

Eva Meunier

Transferevaluation eines Testtrainings für den Test
für medizinische Studiengänge (TMS)

Diplomarbeit

Freie Universität Berlin

Fachbereich Erziehungswissenschaft und Psychologie

Diplomstudiengang Psychologie

Erstgutachter: PD Dr. Hans-Uwe Hohner

Zweitgutachter: Prof. Ernst-H. Hoff

Berlin

13. Januar 2010

Ich erkläre an Eides Statt, dass ich die beiliegende Diplomarbeit selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Mir ist bekannt: Bei Verwendung von Inhalten aus dem Internet habe ich diese zu kennzeichnen und mit Datum sowie der Internet-Adresse (URL) ins Literaturverzeichnis aufzunehmen.

Diese Arbeit hat keiner anderen Prüfungsbehörde vorgelegen.

Ich bin mit der Einsichtnahme im Universitätsarchiv der FU und auszugsweiser Kopie einverstanden. Alle übrigen Rechte behalte ich mir vor. Zitate sind nur mit vollständigen bibliographischen Angaben und dem Vermerk „unveröffentlichtes Manuskript einer Diplomarbeit“ zulässig.

Berlin, den

(Eva Meunier)

Danksagung

Ich danke PD Dr. Hans-Uwe Hohner für die Vergabe des Themas dieser Diplomarbeit und für die Bereitstellung der erhobenen Daten, sowie für die wertvolle Hilfe und Unterstützung während der gesamten Entstehungsphase der Arbeit.

INHALTSVERZEICHNIS

1	EINLEITUNG.....	6
2	FRAGESTELLUNG.....	9
2.1	AUFBAU DER ARBEIT.....	11
3	DER TEST FÜR MEDIZINISCHE STUDIENGÄNGE (TMS)	12
3.1	TESTENTSTEHUNG	12
3.2	TESTAUFBAU.....	13
3.3	TESTABLAUF.....	16
3.4	TESTAUSWERTUNG.....	16
3.5	GESCHICHTE DES TMS	16
4	TESTTRAINING IN DER DEUTSCHSPRACHIGEN LITERATUR.....	18
4.1	STUDIE VON BERNHARD DETER (1982)	18
4.1.1	<i>Anmerkungen zu Deters (1982) Studie</i>	<i>33</i>
4.2	STUDIE VON ERNST FAY (1985)	35
4.2.1	<i>Anmerkungen zu Fay (1985).....</i>	<i>39</i>
4.3	STUDIE KIRCHENKAMP UND MISPELKAMP (1988).....	39
4.3.1	<i>Anmerkungen zu Kirchenkamp und Mispelkamp (1988).....</i>	<i>42</i>
4.4	STUDIE VON R. HOFER & K.-D. HÄNSGEN (1996)	46
4.5	STUDIE VON GÜNTER TROST ET AL. (1998).....	46
4.5.1	<i>Anmerkungen zu Trost (1998).....</i>	<i>51</i>
5	TESTTRAINING IM ANGLOAMERIKANISCHEN RAUM	52
5.1	AUSWAHLVERFAHREN IN DEN USA.....	52
5.2	SAT.....	53
5.3	ACT	54
5.4	„TEST-WISENESS“, „TEST-PRACTICE“ UND „TEST-COACHING“	55
5.5	STUDIEN ZUM THEMA TESTTRAINING AUS DEN USA.....	57
5.6	ANMERKUNGEN ZU DEN STUDIEN ZUM SAT	64
5.7	BEZUG ZUM TMS	65
6	TESTTRAINING.....	66
6.1	KONZEPT	66
6.2	SEMINAR UND GENERALPROBE	67
7	BEFUNDE ZUR TRANSFEREVALUATION.....	71

7.1	EVALUATIONSGEGENSTAND UND FRAGESTELLUNGEN	73
7.2	DATENERHEBUNG, STICHPROBE UND AUSWERTUNGSVERFAHREN	75
7.2.1	<i>Datenerhebung</i>	75
7.2.2	<i>Stichprobe</i>	75
7.2.3	<i>Auswertungsverfahren</i>	78
7.3	DARSTELLUNG DER BEFUNDE	78
7.4	SCHLUSSFOLGERUNGEN AUS DEN ERGEBNISSEN DER EVALUATION	90
8	ZUSÄTZLICHE ANALYSE AUSGEWÄHLTER ASPEKTE	94
8.1	FRAGESTELLUNGEN.....	95
8.2	AUSGEWÄHLTE SPEZIALSTICHPROBE	96
8.3	MESSINSTRUMENTE.....	99
8.4	AUSWERTUNGSVORGEHEN.....	102
8.5	DATENERHEBUNG	105
8.6	BEFUNDE DER ANALYSE DER SPEZIALSTICHPROBE	106
8.7	DARSTELLUNG DER ERGEBNISSE.....	107
9	DISKUSSION UND FAZIT.....	121
10	LITERATUR	128
11	ANHANG.....	131
11.1	ALLGEMEINER FRAGEBOGEN ZUR ERFASSUNG VON DATEN	131
11.2	FRAGEBOGEN ZUR PRÜFUNGSERWARTUNG.....	132
11.3	FRAGEBOGEN „FAKTOREN FÜR EIN GUTES ABSCHNEIDEN IM TMS“.....	134
11.4	FRAGEBOGEN ZUR TESTBEFINDLICHKEIT.....	135
11.5	FRAGEBOGEN ZUR GENERALPROBE.....	138
11.6	EVALUATIONSBOGEN	139

1 Einleitung

Seit der siebten Novelle des Hochschulrahmengesetzes (HRG) im Sommer 2004 haben die Hochschulen die Möglichkeit, die Auswahl ihrer Studierenden in den bundesweit zulassungsbeschränkten Studiengängen aktiv durch Einbeziehungen verschiedener Auswahlverfahren mitzugestalten. Die Auswahl der Studierenden hat zweierlei Gründe: Zunächst einmal können die Hochschulen durch die Auswahl der geeigneten Studenten ihr Profil schärfen und sich somit auf dem nationalen und internationalen Bildungsmarkt besser profilieren. In den USA haben Universitäten mit strengen Auswahl- und Zulassungsverfahren eine höhere Reputation. Außerdem haben die Hochschulen die Möglichkeit, die Studenten auszuwählen, die den Anforderungen ihres Studiengangs am besten entsprechen. Auf der anderen Seite bedeutet ein Auswahlverfahren für den Studieninteressierten einen höheren Anspruch an sein Bewerbungsverhalten. Der Studieninteressierte muss nicht nur seine Hochschulzugangsberechtigung einreichen, sondern er muss sich auch mit einem Auswahlinstrument, wie z. B. einem Auswahlgespräch oder einem Studierfähigkeitstest auseinandersetzen. Um eine zusätzliche Hürde nehmen zu können, muss man sich vorbereiten, so dass der Studieninteressierte schon bei der Vorbereitung die Möglichkeit hat, seinem Studienwunsch zu überdenken. Die intensivere Auseinandersetzung mit dem Studienwunsch kann somit auch dazu führen, dass sich die Studierenden in ihrem Studium wohler fühlen werden. Eine gelungene Auswahl wirkt sich positiv auf das Leistungsverhalten aus (Schuler, 2008), (Deidesheimer Kreis, 1997), (Hell, 2007), (Wissenschaftsrat, 2004), (Tarazona, 2006), (Trost, 2005a), (Trost, 2005b), (Hänsgen, Hofer & Ruefli, 1995).

Für die Auswahl der Studierenden stehen den Hochschulen verschiedene Möglichkeiten zur Verfügung: Schulnoten, Interviews und Studierfähigkeitstests. Anlässlich der aktuellen Entwicklungen an den deutschen Hochschulen der Auswahl ihrer zukünftigen Studenten mehr Gewicht zu verleihen, beschäftigt sich diese Studie mit dem Studierfähigkeitstest als mögliches Auswahlverfahren. Dieses Selektionsinstrument stellt an die Testteilnehmer meistens für sie bisher unbekannte Anforderungen. Auf der einen Seite werden ihre kognitiven Fähigkeiten im äußersten Maße gefordert. Auf der

anderen Seite werden sie psychisch wie physisch durch das oft mehrstündig dauernde Testverfahren starken Belastungen ausgesetzt. Welche Faktoren haben auf eine Testleistung Einfluss? Das Testergebnis spiegelt nicht nur die kognitiven Fähigkeiten wider, sondern noch andere bisher unbekannte Einflussgrößen. Hochschulen, die sich für einen Studierfähigkeitstest als Auswahlverfahren entscheiden, sollten sich darüber in Klarem sein, dass der Studierfähigkeitstest nicht allein kognitive Fertigkeiten misst, sondern eine Menge anderer Faktoren, die einen Einfluss auf die Testleistung haben.

Um einen genauen Einblick in dieses Auswahlinstrument zu erhalten, wird in dieser Arbeit ein anerkannter und erprobter Studierfähigkeitstest vorgestellt. In der Bundesrepublik Deutschland wurde in den Jahren 1986 bis 1997 ein Studierfähigkeitstest für die medizinischen Studiengänge (TMS) durchgeführt, um der großen Zahl von Bewerbern und den damit verbundenen Wartezeiten zu begegnen. Ein viel diskutierter Aspekt bei der Einführung des Studierfähigkeitstests als Selektionsinstrument war damals die Frage, in wieweit der Test trainierbar ist. Unter Testtraining versteht man Vorbereitung auf einen spezifischen Test unter professioneller Anleitung. Zum Thema Testtraining speziell für Studieneignungstests wurden sowohl in Deutschland als auch in anderen Ländern verschiedene Studien veröffentlicht. Vorgestellt werden hier die wichtigsten Untersuchungen zum TMS in Deutschland und wichtige Analysen zu dem wichtigsten Studieneignungstest, dem Scholastic Aptitude Test (SAT), in den USA. Neben der Darstellung der wichtigsten Literatur wird in dieser Arbeit ein spezielles Testtraining zum TMS vorgestellt, das in den Jahren 1986 bis 1995 über 6000 Studienbewerber auf den TMS vorbereitet hat. Dieses spezifische Testtraining wurde von der Career Counseling Laufbahnberatung Dr. Hohner Berlin (CCH) entwickelt.

Ein Testtraining ist eine Interventionsmaßnahme, von der sich die Teilnehmer ein bestimmtes Ergebnis versprechen. Die Besucher eines solchen Trainingsprogramms investieren Zeit und Geld, um durch ein gutes Testergebnis im Studierfähigkeitstest ihre Chancen auf eine baldige Zulassung in ihrem Wunschstudienfach an ihrer Wunschuniversität zu erhöhen. Oft bleibt die Frage unbeantwortet, ob sich die Investition für die Kursteilnehmer gelohnt hat und der versprochene Erfolg eingetreten ist. Im ersten Teil wird daher das Testtraining der CCH-Laufbahnberatung evaluiert

werden. Um die Wirksamkeit einer Interventionsmaßnahme zu überprüfen, stehen verschiedene Instrumente zur Verfügung. Die vorliegende Arbeit wird das Testtraining anhand des von Kirkpatrick (2004) entwickelten Vier-Ebenen-Modells evaluieren. Im Mittelpunkt der Bewertung des Testtrainings steht die Transferevaluation. Die Frage ist, ob die Seminarbesucher in der Lage waren, die im Training erworbenen Kenntnisse in die Praxis zu transferieren.

Aus der Literaturrecherche zum Thema Testtraining haben sich interessante Fragen ergeben. Übereinstimmend ist allen Studien zum Thema Testtraining, dass eine solche Interventionsmaßnahme einen Effekt hat. Über die Größe des Effekts herrscht allerdings Uneinigkeit. Bei der genaueren Analyse der Coachingliteratur wird deutlich, dass alle Studien mit methodischen Schwierigkeiten zu kämpfen haben. Es ist also nicht einfach, Aussagen über die Höhe des Trainingseffekts zu machen. Insbesondere der Punkt der individuellen Voraussetzungen der Testtrainingsteilnehmer bleibt häufig bei den Analysen unberücksichtigt. Im Anschluss an die Transferevaluation werden deshalb anhand einer Spezialstichprobe zusätzliche Analysen durchgeführt werden, um mögliche Einflussfaktoren, die neben dem Testtraining auf die Testleistung wirken, genauer zu betrachten.

2 Fragestellung

Um die in der Einleitung angesprochenen Fragen genauer zu untersuchen, mussten zunächst einmal die Studien zum Einfluss von Testtraining auf die Testleistung im TMS analysiert werden. Die wichtigste Studie in diesem Zusammenhang ist die Studie von Deter (1982). Diese Studie wurde noch vor der damaligen Einführung des besonderen Auswahlverfahrens realisiert. Deter hat den Einfluss von Testtraining mit Hilfe einer experimentellen Anordnung untersucht und später noch weitere Analysen nach der endgültigen Einführung des TMS durchgeführt. Neben Deter (1982) wurden noch weitere Untersuchungen zum Einfluss von Training auf das Testresultat veröffentlicht. Fay (1985) analysierte genauer die Möglichkeiten, die Studenten damals zur Verfügung standen, sich auf den TMS vorzubereiten. Er befragte die Testteilnehmer nach ihrer Art der Vorbereitung und versuchte anhand der gesammelten Daten die Effekte, die durch die unterschiedliche Form der Vorbereitung zustande kamen, miteinander zu vergleichen. Kirchenkamp und Mispelkamp (1988) untersuchten diese Fragestellung einige Jahre später erneut. Im Jahr 1998 fasste Trost in seiner Evaluation des Tests für medizinische Studiengänge (TMS) die zu der Frage der Trainierbarkeit des TMS erstellten Studien in einem Artikel zusammen. Ein weiterer Ausgangspunkt zur Analyse der beschriebenen Fragestellungen war die Prüfung der angloamerikanischen Literatur zum Thema Trainingsforschung. In den USA ist das Thema Testtraining viel präsenter als in Deutschland, da fast alle amerikanischen Universitäten zur Auswahl ihrer Studenten das Resultat eines Studieneignungstest miteinbeziehen. Die Literaturrecherche konzentriert sich auf die Studien zum Scholastic Aptitude Test (SAT). Im Zusammenhang mit der Recherche sowohl im deutsch- als auch im amerikanischen Raum wird deutlich, dass gerade der Nachweis von Trainingseffekten die Wissenschaftler methodisch vor große Schwierigkeiten stellt. So ist es nicht weiter verwunderlich, dass über die Größe des Trainingseffekts Uneinigkeit besteht.

Die vorliegende Arbeit konzentriert sich zunächst auf die Evaluation des Testtrainings. Bei der Entwicklung eines Trainingsprogramms geht man davon aus, dass dieses Programm eine bestimmte Wirkung erzielen soll. Aus den Zielen der Interventionsmaßnahme leiten sich dementsprechend die Fragestellungen ab, die mit

Hilfe der Evaluation überprüft werden sollen. Die Evaluation wird anhand des Vier-Ebenen-Modells von Kirkpatrick durchgeführt werden. Im Vordergrund steht dabei der Gedanke, dass der Lernerfolg eines Trainingsprogramms auf vier verschiedenen Ebenen gemessen werden kann. Aus den Ebenen, die Kirkpatrick „reaction“, „learning“, „behaviour“ und „results“ nannte, ergeben sich folgende Fragestellungen, die in dem Kapitel zur Transferevaluation noch stärker differenziert werden:

1. Wie beurteilten die Seminarteilnehmer die Trainer des Testtrainings der CCH-Laufbahnberatung?
2. Haben die Besucher des Testtrainings ihre Einstellungen gegenüber Prüfungen durch das Seminar verändert?
3. Haben die Teilnehmer von dem Trainingsprogramm profitiert?
4. Wie haben die Kursteilnehmer im TMS abgeschnitten?
5. Wie sahen ihre Ergebnisse in den neun Untertests des TMS aus?
6. Wie sahen ihre Resultate gerade in den kurzfristig trainierbaren Untertests aus?

Eine weitere Untersuchung beschäftigt sich mit einer zusätzlichen Analyse ausgewählter Aspekte an einer Spezialstichprobe, die sich aus Fragen aus der Literaturrecherche ergeben hat. Die Betrachtung der Studien zum Thema Testtraining hat aufgezeigt, dass es methodisch sehr schwierig ist, die Größe des Trainingseffekts zu bestimmen. Die vorliegende Arbeit versucht zunächst durch die Evaluation des Testtrainings die Wirkung des Trainings zu bestimmen. Welche weiteren Einflussfaktoren neben dem Testtraining bei der Testleistung einen Einfluss haben, wird durch die Analyse an einer Spezialstichprobe untersucht. Durch die vorliegenden Daten der CCH-Laufbahnberatung ist es gelungen, eine homogene Gruppe von Probanden zu erhalten, die in dem Faktor Testtraining konstant ist. So ist es möglich zu überprüfen, wie der Einfluss folgender Variablen auf die Testleistung aussieht: Abiturdurchschnittsnote, Kenntnisse in Mathematik und Medizin, Geschlecht und Versagensangst. Anhand der homogenen Gruppe werden folgende Fragestellungen untersucht:

1. Wie wirkt sich die Abiturdurchschnittsnote auf das TMS-Gesamtestresultat und auf die Ergebnisse der einzelnen Untertests aus?

2. Wie wirken sich gut ausgeprägte Mathematikkennntnisse oder ausgeprägtes medizinisches Vorwissen auf die Resultate der Untertests aus, die gerade dieses Fachwissen prüfen?
3. Wie sehen Unterschiede zwischen Frauen und Männer in der TMS-Gesamttestleistung und bei den einzelnen Untertests aus?
4. Unterscheiden sich Personen mit ausgeprägter Prüfungsangst in ihrer Testleistung zu Personen, die keine Prüfungsangst empfinden?

2.1 Aufbau der Arbeit

Zunächst wird der Test für medizinische Studiengänge (TMS) vorgestellt. Die neun Untertests werden kurz erläutert. Es wird erwähnt, welche Fähigkeiten mit dem jeweiligen Test überprüft werden soll. Im theoretischen Teil der Arbeit werden in Kapitel 4 die Studien zum Testtraining für den TMS vorgestellt. In Kapitel 5 wird ein Überblick über das Auswahlverfahren in den USA und eine Zusammenfassung der Diskussion über den Effekt von Coaching gegeben. Im Anschluss wird das spezielle Testtraining vorgestellt, das alle untersuchten Personen durchlaufen haben. Kapitel 7 stellt das Vorgehen der Transferevaluation und ihre Ergebnisse dar. Die empirischen Befunde zur zusätzlichen Analyse an einer ausgewählten Spezialstichprobe werden in Kapitel 8 beschrieben. Am Ende folgen in Kapitel 9 die Diskussion und das Fazit mit einem Ausblick auf mögliche weitere Studien.

3 Der Test für medizinische Studiengänge (TMS)

Da sich die nachfolgende Untersuchung zum Thema Testtraining auf ein Testtraining für den Test für medizinische Studiengänge (TMS) bezieht, wird der TMS in diesem Kapitel eingehend dargestellt werden.

3.1 Testentstehung

In der Bundesrepublik Deutschland hatte sich in den in 70er Jahren des letzten Jahrhunderts eine nicht haltbare Situation für Bewerber der medizinischen Studiengänge, Human-, Zahn- und Tiermedizin entwickelt. Es gab bei weitem mehr Bewerber für ein medizinisches Studium als es tatsächlich Plätze gab. Es wurden nur noch die Kandidaten ausgewählt, die einen extrem guten Notendurchschnitt (meist besser als 1,4) aufweisen konnten. Die Wartezeit für Abiturienten mit einem schlechteren Durchschnitt nahm dadurch ein zeitliches Ausmaß an, dass als nicht mehr tragbar bezeichnet werden konnte. Die Kultusministerkonferenz entschloss sich für die medizinischen Studiengänge ein „besonderes Auswahlverfahren“ einzuführen. Zukünftig wurden die Studienplätze nach folgenden Quoten vergeben:

- Vorab-Quote: 10 Prozent
Härtefälle: z. B. Zweit-Studium-Bewerber, Nicht-EU-Ausländer etc.
- Abitur-Test-Quote: 45 Prozent
Gewichtung: 55 Prozent Abiturnote und 45 Prozent TMS-Wert
- Test-Quote: 10 Prozent
Die erzielte Testleistung war hier ausschlaggebend, auch Testbestenquote genannt.
- Wartezeit-Quote: 20 Prozent
Die Studienplätze werden an Bewerber vergeben, die sich häufiger beworben haben, aber noch nicht angenommen wurden
- Auswahlgespräch-Quote: 15 Prozent
Bewerber, die über die anderen Quoten nicht zum Zuge gekommen sind, hatten per Losverfahren die Möglichkeit auf ein Auswahlgespräch.

Die Entwicklung und Weiterentwicklung des Tests für medizinische Studiengänge (TMS) gaben die Kultusminister und Senatoren aller Länder der Bundesrepublik Deutschland und des Bundesministers für Bildung und Wissenschaft beim Institut für Begabungsforschung in Auftrag. Der TMS sollte die Eignung von Studienbewerbern für die medizinischen Studiengänge, Human-, Zahn- oder Veterinärmedizin testen. Ab dem Wintersemester 1986 war das „besondere Auswahlverfahren Bestandteil des Hochschulzulassungssystems. Ab 1992 galt es auch in den neuen Bundesländern (Roeder, 1998).

3.2 Testaufbau

Der TMS besteht aus neun Untertests. Die Bearbeitungszeit liegt jeweils zwischen 10 bis 60 Minuten. Die Untertests erfassen spezielle Leistungsfähigkeiten bezüglich des räumlichen Vorstellungsvermögens, der visuellen Wahrnehmung und ferner der Merk- und Konzentrationsfähigkeit. Außerdem werden Fertigkeiten zum logischen Schlussfolgern, zur Aufnahme und Verarbeitung von Informationen, zur Planung und Organisation und auch zum medizinischen und naturwissenschaftlichen Verständnis abgeprüft (Hohner, 2009). Nachfolgend werden die Untertests vorgestellt. Dabei werden die Fähigkeiten angeführt, die zum Lösen des Untertests notwendig sind und deren Ausprägung gemessen werden soll.

Muster zuordnen

Der Untertest bestand aus 24 Aufgaben. Es sollten Ausschnitte in einem vielschichtigen Bild identifiziert werden. Es wurde ein Muster vorgegeben und fünf Musterausschnitte. Der Kandidat sollte den Ausschnitt wiedererkennen, der deckungsgleich mit einem Ausschnitt aus dem vorgegebenen Muster ist. Die Teilnehmer hatten für die Bearbeitung dieses Tests 22 Minuten Zeit. Die Ausschnitte waren nicht verändert in der Größe, sie waren nicht gekippt oder gedreht worden.

Medizinisch-naturwissenschaftliches Grundverständnis

Die Bearbeitungszeit für die 24 Aufgaben dieses Untertests betrug 60 Minuten. Bei diesem Test wurde das Verständnis für medizinische und naturwissenschaftliche Fragen abgeprüft. Jede Aufgabe bestand aus einem kurzen Text zu einem medizinischen Sachverhalt. Abschließend wurde eine Frage gestellt und es wurden fünf Lösungs-

vorschläge zu dem medizinischen Problem gemacht. Der Testteilnehmer konnte durch Interpretieren des Inhalts, logisches Schlussfolgern und Abwägen aller Tatsachen zum richtigen Lösungsvorschlag gelangen. Es wurden keine speziellen Kenntnisse im medizinischen oder naturwissenschaftlichen Bereich vorausgesetzt.

Schlauchfiguren

Für diesen Untertest mit seinen 24 Aufgaben hatten die Bewerber 15 Minuten Zeit. Ziel dieses Untertest war es, das räumliche Vorstellungsvermögen zu prüfen. Jede Aufgabe umfasste zwei Abbildungen eines durchsichtigen Würfels, der ein oder zwei Kabel enthält. Das Kabel war in dem Würfel jeweils auf verschiedene Arten geformt. Das erste Bild zeigte den Würfel von der Vorderansicht. Der Kandidat sollte jetzt erkennen aus welcher Position die zweite Abbildung des Würfels aufgenommen worden war: von rechts, links, oben oder hinten.

Quantitative und formale Probleme

Dieser Test bestand aus 24 Aufgaben, für deren Lösen die Testteilnehmer 60 Minuten Zeit hatten. Anhand von medizinischen und naturwissenschaftlichen Fragestellungen, sollte in diesem Untertest überprüft werden, wie gut der Studienbewerber in der Lage ist, korrekt mit Zahlen, Größen, Einheiten und Formeln umzugehen. Es waren jeweils fünf Antwortmöglichkeiten vorgegeben. Die Lösungen konnten mit Hilfe von physikalischem Grundverständnis und verschiedener mathematischer Methoden, z. B. Dreisatz und Bruchrechnung, errechnet werden.

Konzentriertes und sorgfältiges Arbeiten

Dieser Untertest arbeitete mit einer Zeitvorgabe. Die Teilnehmer hatten 8 Minuten Zeit, möglichst viele vorher bestimmte Zeichen zu markieren. Mit diesem Test sollte überprüft werden, inwieweit der Kandidat in der Lage ist unter Zeitdruck möglichst rasch, sorgfältig und konzentriert zu arbeiten. Der Test bestand aus einer DIN A4-Seite. Darauf waren 20 Zeilen mit jeweils 50 Zeichen gedruckt. Jede Zeile bestand dabei nur aus den Buchstaben „b“ und „q“, die in beliebiger Reihenfolge aufeinander folgten. Jeder Buchstabe war dabei mit entweder ober- und/oder unterhalb mit einem bis vier Querstrichen versehen. Es sollten alle „b“ markiert werden, die zwei Querstriche

aufweisen. Wobei es egal war, wo sich die Querstriche befanden. Gearbeitet wurde zeilenweise von links nach rechts.

Figuren lernen

Mit diesem Test wurde die Gedächtnisleistung und Reproduktionsfähigkeit von visuell wahrgenommenen Figuren geprüft. Es sollte deutlich gemacht werden, wie gut der Teilnehmer in der Lage ist, sich Einzelheiten von Gegenständen einzuprägen. Zunächst begann der Test mit einer vierminütigen Einprägphase. Nach einer Stunde, in der ein anderer Test bearbeitet wurde, folgte die fünfminütige Reproduktionsphase. In der Merkphase wurden den Teilnehmern 20 Figuren präsentiert. Bei jeder Figur war ein Teil eingeschwärzt. Diese Stelle sollten sich die Getesteten merken. In der Wiedergabephase sollten sie dann in einem Zeitraum von fünf Minuten an einer Blanco-Figur die entsprechende Stelle einschwärzen. Die Positionen der Figuren entsprachen der Vorlage in der Einprägphase, aber die Reihenfolge der Figuren war verändert.

Fakten lernen

Die Merkfähigkeit von Fakten wurde in diesem Untertest überprüft. Die Kandidaten erhielten 15 Fallbeschreibungen von Patienten, in denen sie jeweils Namen, Altersgruppe, Beruf, Geschlecht, ein zusätzliches Beschreibungsmerkmal und eine Diagnose erfahren. Sie hatten sechs Minuten Zeit zum Einprägen. Nach einer Stunde, in der ein anderer Test bearbeitet wurde, betrug die Wiedergabephase sieben Minuten, um 20 detaillierte Fragen zu den Patientenangaben zu beantworten.

Textverständnis

In diesem Untertest wurde untersucht, wie gut der Bewerber in der Lage ist, komplexes und ausführliches Material aufzunehmen und daraus Schlussfolgerungen zu ziehen. Dieser Test setzte sich zusammen aus vier Texten mit jeweils sechs Fragen. Die Fragen bezogen sich auf den Inhalt des Textes. Die Bewerber mussten den Text gründlich durchlesen, kognitiv verarbeiten und dann die möglichst passende Antwort aus fünf Alternativen bestimmen.

Diagramme und Tabellen

Für die Bearbeitung dieser 24 Aufgaben standen den Studienbewerber insgesamt 60 Minuten zur Verfügung. Dabei sollten die Testteilnehmer Diagramme und Tabellen

analysieren und richtig interpretieren. In jeder Aufgabe war entweder eine Tabelle oder ein Diagramm abgebildet. Der Teilnehmer musste die Darstellung richtigen deuten, um dann die richtige von fünf möglichen Antworten herauszufinden. (Karnaoukh, 2008)

3.3 Testablauf

Für jeden Untertest gibt es genaue Vorgaben für die Bearbeitungszeit. Wenn der eine Test abgeschlossen war, durfte nicht zurück geblättert werden. Sollte ein Teilnehmer vor der für diesen Untertest angesetzten Bearbeitungszeit fertig sein, durfte er nicht mit dem nächsten Untertest beginnen. Es waren keine Hilfsmittel erlaubt, z. B. Taschenrechner. Bei Verstoß wurde der Teilnehmer aus dem Testverfahren ausgeschlossen und durfte den Test auch nicht noch einmal wiederholen. Insgesamt hat der TMS zusammen mit einer Mittagspause von 60 Minuten ca. 5 Stunden gedauert.

3.4 Testauswertung

Bis auf den Untertest „Konzentriertes und sorgfältiges Arbeiten“, der ein Durchstreichtest war, wurden die Aufgaben nach dem Multiple-Choice-Prinzip gelöst. Es war immer nur eine Lösung richtig. Falsche Antworten und nicht bearbeitete Aufgaben wurden nicht berücksichtigt. Zunächst wurden je Aufgabengruppe die richtig gelösten Aufgaben ausgezählt ("Informationsbroschüre," 2009). Pro Aufgabengruppe konnten maximal 20 Punkte erreicht werden, beim Textverständnis nur 18. Der Untertest „Konzentriertes und sorgfältiges Arbeiten“ wurde anders ausgewertet, aber auch hier lag die maximal erreichbare Punktzahl bei 20. Anschließend wurden alle erreichten Punkte addiert. Die Summe wurde dann in eine Skala mit dem Mittelwert 100 und einer Standardabweichung von 10 umgerechnet. Hatte ein Teilnehmer einen Standardwert von über 100 Punkten erreicht, dann war sein Testergebnis über dem Durchschnitt, bei unter 100 Punkten lag das Ergebnis unter dem Durchschnitt.

3.5 Geschichte des TMS

Der TMS, wie er in dieser Arbeit behandelt wird, wurde in Deutschland von 1986 bis 1997 durchgeführt. Der TMS wurde eingestellt, da die Bewerberzahlen für die

medizinischen Studiengänge deutlich zurückgegangen waren. Die Ergebnisse dieser Studie beziehen sich auf diesen Zeitraum.

Seit 2007 wurde der TMS wieder in Baden-Württemberg eingeführt. Allerdings ist der TMS nicht obligatorisch bei der Bewerbung zum Medizinstudium, sondern er dient den Bewerbern dazu, ihre Zulassungschancen zu verbessern. Sollte das Ergebnis nicht wie gewünscht ausfallen, hat das keine negativen Konsequenzen für den Bewerber.

In der Schweiz wurde der Eignungstest für das Medizinstudium (EMS) im Jahr 1997 für alle Studienbewerber eingeführt. In Österreich müssen seit 2006 alle Bewerber den EMS absolvieren.

4 Testtraining in der deutschsprachigen Literatur

4.1 Studie von Bernhard Deter (1982)

Ein ernstzunehmender Kritikpunkt an der Durchführung eines Studierfähigkeitstest zur Auswahl der zukünftigen Studierenden ist die Frage, ob der Test trainierbar ist. Sollte diese Frage bejaht werden, stellt sich sofort die Frage, ob der Studierfähigkeitstest auch das Gütekriterium der Fairness erfüllt. Deter (1982) hat sich in seiner Dissertation „Zum Einfluß von Übung und Training auf die Leistung »Test für medizinische Studiengänge« (TMS)“ speziell mit der Trainierbarkeit des TMS beschäftigt (Deter, 1982).

Zunächst ist Deter (1982) in seiner Arbeit auf das allgemeine Problem eingegangen, das entsteht, wenn breite Kreise der Bevölkerung einem Gruppentestverfahren gegenüberstehen und dieses Verfahren zur Auslese verwandt wird. Natürlich wird jeder Testteilnehmer alles daran setzen in diesem Testverfahren so gut wie möglich abzuschneiden, um seine Aussichten auf Auswahl zu erhöhen. Der Aspekt der Fairness war aber nicht der einzige Grund für die Untersuchung für Deter (1982). Der Test für medizinische Studiengänge sollte auch gleichzeitig hinsichtlich seiner Reliabilität und Validität als psychometrisches Instrument überprüft werden. Denn sollte der Test trainierbar sein, so wären auch seine Reliabilität und somit auch seine Validität eingeschränkt.

Das zentrale Anliegen Deters (1982) war der Vergleich der Effekte von Testübung und Testtraining mit dem Vorbereitungseffekt der Testbroschüre, die jeder Bewerber zusammen mit den Bewerbungsunterlagen zugeschickt bekam.

Deter (1982) ging in seiner Studie nicht davon aus, dass es keinen Trainingseffekt geben würde. Sein Interesse lag darin zu untersuchen, ob der Test die Gütekriterien Fairness, Reliabilität und Validität erfülle. Sollte aber ein Teil der Testteilnehmer durch die Teilnahme an einen Trainingskurs Vorteile gegenüber anderen erwerben, so würde das Kriterium der Fairness nicht erfüllt werden. Eine Möglichkeit, den höchstwahrscheinlich eintretenden Effekt eines Testtrainings für alle Testteilnehmer konstant zu halten, sah Deter (1982) in der Option, allen Bewerbern von Anfang an eine

Testbroschüre mit allen wichtigen Hinweisen zu den Aufgaben, den Instruktionen und sogar Beispielaufgaben zur Verfügung zu stellen. Das Ziel Deters (1982) war, die beiden möglichen Effekte, die einen Einfluss auf das Testergebnis des TMS haben können, zu vergleichen: den Effekt aus Testübung und Testtraining verglichen mit dem Vorbereitungseffekt der Testbroschüre.

Deter (1982) hat sich für seine Studie ausgiebig mit der Trainingsliteratur beschäftigt. Der größte Teil der Literatur stammt aus den USA, da die Universitäten dort neben anderen Kriterien auch mit Hilfe des Testwerts des SAT (Scholastic Aptitude Test) auswählen, wen sie zum Studium zulassen. Schlagwörter zum Thema Testtraining im angloamerikanischen Bereich sind test-wisness, test-practice und test-coaching. Auf diese Begriffe wird im Kapitel 4 noch einmal näher eingegangen. Bei der Durchsicht der internationalen Literatur zum Thema Testtraining wurde allerdings deutlich, dass ein Vergleich der Testtrainings nicht ganz einfach ist. Denn jedes Testtraining ist in Dauer, Umfang und Inhalt unterschiedlich. Interessant ist jeweils die Größe des Effektes. Eine zufallskritische Überprüfung lässt sich allerdings aufgrund der verschiedenen Randbedingungen in den zu Testtraining gemachten Studien nicht durchführen, so Deter (1982).

Zur Messung der Effektgröße schlug Deter (1982) zwei Indizes vor. Er meinte, dass diese beim Vergleich von Untersuchungsergebnissen zu Trainings- und Übungseffekten besonders nützlich sein könnten. Diese Effektgrößen-Indizes heißen Standardgewinne. Deter (1982) unterschied zwei Arten von Standardgewinnen. Je nachdem, um welches Versuchsdesign es sich handelte, unterschied Deter (1982) bei einem Eingruppen-Design in einen Standardgewinn innerhalb einer Versuchsgruppe und bei einem Kontrollgruppen-Design in einen Standardgewinn zwischen zwei Versuchsgruppen. Standardgewinn innerhalb einer Gruppe ist die Differenz zwischen durchschnittlicher Nachtestleistung und durchschnittlicher Vortestleistung, geteilt durch die Standardabweichung. Standardgewinne zwischen den Gruppen ist die Differenz zwischen den durchschnittlichen Nachtestleistungen in den experimentellen Gruppen und der mittleren Testleistung in der unbehandelten Kontrollgruppe, geteilt durch die Standardabweichung in dieser Kontrollgruppe. Hergeleitet hat Deter (1982) diese Indizes über den Index d , den Cohen (1977, zit. n. Deter, 1982) für den Vergleich der

Mittelwerte zweier unabhängiger Stichproben vorgeschlagen hat, und die Definition der Effektgröße nach Smith & Glass (1977, zit. n. Deter, 1982), die das Maß direkt auf den Vergleich zwischen Experimental- und Kontrollgruppe beziehen.

$$\text{Standardgewinn „innerhalb“: } Z_{G(I)} = \frac{\bar{y} - \bar{x}}{s_x}$$

$$\text{Standardgewinn „zwischen“: } Z_{G(Z)} = \frac{\bar{y}_T - \bar{y}_K}{s_{xK}}$$

x = Vortest (bzw. Ersttestung)

y = Nachtest (bzw. Wiederholungstest)

T = Trainingsgruppe

K = Kontrollgruppe

s = Standardabweichung

Mit Hilfe dieser beiden Indizes können die Übungs- und Trainingsgewinne bestimmt werden. Durch sie können Ergebnisse aus verschiedenen Studien verglichen werden. Standardgewinne beziehen sich auf eine Standardabweichung, also $s = 1$. Allerdings sind auch bei diesem deskriptiven Maß immer die Randbedingungen des Vorbereitungsprogramms bzw. des Versuchs zu beachten. Deter (1982) wies ausdrücklich darauf hin, dass es sich um ein deskriptives Maß handelt und nicht um ein inferenzstatistisches handele. Die deskriptive Statistik kommt, so Deter (1982) bei Studien, die den gleichen Untersuchungsgegenstand haben, eine ebenso wichtige Rolle zu wie die schließende Statistik. Carver (1978, zit. n. Deter, 1982) schrieb, dass man durch die deskriptive Statistik erkennen könne, welche Art von Beweisen man für seine Forschungshypothesen habe: starke, schwache oder auch gar keine Unterstützung. Deter (1982) führte dieses Maß eines Effektgrößen-Index ein, um mit Hilfe des Standardgewinns feststellen zu können, ob seine Ergebnisse nützlich sind.

Zu dem Einfluss von Testübung und Testwiederholung kam Deter (1982) aufgrund seiner Recherchen zu nachfolgenden Aussagen. Generell kann man davon ausgehen, dass sich durch Wiederholung eines Tests immer ein Effekt einstellt - ein

Gedächtniseffekt. Dieser Effekt wurde auch bei einer Untersuchung von Fay (1980, zit. n. Deter, 1982) zur zeitlichen Stabilität des TMS gefunden. Der Gedächtniseffekt nimmt ab, wenn der Zeitraum zwischen Erst- und Wiederholungstestung steigt (Lienert, 1969 zit. n. Deter, 1982). Ebenso ist dieser Erinnerungseffekt geringer, je mehr Aufgaben ein Test umfasst (Michel, 1964 zit. n. Deter 1982). Verbunden mit der Testwiederholung ist auch immer der Aspekt der Vertrautheit. Der Testteilnehmer macht sich mit dem Test und den Testaufgaben vertraut und auch die Testsituation ist in der Wiederholung für den Teilnehmer schon viel bekannter als bei der Ersttestung. Auch dieser Aspekt hat laut Deter (1982) neben dem Übungsgewinn einen Einfluss auf das Testergebnis. Diese Gewinne durch Testwiederholung sind allerdings lediglich Übungsgewinne und haben nichts mit Trainingsgewinnen zu tun.

Deter (1982) unterscheidet durch seine Literaturrecherche in Übungs- und Testwiederholungsgewinne und Testtrainingsgewinne. Am schwierigsten ist es hierbei die Trainingseffekte zu vergleichen, da sich die Vorbereitungsprogramme in Inhalt, Umfang und Dauer stark unterscheiden können. Aber nicht nur die Trainingsprogramme unterscheiden sich, sondern auch die Testverfahren und auch die Versuchspersonen sind nicht einfach miteinander vergleichbar. So ist es bezüglich des TMS sinnvoll, so Deter (1982), wenn man sich beim Vergleich bezüglich der Trainingseffekte auf Studien entweder mit Oberprimanern oder Studenten konzentriert, die mit der Testvorbereitung darauf abzielen, ihre Leistungen in einen Studieneignungstest zu verbessern.

Bei den Untersuchungen zu den Vorbereitungskursen für den SAT, laut Deter (1982), zeigte sich, dass je länger ein Trainingskurs dauerte, umso größer waren auch die Lerngewinne. Die Größe des Trainingseffekts hängt demzufolge auch vom Zeitfaktor ab.

Deter (1982) leitete aus der Literaturrecherche verschiedene Hypothesen ab, die er in Hinsicht auf den TMS und seine zu erwartenden Trainingseffekte machen kann. Er bezog sich dabei auf die Aussagen, die über die Übungs- und Testwiederholung hinaus gemacht wurden. In diesem Teil der Arbeit wird deutlich, dass mit Training immer eine Testleistungssteigerung verbunden ist, die aber in der Regel im Zusammenhang mit der Testwiederholung steht und somit als Testübungseffekt zu bezeichnen ist.

Eine wichtige Schlussfolgerung Deters (1982) aus seiner Literaturrecherche war zudem, dass eine Testvorbereitung nicht unbedingt die Validität eines Testwerts reduziert. Es sei denn, die Testleistung wird durch Testbearbeitungsstrategien künstlich erhöht. Dann wäre die Testleistung kein getreues Abbild der Fähigkeiten des Teilnehmers.

Zentrale Bedeutung für eine Leistungssteigerung im Test hat für Deter (1982) der Aspekt der Vertrautheit. Dies führt ihn zu seiner Fragestellung, ob ein leistungssteigernder Effekt hervorgerufen durch die Vertrautheit und vermittelt in einem Vorbereitungskurs auch durch Informations- und Übungsmaterial kompensiert werden könnte?

Zusammenfassend ergab die Literaturrecherche, dass sich keine Schlussfolgerungen über den konkreten Umfang der Lerngewinne machen lassen, da der Einfluss von Trainingsverfahren mit dem Umfang, dem Inhalt und der Dauer der Vorbereitung variiert.

Im ersten Abschnitt des empirischen Teils der Studie Deters (1982) ging es um die Gegenüberstellung der Effekte eines kurzfristigen, intensiven Trainings und dem Effekt der Vorbereitung mit der Testbroschüre. Sein Versuchsplan setzte fünf Versuchsgruppen fest. Details seiner Studie sind in der Tabelle 4.1 dargestellt.

