

EMS

Eignungstest für das Medizinstudium in Österreich

**Medizinische Universität Innsbruck
Medizinische Universität Wien
2007**

Berichterstatter

R. Mallinger; Chr. Holzbaur (Medizinische Universität Wien)

M. Dierich; M. Heidegger (Medizinische Universität Innsbruck)

K.-D. Hänsgen; B. Spicher (Zentrum für Testentwicklung Fribourg/Schweiz)

Die Verantwortung für die psychodiagnostischen Inhalte liegt bei den Mitarbeitern des Zentrums für Testentwicklung Fribourg/Schweiz.

Redaktion: Tanja Läng (Zentrum für Testentwicklung Fribourg/Schweiz)

Information Zulassungsverfahren: www.eignungstest-medizin.at

Informationen zum Test: www.unifr.ch/ztd/ems/

Der EMS wurde 2006 und 2007 gemeinsam an der Universität Basel, der Universität Bern, der Universität Fribourg, der Medizinischen Universität Innsbruck, der Medizinischen Universität Wien und der Universität Zürich durchgeführt. Die Verantwortung für Organisation und Zulassungsverfahren lag für Österreich direkt bei beiden österreichischen Universitäten, das Zentrum für Testentwicklung stellte den Test und die Logistik zur Verfügung und führte die Auswertung durch.

Inhaltsverzeichnis

1	EINLEITUNG	4
2	EIGNUNGSTEST UND NUMERUS CLAUSUS	5
3	BESCHREIBUNG DES EIGNUNGSTESTS.....	14
3.1	Aufbau des Tests.....	14
3.2	Berechnung der Werte	16
3.3	Mittlerer Rangplatz der Untertests	17
4	TESTANWENDUNG IN ÖSTERREICH 2007	18
4.1	Statistische Angaben zur Teilnahme	18
4.2	Verteilungsprüfung der einzelnen Punktwerte	22
4.3	Vergleich der Geschlechter	25
4.4	Vergleichbarkeit der Testorte	31
4.5	Vergleich nach Maturitätsländerquote	32
4.6	Vergleiche für Altersgruppen nach Geburtsjahr	33
4.7	Vergleich nach Disziplin	34
5	ERGEBNISSE ZUR TESTGÜTE.....	35
5.1	Zuverlässigkeit.....	35
5.2	Faktorielle Validität.....	35
5.3	Item-Trennschärfen	40
5.4	Konzentriertes und sorgfältiges Arbeiten	45
6	DIF-ANALYSEN FÜR UNTERGRUPPEN	48
6.1	DIF nach Quoten (EU vs. AT), Humanmedizin	49
6.2	DIF nach Geschlecht (AT, Humanmedizin)	57
7	BEISPIELAUFGABEN FÜR DIE UNTERTESTS.....	67
7.1	Untertest: Quantitative und formale Probleme	67
7.2	Untertest: Schlauchfiguren.....	67
7.3	Untertest: Textverständnis	68
7.4	Untertest: Planen und Organisieren	69
7.5	Untertest: Konzentriertes und Sorgfältiges Arbeiten.....	71
7.6	Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	71
7.7	Untertest: Figuren lernen.....	72
7.8	Untertest: Fakten lernen	73
7.9	Untertest: Muster zuordnen.....	73
7.10	Untertest: Diagramme und Tabellen	74
8	LITERATUR	75
8.1	Originaltest zur Information und Vorbereitung.....	76
8.2	Information im Internet.....	76

1 Einleitung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2007 für das Medizinstudium und die Zulassung zum Studium der Human- und Zahnmedizin ab Wintersemester 2007/2008 an den Medizinischen Universitäten Innsbruck und Wien. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen in der Schweiz herangezogen. **Die Ergebnisse sind nicht repräsentativ für Schlussfolgerungen hinsichtlich einer gesamthaften bzw. regionalen Evaluation von Bildungseinrichtungen und -wegen.**

Aufgrund des Urteils des Europäischen Gerichtshofs vom 7.7.2005, mit dem die bisherige Regelung über den Zugang zu österreichischen Universitäten als europarechtswidrig qualifiziert wurde, und der damit geschaffenen neuen Rechtslage, die zu einem verstärkten Andrang von Studierenden aus dem Ausland, insbesondere aus Deutschland, geführt hat, führen die Medizinischen Universitäten in Wien und Innsbruck auf Basis der vom Nationalrat erlassenen Novelle zum Universitätsgesetz 2002 seit 2006 gemeinsam eine kapazitätsorientierte Studienplatzvergabe für alle Studienwerberinnen und -bewerber durch. Als Zulassungskriterium wird der Eignungstest für das Medizinstudium (EMS) verwendet, der gemeinsam mit der Schweiz durchgeführt wird.

Ein entsprechender Kooperations-Vertrag wurde zwischen der Rektorenkonferenz der Schweizer Universitäten und den beiden Universitäten im Jahre 2006 geschlossen, die Durchführung für beide Länder wird vom Zentrum für Testentwicklung und Diagnostik der Universität Freiburg/Schweiz (ZTD) koordiniert.

In Innsbruck hatten sich 1842 Personen auf 400 Studienplätze angemeldet, in Wien 3488 Personen auf 740 Studienplätze. Es war die Bereitstellung der Testlogistik für 5330 Personen notwendig.

In Österreich haben 3940 Personen (1359 in Innsbruck, 2581 in Wien) ein gültiges Testergebnis erzielt. Die weiterhin hohe Rückzugsrate von 26% ist in diesem Jahr nicht durch Doppelanmeldungen in Graz und Innsbruck/Wien erklärbar, da es ein gemeinsames Anmeldeverfahren gab (2006 betrug sie noch 35%).

Dieser Bericht will sich zugleich an die bildungspolitisch Verantwortlichen und die psychologischen bzw. psychometrischen Experten wenden. Um für letztere bestimmte Entscheidungen nachvollziehbar zu machen, enthält dieser Bericht in den Tabellen auch die für die Beurteilung eines Ergebnisses notwendigen statistischen Prüfgrößen, die nicht näher erläutert werden. In den interpretierenden Texten wird dagegen versucht, die erhaltenen Befunde so allgemeinverständlich wie möglich darzustellen.

Allgemeine Abkürzungen (ohne statistische Prüfgrößen)

EU: Europäische Union (meint die Länderquote der EU)

Nicht-EU: Nicht-EU-Länder (meint die entsprechende Länderquote)

AT: Österreich (meint die Länderquote der Personen mit österreichischer Maturität)

HM: Humanmedizin

ZM: Zahnmedizin

m: Mittelwert

s: Standardabweichung

n: Personenanzahl

2 Eignungstest und Numerus Clausus

Warum der EMS?

Im November 2005 sind beide Medizinuniversitäten an die Rektorenkonferenz der Schweizer Universitäten (CRUS) herangetreten, ob die notwendige Beschränkung der Zulassungen ab 2006 auf der Basis des Eignungstests für das Medizinstudium (EMS) erfolgen kann. Dieser wird in der Schweiz seit 1998 eingesetzt und beruht auf dem deutschen Test für Medizinische Studiengänge (TMS), der sich in Deutschland von 1986 bis 1996 bewährt hatte (300.727 Personen nahmen dort daran teil) und nach einer Pause seit 2007 in einzelnen Universitäten ebenfalls wieder eingesetzt wird.

Bereits ab dem Jahr 2005 waren in Österreich Zulassungsbeschränkungen für die Aufrechterhaltung der Studienqualität notwendig. Die Universitäten waren nicht mehr in der Lage, allen Studierenden Bedingungen anzubieten, die ein qualitativ hochwertiges Studium garantieren. Man musste 2005 das Datum des Poststempels als Kriterium der Zulassung verwenden. Neben organisatorischen Problemen (die Post arbeitet nicht überall gleich) war klar, dass dies nicht noch einmal funktionieren würde, wenn man keine Campingplätze vor den Zulassungsstellen der beiden Universitäten einrichten wollte. Deshalb wurde die Möglichkeit der Übernahme des EMS erwogen. Es gab zwei wesentliche Argumente für alle Verantwortlichen, den EMS auch in Österreich einzusetzen – den objektiven Zeitdruck immer mit berücksichtigend:

- Der Test wurde in Deutschland und der Schweiz umfangreich evaluiert. Die prognostische Validität (diejenigen bevorzugt zulassen, die ein Studium mit hoher Wahrscheinlichkeit in angemessener Zeit und mit guten Ergebnissen abschließen) war in beiden Ländern sehr gut, auch andere Anforderungen für ein solches Verfahren (elaboriertes Informationssystem mit Test Info und im WEB, ausreichend Vorbereitungsmöglichkeiten stehen offiziell zur Verfügung) wurden erfüllt.
- Ein Abgleich der Studienanforderungen zeigte, dass diejenigen von Österreich nicht wesentlich von denen in Deutschland und der Schweiz abweichen. Die seinerzeit in Deutschland durchgeführten Anforderungsanalysen für ein Studium der Medizin wurden sehr aufwändig und von verschiedenen Seiten durchgeführt (siehe Trost, 1989). Auf der Grundlage dieser Analysen wurden ca. 50 Bereiche und Aufgabentypen geprüft und die unter vielen Aspekten am besten geeignet erscheinenden 13 für die Erprobungen im Übergangsverfahren des TMS ausgewählt und empirisch erprobt. Von diesen haben vor allem aufgrund der Ergebnisse zur Validität dann 9 Aufgabengruppen Eingang in den TMS gefunden, eine (Planen und Organisieren) kam im EMS dazu.

In beiden Jahren der Testanwendung trat nun der Effekt auf, dass die Personen aus Österreich insgesamt schlechtere Leistungen erreichten als diejenigen aus Deutschland bzw. der EU-Quote und dass die österreichischen Frauen dabei noch schlechtere Leistungen erreichten als die österreichischen Männer. Durch die Quotenregelung wirkt sich der generelle Unterschied nicht auf die Zulassung aus, wohl aber der Unterschied zwischen den Geschlechtern. Daraufhin hat sich eine notwendige und wichtige Diskussion entwickelt, was die Ursachen für diese Effekte sind. Im Rahmen dieser Diskussion wurde manchmal zu schnell die Ursache auch beim „ausländischen“ Test als „Boten“ der schlechten Nachricht gesehen.

