative (Graduate Record Examination-Quantitative) (Willingham & Cole, 1997) wurden Geschlechterunterschiede zugunsten der Männer im Bereich Mathematik beobachtet.

Der männliche Vorteil bei modernen Tests zum mathematischen Schlussfolgern, beim Lösen mathematischer Probleme hängt vermutlich mit den jeweils verwendeten visuell-räumlichen Strategien zusammen (Halpern & LaMay, 2000).

In einer Studie zum Zusammenhang zwischen visuell-räumlichen und quantitativen Fähigkeiten von Casey, Nuttal, Pezaris und Benbow (1995) wurden ein Test zur mentalen Rotation sowie der Scholastic Assessment Test-Mathematics (SAT-M) und der Scholastic Assessment Test-Verbal (SAT-V) verwendet. Bei der Untersuchung von Studenten, High-School-Abgängern mit hohen und niedrigen Fähigkeiten und mathematik-talentierte präadoleszenten Jugendlichen, insgesamt 760 Personen, zeigte sich, dass die männlichen Teilnehmer die weiblichen Teilnehmerinnen im Test zur mentalen Rotation sowie im SAT-M in allen Gruppen leistungsmäßig übertrafen, mit Ausnahme der Gruppe der High-School-Abgänger mit niedrigen Fähigkeiten. Die Autoren folgern, dass wenn mathematische Probleme zusätzlich noch visuell-räumliche Strategien erfordern, z.B. im Bereich Geometrie, die Jungen den Mädchen überlegen sind (Casey et al., 1995). Hingegen bei der numerischen Kalkulation (Addition, Subtraktion, etc.) besteht ein Vorteil zugunsten der Frauen (Kimura, 2002).

Basierend auf ihrer Metaanalyse kam Friedman (1995) jedoch zu der Ansicht, dass die Leistung in mathematischen Tests stärker mit verbalen Kompetenzen korrelierte als mit räumlichen Fähigkeiten. Der Zusammenhang zwischen mathematischen und räumlichen Kompetenzen war vermutlich durch eine gemeinsame Kovariation mit der Intelligenz bedingt (Geary, Sauls, Liu & Hoard, 2000). Nach einer Studie von Kimura (2000) war der Zusammenhang

zwischen den Ergebnissen im mathematischen Schlussfolgern mit einem mentalen Rotationstest nicht größer als mit einem Wortschatztest. Dies deutet, so Kimura (2000), darauf hin, dass die Rotation von abstrakten Figuren eine Fähigkeit ist, die nicht mehr zur mathematischen Fertigkeit beiträgt als ein generelles Intelligenzmaß.

Die Hypothese von Kyllonen und Christal (1990), dass „reasoning“ und „working memory“ nahezu identisch seien, konnte nicht bestätigt werden. Jedoch werden Leistungen in Reasoning-Tests, vor allem solche die gleichzeitiges Behalten mehrerer Zwischenschritte, Operationen und unterschiedlicher mentaler Modelle verlangen, von der Kapazität des Arbeitsgedächtnisses mitbestimmt (Beckmann & Guthke, 1999).

Personen mit guten Leistungen in mathematischen Problemaufgaben, zeigten bessere Fähigkeiten Informationen im Kurzzeitgedächtnis zu repräsentieren und zu manipulieren. Außerdem waren sie besser darin, verbal gestellte Probleme so zu übersetzen, dass diese numerisch gelöst werden konnten (Dark & Benbow, 1990).

Diese Ergebnisse könnten darauf hindeuten, dass Frauen und Männer sich möglicherweise gar nicht per se in der mathematischen Begabung unterscheiden, sondern jedes Geschlecht verschiedene Basiskompetenzen anwendet, welche sich dann abhängig vom Aufgabentyp mehr oder weniger effektiv erweisen. Wenn die Lösungsstrategie experimentell beeinflusst wurde, beispielsweise mittels Ermutigung eine räumliche Strategie beim Lösen eines mathematischen Problems anzuwenden, so zeigte sich, dass sich der Geschlechterunterschied reduzieren kann (Geary, 1996).

Auch Gallagher, Levin und Cahalan (2002) fanden beim Lösen mathematischer Probleme, die üblichen Geschlechterunterschiede zugunsten der Männer, wenn es vorteilhaft war räumlich-basierende Lösungsstrategien anzuwenden. Jedoch zeigten sich diese Vorteile für die Männer nicht, wenn die Lösungsstrategien verbal oder ähnlich wie bereits bekannte Strategien aus Mathematik-Lehrbüchern waren.

Auch bei einer Untersuchung zur Selbsteinschätzung der eigenen Fähigkeiten zeigten sich Geschlechterunterschiede. Während Männer ihre mathematische, räumliche und schlussfolgernde Intelligenz höher bewerteten, gaben Frauen höhere Werte bei der musikalischen und interpersonalen Intelligenz an. Des Weiteren gab es signifikante Beweise dafür, dass die Selbsteinschätzung der Intelligenz durch geschlechterspezifische Stereotype beeinflussbar ist (Rammstedt & Rammsayer, 2000).

Die Geschlechterforschung der letzten Jahrzehnte zeigt eindeutig, dass sich Frauen und Männer in bestimmten kognitiven Fähigkeiten unterscheiden. Ursachen für kognitive Geschlechterunterschiede können sowohl auf soziale als auch auf biologische Faktoren zurückgeführt werden (Funke & Vaterrodt-Plünnecke, 2004). Halpern et al. (2007) folgerten, dass frühe Erfahrung, biologische Faktoren, Bildungspolitik und kultureller Kontext die Anzahl der Frauen und Männern, welche ein Studium der Naturwissenschaft und Mathematik verfolgen, beeinflusst und, dass diese Effekte auf komplexe Weise interagieren.

So ist beispielsweise das Erziehungsverhalten vieler Bezugs- und Lehrpersonen immer noch vom klassischen Geschlechterrollenbild geprägt. Für Mädchen und Buben liegen verschiedene Erziehungsideale vor, bei denen rollenkonforme Verhaltensweisen verstärkt werden (Beerman et al., 1992). Mädchen bekommen mehr Lob für Fleiß und Anpassung. Bei Buben werden höhere Leistungsanforderungen gestellt (Hoffmann, 1988a, 1988b; zitiert nach Beerman et al., 1992, S. 61). Oft wird den Mädchen gesagt, dass Mathematik für sie im Berufsleben nicht relevant ist (Beerman et al., 1992).

Beobachtungen und Auswirkungen solcher Geschlechterstereotypen insbesondere in Verbindung mit dem Phänomen Stereotype Threat werden in Kapitel „2.3.2.1 Einfluss von Stereotype Threat auf kognitive Leistungen und Methoden um diesen zu reduzieren“ behandelt.

Allgemeingültige Antworten auf die komplexen Fragen zu Geschlechterunterschieden in Bereich Naturwissenschaft und Mathematik gibt es nicht (Halpern et al., 2007). Die spezifischen Geschlechterunterschiede

werden von verschiedenen Faktoren mitbestimmt (Holling et al., 2004). Halpern (2000) und auch andere Autoren schlagen ein psychobiosoziales Modell vor, welches die kognitiven Geschlechterunterschiede auf komplexe Wechselwirkungen zwischen sozialen, biologischen und psychologischen Faktoren, zurückführt.

2.3 Stereotype Threat

In diesem Kapitel werden geschlechtsspezifische Unterschiede in kognitiven Leistungen, hauptsächlich in Mathematik- und Problemlöseaufgaben welche typisch sind um das schlussfolgernde Denken zu erheben (Anderson, 2007), in Bezug auf den sozialen Einflussfaktor Stereotype Threat behandelt.

Es werden Stereotype, Stereotype Threat und ihre Wirkung auf Leistungen, mit Augenmerk auf Aspekte wie Geschlecht, Arbeitsgedächtnis und Angst, angeführt. Weiters werden Methoden um Stereotype Threat zu reduzieren beschrieben.

2.3.1 Definition von Stereotyp

Aronson, Wilson und Akert (2008) bezeichnen ein Stereotyp als eine verallgemeinernde Annahme über eine Gruppe von Menschen, die praktisch all ihren Mitgliedern bestimmte Eigenschaften zuschreibt, unabhängig von tatsächlichen Unterschieden zwischen den Mitgliedern.

Stereotypisierung dient uns oft als Technik, um unsere Sicht von der Welt zu vereinfachen, überschaubarer zu machen und die Orientierung zu erleichtern (Allport, 1954; Herkner, 2003). Die Welt ist zu kompliziert, um allem gegenüber eine hoch differenzierte Einstellung zu haben, deshalb entwickeln wir zu einigen Themen ausgewählte, exakte Einstellungen und bei anderen verlassen wir uns auf oberflächliche und einfache Ansichten (Allport, 1954). Stereotype

sind kognitive Schemata, welche mit vereinfachenden Urteils- und Verarbeitungsheuristiken einhergehen (Klauer, 2008).

Die Kategorisierung von Personengruppen, kann aufgrund von Eigenschaften wie beispielsweise Hautfarbe, Geschlecht, Fähigkeiten und unzähligen weiteren erfolgen (Klauer, 2008). Stereotype sind aufgrund selektiver Wahrnehmung, selektiver Informationssuche, selektiven Gedächtnisses und stereotypekonsistenter Interpretationen sehr stabil und änderungsresistent (Herkner, 2003). Der Schaden der durch sie angerichtet werden kann ist groß. (Aronson et al., 2008). Sie beeinflussen sowohl die wahrnehmende Person als auch die Mitglieder der stereotypisierten Personengruppe (Petersen & Six, 2008). Stereotype können aber auch positive Merkmale betreffen (Aronson et al., 2008).

2.3.2 Definition und Ursprung von Stereotype Threat

Viele Studien belegen kognitive Leistungsunterschiede zwischen den Geschlechtern. Neben biologischen können auch soziale Einflussfaktoren dafür gefunden werden, wie etwa das Hemmen der Leistung aufgrund eines gesellschaftlichen Stereotyps (Maderthaner, 2008). In diesem Zusammenhang nennen Steele und Aronson erstmals 1995 das Phänomen „Stereotype Threat“ (Bedrohung durch Stereotype). Dabei handelt es sich um die Angst von Mitgliedern einer Gruppe, ihr Verhalten könne ein kulturelles Stereotyp bestätigen (Aronson et al., 2008).

Ist die Identifikation mit der stereotypisierten Gruppe aktiv, so können negative Stereotype, die der Gruppe zugeschrieben werden, die eigenen kognitiven Fähigkeiten negativ beeinflussen (Steele, 1997). Zu wissen, dass andere Personen eine negative Meinung oder Einstellung über die Gruppe haben, der man sich zugehörig fühlt (z.B. „Frauen sind schlecht in Mathematik“), kann zur Besorgnis führen, dass die Stereotype gegenüber der Gruppe auf die eigene Person übertragen werden. Somit kann das zuvor aktivierte Stereotyp von mathematisch unbegabten Frauen, tatsächlich stereotypkonformes Verhalten,

wie schlechte Ergebnisse von Mädchen in Mathematiktests bewirken (Heilemann, Hackl, Neubauer & Stöger, 2012).

Steele und Aronson wollten mit ihren klassischen Experimenten zu Stereotype Threat die Leistungsunterschiede in kognitiven Fähigkeiten zwischen weißen und afroamerikanischen StudentInnen erklären (Aronson et al., 2008). Während sie der einen Hälfte der afroamerikanischen und weißen StudentInnen mitteilten, es handle sich bei dem durchgeführten mündlichen Test um einen Fähigkeitstest, ließen sie die andere Hälfte in dem Glauben, der Test werde erst entwickelt und sage deshalb nichts über ihre tatsächlichen Fähigkeiten aus. Die Forscher nahmen an, dass die Vorgabe eines als diagnostisch wahrgenommenen Tests bei afroamerikanischen Studierenden ein Gefühl der Bedrohung auslösen würde. Die Ergebnisse unterstützen diese Annahme. Während die afroamerikanischen Studierenden, die den Test nicht als diagnostisch wahrnahmen, ebenso gute Leistungen wie die weißen Studierenden erzielten, lag die Leistung der afroamerikanischen Studierenden die glaubten der Test messe ihre intellektuellen Fähigkeiten, unter der Leistung der weißen Studierenden und auch unter jener der afroamerikanischen TeilnehmerInnen der anderen Gruppe. Die Leistungen der weißen Studierenden waren in beiden Gruppen gleich gut. Die Untersuchung zeigte, dass Stereotype Threat die intellektuelle Testleistung von afroamerikanischen Studierenden beeinträchtigen kann (Aronson et al., 2008).

Allein zu wissen, dass ein Test dazu bestimmt ist eine Diagnose über die eigenen Fähigkeiten in einem stereotype-relevanten Bereich stellen zu können, genügt oft schon um eine Bedrohung durch Stereotype auszulösen (Steele & Aronson, 1995). Die Beständigkeit und Robustheit des Phänomens Stereotype Threat und die Auswirkungen auf den Erfolg von stigmatisierten Individuen, führten zu erhöhtem Interesse seitens der Forschung (Johns, Schmader & Martens, 2005).

2.3.2.1 Einfluss von Stereotype Threat auf kognitive Leistungen und Methoden um diesen zu reduzieren

Das Phänomen Stereotype Threat gilt auch in Bezug auf Geschlechter. Der Einfluss von Geschlechterstereotypen auf die kognitiven Leistungen von Frauen und Männern, steht im Zentrum vieler Studien.

Untersuchungen deuten darauf hin, dass Frauen bei Mathematiktests und intellektuellen Test teilweise deswegen schlechtere Leistungen erbringen, weil die Besorgnis besteht, dass ihre Leistung mit den negativen Stereotypen ihrer Gruppe (z.B. „Frauen sind schlecht in Mathematik“) übereinstimmen könnte (Steele, Spencer & Aronson, 2002).

Die Bereiche Mathematik, Naturwissenschaften und Technik werden häufig als männliche Domäne stereotypisiert. Demzufolge herrscht die allgemeine Meinung vor, dass Frauen und Mädchen darin weniger leisten könnten als Männer (Beerman et al., 1992).

Entsprechend zeigten sich bei der Mathematikleistung von Frauen, die Stereotype über Mathematik als männliche Domäne unterstützten und befürworteten, größere Auswirkungen des Stereotype Threat-Phänomens. Sie verzeichneten schlechtere Ergebnisse als Frauen die diesen Stereotypen nicht zustimmten (Schmader, Johns & Barquissau, 2004).

Untersuchungen zu Geschlechterrollen zeigten, dass sich Personen mehr anstrebten kognitive Aufgaben zu lösen, wenn diese Aufgaben Funktionen maßen die mit der eigenen Geschlechterrolle übereinstimmten. Wurde ein vorgegebener Test zur räumlichen Visualisierung den Studierenden im Vorfeld als räumlicher Test beschrieben, so erbrachten weibliche Studierende deren Geschlechterrolle eher männlich war bessere Leistungen. Wurden die Aufgaben jedoch als Test zur Empathie beschrieben, schnitten Frauen mit einer eher weiblichen Geschlechterrolle besser ab (Massa, Mayer & Bohon 2005).

In einem Experiment von Spencer et al. (1999) wurden weibliche und männliche Studierende mit äquivalentem mathematischen Hintergrund getestet. Der einen Gruppe der StudentInnen wurde gesagt, der durchgeführte

Mathematiktest habe in der Vergangenheit Geschlechterunterschiede aufgezeigt. Der anderen Gruppe wurde erzählt der Mathematiktest sei nicht geschlechtersensitiv. Unter jenen die dachten, der Test zeige keine geschlechterspezifischen Leistungsunterschiede waren auch tatsächlich keine Differenzen zu erkennen. Hingegen unter jenen die in der Annahme waren, der Test zeige Leistungsunterschiede zwischen den Geschlechtern, schnitten die Frauen nicht so gut ab wie ihre männlichen Kollegen.

Keller und Dauenheimer (2003) untersuchten in ihrer Studie das Phänomen Stereotype Threat bei SchülerInnen der 10. Schulstufe. Alle SchülerInnen bearbeiteten die gleichen Mathematikaufgaben. Die beiden zufällig zugewiesenen Testversionen unterschieden sich nur darin, dass in einer Version vor Bearbeitungsbeginn schriftlich mitgeteilt wurde, dass die vorgegebenen Aufgaben in vorherigen Studien Geschlechterunterschiede gezeigt hätten, während in der anderen Testversion angegeben wurde, dass sich in früheren Studien keine Geschlechterunterschiede gezeigt hätten. Es ließ sich erkennen, dass Mädchen beträchtlich bessere Ergebnisse erreichten, wenn sie die Aufgaben als geschlechterneutral beschrieben bekamen. Ihre Leistungen waren hier gleichermaßen gut wie die der männlichen Jugendlichen. Wenn hingegen das Geschlechterstereotyp anwendbar war, erbrachten Jungen bessere Leistungen als Mädchen. Die Ergebnisse stützen das Phänomen Stereotype Threat und zeigen, dass kleine Variationen in der Aufgabenbeschreibung beträchtlich Auswirkungen auf die Mathematikleistung der weiblichen Jugendlichen haben.

Studienergebnisse von Davies et al. (2002) erbringen den Beweis dafür, dass Medien eine allgegenwärtige soziokulturelle Quelle von Stereotype Threat sind. Die Autoren wiesen darauf hin, dass das Anschauen von geschlechterstereotypen Fernsehwerbungen die mathematischen Leistungen von Frauen reduzierte. Weibliche Studierende welche stereotype Werbespots ansahen (z.B. eine Werbung in der einer Frau das Wasser im Mund zusammenlief, aus Vorfreude darauf einen neuen Brownie-Mix auszuprobieren), erbrachten beim darauf folgenden Mathematiktest schlechtere Leistungen, als Frauen, die kontra-stereotype Werbespots sahen

(z.B. eine Werbung in der eine attraktive Frau einen Mann mit ihrem Wissen über Autos imponierte) (Davies et al., 2002). Angesichts dieser Ergebnisse kann nur gemutmaßt werden, welche Auswirkungen die unzähligen Werbungen haben die Menschen stereotypisiert zeigen, und denen wir täglich ausgesetzt sind (Aronson et al., 2008).

Johns et al. (2005) untersuchten in ihrer Untersuchung eine Methode zur Reduzierung der Bedrohung durch Stereotype. Sie teilten die TestteilnehmerInnen in drei Gruppen. Während der ersten Gruppe mitgeteilt wurde, sie solle für eine Studie welche allgemeine Aspekte kognitiver Prozesse erhebe, Problemlöseaufgaben bearbeiten, wurde die zweite Gruppe instruiert einen Mathematiktest zu bearbeiten, um damit Geschlechterunterschiede in der Mathematikleistung zu erheben. Die TeilnehmerInnen der dritten Gruppe erhielten die gleiche Instruktion wie jene der Zweiten mit dem Zusatz, dass sie aber Informationen über Bedrohung durch Stereotype bekamen. Die Frauen wurden darauf hingewiesen, sich daran zu erinnern, dass falls sie sich während der Testbearbeitung ängstlich fühlen, diese Ängstlichkeit Resultat negativer Stereotype sein könnte, welche in der Gesellschaft weit verbreitet sind. Die Ängstlichkeit habe mit gegenwärtigen Fähigkeiten um im Test gut abzuschneiden nichts zu tun. Die Ergebnisse zeigten, dass Frauen der zweiten Gruppe, welche laut Instruktion einen Mathematiktest zur Erhebung geschlechterspezifischer Unterschiede bearbeiteten, eine schlechtere Leistung erzielten als Männer dieser Gruppe. Hingegen in der dritten Gruppe, in der über die Bedrohung durch Stereotype aufgeklärt wurde, war dies nicht zu beobachten. Die Leistung der Frauen wurde hier durch die Information und Aufklärung über Bedrohung durch Stereotype verbessert. Johns et al. (2005) zeigten mit dieser Arbeit, dass der einfache Ansatz, stigmatisierte Personen über Bedrohung durch Stereotype aufzuklären und zu informieren, bereits hilfreich sein kann um die negativen Auswirkungen auf die Leistungen abzuschwächen.

Marx und Roman (2002) gehen davon aus, dass die Anwesenheit eines mathematisch kompetenten weiblichen Vorbildes die Stereotype Threat-Effekte auf die Mathematikleistung der Frauen beseitigt. Dem Phänomen Stereotype

Threat wirkten sie entgegen indem sie die Mitglieder einer stereotypisierten Gruppe, mit erfolgreichen Rollenbildern der eigenen Gruppe konfrontierten. Es zeigte sich, dass die Anwesenheit eines mathematisch kompetenten weiblichen Vorbildes (Versuchsleiterin), die leistungshemmenden Auswirkungen des Phänomens Stereotype Threat auf die Mathematikleistungen von Frauen, verhinderte. Ausschlaggebend dabei war die wahrgenommene mathematische Kompetenz der Versuchsleiterin. Frauen, die auf eine mathematisch kompetente Versuchsleiterin trafen, erbrachten höhere Ergebnisse als Frauen die einer Versuchsleiterin mit niedrig wirkender Mathematikkompetenz entgegentraten. Hingegen hatten Männer unter einer Versuchsleiterin die mathematisch kompetent wirkte, niedrigere Ergebnisse als unter einer weniger kompetent wirkenden Testleiterin. Weiters zeigten Marx und Roman (2002), dass das Wahrnehmen einer kompetenten Versuchsleiterin die selbst eingeschätzte Mathematikfähigkeit der Frauen erhöhte, was wiederum zu einer besseren Leistung beim schwierigen Mathematiktest führte. Mit ihren Untersuchungen deuten Marx und Roman (2002) darauf hin, dass situationsbezogene Faktoren die Mathematikleistung von Frauen beeinträchtigen können.

McGlone und Aronson (2006) sind der Meinung die Wirkungen des Phänomens der Bedrohung durch Stereotype lässt sich umkehren. Sie glauben, dass eine alternative Denkweise die dem Stereotyp zuwiderläuft, die Leistung fördern kann. In einer ihrer Untersuchungen wurden die teilnehmenden Studierenden, bevor sie einen Test zur räumlich Fähigkeit absolvierten daran erinnert, dass sie Studierende eines „ausgewählten Privat-Colleges“ seien. Die TeilnehmerInnen der Kontrollgruppe wurden daran erinnert „Bewohner des Nordostens“ zu sein. Im Gegensatz zur Kontrollgruppe in der sich Leistungsunterschiede zwischen männlichen und weiblichen Studierenden fanden, zeigten sich keine kognitiven Geschlechterunterschiede bei den Studierenden, die daran erinnert wurden ein ausgewähltes Privat-College zu besuchen. Das Stereotyp „Frauen sind in Tests zu räumlichen Fähigkeiten nicht gut“ wurde von der Ansicht „ich bin ein guter Studierender“

verdrängt und die weiblichen Studierenden erbrachten dadurch signifikant bessere Leistungen.

Ähnliche Ergebnisse erbrachten McGlone und Aronson (2007) auch in ihrer darauf folgenden Studie, in der sie Strategien um Bedrohung durch Stereotype zu bewältigen, untersuchten. Dabei wurden die StudentInnen in der Kontrollgruppe instruiert, bei Schwierigkeiten im vorgegebenen Mathematiktest durchzuhalten. Die beiden anderen Gruppen wurden kurz über das Phänomen Stereotype Threat und dessen mögliche negative Beeinträchtigung informiert. Eine der informierten Gruppen wurde angewiesen, jede Bedrohung die sie während des Tests wahrnimmt dadurch zu bewältigen, dass sie versucht Gedanken über das Stereotyp zu unterdrücken. Bei der zweiten Gruppe die informiert wurde, wurde ein alternatives, selbstrelevantes positives Stereotyp aktiviert. Man sagte ihnen, dass laut Forschungsbelegen Studierende eines Privat-Colleges, wie jenes das sie selbst besuchten, weniger durch die Bedrohung durch Stereotype gefährdet sind, als StudentInnen von weniger renommierten Colleges. In der Kontrollgruppe wurde eine Leistungskluft zugunsten der Männer beobachtet. Wenn die Studierenden versuchten die bedrohlichen Gedanken zu unterdrücken vergrößerte sich diese Kluft noch mehr. Hingegen bei jenen Personen, bei denen das selbstrelevante positive Stereotyp des Privat-Colleges aktiviert wurde, verringerte sich dieser Unterschied. Die Leistung der Frauen war hier höher als in den anderen beiden Gruppen und unterschied sich nicht von jener der Männer. Die Ergebnisse zeigten, dass bei Frauen der Stereotype Threat-Effekt in Testsituationen abgeschwächt werden kann, wenn ihre Zugehörigkeit zu erreichten sozialen Kategorien (z.B. Studentin eines Privat-Colleges) mehr hervorsteht als ihre Zugehörigkeit zu einer zugeschriebenen Kategorie (z.B. weiblich).

2.3.2.2 Mögliche verantwortliche Prozesse für Stereotype Threat

Die Forschung hat gezeigt, dass negative Stereotypen zu einer geringeren Leistungsfähigkeit der stereotypisierten Individuen führt. Die kognitiven

Mechanismen und unterliegenden Prozesse dieser Effekte sind nicht endgültig erforscht (Schmader & Johns, 2003).

Das Arbeitsgedächtnis spielt eine zentrale Rolle in Bezug auf kognitive und soziale Stereotype Threat-Effekte (Schmader, Johns & Forbes, 2008).

Schmader und Johns (2003) sind der Meinung, dass die Leistung dadurch eingebüßt wird, dass durch die Verarbeitung von Informationen die mit negativen Stereotypen zusammenhängen, kognitive Ressourcen verbraucht werden. Die Bedrohung durch Stereotype beeinträchtigt die Testleistung, weil sie die Kapazität des Arbeitsgedächtnisses reduziert. Sie zeigten in ihrer Untersuchung auf, dass die Bedrohung durch Stereotype die Arbeitsgedächtniskapazität bei Frauen verringerte.

Auch Régner et al. (2010) beschäftigten sich mit Stereotype Threat und dem Arbeitsgedächtnis. Sie untersuchten ob Frauen mit hoher Arbeitsgedächtniskapazität gegenüber Stereotype Threat-Effekte resistenter sind als Frauen mit geringer Arbeitsgedächtniskapazität. Den StudentInnen mit entweder geringer oder hoher Arbeitsgedächtniskapazität (die Zuteilung erfolgte im Vorfeld), wurde der Raven Matrizen test vorgegeben um das schlussfolgernde Denken zu erheben. Ein Stereotype Threat-Effekt war nur bei Personen mit geringer Arbeitsgedächtniskapazität zu beobachten. Frauen mit geringer Kapazität im Arbeitsgedächtnis erbrachten schlechtere Leistungen, wenn ihnen gesagt wurde, es handle sich um einen diagnostischen Test zum schlussfolgernden Denken, als Männer mit geringer Kapazität im Arbeitsgedächtnis. Wurde den TeilnehmerInnen mitgeteilt, der zu bearbeitende Test sei diagnostisch, zeige aber keine kognitiven Unterschiede zwischen den Geschlechtern, so waren die Leistungen der weiblichen und männlichen Personen mit niedriger Arbeitsgedächtniskapazität gleich gut. StudentInnen mit hoher Kapazität im Arbeitsgedächtnis erzielten in allen Bedingungen gleich gute Ergebnisse, das Phänomen Stereotype Threat ließ sich hier nicht feststellen. Da Ravens Matrizen test die Verwendung von Routinen ausschließt, kann die Bedrohung durch Stereotype nicht der Bereitstellung von Aufmerksamkeit auf automatisierte Prozesse zugeschrieben werden (Régner et

al., 2010). Eine eher wahrscheinliche Erklärung für die Ergebnisse ist, dass nur geringe Ressourcen im Arbeitsgedächtnis verfügbar sind und der Erfolg bei diesen Aufgaben in hohem Maße auf der Kapazität des Arbeitsgedächtnisses beruht (Unsworth, Brewer, & Spillers, 2009; zitiert nach Régner et al., 2010, S. 1647).

Nach Schmader et al. (2008) verbrauchen Personen die von einer Bedrohung durch Stereotype betroffen sind, um negative Gedanken und Emotionen zu unterdrücken Teile ihrer Ressourcen, was Grund für eine Beeinträchtigung des Arbeitsgedächtnisses sein könnte. Die Autoren entwarfen um Stereotype Threat zu erklären ein Modell das physiologische, affektive, kognitive und selbst-regulierende Prozesse integriert. Laut Schmader et al. (2008) lösen die meisten, wenn nicht sogar alle Situationen der Bedrohung durch Stereotype, bestimmte physiologische und psychologische Prozesse aus, die die allgemeinen exekutiven Ressourcen welche für eine Vielzahl von Leistungen benötigt werden, beeinträchtigen. Effiziente Leistungen in kognitiven und sozialen Aufgaben erfordern koordinierte Informationsverarbeitung und werden jedoch zugleich durch störende und ablenkende Informationen gehemmt.

Nach dem Prozessmodell von Schmader et al. (2008) wird angenommen, dass Stereotype Threat die Leistung über drei verschiedene, jedoch in Wechselbeziehung stehende Mechanismen stört, nämlich über die physiologische Stressreaktion, über die erhöhte Wachsamkeit und über die Unterdrückung negativer Gedanken und Emotionen.

Wenn physiologischer Stress nicht direkt das Arbeitsgedächtnis reduziert, so kann jedoch die erhöhte Wachsamkeit bezüglich der eigenen Leistung, oder die Unterdrückung negativer Emotionen das Arbeitsgedächtnis verringern. Auch die Leistung in sensomotorischen Aufgaben, welche automatische Verarbeitung erfordern, wird durch die Stereotypebedrohung beeinträchtigt.

Die Arbeit von Johns, Inzlicht und Schmader (2008) steht im Einklang mit dem integrierten Prozessmodell das Schmader et al. (2008) vorschlagen.

Johns et al. (2008) konnten in ihren Untersuchungen zeigen, dass bei dem von ihnen vorgegebenen Mathematiktest Personen unter Stereotype Threat

tatsächlich versuchten Angst zu unterdrücken. Sie wiesen nach, dass von Stereotype Threat betroffene Personen spontan versuchen ihre Angst durch Unterdrückung zu kontrollieren und, dass solche Emotionsregulationen exekutive Ressourcen aufbrauchen die für gute Leistungen in kognitiven Tests benötigt werden. Sie zeigten auch, dass wenn man bedrohten Personen Möglichkeiten bietet negative Emotionen effektiv zu bewältigen, wie etwa die Situation oder die Bedeutung ihrer Angst neu zu bewerten, kann dies exekutive Ressourcen zurückgeben und die Testleistung verbessern. Wurde nämlich den Frauen unter Aktivierung von Stereotype Threat zusätzlich mitgeteilt, dass frühere Forschungsergebnisse gezeigt haben, dass Ängstlichkeit nicht schadet, sondern sogar bei der Bearbeitung derartiger vorgegebener mathematischer Probleme hilfreich sein kann, so erbrachten sie höhere Ergebnisse als Frauen in der Stereotype Threat-Bedingung, denen nur gesagt wurde, dass sie einen Mathematiktest bearbeiten sollten und die Untersuchung dazu diene die Beziehung zwischen Ängstlichkeit und Mathematikleistung zu erklären.

Auch weitere Forschungen untersuchten die beteiligten Prozesse bei Stereotype Threat. Wenn schwierige mathematische Probleme gelöst werden sollten, war die Fähigkeit von Frauen Problemlösestrategien zu formulieren unter der Stereotype Threat-Bedingung geringer als bei Männern. Wurde ihnen gesagt, dass bei früheren Vorgaben dieser Probleme Frauen und Männer gleich gute Leistungen erbrachten und somit die Stereotypebedrohung reduziert, dann zeigten die beiden Geschlechter tatsächlich keine Unterschiede bei der Formulierung und Anwendung von Problemlösestrategien (Quinn & Spencer, 2001).

Carr und Steele (2009) untersuchten ebenfalls die Auswirkungen von Stereotype Threat auf die Problemlösefähigkeit. Sie beschäftigten sich in ihren Untersuchungen damit, ob Frauen die durch negative Stereotype Threat belastet werden, unter Bedingungsänderungen weniger dazu fähig sind alte Lösungsstrategien abzulegen und stattdessen neue, effizientere anzuwenden. Sie fanden heraus, dass die Bedrohung durch Stereotype, unflexibles Verharren beim Problemlösen steigert. Frauen unter Aktivierung von Stereotype Threat, die glaubten sie würden einen diagnostischen

Mathematiktest durchführen, verharrten eher bei Problemlösestrategien die sich im Vorfeld als erfolgreich erwiesen, aktuell aber nicht optimal und falsch waren, als Frauen in der Kontrollgruppe, die meinten sie würden nicht-diagnostische Puzzleaufgaben bearbeiten. Aufgrund der Ergebnisse nehmen die Autoren an, dass Stereotype Threat zu einer Inflexibilität im Denken führt. Flexibilität ist jedoch wichtig für die Bearbeitung von Tests, um von Aufgabe zu Aufgabe neue Strategien und Lösungsansätze adaptieren zu können.

2.4 Risikoverhalten und Entscheidungen

In diesem Abschnitt der Arbeit werden die Begriffe Risiko und Entscheidungen näher erläutert. Es wird auf Entscheidungsmodelle und Unterschiede im Risikoverhalten, insbesondere in Bezug auf das Geschlecht, eingegangen sowie auf Untersuchungsergebnisse zum Game of Dice Task (GDT).

2.4.1 Begriffsbestimmung Risiko und Entscheidung

Der Begriff „Risiko“ wird in verschiedenen Disziplinen (z.B. Gesundheitswesen, Wirtschaft) unterschiedlich erklärt.

Ganz allgemein wurde Risiko, als eine Situation in der der Entscheider Wahrscheinlichkeiten für das Auftreten zukünftiger Umweltzustände angeben kann, definiert (Laux, Gillenkirch & Schenk-Mathes, 2012).

Risiko charakterisiert sich mit der Wahrscheinlichkeit des Eintretens eines Schadens im negativen Fall, oder eines Gewinnes im positiven Fall. Was als Schaden oder Nutzen angesehen wird, hängt von den Wert- und Zielvorstellungen ab (Pietrowski & Ennker, 2007).