Tabelle 4.1: Elemente der experimentellen Studie von Deter (1982)

Versuchsplan	Messinstrumente	Stichprobe	Durchführung
Solomon-Plan mit 5 Gruppen		N = 439	
	TMS	Gruppe I = 96	Vortest
	Trainingsmaterial und Trainingsgestaltung	Gruppe II = 47	Training
	Fragebogen zur Begleituntersuchung	Gruppe III = 49	Nachtest
		Gruppe IV = 202	Studium der Broschüre
		Gruppe V = 45	

Anmerkungen: Der TMS war für Vor- und Nachtest der gleiche.

Alle Probanden wurden den Versuchsgruppen nach Zufall zugeordnet. Die Gruppe I als Experimentalgruppe ist mengenmäßig die größte, da Deter (1982) mit Hilfe der Ergebnisse dieser Gruppe überprüfen wollte, ob es durch Einflüsse aus Testübung und Testtrainings dazu kommen kann, dass sich die faktorielle Zusammensetzung der

gemessenen Testleistung verändert. Die Voraussetzung für eine solche Überprüfung setzt eine Mindestgruppengröße voraus. In dieser Arbeit wird auf diesen Aspekt von Deters (1982) Untersuchung nicht näher eingegangen.

In der Untersuchungsstichprobe waren geringfügig mehr Männer als Frauen, was der Studienbewerberpopulation entsprach. Im Mittel waren die Teilnehmer 21,5 Jahre alt. Auch die Verteilung der Abiturdurchschnittsnoten entsprach der Studienbewerberpopulation. Die besseren Abiturdurchschnittsnoten waren stärker repräsentiert.

Deter (1982) gebrauchte bei der Erstellung seines Versuchsplans die Vorlage des Solomon-Viergruppenplan. Der Solomon-Plan (Bortz & Döring, 2006) versucht durch sein Design, mögliche Pretesteffekte zu überprüfen. Auch bei Deter (1982) ging es um die Kontrolle möglicher Störvariablen und um die interne und externe Validität. Der eigentliche Solomon-Plan ist ein Viergruppensdesign zur Untersuchung eines Treatments. Vorher soll aber anhand eines Vortest festgestellt werden, ob die Stichproben überhaupt vergleichbar sind. Dieser Vortest bleibt nicht ohne Einfluss auf die Testleistung. Durch den Solomon-Viergruppenplan kann dieser Effekt kontrolliert werden. Bei Deter (1982) waren solche Pretesteffekte zur Überprüfung von Übungseffekten durch die Wiederholungsprüfung eingeplant. Deters (1982) Versuchsdesign ist in Tabelle 4.2 dargestellt.

Tabelle 4.2: Versuchsplan von Deter (1982)

Gruppe	Versuchsablauf		
I	Vortest	Training	Nachtest
II	-	Training	Nachtest
III	Vortest	-	Nachtest
IV	-	-	Nachtest
V	-	Studium der Broschüre	Nachtest

Anmerkungen: Zwischen Vor- und Nachtest lag ein Zeitraum von 4 Wochen.

Aus der Literaturrecherche ging eindeutig hervor, dass es zu einem Übungseffekt kommen wird. Deter (1982) hat den Plan um eine fünfte Gruppe erweitert. Die fünfte Gruppe war die Gruppe, die die Testbroschüre durcharbeiten sollte.

Die Gestaltung des Trainings war besonders anspruchsvoll. Da Deter (1982) vermutete, dass – sobald der Test für medizinische Studiengänge eingeführt werden würde – kommerzielle Unternehmen Testtrainings für den TMS anbieten würden, sollte dieses Training mit den später angebotenen Trainings nach Möglichkeit gut vergleichbar sein. Zunächst einmal ging Deter (1982) davon aus, dass die Vorbereitungskurse eher kurzfristig sein würden, z. B. Wochenendkurse. Der Vorbereitungskurs, der für diese Untersuchung entwickelt wurde, dauerte zwei Tage à drei Stunden. Ein Schwerpunkt des Trainings lag auf der Testübung. Es wurde aus 17 zusammengestellten Übungstests insgesamt 348 Übungsaufgaben für das Training ausgewählt. Die Testfragen stammten aus verschiedenen Testverfahren, Intelligenz- und Eignungstests, aber hauptsächlich aus Studierfähigkeitstests. Für den Untertest „Fakten lernen“ wurde eine komplette Entsprechung entwickelt, da diese Art von Test bisher im deutschsprachigen Raum noch nicht eingesetzt worden war. Daneben waren die Übungsaufgaben so ausgesucht worden, dass sie den TMS-Aufgaben möglichst ähnlich waren. Die Teilnehmer konnten ihre Lösungen mit einem bereitgestellten Lösungsschlüssel selbstständig vergleichen. Ansonsten wurden den Kursteilnehmern spezifische Informationen zum Verständnis und zur Bearbeitung der jeweiligen Aufgabengruppen gegeben und mit Hilfe von Diskussion über die Lösungswege versucht, den Lernerfolg zu verstärken. Zusätzlich wurden die Versuchspersonen der Gruppe I in bestimmten Mnemotechniken unterwiesen. Alle Hinweise gingen deutlich über das Maß von Information hinaus, das die Gruppe V durch die Bearbeitung der Testbroschüre erhielt.

Die Testbroschüre, die die Teilnehmer der Gruppe V bearbeiteten, enthielt eine Beschreibung des Tests, Bearbeitungshinweise, Beispielaufgaben und die Aufforderung alles gründlich durcharbeiten und sich die Informationen kurz vor der eigentlichen Testsitzung noch einmal in Erinnerung zu rufen. Die Gruppe V beschäftigte sich selbstständig während einer dreistündigen, beaufsichtigten Sitzung mit der Testbroschüre und nahm die Broschüre dann mit nach Hause, um sich dort weiter auf

den Test vorzubereiten. Dabei wurde den Teilnehmern der Gruppe V vorgeschlagen, wie sie bei der Bearbeitung der Broschüre vorgehen sollten.

Zunächst wertete Deter (1982) die Nachtestwerte nach der von Campbell & Stanley (1970, zit. n. Deter, 1982) für den Solomon-Viergruppenplan vorgeschlagenen zweifaktoriellen Varianzanalyse aus. Der Solomon-Viergruppenplan beinhaltete in dieser Studie die Gruppen I bis IV aus. Die beiden Haupteffekte in der zweifaktoriellen Varianzanalyse waren „Testtraining“ und „Testwiederholung“. In Bild 4.1 sind die beiden Haupteffekte und die ordinale Interaktion für den Gesamttest graphisch dargestellt.

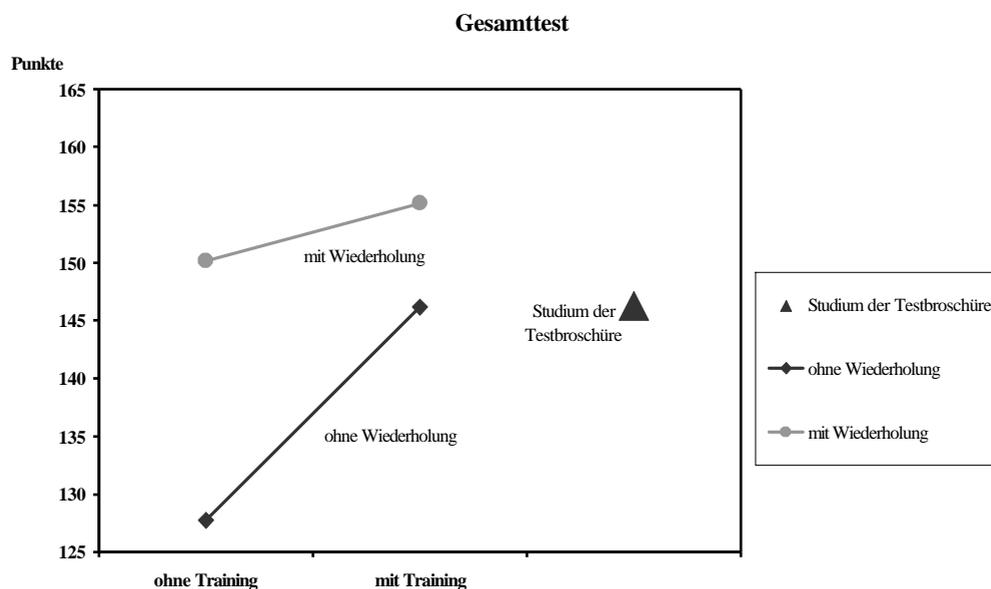


Bild 4.1: Veranschaulichung der Haupteffekte und Interaktion nach den Ergebnissen aus Deters (1982) Studie

Für den Gesamttest waren beide Haupteffekte signifikant. Die Untertests hingegen weisen sowohl bei dem Haupteffekt „Testtraining“ als auch „Testwiederholung“ nicht immer Signifikanzen auf. Beim Gesamttest und auch bei einigen Untertests kann man einige signifikante Interaktionen erfassen. Allerdings war nur die Interaktion beim Untertest „mathematisches Grundverständnis“ disordinal. In diesem Fall ist es besser die signifikanten Haupteffekte - wobei für den Untertest „mathematisches Grundverständnis“ sowieso keiner der Haupteffekte signifikant war - nicht zu interpretieren. Beim Gesamttest und auch bei drei weiteren Untertests handelte es sich um eine

ordinale Interaktion. Die Faktoren konnten nicht global interpretiert werden (Bortz & Döring, 2006).

Die Nachtestwerte hat Deter (1982) mit einer einfaktoriellen Varianzanalyse verglichen. Beim Gesamttest und bei fast allen Untertests - bis auf die Untertests Satzergänzung und Textverständnis - konnten signifikante Behandlungseffekte festgehalten werden. Die besten durchschnittlichen Testleistungen erbrachte die Gruppe I, die Gruppe mit dem Training und der Wiederholungstestung, die Gruppe III, die Wiederholungstestung, folgte auf Platz 2, auf Platz 3 kamen die Gruppen II (nur Training) und Gruppe V (Testbroschüre). Am schlechtesten schnitt im Durchschnitt die Gruppe IV ab, die weder ein Training hatte, noch wiederholt getestet wurde und auch keine Testbroschüre bearbeitet hatte.

Die Ergebnisse der multiplen Mittelwertvergleiche mit Hilfe des Scheffé-Tests zeigten, dass sich signifikante Ergebnisse der Varianzanalyse ausschließlich auf die schwachen Testleistungen der Gruppe IV zurückführen ließen. Die Gruppe V, die für Deters (1982) Untersuchung wichtigste Gruppe, unterschied sich nicht signifikant von der Gruppe II, der Gruppe, die das Testtraining durchlaufen hatte. In Tabelle 4.3 ist eine Übersicht der durchschnittlichen Testleistungen und der Lerngewinne, die in Standardgewinne zwischen den Gruppen, bezogen auf Mittelwert und Standardabweichung der Kontrollgruppe IV unterteilt sind, dargestellt. Deutlich ist zu erkennen, dass die Gruppe I den höchsten Standardgewinn, über eine Standardabweichung, verzeichnen konnte. Die Lerngewinne der Gruppe II (Testtraining) und Gruppe V (Testbroschüre) waren hingegen gleich.

Tabelle 4.3: Durchschnittliche Testleistung und Lerngewinne dargestellt als Standardgewinne nach Deter (1982)

	Gruppe I ^a	Gruppe II ^b	Gruppe III ^c	Gruppe IV ^d	Gruppe V ^e
$M (SD)$	155.16 (18.93)	146.19 (18.48)	150.20 (17.46)	127.69 (20.39)	146.24 (19.05)
${}^f Z_{G(Z)}$	1.35	0.91	1.10		0.91

Anmerkungen: ^aGruppe I: Testtraining + Testwiederholung, ^bGruppe II: Testtraining, ^cGruppe III: Testwiederholung, ^dGruppe IV: Kontrollgruppe, ^eGruppe V: Testbroschüre, ^f $Z_{G(Z)}$: Standardgewinne zwischen den Gruppen

Deter (1982) verglich die Übungs- und Trainingsgewinne mit den Gewinnen aus Studien über ähnliche Testverfahren, wie SAT oder MCAT¹. Die Leistungszunahmen waren vergleichsweise beachtlich. Hierbei könnte sofort die Frage aufkommen, ob der TMS anfälliger auf Trainings- und Wiederholungseffekte reagiert als andere Tests. Die hohen Leistungsgewinne könnten aber auch dadurch zustande gekommen sein, so Deter (1982), dass in seiner Studie der Vortest der gleiche war wie der Nachtest. Bedauerlicherweise gab es keine Paralleltestform. Außerdem lag zwischen den beiden Testungen ein relativ kurzer Zeitraum von nur vier Wochen. Es kommt gewöhnlich zu geringeren Wiederholungseffekten, wenn der Zeitraum zwischen den Testwiederholungen länger ist. Zudem haben deutsche Schüler wenig Testvorerfahrung. Die Testwiederholungseffekte sind umso höher je weniger Testvorerfahrung der Getestete hat.

Deter (1982) wies auch noch einmal darauf hin, dass sich die Studien aufgrund der verschiedenen Randbedingungen nur schwer miteinander vergleichen lassen. Deter (1982) betonte ein weiteres Mal, dass es in seiner Studie nicht darum ginge, nachzuweisen, dass der TMS unempfindlich gegenüber Übungs- und Trainingseinwirkungen sei, sondern zu untersuchen, ob diese Trainings- und Übungseffekte durch geeignetes Zusatzmaterial, wie die Testbroschüre, ausgeglichen werden könnten. In seiner Studie sah Deter (1982) diesen Aspekt bestätigt. Die Gruppe V (Testbroschüre) konnte die gleiche Leistungssteigerung verzeichnen wie die Gruppe II (Testtraining). Einzig eine Gruppe untersuchte Deter (1982), wie er auch selbst erwähnte, nicht: die Gruppe, die sowohl ein Training absolviert hat, als auch an einer Testwiederholung teilgenommen hat und zusätzlich die Testbroschüre durchgearbeitet hat.

Deter (1982) hat in seiner Studie auch andere mögliche Effekte untersucht. So betrachtete er näher, ob es einen Zusammenhang zwischen dem Niveau der Vortestleistungen und der Größe der Trainings- und Übungsgewinne geben könne. Allerdings konnte er diesen eventuellen Effekt aufgrund der zu geringen Teilnehmerzahl in der Gruppe III nicht überprüfen. Zusammengefasst nahm Deter

¹ SAT = Scholastic Aptitude Test für die Zulassung zu Colleges in den USA und MCAT = Medical College Admission Test

(1982) einen positiven Zusammenhang an zwischen Teilnehmern, die im Vortest überdurchschnittlich hoch abgeschnitten hatten, und der Größe des Übungsgewinns. Bei genauerer Betrachtung der Zusammenhänge zwischen dem Merkmal „Geschlecht“ und der Größe der Trainings- und Übungsgewinne konnte kein übereinstimmendes Ergebnis weder für Männer noch für Frauen erkannt werden.

Interessant war die Betrachtung des möglichen Einflusses der Vorerfahrung mit psychologischen Tests auf das Ergebnis im TMS. Durch die Daten aus dem Fragebogen zur Begleituntersuchung konnte Deter (1982) die Teilnehmer in Testunerfahrene und Testerfahrene unterteilen. Die Ergebnisse einer einfaktoriellen Varianzanalyse gebildet aus den Vortestwerten der Gruppen I, Gruppe III und den Nachtestwerten der Gruppe IV zeigte sich signifikant auf dem 5 %-Niveau. Eine zweifaktorielle Varianzanalyse, Testerfahrung und Gruppenzugehörigkeit, für alle Nachtestwerte wurde hingegen nur noch für den Haupteffekt Gruppenzugehörigkeit signifikant. Für Deter (1982) war dies ein Beleg, dass sich die am Anfang existierenden Unterschiede in den Testergebnissen zwischen Teilnehmern mit Erfahrung und denjenigen ohne Erfahrung, die durch die Berechnung der einfaktoriellen Varianzanalyse festgestellt wurden, durch die verschiedenen Treatments aufheben. Deter (1982) verglich mit einem t-Test für Mittelwertvergleiche bei unabhängigen Stichproben die Gruppe V, die die Testbroschüre bearbeitet hatte, mit der Gruppe aller Testerfahrenen aus Gruppe I, III (hier die Vortestwerte) und IV (hier der Nachtestwert), die aber die Testbroschüre nicht kannten. Der t-Wert war auf dem 1 %-Niveau signifikant. Die testerfahrenen Bewerber erreichten einen durchschnittlichen Gesamtestwert von 131.76 Punkten. Die testunerfahrenen Broschürebearbeiter erlangten hingegen einen durchschnittlichen Wert von 146.24 Punkten. Deter (1982) schloss daraus, dass frühere Testerfahrung die Leistung im TMS zwar vorteilhaft beeinflusst, aber die Teilnehmer der Gruppe V durch die Bearbeitung der Testbroschüre im Vergleich dazu erheblich höhere Leistungsgewinne verzeichnen konnten.

Insgesamt sah Deter (1982) sich durch seine experimentelle Studie darin bestätigt, dass ein gewissenhaftes Bearbeiten der Testbroschüre, den Testerfahrungsmangel ausgleichen kann und genauso wie ein kurzfristiges Training zu einer Ergebnissteigerung führen kann. Wobei Deter (1982) aber auch einschränkend

erwähnte, dass die Bearbeitungssituation der Testbroschüren-Gruppe einer Laborsituation glich, da sie die Testbroschüre unter Aufsicht bearbeiten mussten. Die erwarteten Leistungssteigerungen, die durch die Beschäftigung mit der Broschüre zu erwarten seien, in der Realität wahrscheinlich geringer sind. Zusätzlich wies Deter (1982) darauf hin, dass die externe Validität dieser Studie durch die Variablen Inhalt, Umfang und Dauer eines Trainings eingeschränkt ist, die maßgeblich über den Erfolg eines Trainings entscheiden. Häufig könne man Trainingsstudien nicht untereinander vergleichen, da diese Trainingsvariablen unterschiedliche Ausprägungen haben.

Deter (1982) machte in seiner Dissertation zusätzlich zu seiner experimentellen Studie noch Untersuchungen an der Gruppe der Medizinstudienplatzbewerber die sich dem TMS in der dreijährigen Übergangsfrist für das Zulassungsverfahren zu den medizinischen Studiengängen stellte. In diesen drei Jahren hatten die Studienplatzbewerber die Möglichkeit freiwillig an dem Test teilzunehmen, um ihre Chancen auf einen Studienplatz zu verbessern. Ein Teil der Studienplätze wurde in dieser Zeit schon über das Kriterium Testleistung vergeben. Pro Semester konnten 6.000 Bewerber an dem neuen Vergabeverfahren teilnehmen. Das Ganze diente dazu, weitere Aufschlüsse über die prognostische Validität des TMSs zu erhalten. Die Bewerber, die ausgelost wurden, am Test teilnehmen zu können, erhielten mit der Testeinladung einen Begleitfragebogen. Deters (1982) Untersuchungen bezogen sich auf die Auswertung der Daten aus dem Begleitfragebogen.

Die erste Analyse galt dem ersten bundesweiten Testeinsatz im August 1980 für das Wintersemester '80. Anhand der gewonnenen Daten untersuchte Deter (1982) den Einfluss von Vorerfahrung mit psychologischen Tests. Hier wurde das Ergebnis der experimentellen Studie betätigt, dass testerfahrene Teilnehmer im Durchschnitt etwas bessere Ergebnisse hatten als Getestete, die noch nie an einem Eignungs- bzw. Intelligenztest teilgenommen hatten. Trennt man aber die Testteilnehmer nach Frauen und Männern, so fällt auf, dass Männer schon viel häufiger an den genannten Testverfahren teilgenommen hatten. Der Effekt der Testerfahrung verringerte sich, wenn er für männliche und weibliche Bewerber getrennt ermittelt wurde. In der Gesamtgruppe ist der Vorerfahrungseffekt folglich eher überschätzt worden.

Die zweite Analyse erhob Deter (1982) anhand des zweiten bundesweiten Testeinsatzes im Februar 1981 für das Sommersemester 1981. Wieder wurde die Frage untersucht, ob Teilnehmer, die Testerfahrung haben, bessere Testleistungen erbringen als solche, die noch keine Erfahrung mit Eignungs- bzw. Intelligenztests haben. In dem Fragebogen zur Begleituntersuchung war diesmal aber die Frage zur Art der Testvorbereitung aufgenommen. Die Frage zur Art der Testvorbereitung wurde in diesem Fragebogen zweimal gestellt. Sowohl die Frage 13 als auch die Frage 17 zielten auf die Erfassung dieser Daten ab. Nur die Formulierungen der Fragen unterschieden sich, so dass Deter (1982) bei der Auswertung dieser beiden Fragen die Möglichkeit hatte, das Antwortverhalten auf die eine Frage mit Hilfe der Antworten auf die zweite Frage zu kontrollieren. Die Anwärter konnten für ihre Bewerbung zum Sommersemester 1981 unter folgenden Vorbereitungsmaßnahmen auswählen: Trainingskurse, angeboten von kommerziellen Trainingsinstituten, Testbroschüre, kommerzielle Trainingsbücher und die vom ITB² herausgegebene Originalversion des TMS vom Sommer 1980.

Zunächst einmal bestätigte sich wie schon aus der Analyse der Daten aus dem Wintersemester 1980, dass ein geringer Effekt der Testvorerfahrung aus Eignungs- oder Intelligenztests vorhanden ist, dieser sich jedoch auf den höheren Anteil männlicher Studienplatzbewerber zurückführen lässt. Ferner fand Deter (1982) seine Ergebnisse aus der experimentellen Studie bestätigt, dass Testunerfahrenheit durch das Studium der Testbroschüre ausgeglichen werden kann.

Die Untersuchungen zu unterschiedlichen Arten der Vorbereitung ergaben, dass Teilnehmer, die angaben unter anderem einen Trainingskurs besucht zu haben, durchschnittlich am besten abschnitten. Gefolgt von denjenigen, die die Originalversion bearbeitet hatten. An dritter Position kam die Gruppe, die schon einmal eine Vorform des TMS bearbeitet hatte, dicht gefolgt von der Gruppe mit der Testbroschüre. Die Gruppe, die gar nichts zur Vorbereitung unternommen hatte, schnitt 20 Punkte schlechter ab als die Trainingsgruppe und 11 Punkte schlechter als die Gruppe mit der Testbroschüre. Deter (1982) sieht die zentrale Aussage seiner experimentelle Studie bestätigt, dass die Testbroschüre mit ihren Instruktionen, Bearbeitungshinweisen und

² ITB = Institut für Test- und Begabungsforschung in Bonn

Beispielaufgaben zu einer Vertrautheit mit dem TMS führe, die sich in einem durchschnittlich besseren Testergebnis im Vergleich zu den Teilnehmern, die sich nicht vorbereitet haben, nieder schlägt.

Deter (1982) wies bei der Darstellung der Analyseergebnisse zum zweiten bundesweiten Testeinsatz aber darauf hin, dass die von den Testteilnehmern angegebenen Daten nicht immer korrekt sein müssen. So zeigte sich bei der Auswertung der Daten der zweiten Frage (Frage 17 im Fragebogen) zu den Vorbereitungsarten, dass sehr viel mehr Teilnehmer angaben, an einem Trainingskurs teilgenommen bzw. sich mit Hilfe eines Trainingskurses informiert zu haben, als sie es in der Frage 13 getan hatten. Deter (1982) schloss daraus, dass die Unterschiedlichkeit der Angaben eventuell darauf zurückzuführen sei, dass in der Testbroschüre explizit von dem Besuch kommerzieller Trainingskurse abgeraten wurde.

Zudem machte Deter (1982) darauf aufmerksam, dass der Vergleich der durchschnittlichen Testleistung der Gruppen mit den jeweiligen Vorbereitungsmaßnahmen noch andere Einflüsse auf das Testergebnis widerspiegeln als es die Ergebnisse der experimentellen Studie wahrscheinlich taten. In der experimentellen Studie waren die Versuchspersonen dem entsprechenden Treatment per Zufall zugeordnet worden. In der Analyse des zweiten bundesweiten Testeinsatzes könnte die Inanspruchnahme mehrerer intensiver Vorbereitungsmaßnahmen für den Test auch die Motivation zum Erreichen eines möglichst guten Testergebnisses wiedergeben. Höchstwahrscheinlich wird ein Teilnehmer, der hohe finanzielle Ausgaben für die Vorbereitungsmaßnahme auf sich nehmen musste, motivierter sein sich gründlicher vorzubereiten als ein Bewerber, der beschlossen hat, sich nur mit der kostenlosen Testbroschüre vorzubereiten.

Deter (1982) verglich die Testleistung in Bezug auf die Kombination der Vorbereitungsmaßnahmen, wie sie in der Frage 13 gestellt wurde, und kam zu dem Ergebnis, dass diejenigen, die sich überhaupt nicht vorbereitet hatten im durchschnittlichen Gesamtergebnis bei weitem am schlechtesten abschnitten. Beim Vergleich des durchschnittlichen Gesamtergebnisses in Bezug auf die Art der Vorbereitung ergab sich die in Tabelle 4.4 dargestellte Reihenfolge.

Tabelle 4.4: Reihenfolge des durchschnittlichen Gesamtergebnisses beim zweiten bundesweiten Testeinsatz bezogen auf die Art der Vorbereitung (Daten stammen aus der Beantwortung der Frage 13)

Gruppe	Reihenfolge ^a
Trainingskurs + Originalversion + Testbroschüre	1. Platz mit 162.12 Punkten
Originalversion + Testbroschüre	2. Platz mit 158.51 Punkten
Originalversion + Trainingsbuch + Testbroschüre	3. Platz mit 157.46 Punkten
Vorform des TMS + Testbroschüre	4. Platz mit 153.96 Punkten
Testbroschüre	5. Platz mit 150.06 Punkten
Vorform des TMS	6. Platz mit 145.35 Punkten
Trainingsbuch + Testbroschüre	7. Platz mit 145.33 Punkten
Originalversion	8. Platz mit 142.95 Punkten
keine Vorbereitung	9. Platz mit 140.20 Punkten

Anmerkungen: ^aReihenfolge der Platzierung mit der durchschnittlich erreichten Gesamtpunktzahl in dieser Gruppe

Besonders in den Untertests „konzentriertes und sorgfältiges Arbeiten“ und „Figuren lernen“ waren die Vorbereitungseffekte am deutlichsten zu erkennen. Hier wurde eindeutig sichtbar, dass diejenigen, die keine Form der Vorbereitung genutzt hatten, schlechter abschnitten als die Vorbereiteten. Es zeigte sich, dass es bei beiden Untertests von Vorteil war, mit der Aufgabe und der Testinstruktion vertraut zu sein.

Wie schon oben erwähnt, gab es in dem Fragebogen zwei Fragen, die darauf abzielten die Art der Vorbereitung in Erfahrung zu bringen. Die Menge derjenigen, die angaben sich mit Hilfe eines Trainingskurses eingearbeitet zu haben, war bei der Frage 17 im Vergleich zur Beantwortung der Frage 13, die quasi den gleichen Inhalt hatte, doppelt so groß. Deter (1982) konnte bei einem erneuten Mittelwertvergleich des Gesamtergebnisses, ausgewertet aus den Daten der Frage 17, keinen deutlichen Vorteil der Trainingsgruppe erkennen. Deter (1982) weist in Folge des inkonsistenten Antwortstils bezüglich der Teilnahme von Trainingskursen darauf hin, dass man anhand der Daten aus dem zweiten bundesweiten Testeinsatz keine gesicherten Aussagen über die Effektivität von Trainingskursen machen könne.

Eine Untersuchung der Frage 13, ob Teilnehmer eines Trainingskurses tendenziell im Vergleich zu anderen Bewerbern, die andere oder keine Vorbereitungsmaßnahmen

unternommen hatten, häufiger zum Medizinstudium zugelassen werden, zeigte eine leicht erhöhte Chance für diese Bewerber. Dieser Effekt wiederholte sich aber wiederum nicht, wenn die inhaltlich ähnliche Frage 17 bezüglich der Zulassungschancen ausgewertet wurde.

Das für Deter (1982) wichtigste Ergebnis war, dass den Bewerbern durch die Bereitstellung der kostenlosen Testbroschüre und ab dem zweiten bundesweiten Testeinsatz auch durch die im Handel erhältliche veröffentlichte Originalversion die Möglichkeit geboten, sich adäquat auf den Test vorzubereiten und mögliche Nachteile, die durch Testunerfahrenheit hätten entstehen können, auszugleichen. Diese Art der Bereitstellung von Vorbereitungsmaterial ist für Deter (1982) nicht nur hinsichtlich der Chancengleichheit selbstverständlich sondern auch aus Gründen der Validität. Ein Teilnehmer, der die Instruktion nicht richtig verstanden hat, wird kein valides Ergebnis abliefern. Ziel der Einführung eines Studierfähigkeitstests ist aber immer auch, geeigneten Studienbewerber zu übersehen.

Deter (1982) vertritt daher die Ansicht, dass der Effekt der Teilnahme an einem kurzfristigen kommerziellen Trainings durch die Testbroschüre und die veröffentlichte Original-version ausgeglichen ist.

4.1.1 Anmerkungen zu Deters (1982) Studie

In seiner experimentellen Studie betonte Deter (1982), dass er versucht hatte, dass experimentelle Treatment Testtraining so gut wie möglich einem zu erwartenden kommerziellen Testtraining nachzubilden. Der Effekt eines Testtrainings sollte möglichst gut mit dem Effekt der Testbroschüre vergleichbar sein. Deter (1982) hatte gerade in dieser experimentellen Studie die Möglichkeit, die Testbroschüre optimal mit dem Testtraining zu vergleichen. Die Versuchsgruppen waren randomisiert und da die Bewerber noch keinen Nachteil aus einem schlechten Testergebnis haben würden, waren die Chancen, dass sie sich tatsächlich an die Vorgaben halten würden, sehr hoch. Trotzdem hatte Deter (1982) sich entschieden, die für die Gruppe V entwickelte Testbroschüre den Gruppen I und II, die beide ein Testtraining absolvieren würden, nicht zur Verfügung zu stellen. Die Frage ist hier, ob die beiden Effekte überhaupt miteinander vergleichbar sind: die eine Gruppe V wurde mit Aufgaben, die im TMS tatsächlich so konstruiert sind, vertraut gemacht, wobei die anderen Gruppen anhand

von ähnlichen Aufgaben vorbereitet wurden. Die eine Gruppe erhielt eine Beschreibung des Tests und Bearbeitungshinweise für die im TMS enthaltenen Aufgaben. Der anderen Gruppe wurden Bearbeitungshinweise für Aufgaben gegeben, mit denen sie sich während des Testtrainings beschäftigten. Diese Aufgaben waren den TMS-Aufgaben aber nur nachempfunden und entsprachen nicht den tatsächlich im TMS enthaltenen Aufgaben. Es könnte sein, dass die Gruppe V durch die Bearbeitung der Testbroschüre einen viel höheren Übungsgewinn hat als die Gruppe II, die nur ein Testtraining mit ähnlichen Aufgaben durchlief.

Deters (1982) zentrales Anliegen war es zu untersuchen, ob eine mögliche Testerfahrung in Eignungs- bzw. Intelligenztests einen Vorteil mit sich bringt und ob dieser Vorteil durch das Bearbeiten einer Testbroschüre ausgeglichen werden kann. Anzumerken ist, dass Deter (1982) beim Vergleich der Vor- bzw. Nachtestmittelwerte zum Feststellen eines eventuellen Vorteils durch Testerfahrung in die Gruppe der Erfahrenen nur die Personen aus den Gruppen I, III und aus der Gruppe IV eingeschlossen hat. In die Vergleichsgruppe, der Gruppe Testbroschüre, wurden aber alle Kandidaten aufgenommen, die mit und die ohne Erfahrung. Dass sich die Mittelwerte nicht signifikant unterschieden, könnte also auch daran liegen, dass die erfahrenen Testbroschürenbearbeiter durch ihre Testvorerfahrung einfach besser abschnitten und der Mittelwert dadurch höher lag.

Schwierig ist bei der Analyse des zweiten bundesweiten Einsatzes des Testes zudem, wie Deter (1982) bereits erwähnte, dass man nicht davon ausgehen kann, dass die Fragen zur Art der Vorbereitung wahrheitsgemäß beantwortet wurden. Gerade der Umstand, dass in der Testbroschüre extra darauf hingewiesen wurde, von dem Besuch eines kommerziellen Trainingskurses abzusehen, lässt an den Daten zweifeln. Es ist möglich, dass die Studienplatzbewerber, die hoch motiviert waren, ihre Zulassungschancen nicht gefährden wollten und lieber sozial erwünscht antworteten. In diesem Fall entsprach eine sozial erwünschte Antwort den Vorgaben, die aus der Testbroschüre zu entnehmen waren: ein Trainingskurs muss nicht besucht werden und die Broschüre ist ausreichend als Vorbereitung für den Test. So sind die von Deter (1982) zum Testtraining gefundenen Befunde mit Zurückhaltung zu betrachten.

4.2 Studie von Ernst Fay (1985)

Ernst Fay hat in seiner Untersuchung „Vorbereitungsmöglichkeiten auf den «Test für medizinische Studiengänge»: Was gibt es? Wie wird es genutzt? Nutzt es?“ über die verschiedenen Vorbereitungsmaßnahmen für den TMS berichtet (Fay, 1985). Immer wieder hatte es gegen den Einsatz des TMS den Einwand der Trainierbarkeit gegeben. Schon vor dem ersten Testtermin hatte man versucht, durch den Versand der Testbroschüre an alle Studienbewerber diesem Argument entgegenzutreten. Deter (1982) hatte den Effekt der Testbroschüre im Vergleich zu einem kurzfristigen Training untersucht. Die Ergebnisse sind dargestellt in Kapitel 4.1. Fay (1985) verwies in seiner Untersuchung noch einmal darauf, dass die Leistung in einem Test durch die Testwiederholung gesteigert werden könne. Der Lerngewinn sei gerade von der ersten zur zweiten Testung besonders hoch. Mit dem Versand der Testbroschüre sollte erreicht werden, dass alle Kandidaten die Möglichkeit erhielten, ein quasi kostenloses und einheitlich gestaltetes Testtraining absolvieren zu können.

Fay (1985) erwähnte eine weitere Untersuchung zum Thema Trainingsanfälligkeit des TMS von Bartussek (Bartussek et al., 1984 zit. n. Fay, 1985). Einige Untertests waren besonders sensibel gegenüber Testtrainings. Die Studie von Bartussek (1984 zit. n. Fay, 1985), die von der Konferenz der Kultusminister der Länder als Evaluationsstudie in Auftrag gegeben worden war, hatte zur Folge, dass die Testbroschüre nochmals neu gestaltet und für die Testvorbereitung verbessert wurde. Zusätzlich wurde noch weiteres Material vom ITB zur Verfügung gestellt. Die Testanwärter konnten im Buchhandel zu einem moderaten Preis eine Originalversion des Testes erwerben. Aber natürlich gab es noch weitere Angebote von Test-Trainings-Instituten in Form von Trainingskursen und nicht autorisierte Trainingsbüchern, die man kaufen konnte. Fays (1985) beabsichtigte mit seiner Studie, eine Übersicht über die Vorbereitungsmaterialien zu geben, welche Form der Vorbereitung die Studienbewerber wählten und wie diese im Zusammenhang mit ihrer Testleistung standen.

Zur Vorbereitung auf den TMS konnten die Testteilnehmer zwischen vier Möglichkeiten wählen: Testbroschüre, veröffentlichte Originalversion des Tests, kommerzielle Übungsbücher und einem Trainingskurs. Sie hatten natürlich immer auch die Option, die Vorbereitungsmaßnahmen miteinander zu kombinieren. Über die

Testbroschüre wurde schon eingehend berichtet. Eindeutig ist auch, was mit der Originalversion des TMS gemeint ist. Auch zu den Übungsbüchern machte Fay (1985) weitere Ausführungen: Er merkte dabei nicht nur an, dass die Preise häufig überhöht waren, sondern fand auch heraus, dass in mehreren Untersuchungen auch regelrechte Fehler in den Übungsbüchern enthalten waren. Über das Angebot der Trainingsinstitute informierten sich Mitarbeiter des ITB über Anzeigen in der Presse. Teilweise hatten sie die Trainingskurse selbst besucht.

Interessant war der zweite Teil der Zielsetzung von Fays (1985) hier vorgestellten Untersuchung: die Nutzung der verschiedenen Vorbereitungsmöglichkeiten. Die verschiedenen Optionen wurden über einen Zeitraum von August '80 bis August '84 dargestellt. Deutlich war zu erkennen, dass die Anwärter zunehmend auch andere Maßnahmen als nur die Testbroschüre in Anspruch nahmen, wie z. B. des Trainingsbuchs und verstärkt auch Trainingskurse. Einschränkend erwähnte Fay (1985), dass der Fragebogen zur Begleituntersuchung erst zehn Tage vor dem Testtermin bei den Studienbewerbern einging, was eventuell zur Folge hatte, dass viele nicht die Vorbereitungsmaßnahme ergriffen, die sie bei mehr Vorlauf gerne genutzt hätten.

Das dritte Ziel von Fays (1985) Studie war die Untersuchung, ob die unterschiedlichen Arten der Vorbereitung einen Effekt auf die Testleistung hatten. Fay (1985) betonte in diesem Teil der Untersuchung noch einmal, dass man bei einer vorhandenen Korrelation von zwei Variablen nicht automatisch auf eine Kausalität schließen dürfe. Es gebe immer auch alternative Interpretationen für eine solche Korrelation. Fay (1985) nannte hier die drei möglichen Erklärungsansätze für eine Korrelation: Kausalität, Zufall oder der Einfluss einer möglichen Drittvariable, die auf beide Variablen wirkt. Bevor Fay (1985) auf die durchschnittlichen Testwerte einging, erklärte er, warum die Untertests in trainierbare und nicht-trainierbare Untertests unterschieden werden können. Bezüglich seiner möglichen Erklärungsansätze für Korrelationen legte er dar, dass man zwischen den Erklärungsversuchen unterscheiden müsse. Liegt nämlich eine Korrelation vor, die aufgrund einer vorher gemachten Hypothese, die in eine Theorie eingebettet ist und die in einer experimentellen Studie bestätigt wird und die dann in einer weiteren Studie wiederum bekräftigt wird, kann man davon ausgehen, dass dieser Zusammenhang wahrscheinlich nicht mehr zufällig

ist. Genau dieser Fall ist durch die beiden Studien von Deter (1982) und von Bartussek (1984 zit. n. Fay, 1985) eingetreten. Die Befunde zu den trainierbaren Untertests bei Deter (1982) wurden in Bartussek Untersuchung (1984 zit. n. Fay, 1985) bestätigt. Beide Studien waren konsistent, was die Trainierbarkeit bestimmter Untertests betrifft. So macht Fay (1985) eine Unterscheidung in trainierbare und nicht kurzfristig trainierbare Untertests des TMS. Die Aufteilung in die beiden Kategorien ist in der Tabelle 4.5 dargestellt.

Tabelle 4.5: Unterscheidung der Untertests nach Fay (1985) in trainierbare und nicht-trainierbare Untertests

trainierbare Untertests	nicht kurzfristig trainierbare Untertests
Schlauchfiguren	Figuren zusammensetzen
konzentriertes und sorgfältiges Arbeiten	medizinisch-naturwissenschaftliches Grundverständnis
Schattenrisse	Beurteilung formalisierter Information
Satzergänzung	Mathematisches Grundverständnis
Bilder ordnen	Textverständnis
Figuren lernen	
Fakten lernen	
Muster zuordnen	

Ein Trainingseffekt im Gesamtest dürfte sich nur folgendermaßen abzeichnen: Der trainierte Teilnehmer schneidet nur in den trainierbaren Untertests besser ab. Ein leistungsstarker und vorbereiteter Teilnehmer schneidet sowohl in den trainierbaren als auch in den nicht kurzfristig trainierbaren Untertests besser ab.

Beim Vergleich der durchschnittlichen Gesamtestwerte schnitt die Gruppe, die einen Trainingskurs besucht hatte, geringfügig besser ab. In dieser Studie um genau 2.43 Rohpunkte. Da dies genau der Rohpunktunterschied im Untertest „konzentriertes Arbeiten“ ist, den diese Gruppe zu der Gruppe hat, die die Originalversion bearbeitet hat, behauptete Fay (1985), dass die Teilnehmer einen Trainingskurs belegt hatten, um sich einzig allein in diesem von insgesamt dreizehn Untertests zu verbessern.

Ein wichtiger Aspekt, den Fay (1985) in seiner Untersuchung eingehender beleuchtete, sind die Moderatoreffekte. Selbstverständlich hat nicht nur die Art der Vorbereitung

einen Einfluss auf das Testergebnis, sondern es gibt noch eine Reihe anderer Faktoren: Fay (1985) untersuchte hier zusätzlich den Einfluss der Abiturnote, die Art der Hochschulzugangsberechtigung, die in der Schule gewählte Interessensschwerpunkte, das Geschlecht und die Vorbildung bzw. die berufliche Position der Eltern. Fay (1985) hatte in einer seiner früheren Arbeiten herausgefunden, dass es eine Gruppe von Faktoren gibt, die mit der Testleistung in Verbindung stehen (Fay, 1984; Maichle, 1984 zit. n. Fay, 1985). Er überprüfte, ob es einen Zusammenhang zwischen der Gruppenzugehörigkeit bezogen auf die Art und den Umfang der Vorbereitung und der Ausprägung der oben erwähnten Moderatorvariablen gab und ob sich dieser Zusammenhang in dem Testergebnis niederschlägt. Ein erwarteter Zusammenhang zwischen einer schwächeren Abiturnote und höherem Aufwand an Vorbereitungsmaßnahmen bestätigte sich bei dieser Analyse nicht. Insgesamt stellte Fay (1985) fest, dass die erwähnten Moderatorvariablen bei allen Gruppen, die man hinsichtlich ihrer Vorbereitungsmaßnahmen unterscheiden konnte, gleich verteilt waren.

Fay (1985) ging am Ende seiner Analyse noch einmal auf den für die Bewerbung sehr wichtigen Aspekt der Zulassungschance ein. Die Frage hierbei war, ob sich ein positiver Zusammenhang zwischen der Art der Vorbereitung und der Zulassung zum Medizinstudium erkennen lässt. Hierbei war natürlich der Faktor Trainingskurs besonders wichtig. Wären die Trainingskursteilnehmer prozentual bei der Zulassung zum Studium stärker vertreten gewesen, dann wäre der Aspekt der Chancengleichheit nicht gegeben. Im untersuchten Sommersemester 1984 wurden aber von den Studienplätzen nur 25 Prozent (1200 Plätze) unter Berücksichtigung des Testergebnisses vergeben, davon war nur für 400 das Testergebnis entscheidend. Alle anderen wurden über verschiedene Quoten oder notengesteuert zugelassen. Die Art der Vorbereitung war also nicht allein entscheidend für die Zulassung. Fay (1985) erwähnte in diesem Zusammenhang die schon vorher angesprochene Drittvariable: Das Verhalten, das sich auf die Notengebung der Lehrer, das Vorbereitungsverhalten vor dem Test und auf das Testverhalten auswirkt. Dieses Verhalten, das schwierig zu messen ist, sich jedoch in allen Variablen niederschlägt, scheint einen großen Einfluss auf die mögliche Zulassung zum Medizinstudium zu haben.