Genderfairness und Ursachen für die Unterschiede

Genderfairness wurde und wird auch im Rahmen der EMS-Evaluation laufend und ausführlich untersucht. Es war jedem Verantwortlichen klar, dass testbedingte Unterschiede ein Verstoß gegen das Fairnessprinzip wären und wenn möglich sofort auszugleichen sind (siehe auch Bericht über den EMS in Österreich 2006). Oder es müsste ein anderes Zulassungskriterium verwendet werden, wenn sich der Test in Österreich nicht eignen würde.

Folgende Vorannahmen für die Gestaltung des Zulassungsverfahrens sind dabei wichtig:

- Es gibt ein einheitliches Studium der Medizin mit einheitlichen Anforderungen für beide Geschlechter, das Zulassungsverfahren muss das Gleiche sein.
- Bei gleicher Studieneignung muss die gleiche Wahrscheinlichkeit der Zulassung für beide Geschlechter bestehen, um **individuelle Fairness** zu gewährleisten. Jede Quotenregelung hätte den Nachteil, dass Geeigneterer der einen Gruppe zugunsten der anderen abgelehnt werden müssten, um die Quote einzuhalten (sog. Paternostereffekte).

Vier mögliche Ursachenkomplexe für diese Genderunterschiede sind theoretisch zu unterscheiden. Auch innerhalb der Komplexe sind dann noch verschiedenste Ursachen und Bedingungen denkbar:

- **TEST:** Allein testbedingte Unterschiede, die keine Entsprechung im vorherzusagenden Bereich der Studieneignung haben: Durch den Test selbst würden einzelne Gruppen benachteiligt. (Fragen, die genderspezifisch leichter/schwerer zu beantworten wären oder Vor- und Nachteile, die durch die Art und Dauer des Tests entstehen). Wenn der Test eine generelle Ursache wäre, müssten sich diese Unterschiede unter vergleichbaren Bedingungen wiederholen (z.B. anderes Land), bei anderen Tests (z.B. Graz) nicht ebenso finden lassen bzw. in keiner Beziehung zu Unterschieden bei der Studieneignung stehen.
- **Bedingungen:** Ursache können die Abnahmebedingungen sein, z.B. erstmals einen derartigen Test, große Räume mit vielen Personen, strenge Aufsicht und Kontrollen. 2007 wurde großer Wert darauf gelegt, die Bedingungen stressfreier zu gestalten – trotzdem bleibt der Unterschied vorhanden. Wenn man von einer unterschiedlichen Ansprechbarkeit der Geschlechter auf Stress ausgeht, kann die Bedeutung der Abnahmebedingungen nicht ganz ausgeschlossen werden – es ist aber wohl nicht der Hauptfaktor.
- **Studienwahl:** Es handelt sich nicht um repräsentative Stichproben für Männer und Frauen, sondern um Medizinstudiumsbewerbungen. Hier können sich aus beiden Gruppen Personen mit unterschiedlichem Leistungsniveau bewerben. Auch die Tatsache, dass sich generell mehr Frauen bewerben, spricht für Unterschiede. Wenn z.B. Männer strenger benotet werden, könnten leistungsschlechtere Männer vor einer Bewerbung für Medizin eher zurückschrecken (weil es als anspruchsvolles Studium bekannt ist). Würden Frauen besser benotet als es dem realen Leistungsniveau entspricht, würden diese eher den Mut finden, sich für Medizin zu bewerben. Insbesondere die PISA-Befunde könnten für diesen Faktor sprechen: In repräsentativen Stichproben der 15jährigen (vermutlich auch gleiche Schultypen) finden sich diese Leistungsunterschiede offenbar nicht so wie beim EMS.
- **Aktueller wahrer Unterschied:** Es kann einen tatsächlichen **mittleren** Leistungsunterschied zwischen beiden Geschlechtern geben. Der muss nicht unabänderlich sein, sondern kann auch einen aktuellen Unterschied betreffen, der sich später noch ausgleichen lässt. Sollte die unterschiedliche naturwissenschaftliche Orientierung, die unterschiedliche Beschäftigung mit dieser Materie in verschiedenen Schulen wie diskutiert hier eine Rolle spielen, würde dies hierzu gerechnet – ohne damit zu unterstellen, dies wäre durch mehr Beschäftigung nicht ausgleichbar.

Nachfolgend seien alle sachlichen Argumente bezüglich der Genderfairness tabellarisch zusammengetragen. Offensichtlich wirken mehrere Ursachen zusammen, was die Aufklärung erschwert. Man kann jeweils bestimmen, ob sich durch die dargestellten Fakten Ursachen ausschließen lassen oder nicht. Es ist markiert, ob der jeweilige Fakt eher für (+) oder gegen (-) eine der Hypothesen spricht:

	Test	Bedin- gungen	Studien- wahl	Wahrer Unter- schied
Ländervergleich Schweiz : Gleicher Tag, identischer Test: Keine entsprechenden Unterschiede zwischen Geschlechtern beim Test 2006 und 2007.	-	+	+	+
Ländervergleich Deutschland : Gleiche Testaufgaben: Keine entsprechenden Unterschiede in allen Sessionen. Auch 2007 in Baden-Württemberg keine derart großen Unterschiede.	-	+	+	+
2006/07: Gleiche Ergebnisse mit Test eines ganz anderen Typs (Wissenstest) in Graz – gleiche Auswahl-situation, auch dort die Leistung der Männer besser.	-	+	+	+
DIF-Analyse („Differential Item Functioning“) auf Itemebene: Kein Itembias 2006 und 2007, d.h. es gibt keinen Effekt durch Aufgaben die z.B. aufgrund unterschiedlicher Erfahrungen der Geschlechter von einer Gruppe speziell schlechter beantwortet werden.	-	+	+	+
Unterschiede Studienerfolg für Männer, Frauen Med. Univ. Wien (Studien Frischenschlager u.a.) in 2 Jahrgängen Nachweis, dass Männer auch bessere Prüfungsleistungen erreichen, was der Test korrekt vorher sagen muss.	-	-	+	+
Evaluation Wien : Gute Vorhersageleistung SIP1 durch EMS für Kohorte 2006, in Korrelation kein Unterschied Männer zu Frauen, gleicher Mittelwert für die Gruppen nach Nichtbestehen der Prüfung für die Geschlechter. Test ist <i>individuell</i> fair für beide Geschlechter.	-	-	+	(+)
Befunde Linz (Brandstätter) Studienberatung und Auswahl, spezielle Disziplinen: Männer sind auch dort besser als Frauen.	-	-	+	+
Befunde Salzburg (Lengenfelder, Baumann), Studienauswahl Salzburg: Männer ebenfalls besser.	-	-	+	+
PISA-Studie 03 und 07 : Mathematik, Naturwissenschaften und Problemlösen im 15. Jahr bei repräsentativen Stichproben nicht der gleiche Unterschied wie im EMS.	-	-	+	-
Jungen werden bei der Benotung benachteiligt , strenger benotet (Studie Eder, Salzburg), daher weniger Mut zur Bewerbung Medizin bei gleicher Leistung wie Mädchen?	-	-	+	-
Stresssituation Test in AT größer? 2006 Unterschiede am Vormittag größer, geringer, wenn Test leicht erlebt wurde, wenn Tagesform besser – größer, wenn weniger Vorbereitungszeit und in kleineren Testlokalen (Innsbruck). 2007 wurde die Situation allerdings deutlich stressfreier gestaltet, der Effekt bleibt trotzdem.	-	+	-	-

Sagt der EMS die Studieneignung auch in Österreich vorher?

Eine hohe **Prognosekraft bezüglich des Studienerfolgs ist das eigentliche Kriterium** der Brauchbarkeit für den Eignungstest. Würde dieses Ziel nicht erfüllt, wäre jeder Testeinsatz verfehlt und man könnte stattdessen auch losen. Der Test ist ein wettbewerbsorientiertes Reihungsverfahren, es sollen diejenigen bevorzugt einen Studienplatz erhalten, die in der vorgesehenen Zeit das Studium abschließen (und die Kapazitäten wieder frei machen) sowie die besseren Leistungen erzielen. Dass die Kapazitäten beschränkt sind und nicht mehr alle Interessenten zugelassen werden können, ist **Ursache – nicht Folge** des Testeinsatzes. Da wesentlich mehr Interessenten als Studienplätze vorhanden sind (4-5 Bewerbungen auf einen Platz), **hätten auch Abgelehnte das Studium bewältigt** – aber entweder in längerer Zeit oder mit schlechteren Leistungen, wenn der Test richtig funktioniert. Beides muss man im Auge behalten.

Da der Test 2006 erstmalig in Österreich angewendet wurde, können zeitlich bedingt bisher nur die Ergebnisse der SIP-1 zur Analyse der Studienleistungen verwendet werden. Schuler und Mitarbeiter¹ haben eine sehr umfassende internationale Metaanalyse aller Zulassungskriterien vorgelegt, die auch als **Benchmarking** dienen kann, welche Prognosekraft ein gutes Zulassungsverfahren für Studienerfolg überhaupt erreichen kann und muss. Schulnotendurchschnitte und fachspezifische Studierfähigkeitstests teilen sich mit Korrelationen zwischen 0.40 und 0.50 den ersten Platz – die Fairness von Schulnoten für den **Einzelfall** wird z.B. dann relativiert, wenn Gruppen unterschiedlich streng benotet werden (wofür es bekanntlich Anzeichen gibt). Eignungstests haben demgegenüber den Vorteil, dass die Bedingungen für alle Personen vergleichbar und frei von subjektiven Bewertungseinflüssen gestaltet werden können.

Die Korrelationen des SIP-1-Punktwertes (Wien) mit der EMS-Testleistung sind wie folgt („korrigiert“ bezieht sich auf das übliche Verfahren zum Ausgleich der Varianzverringering beim Testwert durch die Zulassung der Besten):

Gesamt	0.42 (p < .000) korrigiert 0.53
Männer	0.40 (p < .000) korrigiert 0.50
Frauen	0.41 (p < .000) korrigiert 0.53

Weitergehende Validierungen stehen bisher nur aus Deutschland und der Schweiz zur Verfügung. Die jüngste Validierung in der Schweiz² ist dabei besonders hervorzuheben, **weil hier bereits umfassende Studienreformen stattfanden**. Auch für die 2. Vorprüfung betragen die Prognose-Korrelationen 0.45. Der Prozentsatz der Personen, welche die 2. Vorprüfung bestehen, hat sich in Universitäten mit Zulassungsbegrenzung erhöht – nähert sich z.B. in Bern 90%.

Für fachspezifische Studierfähigkeitstests liegt die erreichte Höhe im Bereich der erwarteten Korrelationen, die in einer umfangreichen Metaanalyse von Schuler und Mitarbeitern ermittelt worden sind³. Höhere Korrelationen sind also kaum realistisch zu erreichen.