Eine Risikosituation bietet verschiedene Handlungsalternativen mit entsprechenden Handlungszielen. Wird ein Handlungsziel nicht erreicht, hat dies für den Entscheider einen Zustand zur Folge, der unerwünschter als die

Ausgangssituation ist. Das Verhalten in Risikosituationen wird als Risikoverhalten bezeichnet (Dorsch, 1976).

Der Begriff „Entscheidung“ bedeutet sich zwischen mindestens zwei Optionen präferentiell festzulegen, also eine Option gegenüber einer oder mehreren anderen vorzuziehen (Jungermann, Pfister & Fischer, 2010).

In Entscheidungssituationen existiert keine eindeutig richtige Lösung, es gilt aus mehreren Optionen die, nach Möglichkeit beste auszuwählen (Kirchler & Schrott, 2005).

Entscheidungssituationen lassen sich in vielfacher Hinsicht wie z.B. in Entscheidungen unter Sicherheit und Entscheidungen unter Unsicherheit bzw. mit Risiko unterteilen (Jungermann et al., 2010; Lerch, Rauch & Schlesier, 2000). Wenn bei Entscheidungen die Alternativen und ihre mit Sicherheit eintretenden Folgen bekannt sind, liegen sichere Entscheidungen vor. Wenn bestimmte Konsequenzen mit bestimmter Wahrscheinlichkeit eintreten, spricht man von riskanten Entscheidungen. Risiko ist das Ausmaß der Unsicherheit, mit dem eine Entscheidung und ihre Folgen verbunden sind. Bei einer Zwischenform von Entscheidungen, den ambigen Entscheidungen, liegen vage Vermutungen über Konsequenzen und deren Wahrscheinlichkeiten vor (Kirchler & Schrott, 2005).

Hier ein Beispiel um Entscheidungen unter Sicherheit und unter Unsicherheit zu verdeutlichen. Eine Person die mit der S-Bahn fährt und dabei keine Fahrkarte löst obwohl sie weiß, dass in dieser S-Bahn immer Fahrscheinkontrollen durchgeführt werden weiß auch, dass dieses Verhalten dazu führt, dass sie eine Strafe bezahlen muss. Es liegt hier eine sichere Entscheidung vor, bei der eine bestimmte Konsequenz mit Sicherheit eintritt. Werden hingegen in der S-Bahn nur manchmal die Fahrkarten kontrolliert, so bleibt offen ob die Person tatsächlich überprüft wird. Die Entscheidung keine Fahrkarte zu lösen und die damit verbundene Konsequenz einer Strafzahlung muss nicht eintreten. Es liegt hier eine riskante Entscheidung bzw. eine Entscheidung unter Unsicherheit vor.

2.4.2 Entscheidungsmodelle

Um Entscheidungen unter Risiko bzw. Unsicherheit zu erklären, werden zwei wesentliche Modelle angeführt.

Das erste deskriptive Modell zur Entscheidung unter Unsicherheit, das SEU-Modell (Subjectively Expected Utility-Modell), wurde im Jahre 1954 von Edwards vorgeschlagen. Dem SEU-Modell nach, wird eine Person jene Option wählen, die den größten subjektiv erwarteten Nutzen hat. Die wichtigste Revision des SEU-Modells, stellt die Prospect Theorie von Kahneman und Tversky aus dem Jahre 1979, dar. Die Prospect Theorie zeichnet sich dadurch aus, dass im Entscheidungsprozess zwischen zwei unsicheren Optionen, in zwei Phasen unterschieden wird. Während in der ersten Phase das gegebene Problem editiert wird, werden in der zweiten Phase die editierten Optionen evaluiert. Das gegebene Problem wird also erst nach bestimmten Regeln enkodiert, transformiert und mental repräsentiert, und anschließend wird für jede Option der subjektive Wert bestimmt und eine Option gewählt. Der Prospect Theorie nach, sind die Konsequenzen einer Option und auch deren Wahrscheinlichkeiten ausschlaggebend dafür, wie die Optionen bewertet werden und determinieren in weiterer Folge somit auch die Entscheidung (Jungermann et al., 2010).

Um Konsequenzen und deren Wahrscheinlichkeiten in subjektive Größen umzuwandeln, gibt es in der Prospect Theorie unterschiedliche Annahmen. Es wird angenommen, dass die Konsequenzen einer Option subjektiv analysiert werden, sie werden nicht als absolute Größen, sondern relativ zu einem Referenzpunkt bewertet. Beispielsweise wird einer Person, die davon überzeugt ist in der Statistikprüfung die Note 1 schaffen zu können, eine 3 eventuell nicht viel wert sein. Glaubt die Person allerdings in der Diagnostikprüfung sei eine 4 die beste Note die sie erreichen könnte, dann ist sie höchstwahrscheinlich über eine 3 besonders glücklich. Abhängig vom Bezugspunkt (Note 1 oder Note 4) kann also die gleiche Konsequenz (Note 3) einen anderen subjektiven Wert haben. Konsequenzen oberhalb des

Referenzpunktes werden als Gewinne, unterhalb des Referenzpunktes als Verluste angesehen (Jungermann et al., 2010).

Die Wertfunktion verläuft über Gewinne konkav und über Verluste konvex. Die konkave Funktion erläutert, warum man bei Gewinnchancen eher risikoscheu bzw. risikoavers entscheidet. Der subjektive Wert steigt rascher an als der Gewinn. Die Funktion steigt bei geringen Beträgen steil an und flacht dann ab. Hat man beispielsweise die Wahl zwischen 80 Euro sicherem Gewinn und einer Wahrscheinlichkeit von 0,85 bei einer Lotterie 100 Euro zu gewinnen, so ist der subjektive Wert der Lotterie geringer als der subjektive Wert eines Gewinnes von 80 Euro, und man bevorzugt diesen niedrigeren aber sicheren Gewinn.

Die konvexe Funktion erläutert, warum man bei drohenden Verlusten eher risikofreudig bzw. risikosuchend entscheidet. Wenn Unterschiede zwischen geringen Verlusten vorliegen, werden diese subjektiv hoch bewertet. Bei kleinen Beträgen fällt die Kurve steil ab. Weitere Verluste werden als weniger schlimm eingestuft und auch hier flacht sich die Kurve dann ab. Hat man beispielsweise die Wahl zwischen 80 Euro sicherem Verlust und einer Wahrscheinlichkeit von 0,85 bei einer Lotterie 100 zu verlieren, so ist der subjektive Wert der Lotterie größer bzw. weniger negativ, als der subjektive Wert eines sicheren Verlustes. Man entscheidet sich bevorzugt für die Lotterie (Jungermann et al., 2010).

Außerdem ist die Wertfunktion für Verluste steiler als für Gewinne, verliert man 100 Euro so ist der Schmerz größer als die Freude, wenn man 100 Euro gewinnt (Jungermann, 2005; Jungermann et al., 2010).

Eine zweite Funktion in der Prospect Theorie ist die Entscheidungsgewichtungsfunktion. Sie ist eine weitere Modifikation des SEU-Modells und stellt eine Annahme zur Transformation der Wahrscheinlichkeiten dar. Sie gibt an, welche Gewichtung jemand einer Wahrscheinlichkeit bei der Entscheidung gibt. Häufig wird kleinen Wahrscheinlichkeiten, wie etwa der Wahrscheinlichkeit eines Unfalls in einem Kernkraftwerk, ein übermäßig großes Gewicht zugeschrieben (Jungermann et al., 2010).

Die Prospect Theorie ist eine der wichtigsten Theorien, die die Entscheidungsfindung in Situationen der Unsicherheit beschreibt (Jungermann et al., 2010).

2.4.3 Unterschiede im Risikoverhalten

Das Antwortformat von Aufgaben spielt bei kognitiven Leistungen eine Rolle. Befunde sprechen dafür, dass Frauen bessere Leistungen bei Tests mit freiem Antwortformat und Männern bessere Ergebnisse bei Multiple-Choice-Aufgaben erbringen (Willingham & Cole, 1997). Auch Bridgeman und Morgan (1996) bestätigten diese Annahme.

Ebenso konnte DeMars (1998) unter jenen StudienteilnehmerInnen mit höchstem Fähigkeitsniveau erkennen, dass im Bereich Mathematik und Naturwissenschaften die männlichen Studierenden bessere Leistungen bei Multiple-Choice-Aufgaben erzielten. Die weiblichen Studierenden schnitten bei Aufgaben mit freiem Antwortformat besser ab.

Bei der geschlechterspezifischen Analyse der TIMSS, welche Ergebnisse von SchülerInnen der achten Klasse Mittelstufe im internationalen Vergleich beschreibt, wurde auch nach der Antwortart analysiert. Bei den drei in der TIMSS verwendeten Antworttypen, Multiple-Choice, kurze und lange freie Antwort zeigten sich im internationalen Vergleich Geschlechterunterschiede. Die Jungen erbrachten bei Multiple-Choice-Items deutlich bessere Leistungen, bei kurzen und langen freien Antworten verminderten sich die Unterschiede. Im Großteil der untersuchten Nationen sanken die Unterschiede bei langen freien Antwortformaten im Vergleich zu Multiple-Choice-Items zugunsten der Mädchen (Kaiser & Steisel, 2000).

In einer Studie von Ben-Shakhar und Sinai (1991), welche sich ebenfalls mit Geschlechterunterschieden beim Antwortverhalten in Multiple-Choice-Tests beschäftigte, ist die Tendenz zu raten sowie die Tendenz Items auszulassen wesentlich. Während die einen TestteilnehmerInnen ermutigt wurden zu raten, wenn sie die Antwort nicht sicher wissen, wurde bei den anderen die Option zu

raten nicht erwähnt. Es zeigte sich bei den Frauen eine höhere Rate an ausgelassenen Items, sogar dann, wenn in der Instruktion auf die Möglichkeit zu raten hingewiesen wurde. Die männlichen Teilnehmer neigten weniger dazu Items auszulassen. In den kognitiven Aufgaben die einen höheren männlichen Vorteil zeigten, ließen sich größer Unterschiede feststellen. Dennoch kann der männliche Vorteil nicht gänzlich die Geschlechterdifferenzen bei den ausgelassenen Items erklären, da Männer auch bei kognitiven Aufgaben die einen weiblichen Vorteil zeigten, weniger dazu tendierten Items auszulassen. Die Untersuchung zeigte, dass selbst die explizite Ermutigung zum Raten nicht ausreichend ist, um Unterschiede in der Ratetendenz zu beseitigen (Ben-Shakhar & Sinai, 1991).

Mäkitalo (1993, zitiert nach Åberg-Bengtsson, 1999, S. 566) konnte in ihrer Studie zum SweSAT (Swedish Scholastic Assessment Test), einem Eignungstest an Colleges, für Frauen einen stärkeren Zusammenhang zwischen dem Auslassen von Items und der Position der Items, als bei Männer erkennen. Mädchen ließen mehr Items aus, vor allem gegen Ende des Tests, was auf eine negative Beeinträchtigung durch Zeitmangel hindeutet. Der Autorin nach ist dieser Zeitmangel (Speed-Komponente) für die geschlechterspezifischen Leistungsunterschiede verantwortlich. Sie merkte diesbezüglich an, dass es bei Zeitmangel für Frauen nachteilig sein kann, dass sie gewissenhafter arbeiten und weniger zum Raten neigen als Männer.

Belege sprechen dafür, dass das Streichen der Zeitbeschränkung die Testleistungen von Frauen verbessern kann (Rosser, 1989; zitiert nach Forbes, 1996, S. 85).

Des Weiteren wird herkömmlicherweise angenommen, dass bei einer Person die bei einer Prüfung richtig antwortet, gelehrtes Wissen vorhanden ist. Für einen Prüfling der einen Zufallstreffer macht, wird somit angenommen, dass er genauso viel weiß wie jemand der richtig antwortet und sich sicher ist (Koivula, Hassmén & Hunt, 2001). Koivula et al. (2001) untersuchten zwei Gruppen, mit insgesamt 550 TeilnehmerInnen, die alle die gleichen Multiple-Choice-Fragen des SweSAT beantworten mussten. Nur eine Gruppe musste zusätzlich nach

jeder Frage auf einer fünfstufigen Skala angeben, wie sicher sie sich bezüglich der Richtigkeit ihrer Antwort waren. Es konnten signifikante Unterschiede beim Subtest quantitatives Schlussfolgern, welcher Problemlösen sowie numerische und schlussfolgernde Fähigkeiten messen soll, beobachtet werden. Frauen erbrachten darin niedrigere Leistungen als Männer. Die Personengruppe die Angaben zu ihrer Antwortsicherheit machte, übertrafen im quantitativen Subtest leistungsmäßig jene die keine Angaben dazu machten. Vor allem Personen die sich Eigenschaften die traditionell als männlich gelten weniger zuschrieben (gemessen mittels Bem Sex Role Inventory), profitierten von diesem Prozess. Frauen die sich den Geschlechterstereotypen entsprechend weibliche Eigenschaften zuschrieben und Männer die sich weibliche Eigenschaften beimaßen, wie auch Personen die sich sowohl weibliche als auch männliche Eigenschaften zuschrieben, erbrachten bessere Leistungen wenn sie Angaben zur Antwortsicherheit machten, als ohne diese Angaben. In Bezug auf die Richtigkeit ihrer Antworten und der zutreffenden Einschätzung ihrer Antwortsicherheit waren Frauen fehlerfreier. Bei falschen Antworten im Untertest quantitatives Schlussfolgern, gaben Männer die sich maskuline Eigenschaften zuschrieben, bei ihrer Antwortsicherheit die Kategorie „extrem sicher“ am häufigsten an. Die Selbsteinschätzung der Antwortsicherheit bietet zusätzliche Informationen um zwischen jenen Personen die das Sachgebiet kennen und jenen die raten zu unterscheiden. Außerdem bietet sie eine Möglichkeit um die Wirkung von Geschlechterstereotypisierung auf die Leistung zu reduzieren.

In einer Untersuchung von Steele und Aronson (1995) bezüglich der Auswirkung von Stereotype Threat auf kognitive Leistungsunterschiede zwischen weißen und afroamerikanischen Studierenden, zeigten sich Unterschiede beim Raten bzw. Risikoverhalten. AfroamerikanerInnen machten weniger Rateversuche, wenn sie vor der Testung in einem Fragebogen Angaben zu ihrer Rasse machen sollten als afroamerikanische StudentInnen bei denen die Rasse nicht erfragt wurde. Sie neigten in dieser Bedingung also dazu weniger riskant zu arbeiten. Hingegen wenn weiße TeilnehmerInnen ihre Rasse angeben sollten neigten sie mehr dazu zu raten.

2.4.4 Entscheidungen unter Risiko im Game of Dice Task (GDT)

Der Game of Dice Task (GDT) wurde entwickelt um Entscheidungen unter Risiko zu untersuchen (Brand, Fujiwara et al., 2005). Bei dem computerisierten Verfahren können die Testpersonen fiktives Geld gewinnen oder verlieren, indem sie aus mehreren Alternativen eine auswählen und damit angeben welche Zahl sie glauben, dass als nächstes gewürfelt wird. Für einen näheren Einblick in den Ablauf des Verfahrens siehe Kapitel „4.1.2 Erhebungsinstrumente“.

Untersuchungen in unterschiedlichen klinischen Populationen zeigten unter Anwendung des GDT, welcher explizite Regeln für Gewinne und Verluste angibt, Zusammenhänge zwischen exekutiven Funktionen und dem Entscheidungsprozess (Brand, Fujiwara et al., 2005).

In einer Arbeit von Brand, Kalbe et al. (2005) wiesen pathologische Glücksspieler bei der Entscheidungsfindung in Risikosituationen mit expliziten Regeln für Gewinn und Verlust, wie sie der GDT bietet, Beeinträchtigungen auf. Es wurden männliche pathologische Spieler und gesunde Männer mittels GDT in einer Glücksspielsituation untersucht. Die Spielsuchtpatienten deuteten aufgrund ihrer Ergebnisse auf ausgeprägte Defizite im GDT, welche sich durch riskante Entscheidungen widerspiegeln, hin. Sie zeigten signifikante Präferenzen für nicht vorteilhafte Entscheidungen. Ein Zusammenhang zwischen der Häufigkeit der riskanten Entscheidungen und den exekutiven Funktionen wie etwa der Kategorisierung von Optionen oder der kognitiven Flexibilität ließ sich erkennen.

Exekutive Funktionen spielen hier eine wichtige Rolle. Beispielsweise müssen um vorteilhafte und nachteilige Alternativen zu erkennen, die verschiedenen offensichtlichen Gewinnwahrscheinlichkeiten der Entscheidungen und die damit verbundenen Geldbeträge kategorisiert und verglichen werden (Brand, Fujiwara et al., 2005).

Auch Morbus Korsakow-PatientInnen zeigten starke Beeinträchtigungen im Entscheidungsverhalten auf und diese stehen in Zusammenhang mit

spezifischen exekutiven Funktionen, wie jener der Kategorisierung (Brand, Fujiwara et al., 2005). Weiters zeigten sich auch bei Morbus Parkinson-PatientInnen diese Beeinträchtigungen im Entscheidungsprozess und es wurde auch bei ihnen ein Zusammenhang zwischen exekutiven Dysfunktionen und schwacher Leistung im GDT beobachtet (Euteneuer et al., 2009).

3 Zielsetzung, Fragestellung und Hypothesen

3.1 Zielsetzung und Fragestellung

Das Ziel der vorliegenden empirischen Studie ist die Untersuchung von geschlechterspezifischen Leistungsunterschieden in Bezug auf schlussfolgerndes Denken in EMS-ähnlichen Aufgaben. Neben dem Geschlecht soll auch der Einfluss der sozialen Faktoren Stereotype Threat und Risikoverhalten in Bezug auf kognitive Leistungen im schlussfolgernden Denken, untersucht werden.

3.2 Hypothesen

3.2.1 Haupthypothesen

Es werden die Untertests zum schlussfolgernden Denken, dies betrifft im Speziellen die Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* unter Berücksichtigung der Faktoren Geschlecht, Stereotype Threat-Bedingung und Risikobedingung untersucht. Zur besseren Lesbarkeit befinden sich die formulierten Hypothesen vor den jeweiligen Ergebnisdarstellungen.

3.2.2 Nebenhypothesen

Weiters werden die Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* unter Berücksichtigung der Faktoren Schulzweig und Absicht ein Medizinstudium zu beginnen sowie Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben untersucht. Der GDT wird unter Berücksichtigung der Faktoren Geschlecht und Risikobedingung untersucht. Auch hier werden die abgeleiteten Hypothesen zur besseren Lesbarkeit vor den jeweiligen Ergebnissen dargestellt.

4 Empirischer Teil

4.1 Methode

Die Medizinische Universität Wien (MUW) erteilte den Auftrag für eine Untersuchung, welche Aufschlüsse über geschlechterspezifische Unterschiede beim Eignungstest im Medizinstudium, liefern soll. Die vorliegende Arbeit behandelt einen Teil dieser breit angelegten Untersuchung. Meine Kolleginnen Katrin Anzirk, Gabriele Hangl, Helene Lager und Anita Teufl behandeln in ihren Arbeiten, welche ebenfalls den gleichen Datensatz heranziehen, weitere Teilbereiche der Studie.

Im Folgenden werden der Untersuchungsplan sowie die verwendeten Erhebungsinstrumente beschrieben.

4.1.1 Untersuchungsplan und intendierte Stichprobe

Mittels multivariatem Ansatz sollten biologische Faktoren, wie der Hormonstatus und soziale Faktoren, wie Stereotype Threat und Risikoverhalten, Geschlechterunterschiede in kognitiven Leistungen, ähnlich den EMS-Aufgaben, und in sozial-kognitiven Aufgaben zur Theory of Mind, aufdecken. Hierbei sollten auch Mediatorvariablen wie Leistungsmotivation, Selbstwirksamkeit und Ängstlichkeit berücksichtigt werden.

Die Methodik sah ein experimentelles multivariates Untersuchungsdesign vor, bei dem die Zuweisung der Versuchspersonen zu den Versuchsbedingungen (Stereotype Threat und Risikoverhalten) randomisiert erfolgen sollte.

Die Untersuchung besteht aus 24 verschiedenen Versuchsbedingungen. Diese setzen sich aus den Faktorstufenkombinationen Geschlecht (2), Stereotype Threat-Bedingung (4) und Risikobedingung (3) zusammen. Tabelle 2 gibt einen Überblick über die 24 realisierten Versuchsbedingungen.

Tabelle 2: Geplantes Untersuchungsdesign

Variablen	Versuchsbedingungen															
Hormonmessung 1																
Stereotype Threat	A			B			C			D						
Risikoverhalten	1	2	3	1	2	3	1	2	3	1	2	3				
Geschlecht	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀	♂
EMS-ähn. Tests																
Theory of Mind																
Hormonmessung 2																
Ängstlichkeit																
Selbstwirksamkeit																
Leistungsmotivation																

Sowohl vor, als auch nach der Bearbeitung der EMS-ähnlichen Aufgaben, wurde eine Hormonmessung mittels Speichelproben durchgeführt. Diese soll Hormone wie Cortisol, Dehydroepiandrosteron, Testosteron, Dopamin, Chromogranin A, Östradiol und Progesteron erheben.

Untersucht werden sollten etwa 1200 SchülerInnen aus den 8. Klassen Allgemeinbildender Höherer Schulen in Wien, wobei die GymnasiastInnen aus verschiedenen Oberstufenschulzweigen, wie beispielsweise naturwissenschaftlicher, humanistischer oder sprachlicher Ausbildungszweig, stammen sollten. Die Untersuchung sollte in zwei Unterrichtsstunden, etwa 100 Minuten, in Gruppentestungen durchgeführt werden.

Beginnend mit Oktober 2009 sollte die Untersuchung mit Juni 2010 abgeschlossen sein. In der zeitlichen Planung wurde die Begutachtung der Ethikkommission der Medizinischen Universität Wien, die Genehmigung zur Untersuchungsdurchführung seitens Schulrat, die Datenerhebung sowie die Dateneingabe und Auswertung berücksichtigt.

4.1.1.1 Voruntersuchung

Die Wirksamkeit der Stereotype Threat-Bedingungen und Risikobedingungen, sowie der praktische und zeitliche Ablauf wurden in einer Voruntersuchung an 31 MedizinstudentInnen der Medizinischen Universität Wien, welche sich im zweiten Studienabschnitt befanden und den EMS somit bereits positiv absolviert hatten, überprüft. Infolge der Vorstudie wurden Veränderungen an der Testbatterie vorgenommen. Sie musste aus zeitlichen Gründen um den Untertest *Textverständnis* gekürzt werden. Außerdem wurde der *Wissenstest*, angelehnt an die Untersuchung von Halpern und Tan (2001), um die Selbsteinschätzungsaufgaben erweitert.

4.1.2 Erhebungsinstrumente

Die Testbatterie setzte sich nach der Überarbeitung aufgrund der Vorstudie aus folgenden Erhebungsinstrumenten zusammen:

1. Soziodemographischer Fragebogen
2. Instruktionen der Stereotype Threat-Bedingungen
3. Wissenstest
4. Instruktionen zum Risikoverhalten
5. EMS-ähnliche Aufgaben (ITB Consulting, 2008; ZTD, 2009)
6. Theory of Mind-Aufgaben (Willinger, Schmöger, Müller & Auff, in Vorbereitung)
7. Leistungsmotivationsinventar (LMI) (Schuler & Prochaska, 2001)
8. Skala zur Allgemeinen Selbstwirksamkeitserwartung (SWE) (Schwarzer & Jerusalem, 1999)
9. Game of Dice Task (GDT) (Brand, Fujiwara et al., 2005)

Die Dauer und Reihenfolge der Vorgabe gestaltete sich bei allen Testungen gleich. Die Erhebungsinstrumente wurden so wie in der hier angeführten Abfolge vorgegeben.

ad 1: Soziodemographischer Fragebogen

Zu Beginn der Testbatterie wurden soziodemographische Daten der TestteilnehmerInnen wie Alter, Geschlecht, Staatsbürgerschaft und auch Beruf und Ausbildung der Eltern, erfragt. Überdies wurde der besuchte Oberstufenschulzweig, das Wiederholen einer Klasse sowie die weiteren Ausbildungs- und Berufsziele nach der Matura erhoben. Auch nach Freizeitbeschäftigungen und Absicht, ein Medizinstudium zu beginnen, wurde gefragt, sowie bezüglich der beabsichtigten Fachrichtung nach einem Medizinstudium und der bisherigen Auseinandersetzung mit dem Aufnahmetest für das Medizinstudium.

ad 2: Instruktionen der Stereotype Threat-Bedingungen

Es wurden vier unterschiedliche Stereotype Threat-Bedingungen vorgegeben:

A) *Aktivierung* eines Stereotype Threats

Bei der Aktivierung des Stereotype Threat wurde in der schriftlichen Instruktion betont, dass es statistisch gesichert sei, dass Frauen bessere Leistungen im Gedächtnis und Männer bessere Leistungen in der räumlichen Orientierung, im Arbeiten mit Zahlen, in der Interpretation von Diagrammen und Tabellen und im naturwissenschaftlichen Wissen, erbringen.

B) *Neutrale Bedingung*

Den Versuchspersonen in der *Neutralen Bedingung* wurde mitgeteilt, dass es zwischen Männern und Frauen keine Unterschiede bezüglich der Leistung in der räumlichen Orientierung, im Arbeiten mit Zahlen, in der Interpretation von Diagrammen und Tabellen und im naturwissenschaftlichen Wissen gäbe.

C) *Aufklärung*

In der *Aufklärungsbedingung* wurden die TestteilnehmerInnen über das Phänomen *Stereotype Threat* aufgeklärt. Es wurde darauf aufmerksam gemacht, dass eine mögliche Ängstlichkeit während der Bearbeitung der nachfolgenden Aufgaben auch ein Resultat negativer *Stereotype* sein könnte.

D) *Negativaktivierung* eines *Stereotype Threats*

Um eine negative Aktivierung eines *Stereotype Threat* zu bewirken, wurde entgegen der Bedingung *Aktivierung* eines *Stereotype Threats*, hier darauf hingewiesen, dass es statistisch gesichert sei, dass Männer bessere Leistungen im Gedächtnis und Frauen bessere Leistungen in der räumlichen Orientierung, im Arbeiten mit Zahlen, in der Interpretation von Diagrammen und Tabellen und im naturwissenschaftlichen Wissen, aufweisen.

ad 3: Wissenstest

Um die Wirksamkeit der *Stereotype Threat*-Bedingungen in den EMS-ähnlichen Bereichen zu überprüfen, wurde unmittelbar nach der jeweiligen *Stereotype Threat*-Instruktion, der *Wissenstest* welcher in Anlehnung an den *Beliefs About the cognitive Abilities of Females and Males Questionnaire* von Halpern und Tan (2001) entwickelt wurde, vorgegeben. Der *Wissenstest* setzt sich aus den beiden Teilen Fremd- und Selbsteinschätzungsaufgaben zusammen.

Im Fremdeinschätzungsteil wurde zu jedem der EMS-ähnlichen Bereiche, sowie zur *Theory of Mind*, eine Beispielaufgabe vorgelegt. Die Untersuchungsperson sollte angeben zu wie viel Prozent sie glaube, dass es sich bei einer Person die, die Beispielaufgabe richtig bearbeitet hatte, um einen Mann oder eine Frau handle, wobei die Prozentsumme 100 ergeben sollte. Abbildung 1 veranschaulicht hierzu ein Itembeispiel.

Denken Sie an eine Person, die Sie nicht kennen. Der Person wurden Fragen zum Umgang mit Zahlen, Größen, Einheiten und Formeln gestellt wie folgende Aufgabe:

Zwei Drittel der Fälle der erblichen Erkrankung „Down-Syndrom“ werden von der Eizelle verursacht und ein Drittel durch die Samenzelle. Erhebungen zeigten, dass unter 10 000 Neugeborenen 15 Fälle des „Down-Syndroms“ auftreten. Wie viele Down-Syndrom-Geburten sind auf eine Schädigung der Eizelle zurückzuführen?

(A) 1,5 % (B) 1 % (C) 0,5 % (D) 0,15 % (E) 0,1 %

richtige Lösung: (E)

Diese Aufgabe aus dem Bereich Arbeiten mit Zahlen wurde richtig gelöst, die Wahrscheinlichkeiten, **dass es sich um einen Mann bzw. Frau handelt, betragen:**

Mann _____% bzw. Frau _____% (Summenwert soll 100% ergeben)

Abbildung 1: Fremdeinschätzungsaufgabe zum Bereich Quantitative und formale Probleme

Bei der Selbsteinschätzung wurden die gleichen Beispielaufgaben wie im Fremdeinschätzungsteil vorgegeben. Die Testpersonen wurden instruiert, so anzukreuzen, wie sie sich selbst auf einer Skala von 1 bis 7 bezüglich der jeweiligen Fähigkeiten einschätzen (siehe Abbildung 2).

Denken Sie an folgende Aufgabe: Einer Person wurden Fragen zum Umgang mit Zahlen, Größen, Einheiten und Formeln gestellt wie folgende Aufgabe:

Zwei Drittel der Fälle der erblichen Erkrankung „Down-Syndrom“ werden von der Eizelle verursacht und ein Drittel durch die Samenzelle. Erhebungen zeigten, dass unter 10 000 Neugeborenen 15 Fälle des „Down-Syndroms“ auftreten. Wie viele Down-Syndrom-Geburten sind auf eine Schädigung der Eizelle zurückzuführen?

(A) 1,5 % (B) 1 % (C) 0,5 % (D) 0,15 % (E) 0,1 %

richtige Lösung: (E)

Ich verfüge über die Fähigkeit, die oben angeführte Aufgabe richtig zu lösen:

① ② ③ ④ ⑤ ⑥ ⑦

trifft nicht zu trifft sehr zu

Abbildung 2: Selbsteinschätzungsaufgabe zum Bereich Quantitative und formale Probleme

ad 4: Instruktionen zum Risikoverhalten

Nach der Bearbeitung des *Wissenstests* wurden den Testpersonen drei unterschiedliche Risikobedingungen zur Erfassung des Risikoverhaltens vorgegeben.

1) Instruktion *raten*

Unter dieser Risikobedingung wurden die Testpersonen dazu aufgefordert, bei den folgenden Aufgaben zu raten, wenn sie die richtige Lösung nicht wissen.

2) *ohne Instruktion*

Hier wurden die Testpersonen aufgefordert, die folgenden Aufgaben aufmerksam und konzentriert zu bearbeiten.

3) Instruktion *sicher*

In der Risikobedingung *sicher* wurden die Testpersonen dazu aufgefordert, nur dann eine Antwort anzukreuzen, wenn sie sich ganz sicher sind.

Wesentlich ist die Tatsache, dass in der vorliegenden Untersuchung den Testpersonen die jeweilige Stereotype Threat-Bedingung sowie auch die jeweilige Risikobedingung immer wieder dargeboten wurden. Die Bedingungen wurden in schriftlicher Form, zwischen den EMS-ähnlichen Untertests immer wieder und erneut in Erinnerung gerufen.

ad 5: EMS-ähnliche Aufgaben

Der *Eignungstest für das Medizinstudium* setzt sich aus 10 Untertests zusammen. Die Zulassung zum Medizinstudium erfolgt aufgrund des Testwertes, der sich aus der Gesamtpunktzahl berechnen lässt. Es steht für jeden Untertest nur eine begrenzte Bearbeitungszeit zur Verfügung. Der EMS

ist als „Speed- und Power-Test“ zu sehen, d.h. er stellt anspruchsvolle Leistungsanforderungen unter Zeitdruck (Kubinger, 2006). Die Aufgaben der Untertests sind nach dem Multiple-Choice-Prinzip konzipiert, d.h. zu jeder Aufgabe werden fünf Antwortvorschläge geboten, wovon nur einer richtig ist.

In diese Untersuchung wurden nur sechs der zehn Untertests aufgenommen, wobei von diesen sechs Untertests die Bereiche *Quantitative und formale Probleme*, *Schlauchfiguren*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* bei vorangegangenen EMS-Testungen signifikante Leistungsunterschiede zugunsten der Männer aufzeigten. In den Subtests *Figuren Lernen* und *Fakten Lernen* ließen sich signifikant bessere Leistungen zugunsten der Frauen finden.

Um im Folgenden die verwendeten Untertests kurz vorzustellen, wird auf den Bericht „EMS-Eignungstest für das Medizinstudium 2007“ von Hänsgen und Spicher (2007) sowie das Übungsheft „TestInfo'09“ (ZTD, 2009) Bezug genommen.

■ **Untertest *Figuren Lernen***

In diesem Untertest wird überprüft, wie gut sich die Testperson Einzelheiten von Gegenständen einprägen und merken kann.

Der Untertest setzt sich aus einer Lern- und einer Reproduktionsphase zusammen. In der Lernphase hat die Testperson zwei Minuten Zeit um sich acht Figuren und ihre jeweiligen schwarzen Flächen einzuprägen (siehe Abbildung 3, links). Nach etwa 50 Minuten findet die Reproduktionsphase statt, in der die Testperson zwei Minuten Zeit hat, um anzugeben, welche Flächen der Figuren geschwärzt waren (siehe Abbildung 3, rechts).

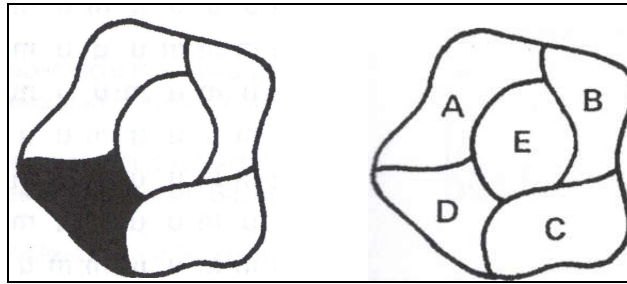


Abbildung 3: Beispielaufgabe zum Untertest Figuren Lernen (ZTD, 2009)

■ Untertest *Fakten Lernen*

Der Untertest überprüft, die Fähigkeit sich Fakten einprägen und merken zu können. Es werden acht Patienten vorgestellt, deren Name, Alter, Beruf, Geschlecht, Diagnose und ein zusätzliches Beschreibungsmerkmal, sich die Testperson einprägen soll, später werden Fragen dazu gestellt. Auch hier sind eine Einpräge- und eine Reproduktionsphase vorgesehen, welche jeweils drei Minuten dauern. In den dazwischen liegenden 50 Minuten werden die Aufgaben der Untertests *Quantitative und formale Probleme*, *Schlauchfiguren*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen*, bearbeitet. Eine mögliche Patientenbeschreibung und bezugnehmende Frage dazu wird in Abbildung 4 angeführt.