4.2.1 Anmerkungen zu Fay (1985)

Hinsichtlich des zuletzt angesprochenen Punktes eines möglichen Zusammenhangs zwischen Testvorbereitung und Zulassungschancen für das Medizinstudium ist zu erwähnen, dass dieser Punkt nicht den Aspekt der Trainierbarkeit des TMS betrifft. Es geht hier allein um die Zulassungschancen für die Bewerber. Die Zulassung zum Medizinstudium setzt sich aus verschiedenen Quoten zusammen wie in Kapitel 2 zum TMS beschrieben wird.

In dem Abschnitt, in dem Fay (1985) die Effekte der verschiedenen Arten der Vorbereitung hinsichtlich ihres durchschnittlichen Gesamtergebnisses vergleicht, bemerkte er, dass die Trainingskursteilnehmer gerade mal 2.43 Rohpunkte besser abschnitten als die Bewerber, die die Originalversion bearbeitet hatten. Dieser Rohpunktwert entspricht genau dem Wert, den die Trainingsteilnehmer im Untertest „konzentriertes Arbeiten“ besser abschneiden als die Originalversionsgruppe. Fay machte daher die Aussage, dass die Teilnehmer nur für diesen einen Untertest zahlen würden. Allerdings, wenn man alle Untertests betrachtet, wird deutlich, dass die Trainingsteilnehmer in sieben von dreizehn Untertests besser abschlossen und diese übereinstimmende Rohpunktzahl wohl eher auf Zufall beruht.

4.3 Studie Kirchenkamp und Mispelkamp (1988)

Kirchenkamp und Mispelkamp haben in ihrer Untersuchung „Beziehungen zwischen Leistungen im Test für medizinische Studiengänge und verschiedene Vorbereitungsmaßnahmen“ noch einmal betont, wie wichtig die Studien über die unterschiedlichen Arten der Vorbereitung auf den TMS sind, da gerade das Problem der möglichen Trainierbarkeit des Tests immer wieder ein Argument gegen die Verwendung des TMS war (Kirchenkamp & Mispelkamp, 1988). Die Kritiker des Tests behaupteten, dass dieser, wenn er trainierbar sei, nicht fair wäre. Aber auch wenn zwischen der Vorbereitung und dem Testergebnis ein enger Zusammenhang bestünde, so kann dieser – wie Fay (1985) betont hatte – nicht kausal ausgelegt werden, da man immer mögliche Moderatorvariablen, die ebenfalls einen Einfluss auf Testleistung und Vorbereitung haben, nicht ausschließen könne.

Aufgrund der Bedeutung der Frage eines möglichen Einflusses der Vorbereitungsart auf das Testresultat führten Kirchenkamp und Mispelkamp im Jahr 1986 erneut eine schriftliche Befragung durch. In dieser Befragung wurden die Studienplatzbewerber detaillierter als in den Studien zuvor zur Form und zum Umfang ihrer Vorbereitung für den Test befragt. Die Studenten gaben bei der Befragung ihre jeweilige TMS Nummer an. Bei der Auswertung wurden die Befragten in verschiedene Gruppen je nach Vorbereitungsart eingeteilt. Es entstanden je nach Analyse sieben bis neun disjunkte Vorbereitungsgruppen. Die Befragten konnten zwischen folgenden Vorbereitungs-möglichkeiten auswählen. Mehrfachantworten waren zugelassen.

- keine Vorbereitung
- Test-Info
- Originalversion(en)
- Trainingsbuch(-bücher)
- Trainingskurs(e)
- Schon einmal an einem TMS teilgenommen
- Sonstiges

Die sich ausschließende Gruppen verglichen Kirchenkamp und Mispelkamp (1988) hinsichtlich ihrer durchschnittlichen Testergebnisse untereinander. Es zeichnete sich ein eindeutiges Bild zu Gunsten der intensiv vorbereiteten Gruppen ab: die Testteilnehmer, die alle möglichen Vorbereitungen genutzt hatten: Testbroschüre, verschiedene Originalversionen des TMS, Trainingsbuch und Trainingskurs. Am schlechtesten schnitt die Gruppe ab, die sich gar nicht vorbereitet hatte. Dies entsprach den Ergebnissen, die in der Studie zu den Vorbereitungs-möglichkeiten dokumentiert worden waren (Fay, 1985). Nach der Bereinigung der Mittelwerte um die Varianzanteile, die auf unterschiedliche schulische Leistungen zurückzuführen sind (Blum & Hensgen, 1987 zit. n. Kirchenkamp & Mispelkamp, 1988), waren die Abstufungen zwischen den Gruppen nicht mehr ganz so gravierend, aber die Reihenfolge blieb gleich. Da die Gruppe, die sich nicht vorbereitet hatte, auch die Gruppe mit dem schlechtesten Abiturdurchschnitt war, hatte diese Bereinigung in der Gruppe einen positiven Effekt. Die Punktzahl stieg. Immer noch schnitten die optimal Vorbereiteten am besten ab und

die Kandidaten ohne jegliches Übungsverhalten am schlechtesten. Die schulischen Leistungen waren operationalisiert durch die Abiturdurchschnittsnoten. Diese Bereinigung der Testresultate um die Schulnoten zeigt eindeutig, so Kirchenkamp und Mispelkamp (1988), dass man mit der einfachen Schlussfolgerung eines kausalen Zusammenhangs zwischen Art der Vorbereitung und Testergebnis vorsichtig sein muss, da andere Variablen immer auch einen Einfluss auf das Testergebnis hätten und diese mit einbezogen werden müssten. Wenn die Abiturnote einen Einfluss auf das Resultat hatte, dann könnten andere mögliche Faktoren ebenfalls einen Einfluss auf die Leistung haben. Sie sind aber in dieser Studie nicht aufgeführt worden. Kirchenkamp und Mispelkamp (1988) wiesen nur noch einmal explizit darauf hin, dass die Gesamtvarianz der Testleistung durch mehrere Faktoren aufgeklärt wird.

Kirchenkamp und Mispelkamp (1988) haben in ihrer Untersuchung die Untertestergebnisse der jeweiligen Vorbereitungsgruppen verglichen. Angelehnt an Fay (1985) unterschieden sie dabei in trainierbare und nicht-trainierbare Untertests. Sie erwarteten, dass die Differenzen zwischen der Gruppe der gut vorbereiteten Teilnehmer und der Gruppe derjenigen, die ohne Vorbereitung am Test teilnahmen, in den trainierbaren Untertests, z. B. „Fakten lernen“, größer sein müsste als in einem nicht-trainierbaren Untertest, z. B. „Medizinisch-naturwissenschaftliches Grundverständnis“. In der Analyse der Daten stellten Kirchenkamp und Mispelkamp (1988) allerdings fest, dass sich die beiden zwar in der Differenz ihrer Mittelwerte in den Untertests unterschieden, aber nicht in der zu erwartenden Unterteilung. Die Differenz in dem nicht-trainierbaren Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“ war größer als die Differenz in dem trainierbaren Untertest „Fakten lernen“. Kirchenkamp und Mispelkamp (1988) schlossen daraus, dass das stärkere Abschließen der Gruppe derjenigen, die sich sehr intensiv vorbereitet hatten – wie schon in der Studie von Fay (1985) erwähnt wurde - nicht allein mit deren besseren Vorbereitungsmaßnahmen erklären ließe, sondern dass andere Faktoren, wie z. B. Leistungsmotivation und –stärke hierbei auch eine Rolle spielten. Die gut Vorbereiteten waren von Anfang an besser, so dass sie auch in den nicht-trainierbaren Untertests allein aufgrund ihres höheren Wissenstandes besser abschlossen.

Kirchenkamp und Mispelkamp (1988) konnten feststellen, dass die Teilnehmer, deren Testergebnisse über dem Durchschnitt lagen, sich eingehender mit dem Test vor der Testabnahme beschäftigt hatten. Das Minimum an Vorbereitung, um überdurchschnittlich abzuschließen, lag hier bei der Bearbeitung des Test-Infos und der Originalversionen aus den Jahren '84 und '86. Die maximale durchschnittliche Gesamtpunktedifferenz betrug hier 2.5 Punkte.

Eine weitere Frage in dem Fragebogen zur Begleituntersuchung hinsichtlich der verschiedenen Vorbereitungsmaßnahmen von Kirchenkamp und Mispelkamp (1988) bezog sich auf die für die Vorbereitung aufgewandete Zeit. Es war ein leichter Anstieg der durchschnittlichen Testleistung mit gleichzeitigem Anstieg der Länge der aufgebrauchten Zeit zu erkennen. Die Abstände zwischen den aufeinanderfolgenden Mittelwerten wurden aber immer kleiner. Der Unterschied in der Testleistung war deutlicher zu erkennen, wenn man die Gruppe, die nur eins bis vier Stunden Vorbereitungszeit investiert hatte, mit der Gruppe verglich, die fünf bis neun Stunden aufgewendet hatte. Zwischen 20 bis 39 Stunden und 40 Stunden und länger war die Differenz der durchschnittlichen Mittelwerte nicht mehr so hoch. Kirchenkamp und Mispelkamp (1988) wiesen aber ausdrücklich daraufhin, dass dieser Zeitfaktor bei der Vorbereitung nicht überinterpretiert werden sollte. Schließlich habe jeder Bewerber sein eigenes Tempo beim Lernen. Die interindividuellen Unterschiede seien jedoch so schwer zu erfassen, dass es schwierig ist, die Ergebnisse eindeutig zu beurteilen. Ein weiterer entscheidender Punkt bei der Vorbereitung war neben dem Zeitfaktor auch die Qualität des Materials. Wenn man sich 40 Stunden und länger mit dem falschen Material vorbereite, werde auch das Ergebnis nicht besser.

4.3.1 Anmerkungen zu Kirchenkamp und Mispelkamp (1988)

Kirchenkamp und Mispelkamp (1988) haben in ihrer Analyse hauptsächlich die durchschnittlichen Testergebnisse zweier Gruppen miteinander verglichen: die Gruppe, die sich gar nicht vorbereitet hat, mit der Gruppe, die alle Arten der Vorbereitung für sich genutzt hatte wie Test-Info, Originalversionen, Trainingsbücher und Trainingskurs. Deutlich wurde bei dem Vergleich, dass die Bewerber, die nichts getan hatten, signifikant schlechter abschnitten als die andere Gruppe, nicht nur im Gesamtest sondern auch in allen Untertests. Eine Schwierigkeit dieses Verfahrens war sicherlich,

dass die beiden Autoren die Mittelwerte eines Untertests zweier verschiedener Gruppen miteinander verglichen, die nicht, wie in einem experimentellen Design, zufällig ihrem Treatment zugeteilt wurden. Man kann also nicht davon ausgehen, dass das Ausgangsniveau in den Gruppen gleich hoch war. Es ist nun schwierig zu erkennen, ob eine Differenz zwischen den Mittelwerten dieser Gruppen auf ein eventuelles Treatment zurückzuführen ist oder ob sich die beiden Gruppen nicht schon in ihrem Ausgangsniveau voneinander unterschieden. Kirchenkamp und Mispelkamp (1988) gingen davon aus, dass sich die Gruppen von Anfang an unterschieden. Die Gruppe, die sich gut vorbereitet hatte, war erkennbar an den besseren Abiturnoten und auch insgesamt leistungsstärker. Da man aber über das Niveau der beiden Gruppen nichts weiß, kann man nicht einfach davon ausgehen, dass die Gruppe derjenigen, die sich nicht vorbereitet hatte, auch automatisch die schlechteren Ausgangswerte hatte. In der

Tabelle 4.6 erkennt man, dass Kirchenkamp und Mispelkamp (1988) für das Aufzeigen des geringen Trainingseffektes einen relativ niedrigen Wert für die trainierbaren Untertests ausgewählt haben.

Betrachtet man alle Differenzen der trainierbaren Untertests zwischen der nicht vorbereiteten und der optimal vorbereiteten Gruppe und bildet aus ihnen eine durchschnittliche Differenz, dann liegt diese bei 2.73. Kirchenkamp und Mispelkamp (1988) haben für ihre Darstellung, dass es sich nicht um einen Trainingseffekt, sondern wahrscheinlich eher um einen schon vorher bestehenden Unterschied in der Leistungsstärke handele, den Wert aus dem Untertest „Fakten lernen“ ausgewählt. Dieser Wert betrug 1.68. Er liegt unter der durchschnittlichen Differenz. Während im Gegensatz dazu die zum Vergleich gewählte Differenz im Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“, nämlich 1.73, höher ist als der Durchschnittswert aller Differenzen der nicht-trainierbaren Untertests, der bei 1.51 liegt. Sie hätten auch andere Differenzen hinzuziehen können, dann hätte sich ein anderes Bild geboten.

Tabelle 4.6: Übersicht über die Differenzen zwischen den beiden Extremgruppen untergliedert in trainierbare und nicht-trainierbare Untertests

Trainierbarkeit	Untertest	nicht vorbereitet	optimal vorbereitet	Differenz
kurzfristig trainierbar	Muster zuordnen	11.69	14.05	2.36
	Schlauchfiguren	10.57	13.13	2.56
	Konzentriertes Arbeiten	12.86	16.97	4.11
	Figuren lernen	9.48	12.44	2.96
	Fakten lernen	10.88	12.56	1.68
nicht kurzfristig trainierbar	Mediz.-naturw. Grundverständnis	11.02	12.75	1.73
	Quantit.-formale Probleme	9.10	10.58	1.48
	Textverständnis	9.18	11.06	1.88
	Diagramme und Tabellen	10.22	11.17	0.95

Aus der Analyse von Kirchenkamp und Mispelkamp (1988) ist ersichtlich, dass sich Vorbereitung auf den Test lohnt. Alle Bewerber, die sich vorher mit dem Test übungsmäßig auseinandergesetzt hatten, schlossen mit einem höheren durchschnittlichen Testergebnis ab, als diejenigen, die unvorbereitet in den Test gingen. Dieses Ergebnis blieb auch bestehen, wenn die durchschnittliche Testleistung durch die Abiturnote bereinigt wurde. Wie Kirchenkamp und Mispelkamp (1988) mehrmals betonten, kann man nicht kausal schließen, dass eine intensive Vorbereitung, also Test-Info, Originalversionen, Trainingsbuch und Trainingskurs, impliziert, dass man gut im Test abschneidet. Es spielen immer noch andere Faktoren beim Zustandekommen des Testergebnisses eine große Rolle. Der Faktor Vorbereitung hat neben anderen Faktoren einen Einfluss auf die Testleistung. Auch wenn ein kausaler Zusammenhang nicht

überprüfbar ist. Der Faktor Vorbereitung darf aber deshalb nicht unbeachtet bleiben. Faktoren - wie Motivation und Leistungsstärke - haben ebenfalls einen Einfluss. Beim Zustandekommen einer Testleistung spielen viele Faktoren eine Rolle.

4.4 Studie von R. Hofer & K.-D. Hänsgen (1996)

Hänsgen und Hofer sind in ihrem Beitrag „Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium in der Schweiz“ noch einmal auf die verschiedenen Untersuchungen bezüglich der Trainierbarkeit der Testleistung beim Test für das Medizinstudium eingegangen (Hofer, 1996). Das Thema war in der Schweiz aktuell geworden, da dort ab dem Jahre 1998 ein Eignungstest für das Medizinstudium (EMS) bindend für alle Bewerber für die Universitäten Basel, Bern, Feiburg und Zürich eingeführt worden war. Alle von Hänsgen und Hofer (1996) in ihrem Artikel erwähnten Studien sind in diesem Kapitel eingehend vorgestellt worden. Einzig allein eine Information aus einer Studie von Mispelkamp (1987 zit. n. Hofer, 1996) klärt einmal, warum so viele Bewerber sich mit Hilfe eines Trainingskurs auf den Test vorbereiten „... die Mehrzahl der Kursbesucher und –besucherinnen sich nicht in der Lage sehen, eigenständig und ohne Anleitung konzentriert über mehrere Stunden das Trainingsmaterial zu bearbeiten“ (Hofer, 1996).

Auch Hänsgen und Hofer (1996) betonten in ihrem Artikel, dass eine Vorabbeschäftigung mit dem Test Vorteile bringe. So kam es auch in der Schweiz zu der Veröffentlichung verschiedener Originalversionen in den gängigen Landessprachen und zur Verschickung der kostenlosen Testbroschüre an alle Studienplatzbewerber.

4.5 Studie von Günter Trost et al. (1998)

Trost ist in seinem Beitrag „Inwieweit ist der Test trainierbar?“ noch einmal auf mehrere Studien und Analysen eingegangen, die bezüglich der Trainierbarkeit des TMS in den Zeitraum 1979 bis 1998 durchgeführt wurden (Trost, 1998). Er betonte noch einmal, warum der Aspekt der Trainierbarkeit und die darum entstandenen Studien so wichtig waren. Zunächst einmal sei der TMS ein Test gewesen, der an fast der gesamten Population der Medizinstudienbewerber angewandt worden war, ausgenommen die Bewerber, die über besondere Quoten wie Wartezeiten etc. zum Medizinstudium

zugelassen worden waren. Da dieser Test zu Selektionszwecken durchgeführt wurde, war es daher wichtig, dass der TMS eine breite Akzeptanz in der Öffentlichkeit fand. Wäre dieser Test trainierbar, so die Kritiker, hätten einige Bewerber höchstwahrscheinlich einen Vorteil. Ein weiterer Grund für die Bedeutung der Studien zur Trainierbarkeit des TMS war, dass mit dem Test die tatsächlich vorhandenen Fähigkeiten und nicht kurzfristig trainierte Fähigkeiten gemessen werden sollten. Der Test sollte Fähigkeit und nicht Fleiß messen.

Zur Beantwortung der Frage der möglichen Trainierbarkeit gab es nach Trost (1998) zwei Ansätze. Der eine Ansatz versuche die Frage der Trainierbarkeit über eine Fragebogenaktion deskriptiv zu beschreiben. Analysen von Fay (1985) und Kirchenkamp und Mispelkamp (1988) hatten bereits einen Überblick darüber gegeben, welche Arten von Vorbereitungsmöglichkeiten es gab, wie sie genutzt wurden und wie sich die vorbereiteten Getesteten bezüglich ihrer Testleistungen unterschieden. Der zweite Ansatz kläre die Frage der Trainierbarkeit mit Hilfe einer experimentellen Untersuchung. Trost (1998) gab in seinem Bericht einen Überblick über die experimentellen Studien zur Trainierbarkeit des TMS sowohl bezogen auf den Gesamttest als auch auf die einzelnen Untertests. Hierbei ging er jeweils auf die Wirksamkeit der bestimmten Vorbereitungsmaßnahmen ein.

Zunächst beschäftigte sich Trost (1998) mit der oben beschriebenen Studie von Kirchenkamp und Mispelkamp (1988). Aus diesen Ergebnissen und den Resultaten der Untersuchung von Fay (1985) kam Trost (1998) zu dem Schluss, dass die Ergebnisse dieser Studien belegen, dass die Hinweise, die das ITB zur Vorbereitung auf den TMS gibt, ausreichend seien, um optimal auf den Test vorbereitet zu sein.

Für den zweiten Ansatz stellte Trost (1998) die beiden Studien vor, die sich mit dem Zusammenhang zwischen Training und Gesamtestleistung auseinandersetzten: Deter (1982) und Bartussek et al. (1984 zit. n. Trost, 1998). Deter (1982) kam, wie oben schon erwähnt, aufgrund seiner Ergebnisse der experimentellen Studie zu dem Schluss, dass erstens eine Vorbereitung mit der Testbroschüre die gleiche Wirksamkeit habe wie ein kurzfristiges Training und dass zweitens einige Untertests trainierbar seien. Die zweite experimentelle Studie von Bartussek et al. (1984 zit. n. Trost, 1998) unterschied sich darin, dass sich in diesem Fall alle Versuchspersonen mit der Testbroschüre

vorbereiteten, aber nur die Experimentalgruppe zusätzlich ein 50stündiges Training erhielt. Alle Teilnehmer verbesserten sich: Die Kontrollgruppe um ca. 2 Prozent der maximal erreichbaren Punktzahl und die Experimentalgruppe insgesamt um ca. 6 Prozent der maximal erreichbaren Punktzahl. Hierbei ist zu beachten, dass beide Gruppen die Testbroschüre bearbeitet hatten. Trost (1998) schloss aus den Ergebnissen beider Studien, dass man durch ein zusätzliches Training neben der Bearbeitung der Testbroschüre nur noch mit einem unerheblichen Leistungsanstieg rechnen könne.

Trost (1998) ging in seinem Beitrag zur Trainierbarkeit des TMS noch auf zwei weitere Studien ein, die sich mit Trainingsmaßnahmen für zwei Untertests befassten. Beide Untertest galten als trainierbare Untertests: „Konzentriertes und sorgfältiges Arbeiten“ und „Fakten lernen“ (Fay, 1985).

Die Studie bezüglich der Trainierbarkeit des Untertests „Konzentriertes und sorgfältiges Arbeiten“ wurde von Fay et al. durchgeführt (1989, zit. n. Trost, 1998). Der KSA³ ist der Untertest, der im TMS am besten trainierbar ist. Trost (1998) merkte an, dass diese leichte Übbarkeit des KSA die Nützlichkeit dieses Untertests nicht einschränkt, wenn vier Voraussetzungen als erfüllt betrachtet werden könnten. Erstens erhielten alle Bewerber den Hinweis, dass dieser Test sehr gut trainierbar sei. Zweitens wurde allen Kandidaten die Möglichkeit gegeben, den KSA zu üben. Dies geschah durch mehrere Übungsexemplare, die den Bewerbern zusammen mit dem Test-Info zugesandt wurden. Drittens bestanden die Unterschiede zwischen den Teilnehmern auch nach dem Training immer noch - nur auf einem höheren Niveau. Die Rangfolge hatte sich aber nicht verändert. Und viertens müssten die Leistungsunterschiede auch in dem Testwert sichtbar werden. Die ersten drei Punkte sah Trost (1998), gerade durch die experimentelle Studie von Deter (1982) und ebenso durch die Studie von Fay et al. (1989, zit. n. Trost, 1998), als verwirklicht an. Nur die vierte Voraussetzung wurde nicht genügend erfüllt. Beim KSA konnte es passieren, dass leistungsstarke Testteilnehmer, die sich zusätzlich noch gut auf den Konzentrationstest vorbereitet hatten, den obersten Wert der Skala erreichten, obwohl sie eigentlich noch besser abschneiden würden, wenn es in dieser Bewertung sichtbar würde. Der KSA ermögliche

³ „Konzentriertes und sorgfältiges Arbeiten“ wird im Folgenden KSA abgekürzt.

es also nicht, die Verschiedenheit der in diesem Test gemessenen Fähigkeiten gerade im oberen Leistungsbereich zum Ausdruck zu bringen. In ihrer Untersuchung zur Trainierbarkeit des KSA verglichen Fay et al. (1989, zit. n. Trost, 1998) eine Experimental- und eine Kontrollgruppe miteinander. Beide Gruppen wiederholten den KSA insgesamt sechs Mal. Die Experimentalgruppe erhielt aber zusätzlich zu der Wiederholung außer bei der ersten und letzten Wiederholung noch zusätzliche Informationen, Tips und die Bearbeitungsbedingungen wurden verändert. Fay et al. (1989, zit. n. Trost, 1998) stellten fest, dass alle Teilnehmer sich verbesserten. Der Lerngewinn der Kontrollgruppe war nicht sehr viel geringer als in der Experimentalgruppe. Die Rangfolge, die die Versuchspersonen nach der ersten Abnahme des KSA aufgezeigt hatten, blieb nahezu unverändert. Eine wichtige Erkenntnis aus dieser Studie hinsichtlich des KSA war, dass es nach dem fünften Durchgang zu keinen erwähnenswerten Verbesserungen in den individuellen Werten kam. Die Testteilnehmer hatten ihr persönliches Leistungsmaximum erreicht. Ein weiteres Üben würde ihren Testwert nicht mehr verbessern.

Eine weitere Studie zur Trainierbarkeit eines Untertests stellte Trost (1998) in seinem Gesamtbericht über die Trainierbarkeit des TMS vor. Die Studie von Klieme und Espey (1992 zit. n. Trost, 1998) beschäftigte sich mit der Trainierbarkeit des Untertests „Fakten lernen“. Auch Klieme et al. (1992, zit. n. Trost, 1998) führten eine Untersuchung mit einer Kontrollgruppe durch. Sowohl eine Kontroll- als auch eine Experimentalgruppe erhielten ein Training. Das Training der Experimentalgruppe war aber bei Weitem intensiver als das Basistraining der Kontrollgruppe, das den Empfehlungen aus dem Test-Info entsprach. Beide Gruppen zeigten hinsichtlich des Vergleichs von Vor- und Nachtestwerten hoch signifikante Werte. Vergleich man die Gruppen allerdings untereinander, unterschieden sie sich nicht bedeutend voneinander. Trost (1998) stellte die weiteren Ergebnisse aus dieser Studie vor, die sich mit der Einschätzung der eigenen Leistung und der Einschätzung der Wirksamkeit und Anwendbarkeit des Trainings beschäftigte. In dieser Arbeit wird auf diese Ergebnisse nicht näher eingegangen. Trost (1998) schloss aus den Ergebnissen von Klieme et al. (1992, zit. n. Trost, 1998), dass die Empfehlungen, die in dem Test-Info gegeben werden, für die optimale Vorbereitung völlig ausreichten.

Als letztes berichtete Trost (1998) über eine Studie, die von Klieme und Maichle (1990 zit. n. Trost, 1998), hinsichtlich der Trainierbarkeit bestimmter medizinrelevanter Fähigkeiten durchgeführt wurde. Die Fragestellung der Untersuchung war, ob einige Fertigkeiten, die im Test begutachtet wurden, durch ein spezielles Training verbessert werden könnten. Wäre das der Fall, so müsste das Test-Info um entsprechende Empfehlungen für die Vorbereitungen der Untertests erweitert werden. Die Untersuchung bezog sich auf die Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“ (MNGV)⁴ und „Quantitative und formale Probleme“ (QFP)⁵. Im Herbst 1988 wurden Medizinbewerber, nachdem sie den TMS absolviert hatten, per Zufall zwei Trainingsgruppen zugewiesen: die erste Gruppe absolvierte ein Training zum schlussfolgernden Denken, die zweite ein Programm zum Problemlösen. Die erste Gruppe zeigte keine statisch signifikanten Veränderungen zwischen Vor- und Nachtest. Allerdings gab es positive Lerngewinne. Die zweite Gruppe zeigte ebenfalls keinen Effekt in dem Vergleich der Vor- und Nachtestwerte hinsichtlich des trainierten Untertests QFP. Es konnte aber ein schwacher Effekt erkannt werden, wenn man die durchschnittliche Leistung von vier Untertests mit einbezog: MNGV, QFP, „Diagramme und Tabellen“ und „Schlauchfiguren“. Es kam also zu einer schwachen positiven Leistungssteigerung. Trost (1998) gab aber zu Bedenken, dass sich erstens Teilnehmer mit niedrigen Vortestwerten verbesserten und sich diejenigen mit hohen Vortestwerten sich verschlechterten. Dieser Umstand ist ein eindeutiges Zeichen für die Regression zur Mitte. Zweitens hatte die Verbesserung die durch den Vortest entstandene Rangfolge nicht verändert. Die guten Bewerber blieben weiterhin oben in der Rangliste, die schlechteren unten, aber mit insgesamt besseren Ergebnissen.

Trost (1998) schloss daraus, dass die Zusammenfassung der Berichte zur Trainierbarkeit des TMS den bisherigen Kenntnisstand bestätigte und folglich eine Überarbeitung des Vorbereitungsmaterials nicht notwendig sei.

⁴ Im folgenden MNGV genannt

⁵ Im folgenden QFP genannt

4.5.1 Anmerkungen zu Trost (1998)

Trost (1998) hat in seiner Zusammenfassung zu der Trainierbarkeit des TMS verschiedene Untersuchungen und ihre Ergebnisse vorgestellt. In allen Studien konnte nachgewiesen werden, dass eine zusätzliche Vorbereitung eine Leistungssteigerung mit sich brachte. Diese Leistungssteigerung, die durch ein Training hervorgerufen werden kann, kann aber durch die kostenlose Test-Info und die Bearbeitung der veröffentlichten Originalversionen ausgeglichen werden. So kann also festgestellt werden, dass der TMS hinsichtlich der allen Teilnehmern zur Verfügung gestellten Vorbereitungs-möglichkeiten fair ist. Die Teilnehmer, die sich noch zu dem Besuch eines extra Trainingskurs oder den Kauf eines nicht autorisierten Trainingsbuchs entschließen, werden keinen Vorteil gegenüber den anderen vorbereiteten Testteilnehmern haben.

Folgt man Trost (1998), ist es jedoch nicht sinnvoll, den Test unvorbereitet zu absolvieren, da in allen Studien aufgezeigt wurde, dass ein Training bzw. eine Vorbereitung zu einer Leistungssteigerung führe. In vielen Studien konnte festgestellt werden, dass die Bewerber, die angaben sich in keinerlei Weise vorbereitet zu haben, sich auch hinsichtlich der durchschnittlichen Schulnote unterschieden. Häufig hatten diese Kandidaten weniger gute Abiturdurchschnittsnoten als diejenigen, die sich gründlich vorbereiteten. So könnte man implizieren, dass sie auch in ihrer bisherigen Schullaufbahn weniger fleißig waren. Der TMS misst also nicht nur Fähigkeiten und Fertigkeiten, die ein Student mitbringen muss, um ein Medizinstudium erfolgreich absolvieren zu können, er misst auch Fleiß. Und dass Fleiß ein Kriterium für den erfolgreichen Abschluss eines Studiums ist, kann nicht geleugnet werden.

5 Testtraining im angloamerikanischen Raum

Wenn große Teile einer Population oder sogar eine gesamte Population mit ein- und demselben standardisierten Instrument zu Selektionszwecken getestet werden, ist auch immer die Frage von Bedeutung, ob der Test trainierbar ist. Nicht nur in Deutschland werden Studierfähigkeitstests zur Auswahl geeigneter Studierender eingesetzt, sondern auch in vielen anderen Ländern, wie z. B. Schweden, Israel, Japan, Großbritannien und in den USA. So werden gerade in den USA seit vielen Jahren Studien durchgeführt, die die Größe des Trainingseffekts bestimmen sollen. Da die Literatur zu diesem Thema in den USA am vielfältigsten ist, wird in dieser Arbeit zum Thema Testtraining exemplarisch nur auf Studien aus dem angloamerikanischen Raum eingegangen. Dieses Kapitel dient dazu, einen Überblick darüber zu geben, dass das Thema Studierfähigkeitstest und mögliche Trainingseffekte nicht nur im deutschsprachigen Raum, sondern weltweit auf großes Interesse stößt. Zunächst wird in dem Kapitel beschrieben, wie sich die Auswahl der Studierenden in den USA vollzieht, dann werden die beiden wichtigsten allgemeinen Studierfähigkeitstests SAT und ACT kurz vorgestellt. Anschließend folgt eine Unterscheidung der drei wichtigsten Begriffe, die es im angloamerikanischen Sprachraum zum Thema Testtraining gibt: „test-practice“, „test-wisness“ und „test-coaching“. Am Schluss folgt eine Darstellung der Ergebnisse der wichtigsten Studien zum Thema Testtraining.

5.1 Auswahlverfahren in den USA

Schon im letzten High-School Jahr bereiten sich die zukünftigen Studenten Amerikas auf die Bewerbung an einer Hochschule vor, indem sie parallel zum letzten Schuljahr an einem Studierfähigkeitstest teilnehmen. Nach dem Abschluss der High-School bewerben sie sich an verschiedenen Hochschulen zur Aufnahme eines Undergraduate-Studiums. In den Vereinigten Staaten wählen die Hochschulen mit wenigen Ausnahmen ihre Studenten mit Hilfe des Ergebnisses eines allgemeinen Studierfähigkeitstests aus. Dies ist ein Ergebnis der Tatsache, dass die Qualität der Schulabschlüsse, das High-School Abschlusszeugnis, in den verschiedenen Staaten des Landes stark variiert. Zusätzlich kommen traditionell viele Bewerber aus dem Ausland

für ein Studium in die USA. Die Hochschulen können durch die Ergebnisse eines standardisierten Tests eine bessere Auswahl unter den Bewerbern treffen. Das Ergebnis des Studierfähigkeitstests ist also für den Studienbewerber sehr wichtig. Zusammen mit den anderen Bewerbungsunterlagen, dem High-School Abschlusszeugnis, Empfehlungsschreiben und manchmal auch einem persönlichen Essay, entscheidet das Testergebnis über die Zulassung an der Wunschuniversität. Jede Hochschule hat ihre eigene Zulassungsabteilung und legt selbstverwaltet ihre Zulassungskriterien fest. Zwischen den Universitäten herrscht großer Wettbewerb um die besten Studenten. Gleichzeitig sind die Bewerber stark unter Druck, an den renommierten Universitäten einen Studienplatz zu erhalten. Da der Testwert des allgemeinen Studierfähigkeitstests eine große Rolle bei der Zulassung an der Hochschule spielt, unternehmen die amerikanischen Studienbewerber eine Reihe von Aktivitäten, um einen möglichst hohen Testwert zu erreichen. So ist es zu erklären, dass das Thema Testtraining⁶ von besonderem Interesse ist.

5.2 SAT

Der Name SAT hat sich aus dem ehemaligen Scholastic Aptitude Test heraus gebildet. Heute spricht man aber nur noch vom SAT. Der SAT ist ein allgemeiner Studierfähigkeitstest. Seine Entstehung geht auf die 40er Jahre des letzten Jahrhunderts zurück. Der Test wird unterstützt vom College Entrance Examination Board (CEEB). Entwickelt wurde er vom Educational Testing Service (ETS). Das ETS ist auch für die Verwaltung des SAT zuständig (Briggs, 2002). Der SAT besteht aus zwei Teilen: dem SAT Reasoning Test, meistens auch als SAT I bezeichnet, und dem SAT Subject Test, auch als SAT II gekennzeichnet. Der SAT Reasoning Test ist ein allgemeiner Test (Heine, 2006). Er überprüft generelle kognitive Kompetenzen, die für das erfolgreiche Abschließen eines Studiums unverzichtbar sind. Der SAT hat seit seinem ersten Einsatz schon mehrere Veränderungen durchlaufen. Die letzte Novellierung fand im Jahr 2005 statt. Heute besteht der SAT I aus drei Teilen: „Critical Reading“, „Mathematics“ und „Writing“. Der Test dauert ca. vier Stunden und enthält hauptsächlich Multiple-Choice-Aufgaben. Er ermöglicht Aufschlüsse über schlussfolgerndes Denken und Problemlöse-

⁶ Der Begriff Testtraining wird in dieser Arbeit synonym mit dem Begriff Coaching verwendet.

fähigkeiten. Die maximale Punktzahl, die ein Bewerber im SAT I erreichen kann, sind 2400 Punkte. Die Spannbreite bei jedem der drei Untertests liegt zwischen 200 bis 800 Punkten. Der SAT II testet vorwiegend schulisches Wissen und dessen Anwendung. Er wird für 20 verschiedene Fächer angeboten. Jeder Fachwissenstest dauert ca. eine Stunde und besteht aus Multiple-Choice-Fragen. Hauptsächlich lassen sich die Bewerber in drei der folgenden fünf Fächern testen: Englisch, Geschichte, Sozialwissenschaften, Mathematik und Naturwissenschaften. Der SAT II kann aber um viele Fächerkomponenten, naturwissenschaftlicher wie sprachlicher, erweitert werden. Das Ergebnis der SAT II-Tests wird in manchen Hochschulen zur Einstufung bei den Kursen verwendet (Heine, 2006). Der SAT kann an sechs Terminen im Jahr in verschiedenen Testzentren, die über das ganze Land und auch im Ausland verteilt sind, absolviert werden. Die Bewerber haben die Möglichkeit, den Test mehrmals abzulegen und sich dann zu entscheiden, welches Testergebnis sie bei den Hochschulen mit ihren Bewerbungsunterlagen einreichen. Briggs berichtete für das Jahr 2008 folgende Durchschnittswerte in den drei Bereichen des SAT I: Mathematik = 515 Punkte, Kritisches Lesen = 502 Punkte und Schreiben = 494 Punkte (Briggs, 2009).

5.3 ACT

Die Abkürzung ACT bezeichnet den American College Test. Dieser allgemeine Studierfähigkeitstest existiert bereits seit 1959. Er wurde entwickelt und wird weiterhin verwaltet von der American College Testing in Iowa (ACT Inc.). Der Test besteht hauptsächlich aus Multiple-Choice-Aufgaben in vier verschiedenen Bereichen: Englisch, Lesen, Mathematik und wissenschaftlicher Argumentation. Im ACT werden schulstoffbezogenes Wissen und der Gebrauch von diesem zusammen mit kognitiven Kompetenzen, wie schlussfolgerndem Denken und Problemlösen, getestet (Heine, 2006). Die Fragen zum schulstoffbezogenen Wissen unterscheiden den ACT vom SAT I, der nur kognitive Kompetenzen abfragt. Auch der ACT wurde im Februar 2005 wie der SAT I um einen Schreibtest erweitert. Es gibt jedes Jahr ebenfalls bis zu sechs Termine in den gesamten USA und verschiedenen anderen Ländern, an denen die Studienbewerber den ACT absolvieren können. Insgesamt dauert der ACT ohne den Schreibtest ca. vier Stunden. Die Kandidaten dürfen den Test bis zu zwölf Mal wiederholen. 55 Prozent derjenigen, die den Test ein zweites Mal ablegten, konnten sich

verbessern. Der maximale Testwert, den die Teilnehmer erreichen können, beträgt 36 Punkte. Für das Jahr 2003 berichtete Briggs (2009) von einem Durchschnittswert von 21.1 Punkten für den ACT.

In der nachfolgenden Tabelle 5.1 sind einige aktuelle Beispieltestwerte für die einzelnen Sektionen des SAT I und den ACT von verschiedenen Colleges in den USA angegeben. Der untere Wert steht für die 25 Prozent Studenten, die zugelassen wurden und einen Wert hatten, der diesem Wert oder einem Wert darunter entsprach. Der obere Wert entspricht den 75 Prozent Studenten, die zugelassen wurden mit diesem Wert oder niedriger. Das bedeutet, dass 25 Prozent der zugelassenen Studenten einen Wert hatten, der über dem oberen Wert lag. Anhand der von den Colleges ausgegebenen Werte können sich die Bewerber informieren: erstens welche Punktzahl sie benötigen, um an ihrem Wunschcollege mit hoher Wahrscheinlichkeit aufgenommen zu werden. Zweitens können sie sich auch entscheiden, wie viel Chancen sie ihren verschiedenen College-Bewerbungen einräumen: risikoreich, passend oder sicher.

Tabelle 5.1: Die aktuellen erforderlichen Testwerte (26.11.2009) einiger ausgewählter Colleges in den USA für den SAT I und den ACT. Quelle: <http://www.collegeboard.com/>

College	SAT-V ^a	SAT-M ^b	SAT-W ^c	ACT ^d
Highland Community College	290 – 480	380 -430	450 – 450	17 -22
Nevada State College	410 – 530	410 – 520	390 – 520	16 – 23
Princeton	690 – 790	700 – 790	690 – 780	31 – 34
Harvard	690 – 800	700 – 780	690 – 790	31 – 35
Yale	750 – 800	700 – 780	700 – 790	30 – 34

Anmerkungen: ^aSAT-V: verbaler Teil;, ^bSAT-M: mathematischer Teil, ^cSAT-W: Schreibtest, In jedem SAT-Untertest beträgt liegt die erreichbare Punktzahl zwischen 200 bis 800 Punkten, ^dACT: maximal erreichbare Punkte: 36

5.4 „test-wiseness“, „test-practice“ und „test-coaching“

Zur Erklärung der drei wichtigen Begriffe, die im angloamerikanischen Sprachraum hinsichtlich Testtrainings eine große Rolle spielen, wird noch einmal auf die Arbeit von

Deter (1982) eingegangen. Deter (1982) hat die drei Begriffe in seiner Arbeit ausgiebig mit Hilfe der angloamerikanischen Literatur beschrieben.

Unter „test-wiseness“ versteht man den Gebrauch der bestmöglichen Vorgehensweise bei der Testbearbeitung. Die Strategien sind losgelöst vom Inhalt der Fragestellung. Der Begriff „test-wiseness“ schließt einzig allein Bearbeitungsmethoden ein, mit denen man den Testwert erhöhen kann, ohne dass das entsprechende Mehr-Wissen vorhanden ist (Deter, 1982). Möchte man einen zukünftigen Testteilnehmer in „test-wiseness“ unterrichten, werden ihm zwei Arten von Strategien vermittelt. Zunächst einmal werden ihm Vorgehensweisen erläutert, die es ermöglichen, von Itemkonstruktionsfehlern und von seinem Wissen über das Ziel des Tests zu profitieren, wie z. B. der Strategie der gewünschten Antwort und der Strategie zur Ausnutzung versteckter Lösungshinweise (Deter, 1982). Außerdem wird der künftige Testkandidat in Arbeitsweisen unterrichtet, die es ihm erlauben, besser mit der Bearbeitung des Tests zu verfahren. Diese Methoden sind unabhängig von Fehlern in der Test- und Aufgabenkonstruktion und getrennt vom Anwendungszweck des Tests. Darunter versteht man Zeiteinteilungsstrategien, Fehlervermeidungsstrategien, Ratestrategien und Strategien für das deduktive Schließen (Deter, 1982). Zum Messen von „test-wiseness“ wurden verschiedene Instrumente entwickelt, die aber immer nur einen Ausschnitt des gesamten Konstrukts messen können. Obwohl es schwer ist „test-wiseness“ zu operationalisieren, sollte die Relevanz der „test-wiseness“ bei der Anwendung von Tests, die zur Auswahl dienen, nicht unterschätzt werden. Es hat sich deutlich gezeigt, dass Personen, die bereits testerfahren waren, sich in ihrer Leistung zu Testunerfahrenen unterschieden.