¹ Hell, Trappmann, Weigand, Hirn und Schuler (2005): Die Validität von Prädiktoren des Studienerfolgs. Eine Metaanalyse. Präsentation als Vorabdruck zum Buch: Studierendenauswahl und Studienentscheidung. Göttingen: Hogrefe (2008)

² <http://www.unifr.ch/ztd/ems/emseval07.pdf>

³ Hell, Trappmann, Weigand, Hirn und Schuler (2005): Die Validität von Prädiktoren des Studienerfolgs. Eine Metaanalyse. Präsentation als Vorabdruck zum Buch: Studierendenauswahl und Studienentscheidung. Göttingen: Hogrefe (2008).

Analyse der Mittelwerte

	EMS Testwert Zeitpunkt Zulassung			nur Personen Teilnahme SIP				
					EMS Testwert		SIP Punktwert	
	m	s	N	N	m	s	m	s
männlich	109.6	7.7	400 (54%)	325 (54%)	110.4	7.6	70.3	16.2
weiblich	107.8	7.8	340 (46%)	273 (46%)	108.7	7.1	64.5	17.3
Gesamt	108.8	7.8	740	598 (81% v. Zugel.)	109.6	7.4	67.6	16.9

Tabelle 1: Kennwerte für EMS und SIP-1 bei Zulassung und für die Evaluationsstichprobe

81% der Zugelassenen haben an der SIP (erst möglicher Termin) teilgenommen.

Das Geschlechterverhältnis entspricht dem der Zulassung, beide Geschlechter treten gleich häufig an. Für den Vergleich Männer zu Frauen ergeben sich folgende Unterschiede:

Testwert-Differenz Zulassung: 1.8 = 23% der Standardabweichung

Testwert Differenz SIP-Teilnahme: 1.7 = 23% der Standardabweichung

SIP-Punkte Differenz 5.8 = 34% der Standardabweichung

Der **bereits bei der Zulassung vorhandene Unterschied** beim mittleren Testwert für die Geschlechter zeigt, dass Männer und Frauen in der Rangreihe der Zugelassenen nicht gleich verteilt sind, es auch anteilig mehr leistungsbessere Männer gibt. Diese Differenz zeigt sich ebenso auch in der SIP-Prüfung, wobei sich die Unterschiede bezogen auf die Standardabweichung etwa um 1/3 verstärken.

Zwei Dinge blieben festzuhalten:

Der EMS-Test sagt die Unterschiede in der Prüfung korrekt voraus.

Die von Frischenschlager und Mitarbeitern (2005) zweifach festgestellten Leistungsunterschiede bei der SIP-1 zugunsten der Männer haben ihre Ursache offenbar nicht nur in unterschiedlichem Lernerfolg während des Studiums. Die Zusammensetzung beider Gruppen ist schon bei Studienantritt unterschiedlich und diese Leistungsunterschiede setzen sich lediglich fort.

Eine interessante Differenzierung ergibt sich, wenn man die mittleren Testwerte für „Bestanden“ und „Nicht bestanden“ (Kriterium Bestehen: 66% der Punkte) vergleicht:

	Gruppen nach Bestehen der SIP								
	nicht bestanden (bis 66 Punkte)			Bestanden (ab 66 Punkte)			Total		
	Testwert			Testwert			Testwert		
	m	s	n	m	s	n	m	s	n
männlich	106.7	6.6	100	112.1	7.4	225	110.4	7.6	325
weiblich	106.3	7.1	122	110.6	6.4	151	108.7	7.1	273
Total	106.5	6.8	222	111.5	7.1	376	109.6	7.4	598

Tabelle 2: Kennwerte für Bestehen und Nicht bestehen der SIP-1

Der Gesamtunterschied im Testwert zwischen Männern und Frauen von 2 findet sich **NUR** bei den Personen, welche die SIP bestanden haben. Die Mittelwerte für die Gruppe „Nicht Bestanden“ sind für Männer und Frauen praktisch identisch – **die „Erfolgsgrenze“ ist für beide Geschlechter gleich**. Dies spricht **dagegen**, dass der Testwert des EMS den Prüfungserfolg unterschätzt hat. Wäre dies der Fall, würden sich die Mittelwerte der Gruppen für

„Nicht bestanden“ jetzt um den Betrag unterscheiden, den der EMS die Frauen hinsichtlich des Erfolges unterschätzt hätte.

Ist die Rangreihe nach dem Testwert bedeutsam für die Rangreihe der SIP? Dann wäre gewährleistet, dass die Zugelassenen bei jedem Grenzwert besser sind als die Nicht-Zugelassenen. In der folgenden Abbildung wurden die mittleren SIP-Punktwerte für jeden diskreten Testwert bestimmt (dünne Linie) und geglättet (dickere Linie). Es zeigt sich vor allem im Bereich unter 66% der SIP-Punkte (Nicht-Bestehen) ein monotoner Abfall der Leistung – der Test selektiert vor allem im Bereich des Nicht-Erfolges.

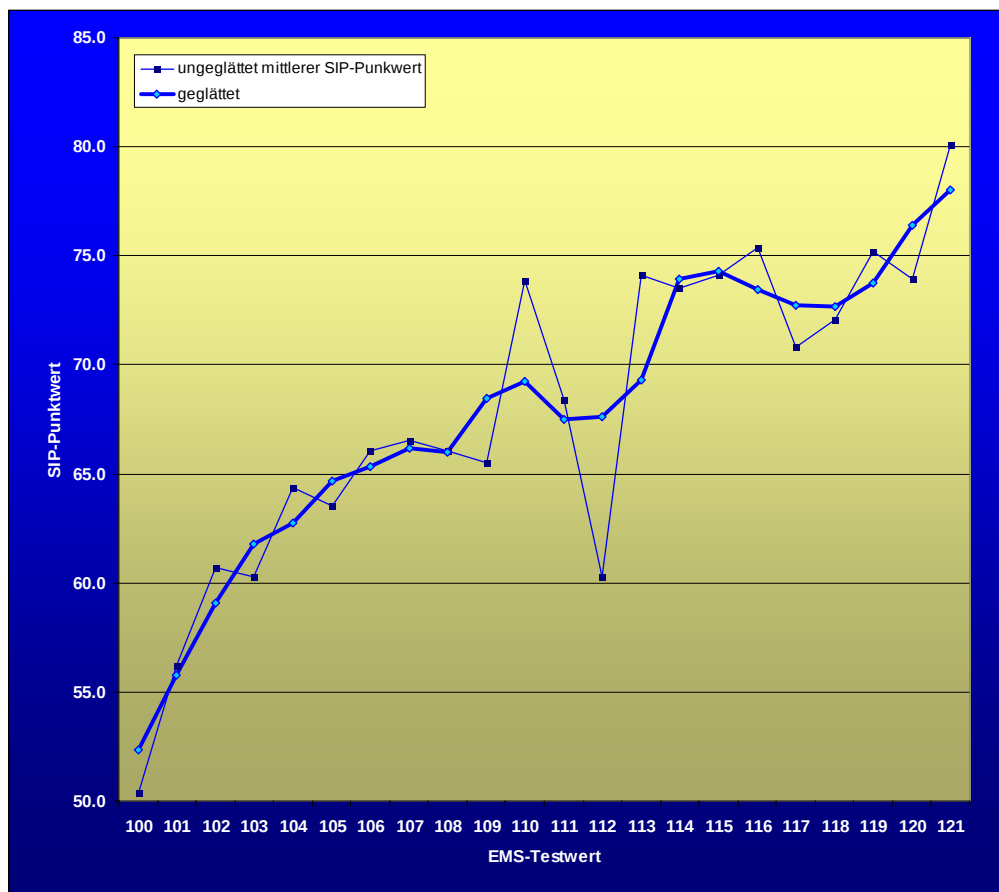


Abbildung 1: Mittlerer SIP-Punktwert für diskrete EMS-Testwerte

Einzelne Sprünge sind auch durch die Quoten bzw. geringen Besetzungen pro Testwert bedingt, diese Auswertung wird mit einer größeren Stichprobe dann wiederholt. Von allen Seiten unbestritten ist die Tatsache, dass Frauen in den SIP-1-Prüfungen unter **vergleichbaren** Bedingungen (erste Antritte) schlechtere Leistungen erreichen als Männer (Mitterauer u.a. 2007, auch frühere Untersuchungen von Frischenschlager u.a. 2005 kamen zu dem Ergebnis). Für Frauen wird festgestellt, dass mit einem Jahr Zeitverlust die Unterschiede aufgeholt werden – es bleibt aber festzuhalten, dass zu Studienbeginn besagte Unterschiede objektiv vorhanden sind. Der EMS kann seinerseits nur den Studienerfolg unter **vergleichbaren** Bedingungen vorhersagen. Insofern entspricht der im Test festgestellte Genderunterschied genau dem in der SIP-1 festgestellten Unterschied und die Studienerfolgsprognose ist insgesamt richtig. Der Test würde falsch prognostizieren, wenn sich dieser Unterschied bei den Prüfungen nicht im Testergebnis widerspiegelt. Dass die Prognosekorrelation für die Geschlechter gleich ist (s.o.), spricht für gleichartige Zusammenhänge zwischen Testergebnis und Studienerfolg.

Sollte es bei Frauen eine Gruppe geben (es sind ja nicht alle), die eine längere Anlaufzeit benötigen – etwa auch, um bestimmte Defizite bei der Aneignung naturwissenschaftlicher Kenntnisse auszugleichen – wird man trotzdem akzeptieren müssen, dass diese in einem objektiven Test vor Studienbeginn, dem besagten „wettbewerbsorientierten Reihungsverfahren“ schlechter abschneiden und ein Ausgleich „weil das ja später anders wird“ **wäre nur auf poli-**

tischem Wege möglich. Dies wäre aber auch nicht problemlos, weil dann **Fairness im Einzelfall** nicht mehr gegeben wäre: Wenn die Kapazität gleich bleibt, müsste man zum Ausgleich Männer **nicht** zulassen, die die Prüfung laut Prognose in der Realität auch eher bewältigen würden.

Misst der Test das Gleiche für Frauen und Männer, für Deutsche und Österreicher?