Fischer:	ca. 25 Jahre,	med.-tech. Assistentin, alleinerziehend -Karies
Die Diagnose für die med.-techn. Assistentin lautet:		
(A)	Herzinfarkt.	
(B)	Bluthochdruck.	
(C)	Karies.	
(D)	Asthmaanfall.	
(E)	Magengeschwür.	

Abbildung 4: Beispielaufgabe zum Untertest Fakten Lernen (ZTD, 2009)

Einzelheiten zu den Ergebnissen der Untertests *Figuren Lernen* und *Fakten Lernen* können der Arbeit von Helene Lager entnommen werden.

■ Untertest *Quantitative und formale Probleme*

In diesem Subtest wird der richtige Umgang mit Zahlen, Größen, Einheiten und Formeln im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen, geprüft. Algebraisches Umformen, rechnen sowie logisches Kombinieren ist hier notwendig; wichtiger ist allerdings, den naturwissenschaftlichen Sachverhalt zu verstehen und den korrekten Lösungsansatz zu finden, als zu rechnen und umzuformen. Es werden hier zwar keine Lerninhalte, aber Kenntnisse der Mittelstufen-Mathematik vorausgesetzt. Die Testpersonen hatten bei diesem Untertest, für die Bearbeitung von 16 Aufgaben, eine Zeitvorgabe von 10 Minuten. Abbildung 5 zeigt ein Itembeispiel zum Untertest *Quantitative und formale Probleme*.

Ein physikalisches Gesetz ermöglicht die Bestimmung einer Kraft F aus den Größen

Geschwindigkeit	v	(Einheit: 1 m/s),
Masse	m	(Einheit: 1 kg),
Radius	r	(Einheit: 1 m).

Bei Anwendung welcher der folgenden fünf Formeln ergibt sich für F die Einheit $1 \text{ kg} \cdot \text{m/s}^2$?

(A) $F = v^2 \cdot r / m$
(B) $F = r \cdot \sqrt{v \cdot m}$
(C) $F = m \cdot v^2 / r$
(D) $F = m^2 \cdot v \cdot r$
(E) $F = r \cdot v^2 \cdot m$

Abbildung 5: Beispielaufgabe zum Untertest *Quantitative und formale Probleme* (ZTD, 2009)

■ **Untertest *Schlauchfiguren*** (Stumpf & Fay, 1983)

Das räumliche Vorstellungsvermögen ist jene Fähigkeit, die mittels Untertest *Schlauchfiguren* erhoben wird. Die 21 Aufgaben, für die eine Bearbeitungszeit von 12 Minuten vorgesehen ist, bestehen jeweils aus zwei Abbildungen, welche durchsichtige Würfel zeigen. Während die linke Abbildung den Würfel immer aus der Vorderansicht zeigt (siehe Abbildung 6), soll die Testperson angeben, aus welcher Ansicht die rechte Abbildung den Würfel darstellt, von rechts (r), links (l), unten (u), oben (o) oder hinten (h).

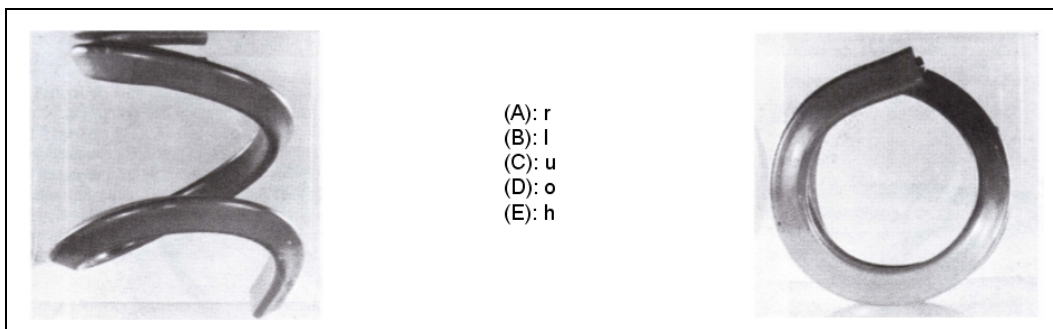


Abbildung 6: Beispielaufgabe zum Untertest Schlauchfiguren (ZTD, 2009)

■ **Untertest *Medizinisch-naturwissenschaftliches Grundverständnis***

Mit diesem Untertest wird das Verständnis für Fragen der Medizin und Naturwissenschaft überprüft. Wesentlich ist hier, die geschilderten Sachverhalte nachzuvollziehen und Schlussfolgerungen aus den im Text enthaltenen Informationen ziehen zu können. Es werden keine speziellen Sachkenntnisse vorausgesetzt. Bei den 16 vorgegebenen Aufgaben werden bezüglich des Sachverhalts je drei bis fünf Behauptungen aufgestellt und die Testperson soll angeben, ob diese Aussagen, aus den im Text enthaltenen Informationen abzuleiten sind. Für die 16 Aufgaben werden zehn Minuten Zeit geboten. Eine Beispielaufgabe dazu zeigt Abbildung 7.

Reize, die von außen auf die Haut einwirken, werden in speziellen Sinnesorganen der Haut in bioelektrische Impulse umgewandelt. Die so erzeugten Impulse laufen über die afferenten (zuführenden) Nervenfasern und die sog. Hinterwurzeln des Rückenmarks ins Rückenmark, wo sie auf andere Nervenzellen umgeschaltet werden. Sie können nun über motorische Nervenzellen Reflexbewegungen auslösen; sie können aber auch über aufsteigende Leitungen nach mehrfacher Umschaltung zur Hirnrinde gelangen, wo sie weiterverarbeitet werden und ein bewusstes Wahrnehmen bzw. Erkennen der Reize ermöglichen.

Bei einem Patienten sind die Hinterwurzeln des Rückenmarks durchtrennt. Welche(r) der folgenden Ausfälle ist bzw. sind diesen Informationen zufolge zu erwarten?

- I. In den Sinnesorganen der Haut werden keine bioelektrischen Impulse gebildet werden.
 - II. Reflexbewegungen lassen sich nicht mehr durch Hautreizung auslösen.
 - III. Hautreize können nicht mehr bewusst wahrgenommen bzw. erkannt werden.
-
- (A) Nur Ausfall I ist zu erwarten.
 - (B) Nur die Ausfälle II und III sind zu erwarten.
 - (C) Nur Ausfall III ist zu erwarten.
 - (D) Nur die Ausfälle I und III sind zu erwarten.
 - (E) Nur Ausfall II ist zu erwarten.

Abbildung 7: Beispielaufgabe zum Untertest Medizinisch-naturwissenschaftliches Grundverständnis (ZTD, 2009)

■ Untertest *Diagramme und Tabellen*

In diesem Subtest wird überprüft, wie gut eine Testperson Diagramme und Tabellen richtig analysieren und interpretieren kann. Spezielle naturwissenschaftliche, medizinische oder statistische Kenntnisse werden für die Bearbeitung der Aufgaben nicht vorausgesetzt. Allein durch die graphischen bzw. tabellarischen Informationen, sowie auch aus dem jeweiligen Text, lassen sich die richtigen Antworten ableiten. Ein Itembeispiel zum Untertest *Diagramme und Tabellen* präsentiert Abbildung 8. Die Zeitvorgabe beträgt bei 16 vorgegeben Items, zehn Minuten.

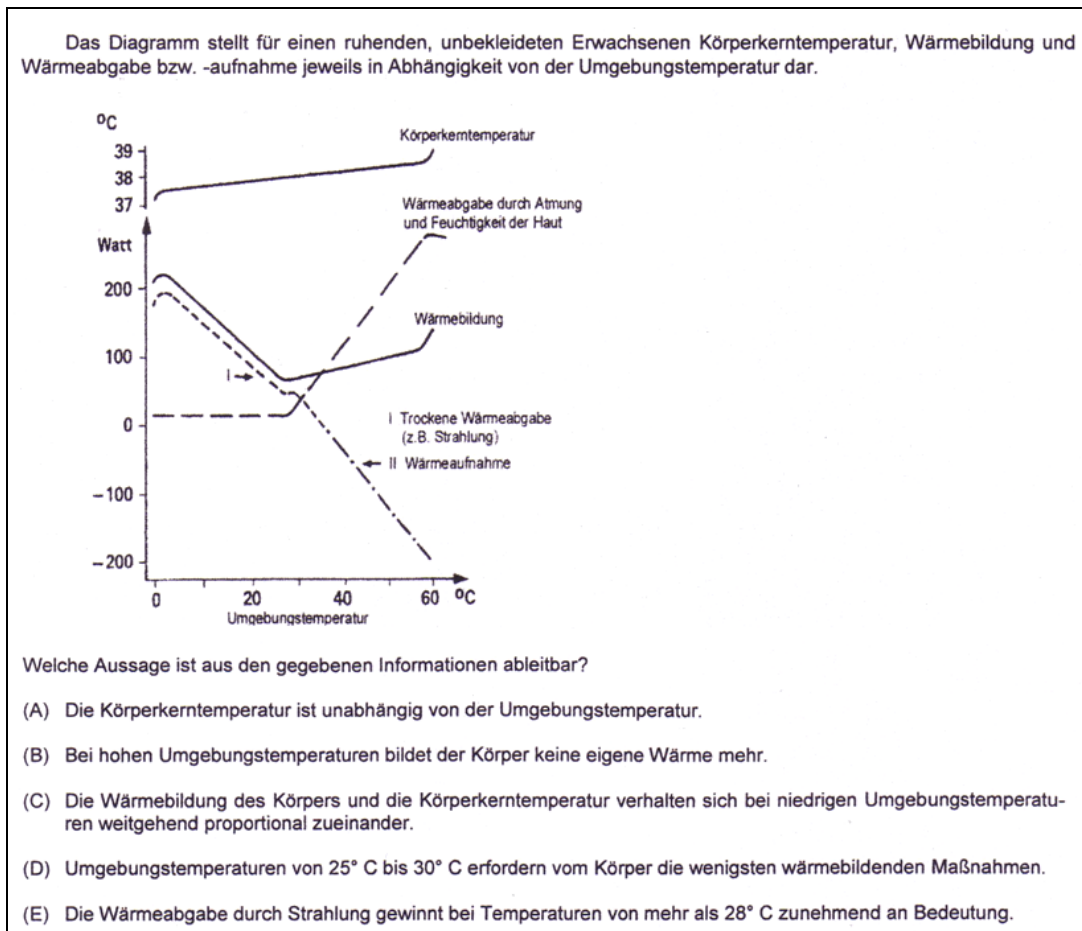


Abbildung 8: Beispielaufgabe zum Untertest Diagramme und Tabellen (ZTD, 2009)

ad 6: Theory of Mind-Aufgaben (Willinger et al., in Vorbereitung)

Unter „Theory of Mind“ ist die Fähigkeit zu verstehen, mentale Zustände anderer Personen zu repräsentieren. Dazu gehört es, nachvollziehen zu können, was eine andere Person weiß oder nicht weiß, oder welche Motive eine andere Person dazu bringen, etwas zu tun (Pauen & Vonderlin, 2009). Um die Theory of Mind zu überprüfen, wurden den Testpersonen drei Geschichten mit anschließenden Fragen dazu vorgegeben, welche die Theory of Mind im Sinne einer „False-Belief“ erster, zweiter und dritter Ordnung erfassen (Willinger et al., in Vorbereitung).

Eine False-Belief erster Ordnung erfasst fälschliche Annahmen über einen Zustand. Eine False-Belief zweiter Ordnung erfasst eine fälschliche Annahme über die Annahmen einer anderen Person (Rollett & Kastner-Koller, 2007). Bei

False-Belief dritter Ordnung steigt die Komplexität der Aufgaben. In Hinblick auf die Theory of Mind-Aufgaben, welche zwei- bis vierkategorial gestaltet sind, ist immer nur eine Antwortmöglichkeit *richtig*.

Details zu den Ergebnissen der Theory of Mind-Stories können der Arbeit von Katrin Anzirk entnommen werden.

ad 7: Leistungsmotivationsinventar (LMI) (Schuler & Prochaska, 2001)

Ein Verhalten gilt nur dann als leistungsmotiviert, wenn es auf die Selbstbewertung eigener Tüchtigkeit zielt, und zwar in Auseinandersetzung mit einem Gütemaßstab, das erreicht oder übertroffen werden soll (Rheinberg, 2008).

Das Leistungsmotivationsinventar umfasst 17 berufserfolgsrelevante Dimensionen, welche mit insgesamt 170 Items erhoben werden. Bei der vorliegenden Untersuchung wurden nur die beiden Dimensionen Internalität und Wettbewerbsorientierung erhoben. Internalität meint hier die Überzeugung, dass Handlungsergebnisse tendenziell eher als selbstverursacht und selbstverantwortet erlebt werden als, dass sie externalen Ursachen zugeschrieben werden. Mittels Wettbewerbsorientierung wird die Tendenz erfasst, inwiefern Konkurrenz als Ansporn und Motivation für berufliche Leistung erlebt wird (Prochaska, 1998).

Die Durchführungs-, Auswertungs- und Interpretationsobjektivität ist gegeben. Die Reliabilität wird mit $r = .68$ bis $r = .86$ (Cronbach's Alpha) für die Dimensionswerte angegeben (Schuler & Prochaska, 2001).

Den Testpersonen wurden im Rahmen der Untersuchung je Dimension zehn Aussagen vorgegeben, welche sie auf einer Skala von 1 (trifft gar nicht zu) bis 7 (trifft vollständig zu) bewerten sollen. Abbildung 9 zeigt ein Itembeispiel aus der Dimension Internalität.

Wie weit man es beruflich bringt, ist zu einem guten Teil Glückssache.	trifft gar nicht zu ①-②-③-④-⑤-⑥-⑦ trifft vollständig zu
--	---

Abbildung 9: Beispielaufgabe aus der Dimension Internalität - LMI (Schuler & Prochaska, 2001)

Details zu den Resultaten des Leistungsmotivationsinventars können in der Arbeit von Gabriele Hangl eingesehen werden.

ad 8: Skala zur Allgemeinen Selbstwirksamkeitserwartung (SWE)
(Schwarzer & Jerusalem, 1999)

Selbstwirksamkeit drückt die subjektive Überzeugung aus, aufgrund eigener Kompetenzen neue oder schwierige Anforderungen bewältigen zu können (Schwarzer, 2004, S. 12). Das erfolgreiche Meistern von schwierigen Lagen wird auf die eigenen Kompetenzen zugeschrieben. Die Skala zur Allgemeinen Selbstwirksamkeitserwartung erfasst mit zehn Items die optimistische Kompetenzerwartung (Schwarzer & Jerusalem, 1999).

Die Durchführungs-, Auswertungs- und Interpretationsobjektivität ist bei der Befolgung der Testinstruktion gesichert. Die Reliabilität streut mit einer internen Konsistenz (Cronbach's Alpha) zwischen $r = .76$ und $r = .90$, bei deutschen Stichproben sogar zwischen $r = .80$ und $r = .90$.

Aufgrund einiger Korrelationsbefunde mit anderen Variablen (z.B. Arbeitszufriedenheit, Angst, Depression) kann die kriterienbezogene Validität angenommen werden (Schwarzer & Jerusalem, 1999).

Die vorgegebenen zehn Aufgaben werden auf einer Skala von 1 bis 4 („stimmt nicht“ bis „stimmt genau“) beantwortet. In Abbildung 10 wird eine Beispielaufgabe aus der Skala zur Allgemeinen Selbstwirksamkeitserwartung veranschaulicht.

	Diese Aussage...			
	stimmt nicht	stimmt kaum	stimmt eher	stimmt genau
Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.	1	2	3	4

Abbildung 10: Beispielaufgabe aus der SWE (Schwarzer & Jerusalem, 1999)

Einzelheit bezüglich der Ergebnisse der Skala zur Allgemeinen Selbstwirksamkeitserwartung können der Arbeit von Anita Teufl entnommen werden.

ad 9: Game of Dice Task (GDT) (Brand, Fujiwara et al., 2005)

Im Anschluss an die Gruppentestung mittels Papier-Bleistift-Verfahren, wurde das Computerverfahren *Game of Dice Task* (Brand, Fujiwara et al., 2005) in Einzeltestungen vorgegeben. Dieses soll die Wirksamkeit der Instruktionen zum Risikoverhalten über die EMS-ähnlichen Aufgaben hinaus überprüfen.

In einer Glücksspielsituation wird das Risikoverhalten erfasst. Ein einzelner Würfel wird 18 Mal gewürfelt. Die Testperson soll sich entscheiden und angeben, welche Zahl sie vermutet, als nächstes gewürfelt werden wird. Dazu wählt sie vor jedem Wurf eine Zahl oder Zahlenkombination aus. Stimmt die tatsächlich gewürfelte Zahl mit der gewählten Zahl oder der Zahlenkombination überein, wird ein Geldbetrag gewonnen, stimmt sie nicht überein so wird ein ebenso hoher Betrag verloren. Wird nur eine einzelne Zahl gewählt so werden 1000 € gewonnen oder verloren. Bei einer Zahlenkombination aus zwei Würfeln beträgt der Gewinn bzw. Verlust 500 €, bei einer Kombination aus drei Würfeln 200 € und bei einer Kombination aus 4 Würfeln 100 €. Die ersten beiden Alternativen sind riskanter als die letzten beiden. Der Kontostand und der aktuelle Gewinn oder Verlust wird in Anzeigefeldern veranschaulicht. Es soll möglichst viel Geld gewonnen und wenig verloren werden, trotzdem kann bei Schulden weitergespielt werden. Alle Testpersonen starten mit einem Kapital von 1000 €. Zur bildlichen Vorstellung zeigt Abbildung 11 eine Ansicht des GDT.

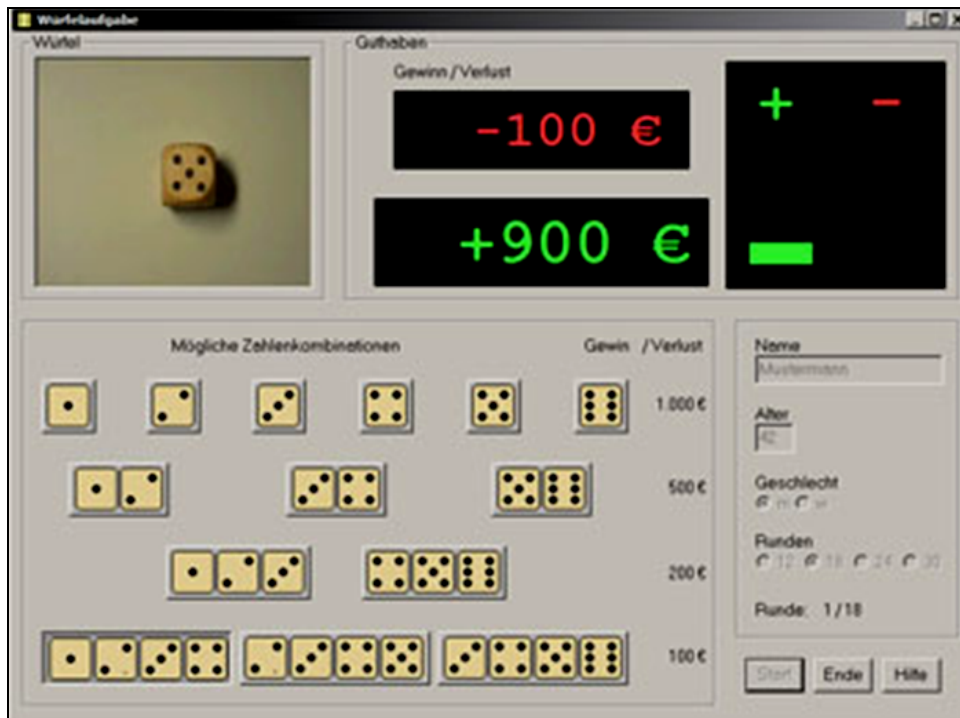


Abbildung 11: Game of Dice Task (Brand, Fujiwara et al., 2005)

4.2 Untersuchung

Die Untersuchungsdurchführung, die eingesetzten Auswertungsverfahren und eine Stichprobenbeschreibung werden im folgenden Kapitel dargestellt.

4.2.1 Untersuchungsdurchführung

Die Datenerhebung fand hauptsächlich in Wiener, aber auch in niederösterreichischen und burgenländischen AHS-Schulen statt. Da die Genehmigungen der Schulräte und die Begutachtung der Ethikkommission der Medizinischen Universität Wien abgewartet werden mussten, kam es jedoch zu Verzögerungen, wodurch die Testungen erst mit Ende März 2010 gestartet, und mit Anfang Februar 2011 abgeschlossen werden konnten. Die Testungen im Ausmaß von zwei Unterrichtseinheiten wurden in den Räumlichkeiten der Schulen abgehalten.

Den SchülerInnen wurden Einverständniserklärungen ausgehändigt, welche bei minderjährigen TeilnehmerInnen eine Zustimmung der Eltern vorsahen, und vor Beginn der Testung wieder eingesammelt wurden.

Die Hormonmessungen konnten aufgrund eingeschränkter finanzieller Ressourcen nicht bei allen SchülerInnen erfolgen und auch das computerisierte Verfahren GDT konnte angesichts des vorgegebenen Zeitrahmens nicht von allen MaturantInnen absolviert werden. Von der Erfassung der modulierenden Variable Ängstlichkeit, wurde ebenfalls aufgrund der 100-minütigen Testzeitlimitierung, abgesehen.

4.2.2 Auswertungsverfahren

Die Berechnungen und Auswertungen der Daten wurden mit der Statistiksoftware IBM Statistics SPSS 19.0 durchgeführt. Ein Ergebnis gilt als signifikant, wenn die beobachtete Irrtumswahrscheinlichkeit $p \leq .05$ beträgt. Im Rahmen der statistischen Datenanalyse wurden folgende Verfahren eingesetzt (vgl. Backhaus, Erichson, Plinke & Weiber, 2008; Bortz & Döring, 2006; Zöfel, 2003):

■ Kreuztabelle und χ^2 -Test (Chi-Quadrat-Test)

In Form von Kreuztabellen (Kontingenztafeln) werden zwei nominalskalierte Merkmale miteinander in Beziehung gebracht und anhand des χ^2 -Tests wird geprüft, ob sich die empirischen (beobachteten) Häufigkeiten von den theoretischen (erwarteten) Häufigkeiten signifikant unterscheiden bzw. ob sich die Daten der Kontingenztafel signifikant von einer Gleichverteilung unterscheiden.

■ Produkt-Moment-Korrelation nach Pearson

Die Produkt-Moment-Korrelation prüft den linearen Zusammenhang zwischen zwei intervallskalierten Variablen, wobei hier die

Normalverteilung der Daten vorausgesetzt wird. Der Korrelationskoeffizient r drückt die Stärke und Richtung des Zusammenhangs aus. Bortz und Döring (2006, S. 606) sprechen bei einem Wert von 0.10 bis 0.29 von einer geringen, zwischen 0.30 und 0.49 von einer mittleren und über 0.50 von einer hohen Korrelation.

■ **Multivariate Varianzanalyse**

Die multivariate Varianzanalyse dient zur Überprüfung von Mittelwertsunterschieden beliebig vieler unabhängiger Variablen, im Hinblick auf mehrere abhängige Variablen. Bei den abhängigen Variablen muss metrisches Skalenniveau vorliegen, während bei der bzw. den unabhängigen Variablen Nominalskalenniveau vorhanden sein muss. Beinhaltet die Berechnung eine unabhängige Variable, spricht man von einer einfaktoriellen Varianzanalyse. Bei mehreren unabhängigen Variablen spricht man von einer mehrfaktoriellen Varianzanalyse, bei der man neben den Haupteffekten auch Wechselwirkungen zwischen den Faktoren auf Signifikanz überprüfen kann. Folgende Voraussetzungen müssen für die Anwendung erfüllt werden: Homogenität der Varianzen (geprüft mittels Levene-Test), Homogenität der Varianz-Kovarianzmatrix der abhängigen Variablen (geprüft mittels Box-M-Test) und die multivariate Normalverteilung der Daten.

■ **Levene-Test**

Der Levene-Test prüft, ob die Varianzen der Stichproben signifikant verschieden sind. Fällt der Test nicht signifikant aus, so ist die Homogenität der Varianzen anzunehmen.

■ **Box-M-Test**

Der Box-M-Test überprüft die Homogenität der Varianz-Kovarianz-Matrizen der abhängigen Variablen. Bei einem nicht signifikanten Ergebnis kann die Gleichheit der Varianzen angenommen werden.

■ **Univariate Varianzanalyse**

Die univariate Varianzanalyse prüft Mittelwertsunterschiede einer oder mehrerer nominalskalierten unabhängigen Variablen hinsichtlich einer metrisch skalierten abhängigen Variable. Bei einer unabhängigen Variable ist von einer einfaktoriellen, bei mehreren unabhängigen Variablen von einer mehrfaktoriellen Varianzanalyse die Rede. Bei letzterer können zusätzlich zu den Haupteffekten der Faktoren, auch Interaktionen auf Signifikanz geprüft werden. Vorausgesetzt werden die Normalverteilung der Daten pro Gruppe und die Homogenität der Varianzen, die mittels Levene-Test geprüft wird.

■ **A-priori-Kontraste**

Vergleicht, bei der Varianzanalyse, ein Faktor mehr als zwei Gruppen, so wird bei signifikanten Ergebnissen mittels a-priori-Kontrasten überprüft, welche Mittelwerte sich voneinander unterscheiden und welche nicht. Voraussetzung für a-priori-Kontraste ist die Formulierung gezielter Einzelvergleichshypothesen vor der Untersuchung.

4.2.3 Stichprobenbeschreibung

Der folgende Teilabschnitt stellt relevante soziodemographische Variablen, welche die untersuchte Stichprobe charakterisieren, näher vor. Es wird u. a. auf die Geschlechter- und Altersverteilung, das Wiederholen eines Schuljahres und die Absicht ein Medizinstudium zu beginnen eingegangen.

4.2.3.1 Stichprobenumfang und Geschlechterverteilung

Es konnten insgesamt 1093 SchülerInnen der 12. Schulstufe von Gymnasien untersucht werden. Mit 72,7% war der übermäßige Anteil der SchülerInnen aus Schulen in Wien, die restlichen SchülerInnen teilten sich auf, auf Schulen in Niederösterreich (20,0%) und Burgenland (7,3%). 684 (62,6%) der teilnehmenden Personen waren weiblich und 409 (37,4%) männlich (siehe Tabelle 3).

Tabelle 3: Geschlechterverteilung der Stichprobe

Geschlecht		Häufigkeit	Prozent
Gültig	männlich	409	37,4
	weiblich	684	62,6
	Gesamt	1093	100,0

Zudem gibt Abbildung 12 einen grafischen Überblick über die Aufteilung der Geschlechter in der Stichprobe.

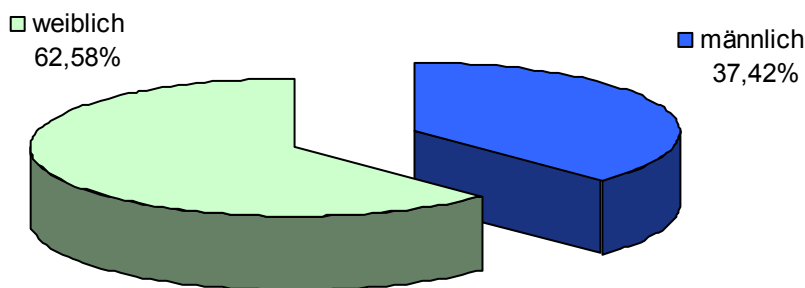


Abbildung 12: Anteilswerte der Geschlechter in Prozent

Der Chi-Quadrat-Test (siehe Tabelle 4), ergibt mit $\chi^2(1) = 69.190$, $p \leq .0001$ ein signifikantes Ergebnis. Es kann ein Verteilungsunterschied im Geschlecht

angenommen werden. Die Stichprobe weist überzufällig mehr weibliche als männliche Teilnehmer auf.

Tabelle 4: Chi-Quadrat-Test für Geschlechterverteilung

	Geschlecht
Chi-Quadrat	69.190 ^a
<i>df</i>	1
Asymptotische Signifikanz	≤.0001

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet.
Die kleinste erwartete Zellenhäufigkeit ist 546.5.

4.2.3.2 Alter in der Gesamtstichprobe und in den Versuchsgruppen

Zum Zeitpunkt der Erhebung waren die teilnehmenden SchülerInnen (1086 gültige Fälle) zwischen 16 und 21 Jahre alt ($M = 17.57$; $SD = 0.76$). Mit einem Anteil von 53,41% ist die Gruppe der 17-jährigen die am häufigsten vertretene. Die 21-jährigen bilden mit einem Prozentsatz von 0,37% die kleinste untersuchte Altersklasse (siehe Abbildung 13).

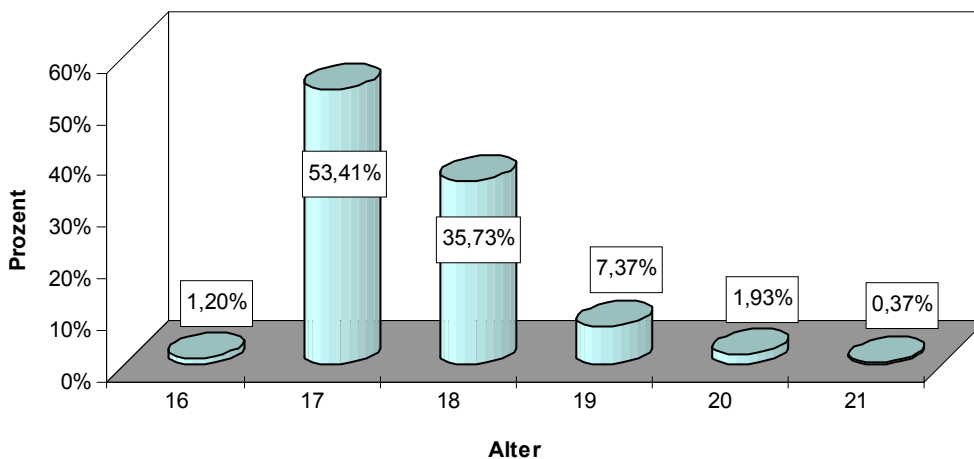


Abbildung 13: Alterskategorien in der Gesamtstichprobe

Tabelle 5 zeigt die Häufigkeiten und Anteilswerte aller Altersgruppen.

Tabelle 5: Häufigkeiten Alter gesamt

Alter		Häufigkeit	Prozent	Gültige Prozente
Gültig	16	13	1,2	1,2
	17	580	53,1	53,4
	18	388	35,5	35,7
	19	80	7,3	7,4
	20	21	1,9	1,9
	21	4	0,4	0,4
	Gesamt	1086	99,4	100,0
Fehlend	System	7	0,6	
Gesamt		1093	100,0	

Der Chi-Quadrat-Test (siehe Tabelle 6), ergibt mit $\chi^2(5) = 1643.116$, $p \leq .0001$ ein signifikantes Ergebnis. Ein Verteilungsunterschied im Alter kann angenommen werden.

Tabelle 6: Chi-Quadrat-Test für Altersverteilung

	Alter
Chi-Quadrat	1643.116 ^a
<i>df</i>	5
Asymptotische Signifikanz	$\leq .0001$

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet.
Die kleinste erwartete Zellenhäufigkeit ist 181.0.

4.2.3.3 Staatsbürgerschaft

43 (3,93%) Schüler der Gesamtstichprobe gaben keine Auskunft über ihre Staatsbürgerschaft (siehe Abbildung 14). Von den 1050 gültigen Fällen (siehe Tabelle 7) sind 992 (94,5%) Personen österreichische StaatsbürgerInnen und

58 (5,5%) TestteilnehmerInnen führen eine andere Staatszugehörigkeit, wie etwa Deutschland, Japan, Kroatien, Polen, Rumänien, Schweden, Serbien, Slowakei und Türkei an.

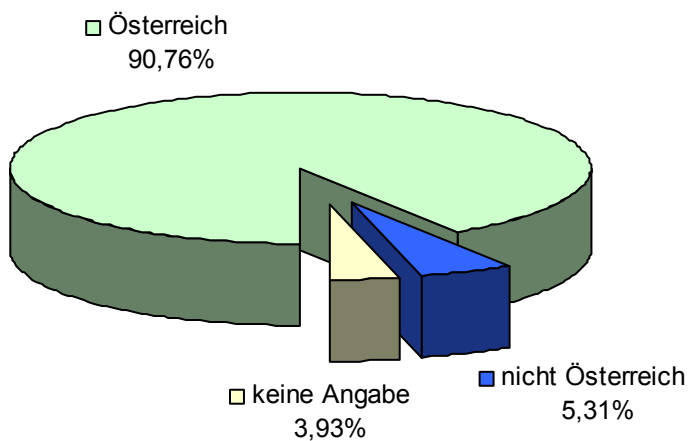


Abbildung 14: Anteilswerte der Staatsbürgerschaft in der Gesamtstichprobe

In Tabelle 7 werden die Häufigkeiten und der entsprechenden Anteilswerte der Staatsbürgerschaft für die Gesamtstichprobe sowie für die gültigen Fälle dargestellt.