Mit „test-practice“ ist die Testwiederholung gemeint. Ein Testteilnehmer, der mit demselben Testverfahren oder einem ähnlichen Testverfahren wiederholt getestet wird, besitzt „test-practice“. Es ist fast immer so, dass ein Testteilnehmer von einer Testwiederholung profitiert.

Der englische Begriffe „test-coaching“ umfasst mehr als nur Testwiederholung. Beim „test-coaching“ handelt es sich um Vorbereitungsprogramme, bei denen es explizit um die Erhöhung des Testwerts geht. Insofern kann „test-coaching“ viele Arten von Testvorbereitungen umfassen: das Lernen von Faktenwissen, das Üben von Testaufgaben, das Erlernen von Teststrategien, das Wiederholen von Schulwissen, also

alles, was den Testwert verbessern kann. Möchte man „test-coaching“ messen, muss der Einfluss systematischer Trainingsprozesse untersucht werden. Eine Schwierigkeit liegt aber darin, dass sich die jeweiligen Programme in vielen Faktoren voneinander unterscheiden.

Spricht man vom „test-coaching“, beinhaltet dieser Begriff immer auch „test-wisenes“ und „test-practice“. In einem „test-coaching“ werden zusätzlich noch Lerninhalte und Lernstrategien vermittelt, z. B. ein Mathematikkurs. Eine Abgrenzung der drei Begriffe voneinander ist insofern schwierig. Wichtig ist, dass die Strategien zur Steigerung von „test-wisenes“ keinen Anstieg der kognitiven Fähigkeiten impliziert. Eine Vermehrung von „test-wisenes“ sollte einem Testteilnehmer bei einem gut konstruierten Test nicht weiterhelfen. Die nachfolgende Darstellung der Ergebnisse der Studien bezieht sich allgemein auf das Thema „test-coaching“.

5.5 Studien zum Thema Testtraining aus den USA

Das Thema Testtraining bzw. Coaching ist in den USA von großer Relevanz. Die meisten amerikanischen Colleges nehmen den Testwert eines allgemeinen Studierfähigkeitstests als eines ihrer Hauptkriterien zur Zulassung zum Studium. Gleichzeitig herrscht in den USA großer Andrang auf einen Studienplatz an einer renommierten Universität, da die meisten Bewerber der Meinung sind, dass sie ihre berufliche Zukunft durch ein Studium an einer dieser Universitäten am besten absichern zu können. Das sichere Erlangen eines Studienplatzes an einer renommierten Hochschule ist am besten über einen überdurchschnittlich hohen Testwert zu erreichen. Dabei reichen manchmal auch schon kleinere Punktgewinne aus, um der Zulassung an seiner Wunschuniversität näher zu kommen.

Die meisten Coaching-Studien in den USA wurden zum SAT I durchgeführt, so dass in diesem Abschnitt nur auf diese Untersuchungen eingegangen wird. Dabei gibt es eine Menge von Einzelstudien aber auch eine größere Anzahl von Meta-Analysen. Die meisten hierzu angefertigten Studien in den USA sind jedoch nicht unabhängig von bestimmten politischen und wirtschaftlichen Interessen. So kann man die durchgeführten Studien grob nach zwei Interessensgemeinschaften einteilen. Zu der einen Gruppe gehören die Wissenschaftler des Educational Testing Service (ETS), die ein

Interesse haben, dass der Coaching-Effekt möglichst klein ist. Zu der anderen Gruppe gehören die Unternehmen, die Coaching-Programme anbieten, und daher daran interessiert sind, dass die Coaching-Effekte eher groß sind. Natürlich gibt es auch noch die Gruppe der Testgegner. Sie sind der Meinung, dass Studierfähigkeitstests nicht fair sind.

In diesem Kapitel werden einzig allein die Ergebnisse der Studien, ihre Ergebnisse und Einschränkungen vorgestellt. Es ist dabei nicht von Bedeutung, welche Intention die Wissenschaftler gehabt haben, ihre Analysen durchzuführen.

Bevor die Studien vorgestellt werden, soll noch einmal auf das Thema der prognostischen Validität des Messinstruments hingewiesen werden. Es gibt immer wieder Diskussionen darüber, dass ein Testtraining auch die Validität eines Studierfähigkeitstests einschränken würde. Bisher sind aber die meisten Wissenschaftler übereingekommen, dass die Validität des SAT I durch einen möglichen Coaching-Effekt nicht eingeschränkt würde. Dass bedeutet aber nicht, dass alle Forscher dieser Überzeugung sind. Immer wieder wird auch darauf hingewiesen, dass durch ein Coaching auch Fähigkeiten trainiert würden (vergleiche im Abschnitt 5.4 die Beschreibung von test-wiseness), die nicht mit dem zu messenden Konstrukt der Studierfähigkeit übereinstimmen. Würde dieses Training der konstruktfernen Fähigkeiten eine Steigerung des Testwerts zur Folge haben, dann wäre die Validität des Studierfähigkeitstests eingeschränkt. Allaouf hat in seiner Arbeit (Allalouf & Ben-Shakhar, 1998) untersucht, ob ein Coaching die prognostische Validität eines Studieneignungstests einschränken würde. Er kam zu dem Ergebnis, dass Coaching zwar einen Effekt hat, aber die prognostische Validität dadurch nicht eingeschränkt wird und es so zu keiner Prognoseverzerrung kommt. In dieser Arbeit kann auf diesen Aspekt aber nicht näher eingegangen werden.

Im Jahr 1981 untersuchten Messick et al. verschiedene Studien zum Thema aufgewendete Vorbereitungszeit für den SAT-I. Sie fassten 19 Studien für den verbalen Teil des SAT-I und 14 Studien für den mathematischen Teil des SAT-I zusammen und setzten die beim Coaching aufgebrauchte Zeit in Zusammenhang mit dem erreichten Testwert. Durch die Berechnung der Rangkorrelation nach Spearman, mit dem man den Zusammenhang zweier ordinalskalierten Merkmale berechnen kann, konnten sie

einen starken Zusammenhang zwischen der aufgebrauchten Zeit und dem erreichten Testwertgewinn erkennen. Der Zusammenhang in Zahlen ist in der Tabelle 5.2 und Tabelle 5.3 dargestellt. Allerdings darf hier nicht außer Acht gelassen werden, dass in den jeweiligen Untertests zwischen 200 bis 800 Punkte erreicht werden können. Um einen Punktgewinn von 30 Punkten im verbalen Teil des SATs zu erreichen, müsste ein Studienanwärter also ca. 260 Stunden investieren. Wenn man davon ausgeht, dass ein Student bis zu fünf Stunden am Tag lernen kann, würde das bedeuten, dass er ca. 50 Tage üben müsste, um am Ende einen Gewinn von ungefähr 30 Punkten zu erzielen.

Tabelle 5.2: Testwertgewinn in der verbalen Sektion des SAT ins Verhältnis gesetzt mit der beim Coaching für diese Sektion aufgebrauchte Zeit nach Messik et al. (Messick & Jungeblut, 1981)

Punktgewinn im SAT-V	mit Coaching verbrachte Zeit
10 Pkt.	12 h
20 Pkt.	57 h
30 Pkt.	260 h

Tabelle 5.3: Testwertgewinn in der mathematischen Sektion des SAT ins Verhältnis gesetzt mit der beim Coaching für diese Sektion aufgebrauchte Zeit nach Messik et al. (Messick & Jungeblut, 1981)

Punktgewinn im SAT-M	mit Coaching verbrachte Zeit
10 Pkt.	8 h
20 Pkt.	19 h
30 Pkt.	45 h
40 Pkt.	107 h

Im mathematischen Teil des SATs ist der Zeitaufwand für einen Punktgewinn nicht ganz so intensiv. Der Zusammenhang ist in den beiden Graphiken in Bild 5.1 noch einmal verdeutlicht.

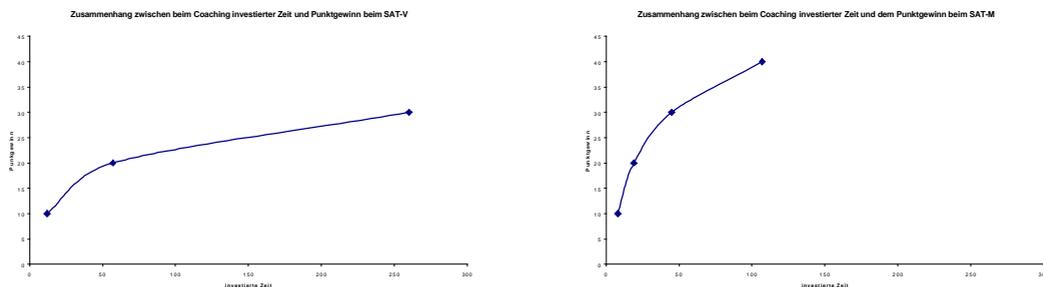


Bild 5.1: Graphische Darstellung des Zusammenhangs zwischen investierter Coaching-Zeit und Punktzahl im SAT-V bzw. SAT-M; Zahlen entnommen aus (Messick & Jungeblut, 1981).

Einschränkend erwähnt Messick et al. (1981) aber, dass es sich um keinen linearen, sondern um einen logarithmischen Zusammenhang handelt. Messick et al. (1981) führen ebenso an, dass fast jede Studie, die sie in die Untersuchung aufgenommen haben, methodische Fehler aufwies, und dass sich die Studien bezüglich Dauer, Inhalt und Qualität unterschieden. Außerdem machen sie darauf aufmerksam, dass sie den oben skizzierten Zusammenhang nur in den von ihnen untersuchten und aggregierten Studien finden konnten. Der vorgefundene Zusammenhang zwischen Coaching-Zeit und Testwert sollte dementsprechend nur vorsichtig generalisiert werden. Briggs (2002) replizierte die Methode von Messick et al. (1981) 20 Jahre später mit neueren Studien. Die Rangkorrelationen verringerten sich und waren nicht mehr signifikant. Das von Messick et al. (1981) entdeckte Muster konnte nicht noch einmal gefunden werden.

Betsy Jane Becker berechnete im Jahr 1990 aus den bis dahin vorhandenen Coaching-Studien eine Meta-Analyse. Die Studie spricht eindeutig von einem Coaching-Effekt, aber in der Analyse wurde betont, dass dieser Effekt sehr gering sei (Becker, 1990). Für den verbalen Teil des SAT-I berechnete sie einen Effekt von 9 – 25 Punkten und für den mathematischen Teil einen Effekt von 15 – 25 Punkten. Becker (1990) stellt auch noch einmal heraus, dass es grundsätzlich sehr schwierig sei, die Studien untereinander zu vergleichen, da sie sich hinsichtlich ihres Versuchsdesigns, ihrer Stichproben und ihrer angewandten Treatments unterscheiden. Alle diese Faktoren besäßen aber einen Einfluss auf die Größe des Trainingseffektes.

Powers und Rock führten im Jahr 1999 eine Beobachtungsstudie mit ca. 4000 Teilnehmern durch. Die beiden Wissenschaftler arbeiteten zu dieser Zeit am ETS und

hatten dadurch Zugriff auf die tatsächlichen SAT-Werte (Powers, 1999). Die 4000 Teilnehmer am SAT-I wurden mit Hilfe eines Fragebogens zu verschiedenen Merkmalen, z. B. Art der Vorbereitung, aber auch Herkunft, Abschlussnoten etc. interviewt. Powers et al. (1999) kamen ebenfalls zu dem Schluss, dass es einen Coaching-Effekt gebe. Dieser aber sehr gering ausfalle. Der Effekt für den verbalen Teil beliefe sich auf 6 – 12 Punkte. Der Effekt für die Mathematik-Sektion sei größer und betrage hier 13 – 18 Punkte. Der Coaching-Effekt für den SAT-I wurde berechnet auf 21 – 34 Punkte. Zusätzlich stellten die Wissenschaftler fest, dass sich die beiden Gruppen, die Coaching-Gruppe und die Nichtcoaching-Gruppe, hinsichtlich ihrer High-School Abschlüsse, ihrer beruflichen Ambitionen und ihrer elterlichen Einkommen unterschieden. Die Besucher der Trainingskurse hatten häufig einen viel besseren High-School Abschluss als die Testteilnehmer, die sich nicht hatten coachen lassen. Powers et al. (1999) erwähnten allerdings, dass eine Fragebogenuntersuchung immer eine Einschränkung der internen Validität bedeutet, da man nie wisse, ob die Befragten auch wahrheitsgemäß geantwortet hätten.

Briggs (2001) untersuchte Daten, die er von der National Longitudinal Survey (NELS) erhalten hatte. Die Stichprobe enthielt Daten von ca. 16.500 Studenten, die von NELS in den Jahren 1988, 1990 und 1992 befragt wurden (Briggs, 2001). Den Studenten wurde unter anderem eine Frage zu ihrer Art der Vorbereitung gestellt. Briggs (2001) wertete die Daten hinsichtlich der verschiedenen Formen der Vorbereitung im Zusammenhang mit einer Steigerung des Testwerts aus. Er stellte eine Testwertsteigerung bei Testteilnehmern, die einen Trainingskurs einer Coaching-Company besucht hatten, von 17 Punkten im SAT-M und von 13 im SAT-V fest. Diese Werte verwendete MacGowan zur Berechnung seiner Voraussetzung, ab wann man überhaupt von einem Trainingsgewinn sprechen könne.

Im Jahr 2002 gab MacGowan in einem Artikel zum Coaching für den SAT-I einen Überblick über die bis dahin verfassten Studien. Da er in seinem Artikel viele methodische Fehler bemängelte, führte er selbst eine Beobachtungsstudie zu diesem Thema durch. MacGowan (2002) wies aber ausdrücklich darauf hin, dass auch diese Studie methodische Mängel aufweise. Aus den bereits vorhandenen Studien zog er folgende Schlüsse: Einige Trainingskurse seien effektiver als andere. Diese Tatsache

komme in den Studien nicht mehr zu Tage, da die meisten Studien die Coaching-Effekte verschiedener Trainingskurse lediglich aggregieren und einen Mittelwert berechnen würden, um diesen als gesamten Coaching-Effekt darzustellen (MacGowan, 2002). MacGowan äußerte die Überzeugung, dass so wichtige Informationen verloren gehen. Die einzig mögliche Art der Untersuchung, ob es einen Coaching-Effekt gebe und wie hoch dieser sei, wäre, wenn man jeden Studenten mit seiner individuellen Art der Vorbereitung als Einzelfallstudie betrachten würde. In seiner Studie befragte MacGowan (2002) Studienbewerber nach ihrer Art der Vorbereitung. MacGowan legte für seine aufgestellte Hypothese fest, ab wann man überhaupt von einem Coaching-Effekt sprechen könne, d.h. wie hoch dieser Punktgewinn sein müsste. Die Voraussetzungen, um auf einen Coaching-Effekt schließen zu können, waren: ein Punktzuwachs in dem verbalen SAT-I, der über 40 Punkten lag, und ein Punktzuwachs im mathematischen Teil von über 90 Punkten. MacGowan (2002) hatte diese Werte berechnet aus einer Gleichung bestehend aus den Werten, die Briggs (2001) berichtet hatte, addiert mit Werten aus einer Studie des College Boards aus dem Jahr 1999, die die Testwertsteigerung berechnet hatte bei einer Wiederholung des SAT-I. Außerdem versuchte er in seiner Fragestellung zu berücksichtigen, dass Testtrainings nicht beliebig vergleichbar seien und auch die Besucher eines Trainingskurses nicht immer auf ein Coaching positiv reagierten. Er stellte also die Hypothese auf, dass wenigstens die Hälfte derjenigen, die Vorbereitungsaktivitäten unternommen hatten, deutliche Punktsteigerungen in beiden Test-Sektionen verzeichnen müssten. Bei der Auswertung seiner Befragung der Studienbewerber stellte MacGowan (2002) fest, dass sich von am Ende

n = 24 in die Studie eingeschlossenen Personen nicht zwei auf die gleiche Art und Weise vorbereitet hatten. Da MacGowan (2002) seine Daten, wie oben schon erwähnt, einzeln aufgeschlüsselt hatte, würde es hier zu weit führen, auf die genaue Darstellung der Ergebnisse einzugehen. Er konnte seine Hypothese bestätigen, dass sich wenigstens bei der Hälfte derjenigen, die sich vorbereitet hatten, die oben erwähnten Punktsteigerungen im SAT-I erfüllt hatten. MacGowan (2002) kam daher zu dem Schluss, dass Vorbereitungsmaßnahmen einen positiven Effekt haben können, aber dass nicht alle Arten von Coaching-Kursen bei jedem Studenten eine Teststeigerung bewirken. Grundsätzlich geht er jedoch davon aus, dass der SAT-I trainierbar ist.

In seiner Dissertation aus dem Jahr 2002 ging Briggs auf das Thema Testtraining ein. Er näherte sich dem Thema über den methodischen Ansatz. Auch er ist fest davon überzeugt, dass es einen Coaching-Effekt gibt (Briggs, 2002). In seiner Arbeit spricht Briggs (2002) aber eher davon, wie man den Coaching-Effekt deutlich machen kann. Er stellte in seiner Studie keine konkrete Zahl für die Größe des Coaching-Effektes vor, sondern regte an, die Größe des Effektes eher als eine Funktion zu sehen. Die Funktion setzt sich aus verschiedenen Elementen zusammen: Merkmale des Coachings sowie Merkmale des Studenten. In seiner Arbeit gab Briggs (2002) einen Überblick über die bisher veröffentlichte Literatur und ging eingehend auf die methodischen Probleme der verschiedenen Studien ein. Briggs (2002) selbst versuchte mit Hilfe zweier verschiedener statistischer Methoden, lineare Regression und Heckman Modell, das Problem der Bias⁷ zu lösen. Ein Bias, der bei fast allen Untersuchungen besteht, ist der „self-selection bias“. Mit „self-selection bias“ ist die Selbstselektion der Probanden gemeint. Die Ausprägung der unabhängigen Variablen, Coaching und Nicht-Coaching, wird den Probanden nicht zufällig zugeteilt, sondern die Probanden wählen selbst, zu welcher Gruppe sie gehören wollen. Man kann also nicht davon ausgehen, dass sich die beiden Gruppen nicht nur zufällig hinsichtlich ihrer Merkmale, z. B. Abschlussnoten, Einkommen der Eltern, etc. unterscheiden. Die Ergebnisse der quasiexperimentellen Untersuchung sind mehrdeutig interpretierbar. Es könnte sein, dass die unabhängige Variable, die Teilnahme am Coaching, mit anderen Variablen konfundiert ist. Diese Variablen, z. B. High-School Abschluss oder Einkommen der Eltern, hat eventuell ebenfalls einen Einfluss auf die abhängige Variable, den Testwert (Bortz & Döring, 2006). Briggs (2002) ging in seiner Arbeit hauptsächlich auf diese methodischen Einschränkungen aller bisher gemachten Coaching-Studien zum SAT-I ein. Briggs (2002) machte aber deutlich, dass er fest davon überzeugt war, dass es einen positiven Coaching-Effekt gibt. Die Höhe des Effekts hängt aber als eine Art Funktion von vielen anderen Faktoren ab.

Im Jahr 2009 veröffentlichte Briggs die neueste Analyse zum Thema Coaching-Effekt. Die Untersuchung ist wiederum eine Literaturzusammenfassung und Vorstellung der bisher gemachten Studien zum Thema Testvorbereitung für das College. Es werden

⁷ Bias = systematischer Verzerrungseffekt, konstanter Fehler

darin auch Studien zum ACT vorgestellt, auf die aber hier nicht eingegangen wird. Briggs (2009) stellte auch in diesem Report erneut fest, wie schwer es sei, eine allgemeine Aussagen über die Höhe eines Coaching-Effekts zu machen. Erschwerend kommt hinzu, dass sich der SAT-I seit 1953 mehrfach verändert hat. Die Coaching-Studien aus den vorangegangenen Jahrzehnten können also nicht einfach zusammengefasst werden. Ein neueres Hemmnis kommt mit der zunehmenden Digitalisierung hinzu. Viele Studenten nutzen inzwischen die Möglichkeit sich via Internet auf die Studierfähigkeitstest vorzubereiten. Wie hoch der Effekt der computerbasierten Vorbereitungsprogramme ist, ist bisher noch nicht erforscht. Briggs (2009) hatte schon in seinen früheren Arbeiten immer wieder erwähnt, dass es einen Coaching-Effekt gebe. Der Effekt sei aber laut Briggs (2009) sogar kleiner als der Standardmessfehler, der beim SAT-I bei ca. 30 Punkten für jede Sektion liege. Für die Zulassung an der Wunschhochschule, so Briggs (2009) reiche manchmal aber auch schon ein kleiner Testgewinn.

5.6 Anmerkungen zu den Studien zum SAT

Gemeinsam ist allen Studien, dass sie immer wieder betonen, dass es eigentlich kaum möglich ist, einen generellen Coaching-Effekt nachzuweisen. Es gibt so gut wie keine Studie ohne methodische Schwächen. Auf der anderen Seite ist es eigentlich nicht möglich ein Zufallsexperiment durchzuführen, damit ein Coaching-Effekt methodisch korrekt beschrieben werden kann. Denn würde man eine genügend große Stichprobe, die natürlich alle möglichen Persönlichkeitsmerkmale der zukünftigen Studenten in den USA abdecken müsste, einer Treatment- und einer Kontrollgruppe zuordnen, begegnet man wieder neuen Problemen. Ein solcher Versuch wäre ethisch und logistisch nicht realisierbar. Man könnte nicht einer Gruppe das Coaching bei einem so wichtigen Test verweigern. Auf der anderen Seite müsste die Motivation, das beste Ergebnis im SAT-I zu erreichen, bei beiden Gruppen gleich hoch sein. Schwierig wäre es auch zu gewährleisten, dass alle Versuchspersonen der Treatmentgruppe immer zum Training gehen würden während sich die Kontrollgruppe gleichzeitig nicht informieren würde. So sind fast alle Studien zum Thema Testtraining für den SAT-I Quasi-experimentelle Studien. Die Studienbewerber haben also entschieden, ob sie an einem Training teilnehmen oder nicht. Wie schon erwähnt, könnte die unabhängige Variable,

Teilnahme am Coaching, mit anderen Variablen konfundiert sein und diese haben ebenfalls eine Auswirkung auf den Trainingsgewinn. Um dies zu verhindern, müssten die Stichproben nach verschiedenen Merkmalen parallelisiert werden. Erst dann könnte man die beiden Gruppen hinsichtlich ihrer Trainingsgewinne inferenzstatistisch miteinander vergleichen. Die aktuellsten Studien weisen immer wieder darauf hin, dass man die Coaching-Programme nicht ohne weiteres miteinander vergleichen könne, da diese sich in vielen Merkmalen, wie z. B. Dauer, Inhalt und Qualität unterscheiden. Zudem reagiert jeder Student anders auf das jeweilige Trainingsprogramm. Gerade beim SAT-I ist es schwierig, die Studien über die Jahrzehnte hinweg zusammenzufassen, da gerade dieser Test fortwährend Veränderungen unterliege, die natürlich auch Auswirkungen auf die Vorbereitungsmaßnahmen haben.

5.7 Bezug zum TMS

Die Studien aus dem angloamerikanischen Raum zeigen noch einmal deutlich, dass auch hier die Forscher das Problem hatten, dass der Trainingseffekt nicht eindeutig zuzuordnen ist. Es gibt immer genügend Drittvariablen, die ebenfalls einen Einfluss haben. Briggs (2009) betonte daher ebenso wie Trost et al., dass alle Studienbewerber angehalten werden sollten, sich auf die Studierfähigkeitstests vorzubereiten.

6 Testtraining

Sowohl in Kapitel 4 als auch in Kapitel 5 wurde bereits deutlich, dass es nicht möglich ist Trainingsprogramme ohne Weiteres untereinander zu vergleichen, da sie sich in Dauer, Umfang und Inhalt stark voneinander unterscheiden. So wird in dieser Arbeit auch nur ein Trainingsprogramm vorgestellt werden. Alle empirischen Daten beziehen sich auf dieses Testtraining.

6.1 Konzept

Die vorliegende Arbeit bezieht sich auf das Testtraining, das von der Carrer-Counseling Laufbahnberatung Dr. Hohner (CCH), einem staatlich eingetragenen Institut für Prüfungsvorbereitung, berufliche Entwicklung und Weiterbildung angeboten wurde. Das Institut wurde im Jahr 1986 gegründet und bietet bis heute verschiedene Beratungen zur Berufsorientierung an. In dieser Arbeit wird nur auf das professionelle Testtraining, das die CCH-Laufbahnberatung zur Vorbereitung auf den Medizinertest (TMS) in den Jahren 1986 bis 1997 angeboten hat, eingegangen.

Deter (1982) unterschied in seiner Arbeit deutlich zwischen Trainingsgewinn und Übungsgewinn. Wie allerdings aus den amerikanischen Studien hervorging, kann man diese beiden Effekte nicht ohne Weiteres voneinander trennen. So wird auch bei der nachfolgenden Beschreibung des Testtrainings deutlich, dass es sich um ein „test-coaching“ handelte. In dem Seminar von der CCH-Laufbahnberatung gab es Elemente von „test-practice“, von „test-wiseness“ und „test-coaching“.

Das Konzept der Testtrainingsseminare für den TMS beruhte auf verschiedenen Prinzipien. Zunächst einmal wurde größte Sorgfalt darauf verwandt, dass sämtliche Trainingsaspekte im höchsten Maße mit der Ernstfallsituation übereinstimmen. Nicht nur im Hinblick auf die Aufgabentypen und die Teststruktur wurde auf möglichst reale Bedingungen geachtet, sondern auch in Bezug auf zeitliche, psychologische und physiologische Belastungen. Trainiert wurde anhand von veröffentlichten Originalversionen sowie anhand von Experten entwickelten Testversionen. Um auf individuelle Schwächen einzelner Teilnehmer eingehen zu können, wurde zum Teil in kleinen Gruppen gearbeitet. Alle Trainingsseminare wurden von qualifizierten Diplom-

Psychologen geleitet. Gleichzeitig wurde von den Teilnehmern aktive Mitarbeit verlangt, um einen möglichst hohen Lerneffekt zu erreichen (Hohner, 1995). Grundsätzlich orientierte sich die Gestaltung der Seminare an folgenden Zielvorstellungen:

- Herstellung einer realitätsbezogenen Leistungsmotivation und einer positiven Einstellung zum Test
- Simulation der Testsituation und Herstellung möglichst wirklichkeitsgetreuer Originalbedingungen
- Systematische Bearbeitung aller Untertests des TMS, Vermittlung von Beantwortungstechniken und Lösungsstrategien anhand von Originalaufgaben
- Verbesserung von Arbeitstechniken und Zeitmanagement
- Bewältigung von Stress und Prüfungsangst
- Anleitungen für die weitere Vorbereitung allein zu Hause (Hohner, 1991)

Das Konzept des Testtrainingseminars sah vor, dass die Kursteilnehmer sich nicht nur mit dem Inhalt der Testaufgaben auseinandersetzen sollten, sondern sie sollten auch lernen, wie sie angemessenen mit ihrer Zeit beim Bearbeiten der Untertests umgehen und wie sie mit eventuell auftretenden Prüfungsängsten fertig werden können.

Die CCH-Laufbahnberatung hat neben dem Testtraining zum Medizinertest zur Auffrischung der Mathematikkenntnisse Mathematik-Workshops und zu verschiedenen Zeitpunkten Generalproben zum Medizinertest angeboten. Die Generalproben entsprachen einer Simulation des TMS in Echtzeit unter realistischer Prüfungsbedingung. Während des Ablaufs der Generalproben orientierte man sich streng an die formalen Rahmenbedingungen des TMS (Hohner, 1995). Bei der empirischen Untersuchung dieser Arbeit sind nur das Testtraining und die Generalprobe relevant.

6.2 Seminar und Generalprobe

Wichtiges Ziel war es, sich mit dem Test vertraut zu machen. Hierfür wurden seine spezifischen Merkmale ausführlich untersucht, wie z. B. die Testsituation, der

Testablauf, Testaufbau- und inhalt, die Bearbeitungszeit, das Bewertungsschema und das Antwortschema. Die Studenten konnten auf diese Weise erkennen, bei welchen Aufgabentypen sie Schwierigkeiten hatten. Ziel war es, spezielle Fähigkeitsmängel zu erkennen und auszugleichen. Mit den Teilnehmern wurden anschließend die Testmerkmale besprochen, das Aufgabenschema betrachtet und Arbeitsanleitungen erörtert. Diese Kurspunkte waren wichtig, damit ausgeschlossen werden konnte, dass die Leistungen im Test durch Instruktionsmissverständnisse oder generelle Bearbeitungsfehler gemindert wurden. Der formale Ablauf des Tests wurde präsentiert, so dass die zukünftigen Testteilnehmer mit dem Testablauf schon einmal vertraut gemacht waren. Sie wurden über den zeitlichen Rahmen der einzelnen Untertests und des Tests insgesamt informiert. Des Weiteren wurden im Laufe des Seminars konkrete Übungsaufgaben bearbeitet und die für die Untertests spezifischen Anforderungen trainiert. Immer wieder wurden die Untertests unter den Bedingungen simuliert, wie sie die Studenten dann im Ernstfall vorfinden würden. Diese Simulationen wurde nicht nur zur Übung genutzt, sondern ihre Ergebnisse gaben auch Aufschluss über die Stärken und Schwächen der Testteilnehmer. So konnte für jeden ein individuelles Leistungsprofil erstellt werden. Für Defizite kognitiver Natur bot das Seminar systematisches Üben an, d.h. das Aufgabenprinzip wurde rekapituliert, die Aufgaben wurden wiederholt geübt, es wurden weitere Fachkenntnisse vermittelt und Gedächtnistrainings durchgeführt. Für Leistungsdefizite motivationaler Natur oder aufgrund von Prüfungsangst wurden den Teilnehmern verschiedene Entspannungs- und Stressbewältigungstechniken vermittelt. Zusätzlich wurden mit ihnen Konzentrationsübungen einstudiert. Jeder Kursteilnehmer hatte stets die Möglichkeit zu einer Einzelberatung.

Da es sich um ein kurzfristiges Training handelte, konzentrierte es sich maßgeblich darauf, die Leistungen in den gut trainierbaren Untertests z. B. „konzentriertes und sorgfältiges Arbeiten“ und „Muster zuordnen“ zu verbessern. Wichtig war es bei diesen Tests, die Zeitkomponente zu beachten und die starke psychische Belastung zu bewerkstelligen. Diese Tests wurden mehrfach im Laufe des Seminars geübt.

Ebenfalls kurzfristig gut trainierbar waren die Untertests, die Merkvermögen „Figuren und Fakten lernen“ und die visuelle Wahrnehmung „Schlauchfiguren“ prüfen. Kaum

Lerngewinne konnten in den Untertests „Textverständnis“, „Quantitative und formale Probleme“ und „Medizinisch-naturwissenschaftliche Probleme“ erzielt werden. Diese Tests waren kurzfristig nicht gut trainierbar. (Maue, 1998).

Daneben wurden den Seminarteilnehmern verschiedene testunabhängige Strategien vermittelt, wie z. B. Strategien zur Zeiteinteilung, Fehlervermeidungsstrategien und Strategien zur Ausnutzung versteckter Hinweise (Maue, 1998).

Allen Teilnehmer wurde geraten, sich weiterhin zu Hause vorzubereiten, da es sich beim CCH-Seminar nur um ein kurzfristiges Training handelte. Insbesondere die nicht so gut trainierbaren Tests sollten zu Hause intensiver vorbereitet werden.

Die Generalprobe wurde unter Echtzeit simuliert. Die Aufgaben waren im Schwierigkeitsgrad vergleichbar mit den Originalaufgaben des TMS. Nach der Simulation wurden die Testhefte eingesammelt. Jeder Teilnehmer bekam dann ein persönliches Testleistungsprofil, auf dem seine Ergebnisse eingetragen waren. So konnten die Teilnehmer ihre Stärken und Schwächen in den einzelnen Untertests selbst einschätzen. Die Kursteilnehmer hatten neben dem persönlichen Leistungsprofil gleichzeitig durch ihre Teilnahme an der Generalprobe einen starken Übungseffekt, der insbesondere in einigen Untertests wie dem Konzentrationstest einen erheblichen Effekt hatte. Ansonsten wurden die Testteilnehmer durch die Generalprobe mit der Testsituation vertraut gemacht und wussten, was auf sie im Ernstfall für eine Belastung zukommen würde.

Die Teilnehmer waren nach Ablauf des Seminars durch die mehrfachen Wiederholungen mit den einzelnen Untertests vertraut, sowohl mit den Testaufgaben als auch mit der Testsituation. Sie kannten zudem die jeweiligen Aufgabeninstruktionen, hatten Mnemotechniken gelernt und konnten durch Stressbewältigungs- und Entspannungsstrategien mit dem enormen Leistungsdruck besser umgehen. Sie hatten sich außerdem eingehend mit den Lösungswegen der Aufgaben beschäftigt. Durch die Teilnahme an der Generalprobe hatten sie noch einmal eine zusätzliche Übung kurz vor dem eigentlichen TMS.

Beides zusammen, Seminar und Generalprobe, waren eine gute Vorbereitung auf den eigentlichen TMS. Die Studenten waren mit dem formalen Ablauf vertraut, sie kannten ihre Stärken und Schwächen, hatten gelernt mit Hilfe von Entspannungsübungen mit

der Prüfungsangst umzugehen und einem Motivationstief mit entsprechenden Konzentrationsübungen zu begegnen. Sie wussten nicht nur inhaltlich, was auf sie zukommen wird, sondern hatten auch ein Fülle von Bewältigungsstrategien erlernt. Die zukünftigen Testteilnehmer hatten um in den Begriffen des angloamerikanischen Raums zu sprechen, „test-practice“ und „test-wiseness“ gelernt und sich mit Hilfe eines „test-coachings“ auch inhaltlich optimal auf den TMS vorbereitet.

7 Befunde zur Transferevaluation

In diesem Kapitel wird das Testtraining der CCH-Laufbahnberatung bewertet. Zielsetzung der Evaluation ist die Überprüfung der Wirksamkeit des Trainings anhand bestimmter Bewertungskriterien (Kuckartz, Ebert, Rädiker & Stefer, 2009). Die folgende Evaluation wird durchgeführt angelehnt an das Vier-Ebenen-Modell von Kirkpatrick (Kirkpatrick, 2004). Dieses Modell gehört zu den nutzungsorientierten Ansätzen. Der Schwerpunkt des Modells liegt auf der Bewertung des ökonomischen Nutzens, den ein Unternehmen hat, wenn es ein Trainingsprogramm, z. B. ein Weiterbildungsprogramm, durchführt. Kirkpatrick (2004) hat verschiedene Techniken zur Evaluation von Trainingsprogrammen entwickelt. Auf jeder der vier Stufen wird das Trainingsprogramm unter einem anderen Aspekt evaluiert. Ausgangspunkt des Modells war die Frage, welchen Zwecken die verschiedenen Formen der Evaluationsergebnisse dienen. Kirkpatrick (2004) schlug daher vor, Ergebnisse auf vier unterschiedlichen Ebenen unter verschiedenen Gesichtspunkten zu beachten (Gollwitzer & Jäger, 2007). Vorrangig geht es bei dem Modell darum, den Nutzen, den ein Trainingsprogramm für das Unternehmen hat, nachzuweisen. Allerdings hat Kirkpatrick (2004) bei der Entwicklung seines Modells nicht außer Acht gelassen, dass ein Gewinn für das Unternehmen nur eintreten wird, wenn das Trainingsprogramm auch einen Lern- und Transfererfolg hat. Lern- und Transfererfolg sind die Voraussetzung dafür, dass überhaupt ein ökonomischer Organisationserfolg erzielt werden kann. Kirkpatrick's Modell setzt erst nach der Beendigung einer Maßnahme ein (Kirkpatrick, 2004). Es bewertet die Trainingswirkung auf den unterschiedlichen Reaktionsebenen einer Maßnahme. Das Vier-Ebenen-Modell ist in folgende vier Ebenen aufgeteilt:

1. Reaction bzw. Reaktionsebene
2. Learning bzw. Lernebene
3. Behaviour bzw. Handlungsebene
4. Results bzw. Erfolgsebene

Auf der Ebene 1 wird die Reaktion der Trainingsteilnehmer auf das Training erfasst. Dies kann geschehen durch Fragen, wie z. B. „Waren Sie zufrieden?“. Ausschlaggebend

ist hier die Zufriedenheit der Trainingsteilnehmer bzw. Kundenzufriedenheit. Auf der Ebene 2 wird der Lernerfolg bewertet. Es wird untersucht, ob die Teilnehmer ihr Wissen vermehrt haben, ihre Einstellungen verändert haben oder ihre Fähigkeiten verbessern konnten aufgrund ihrer Teilnahme an diesem Programm. Lernerfolg kann z. B. mit Vorher-Nachher-Messungen überprüft werden. Die Ebene 3 konzentriert sich auf die Verhaltensänderungen der Teilnehmer bzw. den Transfer von Gelerntem aus dem Training auf das Funktionsfeld, also an den Arbeitsplatz. Die Lerntransferevaluation überprüft, wie das während einer Intervention Gelernte auf andere Situationen außerhalb des Trainingsprogramms übertragen wird und ob es sich in einem veränderten Arbeitsverhalten bzw. in einem verbesserten Arbeitsergebnis ausdrückt (Häring, 2003). Dabei darf nicht unbeachtet bleiben, dass sowohl die Teilnehmermerkmale als auch das Trainingsdesign und auch die Arbeitsumgebung immer einen Einfluss auf den Lerntransfer haben (Häring, 2003). Die Frage ist also: Konnte das Gelernte im Arbeitsalltag umgesetzt werden? Gemessen werden kann das anhand von Fremd- oder Eigenbeurteilung. Die Ebene 4 ist auf den ökonomischen Aspekt ausgerichtet. Betrachtet wird die Wirkung des Trainingsprogramms in seiner Gesamtheit. Profitiert das Unternehmen von dem Training der Mitarbeiter (Ringlstetter, Aschenbach & Kirsch, 2003).

In der Literatur zum Vier-Ebenen-Modell wird häufig diskutiert, ob die Ebenen aufeinander aufbauen müssen. Kirkpatrick selbst hat keine Angaben dazu gemacht, ob die Ebenen voneinander abhängig sind (Häring, 2003). Es gibt allerdings auch einige Autoren, die meinen, dass das Modell von Kirkpatrick in einer bestimmten Reihenfolge durchgeführt werden muss (Newstorm, 1978; Hamblin, 1974 zit. n. Häring, 2003). Für die Verwendung des Modells als hierarchisches Modell gibt es keine empirischen Belege (Häring, 2003). Es gibt allerdings empirische Untersuchungen zu den kausalen Zusammenhängen zwischen den einzelnen Ebenen (Häring, 2003). In dieser Arbeit wird hier nicht näher darauf eingegangen. Es sei nur erwähnt, dass die in dieser Untersuchung durchgeführte Evaluation nach dem Vier-Ebenen-Modell von Kirkpatrick nicht von einem hierarchisch aufgebauten Modell ausgeht. Es wird ebenfalls nicht davon ausgegangen, dass es einen kausalen Zusammenhang zwischen den einzelnen Ebenen geben muss. Jede Ebene wird für sich im Einzelnen betrachtet werden. Wie Dumont du Voitel und Roventa in ihrem Artikel (Ringlstetter et al., 2003) erwähnten, ist das Modell

von Kirkpatrick anpassbar, einfach und einsichtig. Es müssen auch nicht Erhebungen auf allen Stufen vorgenommen werden. Viele Unternehmen ermitteln nur Daten auf der Ebene 1 und Ebene 2.

7.1 Evaluationsgegenstand und Fragestellungen

Evaluationsgegenstand ist das Testtraining der CCH-Laufbahnberatung, das in Kapitel 6 eingehend beschrieben wurde. Evaluationszweck ist die Bewertung des Testtrainings. Das Testtraining soll einerseits hinsichtlich der Zufriedenheit der Teilnehmer beurteilt werden und andererseits in Hinblick auf die konkreten Ergebnisse evaluiert werden, d.h. der TMS-Resultate. Da die Evaluation des Testtrainings nach dem „Vier-Ebenen-Modell“ von Kirkpatrick vorgenommen wird, werden im Folgenden auch die Fragestellungen betreffend der vier Ebenen aufgeteilt.

Ebene 1: Reaktionsebene

Übertragen auf die Evaluation des CCH-Testtrainings geht es hier um die Zufriedenheit der Seminarteilnehmer mit dem von CCH-Laufbahnberatung angebotenen Seminar und der Generalprobe.

- Beurteilten die Teilnehmer das Trainerverhalten als kompetent?
- Beurteilten die Teilnehmer das Trainerverhalten als engagiert?
- Beurteilten die Teilnehmer das Trainerverhalten als sympathisch?

Ebene 2: Lernebene

Auf dieser Ebene geht es darum zu analysieren, ob die Teilnehmer ihr Wissen vermehrt haben oder ihre Einstellung verändert haben. Anhand der vorliegenden Daten wird der Aspekt der Einstellungsänderung überprüft werden. Der Frage, ob sich die Fähigkeiten aufgrund der Teilnahme am Trainingsprogramm verbessert haben, wird analysiert werden, in dem überprüft wird, ob sich die im Seminar gemessenen Ausprägungen der Faktoren, die wichtig sind für ein gutes Abschneiden im TMS, zur Messung der Ausprägung bei der Generalprobe positiv verändert haben.

- Haben sich die Einstellungen in Bezug auf die Prüfungserwartungen der Teilnehmer durch die Teilnahme am Seminar signifikant verändert?
- Haben die Teilnehmer den Eindruck, dass bestimmte Faktoren durch die Teilnahme am Trainingsseminar besser zum Zeitpunkt der Generalprobe ausgeprägt sind als sie es zu Beginn Seminars waren?

Ebene 3: Verhaltensebene

Überprüft wird auf dieser Stufe der Transfer von Gelerntem aus dem Training in den Alltag. Übertragen auf das CCH-Testtraining ist hier die Frage, ob die Teilnehmer vom Training profitiert haben. Können sie das Gelernte auf andere Situation, speziell auf die TMS-Situation, transferieren. Gemessen wird dieses Transferverhalten anhand von Eigenbeurteilungen.