Weil Fairness in der Schweiz wegen der Anwendung des EMS in drei Sprachgruppen besonders wichtig ist, wurde mit den Analysen zum Differential Item Functioning (DIF) eine international eingeführte Methodik zur Identifikation und Korrektur solcher Effekte auch hier eingesetzt. Sie beruht auf der Annahme, dass Items, die Unterschiedliches messen, in einer Gegenüberstellung der Item-Schwierigkeiten für jeweils zwei betrachtete Gruppen sich von einem allgemeinen Trend über alle Items unterscheiden (Delta-Plot). Die Details werden weiter unten bzw. allgemein z.B. bei Hänsgen und Spicher (2007) dargestellt. Es zeigen sich sowohl in der Gegenüberstellung Männer zu Frauen, als auch in der Gegenüberstellung Österreich zu EU keine Hinweise, dass solche Unterschiede eine Bedeutung haben. Wären solche Effekte nachgewiesen worden, wäre auch ein Grund für einen Ausgleich – z.B. in Form von Bonuspunkten – vorhanden gewesen.

Mittels Faktorenanalyse wurde des Weiteren nachgewiesen, dass der TMS, der EMS in der Schweiz und der EMS in Österreich jeweils die gleiche Faktorenstruktur (sogar ohne Ähnlichkeitsrotation) aufweisen – was ein weiterer Hinweis auf Vergleichbarkeit darstellt.

Kann man den Test verkürzen?

Eine Verkürzung wäre grundsätzlich möglich, sogar ohne drastische Einbussen der Prognosekraft. Sowohl in Deutschland als auch in der Schweiz hat man sich aber bisher bewusst für die aktuelle Länge entschieden:

Oft wird kritisiert, dass der EMS die Berufseignung nicht berücksichtige. Diese ist nun nicht nur durch Empathie, sondern auch durch Belastbarkeit, Stressresistenz, Ausdauer und stabile Leistungsfähigkeit über eine längere Zeit (man denke etwa an eine Operation) gekennzeichnet. Auch im Studium werden Anforderungen gestellt, die solche Eigenschaften fordern. Indem der EMS ein volles Eintages-Assessment darstellt, spielen diese Merkmale für eine erfolgreiche Absolvierung auch eine Rolle. Ist der Test zu kurz, verlieren diese Faktoren an Bedeutung.

Es ist des Weiteren in den Evaluationen von 2001 auch nachgewiesen, dass die Einbeziehung jeder Aufgabengruppe noch einen Zugewinn an Reliabilität für den Testwert bringt. Durch Weglassen von „Schlauchfiguren“ oder „Figuren lernen“ wird die Prognosegüte des Studienerfolges für die erste Vorprüfung nicht wesentlich geringer. Diese beiden Aufgaben wurden aber im Test belassen, weil räumliches Denken bzw. räumliche Vorstellungen ggf. in späteren Studienabschnitten stärker gefordert werden.

...oder einen anderen Test verwenden?

Schnell wurde geschlussfolgert: Wenn die Gendereffekte so sind, wie sie sind, brauchen wir einen anderen Test, der die Stärken der Frauen besser berücksichtigt (z.B. Empathie, soziale Fähigkeiten etc. wurden genannt). Das klingt einleuchtend, ist aber „zu kurz gesprungen“. Die Zusammenstellung des Tests kann nicht willkürlich erfolgen, sondern beruht auf einer Anforderungsanalyse. Es muss eine Begründung geben, warum eine Aufgabengruppe verwendet wird. Verwendung „weil eine Gruppe da besser ist als eine andere“ reicht nicht und wäre ein Fairnessverstoss. Über folgende Fragen muss man sich klar werden:

1. **Was soll vorhergesagt werden:** Studieneignung oder Berufseignung oder beides. Die Zulassungsbegrenzung ist letztendlich ein Wettbewerb, wo ein rechtsgleiches und willkürfreies Kriterium benötigt wird, welches gut begründet ist. In Deutschland wie in der Schweiz hat man sich auf „Studieneignung“ als Grundvoraussetzung für die spätere Berufsausübung konzentriert mit der Begründung, dass diejenigen bevorzugt zuzulassen sind, die bessere Chancen haben, das Studium in angemessener Zeit mit bes-

seren Leistungen auch abzuschliessen. Dort gab es genau wie in Österreich Dropout-Raten von bis zu 50% – verschwendete Ausbildungskapazität ebenso wie Lebenszeit der abbrechenden Studierenden.

2. **Welche Merkmale kann man** vor einem Studium in einer definierten Testsituation erheben, die diesen Studienerfolg optimal vorhersagen? Man muss berücksichtigen, dass es sich um 5000 Kandidaten handelt und dass in einer „Grossgruppen-Testsituation“ wie dieser nur bestimmte Aufgaben möglich sind, die zudem nicht anfällig für Training oder eine Antwort nach der sozialen Erwünschtheit sind (was alle Persönlichkeitsfragebögen praktisch ausschliesst).

Immer wieder wird auch die Forderung gestellt, Elemente der **Berufseignung** im Sinne des „guten Arztes“ mit in den Zulassungstest aufzunehmen. Hier kann man sich zwar schnell auf Idealvorstellungen einigen, bis zu einem sachlich begründeten, rechtsgleichen und willkürfreien Test ist es aber von da noch ein weiter Weg. Die nicht-kognitiven Voraussetzungen oder „social skills“ (Empathie, soziale Kompetenz usw.) sind für die einzelnen Fachgebiete der Medizin ganz unterschiedlich zu gewichten, sind wenig analysiert und auch durch Testverfahren nicht mit der gleichen Güte zu messen wie die kognitiven Voraussetzungen. Letztendlich muss der Medizinstudierende auch hier bestimmte Fähigkeiten mitbringen. Die inhaltliche Ausformung und Differenzierung ist Aufgabe des Studiums, wobei eigene Neigungen und Interessen die Wahl des Fachgebietes dann mitbestimmen werden. Insofern wäre es vielleicht ausreichend, soziale Kompetenzen als soziale Fähigkeiten in geeigneter Weise mit zu erfassen und es dann der Ausbildung zu überlassen, dass diese Fähigkeiten auch umgesetzt werden.

Soziale Kompetenzen

Diese sollen grundsätzlich auch erfasst werden – hier gibt es aber zwei Probleme:

- Sie gehen nicht bis kaum in die **Bewertungen** (Notenvergabe) beim Studium ein. Wenn man „Studieneignung“ streng definiert, dürfte man soziale Kompetenzen nicht als Zulassungskriterium verwenden. Man kann nur indirekt darauf schliessen, dass soziale Kompetenzen (wie die Planungskompetenz auch) Anforderungen für eine erfolgreiche Gestaltung des Studiums insgesamt sind. Sie werden also durch Analyse der Anforderungen für die Tätigkeit „Studieren“ begründet – müssen sich in zukünftigen Evaluationen dennoch als Prognoseinstrumente für Studieneignung bewähren.
- Alle bisher betrachteten Tests zur „Sozialen Kompetenz“ sind trainierbar und damit verfälschbar. Man kann sich relativ leicht in Fragebögen als sozial kompetent beschreiben, wenn man die Konstruktionsprinzipien ungefähr kennt. Sie sind daher nicht für ein objektives (willkürfreies) Zulassungsverfahren geeignet. Auch Interviews können nicht für 2 000 Personen vergleichbar objektiv durchgeführt werden. Nur ganz wenige Testkonzepte, die mehr Wert auf die Erfassung sozialer **Fähigkeiten** legen sind eventuell geeignet. Diese sind vergleichsweise anspruchsvoll (z.B. Videodarbietung mit gleichen Bedingungen an allen Testorten) und der zusätzliche Aufwand steht bisher nicht im Verhältnis zum Gewinn bei der Vorhersage von Studieneignung.

Die Erfassung der sozialen Kompetenz bleibt methodische Herausforderung für die Zukunft.

Man kommt hier aber weiter, wenn man **Eignung und Neigung** unterscheidet. Neigung wäre alles, was mit Interesse und Motivation für ein Studienfach umschrieben werden kann. Bei fehlender Neigung wird ein Studium sicher gar nicht aufgenommen. Problematischer sind die Fälle, wo man während des Studiums entdeckt, dass ein Fach nicht den eigenen Neigungen entspricht. Dies kann an fehlenden oder falschen Informationen über Studium und Beruf liegen. Die Erfassung von Neigungen und deren Verwendung als Zulassungskriterium ist wegen der hohen Anfälligkeit für Trainierbarkeit und Verfälschung („sich im besten Licht darstellen“) ebenfalls praktisch unmöglich.

Wie könnte man dafür sorgen, dass die Übereinstimmung eines Studiums mit den eigenen Neigungen besser berücksichtigt würde? Eine in anderen Ländern erfolgreich eingesetzte Methode sind **Studienberatungsinstrumente**, die zur „Selbstberatung“ angeboten werden und meist im Internet realisiert sind. Es werden detaillierte Informationen über die verschie-

densten Anforderungen von Studium und späterem Beruf gegeben und Checklisten sowie Tests angeboten, um ohne Druck eigene Fähigkeiten und Neigungen zu erkunden. Da hier das Interesse der Personen an realistischen Aussagen überwiegt (weil es nicht als Zulassungskriterium verwendet wird), ist es auch nicht sinnvoll, sich in einem „besseren Licht“ darzustellen. Es gibt mittlerweile Universitäten, welche die **Absolvierung** eines solchen Studienberatungstestes zur Pflicht machen (ohne die Ergebnisse selbst für die Zulassung zu verwenden). Es wäre eine Herausforderung mit grossem erwarteten Nutzen, für den Bereich der Medizin ein solches Selbstberatungsinstrument zu entwickeln – die meisten uns bekannten Instrumente beziehen sich bisher auf technische Berufe.

3 Beschreibung des Eignungstests

3.1 Aufbau des Tests

Ausgangspunkt der Testkonstruktion war eine differenzierte Anforderungsanalyse des Medizinstudiums, an der zahlreiche Lehrbeauftragte und Experten in einem Expertenbeirat mitarbeiteten. Als Test für Medizinische Studiengänge (TMS) kam der Test in Deutschland ab 1986 zum Einsatz. Das Ziel des Tests ist die Vorhersage des Studienerfolges, um ein faires und wissenschaftlich begründetes Zulassungsverfahren zu erhalten. Gemäß diesen Anforderungen wurden einzelne Aufgabengruppen (Untertests) konstruiert, die typischen Studienanforderungen entsprechen (eine Art „Probestudium“). Zunächst wurden in Deutschland neun Untertests verwendet. Im Jahre 2005 wurde in der Schweiz der Untertest „Planen und Organisieren“ aufgrund von eigenen Analysen der veränderten Studienanforderungen ergänzt.