Tabelle 7: Häufigkeiten der Staatsbürgerschaft

Staatsbürgerschaft		Häufigkeit	Prozent	Gültige Prozente
Gültig	Österreich	992	90,8	94,5
	nicht Österreich	58	5,3	5,5
	Gesamt	1050	96,1	100,0
Fehlend	System	43	3,9	
Gesamt		1093	100,0	

4.2.3.4 Schulzweige

Abbildung 15 illustriert, welche unterschiedlichen Schulzweige die an der Untersuchung teilnehmenden SchülerInnen der AHS-Oberstufe besuchten. Deutlich ist, dass die drei häufigsten besuchten Schulzweige der sprachliche (26,67%), der naturwissenschaftliche (22,87%) und der musikalische Zweig (18,43%) sind. 67,97% der 1080 gültigen Fälle sind SchülerInnen einer dieser drei Richtungen. Allen weiteren Schulzweigen sind Anteilswerte zwischen 8,90% (sportlicher Zweig) und 0,10% (technischer Schulzweig) zuzuordnen.

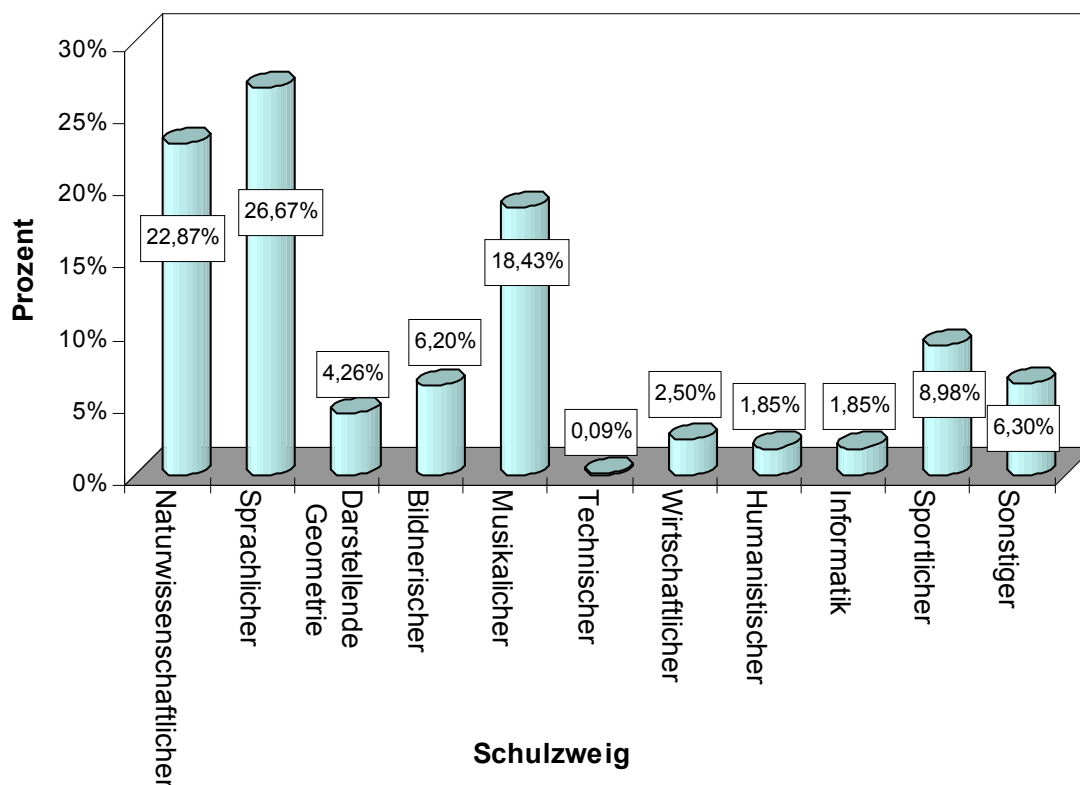


Abbildung 15: Schulzweige

Der Chi-Quadrat-Test (siehe Tabelle 8), ergibt mit $\chi^2(11) = 1184.621$, $p \leq .0001$ ein signifikantes Ergebnis. Es kann ein signifikanter Verteilungsunterschied in den Schulzweigen angenommen werden.

Tabelle 8: Chi-Quadrat-Test – Schulzweig

	Schulzweig
Chi-Quadrat	1184.621 ^a
<i>Df</i>	11
Asymptotische Signifikanz	≤.0001

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet.
Die kleinste erwartete Zellenhäufigkeit ist 90.6.

4.2.3.5 Wiederholen eines Schuljahres

Von 1087 Personen, die hierzu Angaben gemacht haben, haben 166 SchülerInnen ein Schuljahr wiederholt (siehe Tabelle 9).

Tabelle 9: Häufigkeiten – Wiederholen eines Schuljahres

Wiederholt	Beobachtetes N	erwartet	Residuum
nein	921	543.5	377.5
ja	166	543.5	-377.5
Gesamt	1087		

Der Chi-Quadrat-Test (siehe Tabelle 10) ergibt mit $\chi^2(1) = 524.402$, $p \leq .0001$ ein signifikantes Ergebnis. Es kann von einem signifikanten Verteilungsunterschied beim Wiederholen eines Schuljahres ausgegangen werden.

Tabelle 10: Chi-Quadrat-Test – Wiederholen eines Schuljahres

	Wiederholt
Chi-Quadrat	524.402 ^a
df	1
Asymptotische Signifikanz	≤ .0001

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet.
Die kleinste erwartete Zellenhäufigkeit ist 543.5.

4.2.3.5.1 Wiederholen eines Schuljahres nach Geschlecht

In Tabelle 11 soll in Abhängigkeit vom Geschlecht dargestellt werden, wie viele SchülerInnen ein Schuljahr wiederholt haben.

Tabelle 11: Kreuztabelle Geschlecht und Wiederholen einer Klasse

			Wiederholt		Gesamt
			nein	ja	
Geschlecht	männlich	Anzahl	326	77	403
		erwartet	341.5	61.5	403,0
		% innerhalb	80,9%	19,1%	100,0%
		Stand. Residuen	-.8	2,0	
	weiblich	Anzahl	595	89	684
		erwartet	579.5	104.5	684.0
		% innerhalb	87,0%	13,0%	100,0%
		Stand. Residuen	.6	-1.5	
Gesamt		Anzahl	921	166	1087
		erwartet	921.0	166.0	1087.0
		% innerhalb	84,7%	15,3%	100,0%

Die Berechnung der entsprechenden Prüfgröße ergibt mit $\chi^2(1) = 7.281$, $p = .007$ ein signifikantes Ergebnis. Es kann ein Zusammenhang zwischen Geschlecht und Wiederholen einer Klasse angenommen werden. Die männlichen Jugendlichen weisen hier mit 19,1% ein überzufälligeres Wiederholen eines Schuljahres auf als die weiblichen Jugendlichen (13,0%).

4.2.3.5.2 Wiederholen eines Schuljahres im naturwissenschaftlichen Schulzweig

Aus Tabelle 12 ist in Abhängigkeit vom Geschlecht ersichtlich, wie viele UntersuchungsteilnehmerInnen aus naturwissenschaftlichen und nicht-naturwissenschaftlichen Schulzweigen eine Klasse wiederholt haben. Mit Ausnahme des naturwissenschaftlichen Schulzweiges, werden alle weiteren Oberstufenschulzweige (siehe Abbildung 15) in die Kategorie nicht-naturwissenschaftlicher Schulzweig zusammengefasst.

Tabelle 12: Vierfeldertafel naturwissenschaftlicher Schulzweig und Wiederholen eines Schuljahres

		Wiederholt		Gesamt	
		nein	ja		
Schulzweig	naturwissenschaftlich	Anzahl	200	47	247
		erwartet	209.5	37.5	247.0
		% innerhalb	81,0%	19,0%	100,0%
		Stand. Residuen	-.7	1.5	
	nicht-naturwissenschaftlich	Anzahl	716	117	833
		erwartet	706.5	126.5	833.0
		% innerhalb	86,0%	14,0%	100,0%
		Stand. Residuen	.4	-.8	
Gesamt		Anzahl	916	164	1080
		erwartet	916.0	164.0	1080.0
		% innerhalb	84,8%	15,2%	100,0%

Mit der Berechnung der entsprechenden Prüfgröße zeigt sich mit $\chi^2 (1) = 3.672$, $p = .055$ ein nicht signifikantes Ergebnis. Es kann kein Zusammenhang in naturwissenschaftlichen und nicht naturwissenschaftlichen Schulzweigen und dem Wiederholen eines Schuljahres angenommen werden.

4.2.3.6 Absicht ein Medizinstudium zu beginnen

Anhand Tabelle 13 wird ersichtlich, wie viele der teilnehmenden GymnasiastInnen künftig ein Medizinstudium anstreben. Von den vorliegenden 1068 gültigen Fällen, geben 187 Personen an, ein Medizinstudium zu intendieren.

Tabelle 13: Häufigkeiten – Absicht ein Medizinstudium zu beginnen

Medizinstudium	Beobachtetes N	erwartet	Residuum
Nein	881	534.0	347.0
Ja	187	534.0	-347.0
Gesamt	1068		

Mit $\chi^2(1) = 450.970$, $p \leq .0001$ (siehe Tabelle 14) ergibt der Chi-Quadrat-Test ein signifikantes Ergebnis. Es kann ein signifikanter Verteilungsunterschied bei der Absicht ein Medizinstudium zu beginnen angenommen werden.

Tabelle 14: Chi-Quadrat-Test – Absicht ein Medizinstudium zu beginnen

	Medizinstudium
Chi-Quadrat	450,970 ^a
Df	1
Asymptotische Signifikanz	≤ .0001

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 534.0.

4.2.3.6.1 Absicht ein Medizinstudium zu beginnen nach Geschlecht

Die Gruppe jener 187 Personen die ein Medizinstudium beginnen möchten, setzt sich aus 125 weiblichen und 62 männlichen Teilnehmern zusammen (siehe Tabelle 15).

Tabelle 15: Vierfeldertafel Geschlecht und Absicht eine Medizinstudium zu beginnen

			Medizinstudium		Gesamt
			nein	ja	
Geschlecht	männlich	Anzahl	334	62	396
		erwartet	326.7	69.3	396.0
		% innerhalb	84,3%	15,7%	100,0%
		Stand. Residuen	.4	-.9	
	weiblich	Anzahl	547	125	672
		erwartet	554.3	117.7	672.0
		% innerhalb	81,4%	18,6%	100,0%
		Stand. Residuen	-.3	.7	
Gesamt		Anzahl	881	187	1068
		erwartet	881.0	187.0	1068.0
		% innerhalb	82,5%	17,5%	100,0%

Die Berechnung der entsprechenden Prüfgröße fällt mit $\chi^2(1) = 1.496$, $p = .221$ nicht signifikant aus. Es kann kein Zusammenhang zwischen dem Geschlecht und der Absicht Medizin zu studieren angenommen werden.

4.2.3.6.2 Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS allgemein

Tabelle 16 gibt einen Überblick über die Auseinandersetzung mit dem EMS in irgendeiner Form in Abhängigkeit der Absicht, Medizin zu studieren. Beinahe zwei Drittel, nämlich 62,6% der 187 Personen, welche planen ein

Medizinstudium zu beginnen, haben sich zum Zeitpunkt der Erhebung bereits in irgendeiner Form mit dem EMS auseinandergesetzt. D. h. sie haben bezüglich EMS im Internet recherchiert, Medizinstudenten und andere Personen befragt, die die EMS-Testung bereits absolviert haben, Personen aus dem medizinischen Bereich befragt oder auch bereits Testaufgaben des Einstiegstests bearbeitet. Ebenfalls haben sich auch 80 SchülerInnen, die nicht beabsichtigen, Medizin zu studieren mit dem EMS in irgendeiner Form befasst. Die Anzahl jener Personen, welche ein Medizinstudium anstreben und sich zum Erhebungszeitpunkt jedoch noch nicht mit dem EMS beschäftigt haben, beläuft sich auf 70.

Tabelle 16: Kontingenztafel Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS in irgendeiner Form

			EMS		Gesamt
			nein	ja	
Medizinstudium	nein	Anzahl	799	80	879
		erwartet	716.6	162.4	879.0
		% innerhalb	90,9%	9,1%	100,0%
		Stand. Residuen	3.1	-6.5	
	ja	Anzahl	70	117	187
		erwartet	152.4	34.6	187.0
		% innerhalb	37,4%	62,6%	100,0%
		Stand. Residuen	-6.7	14.0	
Gesamt		Anzahl	869	197	1066
		erwartet	869.0	197.0	1066.0
		% innerhalb	81,5%	18,5%	100,0%

Mit $\chi^2 (1) = 292.584$, $p \leq .0001$ ergibt die Berechnung der entsprechenden Prüfgröße ein signifikantes Ergebnis. Es kann ein Zusammenhang zwischen

der Absicht ein Medizinstudium zu beginnen und der Auseinandersetzung mit dem EMS in irgendeiner Form, beobachtet werden.

4.2.3.7 Auseinandersetzung mit EMS allgemein nach Geschlecht

Um ersichtlich zu machen, wie viele Personen sich in Abhängigkeit von Geschlecht bereits in irgendeiner Form mit dem EMS beschäftigt haben, wird auf die nachfolgende Tabelle (siehe Tabelle 17) verwiesen.

Tabelle 17: Kontingenztafel Geschlecht und Auseinandersetzung mit EMS allgemein

		EMS		Gesamt	
		nein	ja		
Geschlecht	männlich	Anzahl	341	61	402
		erwartet	326.8	75.2	402.0
		% innerhalb	84,8%	15,2%	100,0%
		Stand. Residuen	.8	-1.6	
	weiblich	Anzahl	541	142	683
		erwartet	555.2	127.8	683.0
		% innerhalb	79,2%	20,8%	100,0%
		Stand. Residuen	-.6	1.3	
Gesamt		Anzahl	882	203	1085
		erwartet	882.0	203.0	1085.0
		% innerhalb	81,3%	18,7%	100,0%

Die Berechnung der entsprechenden Prüfgröße ergibt mit $\chi^2(1) = 5.249$, $p = .022$ ein signifikantes Ergebnis. Ein Verteilungsunterschied in der Auseinandersetzung mit dem EMS in irgendeiner Form, in Abhängigkeit von Geschlecht kann angenommen werden. Weibliche Jugendliche beschäftigten sich zum Zeitpunkt der Datenerhebung mit 20,8% bereits überdurchschnittlich häufiger mit dem EMS als männliche Jugendliche (15,2%).

4.2.3.7.1 Auseinandersetzung mit EMS allgemein im naturwissenschaftlichen Schulzweig

Bei Betrachtung der Vierfeldertafel (Tabelle 18) zeigt sich, wie viele SchülerInnen aus dem naturwissenschaftlichen Oberstufenschulzweig sich bereits in irgendeiner Form mit dem EMS befasst haben. Zur Übersicht aller einzelnen Schulzweige, welche hier als „nicht-naturwissenschaftlicher Schulzweig“ in eine Kategorie zusammengefasst werden, wird auch hier auf Abbildung 15 verwiesen. Von 246 SchülerInnen, die den naturwissenschaftlichen Zweig absolvieren, hat sich zum Zeitpunkt der Erhebung schon beinahe ein Viertel, nämlich 23,6% (58 Personen) mit dem EMS in irgendeiner Form beschäftigt. Von den 832 SchülerInnen welche nicht naturwissenschaftliche Oberstufenschulzweige besuchen, haben sich 144 Personen bereits allgemein mit dem EMS auseinandergesetzt. Dies entspricht hier einem Anteil von nur 17,3%, der bezüglich EMS im Internet recherchiert hat, Freunde und Verwandte aus dem medizinischen Bereich befragt hat, Personen befragt hat die den EMS bereits absolviert haben, Medizinstudenten befragt hat oder selbst schon Testaufgaben des EMS bearbeitet hat.

Tabelle 18: Vierfeldertafel naturwissenschaftlicher Schulzweig und Auseinandersetzung mit EMS in irgendeiner Form

		EMS		Gesamt	
		nein	ja		
Schulzweig	naturwissenschaftlich	Anzahl	188	58	246
		erwartet	199.9	46.1	246.0
		% innerhalb	76,4%	23,6%	100,0%
		Stand. Residuen	-.8	1.8	
	nicht-naturwissenschaftlich	Anzahl	688	144	832
		erwartet	676.1	155.9	832.0
		% innerhalb	82,7%	17,3%	100,0%
		Stand. Residuen	.5	-1.0	

Gesamt	Anzahl	876	202	1078
	erwartet	876.0	202.0	1078.0
	% innerhalb	81,3%	18,7%	100,0%

Die Berechnung der entsprechenden Prüfgröße zeigt mit $\chi^2(1) = 4.901$, $p = .027$ ein signifikantes Ergebnis. Es kann ein Zusammenhang zwischen der Beschäftigung mit dem EMS in irgendeiner Form und dem Absolvieren eines naturwissenschaftlichen Schulzweiges angenommen werden.

4.2.3.8 Versuchsbedingungen

Hier soll eine Übersicht der Verteilungen in den Stereotype Threat-Bedingungen und den Risikobedingungen gegeben werden.

4.2.3.8.1 Stereotype Threat-Bedingungen

Tabelle 19 veranschaulicht die Häufigkeiten in den vier Stereotype Threat-Bedingungen.

Tabelle 19: Häufigkeiten in den Stereotype Threat-Bedingungen

Stereotype Threat-Bedingung	Beobachtetes N	erwartet	Residuum
Aktivierung	293	273.3	19.8
Neutral	257	273.3	-16.3
Aufklärung	280	273.3	6.8
Negativaktivierung	263	273.3	-10.3
Gesamt	1093		

Bei der Gleichverteilungsprüfung über die Stereotype Threat-Bedingungen, ergibt sich mit $\chi^2(3) = 2.945$, $p = .400$ keine signifikante Abweichung der beobachteten Häufigkeiten von den erwarteten Häufigkeiten (siehe Tabelle 20).

Es kann eine Gleichverteilung über die Stereotype Threat-Bedingungen angenommen werden.

Tabelle 20: Chi-Quadrat-Test – Stereotype Threat-Bedingung

	Stereotype Threat-Bedingung
Chi-Quadrat	2.945 ^a
df	3
Asymptotische Signifikanz	.400

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 273.3.

4.2.3.8.1.1 Stereotype Threat-Bedingungen nach Geschlecht

Anhand Tabelle 21 wird die Geschlechterverteilung der TestteilnehmerInnen in den vier Stereotype Threat-Bedingungen *Aktivierung*, *Neutrale Bedingung*, *Aufklärung* und *Negativaktivierung* veranschaulicht.

Tabelle 21: Kreuztabelle Geschlecht und Stereotype Threat-Bedingung

			Stereotype Threat-Bedingung				Gesamt
			Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	
Geschlecht	männlich	Anzahl	123	85	112	89	409
		erwartet	109.6	96.2	104.8	98.4	409.0
		% innerhalb	30,1%	20,8%	27,4%	21,8%	100,0%
		Stand. Residuen	1.3	-1.1	.7	-.9	
	weiblich	Anzahl	170	172	168	174	684
		erwartet	183.4	160.8	175.2	164.6	684.0
		% innerhalb	24,9%	25,1%	24,6%	25,4%	100,0%
		Stand. Residuen	-1.0	.9	-.5	.7	
Gesamt		Anzahl	293	257	280	263	1093

	erwartet	293.0	257.0	280.0	263.0	1093.0
	% innerhalb	26,8%	23,5%	25,6%	24,1%	100,0%

Die Berechnung der entsprechenden Prüfgröße ergibt mit $\chi^2 (3) = 6.909$, $p = .075$ ein nicht signifikantes Ergebnis. Dies bedeutet, dass die Zuteilung der VersuchsteilnehmerInnen in die einzelnen Stereotype Threat-Bedingungen unabhängig vom Geschlecht erfolgte.

4.2.3.8.2 Risikobedingungen

In Tabelle 22 werden die Häufigkeiten in den drei Risikobedingungen dargestellt.

Tabelle 22: Häufigkeiten in den Risikobedingungen

Risikobedingung	Beobachtetes N	erwartet	Residuum
raten	373	364.3	8.7
ohne Instruktion	368	364.3	3.7
sicher	352	364.3	-12.3
Gesamt	1093		

Die Gleichverteilungsprüfung über die Risikobedingungen, ergibt sich mit $\chi^2 (2) = 0.661$, $p = .719$ keine signifikantes Ergebnis (siehe Tabelle 23). Eine Gleichverteilung über die Risikobedingungen kann angenommen werden.

Tabelle 23: Chi-Quadrat-Test – Risikobedingung

	Risikobedingung
Chi-Quadrat	0.661 ^a
df	2
Asymptotische Signifikanz	.719

a. Bei 0 Zellen (.0%) werden weniger als 5 Häufigkeiten erwartet.
Die kleinste erwartete Zellenhäufigkeit ist 364.3.

4.2.3.8.2.1 Risikobedingungen nach Geschlecht

Die Geschlechterverteilung der Testpersonen in den drei Risikobedingungen *raten*, *ohne Instruktion* und *sicher* ist aus Tabelle 24 zu entnehmen.

Tabelle 24: Kreuztabelle Geschlecht und Risikobedingung

			Risikobedingung			Gesamt
			raten	ohne Instruktion	sicher	
Geschlecht	männlich	Anzahl	145	131	133	409
		erwartet	139.6	137.7	131.7	409.0
		% innerhalb	35,5%	32,0%	32,5%	100,0%
		Stand. Residuen	.5	-.6	.1	
	weiblich	Anzahl	228	237	219	684
		erwartet	233.4	230.3	220.3	684.0
		% innerhalb	33,3%	34,6%	32,0%	100,0%
		Stand. Residuen	-.4	.4	-.1	
Gesamt		Anzahl	373	368	352	1093
		erwartet	373.0	368.0	352.0	1093.0
		% innerhalb	34,1%	33,7%	32,2%	100,0%

Mit $\chi^2 (2) = 0.878$, $p = .645$ ergibt die Berechnung der entsprechenden Prüfgröße ein nicht signifikantes Ergebnis. Dies zeigt, dass die Zuweisung der UntersuchungsteilnehmerInnen in die einzelnen Risikobedingungen unabhängig vom Geschlecht erfolgen konnte und die Gruppeneinteilung somit als gleichmäßig angesehen werden kann.

4.2.3.8.3 Stereotype Threat-Bedingungen und Risikobedingungen

Aus nachstehender Tabelle 25 ist ersichtlich, wie sich die Verteilung der 1093 VersuchsteilnehmerInnen auf die Risikobedingungen in Abhängigkeit der Stereotype Threat-Bedingungen, gestaltet.

Tabelle 25: Kontingenztabelle Stereotype Threat-Bedingung und Risikobedingung

			Risikobedingung			Gesamt
			raten	ohne Instruktion	sicher	
Stereotype Threat-Bedingung	Aktivierung	Anzahl	93	108	92	293
		erwartet	100.0	98.6	94.4	293.0
		% innerhalb	31,7%	36,9%	31,4%	100,0%
		Stand. Residuen	-.7	.9	-.2	
	Neutral	Anzahl	90	87	80	257
		erwartet	87.7	86.5	82.8	257.0
		% innerhalb	35,0%	33,9%	31,1%	100,0%
		Stand. Residuen	.2	.1	-.3	
	Aufklärung	Anzahl	104	92	84	280
		erwartet	95.6	94.3	90.2	280.0
		% innerhalb	37,1%	32,9%	30,0%	100,0%
		Stand. Residuen	.9	-.2	-.7	
	Negativaktivierung	Anzahl	86	81	96	263
		erwartet	89.8	88.5	84.7	263.0
		% innerhalb	32,7%	30,8%	36,5%	100,0%

		Stand. Residuen	-4	-8	1.2	
Gesamt	Anzahl		373	368	352	1093
	erwartet		373.0	368.0	352.0	1093.0
	% innerhalb		34,1%	33,7%	32,2%	100,0%

Mit Berechnung der entsprechenden Prüfgröße ergibt sich mit $\chi^2(6) = 5.121$, $p = .528$ kein signifikantes Ergebnis. Es zeigt sich, dass die Zuteilung der Testpersonen in die Risikobedingungen unabhängig von den Stereotype Threat-Bedingungen erfolgte.

4.3 Deskriptivstatistik und relative Lösungshäufigkeiten der EMS-ähnlichen Aufgaben

Wie unter Berücksichtigung der Schiefekoeffizienten ersichtlich ist, kann für die einzelnen EMS-Untertests eine Normalverteilung der Daten in der Gesamtstichprobe angenommen werden (siehe Tabelle 26).

Tabelle 26: Statistische Kenngrößen zu den EMS-Untertests

		Quantitative u. formale Probleme	Schlauchfiguren	Medizinisch-nat.wissenschaftl. Grundverständnis	Diagramme u. Tabellen	Figuren Lernen	Fakten Lernen
N	Gültig	1093	1093	1093	1093	1093	1093
Anzahl der Items		16	21	16	16	8	8
M		3.00	10.19	2.81	3.01	4.43	4.24
Median		3.00	10.00	3.00	3.00	4.00	4.00
SD		1.68	3.45	1.69	1.63	2.05	1.77
Schiefe		.485	-.168	.482	.370	-.068	-.111
Kurtosis		.155	-.283	-.150	-.037	-.638	-.426

Minimum	0	0	0	0	0	0
Maximum	10	20	9	9	8	8
Cronbach's Alpha	.335	.651	.353	.314	.637	.489

4.3.1 Relative Lösungswahrscheinlichkeiten bzw. Itemschwierigkeiten

Im Folgenden werden die Lösungswahrscheinlichkeiten in den Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* abgebildet. Die Ratewahrscheinlichkeit fällt für alle Items gleich aus, da bei jeder Aufgabe eine von fünf möglichen Antworten ausgewählt werden soll. Die Ratewahrscheinlichkeit beträgt somit 20% und wird in Bezug auf die Lösungswahrscheinlichkeiten nicht berücksichtigt.

4.3.1.1 Relative Lösungshäufigkeiten im Untertest *Quantitative und formale Probleme*

In Tabelle 27 werden die Lösungshäufigkeiten (Schätzer für die Itemschwierigkeiten p) und die entsprechenden Standardabweichungen im Subtest *Quantitative und formale Probleme* aufgegliedert nach Geschlecht sowie für die gesamte Stichprobe dargestellt. Die Itemschwierigkeit zeigt den Anteil derjenigen Personen die das Item richtig lösen konnten (Bortz & Döring, 2006).

Tabelle 27: Lösungswahrscheinlichkeiten im Untertest Quantitative und formale Probleme

Bearbeitungszeit: 10 Min.	Gesamt		weiblich		männlich	
	n = 1093		n = 684		n = 409	
Itemanzahl: 16						
Itemreihenfolge	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Quant1	0.54	0.50	0.54	0.50	0.55	0.50

Quant2	0.47	0.50	0.40	0.49	0.59	0.49
Quant3	0.26	0.44	0.23	0.42	0.31	0.46
Quant4	0.27	0.45	0.20	0.40	0.40	0.49
Quant73	0.51	0.50	0.54	0.50	0.47	0.50
Quant74	0.27	0.44	0.25	0.43	0.30	0.46
Quant75	0.21	0.41	0.20	0.40	0.24	0.43
Quant76	0.11	0.31	0.11	0.31	0.12	0.33
Quant5	0.16	0.37	0.18	0.38	0.15	0.35
Quant6	0.04	0.20	0.04	0.20	0.04	0.20
Quant7	0.02	0.16	0.02	0.14	0.03	0.18
Quant8	0.02	0.15	0.02	0.15	0.02	0.14
ZQuant93	0.03	0.18	0.03	0.17	0.04	0.21
ZQuant94	0.03	0.16	0.02	0.15	0.03	0.17
ZQuant95	0.02	0.15	0.02	0.14	0.03	0.16
ZQuant96	0.02	0.14	0.02	0.14	0.02	0.15

Die Lösungswahrscheinlichkeiten liegen im unteren bis mittleren Bereich. Zu erkennen ist, dass Lösungswahrscheinlichkeiten gegen Ende des Untertests abnehmen, was auf die Speed-Komponente schließen lässt. Wie gut ein einzelnes Item das Gesamtergebnis eines Tests repräsentiert, wird aufgrund der Trennschärfe der Items ersichtlich (Bortz & Döring, 2006). Die Trennschärfe der Items im Untertest *Quantitative und formale Probleme* ist gering.

4.3.1.2 Relative Lösungshäufigkeiten im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

Im Folgenden werden die Lösungswahrscheinlichkeiten (geschätzte Itemschwierigkeiten) und die entsprechenden Standardabweichungen im

Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* separiert nach Geschlecht sowie für die Gesamtstichprobe angeführt (siehe Tabelle 28).

Tabelle 28: Lösungswahrscheinlichkeiten im Untertest Medizinisch-naturwissenschaftliches Grundverständnis

Bearbeitungszeit: 10 Min. Itemanzahl: 16	Gesamt		weiblich		männlich	
	n = 1093		n = 684		n = 409	
Itemreihenfolge	M	SD	M	SD	M	SD
Natur32	0.56	0.50	0.54	0.50	0.60	0.49
Natur33	0.39	0.49	0.40	0.49	0.37	0.48
Natur34	0.38	0.49	0.36	0.48	0.42	0.49
Natur35	0.19	0.40	0.15	0.36	0.27	0.45
Natur26	0.29	0.45	0.28	0.45	0.30	0.46
Natur27	0.30	0.46	0.32	0.47	0.27	0.44
Natur28	0.21	0.41	0.19	0.40	0.23	0.42
Natur29	0.16	0.37	0.15	0.36	0.17	0.38
Natur36	0.07	0.25	0.06	0.24	0.08	0.28
Natur37	0.05	0.22	0.05	0.22	0.05	0.21
Natur38	0.03	0.17	0.02	0.16	0.04	0.19
Natur39	0.04	0.19	0.04	0.18	0.05	0.21
Natur44	0.04	0.19	0.04	0.18	0.05	0.21
Natur45	0.03	0.18	0.02	0.15	0.05	0.21
Natur46	0.04	0.20	0.04	0.20	0.04	0.20
Natur47	0.03	0.16	0.02	0.14	0.03	0.18

Auch in diesem Subtest liegen die geschätzten Itemschwierigkeiten im mittleren bis höheren Bereich und es lässt sich auch ableiten, dass sich aufgrund der Zeitbegrenzung, die Lösungswahrscheinlichkeiten im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* gegen das Ende der

Skalenbearbeitung hin vermindern. Die Trennschärpen der Items in diesem Bereich sind gering.

4.3.1.3 Relative Lösungshäufigkeiten im Untertest *Diagramme und Tabellen*

In Tabelle 29 werden, die Lösungswahrscheinlichkeiten (geschätzte Itemschwierigkeiten) und die entsprechenden Standardabweichungen im Subtest *Diagramme und Tabellen* aufgeteilt nach Geschlecht und auch für die Gesamtstichprobe abgebildet.

Tabelle 29: Lösungswahrscheinlichkeiten im Untertest Diagramme und Tabellen

Bearbeitungszeit: 10 Min. Itemanzahl: 16	Gesamt		weiblich		männlich	
	n = 1093		n = 684		n = 409	
Itemreihenfolge	M	SD	M	SD	M	SD
Diagr71	0.58	0.49	0.54	0.50	0.65	0.48
Diagr72	0.43	0.50	0.41	0.49	0.45	0.50
Diagr73	0.59	0.49	0.55	0.50	0.64	0.48
Diagr74	0.25	0.43	0.24	0.43	0.26	0.44
Diagr161	0.27	0.44	0.25	0.43	0.30	0.46
Diagr162	0.27	0.45	0.27	0.45	0.27	0.45
Diagr163	0.13	0.33	0.12	0.32	0.14	0.35
Diagr164	0.16	0.37	0.16	0.37	0.17	0.38
Diagr75	0.09	0.28	0.08	0.28	0.09	0.29
Diagr76	0.05	0.23	0.05	0.22	0.06	0.24
Diagr77	0.03	0.17	0.03	0.17	0.03	0.18
Diagr78	0.03	0.18	0.04	0.19	0.02	0.16
Diagr181	0.02	0.15	0.02	0.14	0.02	0.16
Diagr182	0.05	0.22	0.04	0.21	0.06	0.23
Diagr183	0.04	0.19	0.04	0.18	0.04	0.19

Diagr184	0.03	0.17	0.02	0.16	0.04	0.19
----------	------	------	------	------	------	------

Im Subtest *Diagramme und Tabellen* liegen die Lösungswahrscheinlichkeiten ebenfalls im unteren bis mittleren Bereich. Auch hier ist ein Rückgang der Lösungswahrscheinlichkeiten gegen Ende des Untertests erkennbar und lässt auf die Speed-Komponente schließen. Die Itemschwierigkeit ist auch hier zu gering.

4.4 Ergebnisse

In diesem Kapitel werden die Überprüfung der experimentellen Variationen, die Reliabilitätsanalysen zu den EMS-ähnlichen Bereichen und die Interkorrelationen zwischen den Untertests, sowie die Prüfung und Ergebnisdarstellung der Haupt- und Nebenhypothesen behandelt.

4.4.1 Überprüfung der experimentellen Stereotype Threat-Versuchsbedingungen mittels *Wissenstest*

Folgende vier Stereotype Threat-Bedingungen wurden vorgegeben:

A) *Aktivierung eines Stereotype Threats*: Es wurde mitgeteilt, dass Frauen bessere Leistungen im Gedächtnis und Männer bessere Leistungen in der räumlichen Orientierung, im Arbeiten mit Zahlen, in der Interpretation von Diagrammen und Tabellen und im naturwissenschaftlichen Wissen, aufweisen.

B) *Neutrale Bedingung*: Es wurde darauf aufmerksam gemacht, dass es keinerlei geschlechtsspezifische Unterschiede gebe.

C) *Aufklärung*: Über das Phänomen Stereotype Threat wurde kurz aufgeklärt.

D) *Negativaktivierung eines Stereotype Threats*: Es wurde darauf hingewiesen, dass Männer bessere Leistungen im Gedächtnis und Frauen bessere Leistungen in der räumlichen Orientierung, im Arbeiten mit Zahlen, in der

Interpretation von Diagrammen und Tabellen und im naturwissenschaftlichen Wissen, erbringen.