- Haben die Teilnehmer vom CCH-Testtraining profitiert?
- Wie fanden sie den Einblick in Aufbau und Logik des TMS und seiner Aufgaben?
- Wie fanden sie die Simulation der Testsituation?
- Haben ihnen die gelernten Bearbeitungsstrategien zu den einzelnen Untertests geholfen?
- Wie fanden sie die Strategien zur Arbeitstechnik, zum Umgang mit der Zeit und mit „Störungen“?
- Fanden sie die Übungen zur Konzentration, Entspannung und Selbstmotivierung nützlich?
- Haben ihnen die Hinweise und Tips zur weiteren Vorbereitung genutzt?
- Wie fanden sie das eingesetzte bzw. ausgegebene Seminarmaterial (z. B. Paralleltest, Würfel)?

Ebene 4: Erfolgsebene

Diese Ebene betrachtet die Wirksamkeit des Trainingsprogramms. Hat das Testtraining den erwarteten Erfolg? Überträgt man die Erfolgsebene auf das CCH-Testtraining, dann ist mit erwünschtem Erfolg in diesem Fall ein gutes TMS-Ergebnis gemeint.

- Wie haben die Testteilnehmer abgeschlossen?
- Wie sahen die Ergebnisse in den einzelnen Untertests aus?
- Wie sahen die Resultate gerade in den kurzfristig trainierbaren Untertests aus?

7.2 Datenerhebung, Stichprobe und Auswertungsverfahren

7.2.1 Datenerhebung

Die Datenerhebung zur Bewertung des Testtrainings nach dem „Vier-Ebenen-Modell“ von Kirkpatrick erfolgte anhand von Fragebögen. Zu Beginn des Seminars wurden ein Fragebogen zur Erhebung von allgemeinen Daten, wie Abiturdurchschnittsnote, Studienfachwunsch etc. und die beiden Fragebögen zur Prüfungserwartung und zur Einschätzung der Faktoren, die wichtig sind für ein gutes Abschneiden im TMS, ausgefüllt. Die beiden letzterwähnten Fragebögen wurden zu Beginn der Generalprobe ein weiteres Mal bearbeitet. Vor der Generalprobe wurde zusätzlich noch der Fragebogen zur Generalprobe ausgefüllt, der Daten z. B. zur Art der Vorbereitung, die zusätzlich zum CCH-Testtraining stattgefunden hatte, erfasste. Nach dem Erhalt der Testergebnisse erhielten alle Trainingskursteilnehmer einen Evaluationsbogen. Hier sollten die Testergebnisse im Einzelnen eingetragen werden und es sollten Fragen zum Testtraining beantwortet werden⁸.

7.2.2 Stichprobe

Insgesamt haben in den Jahren 1986 bis 1995 $N = 6.579$ Personen an einem CCH-Testtraining teilgenommen. Davon waren 2.543 (38,7 Prozent) Männer, 3.232 (49,1 Prozent) Frauen und 804 (12,2 Prozent) machten keine Angaben zum Geschlecht. Die Verteilung der Abiturdurchschnittsnote bei den Seminarteilnehmern ist in Tabelle 7.1 dargestellt.

⁸ Die Fragebögen werden genauer in Kapitel 8.3 beschrieben. Zudem befinden sich alle Fragebögen im Anhang

Tabelle 7.1: Verteilung der Abiturdurchschnittsnoten der Teilnehmer, die von 1986 bis 1995 ein Testtraining für den TMS bei der CCH-Laufbahnberatung besucht haben.

	Häufigkeit	Prozent	kumulierte Prozente
1,00 bis 1,40	290	4,4 %	4,4 %
1,41 bis 1,80	851	12,9 %	17,3 %
1,81 bis 2,20	1503	22,8 %	40,1 %
2,21 bis 2,60	1564	23,8 %	63,9 %
2,61 bis 3,00	970	14,7 %	78,6 %
3,01 bis 3,41	337	5,1 %	83,7 %
3,41 und mehr	74	1,1 %	84,8 %
keine Angabe	990	15,1 %	100 %
Gesamt	6579	100 %	

Den Evaluationsbogen haben 3.138 (47,7 Prozent) Trainingsteilnehmer zurückgeschickt. Nicht zurückgeschickt haben ihn 3.441 (52,3 Prozent) Seminarbesucher. Betrachtet man den Aspekt der Rücksendung bzw. Nicht-Rücksendung hinsichtlich des Geschlechts, dann stellt sich die Verteilung wie in Tabelle 7.2 berichtet und noch einmal graphisch in Bild 7.1 verdeutlich dar.

Tabelle 7.2: Geschlechterverteilung hinsichtlich Rücksendung bzw. Nicht-Rücksendung

Geschlecht	Rücksendung	Nicht-Rücksendung	Gesamt
Männer	1277 (50,2 %)	1266 (49,8 %)	2543 (100 %)
Frauen	1707 (52,8 %)	1525 (47,2 %)	3232 (100 %)
keine Angaben	154 (19,2 %)	650 (80,8 %)	804 (100 %)
Gesamt	3138	3441	6579

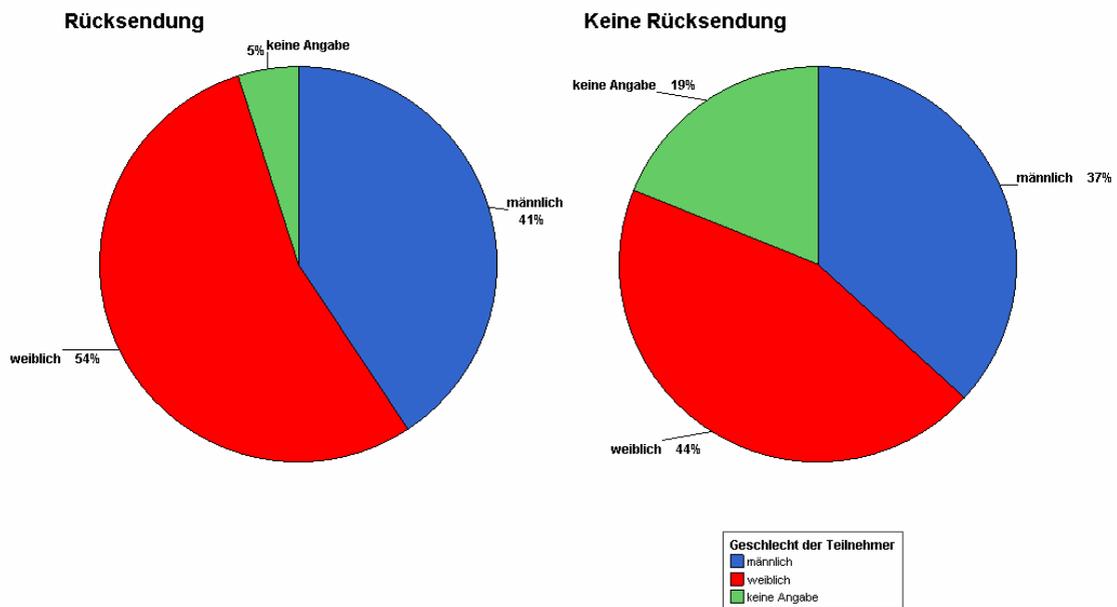


Bild 7.1: Darstellung der Verteilung von Männer, Frauen und keine Angaben hinsichtlich der Rücksendung

Die Verteilung der Abiturnoten im Vergleich zwischen den beiden Gruppen „Rücksendung des Evaluationsbogens“ und „Nicht-Rücksendung des Evaluationsbogen“ wird in Bild 7.2 präsentiert.

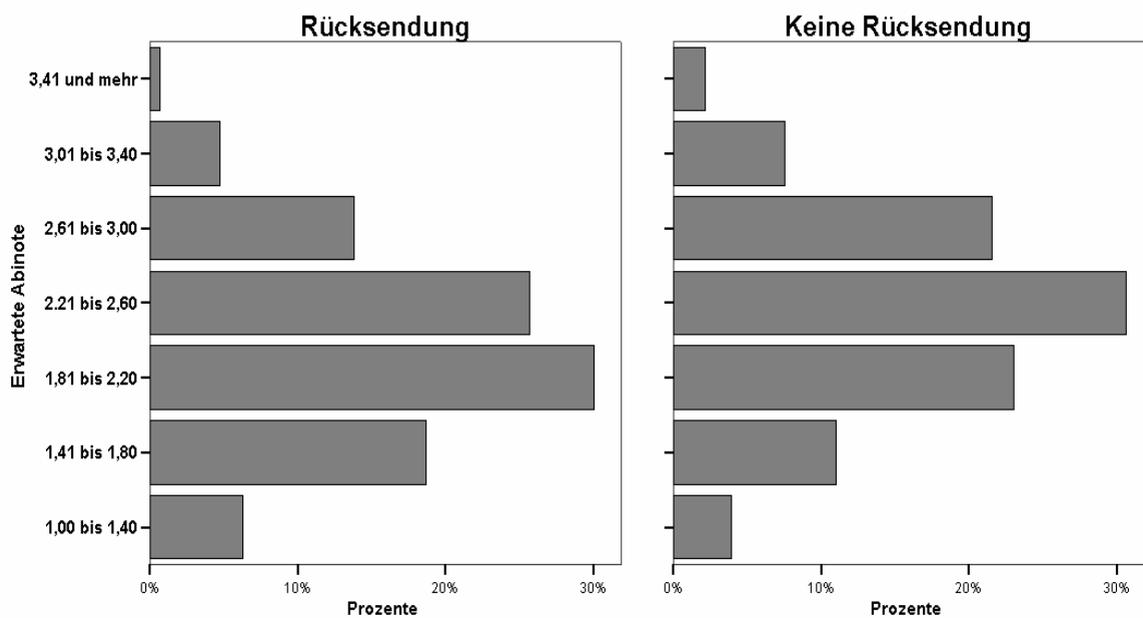


Bild 7.2: Verteilung der Abiturdurchschnittsnoten auf die Rücksender bzw. Nicht-Rücksender

Vergleicht man die Teilnehmer, die ihren Evaluationsbogen zurückgeschickt haben, mit denjenigen, die ihren Bogen nicht zurückgesandt haben, dann kann man eine leichte Tendenz erkennen. Die Seminarbesucher mit den schlechteren Abiturnoten haben den Evaluationsbogen nicht so häufig zurückgeschickt, wie diejenigen mit den guten Abiturnoten. Bei der Rücksendung lagen die Männer und Frauen fast gleich auf. Teilnehmer, von denen man keine Angaben zum Geschlecht aus dem Seminar oder der Generalprobe hatte, haben meistens auch den Evaluationsbogen nicht zurückgesandt.

7.2.3 Auswertungsverfahren

Die Daten zur Evaluations- bzw. Transferevaluation werden mit Hilfe der deskriptiven Statistik dargestellt.

7.3 Darstellung der Befunde

Ebene 1: Reaktionsebene

Auf die Frage „Alles in allem beurteile ich das Trainerverhalten als vollkommen kompetent, weitgehend kompetent, ein wenig kompetent oder gar nicht kompetent“ haben die Probanden wie in Bild 7.3 geantwortet. Das Antwortverhalten von Frauen und Männern wich nicht voneinander ab.

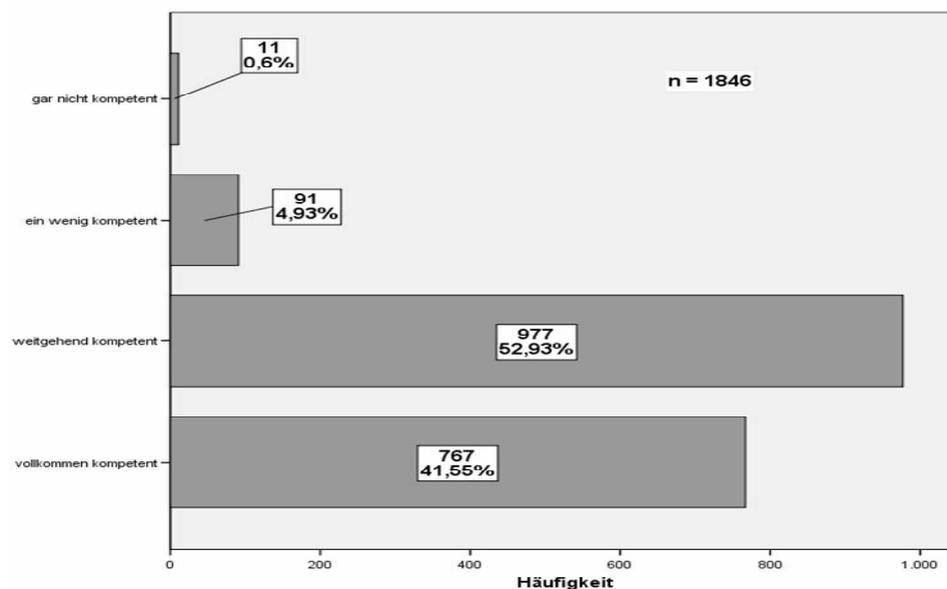


Bild 7.3: Frage nach der Kompetenz der Trainer

Auf die Frage „Alles in allem beurteile ich das Trainerverhalten als vollkommen engagiert, weitgehend engagiert, ein wenig engagiert oder gar nicht engagiert“ haben die Probanden wie in Bild 7.4 geantwortet. Das Antwortverhalten von Frauen und Männern wich auch hier nicht voneinander ab.

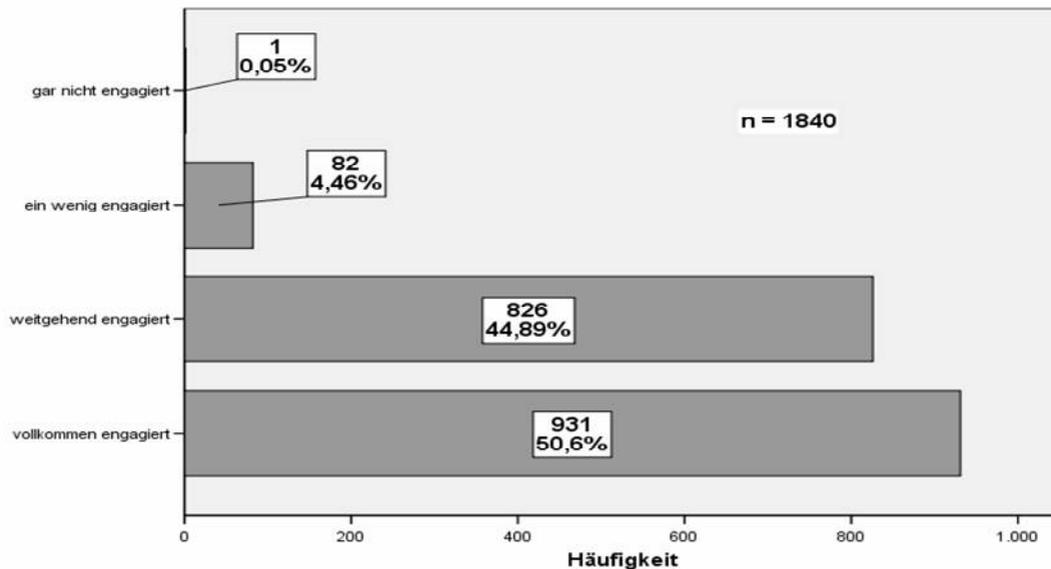


Bild 7.4: Frage nach dem Engagement der Trainer

Auf die Frage „Alles in allem beurteile ich das Trainerverhalten als vollkommen sympathisch, weitgehend sympathisch, ein wenig sympathisch oder gar nicht sympathisch“ haben die Probanden wie in Bild 7.5 geantwortet. Bei dieser Frage haben Männer und Frauen ebenfalls relativ gesehen die gleichen Antworten gegeben.

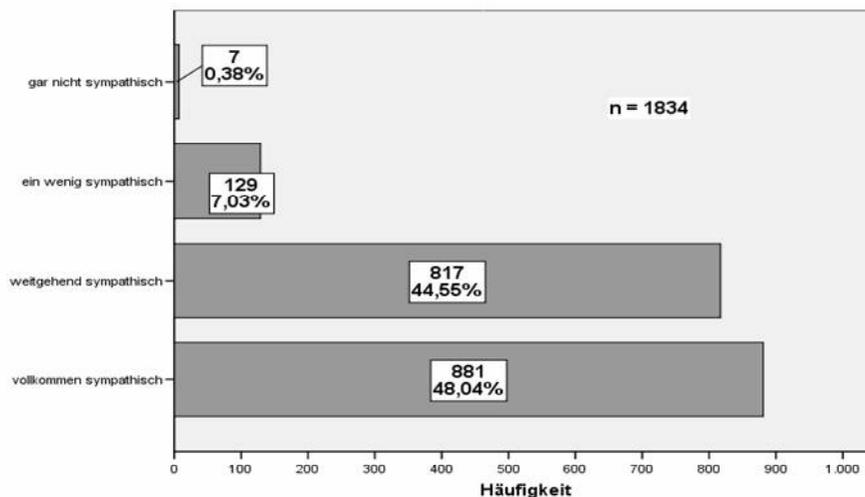


Bild 7.5: Frage, ob die Trainer als sympathisch empfunden wurden

Ebene 2: Lernebene

Die Frage, ob sich die Einstellungen zu Prüfungen bei den Teilnehmern durch die Teilnahme am Seminar signifikant verändert haben, wird anhand einiger ausgesuchter interessanter Fragestellungen durch ein Balkendiagramm abgebildet. Die selektierten Fragen sind dem Fragebogen zur allgemeinen Prüfungserwartung entnommen worden. Verglichen werden die Mittelwerte aus der Befragung vor dem Seminar mit den Mittelwerten aus der Befragung vor der Generalprobe. In den Gruppen, „Befragung vor dem Seminar“ und „Befragung vor der Generalprobe“, sind 984 Probanden in beiden Gruppen enthalten. Folgende Fragen werden in Bild 7.6 dargestellt:

Frage 1: Bedeutung „Dass ich ein möglichst gutes Resultat im TMS erziele, ist für mich von höchster Bedeutung“.

Frage 2: Ungutes Gefühl „Ich habe ein sehr ungutes Gefühl, wenn ich bedenke, dass ich den TMS nur ein einziges Mal absolvieren kann“.

Frage 3: Unsicherheit „Ich bin recht unsicher über mein Abschneiden im TMS“.

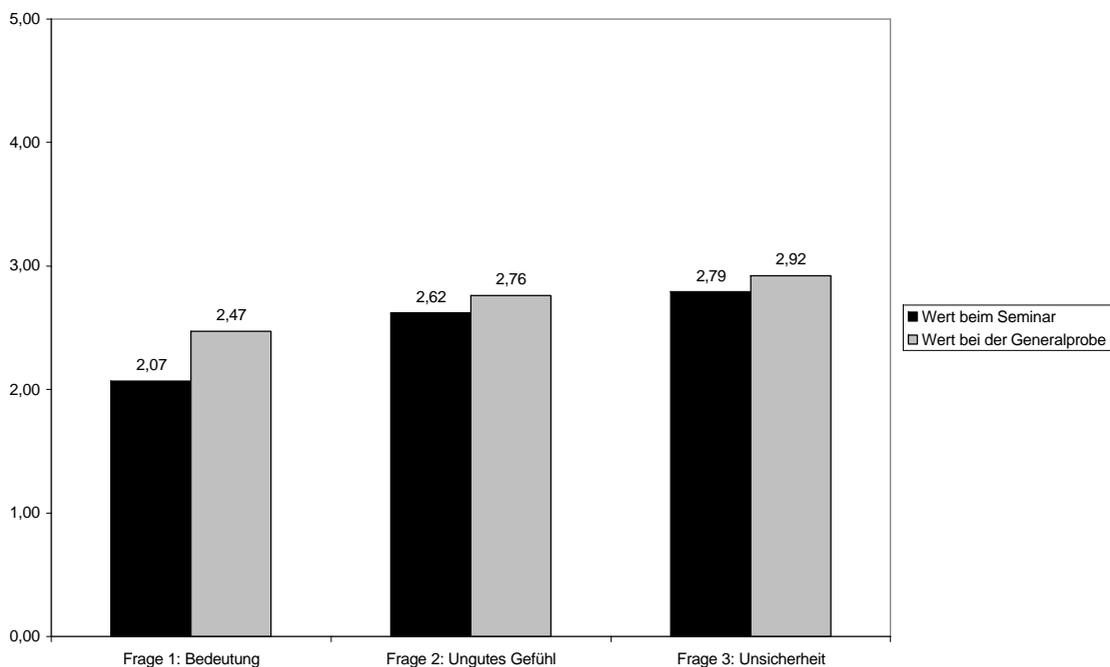


Bild 7.6: Die Skala der Fragen 1-6 reichte von 1 (stimmt genau), 2 (stimmt), 3 (stimmt in etwa), 4 (stimmt kaum) bis 5 (stimmt gar nicht). In der Abbildung sind Mittelwerte dargestellt, die berechnet wurden aus dem Antwortverhalten der Testtrainingsteilnehmer.

In Bild 7.7 ist die Einstellungsänderung in Bezug auf Prüfungen allgemein an nachfolgend aufgelisteten Fragen dargestellt:

Frage 1: Intensität der Vorbereitung „Wie gut man in Prüfungen abschneidet, hängt davon ab, wie intensiv man sich vorbereitet“.

Frage 2: Anstrengung „Gute Leistungen sind das Resultat eigener Anstrengungen“.

Frage 3: Gründlichkeit „Auf wichtigen Prüfungen bereite ich mich gründlich vor“.

Frage 4: Gelassenheit „Auch wichtigen Prüfungen sehe ich gelassen entgegen“.

Frage 5: Krafteinsatz „Prüfungen gut zu bestehen hängt nicht nur vom eigenen Wissen und Können ab, sondern auch davon, ob man seine Fähigkeiten zum richtigen Zeitpunkt voll einsetzen kann“.

Frage 6: Vertrautheit „Wenn ich weiß, was mich erwartet, dann kann ich einer Prüfung relativ gelassen entgegen sehen“.

Frage 7: Mobilisierung „Was immer ich tun kann, um mich optimal auf Prüfungen vorzubereiten, das tue ich in der Regel auch“.

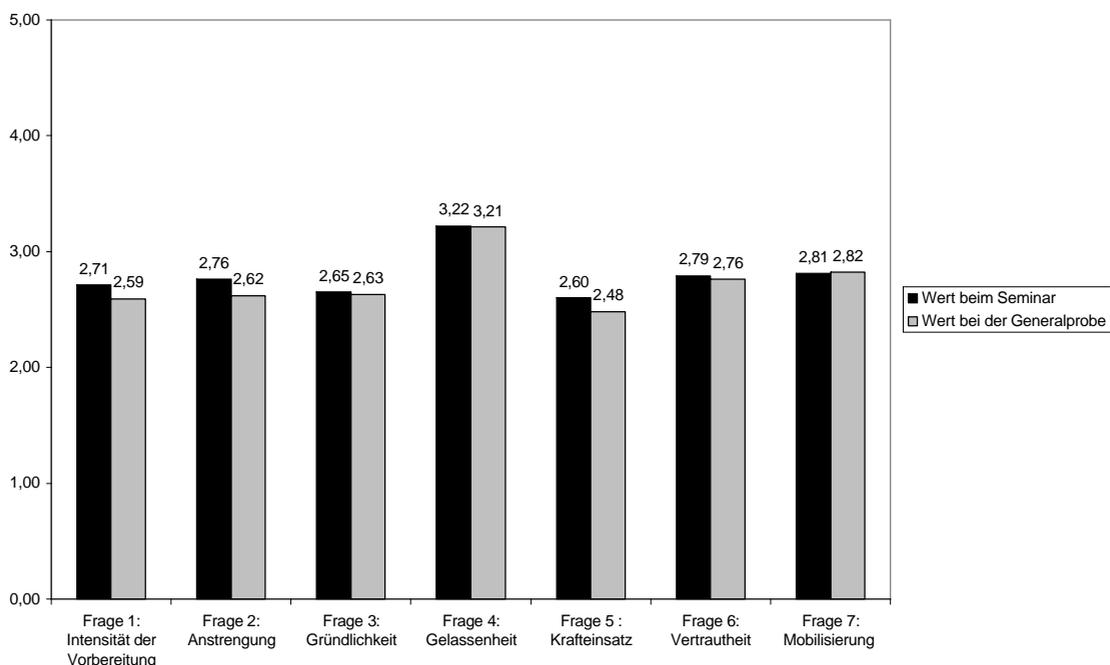


Bild 7.7: Die Skala der Fragen 1-6 reichte von 1 (stimmt genau), 2 (stimmt), 3 (stimmt in etwa), 4 (stimmt kaum) bis 5 (stimmt gar nicht). In der Abbildung sind Mittelwerte dargestellt, die berechnet wurden aus dem Antwortverhalten der Testtrainingsteilnehmer.

In Bild 7.8 wird dargestellt, ob die Trainingsteilnehmer ihre Einstellung von der ersten Befragung, die gleich zu Beginn des Seminars stattfand, bezüglich ihrer Zuversichtlichkeit hinsichtlich ihrer Zulassungschancen zum Medizinstudium bis zur Generalprobe, hier wurde die Befragung wieder gleich zu Beginn durchgeführt, geändert hatten.

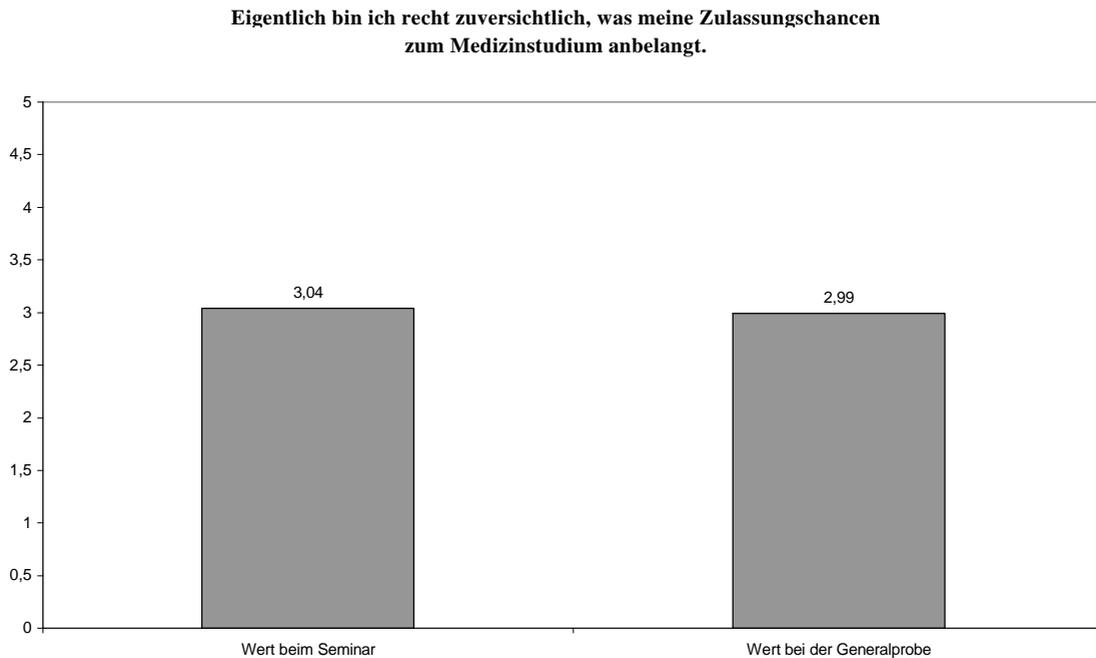


Bild 7.8: Die Skala der Fragen 1-6 reichte von 1 (stimmt genau), 2 (stimmt), 3 (stimmt in etwa), 4 (stimmt kaum) bis 5 (stimmt gar nicht). In der Abbildung sind Mittelwerte dargestellt, die berechnet wurden aus dem Antwortverhalten der Testtrainingsteilnehmer.

Interessant war die Frage, ob die Teilnehmer bestimmte Faktoren nach dem Besuch des Testtrainingsseminars der CCH-Laufbahnberatung in ihrer Wichtigkeit für ein gutes Abschneiden im TMS anders einschätzen. In Bild 7.9 werden die Mittelwerte der beiden Befragungen, einmal vor dem Seminar und einmal vor der Generalprobe, einander gegenübergestellt.

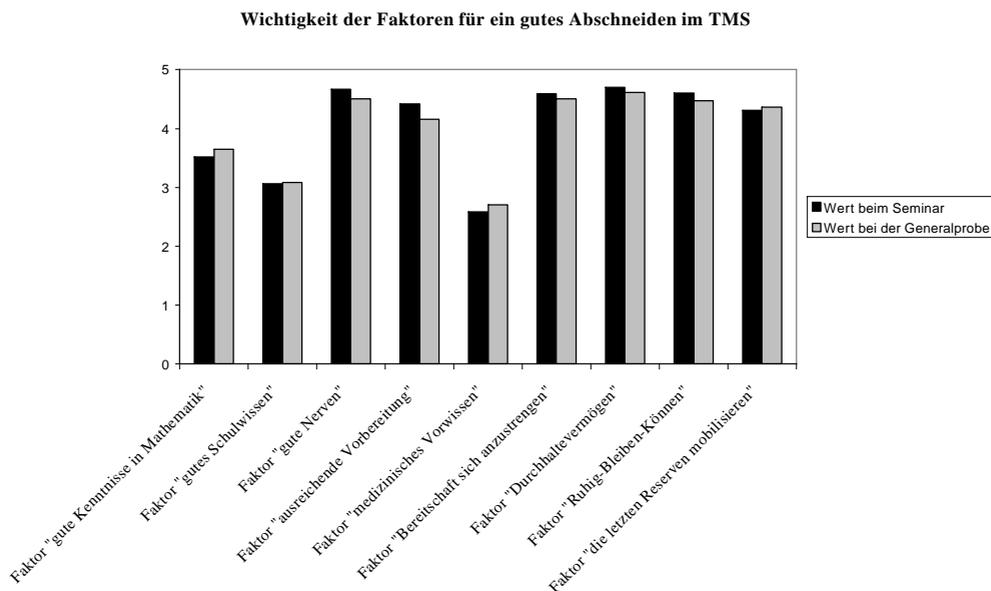


Bild 7.9: Die Skala für die Einschätzung der Wichtigkeit bestimmter Faktoren für ein gutes Abschneiden im TMS reichte von 1 (absolut unwichtig), 2 (unwichtig), 3 (mäßig wichtig), 4 (wichtig) bis 5 (absolut wichtig). In der Abbildung sind Mittelwerte dargestellt, die berechnet wurden aus dem Antwortverhalten der Testtrainingsteilnehmer.

In Bild 7.10 werden die Aussagen aus den beiden Befragungen gegenübergestellt, die die Teilnehmer bezüglich ihrer Ausprägung der ausgewählten Faktoren gemacht haben.

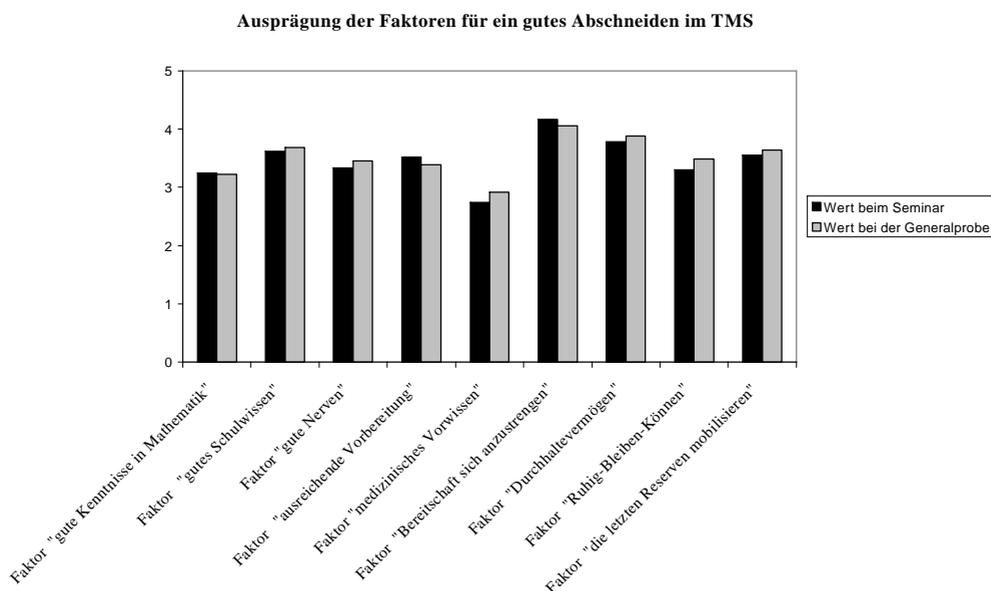


Bild 7.10: Die Skala für die Einschätzung der Wichtigkeit bestimmter Faktoren für ein gutes Abschneiden im TMS reichte von 1 (gar nicht ausgeprägt), 2 (kaum ausgeprägt), 3 (mäßig ausgeprägt), 4 (gut ausgeprägt) bis 5 (sehr gut ausgeprägt). In der Abbildung sind Mittelwerte dargestellt, die berechnet wurden aus dem Antwortverhalten der Testtrainingsteilnehmer.

Ebene 3: Verhaltensebene

Ob die Trainingskursteilnehmer das Gelernte auf andere Situationen übertragen können, wird über die Aussagen beobachtet, die die Teilnehmer bei der Beantwortung des Evaluationsbogens getroffen haben. In Tabelle 7.3 erhält man einen Überblick, ob der Lerntransfer nach Meinung der Teilnehmer funktioniert hat.

Tabelle 7.3: Verteilung der Antworten auf die Fragen im Evaluationsbogen hinsichtlich der Transferierung des Gelernten

	sehr nützlich	nützlich	weniger nützlich	Gesamt
Logik	1339 (42,6 %)	1598 (50,9%)	204 (6,5 %)	3141 (100 %)
Zulassungsfragen	406 (21,7 %)	1126 (60,3 %)	335 (17,9 %)	1867 (100 %)
Simulation	2444 (77,5 %)	637 (20,2 %)	71 (2,3 %)	3152 (100 %)
Strategien	909 (48,8 %)	825 (44,3 %)	130 (7 %)	1864 (100 %)
Arbeitstechniken	677 (36,3 %)	951 (51 %)	236 (12,7 %)	1864 (100 %)
Konzentrations- übungen	256 (13,7 %)	766 (41,1 %)	841 (45,1 %)	1863 (100 %)
Tips	473 (25,5 %)	1160 (62,5 %)	224 (12,1 %)	1857 (100 %)
Seminarmaterial	1119 (60,4 %)	670 (36,1 %)	65 (3,5 %)	1854 (100 %)

Anmerkungen: Die aus dem Evaluationsbogen ausgewerteten Fragen sind in der Tabelle abgekürzt. Den ursprünglichen Wortlaut kann man im Evaluationsbogen nachlesen, der sich im Anhang befindet.

In Bild 7.11 ist der Zusammenhang aus der Tabelle noch einmal graphisch dargestellt.

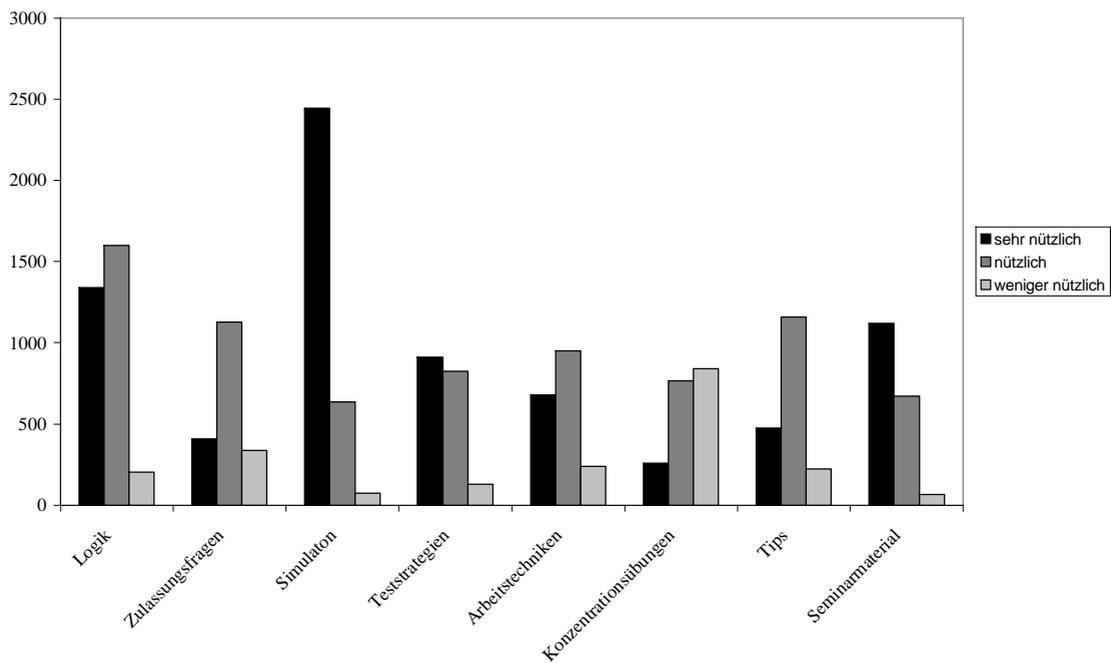


Bild 7.11: Darstellung der Nützlichkeit der Testtrainingselemente für die Teilnehmer

Die Teilnehmer wurden auf dem Evaluationsbogen gebeten darüber Auskunft zu geben, ob sie meinen, dass sie vom Testtraining der CCH-Laufbahnberatung profitiert hätten. Die Beantwortung der Frage ist in dem nachfolgenden Bild 7.12 dargestellt.

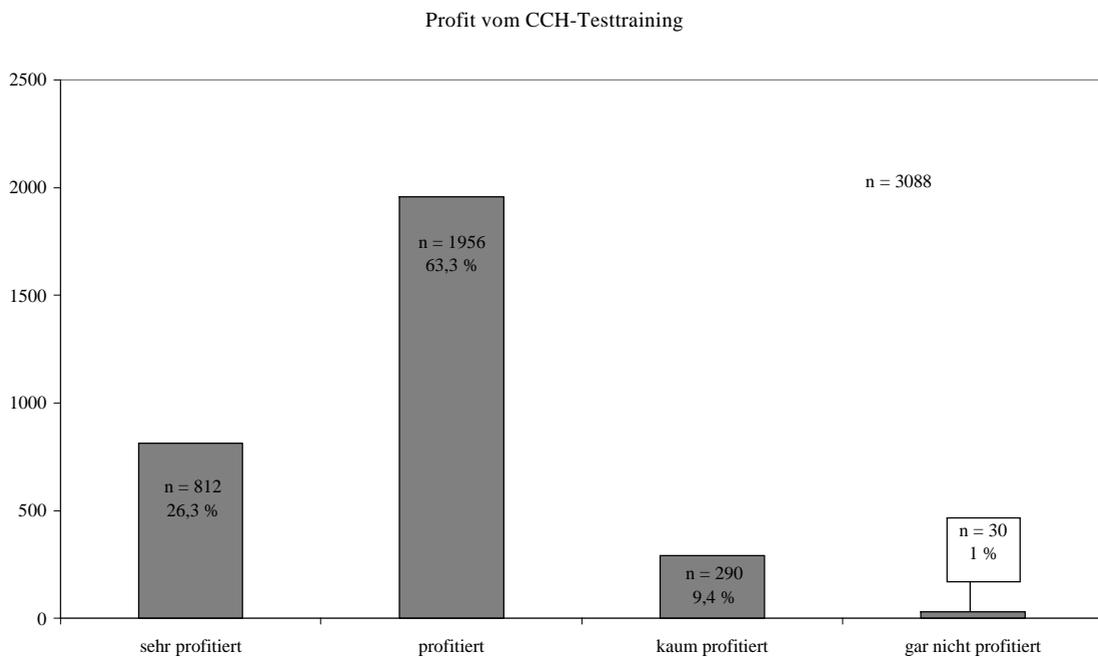


Bild 7.12: Darstellung der Verteilung, wie die Teilnehmer ihren Profit vom CCH-Testtraining einschätzen

Vor der Generalprobe fand eine Befragung statt. Dabei wurden die Teilnehmer explizit danach gefragt, ob sie meinen, dass sie durch die Vorbereitung auf den TMS das Gelernte auch in Hinsicht auf andere Prüfungen profitieren. In Bild 7.13 ist die Antwortverteilung dargestellt.

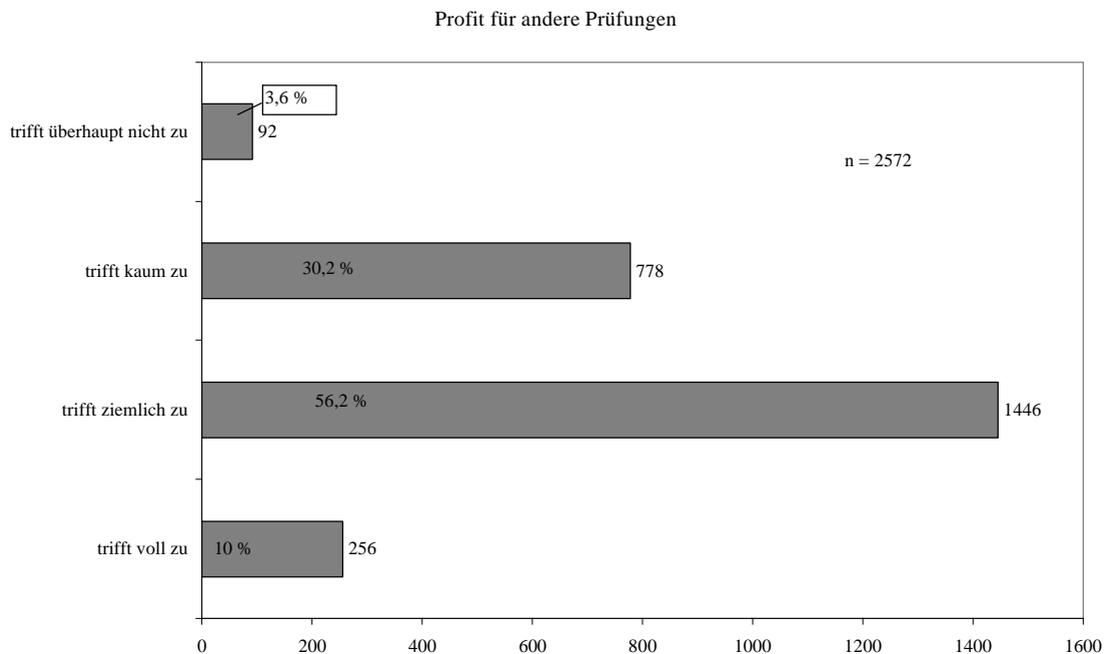


Bild 7.13: Antwortverteilung auf die Frage, ob folgende Aussage „Durch die Vorbereitung auf den TMS kann ich in vielerlei Hinsicht auch für andere Prüfungen profitieren“ auf die Probanden zutrifft.