An der Aufgabenentwicklung nahmen zahlreiche Lehrbeauftragte und Experten teil. Die Aufgaben müssen sehr hohe Qualitätsstandards erfüllen, u.a.

- müssen sie jedes Jahr die Studieneignung gleich zuverlässig messen,
- muss das Schwierigkeitsspektrum aller Aufgaben annähernd vergleichbar sein,
- darf kein spezielles Fachwissen vorausgesetzt werden, um die Trainierbarkeit des Tests gering zu halten sowie die Fairness des Tests nicht zu beeinträchtigen und
- sollen keine sonstigen Bevorteilungen von Personengruppen (Alter, Geschlecht, Schicht usw.) bereits durch die Aufgabenauswahl erfolgen.

Der EMS besteht seit 2005 aus zehn Untertests, die zu einem Gesamtwert verrechnet werden. Im Jahre 2007 wurden sie in folgender Reihenfolge durchgeführt:

Bezeichnung der Untertests	Aufgaben	Max. Punktzahl	Bearbeitungszeit
Quantitative und formale Probleme	20	20	50 min
Schlauchfiguren	20	20	12 min
Textverständnis	18	18	45 min
Planen und Organisieren	20	20	60 min
Konzentriertes und sorgf. Arbeiten	Blatt mit 1600 Zeichen	20	8 min
	Pause		1 Std.
<i>Lernphase zu den Gedächtnistests</i>			
Figuren lernen	Es werden die Vorlagen zum Einprägen gezeigt		4 min
Fakten lernen			6 min
Med.-naturw. Grundverständnis	20	20	50 min
<i>Gedächtnistests:</i>			
Figuren reproduzieren	20	20	5 min
Fakten reproduzieren	20	20	7 min
Muster zuordnen	20	20	18 min
Diagramme und Tabellen	20	20	50 min
Gesamttest	198	198	ca. 5 Std.
Gesamtdauer (inkl. Pause) 9.00 bis ca. 17.00 Uhr			

Tabelle 3: Struktur und Ablauf des EMS 2007

Seit 2004 wird beim Konzentrationstest jährlich eine Vorlage verwendet, die vorher nicht bekannt ist (Zeichen und Durchstreichregel) – nur der Typ der Anforderung und die Zeitdauer

bleiben jeweils gleich. Dadurch werden Effekte von exzessivem Üben weitestgehend vermieden und der Test misst wirklich „Konzentriertes und sorgfältiges Arbeiten“.

Im Jahr 2005 wurde aufgrund einer Anforderungsanalyse der neuen Studienbedingungen der neue Untertest „Planen und Organisieren“ aufgenommen. Diese Anforderung „Planungskompetenz“ wurde in erster Priorität umgesetzt. Damit die Testlänge vergleichbar bleibt, wird seitdem in der Schweiz und Österreich auf nicht gewertete Einstreuaufgaben verzichtet. Sie wurden bis 2004 vor allem beibehalten, um die Testanforderung auch hinsichtlich der Länge mit der deutschen TMS-Anforderung identisch zu halten.

Die Vorteile des EMS lassen sich folgendermaßen zusammenfassen:

- Auswahl der Testanforderungen aus einer größeren Menge möglicher Studienanforderungen aufgrund einer Anforderungsanalyse – ständige Anpassung an sich verändernde Bedingungen;
- Wissenschaftlicher Nachweis der Vorhersagbarkeit von Studienerfolg, was die gesetzlich geforderte Zulassung nach der Eignung zum Studium erlaubt;
- Konstruktion der Aufgaben durch Experten UND anschließende empirische Überprüfung, damit die Aufgaben alle Gütekriterien erfüllen, die gewünschte Fähigkeit tatsächlich messen und optimal „schwierig“ sind;
- Beachtung, dass für die Beantwortung der Aufgaben kein spezielles fachliches Vorwissen notwendig ist, sondern tatsächlich die „Studierfähigkeit“ als aktuelle Fähigkeit zur Wissensaneignung und Problemlösung gemessen wird. Dadurch ist der Test auch wenig trainierbar, was sich positiv auf die Sozialverträglichkeit auswirkt (kein zusätzlicher Gewinn durch zusätzliche Trainingskurse nachgewiesen, wenn die empfohlene Vorbereitung erfolgt);
- Es ist ein ökonomisches und genau kapazitätsentsprechendes Zulassungsverfahren möglich, die Rechtsfähigkeit des Verfahrens hat sich bei mehreren Überprüfungen (auch gerichtlich in der Schweiz) bestätigt.

Wie wurden die Aufgaben entwickelt? Ausgangspunkt war eine differenzierte Anforderungsanalyse des Medizinstudiums, an der zahlreiche Lehrbeauftragte und Experten mitarbeiteten. Das Ziel war die Vorhersage des Studienerfolges, um ein faires und wissenschaftlich begründetes Zulassungsverfahren zu erhalten. Gemäß diesen Anforderungen wurden einzelne Aufgabengruppen (Untertests) konstruiert. Neun davon erfüllten alle notwendigen Anforderungen. Das Resultat ist also bereits eine Auswahl bewährter Aufgabentypen aus mehreren möglichen Alternativen. Jedes Jahr wurden neue Aufgaben für die Untertests entwickelt und in mehreren Schritten überarbeitet. An dieser Aufgabenentwicklung nahmen zahlreiche Lehrbeauftragte und Experten teil.

Die Erprobung neuer Aufgaben für sechs der neun Untertests (siehe Tabelle 3) erfolgte in Deutschland im Rahmen sogenannter „Einstreuaufgaben“. Nur bei ausreichender Bewährung wurden solche Aufgaben in nachfolgenden Testversionen für die Werteberechnung verwendet. Im Unterschied zu vielen „Übungsaufgaben“, die im sogenannten Trainingsmarkt im Umlauf sind, sind die echten EMS-Aufgaben empirisch geprüft, so dass sie bezüglich Lösungs eindeutigkeit und Schwierigkeit optimal sind.

3.2 Berechnung der Werte

Alle Untertests, außer dem „Konzentrierten und sorgfältigen Arbeiten“, liefern eine Summe („Punkte“) richtig gelöster Aufgaben zwischen 0 und 20 bzw. 18 beim „Textverständnis“.

Beim Test „Konzentriertes und sorgfältiges Arbeiten“ müssen seit 2004 insgesamt 1600 Zeichen der Reihe nach bearbeitet werden – 400 davon sind gemäss jährlich wechselnden Instruktionen zu markieren. Es können in der zur Verfügung stehenden Zeit in der Regel nicht alle Zeichen bearbeitet werden. Die Position des **letzten angestrichenen** Zeichens bestimmt, wie viele Zeichen als bearbeitet gewertet werden. Alle übersehenen und fälschlich angestrichenen Zeichen **vor** diesem letzten bearbeiteten Zeichen zählen als Fehler und diese werden von der Menge der insgesamt angestrichenen Zeichen abgezogen. Die verbleibende Menge sind die „Richtigen“, die dann in eine Skala zwischen 0 und 20 transformiert werden, um mit den anderen Tests gleichgewichtig zum Punktwert addiert zu werden.

Alle Punkte der Untertests werden zu einer Summe addiert (**Punktwert**, vgl. Abbildung 2). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Jahre vergleichbar ist. Deshalb findet eine Standardisierung auf den Mittelwert und die Standardabweichung der jeweiligen Testform statt. Dieser **Testwert** liegt zwischen 70 und 130 (der Mittelwert ist 100).

Beim „Planen und Organisieren“ wurden 2007 nur 18 Aufgaben gewertet, weil zwei Aufgaben nicht die notwendigen Kennwerte erreichten. Die Berechnung und Verwendung von Gütekriterien für die Itemselektion ist ein wichtiges Merkmal eines Tests. Insofern bevorzugen wir intern die Strategie, im Zweifel eine Aufgabe lieber auszuschließen und die Testgüte damit ausreichend hoch zu halten. Das Standardisierungsverfahren zur Bildung des Testwertes gewährleistet auch bei unterschiedlichen Aufgabenzahlen vergleichbare Testwerte zwischen den Jahren.

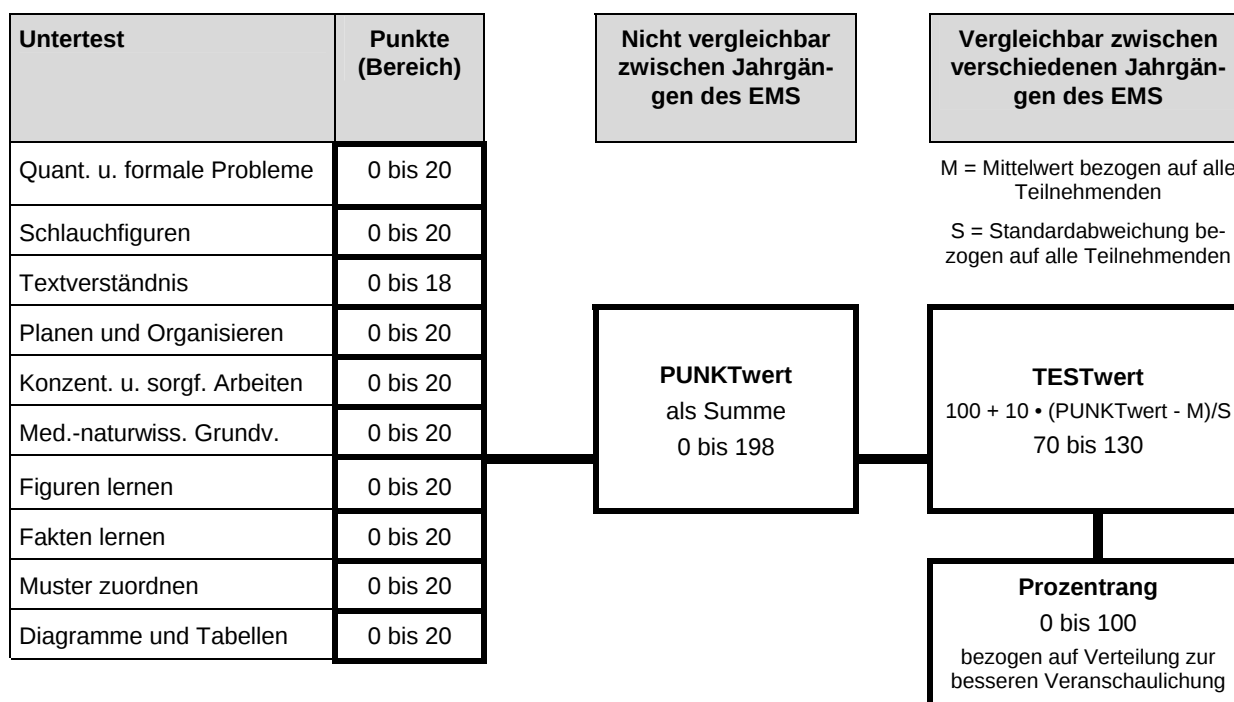


Abbildung 2: Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert (Maximalwerte 2007).