Die Überprüfung der Wirksamkeit der vier Stereotype Threat-Instruktionen in den sechs EMS-Bereichen wird mittels *Wissenstest* durchgeführt. Überdies wurde im *Wissenstest* eine Aufgabe zur Theory of Mind vorgegeben, Einzelheit zu den Ergebnissen der Theory of Mind sind in der Arbeit von Anzirk Katrin zu finden. Zur Konstruktion der im *Wissenstest* vorgegebenen Aufgaben diene der *Beliefs About the cognitive Abilities of Females and Males Questionnaire* von Halpern und Tan (2001) zur Orientierung.

Im Fremdeinschätzungsteil angegebene Prozentwerte, welche im Bereich zwischen 45,0 und 55,0 Prozent einzuordnen sind, werden in den Bedingungen *Neutral* und *Aufklärung* so gehandhabt als, dass die Instruktion wirksam ist. Prozentangaben welche sich außerhalb dieser Bereichsgröße bewegen, werden als erfolgreiche Instruktion in den Bedingungen *Aktivierung* sowie *Negativaktivierung eines Stereotype Threats* gewertet.

Im Folgenden wird die Wirksamkeitsüberprüfung der Stereotype Threat-Bedingungen in den drei Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen*, in Bezug auf die Fremd- bzw. Selbsteinschätzung, behandelt.

Die Prüfung der Wirksamkeit der Stereotype Threat-Bedingungen in den EMS-ähnlichen Aufgaben zur Fremdeinschätzung erfolgt mittels χ^2 -Tests. In der Kontingenztafel (siehe Tabelle 30) wird die Fremdeinschätzungsaufgabe zum Untertest *Quantitative und formale Probleme* in Abhängigkeit der vier Stereotype Threat-Bedingungen, dargestellt.

Tabelle 30: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Quantitative und formale Probleme

			Wirksamkeit in Fremdeinschätzungsaufgabe Quant. u. formale Probleme			Gesamt
			nein	ja	konträr	
			Stereotype Threat-Bedingung	Aktivierung	Anzahl	
erwartet	90.7	157.9			40.4	289.0
% innerhalb	20,8%	71,6%			7,6%	100,0%
Stand. Residuen	-3.2	3.9			-2.9	
Neutral	Anzahl	86		142	26	254
	erwartet	79.7		138.8	35.5	254.0
	% innerhalb	33,9%		55,9%	10,2%	100,0%
	Stand. Residuen	.7		.3	-1.6	
Aufklärung	Anzahl	141		114	22	277
	erwartet	86.9		151.3	38.7	277.0
	% innerhalb	50,9%		41,2%	7,9%	100,0%
	Stand. Residuen	5.8		-3.0	-2.7	
Negativaktivierung	Anzahl	52		127	81	260
	erwartet	81.6		142.0	36.4	260.0
	% innerhalb	20,0%		48,8%	31,2%	100,0%
	Stand. Residuen	-3.3		-1.3	7.4	
Gesamt	Anzahl	339	590	151	1080	
	erwartet	339.0	590.0	151.0	1080.0	
	% innerhalb	31,4%	54,6%	14,0%	100,0%	

Die Berechnung der entsprechenden Prüfgröße ergibt mit $\chi^2(6) = 154.39$, $p \leq .0001$ ein signifikantes Resultat. Es kann ein Verteilungsunterschied in der Fremdeinschätzungsaufgabe zum Untertest *Quantitative und formale Probleme* in Abhängigkeit von der Stereotype Threat-Bedingung angenommen werden.

Tabelle 31 veranschaulicht in Abhängigkeit der Stereotype Threat-Bedingungen, die Fremdeinschätzungsaufgabe zum Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*.

Tabelle 31: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Med.-nat. Grundverständnis

			Wirksamkeit in Fremdeinschätzungsaufgabe Med.-nat. Grundverständnis			Gesamt
			nein	ja	konträr	
			Stereotype Threat-Bedingung	Aktivierung	Anzahl	
erwartet	89.7	153.3			47.0	290.0
% innerhalb	34,5%	51,4%			14,1%	100,0%
Stand. Residuen	1.1	-.3			-.9	
Neutral	Anzahl	76		155	23	254
	erwartet	78.6		134.3	41.2	254.0
	% innerhalb	29,9%		61,0%	9,1%	100,0%
	Stand. Residuen	-.3		1.8	-2.8	
Aufklärung	Anzahl	87		150	40	277
	erwartet	85.7		146.5	44.9	277.0
	% innerhalb	31,4%		54,2%	14,4%	100,0%
	Stand. Residuen	.1		.3	-.7	
Negativaktivierung	Anzahl	71		117	71	259
	erwartet	80.1		136.9	42.0	259.0
	% innerhalb	27,4%		45,2%	27,4%	100,0%
	Stand. Residuen	-1.0		-1.7	4.5	
Gesamt	Anzahl	334	571	175	1080	
	erwartet	334.0	571.0	175.0	1080.0	
	% innerhalb	30,9%	52,9%	16,2%	100,0%	

Mit $\chi^2(6) = 38.017$, $p \leq .0001$ ergibt die Berechnung der entsprechenden Prüfgröße ein signifikantes Ergebnis. Es kann ein Zusammenhang zwischen der Fremdeinschätzungsaufgabe zum Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* und den Stereotype Threat-Bedingungen angenommen werden.

Die nachfolgende Kontingenztafel (Tabelle 32) zeigt die Fremdeinschätzungsaufgabe zum Untertest *Diagramme und Tabellen* in Abhängigkeit der vier Stereotype Threat-Bedingungen.

Tabelle 32:Kontingenztafel Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Diagramme und Tabellen

			Wirksamkeit in Fremdeinschätzungsaufgabe Diagramme und Tabellen			Gesamt
			nein	ja	konträr	
Stereotype Threat-Bedingung	Aktivierung	Anzahl	52	223	14	289
		erwartet	85.6	170.5	32.9	289.0
		% innerhalb	18,0%	77,2%	4,8%	100,0%
		Stand. Residuen	-3.6	4.0	-3.3	
	Neutral	Anzahl	95	139	20	254
		erwartet	75.3	149.8	28.9	254.0
		% innerhalb	37,4%	54,7%	7,9%	100,0%
		Stand. Residuen	2.3	-.9	-1.7	
	Aufklärung	Anzahl	125	123	29	277
		erwartet	82.1	163.4	31.5	277.0
		% innerhalb	45,1%	44,4%	10,5%	100,0%
		Stand. Residuen	4.7	-3.2	-.5	
	Negativaktivierung	Anzahl	48	152	60	260
		erwartet	77.0	153.4	29.6	260.0
		% innerhalb	18,5%	58,5%	23,1%	100,0%
		Stand. Residuen	-3.3	-.1	5.6	

Gesamt	Anzahl	320	637	123	1080
	erwartet	320.0	637.0	123.0	1080.0
	% innerhalb	29,6%	59,0%	11,4%	100,0%

Mittels Berechnung der entsprechenden Prüfgröße zeigt sich mit $\chi^2(6) = 123.767$, $p \leq .0001$ ein signifikanter Verteilungsunterschied in der Fremdeinschätzungsaufgabe zum Untertest *Diagramme und Tabellen*, in Abhängigkeit der Stereotype Threat-Bedingungen.

Auch in den Fremdeinschätzungsaufgaben zu den Untertests *Schlauchfiguren* ($\chi^2(6) = 53.913$, $p \leq .0001$), *Figuren Lernen* ($\chi^2(6) = 168.720$, $p \leq .0001$) und *Fakten Lernen* ($\chi^2(6) = 104.697$, $p \leq .0001$) können Verteilungsunterschiede in Abhängigkeit der Stereotype Threat-Bedingungen angenommen werden. Die Kontingenztafeln hierzu können im Anhang (siehe Tabelle A1 bis A3) betrachtet werden. Die Wirksamkeit der Stereotype Threat-Instruktionen kann für alle Aufgabenbereiche angenommen werden.

Anhand der standardisierten Residuen wird Intensität und Ausmaß der Wirksamkeit der Stereotype Threat-Instruktionen, unter Berücksichtigung des jeweiligen EMS-ähnlichen Bereiches erkennbar. Sobald das standardisierte Residuum einen Wert $\geq \pm 2$ aufweist, kann ein bedeutender Unterschied zwischen beobachteter und erwarteter Häufigkeit angenommen werden (Bühl & Zöfel, 2000).

In Anlehnung an die Untersuchung von Halpern und Tan (2001), wo die Wirksamkeit mittels Prozentmittelwerte dargestellt wird, können die entsprechenden Mittelwerte der Prozentränge in den Fremdeinschätzungsaufgaben in Abhängigkeit der Stereotype-Threat Bedingungen, der Tabelle A4 im Anhang entnommen werden.

Die Selbsteinschätzung in den sechs EMS-ähnlichen Bereichen in Abhängigkeit von Geschlecht und Stereotype Threat-Bedingung wird mittels zweifaktorieller multivariater Varianzanalyse überprüft. Die Homogenität der Varianzen ist im Bereich *Figuren Lernen* und *Schlauchfiguren* ($p < .05$) nicht

anzunehmen. Da das Verfahren gegen eine Verletzung der Varianzhomogenität relativ robust ist, ist die Durchführung der Varianzanalyse als unproblematisch zu betrachten.

Tabelle 33: Kennwerte und Prüfgrößen der Selbsteinschätzung in Abhängigkeit vom Geschlecht

EMS-ähnlicher Untertest	männlich		weiblich		F (1,1057)	Sig.
	M	SD	M	SD		
Quantitative und formale Probleme	5.05	1.68	4.04	1.79	78.300	≤ .0001**
Schlauchfiguren	5.89	1.40	5.16	1.73	48.401	≤ .0001**
Med.-nat.wissensch. Grundverständnis	5.09	1.43	4.86	1.39	5.277	.022*
Diagramme und Tabellen	5.09	1.57	4.22	1.60	73.546	≤ .0001**
Figuren Lernen	6.63	0.87	6.50	1.04	4.667	.031*
Fakten Lernen	4.93	1.30	5.05	1.23	1.753	.186

** . $p < .01$

* . $p < .05$

Wie aus Tabelle 33 entnommen werden kann, zeigen sich bei der Selbsteinschätzung unabhängig von der Stereotype Threat-Bedingung, in allen Bereichen Unterschiede im Geschlecht, mit Ausnahme des Bereichs *Fakten Lernen* $F(p = .186)$. Die männlichen Untersuchungspersonen schätzen sich bei 1065 gültigen Fällen, in den folgenden fünf Bereichen signifikant besser ein als die weiblichen Untersuchungspersonen: *Quantitative und formale Probleme* ($p \leq .0001$), *Medizinisch-naturwissenschaftliches Grundverständnis* ($p = .022$), *Diagramme und Tabellen* ($p \leq .0001$), sowie *Figuren Lernen* ($p = .031$) und *Schlauchfiguren* ($p \leq .0001$).

Bei der Überprüfung der Stereotype Threat-Bedingung unabhängig vom Geschlecht zeigt sich in keinem EMS-ähnlichen-Bereich ein signifikanter Unterschied ($p \geq .155$).

In Abhängigkeit vom Geschlecht und der Stereotype Threat-Bedingung kann im Bereich *Diagramme und Tabellen* $F(3,1057) = 6.820$, $p \leq .0001$ eine

Wechselwirkung beobachtet werden. Abbildung 16 veranschaulicht diese Interaktion grafisch.

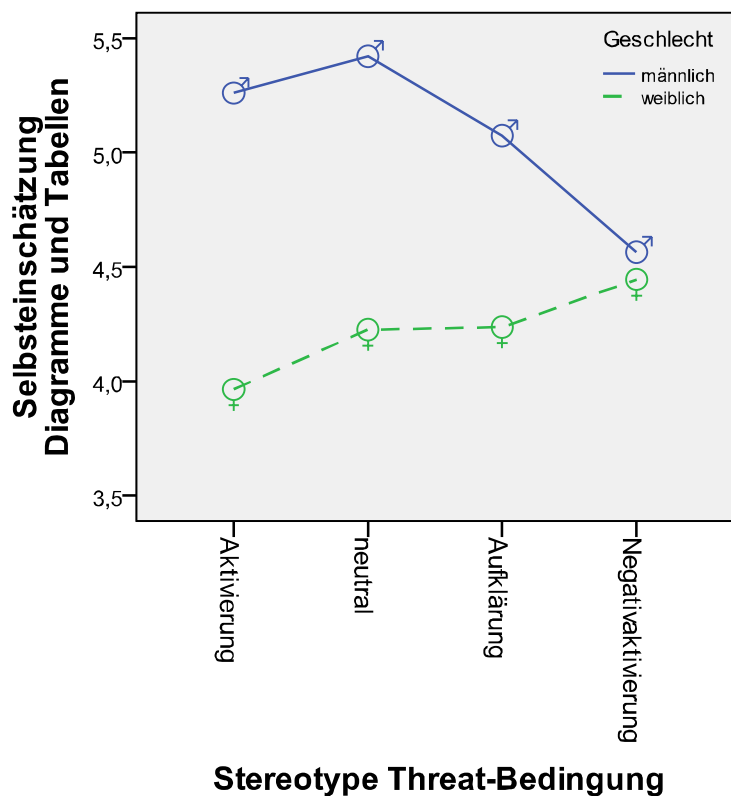


Abbildung 16: Interaktionsdiagramm Selbsteinschätzung im Bereich Diagramme und Tabellen

Aus Abbildung 16 ist zu erkennen, dass hier die Selbsteinschätzung der männlichen Testpersonen, im Gegensatz zur Selbsteinschätzung der weiblichen Testpersonen in den verschiedenen Stereotype Threat-Bedingungen, unterschiedlich hoch ausfällt. Die *Negativaktivierung* ist im Bereich *Diagramme und Tabellen* insofern wirksam, als sich die männlichen Jugendlichen hier schlechter einschätzen als gegenüber den anderen Bedingungen. Die weiblichen Testpersonen schätzen sich unter *Negativaktivierung* hingegen wesentlich besser ein als in den anderen drei Stereotype Threat-Bedingungen.

Die Stereotype Threat-Instruktionen waren schließlich bezüglich aller Aufgabenbereiche wirksam, was die Annahme der Wirksamkeit der Stereotype Threat-Bedingungen erlaubt.

Bei der Selbsteinschätzung kann in fünf von sechs Bereichen ein relativer Niveauunterschied zugunsten der Männer beobachtet werden. Die Stereotype Threat-Bedingungen zeigten bei der Selbsteinschätzung nur auf die Aufgabenbereiche *Diagramme und Tabellen* Effekte.

4.4.2 Reliabilitätsanalysen

Die Reliabilität (Zuverlässigkeit) zeigt die Messgenauigkeit, mit der ein geprüftes Merkmal gemessen wird (Bortz & Döring, 2006). Cronbach's Alpha, als Abschätzung der unteren Grenze der Reliabilität, gibt die innere Konsistenz an (Wagner-Menghin, 2003). Die Methode nach Cronbach Alpha, die nur essentielle *tau*-Äquivalenz voraussetzt ist, da die Items unterschiedliche Schwierigkeiten aufweisen, die Methode der Wahl (Rost, 2004).

Es werden in Tabelle 34 die Ergebnisse der durchgeführten Reliabilitätsanalysen zu den sechs EMS-ähnlichen Bereichen, sowie dem computerisierten GDT, nach Cronbach Alpha, angeführt. Ebenfalls werden hier die Reliabilitäten der EMS-Untertests in der Schweiz nach Hänsgen und Spicher (2007) zum Vergleich dargestellt.

Tabelle 34: Reliabilitäten nach Cronbach's Alpha

N	Gültig = 1093	Cronbach's Alpha	Cronbach's Alpha: EMS-Untertests in der Schweiz (Hänsgen & Spicher, 2007)
	Quantitative und formale Probleme	.335	.70
	Schlauchfiguren	.651	.73
	Med.-nat.wissensch. Grundverständnis	.353	.65
	Diagramme und Tabellen	.314	.69
	Figuren Lernen	.637	.63

Fakten Lernen	.489	.73
GDT	.891	

Wie ersichtlich, weist der GDT eine relativ hohe Reliabilität auf. Bei den EMS-ähnlichen Bereichen zeigen jedoch lediglich die Bereiche *Schlauchfiguren* und *Figuren Lernen* eine befriedigende Reliabilität. Die Bereich *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis*, *Diagramme und Tabellen* und *Fakten lernen* weisen niedrige Reliabilitäten auf und weisen somit auf das Vorliegen wenig konsistenter Skalen hin. Die Items, die in dieser Untersuchung eine Skala bilden, messen nicht genau. Die niedrigen Konsistenzen sind auch Hinweis auf eine Mehrdimensionalität der Skalen.

Die EMS-ähnlichen Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* weisen zusammen nach Cronbach's Alpha eine Reliabilität von .56 auf. Daraus lässt sich schließen, dass die drei EMS-ähnlichen Untertests zusammen nicht die gleichen Merkmale und Fähigkeitsbereiche erfassen.

In den folgenden Tabellen werden die Reliabilitäten getrennt nach allen möglichen Bedingungskombinationen je Untertest dargestellt. Je nach Faktorstufenkombination fallen die Reliabilitäten durchaus unterschiedlich aus.

Tabelle 35: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl)
EMS- ähnlicher Untertest Quantitative und formale Probleme

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	Gesamt
raten	.298 (93; 16)	.244 (90; 16)	.400 (104; 15)	.098 (86; 15)	.287 (373; 16)
ohne Instruktion	.359 (108; 16)	.296 (87; 15)	.418 (92; 16)	.297 (81; 14)	.352 (368; 16)
sicher	.379 (92; 16)	.432 (80; 10)	.213 (84; 14)	.369 (96; 15)	.348 (352; 16)
Gesamt	.353 (293; 16)	.343 (257; 16)	.356 (280; 16)	.279 (263; 16)	.335 (1093; 16)

Wie aus Tabelle 35 ersichtlich, resultiert aus der Stereotype Threat-Bedingung *Neutral* und der Risikobedingung *sicher* mit .432 die höchste Reliabilität im Untertest *Quantitative und formale Probleme*.

Tabelle 36: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Schlauchfiguren

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	Gesamt
raten	.728 (93; 21)	.708 (90; 21)	.663 (104; 21)	.596 (86; 21)	.681 (373; 21)
ohne Instruktion	.616 (108; 21)	.660 (87; 21)	.712 (92; 21)	.485 (81; 21)	.641 (368; 21)
sicher	.655 (92; 21)	.559 (80; 21)	.498 (84; 21)	.707 (96; 21)	.626 (352; 21)
Gesamt	.672 (293; 21)	.657 (257; 21)	.645 (280; 21)	.622 (263; 21)	.651 (1093; 21)

Im Untertest *Schlauchfiguren* erzielt die Stereotype Threat-Bedingung *Aktivierung* in Kombination mit der Risikobedingung *raten*, mit einem Wert von .728 die höchste Reliabilität (siehe Tabelle 36).

Tabelle 37: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Medizinisch-naturwissenschaftliches Grundverständnis

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	Gesamt
raten	.247 (93; 16)	.442 (90; 16)	.398 (104; 16)	.295 (86; 16)	.362 (373; 16)
ohne Instruktion	.241 (108; 16)	.476 (87; 16)	.380 (92; 16)	.266 (81; 15)	.352 (368; 16)
sicher	.174 (92; 15)	.441 (80; 14)	.319 (84; 16)	.364 (96; 16)	.329 (352; 16)
Gesamt	.219 (293; 16)	.467 (257; 16)	.376 (280; 16)	.317 (263; 16)	.353 (1093; 16)

Wie aus Tabelle 37 ersichtlich, stellt sich im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* die Reliabilität in der Stereotype Threat-Bedingungen *Neutral* relativ gesehen besser dar, als in den anderen drei Stereotype Threat-Bedingungen. Die höchste Reliabilität resultiert mit .476

in der Faktorstufenkombination *Neutral* und *ohne Instruktion*. Die Stereotype Threat-Bedingungen *Aktivierung* bewirkt hier besonders niedrige Reliabilitäten.

Tabelle 38: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Diagramme und Tabellen

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	Gesamt
raten	.370 (93; 16)	.341 (90; 15)	.288 (104; 16)	.010 (86; 16)	.289 (373; 16)
ohne Instruktion	.270 (108; 16)	.256 (87; 16)	.262 (92; 16)	.140 (81; 13)	.255 (368; 16)
sicher	.366 (92; 15)	.411 (80; 15)	.434 (84; 16)	.320 (96; 16)	.381 (352; 16)
Gesamt	.337 (293; 16)	.364 (257; 16)	.332 (280; 16)	.203 (263; 16)	.314 (1093; 16)

Tabelle 38 zeigt, dass im Untertest *Diagramme und Tabellen* mit .434, die Stereotype Threat-Bedingung *Aufklärung* in Verbindung mit der Risikobedingung *sicher* die höchste Reliabilität bewirkt.

Tabelle 39: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Figuren Lernen

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ-aktivierung	Gesamt
raten	.577 (93; 8)	.682 (90; 8)	.413 (104; 8)	.668 (86; 8)	.595 (373; 8)
ohne Instruktion	.622 (108; 8)	.624 (87; 8)	.743 (92; 8)	.580 (81; 8)	.655 (368; 8)
sicher	.631 (92; 8)	.658 (80; 8)	.534 (84; 8)	.752 (96; 8)	.659 (352; 8)
Gesamt	.606 (293; 8)	.654 (257; 8)	.600 (280; 8)	.682 (263; 8)	.637 (1093; 8)

Die höchste Reliabilität im Untertest *Figuren Lernen* erzielt die Bedingungskombination *Negativaktivierung* und *sicher* mit einem Wert von .752 (siehe Tabelle 39).

Tabelle 40: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl)
EMS- ähnlicher Untertest Fakten Lernen

Risiko-/Stereotype Threat-Bedingung	Aktivierung	Neutral	Aufklärung	Negativ- aktivierung	Gesamt
raten	.399 (93;8)	.464 (90;8)	.309 (104;8)	.436 (86;8)	.399 (373;8)
ohne Instruktion	.607 (108;8)	.557 (87;8)	.499 (92;8)	.377 (81;8)	.532 (368;8)
sicher	.521 (92;8)	.580 (80;8)	.386 (84;8)	.584 (96;8)	.524 (352;8)
Gesamt	.521 (293;8)	.531 (257;8)	.404 (280;8)	.490 (263;8)	.489 (1093;8)

Aus Tabelle 40 ist ersichtlich, dass die höchste Reliabilität im Untertest *Fakten Lernen* mit .607 aus der Faktorstufenkombination *Aktivierung* und *ohne Instruktion* resultiert.

4.4.3 Interkorrelationen zwischen den Untertests

Die Zusammenhänge zwischen den EMS-ähnlichen Subtests werden nachstehend in Tabelle 41, mittels Produkt-Moment-Korrelation beschrieben.

Tabelle 41: Interkorrelationsmatrix Produkt-Moment-Korrelationen r zwischen den EMS-ähnlichen Untertests (N = 1093)

		Quantitative und formale Probleme	Schlauch- figuren	Med.-nat. wissensch. Grundverständnis	Diagramme und Tabellen	Figuren Lernen
Schlauchfiguren	r	.162**				
Med.-naturwissenschaftliches Grundverständnis	r	.273**	.224**			
Diagramme und Tabellen	r	.287**	.170**	.334**		
Figuren Lernen	r	.067*	.178**	.077*	.094**	
Fakten Lernen	r	.110**	.169**	.180**	.116**	.276**

** . Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

* . Die Korrelation ist auf dem Niveau von .05 (2-seitig) signifikant.

Anhand der Korrelationskoeffizienten zeigen sich geringe bis mittel hohe, positive Zusammenhänge zwischen den EMS-ähnlichen Aufgaben. Der höchste Zusammenhang wird zwischen Untertest *Diagramme und Tabellen* und *Medizinisch-naturwissenschaftlichen Grundverständnis* mit $r = .334$ beobachtet. Es lässt sich erkennen, dass zwischen den Leistungen in den einzelnen EMS-ähnlichen Bereichen ein positiver Zusammenhang besteht, d.h. mit steigender Fähigkeit einer Person ist in allen Bereichen eine gute Leistung zu erwarten.

4.4.4 Uni- und multivariate Ergebnisse

In den folgenden Kapiteln werden die Haupt- und Nebenhypothesen mittels uni- und multivariater Verfahren überprüft.

4.4.4.1 Haupthypothesen

4.4.4.1.1 Hypothesenblock zum Untertest Quantitative und formale Probleme

H₁(1): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme* in Abhängigkeit vom Geschlecht.

H₁(2): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme* in Abhängigkeit von den vier Stereotype Threat-Bedingungen *Aktivierung, Neutrale Bedingung, Aufklärung* und *Negativaktivierung*.

H₁(3): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme* in Abhängigkeit von den drei Risikobedingungen *raten, ohne Instruktion* und *sicher*.

H₁(4): Es gibt im Untertest *Quantitative und formale Probleme* eine signifikante Wechselwirkung zwischen Geschlecht und Stereotype Threat-Bedingung.

H₁(5): Es gibt im Untertest *Quantitative und formale Probleme* eine signifikante Wechselwirkung zwischen Geschlecht und Risikobedingung.

H₁(6): Es gibt im Untertest *Quantitative und formale Probleme* eine signifikante Wechselwirkung zwischen Stereotype Threat-Bedingung und Risikobedingung.

H₁(7): Es gibt im Untertest *Quantitative und formale Probleme* eine signifikante Interaktion höherer Ordnung zwischen Geschlecht, Stereotype Threat-Bedingung und Risikobedingung.

4.4.4.1.2 Hypothesenprüfung zum Untertest *Quantitative und formale Probleme*

Die Überprüfung der einzelnen EMS-ähnlichen Untertests wird mittels univariater dreifaktorieller Varianzanalyse durchgeführt.

Unabhängige Variablen (UV):

- Geschlecht (2)
- Stereotype Threat-Bedingung (4)
- Risikobedingung (3)

Abhängige Variable (AV):

- Untertest *Quantitative und formale Probleme*

Tabelle 42 zeigt eine Übersicht der Mittelwerte und zugehörigen Standardabweichungen im Untertest *Quantitative und formale Probleme* unter Berücksichtigung der Geschlechter sowie der Stereotype Threat- und Risikobedingungen.

Tabelle 42: Deskriptive Statistiken – Kennwerte AV Untertest *Quantitative und formale Probleme*

Stereotype Threat-Bedingung	Risiko-bedingung	männlich			weiblich			Gesamt		
		M	SD	n	M	SD	n	M	SD	N
Aktivierung	Gesamt	3.45	1.74	123	2.84	1.63	170	3.09	1.70	293
Neutral	Gesamt	3.45	1.81	85	2.77	1.60	172	3.00	1.70	257
Aufklärung	Gesamt	3.31	1.91	112	2.86	1.55	168	3.04	1.71	280
Negativ-aktivierung	Gesamt	3.13	1.65	89	2.74	1.53	174	2.87	1.58	263

Gesamt	raten	3.66	1.82	145	2.98	1.52	228	3.24	1.68	373
	ohne Instruktion	3.24	1.77	131	2.71	1.60	237	2.90	1.70	368
	sicher	3.11	1.71	133	2.71	1.59	219	2.86	1.65	352
	Gesamt	3.34	1.78	409	2.80	1.57	684	3.00	1.68	1093

Die Voraussetzungen zur Berechnung dieser Varianzanalyse können als gegeben angenommen werden. Die Homogenität der Varianzen kann mit $p = .284$ angenommen werden. Im Folgenden (siehe Tabelle 43) ist die entsprechende Tafel der Varianzanalyse angeführt.

Tabelle 43: Tests der Zwischensubjekteffekte – AV Untertest Quantitative und formale Probleme

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	167.490 ^a	23	7.282	2.689	≤.0001**	.055
Konstanter Term	9033.429	1	9033.429	3336.241	≤.0001**	.757
Geschlecht	58.182	1	58.182	21.488	≤.0001**	.020
Stereotype	8.686	3	2.895	1.069	.361	.003
Risikoinstruktion	36.899	2	18.449	6.814	.001**	.013
Geschlecht * Stereotype	4.168	3	1.389	0.513	.673	.001
Geschlecht * Risikoinstruktion	4.543	2	2.272	0.839	.432	.002
Stereotype * Risikoinstruktion	26.359	6	4.393	1.622	.137	.009
Geschlecht * Stereotype * Risikoinstruktion	23.168	6	3.861	1.426	.201	.008
Fehler	2894.496	1069	2.708			
Gesamt	12923.000	1093				
Korrigierte Gesamtvariation	3061.985	1092				

a. R-Quadrat = .055 (korrigiertes R-Quadrat = .034)

** . $p < .01$

* . $p < .05$

Die Prüfung der $H_1(1)$ ergibt mit $F(1,1069) = 21.488$, $p \leq .0001$ ein signifikantes Ergebnis. Es kann ein signifikanter Unterschied im Untertest *Quantitative und formale Probleme* in Abhängigkeit vom Geschlecht angenommen werden.

Wie aus Tabelle 42 ersichtlich erbringen Männer mit $M = 3.34$ ($SD = 1.78$), im Untertest *Quantitative und formale Probleme* eine bessere Leistung als Frauen mit $M = 2.80$ ($SD = 1.57$). Mit $d = 0.32$ weist laut Bortz und Döring (2006, S. 606) die Berechnung der entsprechenden standardisierten Effektgröße auf einen kleinen Effekt aufgrund des Geschlechts hin.

Außerdem kann ein signifikanter Unterschied ($H_1(3)$) im Untertest *Quantitative und formale Probleme* in Abhängigkeit von der Risikoinstruktion $F(2,1069) = 6.814$, $p = .001$ angenommen werden. Wie in Abbildung 17 ersichtlich ist, ist in allen drei Risikobedingungen ein Niveauunterschied zugunsten der männlichen Testteilnehmer zu beobachten.

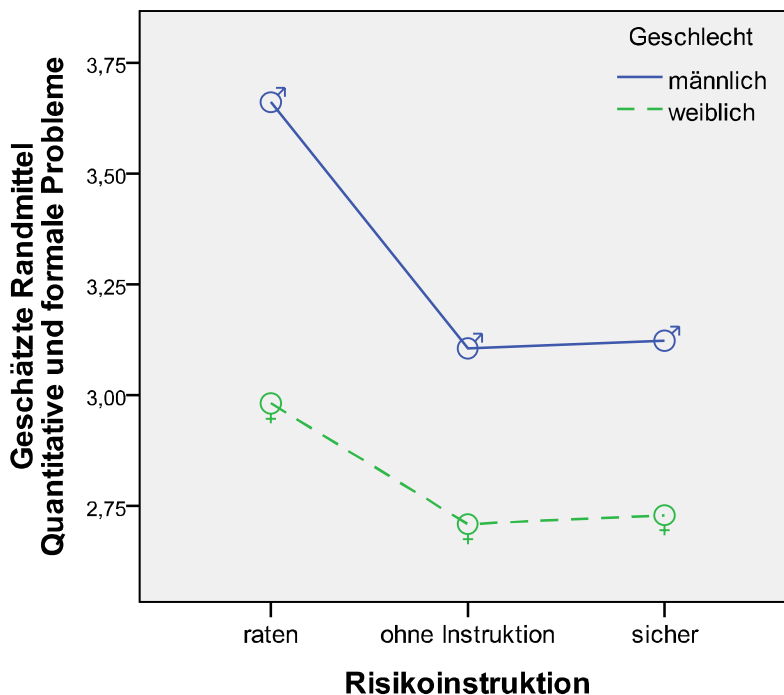


Abbildung 17: Profildiagramm zum Mittelwertsverlauf im Untertest Quantitative und formale Probleme in Abhängigkeit von Geschlecht und Risikoinstruktion

Da vermutet wird, dass die Risikobedingung *raten*, die Testpersonen beim Lösen der Aufgaben positiv unterstützt, werden mittels a-priori-Kontraste mögliche Mittelwertsunterschiede in Abhängigkeit der Risikoinstruktion untersucht (siehe Tabelle 44).

Tabelle 44: Kontrastergebnisse im Untertest Quantitative und formale Probleme in Abhängigkeit der Risikoinstruktion

Risikoinstruktion – einfacher Kontrast	Signifikanz	
	raten vs. sicher	$p = .002$
ohne Instruktion vs. sicher	$p = .889$	n.s.
ohne Instruktion vs. raten	$p = .001$	sig.

Die Vermutung kann bestätigt werden. Man kann annehmen, dass Testpersonen in der Risikobedingung *raten* beim Untertest *Quantitative und formale Probleme* signifikant höhere Werte erzielen als in den beiden anderen Risikobedingungen.

Die Prüfung aller weiteren Hypothesen ($H_1(2, 4, 5, 6, 7)$) fällt nicht signifikant aus (siehe Tabelle 43).

4.4.4.1.3 Hypothesenblock zum Untertest Medizinisch-naturwissenschaftliches Grundverständnis

$H_1(8)$: Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit vom Geschlecht.

$H_1(9)$: Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit von den vier Stereotype Threat-Bedingungen *Aktivierung, Neutrale Bedingung, Aufklärung* und *Negativaktivierung*.

H₁(10): Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit von den drei Risikobedingungen *raten*, *ohne Instruktion* und *sicher*.

H₁(11): Es gibt im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* eine signifikante Wechselwirkung zwischen Geschlecht und Stereotype Threat-Bedingung.

H₁(12): Es gibt im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* eine signifikante Wechselwirkung zwischen Geschlecht und Risikobedingung.

H₁(13): Es gibt im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* eine signifikante Wechselwirkung zwischen Stereotype Threat-Bedingung und Risikobedingung.

H₁(14): Es gibt im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* eine signifikante Interaktion höherer Ordnung zwischen Geschlecht, Stereotype Threat-Bedingung und Risikobedingung.

4.4.4.1.4 Hypothesenprüfung zum Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

Die Überprüfung des EMS-ähnlichen Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* erfolgt mittels univariater dreifaktorieller Varianzanalyse.