Ebene 4: Erfolgsebene

In diesem Abschnitt werden die Ergebnisse der Trainingsteilnehmer betrachtet. Dabei wird im Vergleich die Standardnormalverteilung der TMS-Teilnehmer betrachtet. Wie eingangs bei der Stichprobenbeschreibung erwähnt, haben nicht alle Teilnehmer den Evaluationsbogen zurückgesandt, so dass nicht von allen Teilnehmern TMS-Werte vorliegen. Die Zahl der TMS-Daten schwankt immer leicht, da die Bögen zum Teil nicht vollständig und korrekt ausgefüllt wurden. Manchmal war nur der Prozentrang eingetragen oder einzelne Werte für die verschiedenen Untertests.

Bild 7.14 zeigt die Standardverteilung aller Testteilnehmer. Der Mittelwert liegt bei der Gruppe der Testnehmenden bei 100 Punkten.

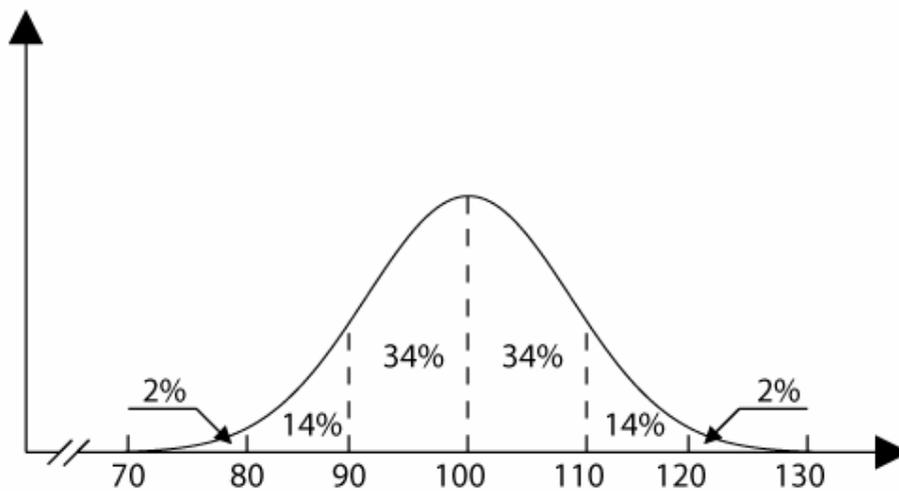


Bild 7.14: Verteilung der TMS-Standardwerte aller Teilnehmer am TMS (Quelle: <http://www.tms-info.org/index.php?ID=100>)

Vergleicht man die Verteilung der Testwerte aller Testteilnehmer mit der Verteilung der Gruppe der Testtrainingsteilnehmer, dargestellt in Bild 7.15, dann wird deutlich, dass die Seminarteilnehmer zu der Gruppe gehören, die überdurchschnittlich abgeschnitten haben. Der Mittelwert der Trainingsteilnehmer lag bei 108,34 Punkten im Vergleich zum Mittelwert aller TMS-Teilnehmer, der bei 100 Punkten liegt.

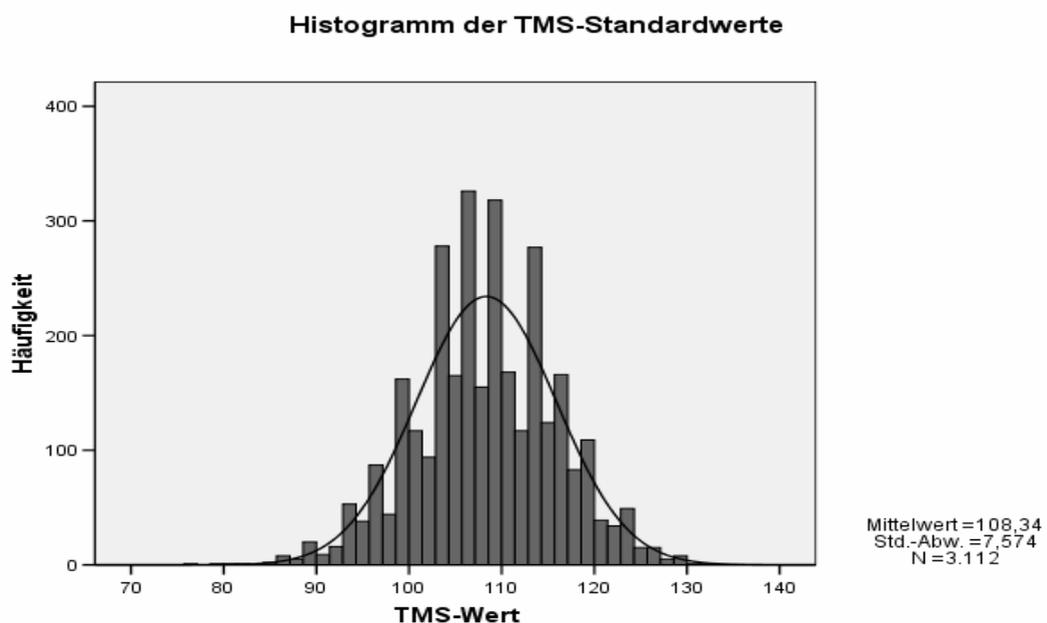


Bild 7.15: Verteilung der TMS-Standardwerte der Teilnehmer am Trainingskurs von der CCH-Laufbahnberatung

Bild 7.16 gibt einen Überblick über die Verteilung der TMS-Wert der Testtrainingsteilnehmer. Zur besseren Übersicht wurden die TMS-Werte in Kategorien zusammengefasst.

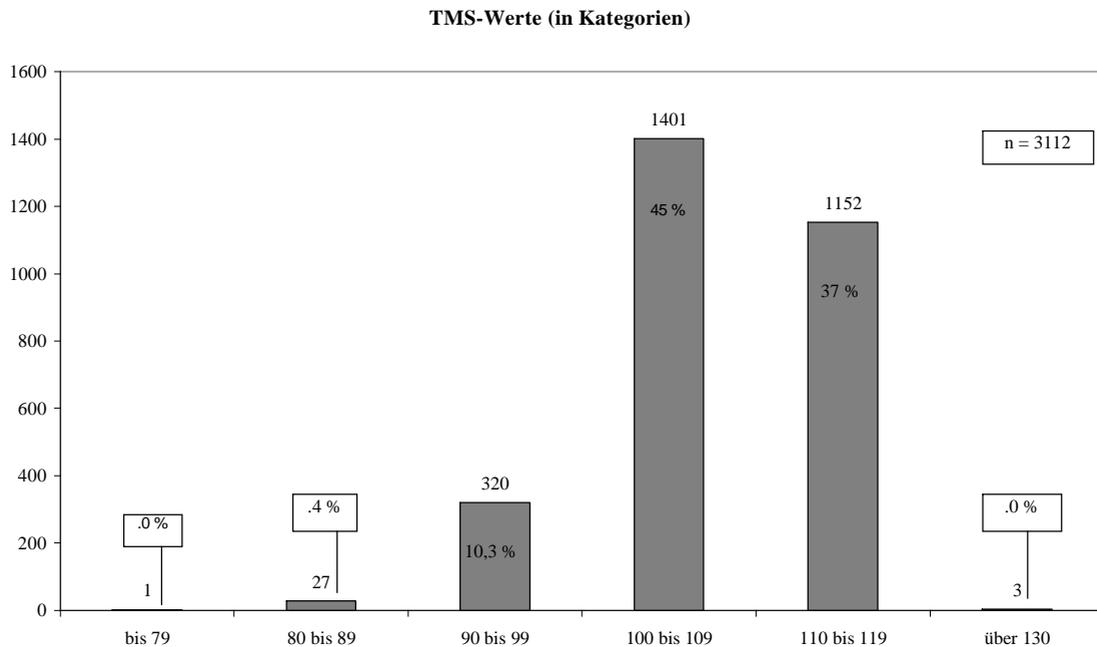


Bild 7.16: Die von den Testtrainingsteilnehmern erreichten TMS-Werte (in Kategorien aufgeteilt).

Da es sich um beim CCH-Testtraining um ein kurzfristiges Testtraining handelte, ist es bei der Evaluation der Erfolgsebene wichtig, die Ergebnisse der einzelnen Untertests zu betrachten. Die Trainingsteilnehmer haben sich gerade in den kurzfristig trainierbaren Untertests hohe Erfolgsquoten, also überdurchschnittliche Leistungen, im TMS versprochen. Tabelle 7.4 zeigt die Mittelwerte und Standardabweichungen für die neun Untertests des TMS.

Tabelle 7.4: Übersicht über die Mittelwerte und der Standardabweichungen der neun Untertests

Skalen ^a	n	M	SD	maximal erreichbare Punkte
MZ	3096	12,37	2,915	20
MNGV	3092	11,45	3,316	20
SF	3095	15,53	2,994	20
QFP	3095	10,68	3,667	20
KONZ	3100	17,80	2,379	20
TV	3094	9,48	3,455	18
FIG	3093	13,68	3,353	20
FAK	3095	12,35	3,915	20
DT	3095	11,17	3,297	20

Anmerkungen: ^aSkalen: MZ: Muster zuordnen, MNGV: Medizinisch-naturwissenschaftliches Grundverständnis, SF: Schlauchfiguren, QFP: Quantitative und formale Probleme, KONZ: Konzentriertes und sorgfältiges Arbeiten, TV: Textverständnis, FIG: Figuren lernen, FAK: Fakten lernen, DT: Diagramme und Tabellen

Als kurzfristig trainierbare Untertests gelten: „Muster zuordnen“, „Schlauchfiguren“, „Konzentriertes und sorgfältiges Arbeiten“, „Figuren lernen“ und „Fakten lernen“. Betrachtet man die Mittelwerte, dann liegen die Mittelwerte der Untertests der Testtrainingsteilnehmer, die den Evaluationsbogen zurückgesandt haben, bei jedem Untertest über dem Durchschnitt (Der Durchschnittswert des Untertests Textverständnis liegt bei 9 Punkten). Bei den beiden Untertests „Muster zuordnen“ und „Konzentriertes und sorgfältiges Arbeiten“ lagen die Testteilnehmer sogar weit über dem Durchschnitt.

In Bild 7.17 sind die Mittelwerte der TMS-Standardwerte für vier Gruppen, Männer, Frauen, gute Abiturienten und schlechte Abiturienten, dargestellt.

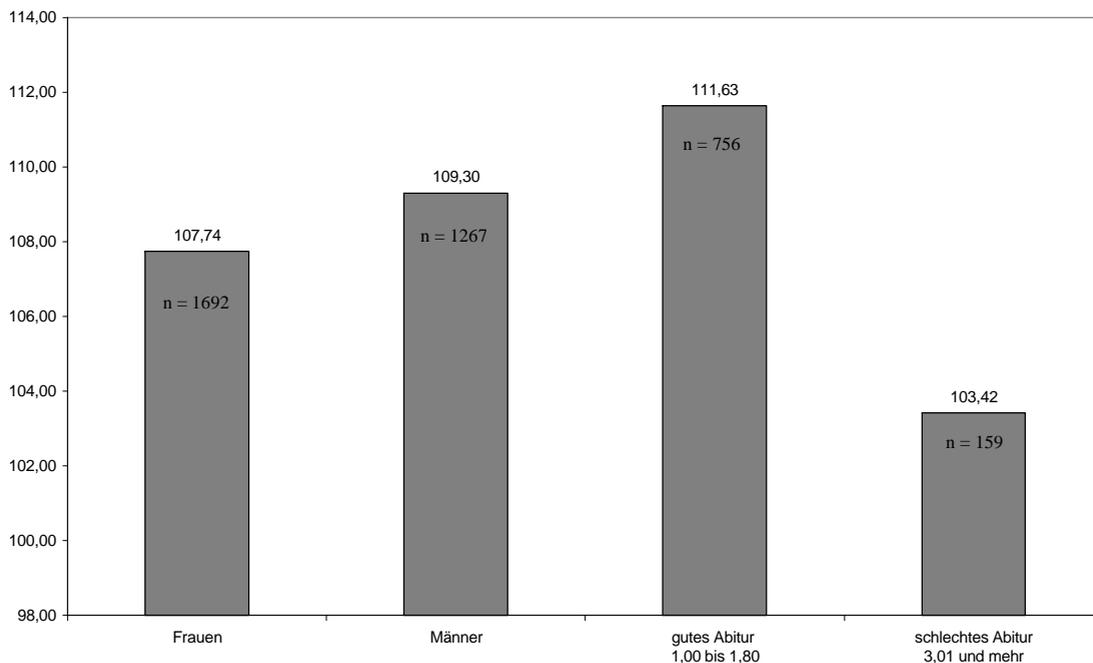


Bild 7.17: Verteilung der durchschnittlichen TMS-Resultate bei ausgewählten Untergruppen

7.4 Schlussfolgerungen aus den Ergebnissen der Evaluation

Die Frage, ob sich die Seminarteilnehmer gut betreut gefühlt haben, kann bejaht werden. Die meisten Besucher des CCH-Laufbahntrainings haben ihre Trainer als kompetent, engagiert und sympathisch beurteilt. Damit ist auf jeden Fall eine Grundvoraussetzung für ein gutes Lernklima gegeben gewesen.

Der Lernerfolg im Laufe des Seminars war durch die vorliegenden Daten nicht ohne Weiters überprüfbar. Mit den Daten waren keine Vorher-Nachher-Messungen hinsichtlich der kognitiven Leistungen durchführbar. Es konnte allerdings anhand der zu zwei verschiedenen Zeitpunkten verteilten Fragebögen eine Vorher-Nachher-Messung in Bezug auf eventuelle Einstellungsänderungen vorgenommen werden. Die Messung erfolgt allerdings rein deskriptiv. Ziel des Testtrainings war es unter anderem, dass die Seminarteilnehmer die Unsicherheit gegenüber dem TMS verlieren. Die ängstlichen Gefühle und Einstellungen gegenüber dem TMS konnten leicht zum positiven verändert werden, d.h. die Trainingsteilnehmer waren nicht mehr so unsicher wie zu Beginn des Seminars über ihr Abschneiden beim TMS. Allerdings kann man hier nur von leichten Tendenzen sprechen. Auch die Einstellungen gegenüber Prüfungen konnte etwas entspannt werden, aber nur sehr geringfügig. Eine tendenzielle Verschiebung der

Wichtigkeit von bestimmten Faktoren für ein gutes Abschneiden im TMS in spezielle vorhersagbare Richtungen war nicht zu erkennen. Auch bei der Ausprägung der Faktoren für ein gutes Abschneiden gab es bei den beobachtbaren Veränderungen kein einheitliches Bild.

Auf der Verhaltensebene beurteilten die Seminarteilnehmer die Nützlichkeit des Trainingsprogramms, nachdem sie das Gelernte angewandt hatten. Im Großen und Ganzen kann man von einer positive Bilanz zu den praktischen Anwendungsbereichen und Lernbereichen sprechen. Die Simulation empfanden fast alle, die den Evaluationsbogen zurückgeschickt hatten, als sehr nützlich. Auch die Lernbereiche Logik und Bearbeitungstrategien wurden positiv beurteilt. Das Seminarmaterial war auch erfolgreich. Es gab allerdings Bereiche, die weniger gut ankamen: Konzentrations- und Entspannungsübungen, Tips zur weiteren Vorbereitung und der Bereich zur Testauswertung und zu Zulassungsfragen. Die Teilnehmer waren scheinbar mehr an der konkreten Testübung und Testbewältigung interessiert.

Der wichtigste Punkt der Evaluation war die Frage, ob das Testtraining für die Teilnehmer ein Erfolg war. Gemessen werden konnte das am tatsächlichen Testergebnis und an den Ergebnissen zu den neun Untertests. Da viele der Befragten den Evaluationsbogen zurückgesandt haben, war es hier möglich, diesen so wichtigen Punkt zur Messung des Erfolgs zu beurteilen. Die Trainingskursteilnehmer haben eindeutig über dem Durchschnitt im Vergleich zur Gesamtgruppe der Testteilnehmer abgeschnitten. Der Mittelwert der Trainingsteilnehmer lag fast eine Standardabweichung über dem Mittelwert der Gesamtgruppe der Testteilnehmer. Vergleicht man die Mittelwerte einzelner Untergruppen, dann lagen sogar die schlechten Abiturienten mit ihrem Mittelwert knapp über dem Durchschnitt. Ob dieses kurzfristige Training erfolgreich war, ist aber auch anhand der kurzfristig trainierbaren Untertests beobachtbar. Die Teilnehmer lagen ebenfalls bei allen kurzfristig trainierbaren Untertest deutlich über dem Durchschnitt und zum Teil sogar weit darüber.

Abschließend kann zusammengefasst werden, dass das Testtraining der CCH-Laufbahnberatung als sehr nützlich bewertet werden kann. Die Trainingsteilnehmer haben überdurchschnittlich gut abgeschnitten. Eine langfristige Einstellungsänderung

gegenüber Prüfungen wird in diesem Trainingskurs eher nicht erreicht worden sein, aber die Teilnehmer haben sich zufrieden über ihre Trainer geäußert. Die Zufriedenheit der Kursteilnehmer ist auf jeden Fall eine gute Voraussetzung für den gewünschten Lernerfolg durch eine Trainingsmaßnahme.

Teilnehmer hinsichtlich ausgewählter Kriterien in einer Fallübersicht

Abiturnote	Profit ^a	Simulation ^b	Kompetenz ^c	Engagement ^e	Sympathie ^e	Aussage ^d	Studienwunsch ^e
1,81 bis 2,20	profitiert	sehr nützlich	weitgehend kompetent	weitgehend engagiert	weitgehend sympathisch	trifft kaum zu	mehreren Jahren
1,81 bis 2,20	sehr profitiert	nützlich	vollkommen kompetent	vollkommen engagiert	vollkommen sympathisch	trifft kaum zu	über einem Jahr
1,41 bis 1,80	profitiert	nützlich	weitgehend kompetent	vollkommen engagiert	vollkommen sympathisch	trifft voll zu	mehreren Jahren
3,01 bis 3,40	profitiert	nützlich	vollkommen kompetent	vollkommen engagiert	weitgehend sympathisch	---	mehreren Jahren
1,81 bis 2,20	profitiert	sehr nützlich	gar nicht kompetent	ein wenig engagiert	ein wenig sympathisch	trifft ziemlich zu	über einem Jahr

zusamt habe ich vom CCH-Testtraining...“, ^bSimulation: „Im einzelnen fand ich die Simulation der Testsituation...“, ^cKompetenz:
 das Trainerverhalten als...“, ^dAussage: „Inwieweit trifft die folgende Aussage auf Sie zu? „Durch die Vorbereitung auf den TMS
 (z. B. Arbeitstechnik, Zeitmanagement, Blitzentspannungstechniken) auch für andere Prüfungen profitieren.“, ^eStudienwunsch:
 ich seit...“

8 Zusätzliche Analyse ausgewählter Aspekte

Die nachfolgenden Untersuchungen und Berechnungen haben zur Grundlage die Daten, die die CCH-Laufbahnberatung während ihrer Vorbereitungsseminare auf den TMS erhoben haben. Wie sich schon durch die Literaturrecherche in Kapitel 4 und 5 zeigte, ist es aufgrund der verschiedenen Randbedingungen der Vorbereitungsmaßnahmen kaum möglich, verschiedene Testtrainings miteinander zu vergleichen. In dieser Arbeit werden also nur die Teilnehmer des CCH-Laufbahnberatungsseminars betrachtet, die alle am Ende den gleichen Test, den TMS, abgelegt haben.

Im theoretischen Teil der Arbeit wurde deutlich, dass es nicht möglich ist, anhand eines Pre- und Posttest im Kontrollgruppendesign zu schlussfolgern, dass eine eventuelle Leistungssteigerung im Posttest allein auf ein Treatment, in diesem Fall ein Testtraining, zurückzuführen ist. Vielmehr spielen viele andere Faktoren beim Zustandekommen einer Leistung eine Rolle. In dieser Arbeit wird versucht auf diese anderen Faktoren einzugehen, die bei der Testleistung ebenfalls relevant sein könnten. Da die Individuen verschieden auf das Coaching ansprechen, werden in der empirischen Überprüfung verschiedene Extremgruppen bezüglich ihrer Gesamtestleistung bzw. ihrer Leistung in den verschiedenen Untertests miteinander verglichen. Dies ist ein Versuch, der Schlussfolgerung aus der Literaturrecherche in Kapitel 4 und 5 gerecht zu werden. Denn nicht alle Seminarteilnehmer bringen die gleichen Voraussetzungen mit, so dass man nicht davon ausgehen kann, dass das Training auf alle gleich wirkt. Der empirische Teil dieser Arbeit wird sich also mit der Darstellung dieser Unterschiedlichkeit in den Merkmalsausprägungen verschiedener Faktoren beziehen. Interessant sind hierbei die Faktoren: Abiturdurchschnittsnote, Vorbereitung, Vorwissen, Geschlecht, Belastbarkeit in Prüfungssituationen und Prüfungsangst. Bezüglich der Ausprägung in diesen aufschlussreichen Faktoren werden Extremgruppen gebildet. Die Frage ist, ob sich die Ausprägung der abhängigen Variable, z. B. der TMS-Gesamtestwert, signifikant unterscheidet, wenn die unabhängige Variable eine extrem hohe Ausprägung bzw. eine extrem niedrige Ausprägung aufweist.

8.1 Fragestellungen

Wie aus der Literaturanalyse deutlich hervorgeht, kann man davon ausgehen, dass es einen Zusammenhang gibt zwischen der Abiturnote und der Leistung im TMS. Es ist also anzunehmen, dass die Seminarteilnehmer mit einem exzellenten (1,00 bis 1,40) Abitur im Vergleich zu den Kursteilnehmern mit einem sehr schlechten (3,01 und mehr) Abiturdurchschnitt im TMS bessere Werte haben.

Wie schon Fay (1985) und auch Messick (1981) dargestellt haben, gibt es einen Zusammenhang zwischen der Vorbereitungszeit und der Testleistung. Personen, die sich neben dem Seminar zusätzlich extrem gut vorbereiten, werden einen höheren TMS-Wert haben als die Personen, die sich neben dem Seminar gar nicht oder kaum vorbereiten.

Fay (1985) vermutete ebenfalls eine Beziehung zwischen den Interessensschwerpunkten und der Testleistung. Es wird angenommen, dass Personen, die ihre Fähigkeiten in Mathematik als sehr ausgeprägt einschätzen in den mathematisch orientierten Untertests besser abschließen als Bewerber, die ihre Fähigkeiten in diesem Bereich als gering ausgeprägt einschätzen. Analog hierzu wird vermutet, dass Studienbewerber, die meinen, sie hätten ein sehr gut ausgeprägtes Vorwissen im medizinischen Bereich, besser in dem entsprechenden Untertest abschließen als Kandidaten, die der Auffassung sind, sie hätten kaum solche Qualitäten.

Ebenfalls Fay (1985) bezog sich auf den Faktor Geschlecht als Moderatoreffekt. Es wird untersucht, ob es einen Zusammenhang zwischen dem Faktor Geschlecht und dem TMS-Wert bzw. Werte in den verschiedenen Untertests gibt.

Zudem wird angenommen, dass es eine Verbindung zwischen dem Faktor Prüfungsangst und Leistung gibt. Es wird vermutet, dass Personen, die nicht unter Prüfungsangst leiden, besser im Test abschneiden als Personen mit großer Prüfungsangst. Außerdem wird erwartet, dass die Leistung bei Kandidaten, die meinen, dass sie körperlich wie psychisch sehr belastbar in Prüfungssituationen seien, auch besser abschneiden als diejenigen, die der Überzeugung sind, dass sie diese Fähigkeiten nicht hätten.

Daneben wird die Annahme überprüft, dass sich diejenigen, die sehr gut im TMS abgeschnitten haben, von denjenigen, die unterdurchschnittlich im TMS abgeschnitten haben, im Hinblick auf folgende Faktoren unterscheiden: Geschlecht, Ausprägung der Mathematikkennnisse, Versagensangst, Dauer des Wunsches Medizin zu studieren, Ausprägung im medizinischen Vorwissen, Vorbereitung und Abiturdurchschnittsnote.

8.2 Ausgewählte Spezialstichprobe

In den Jahren 1986 bis 1995 hat die CCH-Laufbahnberatung Vorbereitungskurse für den TMS und zu einem gesonderten Termin sogenannte Generalproben angeboten. Insgesamt haben in diesem Zeitraum über 6.000 Personen an einem Seminar mit Generalprobe, nur einem Seminar oder nur einer Generalprobe teilgenommen. Alle Teilnehmer beabsichtigten die Aufnahme eines Studiums entweder im Bereich Humanmedizin, Zahnmedizin oder Tiermedizin. Um ihre Zulassungschancen auf einen der begehrten Studienplätze zu erhöhen, hatten sich diese Personen entschieden an einem Trainingskurs für den TMS teilzunehmen. Sie versprachen sich davon eine deutliche Steigerung ihrer Testleistung.

Zu Beginn des Seminarbesuchs und ebenfalls vor der Generalprobe wurden die Kursteilnehmer gebeten, verschiedene Fragebögen zu beantworten. Nachdem die Studienbewerber den TMS abgelegt hatten, wurden sie mit Hilfe eines Evaluationsfragebogens nachträglich zu dem Seminar befragt. Hierbei wurden die Kandidaten auch nach ihren Ergebnissen im TMS befragt.

Die untersuchte Stichprobe besteht aus $N = 3.121$ Versuchspersonen. Diese Personen haben in den Jahren von 1991 bis 1994 an einem Seminar und/oder einer Generalprobe bei der CCH-Laufbahnberatung teilgenommen. Für die nachfolgende Überprüfung der Hypothesen werden aus der gesamten Stichprobe zunächst einmal nur die Personen ausgewählt, die den Evaluationsfragebogen mit einem eingetragenen TMS-Wert an die CCH-Laufbahnberatung zurückgeschickt haben. Die Stichprobe reduziert sich daher von 3.121 auf 1.424 Probanden. Zudem wurden nur Seminarteilnehmer ausgewählt, die sowohl an einem Seminar als auch an der Generalprobe teilgenommen hatten. Außerdem sollten beide Fragebögen, die während des Seminars verteilt wurden, und der Fragebogen der Generalprobe ausgefüllt worden sein. Zur optimalen Vergleichbarkeit

werden von 1.424 Teilnehmern 967 von der Untersuchung ausgeschlossen. Die Stichprobe reduziert sich also auf 457. Obwohl zur Analyse der Daten von den Jahrgängen 1991 bis 1994 eingeschlossen wurden, waren nach endgültiger Auswahl der Probanden nur noch Personen dabei, die in den Jahren 1993 und 1994 ein Seminar und die Generalprobe bei der CCH-Laufbahnberatung besucht hatten. Von den anderen Teilnehmern fehlten entscheidende Daten, die hinsichtlich der Vergleichbarkeit der Gruppen notwendig gewesen wären. So wurden in die Stichprobe 263 Personen (57,5 Prozent) aufgenommen, die ein Seminar im Jahr 1993 besucht hatten, und 194 Personen (42,5 Prozent), die an einem Seminar im Jahr 1994 teilgenommen hatten. Die Stichprobe besteht aus 209 Männern (45,7 Prozent) und 248 Frauen (54,3 Prozent). In Tabelle 8.1 sind die durchschnittlichen Abiturnoten, die die Seminarteilnehmer entweder erwarten oder bereits erreicht haben, aufgeführt. Bild 8.1 zeigt die „Erwartete Abiturnote“ zur besseren Übersichtlichkeit noch einmal als Balkendiagramm.

Tabelle 8.1: Übersicht über die erwarteten bzw. schon erhaltenen Abiturnoten der Stichprobe

	Häufigkeit	Prozent	kumulierte Prozente
1,00 bis 1,40	28	6,1 %	6,1%
1,41 bis 1,80	85	18,6%	24,7%
1,81 bis 2,20	156	34,1%	58,9%
2,21 bis 2,60	112	24,5%	83,4%
2,61 bis 3,00	57	12,5%	95,8%
3,01 bis 3,41	14	3,1%	98,9%
3,41 und mehr	3	0,7%	99,6%
keine Angabe	2	0,4%	100%
Gesamt	457	100%	

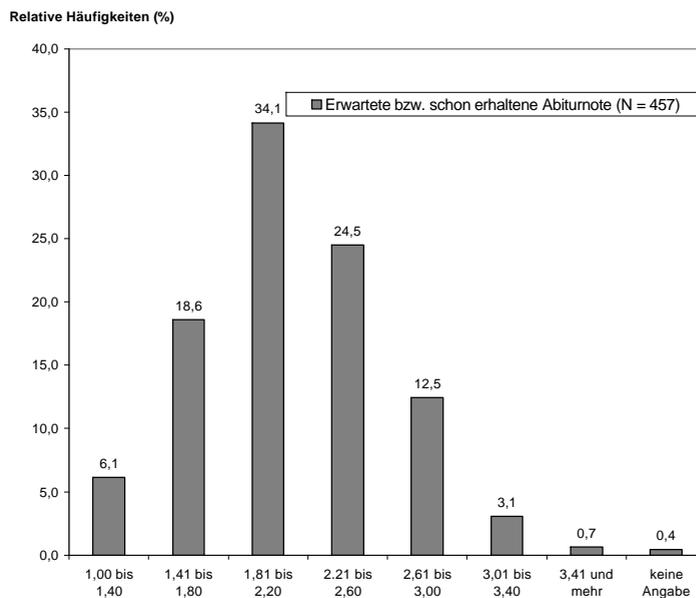


Bild 8.1: Verteilung der erwarteten bzw. schon erhaltenen Abiturnoten (Angaben in Prozent)

Von den 457 Personen wollten 357 Teilnehmer Humanmedizin (78,1 Prozent) und 63 wollten Zahnmedizin studieren (13,8 Prozent). Achtundzwanzig gaben als Studienwunsch Tier-medicin an (6,1 Prozent) und 9 (2 Prozent) machten keine Angaben. Bild 8.2 zeigt anhand eines Balkendiagramms eine Übersicht über die Vorbereitungszeit, die die Studienbewerber neben Seminar und Generalprobe zusätzlich noch selbstständig aufgebracht haben.

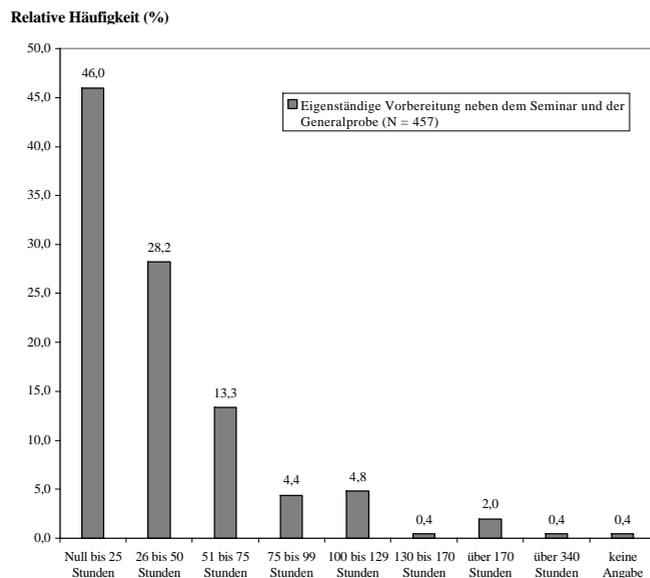


Bild 8.2: Übersicht über die Vorbereitungszeit neben dem Seminar und der Generalprobe (Angaben in Prozent)

Von den Seminarteilnehmern hatten 53 Personen (11,6 Prozent) erst seit kurzem den Wunsch ein Studium im medizinischen Bereich zu beginnen. Seit über einem Jahr haben 148 (32,4 Prozent) Teilnehmer schon über ihren Wunsch nachgedacht und 251 Personen (54,9 Prozent) wussten es schon seit mehreren Jahren. Nur fünf Personen (1,1 Prozent) konnten über die Dauer ihres Wunsches keine Angaben machen. Die Tabelle 8.5 gibt noch einmal einen Überblick über die Verteilung der erreichten standardisierten TMS-Werte, aufgeteilt in Kategorien, in der beobachteten Gesamtstichprobe.

Tabelle 8.2: Überblick über die Verteilung der TMS-Werte in Kategorien

	Häufigkeit	Prozent	kumulierte Prozente
80 -89	4	0,9%	0,9%
90 – 99	56	12,3%	13,1%
100 – 109	213	46,6%	59,7%
110 – 119	159	34,8%	94,5%
120 – 129	25	5,5%	100%
Gesamt	457	100%	

8.3 Messinstrumente

Insgesamt standen vier Fragebögen als Messinstrumente zur Verfügung. Die Fragebögen wurden von der CCH-Laufbahnberatung entwickelt. Zunächst einmal wurden alle Teilnehmer gebeten, einen allgemeinen Fragebogen auszufüllen. Auf diesem Fragebogen sollten die Teilnehmer neben ihren persönlichen Angaben auch Auskunft darüber geben, welches medizinische Fach sie studieren möchten, wie lange der Studienwunsch bereits besteht, wann sie Abitur gemacht haben oder machen werden und welche Durchschnittsnote sie dabei erzielt bzw. erwarten.

Die mit Hilfe der Fragebögen erhobenen Daten wurden nicht nur zu wissenschaftlichen Zwecken benötigt, sondern konnten bei den Einzelgesprächen Aufschluss darüber geben, wo der einzelne Teilnehmer Probleme hat und wie diese behoben werden

könnten. Zudem hatten die Teilnehmer anhand der Fragebögen die Möglichkeit, ihre Einstellungen und Befindlichkeiten selbst zu reflektieren.

Fragebogen „Prüfungserwartung“

Der „Fragebogen zur allgemeinen und spezifischen Prüfungserwartung“⁹, den die Teilnehmer im Seminar beantwortet haben, umfasste insgesamt 16 Fragen. Die Fragen waren in zwei Abschnitte aufgeteilt. Im ersten Abschnitt wurden 11 Fragen zur allgemeinen Prüfungserwartung gestellt. Die Fragen 12 – 16 im zweiten Abschnitt bezogen sich auf den Test für Medizinische Studiengänge (TMS) und auf die Zulassung zum Medizinstudium. Die Items waren von 1 bis 5 auf einer Likert-Skala skaliert („stimmt genau, stimmt, stimmt in etwa, stimmt kaum, stimmt gar nicht“). Im ersten Abschnitt sollten die Kandidaten an Prüfungen allgemein denken, wie z. B. Abitur- oder Führerscheinprüfung. Mit diesen Hintergrundgedanken sollten sie dann Fragen beantworten, z. B. ob das Prüfungsergebnis von der Intensität der Vorbereitung abhängt oder wie gelassen sie Prüfungen entgegen sehen. Im zweiten Teil des Fragebogens wurden Fragen derart gestellt, wie z. B. „Ich bin recht unsicher über mein Abschneiden im TMS“ oder „Eigentlich bin ich recht zuversichtlich was meine Zulassungschancen zum Medizinstudium anbelangt“. Der Fragebogen wurde zunächst zu Beginn des Seminars ausgeteilt und ein zweites Mal zu Beginn der Generalprobe. Für die nachfolgende Untersuchung sind folgende Fragen von Interesse:

- Item 7: „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“.

Fragebogen „Faktoren für ein gutes Abschneiden im TMS“

Den Fragebogen beantworteten die Teilnehmer ebenfalls im Laufe des Seminars. Dieser Fragebogen „Faktoren für ein gutes Abschneiden im TMS“¹⁰ war ebenfalls in zwei Abschnitte geteilt. Beide Teile bestehen jeweils aus den gleichen 10 Items. Im ersten Abschnitt werden die Kursteilnehmer gefragt, für wie wichtig sie die auf dem Fragebogen genannten Faktoren für ein optimales Abschneiden im TMS befinden.

⁹ Fragebogen „Prüfungserwartung“ befindet sich im Anhang

¹⁰ Fragebogen „Faktoren für ein gutes Abschneiden im TMS“ befindet sich im Anhang

Die Fragen werden mit Hilfe einer Likert-Skala von 1 – 5 beantwortet („absolut wichtig“, „unwichtig“, „mäßig wichtig“, „wichtig“ und „absolut wichtig“). Im zweiten Abschnitt wird dann gefragt, in welcher Weise diese Faktoren bei ihnen ausgeprägt sind. Hier werden die Ausprägungen ebenfalls mit Hilfe einer Likert-Skala von 1 – 5 beantwortet („gar nicht ausgeprägt“, „kaum ausgeprägt“, „mäßig ausgeprägt“, „gut ausgeprägt“, „sehr gut ausgeprägt“). Der erste Teil des Fragebogens befasst sich also mit der Wichtigkeit der Faktoren, der zweite mit der Ausprägung dieser Faktoren bei den einzelnen Teilnehmern. Der Fragebogen wurde zunächst zu Beginn des Seminars ausgeteilt und ein zweites Mal zu Beginn der Generalprobe. Für die nachfolgende Untersuchung sind folgende Fragen von Interesse:

- Item 1 (Ausprägung): gute Kenntnisse in Mathematik
- Item 3 (Ausprägung): gute Nerven
- Item 4 (Ausprägung): ausreichende Vorbereitung
- Item 5 (Ausprägung): medizinisches Vorwissen
- Item 7 (Ausprägung): Durchhaltevermögen
- Item 8 (Ausprägung): Ruhig-Bleiben-Können
- Item 9 (Ausprägung): die letzten Reserven mobilisieren können

Fragebogen „Testbefindlichkeit“

Der „Fragebogen zur Testbefindlichkeit“¹¹ wurde den Kursteilnehmern nach der Testsimulation vorgelegt. Insgesamt bestand der Fragebogen aus 23 Fragen zur Testbefindlichkeit. Er war in drei Abschnitte eingeteilt. Alle Fragen wurden auf einer Likert-Skala von 1 – 5 beantwortet („stimmt genau“, „stimmt“, „stimmt in etwa“, „stimmt kaum“ und „stimmt gar nicht“). Im ersten Abschnitt wurden Fragen zur aktuellen körperlichen und geistigen Befindlichkeit nach absolviertem Test gestellt. Abschnitt zwei beschäftigte sich mit Fragen zur Testsituation, z. B. „bestimmte Geräusche haben mich immer wieder gestört (z. B. Naseputzen, Klimaanlage, Straßenlärm ...)“. Im dritten Abschnitt wurden Fragen zum Test und zu den

¹¹ Fragebogen „Testbefindlichkeit“ befindet sich im Anhang

Testeindrücken gestellt. Der Fragebogen wird aus Gründen der Vollständigkeit hier vorgestellt. Für die nachfolgende Untersuchung ist er nicht relevant.

Fragebogen „Fragebogen zur Generalprobe“

Dieser „Fragebogen zur CCH-Generalprobe“¹² wurde den Teilnehmern am Tag der Generalprobe vorgelegt. Er diente zur Feststellung der Vorbereitungsart und –zeit der Medizinstudienplatzanwärter. Gefragt wurde nach der Art des Testtrainingsseminars, sowie nach der eigenständigen Vorbereitungszeit, die unabhängig von jeglichem Coachingkurs stattgefunden hatte.

Evaluationsbogen zum Testtraining

Der Evaluationsfragebogen¹³ war ein Fragebogen zum TMS und zum CCH-Test-Training. Dieser Fragebogen wurde den Teilnehmern jeweils mit einem beiliegenden Freiumschlag zugeschickt, nachdem der bundesweite Einsatz des TMS stattgefunden hatte. Die Kursteilnehmer wurden gefragt, ob sie am TMS teilgenommen hatten und wurden gebeten, ihre Ergebnisse in die entsprechenden Felder des Fragebogens einzutragen. Außerdem wurden ihnen Fragen zum TMS, zum CCH-Test-Training und zum Zulassungsverfahren für die medizinischen Studiengänge gestellt. Für die nachfolgende Untersuchung sind folgende Fragen von Interesse:

- Erzielter Testwert (TMS-Wert)
- Erreichter Prozentrang
- Gesamtpunktzahl
- und die in den einzelnen Untertests erzielten Punkte

8.4 Auswertungsvorgehen

Zur Überprüfung der Hypothesen, bei denen man von einer Normalverteilung ausgehen kann und bei denen die Stichprobe nicht in Extremgruppen aufgeteilt wird, wird ein t-Test gerechnet. Der t-Test gehört zu den gebräuchlichsten statistischen Analysen. Er

¹² Fragebogen „Fragebogen zur Generalprobe“ befindet sich im Anhang

¹³ Der Evaluationsfragebogen befindet sich im Anhang

gehört zu den parametrischen Verfahren und bietet eine Entscheidungsregel auf mathematischer Grundlage. Der t-Test prüft, ob sich zwei empirisch gefundene Mittelwerte systematisch unterscheiden (Rasch, Friese, Hofmann & Naumann, 2006a). In dem Verfahren werden anhand der Stichprobe die Populationsparameter, Streuung und arithmetisches Mittel, geschätzt. Voraussetzungen zur Anwendung des t-Tests sind: intervallskalierte Daten, normalverteilte Grundgesamtheiten und Varianzhomogenität der Stichproben. Der t-Test ist allerdings auch bei Verletzung der Voraussetzungen relativ robust. Wichtig ist nur, dass die Stichproben nahezu dieselbe Größe haben und dass sie nicht zu klein sind ($n_1 = n_2 = 30$). Bei kleineren Stichproben oder sehr ungleich großen Stichproben könnte das Ergebnis des t-Tests unkorrekt sein. Sollten die Stichproben also zu klein oder ihre Größe zu unterschiedlich sein, sollte auf ein verteilungsfreies Verfahren zur Überprüfung der Hypothesen zurückgegriffen werden. In der vorliegenden Arbeit werden, falls notwendig, auch die Ratingskalen als intervallskaliert interpretiert. Die Voraussetzung für die Varianzgleichheit wird hier mit dem Levene-Test überprüft werden. Für die statistische Untersuchung wird gegebenenfalls der t-Test für unabhängige Stichproben gewählt. Unabhängigkeit der Gruppen bedeutet, dass die Werte der einen Gruppen keine Einwirkung auf die Werte der anderen Gruppe hat. Wenn sich in den Gruppen unterschiedliche Versuchspersonen befinden, dann sind die Stichproben meistens unabhängig voneinander.