3.3 Mittlerer Rangplatz der Untertests

Gefordert ist ein Kriterium, welches die Studieneignung vorhersagt. Dies erfüllt der bisher verwendete Testwert nachweislich. Das Gewicht der Untertests in diesem Gesamtwert wird von der Varianz in diesen Untertests beeinflusst. Da die für die Studieneignung prognoserelevantesten Untertests zugleich die höchste Streuung aufweisen, erhöht dieser Effekt auch die Prognosekraft des Testwertes.

Ziel des EMS ist es, eine kapazitätsentsprechende Vergabe der Studienplätze zu erreichen. Es war in der Schweiz vorgekommen, dass mehr Personen den gleichen Testwert aufwiesen als Plätze von der Kapazität her noch zur Verfügung standen. Zugleich soll keine „Scheindifferenzierung“ erfolgen, hinter der Wertestreueung sollen wirkliche Fähigkeitsunterschiede stehen.

Bei **gleichem Testwert** werden die Personen deshalb in der Reihenfolge des **mittleren Rangplatzes aller Untertests** berücksichtigt und es werden nur so viele Personen zugelassen, wie freie Plätze tatsächlich vorhanden sind. In Österreich hat man sich diesem Verfahren angeschlossen.

Der mittlere Rangplatz wird ebenfalls auf dem Testbescheid mitgeteilt. In der Praxis gibt es pro Untertest für jeden Punktwert auch einen Rangplatz. Der Rangplatz ist umso niedriger, je höher die Punktzahl ist und je weniger Personen insgesamt einen besseren Punktwert erreicht haben. Rangplätze belohnen gute Leistungen in einem schwierigen Untertest besser, indem die dort vergebenen Ränge höheres Gewicht erhalten:

Beispiel: Werden in 2 Untertests maximal 20 Punkte erzielt, gehen immer 40 Punkte in den Punktwert ein. Haben in einem Untertest 100 Personen diesen Wert erreicht, erhalten diese Personen den mittleren Rangplatz 50.5. Haben im anderen Untertest nur 10 Personen diesen Wert erreicht, ist der Rangplatz 5.5. Letzterer wird höher bewertet, weil die 20 Punkte im zweiten Untertest schwerer zu erzielen waren und deshalb wertvoller sind.

Zwecks Vergleichbarkeit der Rangplätze zwischen den Jahren wird ein mittlerer Rangplatz auf der Basis von 1000 Personen verwendet.

- Der Rangplatz bewegt sich 2007 zwischen 50 und 920. Niedrige Werte stehen für bessere Leistungen.
- Er korreliert mit dem Testwert mit -0.991 , was praktisch für eine Äquivalenz spricht – siehe auch Abbildung 3.
- Er differenziert allerdings besser zwischen den Personen und lässt eine kapazitätsgenaue Auswahl zu.

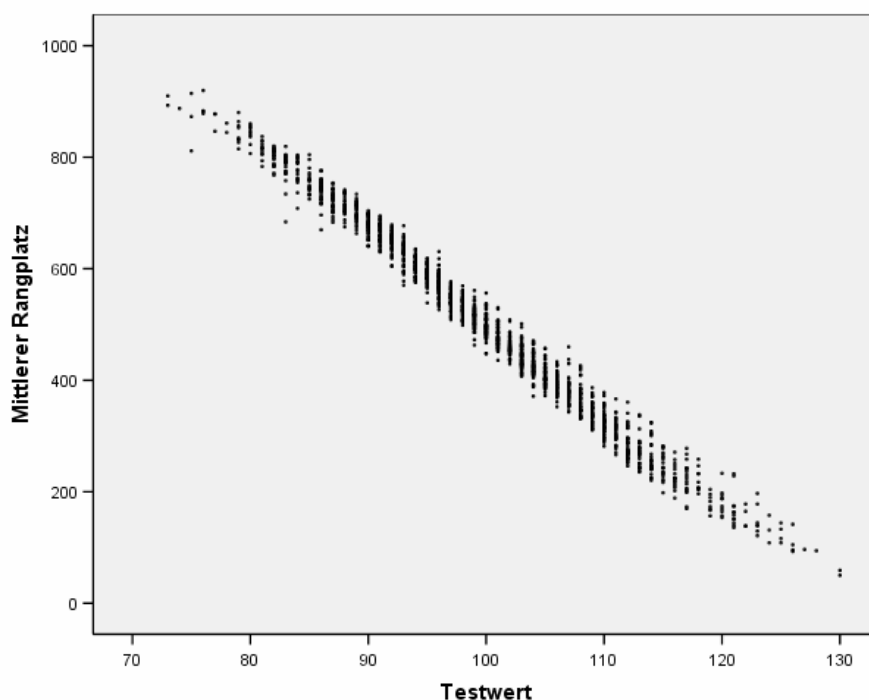


Abbildung 3: Beziehung zwischen Testwert und mittlerem Rangplatz für 2007.

4 Testanwendung in Österreich 2007

4.1 Statistische Angaben zur Teilnahme

			EU	nicht EU	Österreich	Gesamt	
Innsbruck	männlich	geb. vor 1984	43	1	26	70	
		geb. 1984-1986	186	3	48	237	
		geb. nach 1986	104	0	173	277	
		Gesamt	333	4	247	584	
	weiblich	geb. vor 1984	29	3	20	52	
		geb. 1984-1986	194	2	32	228	
		geb. nach 1986	244	2	249	495	
		Gesamt	467	7	301	775	
	Wien	männlich	geb. vor 1984	47	16	73	136
			geb. 1984-1986	194	13	147	354
geb. nach 1986			119	18	481	618	
Gesamt			360	47	701	1108	
weiblich		geb. vor 1984	47	8	89	144	
		geb. 1984-1986	189	12	117	318	
		geb. nach 1986	197	29	785	1011	
		Gesamt	433	49	991	1473	
Gesamt			1593	107	2240	3940	

Tabelle 4: Aufteilung der Teilnehmenden nach Testort, Geschlecht und Geburtsjahr

		Innsbruck	Wien	Gesamt
EU	Humanmedizin	714	696	1410
	Human- und Zahnmedizin	35	49	84
	Zahnmedizin	51	48	99
	Gesamt	800	793	1593
Nicht EU	Humanmedizin	11	74	85
	Human- und Zahnmedizin	0	5	5
	Zahnmedizin	0	17	17
	Gesamt	11	96	107
Österreich	Humanmedizin	480	1491	1971
	Human- und Zahnmedizin	36	113	149
	Zahnmedizin	32	88	120
	Gesamt	548	1692	2240
Gesamt		1359	2581	3940