Unabhängige Variablen (UV):

- Geschlecht (2)
- Stereotype Threat-Bedingung (4)
- Risikobedingung (3)

Abhängige Variable (AV):

- Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

Tabelle 45 fasst die Mittelwerte und zugehörigen Standardabweichungen im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*, separiert

nach den Stereotype Threat- und Risikobedingungen sowie den Geschlechtern zusammen.

Tabelle 45: Deskriptive Statistiken – Kennwerte AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis

Stereotype Threat-Bedingung	Risiko-bedingung	männlich			weiblich			Gesamt		
		M	SD	n	M	SD	n	M	SD	n
Aktivierung	Gesamt	2.94	1.56	123	2.68	1.54	170	2.79	1.55	293
Neutral	Gesamt	3.11	1.87	85	2.77	1.84	172	2.88	1.85	257
Aufklärung	Gesamt	3.19	1.72	112	2.69	1.72	168	2.89	1.74	280
Negativ-aktivierung	Gesamt	2.82	1.64	89	2.62	1.60	174	2.69	1.61	263
Gesamt	raten	3.34	1.73	145	2.81	1.73	228	3.02	1.75	373
	ohne Instruktion	2.89	1.64	131	2.51	1.62	237	2.65	1.64	368
	sicher	2.78	1.65	133	2.77	1.67	219	2.77	1.66	352
	Gesamt	3.02	1.69	409	2.69	1.68	684	2.81	1.69	1093

Erfüllte Voraussetzungen für die Varianzanalyseberechnung können angenommen werden: mit $p = .202$ kann die Varianzhomogenität bestätigt werden. Die entsprechende Tafel der Varianzanalyse ist in Tabelle 46 ersichtlich.

Tabelle 46: Tests der Zwischensubjekteffekte – AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	108.776 ^a	23	4.729	1.685	.023*	.035
Konstanter Term	7856.262	1	7856.262	2798.376	≤ .0001**	.724
Geschlecht	19.976	1	19.976	7.116	.008**	.007

Stereotype	8.717	3	2.906	1.035	.376	.003
Risikoinstruktion	28.659	2	14.330	5.104	.006**	.009
Geschlecht * Stereotype	2.456	3	0.819	0.292	.831	.001
Geschlecht * Risikoinstruktion	11.111	2	5.555	1.979	.139	.004
Stereotype * Risikoinstruktion	27.140	6	4.523	1.611	.141	.009
Geschlecht * Stereotype * Risikoinstruktion	5.167	6	0.861	0.307	.934	.002
Fehler	3001.149	1069	2.807			
Gesamt	11761.000	1093				
Korrigierte Gesamtvariation	3109.925	1092				

a. R-Quadrat = .035 (korrigiertes R-Quadrat = .014)

** . $p < .01$

* . $p < .05$

Bei der Überprüfung der $H_1(8)$ ergibt sich mit $F(1,1069) = 7.116$, $p = .008$ ein signifikantes Resultat. Dies bedeutet, dass ein Zusammenhang zwischen der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* und dem Geschlecht beobachtet werden kann.

Die männlichen Teilnehmer weisen aufgrund des Mittelwerts $M = 3.02$ ($SD = 1.69$) im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* höhere Leistungen auf als die weiblichen Testpersonen ($M = 2.69$, $SD = 1.68$; siehe Tabelle 45). Die Berechnung der entsprechenden standardisierten Effektgröße weist mit $d = 0.20$ auf einen kleinen Effekt hin (Bortz & Döring, 2006, S. 606).

Mit $F(2,1069) = 5.104$, $p = .006$ ergibt die Prüfung der $H_1(10)$ ein signifikantes Ergebnis. Ein signifikanter Unterschied im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit der Risikoinstruktion kann angenommen werden.

Abbildung 18 veranschaulicht den Verlauf der Mittelwerte im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit von Geschlecht und Risikoinstruktion. Es ist ersichtlich, dass die Schüler hier in den

Risikobedingungen *raten* und *ohne Instruktion* höhere Leistungen erbringen als die Schülerinnen, während in der Risikobedingung *sicher* kein Unterschied zwischen den Geschlechtern zu beobachten ist.

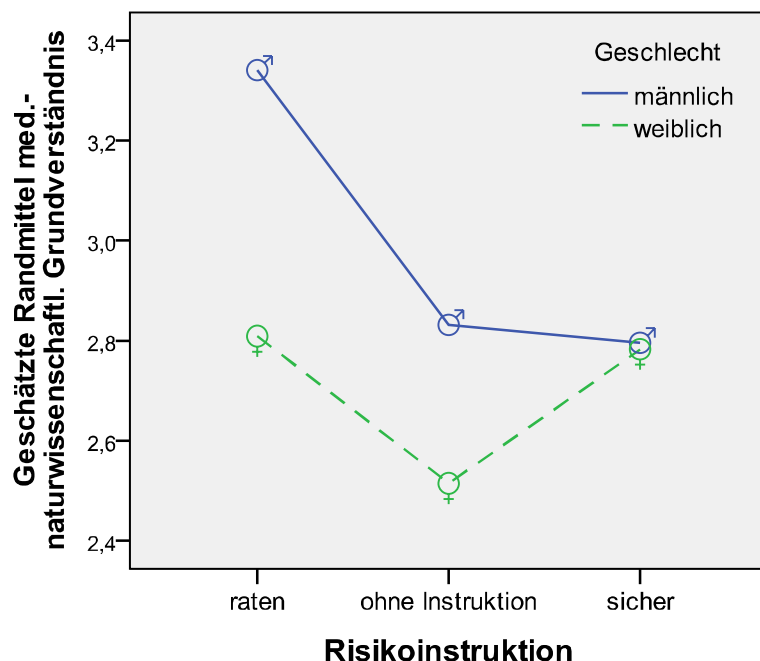


Abbildung 18: Mittelwertsverlaufdiagramm im Untertest Medizinisch-naturwissenschaftliches Grundverständnis in Abhängigkeit von Geschlecht und Risikoinstruktion

Aufgrund der Erwartung, dass die Risikobedingung *raten* für das Lösen der Aufgaben förderlich ist, werden a-priori-Kontraste berechnet. Diese sollen untersuchen, zwischen welchen Risikoinstruktionen Unterschiede anzunehmen sind (siehe Tabelle 47).

Tabelle 47: Kontrastergebnisse im Untertest Medizinisch-naturwissenschaftliches Grundverständnis in Abhängigkeit der Risikoinstruktion

Risikoinstruktion – einfacher Kontrast	Signifikanz	
ohne Instruktion vs. raten	$p = .002$	sig.
sicher vs. raten	$p = .029$	sig.
ohne Instruktion vs. sicher	$p = .390$	n.s.

Es kann angenommen werden, dass unter der Risikoinstruktion *raten* signifikant höhere medizinisch-naturwissenschaftliche Grundverständnisleistungen erreicht werden als unter den anderen Risikoinstruktionen.

Wie aus Tabelle 46 ersichtlich, fallen alle weiteren Haupteffekte und Wechselwirkungen nicht signifikant aus.

4.4.4.1.5 Hypothesenblock zum Untertest *Diagramme und Tabellen*

H₁(15): Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen* in Abhängigkeit vom Geschlecht.

H₁(16): Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen* in Abhängigkeit von den vier Stereotype Threat-Bedingungen *Aktivierung, Neutrale Bedingung, Aufklärung* und *Negativaktivierung*.

H₁(17): Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen* in Abhängigkeit von den drei Risikobedingungen *raten, ohne Instruktion* und *sicher*.

H₁(18): Es gibt im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung zwischen Geschlecht und Stereotype Threat-Bedingung.

H₁(19): Es gibt im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung zwischen Geschlecht und Risikobedingung.

H₁(20): Es gibt im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung zwischen Stereotype Threat-Bedingung und Risikobedingung.

H₁(21): Es gibt im Untertest *Diagramme und Tabellen* eine signifikante Interaktion höherer Ordnung zwischen Geschlecht, Stereotype Threat-Bedingung und Risikobedingung.

4.4.4.1.6 Hypothesenprüfung zum Untertest Diagramme und Tabellen

Auch hier wird die Prüfung des EMS-ähnlichen Untertests mit Hilfe einer univariaten dreifaktoriellen Varianzanalyse durchgeführt.

- Unabhängige Variablen (UV):
- Geschlecht (2)
 - Stereotype Threat-Bedingung (4)
 - Risikobedingung (3)

- Abhängige Variable (AV):
- Untertest *Diagramme und Tabellen*

Einen Überblick über die Testkennwerte im Untertest *Diagramme und Tabellen*, getrennt nach Geschlecht, sowie den Stereotype Threat- und Risiko-bedingungen, gibt nachfolgende Tabelle 48.

Tabelle 48: Deskriptive Statistiken – Kennwerte AV Untertest Diagramme und Tabellen

Stereotype Threat-Bedingung	Risiko-bedingung	männlich			weiblich			Gesamt		
		M	SD	n	M	SD	n	M	SD	n
Aktivierung	Gesamt	3.33	1.68	123	2.94	1.65	170	3.11	1.67	293
Neutral	Gesamt	3.40	1.79	85	2.87	1.644	172	3.04	1.71	257
Aufklärung	Gesamt	3.15	1.61	112	2.86	1.64	168	2.98	1.63	280
Negativ-aktivierung	Gesamt	3.08	1.60	89	2.80	1.47	174	2.90	1.52	263
Gesamt	raten	3.50	1.68	145	3.02	1.62	228	3.21	1.66	373
	ohne Instruktion	3.06	1.53	131	2.89	1.56	237	2.95	1.55	368
	sicher	3.14	1.76	133	2.68	1.60	219	2.85	1.67	352
	Gesamt	3.24	1.67	409	2.87	1.60	684	3.01	1.63	1093

Die zur Berechnung der Varianzanalyse erforderlichen Voraussetzungen können als gegeben betrachtet werden. Die Voraussetzung der Varianzhomogenität kann mit einem nicht signifikanten Ergebnis im Levene-

Test $p = .137$ angenommen werden. Tabelle 49 zeigt die diesbezügliche Tafel der Varianzanalyse.

Tabelle 49: Tests der Zwischensubjekteffekte – AV Untertest Diagramme und Tabellen

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	144.957 ^a	23	6.302	2.435	≤ .0001**	.050
Konstanter Term	8959.849	1	8959.849	3461.558	≤ .0001**	.764
Geschlecht	27.131	1	27.131	10.482	.001**	.010
Stereotype	8.510	3	2.837	1.096	.350	.003
Risikoinstruktion	29.615	2	14.808	5.721	.003**	.011
Geschlecht * Stereotype	2.663	3	0.888	0.343	.794	.001
Geschlecht * Risikoinstruktion	7.526	2	3.763	1.454	.234	.003
Stereotype * Risikoinstruktion	38.733	6	6.455	2.494	.021*	.014
Geschlecht * Stereotype * Risikoinstruktion	16.667	6	2.778	1.073	.377	.006
Fehler	2766.985	1069	2.588			
Gesamt	12797.000	1093				
Korrigierte Gesamtvariation	2911.941	1092				

a. R-Quadrat = .050 (korrigiertes R-Quadrat = .029)

** . $p < .01$

* . $p < .05$

Die Berechnung der entsprechenden Prüfgröße ergibt mit $F(1,1069) = 10.482$, $p = .001$ ein signifikantes Ergebnis, woraus folgt, dass entsprechend der $H_1(15)$, ein signifikanter Unterschied im Untertest *Diagramme und Tabellen*, abhängig vom Geschlecht, angenommen werden kann.

Wie sich aus Tabelle 48 zeigt, erreichen die männlichen Schüler mit $M = 3.24$ ($SD = 1.67$) einen höheren Wert im Untertest *Diagramme und Tabellen* im Vergleich zu den weiblichen Schülerinnen ($M = 2.87$, $SD = 1.60$). Mit $d = 0.23$

kann mittels Berechnung der entsprechenden standardisierter Effektgröße auf einen kleinen Effekt geschlossen werden (Bortz & Döring, 2006, S. 606).

Die Überprüfung der $H_1(17)$ erbringt mit $F(2,1069) = 5.721$, $p = .003$ ein signifikantes Ergebnis. Im Untertest *Diagramme und Tabellen* kann somit ein signifikanter Unterschied, abhängig von der Risikobedingung, beobachtet werden.

In Abbildung 19 ist zu erkennen, dass die männlichen Testpersonen unter allen drei Risikobedingungen höhere Leistungen im Untertest *Diagramme und Tabellen* erzielen als die weiblichen Untersuchungsteilnehmer.

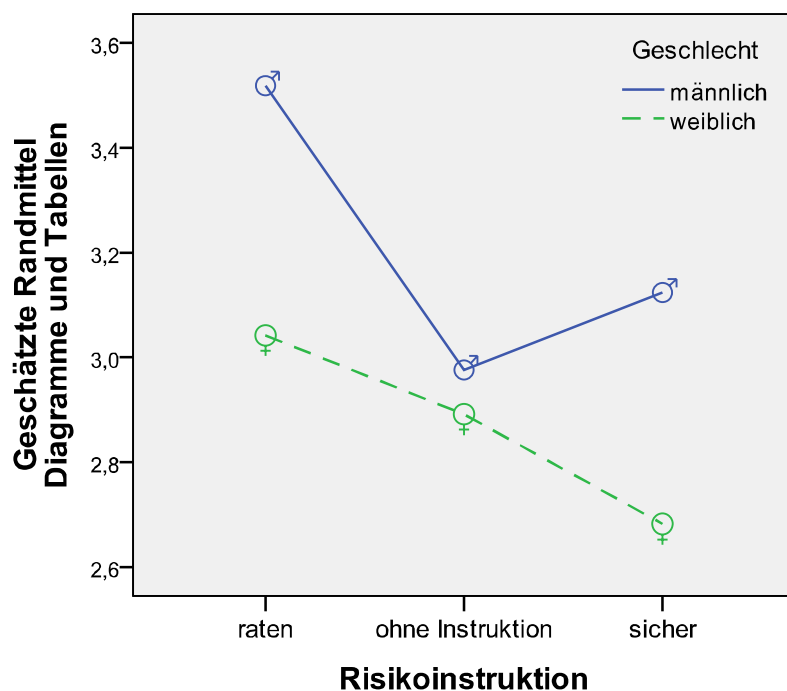


Abbildung 19: Mittelwertsverlaufdiagramm im Untertest Diagramme und Tabellen in Abhängigkeit von Geschlecht und Risikoinstruktion

Es kann auch hier die Vermutung aufgestellt werden, dass sich die Risikobedingung *raten* von den beiden anderen Risikobedingungen unterscheidet, wobei dies mittels a-priori-Kontraste untersucht wird (siehe Tabelle 50).

Tabelle 50: Kontrastergebnisse im Untertest Diagramme und Tabellen in Abhängigkeit der Risikoinstruktion

Risikoinstruktion – einfacher Kontrast	Signifikanz	
	raten vs. sicher	$p = .003$
ohne Instruktion vs. sicher	$p = .811$	n.s.
ohne Instruktion vs. raten	$p = .006$	sig.

Die Vermutung, dass unter der Risikobedingung *raten* eine signifikant höhere Leistung im Untertest *Diagramme und Tabellen* anzunehmen ist, kann bestätigt werden. Wie in den Untertests *Quantitative und formale Probleme* und *Medizinisch-naturwissenschaftliches Grundverständnis* zeigt sich auch hier, dass sich die Risikobedingung *raten* signifikant von den Risikobedingungen *ohne Instruktion* und *sicher*, unterscheidet.

Darüber hinaus kann mit $F(6,1069) = 2.494$, $p = .021$ eine signifikante Wechselwirkung aus Stereotype Threat-Bedingung und Risikoinstruktion beobachtet werden ($H_1(20)$).

Abbildung 20 veranschaulicht, in Abhängigkeit von Stereotype Threat- und Risikobedingung, die Wechselwirkung zum Untertest *Diagramme und Tabellen*.

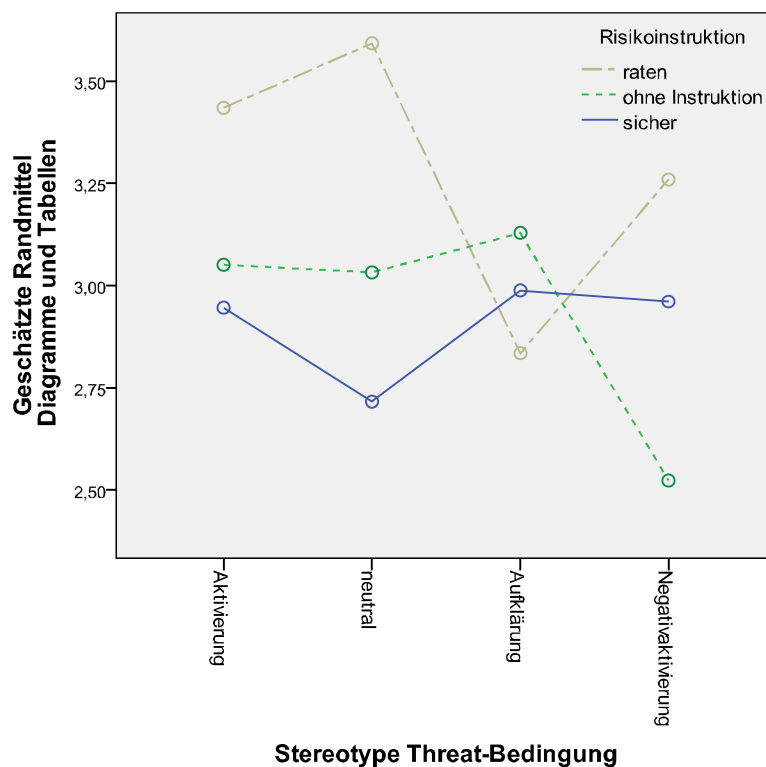


Abbildung 20: Wechselwirkungsdiagramm zum Untertest Diagramme und Tabellen in Abhängigkeit von Stereotype Threat- und Risikobedingung

Bei dieser Interaktion sind Differenzen innerhalb der Risikobedingungen in Abhängigkeit von der Stereotype Threat-Bedingung signifikant unterschiedlich. Somit sind die Unterschiede aufgrund des Hauptfaktors Risikobedingung nur eingeschränkt interpretierbar.

Unter der Stereotype Threat-Bedingung *Aufklärung* sind Unterschiede im Risikoverhalten verschieden im Vergleich zur Bedingung *Neutral* oder *Negativaktivierung*. Die Risikoinstruktion *raten* ist unter der Stereotype Threat-Bedingung *Aufklärung* ist nicht so wirksam wie unter den anderen drei Stereotype Threat-Bedingungen. Bestimmte Faktorstufenkombinationen ergeben überadditive Effekte, sie können leistungsfördernd bzw. leistungshemmend wirken.

Alle anderen Haupteffekte und Wechselwirkungen, so auch in Tabelle 49 ersichtlich, fallen nicht signifikant aus.

4.4.4.2 Nebenhypothesen zu den EMS-ähnlichen Aufgaben

Der folgende Teilabschnitt behandelt die Nebenhypothesen zu den EMS-ähnlichen Untertests mit Bezugnahme zur Absicht, ein Medizinstudium zu beginnen, zur tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben und zum Schulzweig.

Tabelle 51 zeigt, wie viele Personen bereits vor dem Erhebungszeitpunkt Testaufgaben des EMS bearbeitet haben, in Abhängigkeit der Absicht, Medizin zu studieren. 203 Personen (61 Schüler und 142 Schülerinnen von 1085 gültigen Fällen; siehe Tabelle 17) haben sich bereits in irgendeiner Form mit dem EMS auseinandergesetzt. 53 Personen (5,0%), bei 1066 gültigen Fällen hatten bereits zum Untersuchungszeitpunkt Testaufgaben des EMS bearbeitet, wovon 34 SchülerInnen auch ein Medizinstudium beginnen wollten.

Tabelle 51: Kreuztabelle Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet

			Auseinandersetzung mit EMS: Testaufgaben bearbeitet		Gesamt
			nein	ja	
Medizinstudium	nein	Anzahl	860	19	879
		erwartet	835.3	43.7	879.0
		% innerhalb	97,8%	2,2%	100,0%
		Stand. Residuen	.9	-3.7	
	ja	Anzahl	153	34	187
		erwartet	177.7	9.3	187.0
		% innerhalb	81,8%	18,2%	100,0%
		Stand. Residuen	-1.9	8.1	
Gesamt		Anzahl	1013	53	1066
		erwartet	1013.0	53.0	1066.0
		% innerhalb	95,0%	5,0%	100,0%

Mit $\chi^2 (1) = 83.761$, $p \leq .0001$ ergibt die Berechnung der entsprechenden Prüfgröße ein signifikantes Ergebnis und weist somit auf einen deutlichen Zusammenhang zwischen der Absicht Medizin zu studieren und der tatsächlichen Auseinandersetzung mit dem EMS, im Sinne der Bearbeitung von Testaufgaben.

4.4.4.2.1 Nebenhypothesen zum Untertest *Quantitative und formale Probleme*

H₁(22): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen.

H₁(23): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme*, in Abhängigkeit der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

H₁(24): Es gibt signifikante Unterschiede in der Leistung im Untertest *Quantitative und formale Probleme*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen und der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

4.4.4.2.2 Nebenhypothesenprüfung zum Untertest *Quantitative und formale Probleme*

Unabhängige Variablen (UV):

- Absicht, ein Medizinstudium zu beginnen (2)
- Auseinandersetzung mit dem EMS, hinsichtlich Bearbeitung von Testaufgaben (2)

Abhängige Variable (AV):

- Untertest *Quantitative und formale Probleme*

Die Überprüfung der Hypothesen wird durch eine univariate zweifaktorielle Varianzanalyse vorgenommen. Die Voraussetzung der homogenen Varianzen kann mit $p = .982$ im Levene-Test als erfüllt betrachtet werden.

Die Prüfungen der $H_1(22)$ ($p = .716$), der $H_1(23)$ ($p = .133$) und der $H_1(24)$ ($p = .889$) ergeben nicht signifikante Ergebnisse. Es können keine signifikanten Unterschiede und Wechselwirkungen in der Leistung im Untertest *Quantitative und formale Probleme*, in Abhängigkeit der Absicht, ein Medizinstudium zu beginnen und der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben, beobachtet werden. Tabelle 52 zeigt die entsprechende Tafel der Varianzanalyse.

Tabelle 52: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Quantitative und formale Probleme

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	9.107 ^a	3	3.036	1.101	.348
Konstanter Term	1816.101	1	1816.101	658.516	≤ .0001**
Medizinstudium	0.366	1	0.366	0.133	.716
Auseinandersetzung mit EMS: Testaufgaben	6.227	1	6.227	2.258	.133
Medizin * Auseinandersetzung mit EMS: Testaufgaben	0.054	1	0.054	0.019	.889
Fehler	2928.859	1062	2.758		
Gesamt	12568.000	1066			
Korrigierte Gesamtvariation	2937.966	1065			

a. R-Quadrat = .003 (korrigiertes R-Quadrat = .000)

** . $p < .01$

* . $p < .05$

4.4.4.2.3 Nebenhypothesen zum Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

H₁(25): Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen.

H₁(26): Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*, in Abhängigkeit der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

H₁(27): Es gibt signifikante Unterschiede in der Leistung im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen und der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

4.4.4.2.4 Nebenhypothesenprüfung zum Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

Unabhängige Variablen (UV):

- Absicht, ein Medizinstudium zu beginnen (2)
- Auseinandersetzung mit dem EMS, hinsichtlich Bearbeitung von Testaufgaben (2)

Abhängige Variable (AV):

- Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*

Die Hypothesen werden anhand einer univariate zweifaktorielle Varianzanalyse überprüft. Die Gleichheit der Varianzen kann mit $p = .313$ im Levene-Test angenommen werden.

Die Prüfungen der H₁(25), H₁(26) und H₁(27) weisen keine signifikanten Resultate auf. Es können keine signifikanten Unterschiede in der Leistung im

Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*, in Abhängigkeit von der Absicht Medizin zu studieren ($p = .838$) und ebenso nicht aufgrund der Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben ($p = .309$), beobachtet werden. Außerdem sind diesbezüglich keine signifikanten Wechselwirkungen in diesem Untertest festzustellen ($p = .831$). Die entsprechende Tafel der Varianzanalyse zeigt Tabelle 53.

Tabelle 53: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	3.087 ^a	3	1.029	0.364	.779
Konstanter Term	1524.849	1	1524.849	539.349	≤ .0001**
Medizinstudium	0.119	1	0.119	0.042	.838
Auseinandersetzung mit EMS: Testaufgaben	2.927	1	2.927	1.035	.309
Medizin * Auseinandersetzung mit EMS: Testaufgaben	0.128	1	0.128	0.045	.831
Fehler	3002.490	1062	2.827		
Gesamt	11409.000	1066			
Korrigierte Gesamtvariation	3005.577	1065			

a. R-Quadrat = .001 (korrigiertes R-Quadrat = -.002)

** . $p < .01$

* . $p < .05$

4.4.4.2.5 Nebenhypothesen zum Untertest *Diagramme und Tabellen*

H₁(28): Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen.

H₁(29): Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen*, in Abhängigkeit der tatsächlichen

Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

$H_1(30)$: Es gibt signifikante Unterschiede in der Leistung im Untertest *Diagramme und Tabellen*, in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen und der tatsächlichen Auseinandersetzung mit dem EMS, hinsichtlich der Bearbeitung von Testaufgaben.

4.4.4.2.6 Nebenhypothesenprüfung zum Untertest *Diagramme und Tabellen*

Unabhängige Variablen (UV):

- Absicht, ein Medizinstudium zu beginnen (2)
- Auseinandersetzung mit dem EMS, hinsichtlich Bearbeitung von Testaufgaben (2)

Abhängige Variable (AV):

- Untertest *Diagramme und Tabellen*

Die Überprüfung der Hypothesen wird mittels univariater zweifaktorieller Varianzanalyse durchgeführt. Homogene Varianzen können angesichts $p = .441$ im Levene-Test angenommen werden. Die Voraussetzungen zur Berechnung dieser Varianzanalyse können als erfüllt betrachtet werden.

Auch die Überprüfungen der $H_1(28)$ mit $p = .375$, der $H_1(29)$ mit $p = .746$ und der $H_1(30)$ mit $p = .750$, zeigen keine signifikanten Ergebnisse auf. Bezüglich der Leistung im Untertest *Diagramme und Tabellen* können keine signifikanten Unterschiede, abhängig von der Absicht, Medizin zu studieren oder abhängig davon, ob jemand bereits Testaufgaben bearbeitet hat, aufgezeigt werden. Auch kann keine signifikante Wechselwirkung im Untertest *Diagramme und Tabellen* angenommen werden. Tabelle 54 zeigt die diesbezügliche Tafel der Varianzanalyse.

Tabelle 54: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Diagramme und Tabellen

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Korrigiertes Modell	4.980 ^a	3	1.660	0.626	.598
Konstanter Term	1721.713	1	1721.713	649.270	≤ .0001**
Medizinstudium	2.085	1	2.085	0.786	.375
Auseinandersetzung mit EMS: Testaufgaben	0.278	1	0.278	0.105	.746
Medizin * Auseinandersetzung mit EMS: Testaufgaben	0.270	1	0.270	0.102	.750
Fehler	2816.175	1062	2.652		
Gesamt	12596.000	1066			
Korrigierte Gesamtvariation	2821.156	1065			

a. R-Quadrat = .002 (korrigiertes R-Quadrat = -.001)

** . $p < .01$

* . $p < .05$

4.4.4.2.7 Nebenhypothese zu den EMS-ähnlichen Leistungen in Abhängigkeit des Schulzweiges sowie Nebenthesenprüfung

H₁(31): Es gibt signifikante Unterschiede in den Leistungen betreffend der drei EMS-ähnlichen Aufgabenbereiche *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* (= Untertests zum schlussfolgernden Denken) in Abhängigkeit des Schulzweiges (naturwissenschaftliche vs. nicht-naturwissenschaftliche Schulzweige).

Unabhängige Variable (UV): ● Schulzweig (2)

Abhängige Variablen (AV): ● Untertest *Quantitative und formale Probleme*
 ● Untertest *Medizinisch-naturwissenschaftliches Grundverständnis*
 ● Untertest *Diagramme und Tabellen*

Tabelle 55 zeigt eine Übersicht der Mittelwerte und zugehörigen Standardabweichungen in den Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* sowie *Diagramme und Tabellen* nach Schulzweig.

Tabelle 55: Deskriptivstatistik der Kennwerte in den EMS-ähnlichen Bereichen *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* sowie *Diagramme und Tabellen* nach Schulzweig

EMS-ähnlicher Bereich	Schulzweige	M	SD	n
Quantitative und formale Probleme	naturwissensch. Schulzweige	3.29	1.72	247
	nicht-naturwissensch. Schulzweige	2.92	1.65	833
	Gesamt	3.01	1.67	1080
Med-nat.wissensch. Grundverständnis	naturwissensch. Schulzweige	2.87	1.72	247
	nicht-naturwissensch. Schulzweige	2.81	1.68	833
	Gesamt	2.82	1.69	1080
Diagramme und Tabellen	naturwissensch. Schulzweige	3.11	1.65	247
	nicht-naturwissensch. Schulzweige	2.98	1.63	833
	Gesamt	3.01	1.63	1080

Die Prüfung der Hypothese wird mittels multivariater Varianzanalyse durchgeführt. Die Homogenität der Varianzen kann mit $p \geq .493$ angenommen werden. Die Homogenität der Varianz-Kovarianzenmatrizen der abhängigen Variablen, überprüft mittels Box-M-Test, kann mit $p = .592$ angenommen werden.

Die Prüfung mittels MANOVA zeigt im Untertest *Quantitative und formale Probleme* mit $p = .003$ signifikante Unterschiede in Abhängigkeit des Schulzweiges. SchülerInnen im naturwissenschaftlichen Schulzweig erbringen hier bessere Leistungen ($M = 3.29$, $SD = 1.72$), als SchülerInnen der anderen, nicht-naturwissenschaftlichen Schulzweige ($M = 2.92$, $SD = 1.65$; siehe Tabelle 55). Mit $d = 0.22$ weist die Berechnung der entsprechenden standardisierten Effektgröße auf einen kleinen Effekt hin (Bortz & Döring, 2006, S. 606).

In den Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* ($p = .595$) und *Diagramme und Tabellen* ($p = .271$) zeigen sich keine Unterschiede in Abhängigkeit des Schulzweiges. Tabelle 56 zeigt die diesbezügliche Tafel der multivariaten Varianzanalyse.

Tabelle 56: Tests der Zwischensubjekteffekte – Schulzweig – AV Untertest Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis sowie Diagramme und Tabellen

Quelle	Abhängige Variable	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	Quant. u. formale Probleme	25.114a	1	25.114	9.095	.003**	.008
	Med-nat. Grundverständnis	0.803b	1	0.803	0.283	.595	.000
	Diagramme u. Tabellen	3.228c	1	3.228	1.212	.271	.001
Konstanter Term	Quant. u. formale Probleme	7351.159	1	7351.159	2662.080	≤ .0001**	.712
	Med-nat. Grundverständnis	6137.592	1	6137.592	2160.539	≤ .0001**	.667
	Diagramme u. Tabellen	7080.876	1	7080.876	2659.099	≤ .0001**	.712
Schulzweig	Quant. u. formale Probleme	25.114	1	25.114	9.095	.003**	.008
	Med-nat. Grundverständnis	0.803	1	0.803	0.283	.595	.000
	Diagramme u. Tabellen	3.228	1	3.228	1.212	.271	.001
Fehler	Quant. u. formale Probleme	2976.826	1078	2.761			
	Med-nat. Grundverständnis	3062.349	1078	2.841			
	Diagramme u. Tabellen	2870.591	1078	2.663			
Gesamt	Quant. u. formale Probleme	12770.000	1080				
	Med-nat. Grundverständnis	11654.000	1080				
	Diagramme u. Tabellen	12678.000	1080				
Korrigierte	Quant. u. formale Probleme	3001.941	1079				

Gesamtvariation	Med-nat. Grundverständnis	3063.152	1079				
	Diagramme u. Tabellen	2873.819	1079				

a. R-Quadrat = .008 (korrigiertes R-Quadrat = .007)

b. R-Quadrat = .000 (korrigiertes R-Quadrat = -.001)

c. R-Quadrat = .001 (korrigiertes R-Quadrat = .000)

** . $p < .01$

* . $p < .05$

4.4.4.3 Nebenhypothesen zum GDT

H₁(32): Es gibt signifikante Unterschiede im Risikoverhalten, operationalisiert mittels GDT, in Abhängigkeit vom Geschlecht.

H₁(33): Es gibt signifikante Unterschiede im Risikoverhalten, operationalisiert mittels GDT, in Abhängigkeit von der vorangegangenen Risikoinstruktion.

H₁(34): Es gibt im Risikoverhalten, operationalisiert mittels GDT, eine signifikante Wechselwirkung zwischen dem Geschlecht und der vorangegangenen Risikoinstruktion.

4.4.4.3.1 Nebenhypothesenprüfung zum GDT

Mit Hilfe der univariaten zweifaktoriellen Varianzanalyse wird die Überprüfung der Nebenhypothesen den GDT betreffend, durchgeführt, wobei hier eine Teilstichprobe von 471 gültigen Fällen herangezogen werden kann. Die Normalverteilung der Daten je Gruppen kann angenommen werden.

Unabhängige Variablen (UV):

- Geschlecht (2)
- Risikobedingung (3)

Abhängige Variable (AV):

- Nettoscore des GDT

Einen Überblick über die Kennwerte des Nettoscore des GDT, getrennt nach Geschlecht sowie gesamt, zeigt Tabelle 57.

Tabelle 57: Deskriptive Statistiken – AV Nettoscore im GDT

Risiko- bedingung	Gesamt			weiblich			männlich		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
raten	2.72	9.67	157	3.66	9.60	96	1.25	9.71	61
ohne Instruktion	5.28	9.31	163	5.66	9.08	102	4.64	9.74	61
sicher	4.91	10.01	151	5.63	9.98	90	3.85	10.04	61
Gesamt	4.31	9.71	471	4.98	9.55	288	3.25	9.88	183

Die zur Berechnung dieser Varianzanalyse notwendigen Voraussetzungen können als erfüllt betrachtet werden, die Homogenität der Varianzen kann mit $p = .922$ angenommen werden.