Da es sich bei den Untersuchungen zum Teil um Extremgruppenvergleiche handelt, eine spezielle Abwandlung des quasiexperimentellen Zweigruppenplans, wird in diesem Fall nicht der t-Test zur Überprüfung der Hypothesen gewählt (Bortz & Döring, 2006). Bei einem Extremgruppenvergleich gehen nur Versuchspersonen in die Untersuchung ein, die bezüglich der unabhängigen Variablen besonders hohe oder besonders niedrige Ausprägungen aufweisen. Nach Bortz (2006) gehören solche Extremgruppenvergleiche eher zu der Klasse der explorativen Studien. Ihre Intention ist es, herauszufinden, ob eine unabhängige Variable einen möglichen Erklärungswert für eine abhängige Variable hat. Es kann also passieren, dass die Ergebnisse die Bedeutung der untersuchten unabhängigen Variablen zu hoch einschätzen. Ein verteilungsfreies Verfahren ist zum Vergleich besser geeignet, da ein Extremgruppenvergleich mit dem t-Test oft nicht die Voraussetzungen des t-Tests erfüllt (Bortz & Döring, 2006). Verteilungsfreie Verfahren

werden eingesetzt, wenn ordinalskalierte Daten analysiert werden sollen oder wenn die Voraussetzungen für einen t-Test oder eine Varianzanalyse verletzt sind.

Da man bei dem Extremgruppenvergleich zusätzlich nicht davon ausgehen kann, dass die untersuchten Merkmale normalverteilt sind, ist es ebenfalls sinnvoller, solche Daten mit einem nonparametrischen Verfahren zu untersuchen (Nachtigall & Wirtz, 2006). Diese Verfahren benötigen weniger Voraussetzungen. Allerdings ist der β -Fehler auch größer. Es besteht also eher die Gefahr, die H_1 zu verwerfen, obwohl die H_1 gilt. Sind also die Voraussetzungen für ein parametrisches Verfahren nicht gegeben, wird die zentrale Tendenz zweier Stichproben mit Hilfe von Rangetests verglichen. Es werden nicht mehr die Abstände zwischen zwei Messwerten verglichen sondern die Rangplätze. Dabei gehen selbstverständlich auch Informationen verloren, aber sind unempfindlicher gegenüber Ausreißern (Bühl, 2006). In der nachfolgenden Untersuchung der Extremgruppen wird zum Vergleich von zwei unabhängigen Stichproben hinsichtlich ihrer zentralen Tendenz der Mann-Whitney U-Test verwendet. Der Test stützt sich auf eine gemeinsame Rangreihe der Werte beider Stichproben. Der U-Test stellt künstlich einen gleichen Abstand zwischen den Werten her. Auf diese Weise sind mathematische Operationen, wie z. B. die Mittelwertbildung möglich (Rasch, Friese, Hofmann & Naumann, 2006b). Der U-Test als nichtparametrisches Verfahren bzw. Rangverfahren arbeitet nicht mit Populationsparametern und -verteilungen. Er legt seine eigene Verteilung, die Verteilung der Rangplätze, fest. Es gibt also wenige mathematische Voraussetzungen für seine Anwendung: Rangskaliertheit der Daten und zufällige Anordnung der Versuchspersonen zu den Gruppen. Zudem ist er eine Alternative, wenn die Voraussetzungen des t-Tests verletzt werden und bzw. oder wenn die Stichproben sehr klein sind. Der U-Test ist ein Verfahren zur Prüfung, ob sich zwei Gruppen in einer unabhängigen Variable unterscheiden. Getestet wird, ob die Unterschiede zwischen zwei Gruppen hinsichtlich einer abhängigen Variable von zufälligen oder systematischen Einflüssen bestimmt werden. Dabei werden nicht die Messwerte direkt sondern die ihnen zugeordneten Rangplätze untersucht. Ein Vorteil dieses verteilungsfreien Verfahrens ist, dass nach der Zuordnung der Rangplätze die Signifikanzprüfung unabhängig von der Form der Verteilung der Messwerte in der Population ist. Mit dem U-Test kann man eine Aussage darüber treffen, ob ein

signifikanter Unterschied zwischen den beiden analysierten Stichproben hinsichtlich der Verteilung der Rangplätze besteht.

8.5 Datenerhebung

Zunächst einmal wurden alle Teilnehmer gebeten, den allgemeinen Fragebogen zur Erhebung von generellen Daten auszufüllen. Die Instruktionen zur Beantwortung aller Fragebögen wurden den Seminarteilnehmern sowohl in mündlicher als auch in schriftlicher Form gegeben. Alle Teilnehmer erhielten zu Beginn des Seminars folgende Fragebögen vorgelegt: „Fragebogen zur allgemeinen und spezifischen Prüfungserwartung“ und „Faktoren für ein gutes Abschneiden im TMS“. Den Fragebogen zur Testbefindlichkeit erhielten die Seminarteilnehmer direkt in Anschluss an eine Testsimulation. Die Seminare begannen fast immer mit einer Testsimulation. Vor der Generalprobe, die an einem gesonderten Termin zeitnah zum TMS stattfand, wurden die Teilnehmer gebeten, den „Fragebogen zur Generalprobe“, Fragebogen „Prüfungserwartung“ und den Fragebogen „Faktoren für ein gutes Abschneiden im TMS“ auszufüllen. Die Nachbefragung mit dem Evaluationsbogen fand einige Zeit später statt, nachdem die Bewerber bereits ihre Testergebnisse erhalten hatten. Der Ablauf der Datenerhebung mit ungefähren Zeitangaben ist der Tabelle 8.3 zu entnehmen.

Tabelle 8.3: Darstellung des Versuchablaufs mit ungefähren Zeitangaben

Versuchsdurchführung				
Messzeitpunkt	Seminar (Zeitpunkt I)	Generalprobe (Zeitpunkt II)	TMS (Zeitpunkt III)	TMS-Ergebnis (Zeitpunkt IV)
	(z. B. 2-3 Tage im Zeitraum Juni bis Oktober)	(z. B. Ende Oktober bis Anfang November)	(Anfang November)	(Mitte bis Ende Dezember)
Messinstrumente	Fragebögen ^a	Fragebögen ^b		Evaluationsbogen

Anmerkungen: ^aFragebögen :Allgemeiner Fragebogen & Fragebögen: „Prüfungserwartung“ „Faktoren für ein gutes Abschneiden im TMS“ „Testbefindlichkeit“, ^bFragebögen: Fragebogen zur Generalprobe & Fragebögen: „Prüfungserwartung“ „Faktoren für ein gutes Abschneiden im TMS“

8.6 Befunde der Analyse der Spezialstichprobe

In diesem Kapitel werden die in Abschnitt 8.1 vorgestellten Fragestellungen mit Hilfe statistischer Verfahren überprüft. Wie schon im Abschnitt 8.4 erwähnt, werden hierfür maßgeblich zwei Testverfahren eingesetzt: der t-Test für unabhängige Stichproben und der Mann-Whitney U-Test für kleine Stichproben, die bei einer Überprüfung mit einem parametrischen Verfahren die Voraussetzungen des t-Tests verletzen würden. Mit dem Mann-Whitney U-Test werden im Folgenden die Hypothesen zu den Extremgruppen überprüft. Bei der Darstellung der Ergebnisse werden die jeweiligen Hypothesen noch einmal angeführt. Da es sich jeweils um relativ kleine Stichproben handelt und aufgrund des eingeschränkten experimentellen Designs, wird ein Signifikanzniveau von 1 % festgelegt. Alle empirischen Überprüfungen beziehen sich auf die ausgewerteten Daten aus den Fragebögen, die im Laufe des Seminars verteilt wurden. Die Fragebögen waren nicht nur zu wissenschaftlichen Zwecken entwickelt worden, sondern vielmehr dazu, die Seminarteilnehmer für Probleme und Schwierigkeiten, die eine Prüfungssituation bestimmen können, zu sensibilisieren. Die Teilnehmer sollten anhand der Fragebögen erkennen, insbesondere beim Fragebogen II „Faktoren für ein gutes Abschneiden im TMS“, welche Aspekte wichtig sind, um in einer Prüfung und speziell beim TMS gut abzuschneiden. Im Laufe des Seminars sollten sich die Kursteilnehmer mit den Faktoren eingehender beschäftigen. Da es folglich sein kann, dass die Seminarteilnehmer bei der Beantwortung desselben Fragebogens andere Angaben machten als bei der ersten Befragung im Seminar, eventuell auch um sich selbst stärker zu motivieren, denn der Testtermin stand kurz bevor, wird in dieser empirischen Analyse nur mit den Daten aus den Fragebögen gearbeitet werden, die im Seminar verteilt wurden. Im Seminar wurden die Fragebögen ohne Prüfungsdruck und Leistungsdruck beantwortet. Es wird in dieser Arbeit davon ausgegangen, dass die Beantwortung ehrlich und wahrhaftig war, da sie noch frei von Prüfungsdruck bzw. Leistungsdruck stattfand. Die Teilnehmer kamen schließlich in das Seminar, um zu lernen und wollten bestens auf den TMS vorbereitet werden. Es gab für sie zu diesem Zeitpunkt keine Veranlassung, die Fragen nicht ehrlich zu beantworten.

Wenn die Stichprobe genügend groß ist, wird der t-Test für ungepaarte Stichproben angewandt. Dies ist z. B. der Fall bei der Überprüfung der Annahme, dass sich Frauen

und Männer hinsichtlich ihrer empirisch gefundenen Mittelwerte im TMS unterscheiden.

8.7 Darstellung der Ergebnisse

Hypothese I „Einfluss der Abiturnote“

Die Hypothese I „Die Gruppe der sehr guten Abiturienten unterscheidet sich in der Verteilung der Rangplätze hinsichtlich des TMS-Gesamtwerts und auch aller neun Untertests signifikant von der Gruppe der schlechten Abiturienten“ wird mit Hilfe des Mann-Whitney U-Tests überprüft. Zunächst einmal muss allerdings eine Einteilung in sehr gute und schlechte Abiturienten geschehen. Wie aus Tabelle 8.1 ersichtlich wird, gibt es 28 Personen in der Stichprobe, die mit der Abiturdurchschnittsnote 1,00 bis 1,40 rechnen oder die ihr Abitur bereits mit dieser Note abgeschlossen haben. Auf der anderen Seite gibt es 17 Personen, die mit einem Abiturdurchschnitt über 3,01 rechnen oder die Schule bereits mit diesem Notendurchschnitt beendet haben. Diese beiden Gruppen werden miteinander verglichen: hinsichtlich des Gesamtwertes, aber auch bezüglich aller Untertests. Die Ergebnisse des U-Tests sind in der Tabelle 8.4 dargestellt. Aus der Gesamtstichprobe von $N = 457$ werden zwei Stichproben für die Extremgruppen gebildet. Die Gruppe der Seminar Teilnehmer, die einen sehr guten Abiturdurchschnitt haben oder haben werden, besteht aus $n = 28$ Personen. Das sind 6,1 Prozent der Gesamtstichprobe. Die andere Gruppe mit einem Abiturdurchschnitt über 3,01 umfasst $n = 17$. Das sind 3,8 Prozent der Gesamtstichprobe.

Tabelle 8.4: Vergleich der TMS-Werte der Gruppen „sehr gute“ Abiturienten und „schlechte“ Abiturienten (U-Test von Mann-Whitney)

Skala ^a		Gruppe 1 ^b	Gruppe 2 ^c	U
		MR	MR	
TMS-Wert	n ₁ = 28 n ₂ = 17	29,96	11,53	43,000*
MZ	n ₁ = 27 n ₂ = 16	24,96	17,00	136,000
MNGV	n ₁ = 26 n ₂ = 16	26,35	13,63	82,000*
SF	n ₁ = 27 n ₂ = 16	25,02	16,91	134,500
QFP	n ₁ = 26 n ₂ = 17	28,52	12,03	51,5000*
KONZ	n ₁ = 27 n ₂ = 17	23,41	21,06	205,000
TV	n ₁ = 27 n ₂ = 16	26,80	13,91	86,500*
FIG	n ₁ = 27 n ₂ = 16	23,89	18,81	165,000
FAK	n ₁ = 27 n ₂ = 16	26,72	14,03	88,500*
DT	n ₁ = 27 n ₂ = 16	26,37	14,63	98,000*

Anmerkungen: ^aSkalen: TMS-Wert: *standardisierter TMS-Gesamttestwert*, MZ: *Muster zuordnen*, MNGV: *Medizinisch-naturwissenschaftliches Grundverständnis*, SF: *Schlauchfiguren*, QFP: *Quantitative und formale Probleme*, KONZ: *Konzentriertes und sorgfältiges Arbeiten*, TV: *Textverständnis*, FIG: *Figuren lernen*, FAK: *Fakten lernen*, DT: *Diagramme und Tabellen*, ^bAbiturdurchschnittsnote zwischen 1,00 bis 1,40, ^cAbiturdurchschnittsnote über 3,01, * p < .01 (2-seitig)

Von den sehr guten Abiturienten hat keiner unterdurchschnittlich, d.h. unter 100 Punkten, beim TMS-Gesamttestwert abgeschnitten. In der Gruppe der schlechten Abiturienten waren immerhin 3 von 17 unter 100 Punkten, also 17,6 Prozent. Der maximale Testwert lag in dieser Gruppe bei 110 Punkten. Im Vergleich hierzu hatte der Beste aus der Gruppe der sehr guten Abiturienten einen Wert von 129. In der Gruppe der sehr guten Abiturienten hatten 67,9 Prozent einen Wert von 110 und höher. Beim Untertest „medizinisch-naturwissenschaftliches Grundverständnis“ lagen von den schlechten Abiturienten 62,5 Prozent unter der Hälfte der erreichbaren Punkte. Von den Teilnehmern der Gruppe der sehr guten Abiturienten erreichten 88,5 Prozent mehr als die Hälfte der Punkte in diesem Test. Beim Untertest „Konzentriertes und sorgfältiges Arbeiten“ ergab sich ein anderes Bild: Hier lagen beide Gruppen deutlich über der durchschnittlich erreichbaren Punktzahl. Während in der Gruppe mit dem schlechten Abitur fünfzehn Personen (88,3 Prozent) über 18 Punkte erreichten, waren es in der Gruppe der sehr guten Abiturienten neunzehn (70,4 Prozent).

Hypothese II „Einfluss zusätzlicher Vorbereitungszeit“

Die Hypothese II „Die Gruppe der intensiv Vorbereiteten unterscheidet sich in der Verteilung der Rangplätze bezüglich des TMS-Gesamtwerts signifikant von der Gruppe der schlecht oder wenig Vorbereiteten“ wird ebenfalls mit Hilfe des Mann-Whitney U-Tests überprüft. Die Frage nach der Vorbereitung kann anhand der Daten aus den Fragebögen auf verschiedene Arten empirisch überprüft werden. Zunächst einmal wurde die Frage nach der zusätzlichen selbständigen Vorbereitung bei der Befragung zur Generalprobe gestellt. Die Seminarteilnehmer sollten aber auch selbst auf dem Fragebogen „Faktoren für ein gutes Abschneiden im TMS“ einschätzen, wie stark der Faktor „ausreichende Vorbereitung“ bei ihnen ausgeprägt ist.

Zunächst wurden die Daten aus der Befragung während der Generalprobe überprüft. Die Gesamtstichprobe von $N = 457$ wurde in zwei Extremgruppen aufgeteilt. Die eine Gruppe enthielt die Personen, die sich neben dem Seminar und der Generalprobe nur noch mäßig eigenständig vorbereitet hatte, nämlich Null bis 25 Stunden. Sie enthielt $n = 210$ Personen (46 Prozent der Gesamtstichprobe). Auf der anderen Seite enthielt die andere Extremgruppe $n = 35$ Seminarteilnehmer (7,7 Prozent der Gesamtstichprobe). Dies ist die Gruppe derjenigen, die sich neben dem Seminar und der Generalprobe zu

Hause noch über 100 Stunden vorbereitet hatte. Die Ergebnisse des U-Tests sind in Tabelle 8.5 dargestellt.

Tabelle 8.5: Vergleich der TMS-Werte der Extremgruppen „Sehr-gut-Vorbereitete“ und „Kaum-Vorbereitete“ (U-Test von Mann-Whitney)

	Gruppe 1 ^b	Gruppe 2 ^c	
Skala ^a	(n = 35)	(n = 210)	U
	MR	MR	
TMS-Wert	130,59	121,74	3409,500

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt II und III (IV) (siehe Tabelle 8.3), ^aSkala: TMS-Wert: standardisierter TMS-Gesamtestwert, ^bzusätzliche Vorbereitungszeit neben dem Seminar und der Generalprobe: über 100 Stunden, ^czusätzliche Vorbereitungszeit neben dem Seminar und der Generalprobe: Null bis 25 Stunden

Zur Überprüfung der empirischen Daten hinsichtlich der selbsteingeschätzten Ausprägung auf dem Faktor „ausreichende Vorbereitung“ wurden die Extremgruppen aus der Beantwortung des Fragebogens „Faktoren für ein gutes Abschneiden im TMS“ aus dem Seminar verwendet. Zwanzig Personen hatten geantwortet, dass sie meinten, dieser Faktor wäre bei ihnen sehr gut ausgeprägt. Das sind 4,4 Prozent der Gesamtstichprobe. Dagegen meinten siebzehn Teilnehmer, dieser Faktor wäre bei ihnen gar nicht ausgeprägt. Das entsprach 3,7 Prozent der Gesamtstichprobe. Der statistische Vergleich der beiden Extremgruppen ist in Tabelle 8.6 dargestellt.

Tabelle 8.6: Vergleich der TMS-Werte der Gruppen „Sehr gut“ und „Gar nicht“ bei der Befragung des Ausprägungsgrads des Faktors „ausreichende Vorbereitung“ / Befragung im Laufe des Seminars (U-Test von Mann-Whitney)

	Gruppe 1 ^b	Gruppe 2 ^c	
Skala ^a	(n = 20)	(n = 17)	U
	MR	MR	
TMS-Wert	19,50	18,41	160,000

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt I und III (IV) (siehe Tabelle 8.3), ^aSkala: TMS-Wert: *standardisierter TMS-Gesamtttestwert*, ^bAusprägung auf dem Faktor „ausreichende Vorbereitung“ = sehr gut ausgeprägt, ^c Ausprägung auf dem Faktor „ausreichende Vorbereitung“ = gar nicht ausgeprägt

Hierbei ist interessant, dass sich in der Gruppe derjenigen, die meinen, dass dieser Faktor bei ihnen „gar nicht“ ausgeprägt sei, kein Seminarteilnehmer findet, der einen Abiturdurchschnitt von über 3,01 hat. Allerdings fand sich auch kein „schlechter“ Abiturient in der Gruppe derjenigen, die meinen, dass dieser Faktor bei ihnen „sehr gut“ ausgeprägt sei. Dieses Detail könnte ein Indiz dafür sein, dass die Teilnehmer höchstwahrscheinlich wahrheitsgemäß geantwortet haben. Man würde bei einem Abiturient mit der Durchschnittsnote 3,01 und schlechter nicht erwarten, dass er der Auffassung ist, dass der Faktor „ausreichende Vorbereitung“ bei ihm besonders ausgeprägt sei.

Hypothese III „Einfluss ausgeprägter Mathematikkenntnisse“

Die Hypothese III „Die Gruppe derjenigen, die ihre mathematischen Kenntnisse als sehr gut ausgeprägt einschätzen, unterscheidet sich in der Verteilung der Rangplätze in Bezug sowohl auf den Untertest „Quantitative und formale Probleme“ als auch auf den Untertest „Diagramme und Tabellen“ signifikant von der Gruppe, die ihre mathematischen Kenntnisse als „gar nicht ausgeprägt“ einschätzen, wird ebenfalls mit dem Mann-Whitney U-Tests überprüft. Für die Bildung der Extremgruppen wurden aus der Gesamtstichprobe N = 457 zwei Gruppen gebildet. Die eine Gruppe hatte im Fragebogen I angekreuzt, dass sie ihre Ausprägung „gute Kenntnisse in Mathematik“ als sehr gut einschätzen. Dies waren 31 Personen (6,8 Prozent der Gesamtstichprobe). Für die entgegen gesetzte Extremgruppe wurden die Kursteilnehmer zusammengefasst,

die ihre „guten Kenntnisse in Mathematik“ als „gar nicht“ bzw. „kaum“ ausgeprägt empfanden. Es bildete sich eine Gruppe von 47 Personen (10,3 Prozent der Gesamtstichprobe). Die Ergebnisse des U-Tests sind in Tabelle 8.7 dargestellt. Schon die Mittelwerte der guten Mathematiker liegen weit auseinander: Die „guten“ Mathematiker haben im Untertest QFP einen Mittelwert von 13,48 (3,375) und im Untertest DT einen Mittelwert von 13,12 (3,334) im Vergleich zu den „schlechten“ Mathematikern, die einen Mittelwert in QFP von 8,60 (2,864) und in DT einen Mittelwert von 10,80 (3,422) haben. Deutlich wird das auch daran, dass von denjenigen, die ihre Mathematikkenntnisse als „sehr gut“ einschätzten, über 90 Prozent über 10 Punkte bei dem Untertest QFP, also über die Hälfte der erreichbaren Punkte, erlangten. Hingegen bei denjenigen, die meinten, sie hätten nur „schwach“ ausgeprägte Mathematikkenntnisse, blieben ca. 60 Prozent unter dieser Marke.

Tabelle 8.7: Vergleich der TMS-Werte der Gruppen „sehr gute“ Mathematiker und „schlechte“ Mathematiker (U-Test von Mann-Whitney)

Skala ^a		Gruppe 1 ^b	Gruppe 2 ^c	U
		MR	MR	
QFP	n ₁ = 31	56,82	28,07	191,500*
	n ₂ = 47			
DT	n ₁ = 31	47,11	32,57	430,500*
	n ₂ = 45			

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt I und III (IV) (siehe Tabelle 8.3), ^aSkalen: QFP: *Quantitative und formale Probleme*, DT: *Diagramme und Tabellen*, ^bAusprägung auf dem Faktor „gute Kenntnisse in Mathematik“ = sehr gut ausgeprägt, ^cAusprägung auf dem Faktor „gute Kenntnisse in Mathematik“ = gar nicht bzw. kaum ausgeprägt, * p < .01 (2-seitig)

Hypothese IV „Einfluss ausgeprägten medizinischen Vorwissens“

Die Hypothese IV „Die Gruppe derjenigen, die ihr medizinisches Vorwissen als „sehr gut“ ausgeprägt einschätzen, unterscheidet sich in der Verteilung der Rangplätze hinsichtlich des Untertests „medizinisch-naturwissenschaftliches Grundverständnis“ signifikant von der Gruppe, die ihr medizinisches Vorwissen als „gar nicht“ ausgeprägt einschätzen“ wird mit Hilfe des Mann-Whitney U-Tests überprüft. Aus der Befragung ging hervor, dass aus der Gesamtstichprobe N = 457 zehn Kursteilnehmer ihr „medizinisches Vorwissen“ als „sehr gut“ ausgeprägt einschätzten. Das waren 2,2 Prozent der Gesamtstichprobe. Dass ihr „medizinisches Vorwissen“ gar nicht ausgeprägt sei, meinten 44 Seminarteilnehmende (9,6 Prozent der Gesamtstichprobe). Die Ergebnisse des Tests sind in Tabelle 8.8 dargestellt.

Tabelle 8.8: Vergleich der TMS-Werte der Gruppen „sehr gute“ Mediziner und „sehr schlechte“ Mediziner (U-Test von Mann-Whitney).

	Gruppe 1 ^b	Gruppe 2 ^c	
Skala ^a	n = 10	n = 44	U
	MR	MR	
MNGV	37,85	25,15	116,500

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt I und III (IV) (siehe Tabelle 8.3), ^aSkala: MNGV: *Medizinisch-naturwissenschaftliches Grundverständnis*, ^bAusprägung auf dem Faktor „medizinisches Vorwissen“ = sehr gut ausgeprägt, ^cAusprägung auf dem Faktor „medizinisches Vorwissen“ = gar nicht ausgeprägt

Die Mittelwerte der beiden Extremgruppen „Sehr gut ausgeprägtes medizinisches Vorwissen“ und „gar nicht ausgeprägtes medizinisches Vorwissen“ unterscheiden sich auf jeden Fall in ihren Mittelwerten: die Gruppe mit den „sehr guten Medizinkenntnissen“ hat einen Mittelwert von 13,90 (2,132) und Gruppe, die „gar keine Medizinkenntnisse“ hatte, einen Mittelwert von 11,27 (3,223). Insgesamt haben alle Kursteilnehmer mit sehr guten medizinischen Vorkenntnissen in der oberen Hälfte in dem Untertest MNGV abgeschlossen. Bei der Gruppe der „schlechten Mediziner“ lagen immerhin noch ca. 30 Prozent unter der Hälfte der erreichbaren Punktzahl in diesem Untertest. Bei der Gruppe der „guten Mediziner“ hat keiner in der unteren Hälfte

abgeschlossen. Die höchste Punktzahl, die von den beiden Extremgruppen erreicht worden war lag, bei 17 Punkten. Es erreichten zwei Personen aus der Gruppe der „guten Mediziner“, aber auch zwei Personen aus der Gruppe der „schlechten Mediziner“ diese Punktzahl.

Hypothese V „Einfluss des Faktors Versagensangst“

Die Hypothese V „Die Gruppe derjenigen, die finden, dass die Aussage „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“ trifft auf sie gar nicht zu, unterscheidet sich in der Verteilung der Rangplätze in Bezug auf den TMS-Gesamtwert signifikant von der Gruppe, die meinen, dass diese Aussage „ganz genau“ auf sie zutrifft, wird mit Hilfe des Mann-Whitney U-Tests überprüft. Die Auswertung dieser Frage in der Gesamtstichprobe $n = 457$ ergab, dass 31 Personen, meinten, diese Aussage würde für sie „gar nicht“ zutreffen. Das waren 6,8 Prozent der Gesamtstichprobe. Hingegen meinten 55 Studienanwärter, dass diese Aussage auf sie „genau“ passen würde (12 Prozent der Gesamtstichprobe). Die Daten der empirischen Überprüfung stehen in Tabelle 8.9.

Tabelle 8.9: Vergleich der TMS-Werte der Gruppen „keine Versagensängste“ und „große Versagensängste“ (U-Test von Mann-Whitney)

	Gruppe 1 ^b	Gruppe 2 ^c	
Skala ^a	n = 31	n = 55	U
	MR	MR	
TMS-Wert	54,71	37,18	505,000*

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt I und III (IV) (siehe Tabelle 8.3), ^aSkala: TMS-Wert: *standardisierter TMS-Gesamtwert*, ^bdie Aussage „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“ stimmt gar nicht, ^cdie Aussage „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“ stimmt ganz genau, * $p < .01$ (2-seitig)

Von denjenigen, die angaben, dass sie Angst davor hätten, im entscheidenden Moment zu versagen, haben 20 Prozent unterdurchschnittlich im TMS abgeschnitten. Von denjenigen, die angaben, sie hätten solche Ängste gar nicht, blieb keiner unter 100 Punkten im Test.

Hypothese VI „Einfluss der Ausprägung beim Faktor Prüfungsbelastbarkeit“

Die Hypothese VI „Die Gruppe derjenigen, die bei der Zusammenfassung der vier Faktoren, die zur Definition von hoher Prüfungsbelastung summiert wurden, hohe Werte erreicht haben, unterscheidet sich in der Verteilung der Rangplätze hinsichtlich des TMS-Gesamttestwerts signifikant von der Gruppe derjenigen, die bei der Zusammenfassung der vier Faktoren, die zur Definition von hoher Prüfungsbelastung summiert wurden, niedrige Werte erreicht haben“ wird wiederum mit Hilfe des Mann-Whitney U-Tests überprüft.

Tabelle 8.10: Vergleich der TMS-Werte der Gruppen „hohe Belastbarkeit bei Prüfungen“ und „geringe Belastbarkeit bei Prüfungen“ (U-Test von Mann-Whitney)

	Gruppe 1 ^b	Gruppe 2 ^c	
Skala ^a	n = 57	n = 11	U
	MR	MR	
TMS-Wert	34,45	34,77	310,500

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt I und III (IV) (siehe Tabelle 8.3), ^aSkala: TMS-Wert: *standardisierter TMS-Gesamttestwert*, ^bWerte auf der neuen Skala von über 18, ^cWerte auf der neuen Skala von 7 bis 10

Zusammengefasst wurden für diese neue Variable „Prüfungsbelastbarkeit“ Variablen aus dem Fragebogen „Faktoren für ein gutes Abschneiden im TMS“. Hierzu wurden die Skalenwerte der vier Items aufsummiert, die Aussagen zu den Ausprägungen auf den Faktoren: „gute Nerven“, „Durchhaltevermögen“, „Ruhig-Bleiben-Können“ und „die letzten Reserven mobilisieren“ enthalten. Die neu entstandene Variable „Prüfungsbelastung“ wurde mit dem Bereichseinteiler in eine 5 stufige Variable eingeteilt. Maximal konnten auf dieser Skala 20 Punkte erreicht werden. Bei der Bildung der beiden Extremgruppen entstand so eine Gruppe von 57 Personen (12,5 Prozent der Gesamtstichprobe), deren Summenwerte auf der neuen Skala über 18 Punkten lagen und eine Gruppe von 11 Personen (2,4 Prozent der Gesamtstichprobe), deren Summenwerte sich im Bereich zwischen 7 bis 10 Punkten bewegte. Die Gruppe mit den Summenwerten über 18 wurde als die Gruppe bezeichnet, die eine hohe Belastbarkeit bei Prüfungen aufwies. Die Gruppe mit Summenwerten zwischen 7 bis 10

Punkten war die Gruppe, die eine geringe Belastbarkeit bei Prüfungen aufzeigte. Die Ergebnisse des U-Tests sind in Tabelle 8.10 dargestellt. Die Mittelwert des TMS unterschieden sich für die beiden Gruppen „hohe Belastbarkeit in Prüfungen“ und „geringe Belastbarkeit in Prüfungen“ nicht.

Hypothese VII „Einfluss des Faktors Geschlecht“

Die Hypothese VII, dass sich die empirisch gefundenen Mittelwerte im Gesamtwert des TMS und in den Untertests von Männern und Frauen systematisch voneinander unterscheiden, wird mit Hilfe eines t-Test für unabhängige Stichproben überprüft. Die Ergebnisse sind in der Tabelle 8.11 dargestellt. Die Gesamtstichprobe der Seminar Teilnehmer unterteilte sich in $n = 209$ Männer und $n = 248$ Frauen.

Der Untertest „Schlauchfiguren“ wird nicht interpretiert, da bei der Überprüfung der Voraussetzung für die Anwendung des t-Tests für unabhängige Stichproben zwei Voraussetzungen verletzt wurden: Es besteht keine Varianzhomogenität, überprüft mit dem Levene-Test, und man kann nicht von der Normalverteilung der Variable ausgehen, überprüft mit Kolmogorov-Smirnov-Test.

Die Männer schnitten im TMS-Gesamtwert besser als die Frauen ab. Die Hälfte der Männer hatte im TMS über 110 Punkte erreicht. Im Vergleich zu nur einem Drittel der Frauen, obwohl die Frauen im Abitur besser abgeschnitten hatten als die Männer. In allen Tests, in denen die empirisch gefunden Mittelwertunterschiede signifikant auf dem 1 % Niveau wurden, erzielten die Männer die besseren Ergebnisse. Die Mittelwerte lagen jeweils über den Mittelwerten der Frauen. In zwei Tests allerdings schlossen die Frauen besser ab: KONZ und FAK. Die Mittelwertunterschiede sind aber in dieser Stichprobe nicht signifikant geworden.

Tabelle 8.11: Vergleich der Mittelwerte (Standardabweichungen) TMS-Werte und acht der neun Untertest (t-Test für unabhängige Stichproben)

Skala ^a		Gruppe 1 ^b	Gruppe 2 ^c	t-Wert	df	Signifikanz (p)
		M	M			
TMS-Wert	n ₁ = 209	109,85	105,96	5,553	455	.000*
	n ₂ = 248	(7,320)	(7,583)			
MZ	n ₁ = 207	12,18	11,52	2,303	447	.022
	n ₂ = 242	(3,121)	(2,969)			
MNGV	n ₁ = 205	12,13	11,22	2,977	444	.003*
	n ₂ = 241	(3,253)	(3,193)			
QFP	n ₁ = 205	11,67	9,57	6,491	447	.000*
	n ₂ = 244	(3,516)	(3,336)			
KONZ	n ₁ = 206	17,73	18,02	-1,274	448	.203
	n ₂ = 244	(2,486)	(2,438)			
TV	n ₁ = 206	10,91	9,68	3,794	447	.000*
	n ₂ = 243	(3,475)	(3,373)			
FIG	n ₁ = 206	14,04	13,96	,275	446	.783
	n ₂ = 242	(3,393)	(3,140)			
FAK	n ₁ = 206	12,61	12,99	-1,194	447	.233
	n ₂ = 243	(3,411)	(3,331)			
DT	n ₁ = 206	12,71	10,57	6,973	446	.000*
	n ₂ = 242	(3,121)	(3,318)			

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt II und III (IV) (siehe Tabelle 8.3), ^aSkalen: TMS-Wert: *standardisierter TMS-Gesamtwert*, MZ: *Muster zuordnen*, MNGV: *Medizinisch-naturwissenschaftliches Grundverständnis*, QFP: *Quantitative und formale Probleme*, KONZ: *Konzentriertes und sorgfältiges Arbeiten*, TV: *Textverständnis*, FIG: *Figuren lernen*, FAK: *Fakten lernen*, DT: *Diagramme und Tabellen*, ^bMänner, ^c Frauen, * p < .01 (2-seitig)

Hypothese VIII „Unterscheidung der überdurchschnittlich guten Testnehmer und der unterdurchschnittlichen Testnehmer in ausgewählten Faktoren“

Die Hypothese „die Gruppe derjenigen, die „sehr gut“ im Test abgeschnitten haben, unterscheidet sich in der Verteilung der Rangplätze hinsichtlich der Faktoren: Geschlecht, Eigenständige Vorbereitung, Versagensangst, Ausprägung in den Mathematikkenntnissen und Medizin, in ihrer Dauer des Studienwunsches und in der erwarteten Abiturnote signifikant von der Gruppe derjenigen, die im Test „unterdurchschnittlich“ abgeschnitten haben“ wird ebenfalls mit dem U-Test überprüft.

Die Ergebnisse sind in Tabelle 8.12 wiedergegeben. Zu der Gruppe, die in dem Test „sehr gut“ abgeschnitten hat, gehörten all diejenigen aus dem Seminar, die einen Wert über 120 Punkte erreicht hatten. Das waren $n = 25$ Personen (5,5 Prozent der Gesamtstichprobe). In die Gruppe derjenigen, die „unterdurchschnittlich“ abgeschnitten hatten, wurden die Teilnehmer gewählt, die einen Testwert von unter 100 Punkten aufwiesen. Das waren $n = 60$ Studienanwärter (13,2 Prozent der Gesamtstichprobe). In der Gruppe der Testteilnehmer, die im Test über 120 Punkte erreicht hatten, waren 76 Prozent Männer. Genau umgekehrt stellt sich das Verhältnis in der Gruppe der Testteilnehmer dar, die unter 100 Punkten lagen. In dieser Gruppe waren 75 Prozent Frauen.

Tabelle 8.12: Vergleich der Gruppe mit sehr guten Testwerten und der Gruppe mit unterdurchschnittlichen TMS-Werten (U-Test von Mann-Whitney)

Skala ^a		Gruppe 1 ^b	Gruppe 2 ^c	U
		MR	MR	
Geschlecht	n ₁ = 25	27,70	49,38	367,500*
	n ₂ = 60			
Vorb.	n ₁ = 25	47,22	41,24	644,500
	n ₂ = 60			
Angst	n ₁ = 25	49,96	40,10	576,000
	n ₂ = 60			
Mathematik	n ₁ = 25	53,14	38,78	496,500*
	n ₂ = 60			
med. Vorwissen	n ₁ = 25	47,22	41,24	644,000
	n ₂ = 60			
Wunsch	n ₁ = 25	28,44	48,46	386,000*
	n ₂ = 59			
Abiturnote	n ₁ = 25	29,24	48,73	406,000*
	n ₂ = 60			

Anmerkungen: Alle Angaben beziehen sich auf Erhebungszeitpunkt II und III (IV) (siehe Tabelle 8.3), ^aSkalen: Geschlecht: Männer und Frauen, Vorb.: eigenständige Vorbereitung zusätzlich zu Seminar und Generalprobe, Angst: Aussage: „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“, Mathematik: Ausprägung des Faktors Mathematik, med. Vorwissen: Ausprägung medizinisches Vorwissen, Wunsch: Seit wann besteht der Studienwunsch seit kurzem, über einem Jahr oder seit mehr als drei Jahren, Abiturnote: Erwartete Abiturnote liegt bzw. dürfte im Bereich ... liegen, ^bTMS-Wert über 120 ($M = 123,56$; $SD = 2,902$), ^cTMS-Wert unter 100 ($M = 95,28$; $SD = 3,971$), * $p < .01$ (2-seitig)

Die beiden Gruppen unterscheiden sich nicht auffällig hinsichtlich ihrer Vorbereitungszeit, die sie neben Seminar und Generalprobe aufgebracht haben. Von denjenigen, die „sehr gut“ im Test abschnitten, meinten 64 Prozent, dass ihre Mathematikkenntnisse gut und besser ausgeprägt seien. Das behaupteten andererseits von denjenigen, die unter-

durchschnittlich im TMS abgeschnitten hatten, nur 30 Prozent. Tendenziell kann man sagen, dass unter denjenigen, die „schlecht“ im TMS abschnitten, auch mehr Personen mit Versagensängsten waren. Die Aussage „Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen“ würde für sie „genau stimmen“ oder „stimmen“ kreuzten hier 48 Prozent der Teilnehmer an. Im Gegensatz dazu haben dieser Aussage nur 28 Prozent in der Gruppe derjenigen, die den TMS mit über 120 Punkten abgeschlossen hatten, zugestimmt. Dennoch wurde der Unterschied nicht signifikant. In ihrer Einschätzung ihres medizinischen Vorwissens unterschieden sich die beiden Gruppen nicht signifikant. Interessanterweise hatte die Gruppe, die unter 100 Punkte im TMS-Gesamttest erreicht hatte, viel länger den Wunsch, ein Medizinstudium aufzunehmen als die Gruppe, die mit über 120 Punkten so viel besser abschnitt. In der schlechteren Gruppe hatten 70 Prozent bereits seit mehreren Jahren den Wunsch, ein Medizinstudium aufzunehmen. In der überdurchschnittlich guten Gruppe verfolgten nur 28 Prozent diesen Wunsch seit mehreren Jahren. Im Abiturdurchschnitt unterschieden sich die beiden Gruppen ebenfalls signifikant. Die Gruppe mit dem guten TMS-Ergebnis hatte auch den besseren Abiturdurchschnitt zu erwarten bzw. bereits gemacht: 96 Prozent hatten einen Abiturdurchschnitt von maximal 2,60 erreicht bzw. zu erwarten. Die fehlenden 4 Prozent gingen einzig allein auf eine Person zurück, die keine Angaben bezüglich des Abiturdurchschnitts gemacht hatte. In der Gruppe mit den unterdurchschnittlichen TMS-Resultat hatten ca. 67 Prozent maximal einen Abiturdurchschnitt von 2,60 erreicht bzw. zu erwarten. Einen schlechteren Abiturdurchschnitt erwarteten bzw. hatten bereits 30 Prozent aus dieser Gruppe.

9 Diskussion und Fazit

Diese Arbeit hatte zwei Zielsetzungen. Auf der einen Seite sollte das Testtraining der CCH-Laufbahnberatung evaluiert werden und auf der anderen Seite sollten anhand einer ausgewählten Spezialstichprobe mögliche Einflussfaktoren der Testleistung im TMS aufgezeigt werden. Der Ausgangspunkt dieser Arbeit war die Veränderung des Hochschulrahmengesetzes und die damit verbundene Freiheit der Hochschulen einen großen Teil ihrer zukünftigen Studenten nach eigenen Auswahlkriterien zum Studium zuzulassen. Ein mögliches Auswahlverfahren ist der Studieneignungstest. In dieser Arbeit wurde der Test für medizinische Studiengänge (TMS) als spezifischer Studierfähigkeitstest vorgestellt. Das Thema Studierfähigkeitstest beinhaltet auch immer die Frage der Trainierbarkeit eines solchen Tests. Der theoretische Teil der Arbeit bezieht sich auf den Effekt von Testtrainings und die empirische Überprüfbarkeit dieses Effekts. Zunächst wurde die Literatur bezüglich der Trainierbarkeit des TMS analysiert. Die Literaturrecherche wurde dann um die Trainingsliteratur im angloamerikanischen Sprachraum zum Scholastic Aptitude Test (SAT) erweitert.