Tabelle 5: Zulassungsquoten und Disziplinwahl

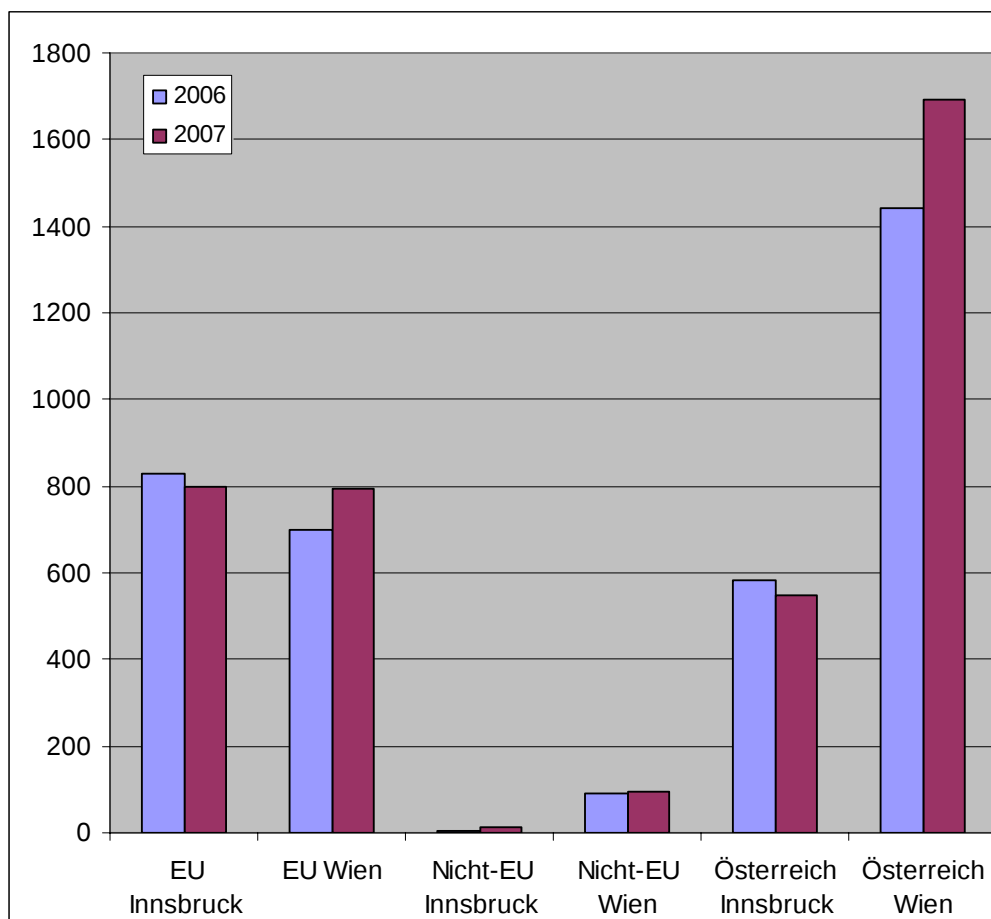


Abbildung 4: Quoten im Vergleich 2006 und 2007, Absolutzahlen

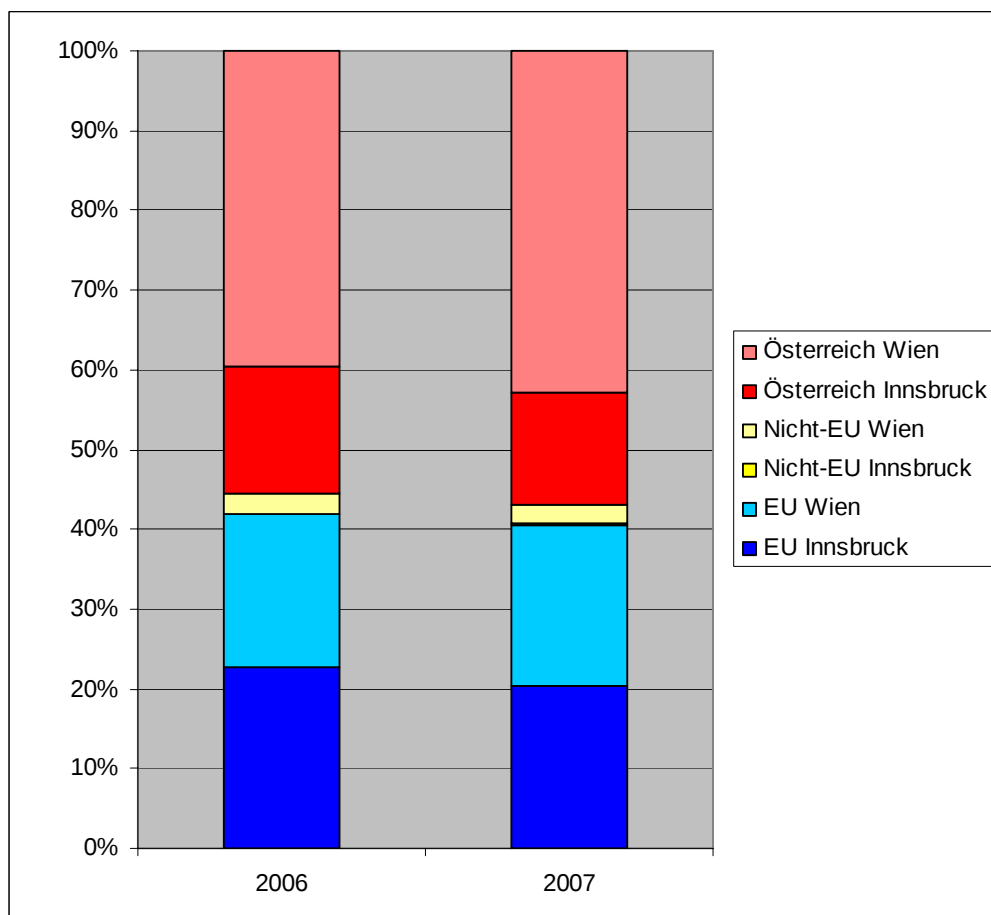


Abbildung 5: Quoten im Vergleich 2006 und 2007, prozentuale Aufteilung

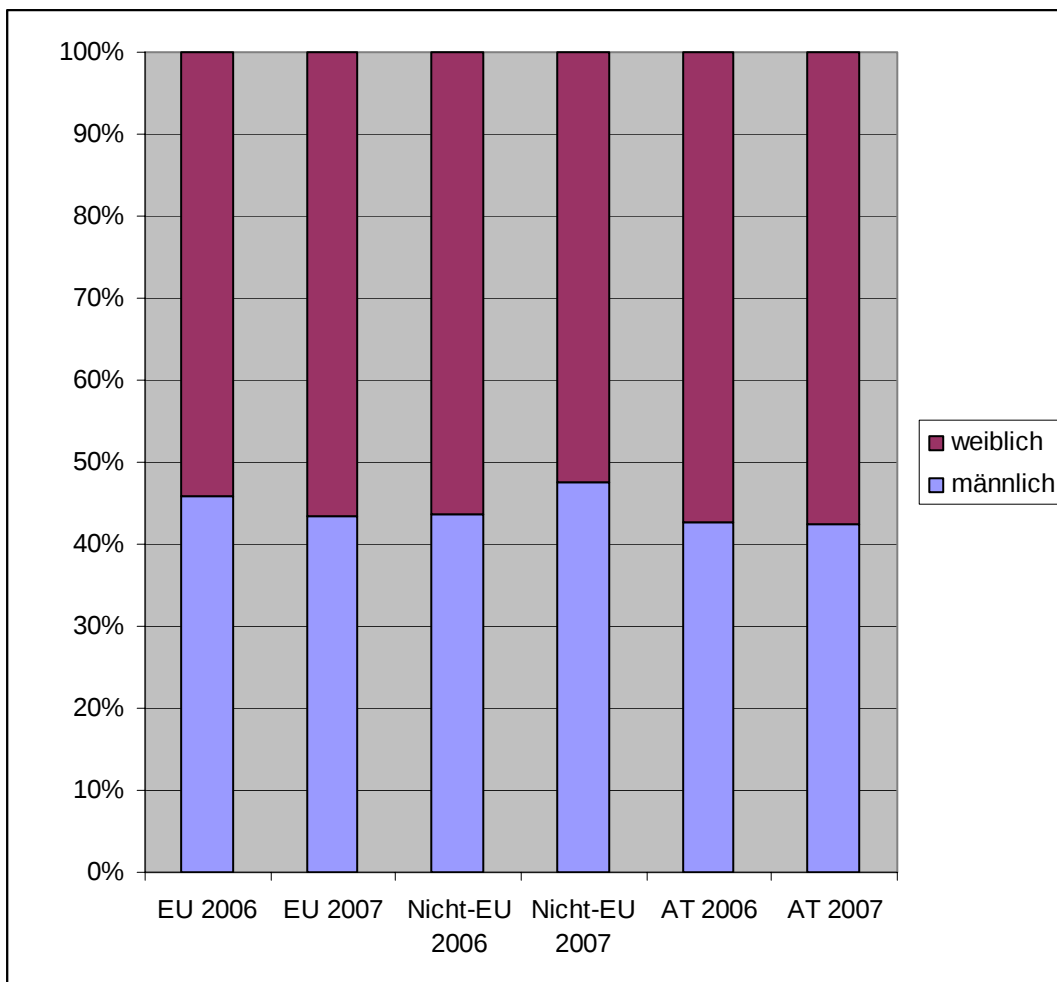


Abbildung 6: Geschlechterverhältnisse pro Quote für 2006 und 2007

Maturitätsjahr (Jahr des Abschlusses)																															
Geburtsjahr	1976	1977	1978	1981	1982	1983	1984	1985	1986	1987	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Gesamt	
1948																													1	1	
1958	2																													2	
1959		2																												2	
1960			1																											1	
1961																												1		1	
1962				1																										1	
1963					1																								1	2	
1964						1														1								1		3	
1965							2		1																				1	4	
1966								1	1																				1	3	
1967																	1													1	
1968									1	1																			2	4	
1969										1	1			1															2	5	
1970											1	1																1		3	
1972												1	1																3	5	
1973															1						1						2	1	4	9	
1974												1		1		2	2												2	8	
1975															1		1	1											2	5	
1976																2	4	1			1						1		6	15	
1977																	3	2	3	1				1			1	1	7	19	
1978																		1	1	1			1					3	8	15	
1979																			3	3	2	1		1				2	13	25	
1980																				4	6	3	3	1		1	1		11	30	
1981																					3	2	7	4		1	2	3	2	42	
1982																					1	7	9	17	5	4	1	4	37	85	
1983																							1	16	25	12	4	7	37	111	
1984																						1	1	13	42	48	15	9	71	200	
1985																									8	65	11	5	15	338	
1986																									3	2	13	241	232	599	
1987																											46	344	445	835	
1988																											3	222	741	966	
1989											1																	1	3	586	591
1990																												1	8	9	
Gesamt	2	2	1	1	1	1	2	1	3	2	3	3	1	2	2	4	11	5	7	10	14	14	31	53	83	151	290	894	2346	3940	

Tabelle 6: Zuordnung der Kandidaten zu den Gruppen „Alter und Maturität“

4.2 Verteilungsprüfung der einzelnen Punktwerte

	Punktwert	Quantitative und formale Probleme	Schlauchfiguren	Textverständnis	Planen und Organisieren	Med.-naturwiss. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme und Tabellen	Konzentr. und sorgf. Arbeiten
Mittelwert	96.23	8.54	12.35	8.22	8.95	9.44	10.02	10.01	10.07	8.44	10.19
Median	96	8	12	8	9	9	10	10	10	8	10
Modalwert	90	6	13	8	8	9	11	9	10	8	11
Stand.-abweich.	23.52	3.77	3.85	3.37	3.58	3.45	3.55	3.75	3.05	3.59	4.70
Spannweite	141	20	20	18	20	20	20	20	20	20	20
25. Perzentil	79	6	10	6	6	7	7	7	8	6	7
75. Perzentil	113	11	15	10	11	12	12	13	12	11	13

Tabelle 7: Kennwerte des Punktwerts und für die Untertests

Maßgeblich für die Zulassung ist der Testwert, welcher auf dem Punktwert basiert. Dieser ist normalverteilt, metrische Verfahren sind also angemessen. Die einzelnen Untertests sind zu Vergleichszwecken mit angegeben. Der Test „Planen und Organisieren“ war im Jahre 2006 vergleichsweise schwierig. Die 2007 verwendete Version entspricht bezüglich Schwierigkeit wieder besser den übrigen Untertests.

Die Verteilungen der weiteren Untertests sind ebenfalls eher symmetrisch, was für eine gute Differenzierungsfähigkeit spricht. Im Untertest „Konzentriertes und sorgfältiges Arbeiten“ gibt es wie in der Schweiz keine Hinweise darauf, dass der Test durch exzessives Üben beeinflusst werden kann. Vor 2004 gab es eine Häufung bei den hohen Punktzahlen, weil der Test von vielen Personen sehr oft wiederholt wurde und dann fast vollständig bearbeitet werden konnte. Dies gelingt durch den Wechsel der Anforderungen, die vorher nicht bekannt gegeben werden, nun nicht mehr.

Der Punktwert wird nach der Formel von Seite 16 in den Testwert umgerechnet. Dieser hat dann den Mittelwert 100 und die Standardabweichung 10.