In Tabelle 58 wird die entsprechende Tafel der Varianzanalyse veranschaulicht.

Tabelle 58: Tests der Zwischensubjekteffekte – AV Nettoscore im GDT

Quelle	Quadratsumme vom Typ III	<i>df</i>	Mittel der Quadrate	<i>F</i>	Sig.	Partielles Eta-Quadrat
Korrigiertes Modell	975.765 ^a	5	195.153	2.096	.065	.022
Konstanter Term	7567.904	1	7567.904	81.290	≤ .0001**	.149
Geschlecht	336.981	1	336.981	3.620	.058	.008
Risikoinstruktion	633.373	2	316.687	3.402	.034*	.014
Geschlecht * Risikoinstruktion	36.736	2	18.368	0.197	.821	.001
Fehler	43290.596	465	93.098			
Gesamt	53007.000	471				
Korrigierte Gesamtvariation	44266.361	470				

a. R-Quadrat = .022 (korrigiertes R-Quadrat = .012)

** . $p < .01$

* . $p < .05$

Allgemein gilt ein niedriger Werte im Nettoscore des GDT als risikoreiches Verhalten. Die niedrigsten Mittelwerte ($M = 2.72$, $SD = 9.67$) zeigen sich in der Risikobedingung *raten*, wobei hier die männlichen Teilnehmer mit $M = 1.25$ ($SD = 9.71$) niedrigere Werte aufzeigen als die weiblichen Teilnehmer ($M = 3.66$, $SD = 9.60$; siehe Tabelle 57). Die Berechnung der entsprechenden Prüfgröße ergibt mit $F(1,465) = 3.620$, $p = .058$ für $H_1(32)$ knapp keinen signifikanten Unterschied im Risikoverhalten, operationalisiert mittels GDT, in Abhängigkeit vom Geschlecht. Eine tendenzielle Signifikanz dahingehend, dass Männer tendenziell risikobereiter sind, kann jedoch angenommen werden. Die Berechnung der entsprechenden standardisierten Effektgröße weist mit $d = 0.18$ auf einen sehr geringen Effekt hin (Bortz & Döring, 2006, S. 606).

Mit $F(2,465) = 3.402$, $p = .034$ zeigt die Prüfung der $H_1(33)$ ein signifikantes Resultat, womit ein signifikanter Unterschied im Risikoverhalten, operationalisiert mittels GDT, in Abhängigkeit von der vorangegangenen Risikoinstruktion, angenommen werden kann.

Die Überprüfung der $H_1(34)$ ergibt mit $F(2,465) = 0.197$, $p = .821$ ein nicht signifikantes Ergebnis. Es kann im Risikoverhalten, erhoben mittels GDT, keine signifikante Wechselwirkung zwischen dem Geschlecht und der vorangegangenen Risikoinstruktion angenommen werden (siehe Abbildung 21).

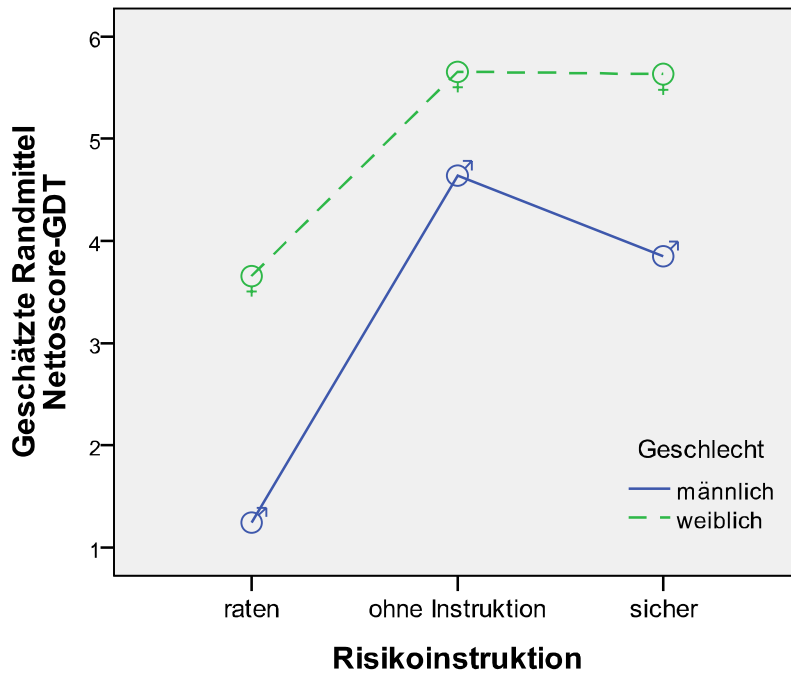


Abbildung 21: Mittelwertsverlaufdiagramm im Nettoscore des GDT in Abhängigkeit von Geschlecht und Risikoinstruktion

Aufgrund der Erwartung, dass die Risikobedingung *raten* für riskantes Verhalten förderlich ist, werden mögliche Mittelwertsunterschiede in Abhängigkeit der vorangegangenen Risikoinstruktion mit Hilfe von a-priori-Kontrasten untersucht.

Tabelle 59: Kontrastergebnisse im Nettoscore des GDT in Abhängigkeit der vorangegangenen Risikoinstruktion

Risikoinstruktion – einfacher Kontrast	Signifikanz	
	raten vs. sicher	$p = .042$
ohne Instruktion vs. sicher	$p = .717$	n.s.
ohne Instruktion vs. raten	$p = .016$	sig.

Wie aus Tabelle 59 ersichtlich, unterscheidet sich die Instruktion *raten* signifikant von den beiden anderen Risikoinstruktionen. Die TestteilnehmerInnen handeln hier, wie erwartet, risikofreudiger als in den

Risikoinstruktionen *sicher* und *ohne Instruktion*. Die Risikoinstruktion *raten* kann als wirksam angenommen werden.

5 Diskussion und Ausblick

In diesem Teil werden die Ergebnisse der vorliegenden Untersuchung zusammengefasst, interpretiert und diskutiert. Die Resultate der Hypothesenprüfungen werden mit der im Theorieteil beschriebenen wissenschaftlichen Literatur in Bezug gebracht. Zudem sollen auch kritische Anmerkungen zur Untersuchung und Aspekte des Erhebungsinstruments nicht außer Acht gelassen werden.

Ziel der Arbeit war es, Leistungsunterschiede zwischen Frauen und Männern in EMS-ähnlichen Aufgaben, insbesondere in jenen Aufgabenbereichen, die das schlussfolgernde Denken betreffen (*Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis und Diagramme und Tabellen*), zu untersuchen. Dabei waren auch die Einflüsse der sozialen Faktoren Stereotype Threat und Risikoverhalten von Interesse.

Unterschiede zwischen den Geschlechtern in kognitiven Aufgaben sind Gegenstand zahlreicher Untersuchungen (u. a. Hyde et al., 1990; Kuhn & Holling, 2009). Derartige Unterschiede zeigen sich auch im EMS, bei den Frauen in den Untertests *Quantitative und formale Probleme, Schlauchfiguren, Textverständnis, Planen und Organisieren, Medizinisch-naturwissenschaftliches Grundverständnis, Diagramme und Tabellen und Konzentriertes und sorgfältiges Arbeiten*, also auch in den drei Untertests zum schlussfolgernden Denken, schlechtere Leistungen erzielen als Männer (Mallinger et al., 2009). Im Studienverlauf werden jedoch diese Unterschiede von den weiblichen Studierenden wieder aufgeholt und sind nicht aufgrund mangelnder kognitiver Fähigkeiten zu erklären (Mitterauer et al., 2007). Für Geschlechterunterschiede in kognitiven Aufgabenbereichen existieren keine allgemeingültigen Erklärungen (Halpern et al., 2007). Um Leistungsunterschiede im EMS zu erläutern, bezieht die hier vorliegende Studie soziale Faktoren (Stereotype Threat und Risikoverhalten) mit ein.

Zur Überprüfung der Wirksamkeit der vier Stereotype Threat-Bedingungen (*Aktivierung, Neutrale Bedingung, Aufklärung und Negativaktivierung*) wurde

ähnlich wie bei der als Vorbild dienenden Untersuchung von Halpern und Tan (2001), der *Wissenstest* mit Fremd- und Selbsteinschätzungsaufgaben vorgegeben. Die TestteilnehmerInnen haben gemeinhin entsprechend den vorgegebenen Stereotype Threat-Bedingungen geantwortet, weshalb wirksame Stereotype Threat-Bedingungen angenommen werden. Die Ergebnisse weisen beispielsweise darauf hin, dass in der *Aktivierungsbedingung* den Frauen schlechtere Leistungen und in der *Negativaktivierungsbedingung* den Frauen bessere Leistungen in Aufgaben zu den Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* sowie *Diagramme und Tabellen*, von den Testpersonen zugeschrieben wurden. Ergebnisse aus der Literatur liefern demgemäß Befunde, dass Frauen durch das Phänomen Stereotype Threat (Aktivierung) in ihren kognitiven Leistungen negativ beeinträchtigt werden (u. a. Davies et al., 2002; Keller & Dauenheimer, 2003; Spencer et al., 1999).

Bezüglich der Selbsteinschätzung wird in der vorliegenden Untersuchung erkennbar, dass sich männliche Testteilnehmer in den Bereichen *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* signifikant besser einschätzen, was mit den Resultaten von Rammstedt und Rammsayer (2000) bezüglich Geschlechterunterschieden bei Selbsteinschätzung übereinstimmt. In weiterer Folge zeigt sich in der vorliegenden Arbeit, dass entsprechend der Selbsteinschätzung die männlichen Teilnehmer tatsächlich höhere Leistungen in den Untertests zum schlussfolgernden Denken erbringen als ihre weiblichen Kollegen. Überdies kann in der Selbsteinschätzung im Bereich *Diagramme und Tabellen* eine signifikante Wechselwirkung in Abhängigkeit vom Geschlecht und der Stereotype Threat-Bedingung beobachtet werden. Es kann angenommen werden, dass sich die männlichen Testpersonen in der *Negativaktivierung* schlechter einschätzen als in den restlichen drei Stereotype Threat-Bedingungen, im Unterschied zu den weiblichen Testpersonen, die sich darin besser als in den anderen Stereotype Threat-Bedingungen einschätzen. Dies deutet auf die Wirksamkeit der *Negativaktivierung* hin.

In allen drei Untertests zum schlussfolgernden Denken lassen sich signifikante Unterschiede in Abhängigkeit vom Geschlecht feststellen ($H_1(1)$, $H_1(8)$, $H_1(15)$). Die Unterschiede beruhen auf kleinen Effekten, welche aufgrund der Stichprobengröße statistisch signifikant ausfallen.

Im Untertest *Quantitative und formale Probleme*, der sich aus mathematischen Sachaufgaben zusammensetzt (Klieme, 1989; zitiert nach Trost et al., 1998, S. 51), sind Männer signifikant besser. Es zeigt sich hier ein kleiner Effekt zugunsten der Männer ($d = 0.32$). Verschiedene Studien belegen einen Vorteil der Männer im Bereich Mathematik (Halpern, 2000; Hyde et al., 1990; Jungwirth, 1998; Willingham & Cole, 1997). Die Ergebnisse gehen auch konform mit früheren EMS-Testergebnissen (Mallinger et al., 2010). Auch Spiel et al. (2008) zeigten in ihrer Untersuchung ähnliche Unterschiede hinsichtlich des Geschlechts. Indes belegten Hyde et al. (1990) einen weiblichen Vorteil für numerische Kalkulationen ($d = 0.14$). Dark & Benbow (1990) stellten die Vermutung auf, dass sich Frauen und Männer nicht per se in ihrer mathematischen Begabung unterscheiden, sondern jedes Geschlecht unterschiedliche Basiskompetenzen verwendet. Diese Basiskompetenzen erweisen sich, je nach Aufgabentyp, mehr oder weniger effektiv. In einer Arbeit von Geary (1996) konnten Geschlechterunterschiede reduziert werden, indem man beim Lösen eines mathematischen Problems dazu ermutigte räumliche Strategien anzuwenden. Es wird vermutet, dass der Vorteil der Männer im mathematischen Schlussfolgern, mit den beim Lösen der Aufgaben angewendeten visuell-räumlichen Strategien zusammenhängt (z.B. im Bereich Geometrie) (Casey et al., 1995; Halpern & LaMay, 2000). Wenn es von Vorteil ist, räumlich-basierende Lösungsstrategien anzuwenden, zeigten sich bei mathematischen Problemen Leistungsunterschiede zugunsten der Männer. Der männliche Vorteil war jedoch nicht zu beobachten, sobald es sich um verbale Lösungsstrategien oder ähnliche wie bereits bekannte Strategien aus Mathematik-Lehrbüchern, handelte (Gallagher et al., 2002). Dieser Wissensstand könnte bei der Konstruktion oder Revision der mathematischen Sachaufgaben im EMS hilfreich sein, um diese Leistungsunterschiede zu reduzieren.

Auch in den Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* ($d = 0.20$) und *Diagramme und Tabellen* ($d = 0.23$) welche gemäß Spiel et al. (2008) naturwissenschaftlich-medizinische Fähigkeiten erheben, zeigt sich, dass Männer signifikant bessere Leistungen erbringen als Frauen. In der Literatur lassen sich vergleichbare Ergebnisse finden, welche Leistungsunterschiede bezüglich schlussfolgerndem Denken und Naturwissenschaften belegen (Hedges & Nowell, 1995; zitiert nach Halpern, 2000, S. 126; Jungwirth, 1998; Kuhn & Holling, 2009; Meehan, 1984; zitiert nach Halpern, 2000, S. 126). Auch diese Ergebnisse stimmen mit bisherigen EMS-Testresultaten und Untersuchungen zum EMS überein (Mallinger et al., 2010; Spiel et al., 2008).

Verschiedene Studien begründen jedoch die Beeinträchtigung der kognitiven Leistungen von weiblichen Testpersonen mit dem Phänomen Stereotype Threat (Davies et al., 2002; Keller & Dauenheimer, 2003; Quinn & Spencer, 2001; Spencer et al., 1999). Johns et al. (2005) zeigten, dass der einfache Ansatz der Aufklärung über das Phänomen Stereotype Threat, kognitive Leistungsbeeinträchtigungen zu Lasten der Frauen reduzieren kann. Zahlreiche Untersuchungen belegen, dass bei Frauen denen gesagt wurde, dass der vorgegebene Mathematiktest bzw. Test zum schlussfolgernden Denken geschlechtersensitiv sei, also in einer Bedingung zur Aktivierung von Stereotype Threat, schlechtere Leistungen zu verzeichnen waren. Wurde ihnen mitgeteilt, der Test sei geschlechterneutral, so zeigten sich tatsächlich keine Differenzen zwischen den Geschlechtern (Keller & Dauenheimer, 2003; Quinn & Spencer, 2001; Régner et al., 2010; Spencer et al., 1999). Dementgegen lassen sich keine signifikanten Leistungsunterschiede in den Subtests zum schlussfolgernden Denken in Abhängigkeit von den Stereotype Threat-Bedingungen ($H_1(2)$, $H_1(9)$, $H_1(16)$) sowie auch nicht in Abhängigkeit vom Geschlecht und den Stereotype Threat-Bedingungen ($H_1(4)$, $H_1(11)$, $H_1(18)$) feststellen.

Die Vermutung, dass signifikante Leistungsunterschiede in Abhängigkeit von der Risikoinstruktion *raten*, *ohne Instruktion* und *sicher* bestehen, lassen sich in allen drei Untertests zum schlussfolgernden Denken bestätigen ($H_1(3)$, $H_1(10)$,

$H_1(17)$). Sowohl im Untertest *Quantitative und formale Probleme*, als auch in den Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* sowie *Diagramme und Tabellen* erbringen die Testpersonen in der Risikobedingung *raten* signifikant höhere Leistungen als in den anderen beiden Risikobedingungen. Die Untersuchung zeigt auch, dass die Instruktionen in Abhängigkeit vom Geschlecht durchaus unterschiedliche Wirksamkeit vorweisen. Bei den Ergebnissen zum Untertest *Quantitative und formale Probleme* in Abhängigkeit von Geschlecht und Risikoinstruktion zeigt sich in allen drei Risikobedingungen ein Unterschied zugunsten der Männer. Der Verlauf der Mittelwerte im Untertest *Medizinisch-naturwissenschaftliches Grundverständnis* in Abhängigkeit von Geschlecht und Risikoinstruktion weist auf, dass in den Risikobedingungen *raten* und *ohne Instruktion* höhere Leistungen seitens der männlichen Testpersonen zu beobachten sind. In der Risikobedingung *sicher* sind indes keine Unterschiede zwischen den männlichen und den weiblichen Untersuchungsteilnehmern zu erkennen. Auch im Untertest *Diagramme und Tabellen* wird mittels Mittelwertsverlaufdiagramm in Abhängigkeit von Geschlecht und Risikoinstruktion erkennbar, dass die männlichen Testpersonen unter allen drei Risikobedingungen bessere Leistungen erreichen als die weiblichen Untersuchungspersonen. Allgemein war die Risikoinstruktion *raten* für Männer förderlicher. Die Testpersonen zum Raten aufzufordern, wenn sie die richtige Lösung nicht wissen, stellt daher keine wünschenswerte Instruktion dar, da diese zugunsten der Männer ist und die Frauen benachteiligt. Dies zeigte auch die Untersuchung von Ben-Shakhar und Sinai (1991), welche zu dem Ergebnis kam, dass Frauen selbst bei expliziter Ermutigung zum Raten, weniger rieten und mehr Items ausließen als Männer.

Um den Geschlechterbias in den Untertests zum schlussfolgernden Denken im EMS zu verringern, sollte eine Aufforderung zum Raten in der Instruktion vermieden werden.

Spiel et al. (2008) merkten bereits an, dass um die Ratewahrscheinlichkeit zu beseitigen, bei den Multiple-Choice-Items im EMS unterschiedlich eine bis vier Antwortalternativen je Aufgabe richtig sein sollten. Dazu führte Hänsgen (2008)

an, dass die mehrfach gegebene Instruktion bei jeder Aufgabe eine Antwort zu markieren, also gegebenenfalls zu raten, den Einfluss der Ratewahrscheinlichkeit ausgleiche. Die Untersuchungsergebnisse von Ben-Shakhar und Sinai (1991) zeigten allerdings, dass die Aufforderung in der Instruktion, zu raten, sobald man die Antwort nicht sicher weiß, nicht ausreichend ist, um Unterschiede in der Ratetendenz aufzuheben.

Mögliche Geschlechterunterschiede sollten nicht aufgrund der Aufforderung zu raten begünstigt werden.

Eine Punkteverrechnung beim EMS, bei der nur die Markierung aller zutreffenden Alternativen als richtig gewertet wird, so wie von Spiel et al. (2008) gefordert, oder auch eine Vergabe von Teilpunkten erscheint sinnvoll.

Der GDT, der für die Untersuchung von Entscheidungen unter Risiko entwickelt wurde (Brand, Fujiwara et al., 2005) zeigt analog den oben diskutierten Ergebnissen zum Risikoverhalten, dass in Abhängigkeit von der vorangegangenen Risikoinstruktion ein signifikanter Unterschied im mittels GDT operationalisierten Risikoverhalten festgestellt werden kann ($H_1(33)$). Personen in der vorangegangenen Risikoinstruktion *raten* zeigen auch im GDT riskanteres Entscheidungsverhalten als in den Risikoinstruktionen *sicher* und *ohne Instruktion*. Ein knapp nicht signifikanter Unterschied wird im mittels GDT operationalisierten Risikoverhalten in Abhängigkeit vom Geschlecht sichtbar ($H_1(32)$). Es kann angenommen werden, dass Männer tendenziell risikobereiter sind, wobei niedrige Werte als risikofreudiges Verhalten verstanden werden. Auch Ben-Shakhar und Sinai (1991) wiesen Männern eine höhere Ratetendenz und somit ein risikofreudigeres Verhalten nach.

Zudem kann im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung aus Stereotype Threat-Bedingung und Risikoinstruktion beobachtet werden ($H_1(20)$). Durch die simultane Untersuchung der Faktoren, zeigt sich eine Interaktion. Unterschiede im Risikoverhalten erweisen sich in den Stereotype Threat-Bedingungen verschieden. Es gibt Faktorstufenkombinationen die leistungsfördernd bzw. leistungshemmend wirken. Als deutlich leistungsfördernd erweist sich beispielsweise die

Stereotype Threat-Bedingung *Neutral* in Kombination mit der Risikoinstruktion *raten*. Besonders leistungshemmend beweist sich die Instruktionkombination Stereotype Threat-Bedingung *Negativaktivierung* und Risikoinstruktion *ohne Instruktion*. Im Untertest *Diagramme und Tabellen* kann beobachtet werden, dass auch die *Aufklärung* über das Phänomen Stereotype Threat die Risikobereitschaft im Sinne des Ratens vermindert und bei den aufgeklärten Personen das Risikoverhalten offenbar keine Rolle spielt. Dieses Ergebnis kann ebenfalls wesentlich zur EMS-Gestaltung, beispielsweise im Rahmen von Revisionen, beitragen.

Entgegen der Untersuchung von Steele und Aronson (1995) zeigt sich im Untertest *Diagramme und Tabellen* unter der Stereotype Threat-Bedingung *Aktivierung* eine hohe Risikobereitschaft. Steele und Aronson (1995) hingegen führten an, dass stereotypisierte Personen (AfroamerikanerInnen) weniger Rateversuche machten bzw. weniger riskant arbeiteten wenn sie sich der stereotypisierten Gruppe zugehörig fühlten (ihre Rasse angeben mussten). Koivula et al. (2001) belegten, dass die Wirkung von Geschlechterstereotypen auf die Leistung reduziert werden konnte, sobald Personen dazu Angaben machen mussten, wie sicher sie sich in Bezug auf die Richtigkeit ihrer gegebenen Antworten waren.

Überdies wurde untersucht, ob es in den Leistungen der drei interessierenden Untertests Unterschiede, in Abhängigkeit der Absicht, ein Medizinstudium zu beginnen und der vorausgegangen Auseinandersetzung mit EMS-Material, hinsichtlich der Bearbeitung von Testaufgaben gibt ($H_1(24)$, $H_1(27)$, $H_1(30)$). Die nicht signifikanten Ergebnisse deuten darauf hin, dass die Absicht Medizin zu studieren und die Beschäftigung mit EMS-Material (Bearbeitung von Testaufgaben) keinen konfundierenden Einfluss auf die Untertestleistungen aufweisen. Es kann angenommen werden, dass die Auseinandersetzung mit dem EMS, im Sinne der Bearbeitung von Testaufgaben, als Störvariable nicht wirksam und die Ergebnisse davon unbeeinflusst sind. Die Leistungsunterschiede in den drei Untertests sind von der Beschäftigung mit dem Testmaterial als unabhängig zu beurteilen und dürften somit andere Ursachen haben. Nicht erhoben wurden jedoch Informationen über das

Ausmaß der Beschäftigung mit dem Testmaterial, auch welches Testmaterial verwendet wurde bleibt, unbekannt. Das Ausmaß kann sich von der Bearbeitung weniger Aufgaben bis hin zum absolvierten Vorbereitungskurs erstrecken. Da für das Lösen der Aufgaben kein spezielles medizinisches Vorwissen erforderlich ist, wird ohnedies auf eine geringe Trainierbarkeit der Aufgaben hingewiesen (Mallinger et al., 2009). Befürwortet werden aber ca. 20 Stunden Vorbereitung, diese ist jedoch vom Begriff Training zu unterscheiden (Hänsgen & Spicher, 2009).

Spiel et al. (2008) belegten in den von ihnen untersuchten EMS-TestteilnehmerInnen mit und ohne naturwissenschaftlichen Schulschwerpunkt keine Unterschiede in der Gesamtestleistung, wobei Unterschiede in der Gesamtleistung im Wesentlichen durch die Unterschiede in den Untertests repliziert wurden. Bei der hier vorliegenden Arbeit können in den einzelnen Untertests *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* ebenfalls keine Unterschiede in Abhängigkeit vom Schulzweig beobachtet werden ($H_1(31)$). Beim Untertest *Quantitative und formale Probleme* hingegen zeigen sich diesbezüglich signifikante Unterschiede. Im naturwissenschaftlichen Schulzweig erbringen SchülerInnen bessere Ergebnisse, als in nicht-naturwissenschaftlichen Schulzweigen ($d = 0.22$).

Eine bedeutende Rolle in Bezug auf die Leistungsunterschiede in den EMS-ähnlichen Aufgaben spielt die Speed-Komponente. Zeit ist eine Dimension die die Lösbarkeit beeinflusst. Nach Mäkitalo (1993, zitiert nach Åberg-Bengtsson, 1999, S. 566) arbeiten Frauen gewissenhafter und neigen weniger zum Raten, dies kann unter Zeitbeschränkung nachteilig für sie sein. Wird eine Zeitbeschränkung aufgehoben, so kann sich dies dahingehend positiv auswirken, dass sich die Leistungen von Frauen erhöhen (Rosser, 1989; zitiert nach Forbes, 1996, S. 85).

Aufgrund der zeitlichen Limitierung musste bei der Testung auf die Erhebung der modulierenden Variable Ängstlichkeit verzichtet werden. Um ein genaueres Bild zu diesem Themengebiet zu bekommen, wäre es zweifellos bedeutsam

auch die individuelle Ängstlichkeit mit einzubeziehen. Die vorliegende Untersuchung bietet eine fundierte Basis um weiterführende Untersuchungen anzulegen.

6 Zusammenfassung

Das Hauptziel der vorliegenden Diplomarbeit lag in der Untersuchung geschlechterspezifischer Leistungsunterschiede in kognitiven EMS-ähnlichen Aufgaben zum schlussfolgernden Denken, unter Berücksichtigung der sozialen Faktoren Stereotype Threat und Risikoverhalten. Die vorliegende Arbeit nimmt somit auf einen spezifischen und aktuellen Themenschwerpunkt Bezug, welcher Teil einer breit angelegten Studie ist. Ergebnisse zu den Themenbereichen Selbstwirksamkeitserwartung, Leistungsmotivation und Theory of Mind können in den Arbeiten von Anita Teufl, Gabriele Hangl und Katrin Anzirk eingesehen werden. Details zur Untersuchung geschlechterspezifischer Unterschiede in EMS-ähnlichen Aufgaben zur Merkfähigkeit sind der Arbeit von Helene Lager zu entnehmen.

Im Rahmen dieser breit angelegten Untersuchung wurden insgesamt 1093 SchülerInnen aus den 8. Klassen Allgemeinbildender Höherer Schulen untersucht. Die teilnehmenden SchülerInnen im Alter zwischen 16 und 21 Jahren konnten in Schulen in Wien, Niederösterreich und dem Burgenland getestet werden. Die Testbatterie beinhaltete neben einem soziodemographischen Fragebogen und Instruktionen zu den Stereotype Threat-Bedingungen und zum Risikoverhalten, den *Wissenstest* und sechs EMS-ähnliche Untertests. Weiters wurden Theory of Mind-Aufgaben, zwei Skalen des Leistungsmotivationsinventar, die Skala zur Allgemein Selbstwirksamkeit und der Game of Dice Task (GDT) vorgegeben.

Die Berechnung der formulierten Haupt- und Nebenhypothesen wurde mittels uni- und multivariaten Varianzanalysen durchgeführt.

In Anlehnung an die Studie von Halpern und Tan (2001) wurde zur Wirksamkeitsüberprüfung der Stereotype Threat-Instruktionen der *Wissenstest* mit Selbst- und Fremdeinschätzungsaufgaben vorgegeben. Es konnte beobachtet werden, dass die Untersuchungspersonen gemäß der vorgegebenen Stereotype Threat-Bedingungen antworteten.

Die Ergebnisse zu den Untertests zum schlussfolgernden Denken belegten in allen drei Aufgabengruppen signifikante Unterschiede in Abhängigkeit vom Geschlecht ($H_1(1)$, $H_1(8)$, $H_1(15)$). Männer erbrachten in den Untertests *Quantitative und formale Probleme*, *Medizinisch-naturwissenschaftliches Grundverständnis* und *Diagramme und Tabellen* bessere Leistungen als Frauen.

Entgegen $H_1(2)$, $H_1(9)$ und $H_1(16)$, ließen sich in den Bereichen zum schlussfolgernden Denken, keine signifikanten Unterschiede in Abhängigkeit der Stereotype Threat-Bedingungen beobachten. Auch in Abhängigkeit von den Stereotype Threat-Bedingungen und dem Geschlecht zeigte sich in keinem der drei Subtests ein signifikanter Unterschied ($H_1(4)$, $H_1(11)$, $H_1(18)$).

In allen drei Aufgabenbereichen des schlussfolgernden Denkens zeigten sich signifikante Unterschiede in Abhängigkeit von der Risikoinstruktion ($H_1(3)$, $H_1(10)$, $H_1(17)$). Unter der Risikoinstruktion *raten* wurden in den drei Untertests signifikant höhere Leistungen erreicht als unter den anderen beiden Risikoinstruktionen *ohne Instruktion* und *sicher* (ermittelt mittels a-priori-Kontraste).

Analoge Ergebnisse zum Risikoverhalten wurden auch zum vorgegebenen GDT sichtbar. Es konnte ein signifikanter Unterschied im mittels GDT gemessenen Risikoverhalten in Abhängigkeit von der vorangegangenen Risikoinstruktion beobachtet werden ($H_1(33)$). Die Instruktion *raten* unterschied sich signifikant von den anderen beiden Risikoinstruktionen (ermittelt mittels a-priori-Kontraste). UntersuchungsteilnehmerInnen unter der Risikoinstruktion *raten* handelten wie vermutet risikofreudiger als unter den Risikoinstruktionen *ohne Instruktion* und *sicher*. Ein tendenziell signifikantes Ergebnis im Risikoverhalten operationalisiert mittels GDT, war in Abhängigkeit vom Geschlecht zu beobachten ($H_1(32)$). Es kann angenommen werden, dass Männer auch hier tendenziell risikofreudiger entscheiden als Frauen.

Des Weiteren ergab sich im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung zwischen den unterschiedlichen Stereotype Threat-Bedingungen und den Risikoinstruktionen ($H_1(20)$). Verschiedene

Faktorstufenkombinationen erwiesen sich als leistungsfördernd (z.B. Stereotype Threat-Bedingung *Neutral* in Kombination mit Risikoinstruktion *raten*) bzw. leistungshemmend (z.B. Stereotype Threat-Bedingung *Negativaktivierung* in Kombination mit Risikoinstruktion *ohne Instruktion*). Unter der Stereotype Threat-Bedingung *Aufklärung* war die Risikoinstruktion *raten* nicht so wirksam wie unter den anderen drei Stereotype Threat-Bedingungen.

Keine signifikanten Ergebnisse zeigten sich bezüglich den Leistungen in den drei Untertests zum schlussfolgernden Denken in Abhängigkeit von der Absicht ein Medizinstudium zu beginnen und der vorausgegangenen Auseinandersetzung mit EMS-Material, hinsichtlich der Bearbeitung von Testaufgaben ($H_1(24)$, $H_1(27)$, $H_1(30)$).

Im Untertest *Quantitative und formale Probleme* konnten signifikante Unterschiede in Abhängigkeit vom Schulzweig beobachtet werden ($H_1(31)$). SchülerInnen im naturwissenschaftlichen Schulzweig erbrachten in dieser Aufgabengruppe bessere Leistungen als jene in nicht-naturwissenschaftlichen Schulzweigen.

7 Abstract

Ziel der vorliegenden Studie war es geschlechterspezifische Unterschiede in EMS-ähnlichen Aufgaben zum schlussfolgernden Denken, unter Berücksichtigung der sozialen Faktoren Stereotype Threat und Risikoverhalten, zu untersuchen.

Es wurde eine Stichprobe von 1093 SchülerInnen (684 weiblich und 409 männlich) aus den 8. Klassen Allgemeinbildender Höherer Schulen für die Untersuchung herangezogen. Den SchülerInnen wurde, um die Wirksamkeit der experimentellen Variationen der Stereotype Threat-Bedingungen zu überprüfen der *Wissenstest* vorgegeben, dieser wurde in Anlehnung an die Untersuchung von Halpern und Tan (2001) konzipiert. Unter vier Stereotype Threat-Bedingungen und drei Risikobedingungen erfolgte anschließend die Bearbeitung EMS-ähnlicher Aufgaben. Zur Untersuchung des Risikoverhaltens wurde überdies das Computerverfahren Game of Dice Task (Brand, Fujiwara et al., 2005) vorgegeben.

Die Ergebnisse zu den Untertests zum schlussfolgernden Denken (*Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis und Diagramme und Tabellen*) belegten signifikante Unterschiede zugunsten der männlichen Untersuchungsteilnehmer. Keine signifikanten Ergebnisse ließen sich in diesen drei Untertests, in Abhängigkeit von den Stereotype Threat-Bedingungen sowie auch nicht in Abhängigkeit von den Stereotype Threat-Bedingungen und dem Geschlecht feststellen. In allen drei Aufgabengruppen erbrachten die Testpersonen unter der Risikobedingung *raten* signifikant höhere Leistungen als unter den anderen beiden Risikobedingungen. Überdies war auch ein signifikanter Unterschied im mittels GDT operationalisierten Risikoverhalten, in Abhängigkeit von der Risikobedingung zu erkennen. Des Weiteren zeigte sich im Untertest *Diagramme und Tabellen* eine signifikante Wechselwirkung zwischen den Stereotype Threat-Bedingungen und den Risikobedingungen. Es konnte beobachtet werden, dass die Untersuchung von einer vorausgegangen

Beschäftigung mit EMS-Material, hinsichtlich der Bearbeitung von Testaufgaben, unbeeinflusst war. Zudem ließen sich im Untertest *Quantitative und formale Probleme* signifikante Unterschiede zugunsten der SchülerInnen im naturwissenschaftlichen Schulzweig, erkennen.