Zunächst einmal wurde in dieser Arbeit das ausführlich vorgestellte Testtraining der CCH-Laufbahnberatung evaluiert. Die Wirksamkeit eines Testtrainings ist nach dem Vier-Ebenen-Modell nicht nur in einem Endresultat, also im TMS-Ergebnis, zu erkennen, sondern der Lernerfolg kann auf insgesamt vier Stufen erfasst werden. Die Ergebnisse zeigten deutlich, dass die Seminarbesucher auf der Reaktionsebene sehr zufrieden waren. Der größte Teil der Trainingsteilnehmer bewertete das Trainerverhalten als vollkommen oder weitgehend engagiert, kompetent und sympathisch. Diese Zufriedenheit ist auf jeden Fall ein gutes Zeichen für eine angenehme Lernatmosphäre. Die Studienbewerber hatten Zeit und Geld investiert und erwarteten gute Betreuung, die sie nach Meinung der meisten auch erhalten hatten. Die Wirkung des Trainingsprogramms auf der Lernebene zu beurteilen, war schwieriger. Zur Messung der reinen Wissensvermehrung lagen keine geeigneten Daten vor. Ein Pre-Posttest-Design war aufgrund von fehlenden Ergebnissen aus Vor- und Nachtest nicht möglich. Gemessen werden konnte allerdings, ob die Trainingsteilnehmer ihre Einstellung gegenüber Prüfungen allgemein und dem TMS im speziellen geändert

haben. Es konnte eine leichte Entspannung hinsichtlich Prüfungen und dem TMS festgestellt werden. Diese Einstellungsänderung geht sicher zum größten Teil darauf zurück, dass die Seminarbesucher mit der auf sie zukommenden Testsituation vertraut gemacht wurden. Zunächst einmal wurden im Laufe des Seminars mehrere Testsimulationen durchgeführt, dann wurden die Untertests, die dazugehörigen Instruktionen und Lösungswege eingehend bearbeitet. Mit dem Vertrautwerden änderte sich auch die Einstellung der Teilnehmer gegenüber Prüfungen und natürlich speziell dem TMS gegenüber. Die Verhaltensebene, die sich mit dem Transfer des Gelernten in die Praxis beschäftigte, zeigte ebenfalls positive Ergebnisse für das Interventionsprogramm. Mit Praxis war in dieser Situation die konkrete Testsituation des TMS gemeint. Im Vergleich der Nützlichkeit der vermittelten Strategien, Techniken, Materialien etc. zeigte die Testsimulation die deutlichste Zustimmung. Das spricht auch wieder für die schon auf der Lernebene gefundenen Einstellungsänderungen, die wahrscheinlich zum größten Teil auf die wachsende Vertrautheit mit dem TMS zurückzuführen ist. Die Analyse der von den Kursteilnehmern eingesandten Testergebnisse, die auf der Erfolgsebene erfolgte, zeigte eindeutig, dass die Trainingsteilnehmer im Mittel besser abgeschnitten hatten als die Gesamtheit der TMS-Teilnehmer. Die Wirkung des Trainings konnte auch in den Mittelwerten der neun Untertests bestätigt werden. Die erwarteten höheren Werte in den kurzfristig trainierbaren Untertests konnten deutlich aufgezeigt werden. Wobei die Seminarteilnehmer auch in den als nicht kurzfristig trainierbaren Untertests im Mittel überdurchschnittlich abgeschnitten hatten. Die Evaluation des Testtrainings weist im Ganzen darauf hin, dass das Trainingsprogramm für die Teilnehmer zum erhofften Erfolg führte.

Die zweite Zielsetzung der Arbeit war die Untersuchung von Einflussfaktoren, die neben dem Faktor Testtraining bzw. Vorbereitung, auf die Testleistung wirken. Bei der Literaturrecherche war deutlich geworden, dass die eindeutige Bestimmung der Größe der Trainingseffekte aufgrund der methodischen Besonderheiten nicht immer möglich ist. Mit dieser wurde versucht, die Einschränkungen einer solchen Analyse darzustellen. Da hier nur ein Testtraining zu einem Test vorgestellt, daher waren die Randbedingungen, d.h. Inhalt, Umfang und Qualität des Trainings für die beobachtete Gruppe gleich. So war es möglich Hypothesen darüber aufzustellen, welche Faktoren

neben dem Testtraining einen Einfluss auf die Testleistung haben könnten. Da es sich um eine explorative Studie handelte, konnte der Einfluss möglicher Faktoren mit Hilfe von künstlich gebildeten Extremgruppen untersucht werden, d.h. es wurden immer nur zwei Gruppen bezüglich einer abhängigen Variable miteinander verglichen, die in ihrer Ausprägung in der interessierenden unabhängigen Variable absolut entgegengesetzt waren. Es wurde so überprüft, ob eine erklärende Variable einen potenziellen Ursachenwert für eine abhängige Variable hat. Wichtig ist es hierbei zu beachten, dass es bei dieser Art von Untersuchung immer dazu kommen kann, dass die Ergebnisse die Wichtigkeit der unabhängigen Variable überschätzen (Bortz & Döring, 2006).

Die Hypothese von Fay (1985) und Briggs (2002), dass neben der Art der Vorbereitung noch andere Faktoren maßgeblichen Einfluss auf die Testleistung haben, wurde durch die Ergebnisse mehrfach bestätigt. Jedoch waren nicht alle vorher aufgestellten explorativen Hypothesen zutreffend.

Fay (1985) hatte neben anderen Faktoren die Abiturnote als Einflussvariable ausgemacht. Die Ergebnisse der Hypothese I bestätigten zum Teil diesen Einfluss. Die beiden Gruppen unterschieden sich signifikant in den als kurzfristig nicht so gut trainierbar angesehenen Untertests: MNGV, QFP, TV, DT und im Gesamttest. Die Ergebnisse der statistischen Überprüfung konnten hingegen keinen signifikanten Unterschied zwischen den beiden Abitextremgruppen bei den als kurzfristig gut trainierbar geltenden Untertests: MZ, SF, KONZ und FIG finden. Einzig allein im als kurzfristig gut trainierbar geltenden Untertest FAK war der Unterschied zwischen den beiden Extremgruppen signifikant. Die Ergebnisse der Hypothese I erbrachte die erwarteten Signifikanzen. Es wird davon ausgegangen, dass die Abiturdurchschnittsnote ein guter Prädiktor für das Abschneiden im TMS ist. Diese Vermutung konnte mit den Ergebnissen bestätigt werden. Gerade in den als kurzfristig nicht trainierbaren Untertests schnitten die Bewerber besser ab, die einen besseren Abiturdurchschnitt hatten. Hingegen waren die Unterschiede in der Verteilung der Rangplätze bei den als kurzfristig trainierbar geltenden Untertests, mit Ausnahme des FAK, nicht signifikant. Dieses Ergebnis spricht dafür, dass in diesem Fall auch die schlechten Abiturienten durch das Testtraining ihre Leistung insoweit verbessern konnten, dass sie sich nicht mehr von den guten Abiturienten signifikant unterschieden. Die Ergebnisse der

statistischen Überprüfung der Hypothese I bestätigte somit die Vermutung sowohl von Deter (1982) als auch von Fay (1985), dass einige Tests leichter kurzfristig trainierbar waren aber auch dass, wie Fay (1985) angenommen hat, höchstwahrscheinlich die Abiturnote einen Einfluss auf das Testergebnis hat. In den kurzfristig nicht trainierbaren Untertests haben die Teilnehmer mit den besseren Abiturnoten bedeutend besser abgeschnitten.

Nicht bestätigt werden konnte die Hypothese II, dass die Dauer der zusätzlichen Vorbereitung, einen Einfluss auf die Testresultate hat. Somit konnte der von Messick (1981) gefundene Zusammenhang nicht bestätigt werden. Das lag aber sicher auch daran, dass die Vorbereitungszeit in diesem Fall größtenteils in dem Seminar stattgefunden hat und somit quasi für alle gleich war. Die zusätzliche Vorbereitungszeit neben dem Seminar und der Generalprobe hatte scheinbar keinen leistungssteigernden Effekt.

Fay (1985) vermutete, dass auch die gewählten Interessen der Bewerber eventuell einen Einfluss auf das Testergebnis haben könnten. Interesse wurde in dieser Untersuchung operationalisiert als „Kenntnisse in Mathematik“ und „medizinisches Vorwissen“. Die Hypothese III konnte bestätigt werden. Diejenigen, die sehr gut ausgeprägte Kenntnisse in Mathematik hatten, waren signifikant besser in den mathematisch orientierten Untertests als diejenigen, die kein Interesse an diesem Fach hatten. Allerdings konnte die Hypothese IV für das Interessengebiet Medizin nicht bestätigt werden. Hier kam es zu keinen signifikanten Unterschieden in dem medizinisch orientierten Untertest. Das bedeutet auch gleichzeitig, dass die Testkonstrukteure des TMS mit ihrer Behauptung in der Testbroschüre Recht hatten, dass es für diesen Untertest nicht notwendig sei, medizinische Vorkenntnisse zu haben.

Im Zuge des explorativen Designs wurden hier auch die beiden Extremgruppen „hohe Versagensangst“ und „keine Versagensangst“ bezüglich ihres Testergebnisses miteinander verglichen. Diese Hypothese V konnte bestätigt werden. Die beiden Gruppen unterschieden sich signifikant. Die Gruppe mit den größeren Versagensängsten hatte auch tendenziell das schlechtere Abitur. Hingegen konnte sich die rein explorativ durchgeführte Hypothese VI, dass sich die Gruppe derjenigen, die sich als sehr prüfungsbelastbar einschätzt, im Vergleich zu der Gruppe, die sich als

kaum belastbar einschätzt, in ihren Ergebnissen signifikant unterscheiden nicht bestätigt werden.

Fay (1985) vermutete ebenfalls, dass auch das Geschlecht einen möglichen erklärenden Einfluss auf das Testergebnis haben könnte. Diese Hypothese VII wurde mit dem t-Test für unabhängige Stichproben gerechnet, da die beiden Gruppen einen genügend großen Stichprobenumfang aufwiesen. Die Hypothese konnte für den Gesamttest und vier Untertests bestätigt werden. Hier lagen die Männer in allen Untertests, die als nicht kurzfristig trainierbar gelten vorne: MNGV, QFP, TV und DT über den Werten der Frauen. Interessanterweise hatten die Frauen tendenziell das bessere Abitur. Dies stützt nicht die Annahme, dass diejenigen, die bessere Abiturnoten haben, gerade in den nicht-trainierbaren Untertests besser abschneiden sollten. Es könnte aber vielleicht daran liegen, dass Frauen tendenziell mehr Versagensängste haben als Männer.

Bei einem Vergleich der beiden Extremgruppen „überdurchschnittlich gutes Testergebnis“ und „unterdurchschnittliches Testergebnis“ wurden die vorher bestätigten Hypothesen in gewisser Weise noch einmal bestätigt. Es gab mehr Männer als Frauen in der guten Gruppe. Der Unterschied war signifikant. Auch die mathematischen Kenntnisse hatten einen hohen Einfluss, diesmal auch auf das Gesamtergebnis. Und auch der Einfluss der Abiturdurchschnittsnote wurde als guter Prädiktor für das Testergebnis noch einmal bestätigt. Nicht belegt werden konnte der vorher gefundene Zusammenhang zwischen Versagensangst und Testergebnis. Interessant war, dass die Dauer des Studienwunschs in die entgegen gesetzte Richtung wies als vermutet. Die Bewerber mit dem schlechten Testergebnis hatten schon viel länger den Wunsch zum Medizinstudium als diejenigen mit einem guten Testergebnis. Dies kann natürlich ein Ergebnis bedingt durch die gewählte Stichprobe sein, aber eventuell auch ein Hinweis darauf, was für einen enormen Druck ein Studieneignungstest für manche Bewerber bedeutet. Gerade wenn man sich jahrelang auf einen solchen Test vorbereitet, ist der Moment der Testabnahme vielleicht auch eine zu starke psychische Belastung, so dass im entscheidenden Moment nicht mehr die volle Leistung erbracht werden kann.

Da sich die guten und schlechten Abiturienten in den kurzfristig trainierbaren Tests nicht signifikant unterscheiden, kann man vermuten, dass das Testtraining für die schlechten Abiturienten einen positiven Effekt hatte, sowie es das sicher auch für die

guten Abiturienten hatte. Da zum Medizinstudium eine starke Bereitschaft gehört sich anzustrengen, sich zu motivieren und seine Potentiale auszunutzen kann man in dieser Untersuchung gut erkennen, dass auch Abiturienten mit einem weniger guten Abiturdurchschnitt in der Lage sind, diese Qualitäten zu zeigen. Sowohl die guten als auch die weniger guten Abiturienten unterschieden sich in den trainierbaren Untertests nicht signifikant. Beide hatten das gleiche Trainingsprogramm durchlaufen und somit gleich große Trainingseffekte. Die schlechten Abiturienten konnten in der Kürze der Zeit große Wissenslücken nicht mehr ausgleichen, aber sie konnten durch Motivation und Anstrengung zumindest in den trainierbaren Untertests mit den guten Abiturienten gleichziehen.

Die Ergebnisse dieser Untersuchung können nicht auf die Population der Studienanwärter generalisiert werden. Sie müssen im engen Zusammenhang mit der untersuchten Stichprobe gesehen werden, da es sich um eine rein explorative Arbeit handelt. Die gefundenen Signifikanzen beziehen sich alle auf die Stichprobe der Seminarteilnehmer. Allerdings zeigen die Ergebnisse trotz ihres vorläufigen Charakters mögliche Tendenzen, welche Faktoren einen Einfluss auf das Resultat eines Studierfähigkeitstests haben können.

Fazit

Zunächst einmal stand im Vordergrund die Evaluation des beschriebenen Testtrainings. Dieses Interventionsprogramm kann anhand der gefundenen Ergebnisse als erfolgreich bewertet werden. Für zukünftige Untersuchungen wäre es interessant zwei Testtrainings, die auf den gleichen Test vorbereiten, anhand von interessanten Fragestellungen miteinander zu vergleichen. Dies kann, wie in dieser Untersuchung, auch mit dem Vier-Ebenen-Modell von Kirkpatrick geschehen, so dass jede einzelne Ebene für sich verglichen werden kann. Es ist nicht immer gesagt, dass die Zufriedenheit der Trainingsteilnehmer auch einher geht mit einem Lernerfolg. Sollten die Universitäten zukünftig mehr Studierfähigkeitstest bei der Auswahl ihrer Studierenden einsetzen, dann wird es sicher auch vermehrt zu Trainingskursangeboten kommen. Zukünftige Studenten haben dann die Möglichkeit sich mit Hilfe solcher Kurse vorzubereiten, aber gleichzeitig besteht auch immer die Gefahr, dass ein Trainingsprogramm nicht das hält, was es verspricht. Insofern müsste es für seriöse

Anbieter von Trainingsprogrammen selbstverständlich sein, dass sie ihr Kursangebot evaluieren, um den Studienbewerbern die Möglichkeit zu bieten, das für sie geeignetste Programm mit Hilfe von objektiven Daten auszuwählen.

Bei der Untersuchung der Spezialstichprobe wurden deutliche Hinweise darauf gefundenen, dass es weitere Einflussfaktoren, neben der Vorbereitung, für die Leistung im TMS gibt: Geschlecht, Abiturdurchschnittsnote und Kenntnisse in Mathematik. Diese schon in anderen Untersuchungen zum Teil bestätigten Einflussfaktoren könnten dazu führen, dass ein Test als nicht fair gilt. Die Hochschulen müssen sich vor der Einführung eines Studierfähigkeitstest darüber in Klaren sein, dass es neben dem viel gescholtenen Testtraining auch immer noch andere Einflussfaktoren auf die Testleistung geben kann, die eine ganze Gruppe von Testnehmern benachteiligen könnte.

Um der Frage nach der Größe des Testtrainingseffekts auf den Grund zu gehen, sollten eventuell weitere Untersuchungen in einem experimentellen Design durchgeführt werden. Ein weiterer spannender Aspekt ist die Frage, ob die im Coaching erlernten Fähigkeiten sich auch ins Studium transferieren lassen. Es wäre interessant zu untersuchen, ob sich gecoachte Studenten in ihrer Leistung im Studium oder auch in ihrer Lern- und Arbeitshaltung von anderen Studenten unterscheiden.

Abschließend ist zu der Studie über das hier vorgestellte Testtraining zum TMS anzumerken, dass ca. 80 Prozent der Seminarteilnehmer überdurchschnittliche Testergebnisse erreicht haben. Ob die Resultate ähnlich gut gewesen wären, wenn die Testnehmer sich ohne Anleitung vorbereitet hätten, bleibt zu beweisen. Man sollte immer bedenken, dass eine eigenständige Vorbereitung enorme Disziplin erfordert, die viele Menschen nicht in der Lage sind, alleine aufzubringen. Darüber hinaus könnte gerade ein Testtraining den Teilnehmern helfen, die unter anderen Umständen unter ihren Möglichkeiten geblieben wären.

10 Literatur

- Allalouf, A. & Ben-Shakhar, G. (1998). The Effect of Coaching on the Predictive Validity of Scholastic Aptitude Tests. *Journal of Educational Measurement*, 35(1), 31-47.
- Becker, B. J. (1990). Coaching for the Scholastic Aptitude Test: Further Synthesis and Appraisal. *Review of Educational Research*, 60(3), 373-417.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler : mit 87 Tabellen* (4., überarb. Aufl.). Heidelberg: Springer.
- Briggs, D. C. (2001). The Effect of Admission Test Preparation: Evidence from NELS:88. *Chance*, 14(1), 10-18.
- Briggs, D. C. (2002). *SAT coaching, bias and causal inference*. Dissertation Abstracts International, University of California-Berkeley.
- Briggs, D. C. (2009). Preparation for College Admission Exams. 2009 NACAC Discussion Paper. *National Association for College Admission Counseling*, 33pp.
- Bühl, A. (2006). *SPSS 14 : Einführung in die moderne Datenanalyse ; [Klassifikationsanalyse]* (10., überarb. und erw. Aufl.). München [u.a.]: Pearson Studium.
- Deidesheimer Kreis. (1997). *Hochschulzulassung und Studieneignungstests : studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus-clausus- und andere Studiengänge*. Göttingen <<[u.a.]>>: Vandenhoeck & Ruprecht.
- Deter, B. (1982). *Zum Einfluß von Übung und Training auf die Leistung im "Test für medizinische Studiengänge" (TMS)*. Zugl Mannheim Univ Diss 1982, Agentur Pedersen, Braunschweig.
- Fay, E. (1985). Vorbereitungsmöglichkeiten auf den "Test für medizinische Studiengänge": Was gibt es? Wie wird es genutzt? Nutzt es? *Auswertung zum achten und neunten Testtermin und Ergebnisse weiterer Begleituntersuchungen zum TMS*, 86-116.
- Gollwitzer, M. & Jäger, R. S. (2007). *Evaluation : Workbook* (1. Aufl.). Weinheim [u.a.]: Beltz, Psychologie-Verl.-Union.
- Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995). Der Eignungstest für das Medizinstudium in der Schweiz. *Schweizerische Ärztezeitung*, 37.
- Häring, K. (2003). *Evaluation der Weiterbildung von Führungskräften Anspruch und Realität des Effektivitätscontrolling in deutschen Unternehmen*. Wiesbaden: Deutscher Universitäts-Verl.
- Heine, C. (2006). *Bestandsaufnahme von Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern*. Hannover.

- Hell, B. T. S. S. H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. A metaanalytic investigation of discipline-specific admission tests in German-speaking countries. *Empirische Paedagogik*, 21(3), 251-270.
- Hofer, R. H., K.-D. (1996). Zur Trainierbarkeit eines Eignungstests für das Medizinstudium. . *Tagungsband. Bericht des ZTD Band 3*.
- Hohner, H.-U. (1991). Der Medizinertest und die Zulassung zum Medizinstudium. *Psychomed*, 3, 47-52.
- Hohner, H.-U. (1995). *CCH-Test-Info*: CCH Dr. Hohner.
- Hohner, H.-U. (2009). Das CCH-Testtraining.
- Informationsbroschüre. (2009). *Test für medizinische Studiengänge TMS 2009*
- Karnaoukh, K. (2008). *Der Zusammenhang von Prüfungsangst, Vorbereitung und Leistung am Beispiel des Tests für Medizinische Studiengänge*. Diplomarbeit, Universität Koblenz-Landau, Koblenz.
- Kirchenkamp, T. & Mispelkamp, H. (1988). Beziehungen zwischen Leistungen im Test für medizinische Studiengänge und verschiedene Vorbereitungsmaßnahmen, Einstellungen und Vergabeverfahren sowie links- bzw. rechtshändiger Schreibweise. *Test für medizinische Studiengänge (TMS) Studien zur Evaluation 12. Arbeitsbericht*, 248-279.
- Kirkpatrick, D. L. (2004). *Evaluating training programs : the four levels* (2. ed., [Nachdr.]). San Francisco, Calif.: Berrett-Koehler.
- Kuckartz, U., Ebert, T., Rädiker, S. & Stefer, C. (2009). *Evaluation online : Internetgestützte Befragung in der Praxis*. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden.
- MacGowan, B. R. (2002). Test prep and the SAT I: The activities and their effects (Bd. <http://home.comcast.net/~bmacgowan/wsb/problems.html>).
- Maue, P. (1998). *Faktoren für ein erfolgreiches Abschneiden in standardisierten Hochschuleingangstests aus Sicht der Kandidaten*. Diplomarbeit, Freie Universität Berlin, Berlin.
- Messick, S. & Jungeblut, A. (1981). Time and Method in Coaching for the SAT. *Psychological Bulletin*, 89(2), 191-216.
- Nachtigall, C. & Wirtz, M. (2006). *Wahrscheinlichkeitsrechnung und Inferenzstatistik. Statistische Methoden für Psychologen* (4., überarb. Aufl. Bd. 2).
- Powers, D. E. a. R., Donald A. . (1999). Effects of Coaching on SAT : Reasoning Test Scores. *Journal of Educational Measurement*, 36(2), 93-118.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006a). *Quantitative Methoden 2 Einführung in die Statistik* (2., erweiterte Auflage). Berlin, Heidelberg: Springer Medizin Verl.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2006b). *Quantitative Methoden Einführung in die Statistik* (2., erweiterte Auflage). Berlin, Heidelberg: Springer Medizin Verl.
- Ringlstetter, M. J., Aschenbach, M. & Kirsch, W. (2003). *Perspektiven der strategischen Unternehmensführung Theorien, Konzepte, Anwendungen ; Werner Kirsch zum 65. Geburtstag*. Wiesbaden: Gabler.

- Roeder, O. (1998). *Prüfungsbelastung und Leistung bei standardisierten Hochschuleingangstests*. Diplomarbeit, Freie Universität Berlin, Berlin.
- Schuler, H. (2008). *Studierendenauswahl und Studienentscheidung*. Göttingen u.a.: Hogrefe.
- Tarazona, M. (2006). Berechtigte Hoffnung auf bessere Studierende durch hochschuleigene Studierendenauswahl? Eine Analyse der Erfahrungen mit Auswahlverfahren in der Hochschulzulassung. *Beiträge zur Hochschulforschung*, 2(28).
- Trost, G. (1998). *Evaluation des Tests für medizinische Studiengänge (TMS): Synopse der Ergebnisse*. Bonn: Institut für Test- und Begabungsforschung.
- Trost, G. (2005a). Studierendenauswahl durch die Hochschulen: Welche Schritte sind zu tun? *Psychologische Rundschau*, 56(2), 140-142.
- Trost, G. (2005b). Studierendenauswahl durch die Hochschulen: Welche Verfahren kommen prinzipiell in Betracht, welche nicht? *Psychologische Rundschau*, 56(2), 138-140.
- Wissenschaftsrat. (2004). *Empfehlungen und Stellungnahmen* (Bd. 2004 -). Köln.

11.2 Fragebogen zur Prüfungserwartung

FRAGEBOGEN ZUR ALLGEMEINEN UND SPEZIFISCHEN PRÜFUNGSERWARTUNG
 FB-ERWART CCH 6/07

Name:
 Geburtsdatum:
 CCH-TEST-TRAINING in _____

Die folgenden Fragen beziehen sich auf Ihre Orientierungen und Einstellungen zu wichtigen Prüfungen im Allgemeinen und auf den Test für medizinische Studiengänge im besonderen. Ihre Beantwortung kann bedeutsame Hinweise darauf liefern, in welcher Weise Sie mit Prüfungen umgehen. Bitte beantworten Sie die folgenden Fragen zügig. Kreuzen Sie diejenige Alternative an, die Ihnen spontan in den Sinn kommt. Es gibt bei diesen Fragen keine richtigen und keine falschen, keine erwünschten und keine unerwünschten Antworten.

Fragen zur allgemeinen Prüfungserwartung

Denken Sie bei der Beantwortung dieser Fragen bitte an Prüfungen wie wichtige Klausuren, an die Führerscheinprüfung oder an Abiturteilprüfungen, also an Prüfungen, von denen für Sie doch einiges abhängt.

1) Wie gut man in Prüfungen abschneidet hängt davon ab, wie intensiv man sich darauf vorbereitet.

5 stimmt	4 stimmt	3 stimmt	2 stimmt	1 stimmt
genau		in etwa	kaum	gar nicht

2) Ein schlechtes Prüfungsergebnis hängt in der Regel mit mehreren Faktoren zusammen.

5 stimmt	4 stimmt	3 stimmt	2 stimmt	1 stimmt
genau		in etwa	kaum	gar nicht

3) Gute Leistungen sind das Resultat eigener Anstrengungen.

5 stimmt	4 stimmt	3 stimmt	2 stimmt	1 stimmt
genau		in etwa	kaum	gar nicht

4) Auf wichtige Prüfungen bereite ich mich gründlich vor.

5 stimmt	4 stimmt	3 stimmt	2 stimmt	1 stimmt
genau		in etwa	kaum	gar nicht

5) Das Resultat von Prüfungen hängt häufig entscheidend von Zufällen und Unwägbarkeiten ab.

1 stimmt	2 stimmt	3 stimmt	4 stimmt	5 stimmt
genau		in etwa	kaum	gar nicht

6) Auch wichtigen Prüfungen sehe ich gelassen entgegen.

5 stimmt	4 stimmt	3 stimmt	2 stimmt	1 stimmt
genau		in etwa	kaum	gar nicht

7) Ich habe bei Prüfungen manchmal Angst davor, im entscheidenden Moment zu versagen.

1 stimmt	2 stimmt	3 stimmt	4 stimmt	5 stimmt
genau		in etwa	kaum	gar nicht

- 8) Prüfungen gut zu bestehen hängt nicht nur vom eigenen Wissen und Können ab, sondern auch davon, ob man seine Fähigkeiten zum richtigen Zeitpunkt voll einsetzen kann.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 9) Wenn ich weiß, was mich erwartet, dann kann ich einer Prüfung relativ gelassen entgegensehen.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 10) Grundsätzlich hängt mein Abschneiden in Prüfungen von verschiedenen Faktoren ab. Einige kann ich beeinflussen, z.B. indem ich mich gut vorbereite, andere dagegen nicht.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 11) Was immer ich tun kann, um mich optimal auf Prüfungen vorzubereiten, das tue ich in der Regel auch.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

Punktzahl (Fragen 1-11):

Fragen zum Test für Medizinische Studiengänge

Die folgenden Fragen beziehen sich auf den TMS/EMS bzw. auf die Zulassung zum Medizinstudium.

- 12) Dass ich ein möglichst gutes Resultat im TMS/EMS erziele ist für mich von höchster Bedeutung.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 13) Ich habe ein sehr ungutes Gefühl, wenn ich bedenke, dass ich den TMS nur ein einziges Mal absolvieren kann.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 14) Ich bin recht unsicher über mein Abschneiden im TMS.

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

- 15) Eigentlich bin ich recht zuversichtlich was meine Zulassungschancen zum Medizinstudium anbelangt.

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

- 16) Ein gutes Resultat im TMS/EMS ist das Ergebnis aus konsequenter Vorbereitung, guten Fähigkeiten, der nötigen Gelassenheit und einem Quantum Glück.

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

Punktzahl (Fragen 12-16):

11.3 Fragebogen „Faktoren für ein gutes Abschneiden im TMS“

Anhang XI

Faktoren für ein gutes Abschneiden im TMS

Bitte schätzen Sie im folgenden ein, welche Faktoren für ein optimales Abschneiden im TMS Ihrer Meinung nach wie wichtig sind

gute Kenntnisse in Mathematik	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
gutes Schulwissen	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
gute Nerven	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
ausreichende Vorbereitung	1	2	3	4	<input checked="" type="checkbox"/> 5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
medizinisches Vorwissen	1	<input checked="" type="checkbox"/> 2	3	4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
Bereitschaft sich anzustrengen	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
Durchhaltevermögen	1	2	3	4	<input checked="" type="checkbox"/> 5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
Ruhig-Bleiben-Können	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
die letzten Reserven mobilisieren	1	2	3	4	<input checked="" type="checkbox"/> 5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>
sonstiges (bitte bezeichnen):	1	2	3	4	5
	<small>absolut unwichtig</small>	<small>unwichtig</small>	<small>mäßig wichtig</small>	<small>wichtig</small>	<small>absolut wichtig</small>

Bitte schätzen Sie nun ein, in welcher Weise diese Faktoren bei Ihnen ausgeprägt sind

gute Kenntnisse in Mathematik	<input checked="" type="checkbox"/> 1	2	3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
gutes Schulwissen	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
gute Nerven	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
ausreichende Vorbereitung	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
medizinisches Vorwissen	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
Bereitschaft sich anzustrengen	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
Durchhaltevermögen	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
Ruhig-Bleiben-Können	1	2	<input checked="" type="checkbox"/> 3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
die letzten Reserven mobilisieren	1	2	3	<input checked="" type="checkbox"/> 4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>
sonstiges (bitte bezeichnen):	1	2	3	4	5
	<small>gar nicht ausgeprägt</small>	<small>kaum ausgeprägt</small>	<small>mäßig ausgeprägt</small>	<small>gut ausgeprägt</small>	<small>sehr gut ausgeprägt</small>

© CCH Laufbahnberatung Dr. Hohner Berlin

11.4 Fragebogen zur Testbefindlichkeit

FRAGEBOGEN ZUR TESTBEFINDLICHKEIT 07FBBEFIND och 6/07

Name:
heutiges Datum:

Bitte beantworten Sie die folgenden Fragen spontan und ohne allzulanges Nachdenken. Es geht darum, Ihre jetzige Befindlichkeit und Ihren aktuellen Testeindruck zu erfassen. Dadurch können wichtige Hinweise gewonnen werden für Ihre optimale Vorbereitung auf eine stundenlange Prüfungssituation. Es gibt keine richtigen und keine falschen, keine günstigen und keine ungünstigen Antworten.

I Fragen zu Ihrer aktuellen körperlichen und geistigen Befindlichkeit nach absolviertem Test

- 1) Im Moment fühle ich mich ausgelaugt

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 2) Ich kann mich auch jetzt sehr gut konzentrieren

1 stimmt genau	2 stimmt	3 stimmt in etwa	4 stimmt kaum	5 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 3) Im Moment habe ich Kopfschmerzen

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 4) Im Moment tun mir die Augen weh

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 5) Ich merke, dass meine Schultern verspannt sind

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 6) Mein Nacken schmerzt

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 7) Mein Kopf ist absolut leer

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 8) Auch jetzt fühle ich mich ziemlich frisch

1 stimmt genau	2 stimmt	3 stimmt in etwa	4 stimmt kaum	5 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 9) Ich habe einen starken Drang, mich jetzt zu bewegen

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------
- 10) Ich merke, dass ich gereizt bin

5 stimmt genau	4 stimmt	3 stimmt in etwa	2 stimmt kaum	1 stimmt gar nicht
-------------------	----------	---------------------	------------------	-----------------------

11) Ich bin im Moment ganz zufrieden

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

12) Jetzt bin ich recht abgeschlafft

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

Punktwert (Fragen 1 - 12):

II Fragen zur Testsituation

13) Bestimmte Geräusche haben mich immer wieder gestört
(z.B. Naseputzen, Klimaanlage, Straßenlärm...)

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

14) Ich wusste immer wieviel Zeit mir noch zur Verfügung stand

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

15) Es störte mich, wenn mir der Trainer über die Schulter
schaute

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

16) Durch äußere Einflüsse ließ ich mich nicht ablenken

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

17) Mit dem Licht, der Helligkeit, der Beleuchtung hatte
ich keinerlei Schwierigkeiten

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

Punktwert (Fragen 13 - 17):

III Fragen zum Test und zu Ihren Testeindrücken

18) Ich denke, dass ich insgesamt recht gut abgeschnitten habe

1 stimmt 2 stimmt 3 stimmt 4 stimmt 5 stimmt
genau in etwa kaum gar nicht

19) Im Verlaufe des Tests hat meine Konzentrationsfähigkeit
abgenommen

5 stimmt 4 stimmt 3 stimmt 2 stimmt 1 stimmt
genau in etwa kaum gar nicht

20) Im großen und ganzen fand ich die Testaufgaben nicht sehr schwer

1 stimmt genau 2 stimmt 3 stimmt in etwa 4 stimmt kaum 5 stimmt gar nicht

21) Ich hatte bei einigen Aufgaben große Probleme mit dem Zeitdruck

5 stimmt genau 4 stimmt 3 stimmt in etwa 2 stimmt kaum 1 stimmt gar nicht

22) Mit der Zeiteinteilung hatte ich keine großen Probleme

1 stimmt genau 2 stimmt 3 stimmt in etwa 4 stimmt kaum 5 stimmt gar nicht

23) Bei manchen Aufgaben wurde ich immer unruhiger

5 stimmt genau 4 stimmt 3 stimmt in etwa 2 stimmt kaum 1 stimmt gar nicht

Punktwert (Fragen 19 - 23):

Bitte schätzen Sie nun Ihre Leistung für die einzelnen Untertests ein:

	keine Probleme	geringe Probleme	große Probleme	sehr große Probleme	
Muster zuordnen	1	2	3	4	5
Medizin. nat.wiss. Grundverständnis	1	2	3	4	5
Schlauchfiguren	1	2	3	4	5
Quantitative und formale Probleme	1	2	3	4	5
Konzentrationstest	1	2	3	4	5
- Mittagspause -					
Textverständnis	1	2	3	4	5
Figuren lernen	1	2	3	4	5
Fakten lernen	1	2	3	4	5
Diagramme und Tabellen	1	2	3	4	5

Punktwert:

Folgende Aspekte des Tests oder der gesamten Situation bereiteten mir beträchtliche Schwierigkeiten (bitte kurz schildern):

11.5 Fragebogen zur Generalprobe

Anhang IX

Fragebogen zur CCH-Generalprobe

Bitte beantworten Sie die folgenden Fragen vollständig und möglichst genau. Sämtliche Angaben werden streng vertraulich behandelt.

Name:	
Vorname:	
GP-Nummer (steht auf Ihrer Einladung): _ _ _ _ _	
Geburtsdatum (bitte in jedem Falle ausfüllen, dient ggfs. als Code):	
Ich habe an insgesamt ... Tagen CCH-TEST-TRAINING (ohne Generalprobe) teilgenommen (bitte zutreffendes ankreuzen)	
<input type="radio"/> 2 Tagen <input type="radio"/> 3 Tagen <input checked="" type="radio"/> 4 Tagen <input type="radio"/> 5 Tagen <input type="radio"/> 6 Tagen <input type="radio"/> 7 Tagen <input type="radio"/> 8 Tagen <input type="radio"/> 9 und mehr Tagen	
Außerdem habe ich Test-Trainings-Seminare im Umfang von xy Tagen besucht bei XYZ (bitte Veranstalter und Anzahl der Tage angeben)	
bei Veranstalter Tage	
bei Veranstalter Tage	

Zusätzlich zum professionellen, schulischen u.ä. TEST-TRAINING habe ich für <u>meine eigenständige Vorbereitung</u> auf den TMS ca ... Stunden aufgewendet (bitte die zutreffende Stundenanzahl ankreuzen).			
<input checked="" type="radio"/>	Null bis 25 Stunden	<input type="radio"/>	26-50 Stunden
<input type="radio"/>	51-75 Stunden	<input type="radio"/>	75-99 Stunden
<input type="radio"/>	100-129 Stunden	<input type="radio"/>	130-170 Stunden
<input type="radio"/>	über 170 Stunden	<input type="radio"/>	über 340 Stunden
Wenn Sie Ihre Zeit für die <u>eigenständige Vorbereitung</u> (ohne Test-Training) aufschlüsseln, welche Prozentanteile entfallen dann auf die aufgeführten Inhaltsbereiche			
- auf Analyse von Aufgaben/ "Aufgabenlogik"/ "Testlogik":		ca. <u>20</u> %	
- auf Bearbeitung von Aufgaben (z.B. Matheaufgaben lösen) ohne Zeitdruck:		ca. <u>40</u> %	
- auf Übungsdurchgänge mit exakter Zeitvorgabe:		ca. <u>40</u> %	
- auf das Einüben von Motivations- und Konzentrationsübungen:		ca. <u> </u> %	
- auf sonstiges, nämlich		ca. <u> </u> %	
Inwieweit trifft die folgende Aussage auf Sie zu? "Durch die Vorbereitung auf den TMS kann ich in vielerlei Hinsicht (z.B. Arbeitstechnik, Zeitmanagement, Blitzenzspannungstechniken) auch für andere Prüfungen profitieren." (bitte zutreffendes ankreuzen)			
<input type="radio"/> trifft voll zu <input checked="" type="radio"/> trifft ziemlich zu <input type="radio"/> trifft kaum zu <input type="radio"/> trifft überhaupt nicht zu			
Von der Generalprobe verspreche ich mir vor allem ... (bitte in eigenen Worten ergänzen)			
<i>eine weitere Übungsmöglichkeit</i>			
Für das CCH-Test-Training habe ich mich in erster Linie entschieden, weil... (bitte in eigenen Worten ergänzen):			
<i>ich selbstständig wahrscheinlich nicht so konzentriert üben hätte und ich Strategien lernen wollte.</i>			

11.6 Evaluationsbogen

Bitte im beiliegenden Freiumschlag umgehend an CCH zurücksenden!



Fragebogen zum TMS und zum CCH-TEST-TRAINING 1995

Nachname: _____ Vorname: _____

PLZ und Ort: _____ Straße: _____

Geburtsdatum: ____/____/____
(Geburtsdatum bitte auf jeden Fall angeben; dient ggf. als Code)

Ich habe am TMS am 8.11.1995 teilgenommen: ja nein

Ergebnis im Test für medizinische Studiengänge (TMS)

Bitte übertragen Sie im folgenden die entsprechenden Werte aus Ihrem Testbescheid, den Sie von der ZVS erhalten haben.

Erzielter Testwert (TMS-Wert): _____ Erreichter Prozentrang: _____

In den einzelnen Untertests erzielte Punkte:

Muster zuordnen: _____
 Medizinisch-naturwiss. Grundverständnis: _____
 Schlauchfiguren: _____
 Quantitative und formale Probleme: _____
 Konzentriertes und sorgfältiges Arbeiten: _____
 Textverständnis: _____
 Figuren lernen: _____
 Fakten lernen: _____
 Diagramme und Tabellen: _____
 Gesamtpunktzahl: _____

Wie schätzen Sie die Schwierigkeit des TMS vom 8. November 1995 im Vergleich zu den von Ihnen im Testtraining bearbeiteten Original- und Paralleltests ein?

1) Alles in allem fand ich den TMS vom 8.11.1995...

schwieriger etwa gleich schwierig einfacher weiß nicht

2) Unterschiede zu den mir vorher bekannten Testversionen aus dem Training oder der Generalprobe sehe ich vor allem in folgenden Untertests/Teilaspekten (falls zutreffend bitte kurz skizzieren):

Rückmeldung zum CCH-TEST-TRAINING

Bitte beantworten Sie die folgenden Fragen durch Ankreuzen bzw. in einigen kurzen Stichworten. Für zusätzliche Hinweise, Stellungnahmen und Kritik (ggf. auf einem separaten Blatt) sind wir sehr dankbar.

3) Insgesamt gesehen habe ich vom CCH-TEST-TRAINING ...
 sehr profitiert profitiert kaum profitiert gar nicht profitiert

Im einzelnen fand ich ...

- 4) ... den Einblick in Aufbau und Logik des TMS und seiner Aufgaben
 sehr nützlich nützlich weniger nützlich
- 5) ... die Informationen zur Testauswertung, zu Zulassungsfragen
 sehr nützlich nützlich weniger nützlich
- 6) ... die Simulation der Testsituation
 sehr nützlich nützlich weniger nützlich
- 7) ... die Bearbeitungsstrategien zu den einzelnen Untertests
 sehr nützlich nützlich weniger nützlich
- 8) ... die Strategien zur Arbeitstechnik, zum Umgang mit der Zeit und mit "Störungen"
 sehr nützlich nützlich weniger nützlich
- 9) ... die Übungen zur Konzentration, Entspannung und Selbstmotivierung
 sehr nützlich nützlich weniger nützlich

bitte wenden->

Fragebogen zum TMS und CCH-TEST-TRAINING Seite 2

- 10) ... die Hinweise und Tips zur weiteren Vorbereitung
 sehr nützlich nützlich weniger nützlich
- 11) ... das eingesetzte bzw. ausgegebene Seminarmaterial (z. B. Parallellists, Würfel)
 sehr nützlich nützlich weniger nützlich
- 12) ... sonstiges, nämlich:
 sehr nützlich nützlich weniger nützlich

nur für Teilnehmer an der Generalprobe 1995:

- 13) ... die Teilnahme an der Generalprobe fand ich
 sehr nützlich nützlich weniger nützlich

Alles in allem beurteile ich das Trainerverhalten als ...

- 14) vollkommen kompetent weitgehend kompetent ein wenig kompetent gar nicht kompetent
- 15) vollkommen engagiert weitgehend engagiert ein wenig engagiert gar nicht engagiert
- 16) vollkommen sympathisch weitgehend sympathisch ein wenig sympathisch gar nicht sympathisch

- 17) Was hat Ihnen persönlich am CCH-TEST-TRAINING besonders gut gefallen und was weniger?
 (ggf. Extrablatt)

- 18) Welche Elemente des CCH-TEST-TRAININGS sollte man noch stärker betonen? Welche Elemente könnte man evtl. ganz weglassen? (ggf. Extrablatt)

- 19) Welche sonstigen Verbesserungen würden Sie sich im nachhinein wünschen? (ggf. Extrablatt)

Zusätzliche Fragen zum aktuellen Zulassungsverfahren zum Medizinstudium:

- 20) Das derzeitige Zulassungsverfahren nach verschiedenen Quoten (Abi-Test-, Test-, Auswahlgesprächs- und Bewerbungssemesterquote) halte ich im allgemeinen für
- vollkommen angemessen angemessen kaum angemessen gar nicht angemessen
- weil...

- 21) Wie könnte Ihrer Meinung nach das ideale Auswahlverfahren zum Medizinstudium aussehen (sie können auch mehrere Auswahlkriterien ankreuzen):

- Auswahl nach:
- Testergebnis
 - Abiturschnitt
 - Bewerbungssemester
 - Auswahlgespräch
 - Los
 - sonstiges, nämlich...

Zusätzliche Fragen zur weiteren Zusammenarbeit mit der CCH Laufbahnberatung:

- 22) Wären Sie daran interessiert im Zeitraum zwischen Juni und November 1996 als ehemalige TMS- und Testtrainingsteilnehmer an unseren Seminaren teilzunehmen und über Ihre Erfahrungen zu berichten? Sie erhalten dafür ein entsprechendes Entgelt.
 ja nein
- 23) Wären Sie nach entsprechender Absprache dazu bereit Ihre Adresse und Telefonnummer als Referenz für die CCH Laufbahnberatung zur Verfügung zu stellen und zukünftigen Teilnehmern telefonisch über das Testtraining Auskunft zu geben?
 ja nein

Danke! Ihr CCH-TEAM.