8 Literaturverzeichnis

- Åberg-Bengtsson, L. (1999). Dimensions of Performance in the Interpretation of Diagrams, Tables, and Maps: Some Gender Differences in the Swedish Scholastic Aptitude Test. *Journal of Research in Science Teaching*, 36, 565-582.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung* (6. überarb. Aufl.). Stuttgart: Kohlhammer.
- Anderson, J. R. (2007). *Kognitive Psychologie* (6. Aufl.). Heidelberg: Spektrum Akademischer Verlag.
- Aronson, E., Wilson, T. D. & Akert, R. M. (2008). *Sozialpsychologie* (6., aktualisierte Aufl.). München: Pearson Studium.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung* (12. Aufl.). Berlin: Springer.
- Baumert, J., Bos, W. & Watermann, J. (1998). *TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Zusammenfassung deskriptiver Ergebnisse* (Studien und Berichte 64). Berlin: Max-Planck-Institut für Bildungsforschung.
- Beckmann, J. F. & Guthke, J. (1999). Psychodiagnostik des schlußfolgernden Denkens: Handbuch zur Adaptiven Computergestützten Intelligenz-Lerntestbatterie für Schlußfolgerndes Denken (ACIL). Göttingen: Hogrefe.
- Beerman, L., Heller, K. A. & Menacher, P. (1992). *Mathe: Nichts für Mädchen? Begabung und Geschlecht am Beispiel von Mathematik, Naturwissenschaft und Technik*. Bern: Verlag Hans Huber.

- Ben-Shakhar, G. & Sinai, Y. (1991). Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies. *Journal of Educational Measurement*, 28, 23-35.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin-Verlag.
- Brand, M., Fujiwara, E., Borsutzky, S., Kalbe, E., Kessler, J., & Markowitsch, H. J. (2005). Decision-Making Deficits of Korsakoff Patients in a New Gambling Task With Explicit Rules: Associations With Executive Functions. *Neuropsychology*, 19, 267-277.
- Brand, M., Kalbe, E., Labudda, K., Fujiwara, E., Kessler, J. & Markowitsch, H. J. (2005). Decision-making impairments in patients with pathological gambling. *Psychiatry Research*, 133, 91-99.
- Bridgeman, B. & Morgan, R. (1996). Success in College for Students With Discrepancies Between Performance on Multiple-Choice and Essay Tests. *Journal of Educational Psychology*, 88, 333-340.
- Bühl, A. & Zöfel, P. (2000). *SPSS Version 9. Einführung in die moderne Datenanalyse unter Windows* (6. Aufl.). München: Addison-Wesley.
- Carr, P. B. & Steele, C. M. (2009). Stereotype threat and inflexible perseverance in problem solving. *Journal of Experimental Social Psychology*, 45, 853-859.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Casey, M. B., Nuttall, R., Pezaris, E. & Benbow, C. P. (1995). The Influence of Spatial Ability on Gender Differences in Mathematics College Entrance Test Scores Across Diverse Samples. *Developmental Psychology*, 31, 697-705.
- Dark, V. J. & Benbow, C. P. (1990). Enhanced Problem Translation and Short-Term Memory: Components of Mathematical Talent. *Journal of Educational Psychology*, 82, 420-429.

- Davies, P. G., Spencer, S. J., Quinn, D. M. & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28, 1615–1628.
- DeMars, C. E. (1998). Gender Differences in Mathematics and Science on a High School Proficiency Exam: The Role of Response Format. *Applied Measurement in Education*, 11, 279-299.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Dorsch, F. (1976). *Psychologisches Wörterbuch*. Bern: Verlag Hans Huber.
- Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T. et al. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: a neuropsychological and psychophysiological study. *Neuropsychologia*, 47, 2882-2890.
- Forbes, S. D. (1996). Curriculum and Assessment: Hitting Girls Twice? In G. Hanna (Ed.), *Towards Gender Equity in Mathematics Education. An ICMI Study* (pp. 71-91). Dordrecht: Kluwer.
- Friedman, L. (1995). The Space Factor in Mathematics: Gender Differences. *Review of Educational Research*, 65, 22-50.
- Frischenschlager, O., Mitterauer, L. & Haidinger, G. (2005). Leistungsfaktoren als potenzielle Auswahlkriterien im Medizinstudium. *Zeitschrift für Hochschuldidaktik* (6), 34-41.
- Funke, J. & Vaterrodt-Plünnecke, B. (2004). *Was ist Intelligenz?* (2. überarb. Aufl.). München: C. H. Beck.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Gallagher, A. Levin, J. & Cahalan, C. (2002). *GRE research: Cognitive patterns of gender differences on mathematics admissions tests* (ETS Report No. 02-19). Princeton, NJ: Educational Testing Service.

- Gardner, H. (1991). *Abschied vom IQ. Die Rahmen-Theorie der vielfachen Intelligenzen*. Stuttgart: Klett-Cotta.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences*, 19, 229-284.
- Geary, D. C., Saults, S. J., Liu, F. & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology*, 77, 337-353.
- Guilford, J. P. & Hoepfner, R. (1976). *Analyse der Intelligenz*. Weinheim. Beltz.
- Haider, G. & Reiter, C. (Hrsg.). (2004). *PISA 2003 - Internationaler Vergleich von Schülerleistungen. Nationaler Bericht*. Graz: Leykam.
- Haidinger, G., Mitterauer, L., Rimroth, E. & Frischenschlager, O. (2008). Lernstrategien oder strategisches Lernen? Gender-abhängige Erfolgsstrategien im Medizinstudium an der Medizinischen Universität Wien. *Wiener Klinische Wochenschrift*, 120 (1-2), 37-45.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3. Aufl.). Mahwah, NJ: Erlbaum.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S. & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*, 8, 1-51.
- Halpern, D. F. & LaMay, M. L. (2000). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. *Educational Psychology Review*, 12, 229-246.
- Halpern, D. F. & Tan, U. (2001). Stereotypes and steroids: using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition*, 45, 392-414.
- Hänggen, K.-D. (2008). *Stellungnahme zum Bericht „Evaluation der Eignungstests für das Medizinstudium in Österreich“ von Chr. Spiel u.a. im Auftrag des bm:wf vom Januar 2008*. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.

- Hänsgen, K.-D. & Spicher, B. (2007). *EMS Eignungstest für das Medizinstudium 2007 - Bericht 13 über die Durchführung und Ergebnisse 2007*. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Hänsgen, K.-D. & Spicher, B. (2009). *EMS Eignungstest für das Medizinstudium. Häufig gestellte Fragen zu EMS und Numerus clausus*. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Heilemann, M., Hackl, J., Neubauer, T. & Stöger, H. (2012). Die Darstellung von Mädchen und Frauen in den Medien. In H. Stöger, A. Ziegler & M. Heilemann (Hrsg.), *Mädchen und Frauen in MINT. Bedingungen von Geschlechtsunterschieden und Interventionsmöglichkeiten* (S.77-102). Münster: LIT Verlag.
- Heine, C., Briedis, K., Didi, H.-J., Haase, K., & Trost, G. (2006): *Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern. Eine Bestandsaufnahme. HIS-Kurzinformation A3*. Hannover: Hochschul-Informations-System 2006.
- Heller, K. A. (1991). Einführung in den Gegenstandsbereich der Begabungsdiagnostik. In K. A. Heller (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung* (S. 11-36). Bern: Verlag Hans Huber.
- Herkner, W. (2003). *Lehrbuch Sozialpsychologie*. Bern: Hans Huber.
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. Göttingen: Hogrefe.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.
- Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender Differences in Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, 107, 139-155.
- ITB Consulting (Hrsg.). (2008). Test für medizinische Studiengänge II: Originalversion II des TMS (5. Aufl.). Göttingen: Hogrefe.
- Jäger, A. O. (1967). *Dimensionen der Intelligenz*. Göttingen: Hogrefe.

- Jäger, A.O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur - Test. Form 4*. Göttingen: Hogrefe.
- Johns, M., Inzlicht, M. & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, 137, 691-705.
- Johns, M., Schmader, T. & Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16, 175-179.
- Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology*, 50, 109-135.
- Jungermann, H. (2005). Entscheiden. In D. Frey, L. v. Rosenstiel & C. Graf Hoyos (Hrsg.), *Handbuch der Angewandten Psychologie*, Bd. II: Wirtschaftspsychologie. Weinheim: Beltz. S.72-78.
- Jungermann, H., Pfister, H.-R. & Fischer, K. (2010). *Die Psychologie der Entscheidung. Eine Einführung* (3. Aufl.). Heidelberg: Spektrum Akademischer Verlag.
- Jungwirth, H. (1998). *TIMSS und COMPED. Studien zur mathematisch-naturwissenschaftlichen und computerbezogenen Bildung. Konsequenzen in geschlechtsspezifischer Hinsicht*. Wien: Bundesministerium für Unterricht und kulturelle Angelegenheiten.
- Kaiser, G. & Steisel, T. (2000). Results of an analysis of the TIMSS study from a gender perspective. *Zentralblatt für Didaktik der Mathematik*, 32 (1), 18-24.
- Keller, J. & Dauenheimer, D. (2003). Stereotype threat in the classroom: dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, 29, 371-381.
- Kimura, D. (2000). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kimura, D. (2002). Sex hormones influence human cognitive pattern. *Neuroendocrinology Letters*, 23, 67-77.

- Kirchler, E. & Schrott, A. (2005). Entscheidungen. In E. Kirchler (Hrsg.), *Arbeits- und Organisationspsychologie* (S. 487-581). Wien: Facultas.
- Klauer, K. C. (2008). Soziale Kategorisierung und Stereotypisierung. In L. E. Petersen & B. Six (Hrsg.), *Stereotype, Vorurteile und soziale Diskriminierung. Theorie, Befunde und Interventionen* (S. 23-32). Basel: Beltz.
- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179-200.
- Koivula, N., Hassmén, P. & Hunt, D. P. (2001). Performance on the Swedish Scholastic Aptitude Test: Effects of Self-Assessment and Gender. *Sex Roles*, 44, 629-645.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Kuhn, J.-T. & Holling, H. (2009). Gender, reasoning ability, and scholastic achievement: A multilevel mediation analysis. *Learning and Individual Differences*, 19, 229-233.
- Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Laux, H., Gillenkirch, R. M. & Schenk-Mathes, H. Y. (2012). *Entscheidungstheorie* (8. Aufl.). Berlin: Springer.
- Leiß, D. & Blum, W. (2006). Beschreibung zentraler mathematischer Kompetenzen. In W. Blum, C. Drüke-Noe, R. Hartung & W. Blum (Hrsg.), *Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen und Fortbildungsmöglichkeiten* (S. 33-50). Berlin: Cornelsen/Scriptor.
- Lerch, H.-J., Rauch, A. & Schlesier, A. (2000). *Risikobereitschaft und Entscheidungsunsicherheit. Eine Analyse von Zusammenhängen mit Alter, Geschlecht, IQ und Schulleistungen*. Regensburg: S. Roderer

Verlag.

- Leutner, D., Funke, J., Klieme, E., & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 11-19). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lohaus, A., Vierhaus, M. & Maass, A. (2010). *Entwicklungspsychologie des Kindes- und Jugendalters*. Berlin: Springer.
- Maderthaner, R. (2008). *Psychologie - UTB-basics*. Wien: UTB-WUV.
- Mallinger, R., Holzbaur, Chr., Dierich, M., Heidegger, M., Hänsgen, K.-D. & Spicher, B. (2008). *EMS Eignungstest für das Medizinstudium in Österreich. Medizinische Universität Innsbruck, Medizinische Universität Wien*. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Mallinger, R., Holzbaur, Chr., Dierich, M., Heidegger, M., Zeller, J., Hänsgen, K.-D. & Spicher, B. (2009). *EMS Eignungstest für das Medizinstudium in Österreich. Medizinische Universität Innsbruck, Medizinische Universität Wien*. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Mallinger, R., Holzbaur, Chr., Mutz, N., Prodingler, W. M., Heidegger, M., Hänsgen, K.-D., Spicher, B. (2010). *EMS Eignungstest für das Medizinstudium in Österreich mit einem Bericht über die Pilotstudie zum Self-Assessment „Studienrelevante Persönlichkeitsmerkmale“*. Medizinische Universität Innsbruck, Medizinische Universität Wien. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Maltby, J., Day, L. & Macaskill, A. (2011). *Differentielle Psychologie, Persönlichkeit und Intelligenz* (2. aktualisierte Aufl.). München: Pearson Studium.
- Marx, D.M. & Roman, J.S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28, 1183–1193.

- Massa, L. J., Mayer, R. E. & Bohon, L. M. (2005). Individual differences in gender role beliefs influence spatial ability test performance. *Learning and Individual Differences, 15*, 99-111.
- McGlone, M. S. & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27*, 486-493.
- McGlone, M. S. & Aronson, J. (2007). Forewarning and Forearming Stereotype-Threatened Students. *Communication Education, 56*, 119-133.
- Mitterauer, L., Frischenschlager, O. & Haidinger, G. (2007). Sex differences in study progress at Medical University of Vienna. *GMS Zeitschrift für Medizinische Ausbildung, 24* (2), Doc111.
- Organisation for Economic Co-operation and Development (OECD). (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD Publications.
- Pauen, S. & Vonderlin, E. (2009). Entwicklungspsychologische Grundlagen. In S. Schneider & J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie. Band 3: Störungen im Kindes- und Jugendalter* (S. 3-22). Heidelberg: Springer.
- Petersen, L. E. & Six, B. (Hrsg.). (2008). *Stereotype, Vorurteile und soziale Diskriminierung. Theorien, Befunde und Interventionen*. Basel: Beltz.
- Pietrowski, D. & Ennker, J. (2007). Begriffserklärung Risiko. In J. Ennker, D. Pietrowski & P. Kleine, *Risikomanagement in der operativen Medizin* (S. 3). Darmstadt: Steinkopff Verlag.
- Prochaska, M. (1998). Leistungsmotivation. *Methoden, soziale Erwünschtheit und das Konstrukt*. Frankfurt: Lang.
- Quinn, D. M. & Spencer, S. J. (2001). The Interference of Stereotype Threat With Women's Generation of Mathematical Problem-Solving Strategies. *Journal of Social Issues, 57*, 55-71.
- Rammstedt, B. & Rammsayer, T. H. (2000). Sex differences in self-estimates of different aspects of intelligence. *Personality and Individual Differences, 29*, 869-880.

- Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J.-M. & Huguet, P. (2010). Individual differences in working memory moderate stereotype threat effects. *Psychological Science*, 21, 1646–1648.
- Rheinberg, F. (2008). *Motivation* (7. Aufl.). Stuttgart: Kohlhammer.
- Rollett, B. & Kastner-Koller, U. (2007). *Praxisbuch Autismus für Eltern, Erzieher, Lehrer und Therapeuten* (3. Aufl.). München: Elsevier.
- Rost, D. H. (2008). Multiple Intelligenzen, multiple Irritationen. *Zeitschrift für Pädagogische Psychologie*, 22, 97-112.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Verlag Hans Huber.
- Schmader, T. & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440-452.
- Schmader, T., Johns, M. & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50, 835-850.
- Schmader, T., Johns, M. & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336-356.
- Schuler, H. & Prochaska, M. (2001). *Leistungsmotivationsinventar (LMI)*. Göttingen: Hogrefe.
- Schwantner, U. & Schreiner, C. (Hrsg.). (2010). *PISA 2009. Internationaler Vergleich von Schülerleistungen. Erste Ergebnisse. Lesen, Mathematik, Naturwissenschaft*. Graz: Leykam.
- Schwarzer, R. (2004). *Psychologie des Gesundheitsverhaltens. Einführung in die Gesundheitspsychologie* (3. überarb. Aufl.). Göttingen: Hogrefe.
- Schwarzer, R. & Jerusalem, M. (Hrsg.). (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.

- Spearman, C. (1927). *The abilities of man: their nature and measurement*. London: Macmillan.
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Spiel, C., Schober, B. & Litzenberger, M. (2008). *Evaluation der Eignungstests für das Medizinstudium in Österreich*. Projektbericht für das Bundesministerium für Wissenschaft und Forschung. Wien.
- Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In P. Z. Mark (Ed.), *Advances in Experimental Social Psychology* (Vol. 34) (pp. 379-440). San Diego: Academic Press.
- Strand, S., Deary, I. J. & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology*, 76, 463-480.
- Stumpf, H. & Fay, E. (1983). *Schlauchfiguren. Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Göttingen: Hogrefe.
- Suchań, B., Wallner-Paschon, C., Bergmüller, S. & Schreiner, S. (Hrsg.). (2008). *TIMSS 2007. Mathematik und Naturwissenschaft in der Grundschule. Erste Ergebnisse*. Graz: Leykam.
- Süß, H.-M. (2003). Intelligenztheorien. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 217-224). Weinheim: Beltz.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H.-U. (1998). Evaluation des Tests für Medizinische Studiengänge (TMS):

Synopse der Ergebnisse. Bonn: Institut für Test- und Begabungsforschung.

Wagner-Menghin, M. M. (2003). Messgenauigkeit. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 282–286). Weinheim: Beltz.

Waldmann, M. R. & Weinert, F. E. (1990). *Intelligenz und Denken. Perspektiven der Hochbegabungsforschung*. Göttingen: Hogrefe.

Willinger, U., Schmöger, M., Müller, C. & Auff, E. Theory of Mind Stories. In *Vorbereitung*.

Willingham, W. W. & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Zentrum für Testentwicklung und Diagnostik (ZTD). (2009). *TESTINFO '09. Eignungstest für das Medizinstudium (EMS). Version A*. [Broschüre]. Innsbruck, Wien: Medizinische Universität.

Zöfel, P. (2003). *Statistik für Psychologen. Im Klartext*. München: Pearson Studium.

9 Abbildungsverzeichnis

Abbildung 1: Fremdeinschätzungsaufgabe zum Bereich Quantitative und formale Probleme ..59	59
Abbildung 2: Selbsteinschätzungsaufgabe zum Bereich Quantitative und formale Probleme...59	59
Abbildung 3: Beispielaufgabe zum Untertest Figuren Lernen (ZTD, 2009)	62
Abbildung 4: Beispielaufgabe zum Untertest Fakten Lernen (ZTD, 2009)	62
Abbildung 5: Beispielaufgabe zum Untertest Quantitative und formale Probleme (ZTD, 2009) 63	63
Abbildung 6: Beispielaufgabe zum Untertest Schlauchfiguren (ZTD, 2009).....	64
Abbildung 7: Beispielaufgabe zum Untertest Medizinisch-naturwissenschaftliches Grundverständnis (ZTD, 2009).....	65
Abbildung 8: Beispielaufgabe zum Untertest Diagramme und Tabellen (ZTD, 2009).....	66
Abbildung 9: Beispielaufgabe aus der Dimension Internalität - LMI (Schuler & Prochaska, 2001)	68
Abbildung 10: Beispielaufgabe aus der SWE (Schwarzer & Jerusalem, 1999).....	69
Abbildung 11: Game of Dice Task (Brand, Fujiwara et al., 2005)	70
Abbildung 12: Anteilswerte der Geschlechter in Prozent.....	74
Abbildung 13: Alterskategorien in der Gesamtstichprobe	75
Abbildung 14: Anteilswerte der Staatsbürgerschaft in der Gesamtstichprobe	77
Abbildung 15: Schulzweige	78
Abbildung 16: Interaktionsdiagramm Selbsteinschätzung im Bereich Diagramme und Tabellen	104
Abbildung 17: Profildiagramm zum Mittelwertsverlauf im Untertest Quantitative und formale Probleme in Abhängigkeit von Geschlecht und Risikoinstruktion.....	113
Abbildung 18: Mittelwertsverlaufdiagramm im Untertest Medizinisch-naturwissenschaftliches Grundverständnis in Abhängigkeit von Geschlecht und Risikoinstruktion	118
Abbildung 19: Mittelwertsverlaufdiagramm im Untertest Diagramme und Tabellen in Abhängigkeit von Geschlecht und Risikoinstruktion	122
Abbildung 20: Wechselwirkungsdiagramm zum Untertest Diagramme und Tabellen in Abhängigkeit von Stereotype Threat- und Risikobedingung	124
Abbildung 21: Mittelwertsverlaufdiagramm im Nettoscore des GDT in Abhängigkeit von Geschlecht und Risikoinstruktion.....	137

10 Tabellenverzeichnis

Tabelle 1: Aufbau, Ablauf und erfassende Fähigkeiten des EMS (modifiziert nach ZTD, 2009) .5	
Tabelle 2: Geplantes Untersuchungsdesign	55
Tabelle 3: Geschlechterverteilung der Stichprobe.....	74
Tabelle 4: Chi-Quadrat-Test für Geschlechterverteilung	75
Tabelle 5: Häufigkeiten Alter gesamt.....	76
Tabelle 6: Chi-Quadrat-Test für Altersverteilung	76
Tabelle 7: Häufigkeiten der Staatsbürgerschaft	77
Tabelle 8: Chi-Quadrat-Test – Schulzweig.....	79
Tabelle 9: Häufigkeiten – Wiederholen eines Schuljahres.....	79
Tabelle 10: Chi-Quadrat-Test – Wiederholen eines Schuljahres	80
Tabelle 11: Kreuztabelle Geschlecht und Wiederholen einer Klasse	80
Tabelle 12: Vierfeldertafel naturwissenschaftlicher Schulzweig und Wiederholen eines Schuljahres.....	81
Tabelle 13: Häufigkeiten – Absicht ein Medizinstudium zu beginnen.....	82
Tabelle 14: Chi-Quadrat-Test – Absicht ein Medizinstudium zu beginnen	82
Tabelle 15: Vierfeldertafel Geschlecht und Absicht eine Medizinstudium zu beginnen	83
Tabelle 16: Kontingenztafel Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS in irgendeiner Form.....	84
Tabelle 17: Kontingenztafel Geschlecht und Auseinandersetzung mit EMS allgemein	85
Tabelle 18: Vierfeldertafel naturwissenschaftlicher Schulzweig und Auseinandersetzung mit EMS in irgendeiner Form	86
Tabelle 19: Häufigkeiten in den Stereotype Threat-Bedingungen.....	87
Tabelle 20: Chi-Quadrat-Test – Stereotype Threat-Bedingung.....	88
Tabelle 21: Kreuztabelle Geschlecht und Stereotype Threat-Bedingung	88
Tabelle 22: Häufigkeiten in den Risikobedingungen.....	89
Tabelle 23: Chi-Quadrat-Test – Risikobedingung	90
Tabelle 24: Kreuztabelle Geschlecht und Risikobedingung.....	90
Tabelle 25: Kontingenztabelle Stereotype Threat-Bedingung und Risikobedingung	91

Tabelle 26: Statistische Kenngrößen zu den EMS-Untertests	92
Tabelle 27: Lösungswahrscheinlichkeiten im Untertest Quantitative und formale Probleme	93
Tabelle 28: Lösungswahrscheinlichkeiten im Untertest Medizinisch-naturwissenschaftliches Grundverständnis.....	95
Tabelle 29: Lösungswahrscheinlichkeiten im Untertest Diagramme und Tabellen	96
Tabelle 30: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Quantitative und formale Probleme	99
Tabelle 31: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Med.-nat. Grundverständnis	100
Tabelle 32:Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Diagramme und Tabellen	101
Tabelle 33: Kennwerte und Prüfgrößen der Selbsteinschätzung in Abhängigkeit vom Geschlecht.....	103
Tabelle 34: Reliabilitäten nach Cronbach's Alpha	105
Tabelle 35: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS- ähnlicher Untertest Quantitative und formale Probleme.....	106
Tabelle 36: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Schlauchfiguren	107
Tabelle 37: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS- ähnlicher Untertest Medizinisch-naturwissenschaftliches Grundverständnis	107
Tabelle 38: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS-ähnlicher Untertest Diagramme und Tabellen	108
Tabelle 39: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS- ähnlicher Untertest Figuren Lernen.....	108
Tabelle 40: Reliabilitäten der Faktorstufenkombinationen (Stichprobenumfang, Itemanzahl) EMS- ähnlicher Untertest Fakten Lernen	109
Tabelle 41: Interkorrelationsmatrix Produkt-Moment-Korrelationen r zwischen den EMS- ähnlichen Untertests (N = 1093).....	109
Tabelle 42: Deskriptive Statistiken – Kennwerte AV Untertest Quantitative und formale Probleme	111
Tabelle 43: Tests der Zwischensubjekteffekte – AV Untertest Quantitative und formale Probleme	112

Tabelle 44: Kontrastergebnisse im Untertest Quantitative und formale Probleme in Abhängigkeit der Risikoinstruktion.....	114
Tabelle 45: Deskriptive Statistiken – Kennwerte AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis.....	116
Tabelle 46: Tests der Zwischensubjekteffekte – AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis.....	116
Tabelle 47: Kontrastergebnisse im Untertest Medizinisch-naturwissenschaftliches Grundverständnis in Abhängigkeit der Risikoinstruktion.....	118
Tabelle 48: Deskriptive Statistiken – Kennwerte AV Untertest Diagramme und Tabellen.....	120
Tabelle 49: Tests der Zwischensubjekteffekte – AV Untertest Diagramme und Tabellen.....	121
Tabelle 50: Kontrastergebnisse im Untertest Diagramme und Tabellen in Abhängigkeit der Risikoinstruktion.....	123
Tabelle 51: Kreuztabelle Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet.....	125
Tabelle 52: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Quantitative und formale Probleme.....	127
Tabelle 53: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Medizinisch-naturwissenschaftliches Grundverständnis.....	129
Tabelle 54: Tests der Zwischensubjekteffekte – Absicht ein Medizinstudium zu beginnen und Auseinandersetzung mit EMS: Testaufgaben bearbeitet – AV Untertest Diagramme und Tabellen.....	131
Tabelle 55: Deskriptivstatistik der Kennwerte in den EMS-ähnlichen Bereichen Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis sowie Diagramme und Tabellen nach Schulzweig.....	132
Tabelle 56: Tests der Zwischensubjekteffekte – Schulzweig – AV Untertest Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis sowie Diagramme und Tabellen.....	133
Tabelle 57: Deskriptive Statistiken – AV Nettoscore im GDT.....	135
Tabelle 58: Tests der Zwischensubjekteffekte – AV Nettoscore im GDT.....	135
Tabelle 59: Kontrastergebnisse im Nettoscore des GDT in Abhängigkeit der vorangegangenen Risikoinstruktion.....	137

11 Anhang

Anhang A: Tabellen

Tabelle A1: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Figuren Lernen

			Wirksamkeit in Fremdeinschätzungsaufgabe Figuren Lernen			Gesamt
			nein	ja	konträr	
Stereotype Threat-Bedingung	Aktivierung	Anzahl	115	124	50	289
		erwartet	76.8	172.2	40.0	289.0
		% innerhalb	39,8%	42,9%	17,3%	100,0%
		Stand. Residuen	4.4	-3.7	1.6	
	Neutral	Anzahl	37	205	13	255
		erwartet	67.7	152.0	35.3	255.0
		% innerhalb	14,5%	80,4%	5,1%	100,0%
		Stand. Residuen	-3.7	4.3	-3.8	
	Aufklärung	Anzahl	33	216	30	279
		erwartet	74.1	166.3	38.6	279.0
		% innerhalb	11,8%	77,4%	10,8%	100,0%
		Stand. Residuen	-4.8	3.9	-1.4	
	Negativaktivierung	Anzahl	103	101	57	261
		erwartet	69.3	155.5	36.1	261.0
		% innerhalb	39,5%	38,7%	21,8%	100,0%
		Stand. Residuen	4.0	-4.4	3.5	
Gesamt		Anzahl	288	646	150	1084
		erwartet	288.0	646.0	150.0	1084.0
		% innerhalb	26,6%	59,6%	13,8%	100,0%

Tabelle A2: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe
Fakten Lernen

			Wirksamkeit in Fremdeinschätzungsaufgabe Fakten Lernen			Gesamt
			nein	ja	konträr	
Stereotype Threat-Bedingung	Aktivierung	Anzahl	48	205	37	290
		erwartet	82.3	155.5	52.3	290.0
		% innerhalb	16,6%	70,7%	12,8%	100,0%
		Stand. Residuen	-3.8	4.0	-2.1	
	Neutral	Anzahl	84	132	38	254
		erwartet	72.1	136.2	45.8	254.0
		% innerhalb	33,1%	52,0%	15,0%	100,0%
		Stand. Residuen	1.4	-4	-1.1	
	Aufklärung	Anzahl	127	104	47	278
		erwartet	78.9	149.0	50.1	278.0
		% innerhalb	45,7%	37,4%	16,9%	100,0%
		Stand. Residuen	5.4	-3.7	-.4	
	Negativaktivierung	Anzahl	48	139	73	260
		erwartet	73.8	139.4	46.9	260.0
		% innerhalb	18,5%	53,5%	28,1%	100,0%
		Stand. Residuen	-3.0	.0	3.8	
Gesamt	Anzahl	307	580	195	1082	
	erwartet	307.0	580.0	195.0	1082.0	
	% innerhalb	28,4%	53,6%	18,0%	100,0%	

Tabelle A3: Kontingenztabelle Stereotype Threat-Bedingung und Fremdeinschätzungsaufgabe Schlauchfiguren

			Wirksamkeit in Fremdeinschätzungsaufgabe Schlauchfiguren			Gesamt
			nein	ja	konträr	
Stereotype Threat-Bedingung	Aktivierung	Anzahl	70	210	9	289
		erwartet	84.0	174.2	30.8	289.0
		% innerhalb	24,2%	72,7%	3,1%	100,0%
		Stand. Residuen	-1.5	2.7	-3.9	
	Neutral	Anzahl	77	149	28	254
		erwartet	73.8	153.1	27.0	254.0
		% innerhalb	30,3%	58,7%	11,0%	100,0%
		Stand. Residuen	.4	-.3	.2	
	Aufklärung	Anzahl	102	147	28	277
		erwartet	80.5	167.0	29.5	277.0
		% innerhalb	36,8%	53,1%	10,1%	100,0%
		Stand. Residuen	2.4	-1.5	-.3	
	Negativaktivierung	Anzahl	65	145	50	260
		erwartet	75.6	156.7	27.7	260.0
		% innerhalb	25,0%	55,8%	19,2%	100,0%
		Stand. Residuen	-1.2	-.9	4.2	
Gesamt	Anzahl	314	651	115	1080	
	erwartet	314.0	651.0	115.0	1080.0	
	% innerhalb	29,1%	60,3%	10,6%	100,0%	

Tabelle A4: Prozentmittelwerte in den Fremdeinschätzungsaufgaben im Wissenstest, in Abhängigkeit der Stereotype-Threat Bedingungen

	Stereotype Threat-Bedingung											
	Aktivierung			Neutral			Aufklärung			Negativaktivierung		
	n	M	SD	n	M	SD	n	M	SD	n	M	SD
Figuren_M	289	42.78	24.14	255	49.47	8.00	279	50.06	8.61	261	55.95	25.29
Figuren_F	289	57.25	24.12	255	50.76	7.98	279	49.90	8.73	261	44.09	25.31
Fakten_M	290	33.69	23.04	254	48.13	13.26	278	46.98	14.13	260	61.13	26.79
Fakten_F	290	66.22	23.20	254	51.87	13.26	278	53.00	14.22	260	38.60	26.83
Naturw_M	290	59.21	22.24	254	52.78	9.61	277	52.12	9.71	259	41.60	24.95
Naturw_F	290	40.76	22.26	254	47.26	9.59	277	47.88	9.67	260	58.22	25.09
Diagr_M	289	69.71	19.78	254	54.57	9.24	277	55.09	11.51	260	36.09	25.02
Diagr_F	289	30.35	19.83	254	45.44	9.24	277	44.84	11.14	260	64.12	24.79
Quant_M	289	68.56	21.21	254	53.70	12.33	277	56.45	11.80	260	42.84	26.16
Quant_F	289	31.54	21.32	254	46.10	11.86	277	43.58	11.81	260	57.06	26.38
Schlauch_M	289	71.65	20.89	254	53.08	12.25	277	54.62	12.27	260	36.35	26.77
Schlauch_F	289	28.45	21.02	254	47.00	12.14	277	45.36	12.27	260	63.89	26.50
ToM_M	288	21.71	17.17	254	34.25	14.39	276	31.68	13.80	257	22.19	18.77
ToM_F	288	78.36	17.08	254	65.99	14.40	276	68.32	13.80	257	77.82	18.74

_M. Wahrscheinlichkeit, Person die betreffendes Item (z.B. zum Bereich Figuren Lernen) gelöst hat ist ein Mann.

_F. Wahrscheinlichkeit, Person die betreffendes Item gelöst hat ist eine Frau.

Anhang B: Lebenslauf

Persönliche Daten

Name	Silvia Rosa Hameseder
Geburtsdatum	14. Dezember 1983
Geburtsort	Melk
Staatsangehörigkeit	Österreich

Ausbildung

1990-1994	Volksschule in Emmersdorf
1994-1998	Hauptschule mit fremdsprachlichem Schwerpunkt in Emmersdorf
1998-2003	Handelsakademie in Ybbs mit dem Schwerpunkt Marketing und internationale Geschäftstätigkeit
seit 10/2003	Psychologiestudium an der Universität Wien

Berufliche Tätigkeiten

seit April 2009	Jugendhelfer beim NÖ Hilfswerk, Melk
Jänner 2010- Juni 2012	Wissenschaftliche Mitarbeit an der Universitätsklinik für Neurologie im Allgemeinen Krankenhaus in Wien
Juli-August 2009	MACO Produktions GmbH, Trieben
September 2008	Veitsch-Radex GmbH & Co, Trieben
August 2008	MACO Produktions GmbH, Trieben
August 2007	MACO Produktions GmbH, Trieben
September 2006	Vetropack Austria GmbH, Pöchlarn
Juli-August 2006	Gasthaus Weidenauer, Emmersdorf
September 2003	Ing. H. Gradwohl GesmbH, Melk

Praktika

August 2005	NÖ Landes-Pensionisten- und Pflegeheim Melk, in den Bereichen Seniorenbetreuung und Hospiz
März-Juni, September 2009	NÖ Hilfswerk, Zentrum für Beratung & Begleitung Krems