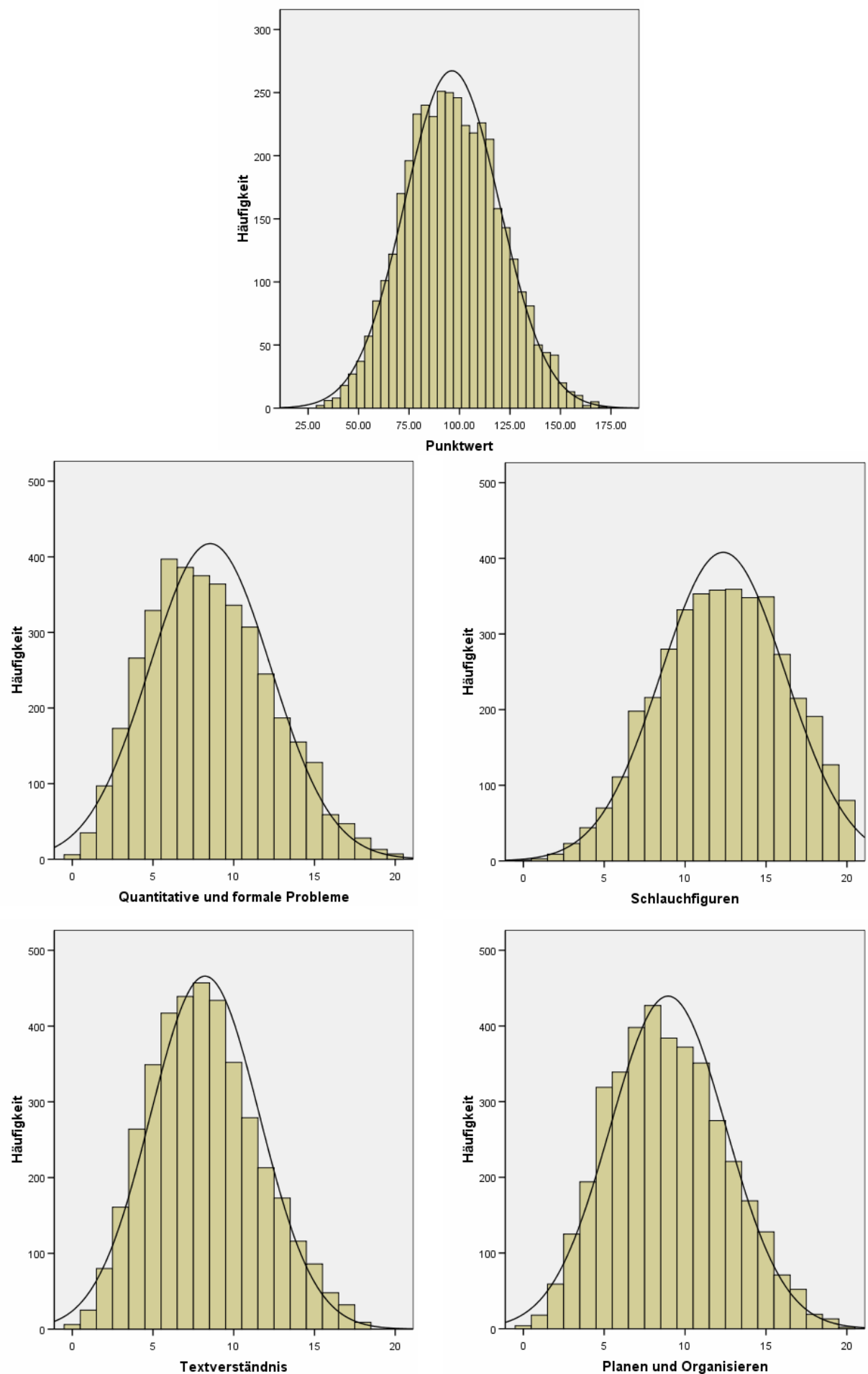


Abbildung 7: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (1)

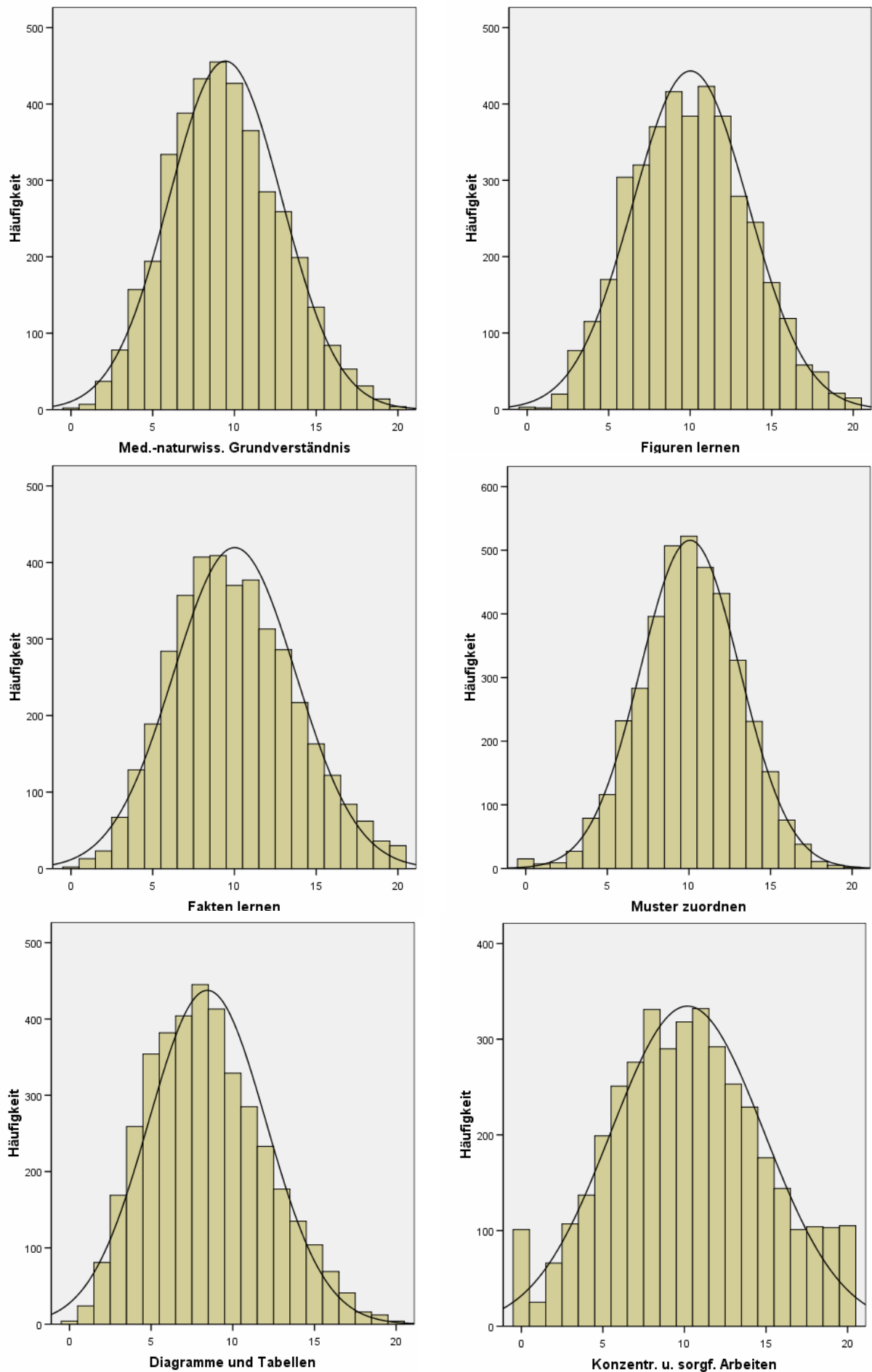


Abbildung 8: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (2)

4.3 Vergleich der Geschlechter

Tabelle 8 beschreibt die Geschlechtsunterschiede auf Untertestebene. Mit Ausnahme der Untertests „Figuren lernen“ und „Muster zuordnen“ erzielen die Bewerberinnen signifikant geringere Ergebnisse als die Bewerber.

Untertest	Geschlecht	m	s	Levene's Test Varianz			t-Test Mittelwert		
					F	Sig.	t	df	Sig.
Testwert	männlich	102.18	10.06	nicht homogen	3.96	.047	12.02	3559.36	.000
	weiblich	98.35	9.65						
Quantitative und formale Probleme	männlich	9.63	3.86	nicht homogen	24.45	.000	16.01	3425.69	.000
	weiblich	7.72	3.48						
Schlauchfiguren	männlich	13.44	3.74	homogen	0.00	.998	15.88	3938.00	.000
	weiblich	11.53	3.73						
Textverständnis	männlich	8.78	3.43	nicht homogen	4.11	.043	9.07	3551.46	.000
	weiblich	7.80	3.27						
Planen und Organisieren	männlich	9.37	3.61	homogen	0.60	.437	6.45	3938.00	.000
	weiblich	8.63	3.52						
Med.-naturwiss. Grundverständnis	männlich	10.03	3.50	nicht homogen	5.87	.015	9.28	3553.23	.000
	weiblich	9.00	3.34						
Figuren lernen	männlich	10.12	3.58	homogen	0.31	.577	1.6	3938.00	.110
	weiblich	9.94	3.52						
Fakten lernen	männlich	9.71	3.80	homogen	1.02	.313	-4.32	3938.00	.000
	weiblich	10.23	3.69						
Muster zuordnen	männlich	10.07	3.01	homogen	0.74	.390	-0.03	3938.00	.973
	weiblich	10.07	3.07						
Diagramme und Tabellen	männlich	9.61	3.72	nicht homogen	47.31	.000	18.12	3342.78	.000
	weiblich	7.56	3.23						
Konzentr. und sorgf. Arbeiten	männlich	10.61	4.77	homogen	2.42	.120	4.87	3938.00	.000
	weiblich	9.87	4.62						

Tabelle 8: Geschlechterbezogene Unterschiede in Untertests und Signifikanzprüfung mittels t-Test

Ein Vergleich über alle bisherigen Testdurchführungen in der Schweiz und in Österreich macht deutlich, dass die geschlechtsspezifischen Unterschiede auf Untertestebene im Trend den bisher bekannten Unterschieden in der Schweiz folgen, in vielen Fällen aber deutlicher ausgeprägt sind (Abbildung 9). Die Tatsache, dass sich Untertestergebnisse überhaupt unterscheiden, deckt sich mit zahlreichen Literaturbefunden über geschlechtsspezifische Unterschiede bei der Bewältigung von Leistungsanforderungen. Der mögliche Diskussionspunkt ist hier der nach den Ursachen, die auf sehr verschiedenen Ebenen liegen können – von tatsächlichen Leistungsunterschieden über Erziehungsfaktoren bis zu möglichen Unterschieden der Berücksichtigung eigener Leistungen bei der Studien- und Berufswahl und unterschiedlicher Selbst-Selektionsprozesse bei der Studienwahl.

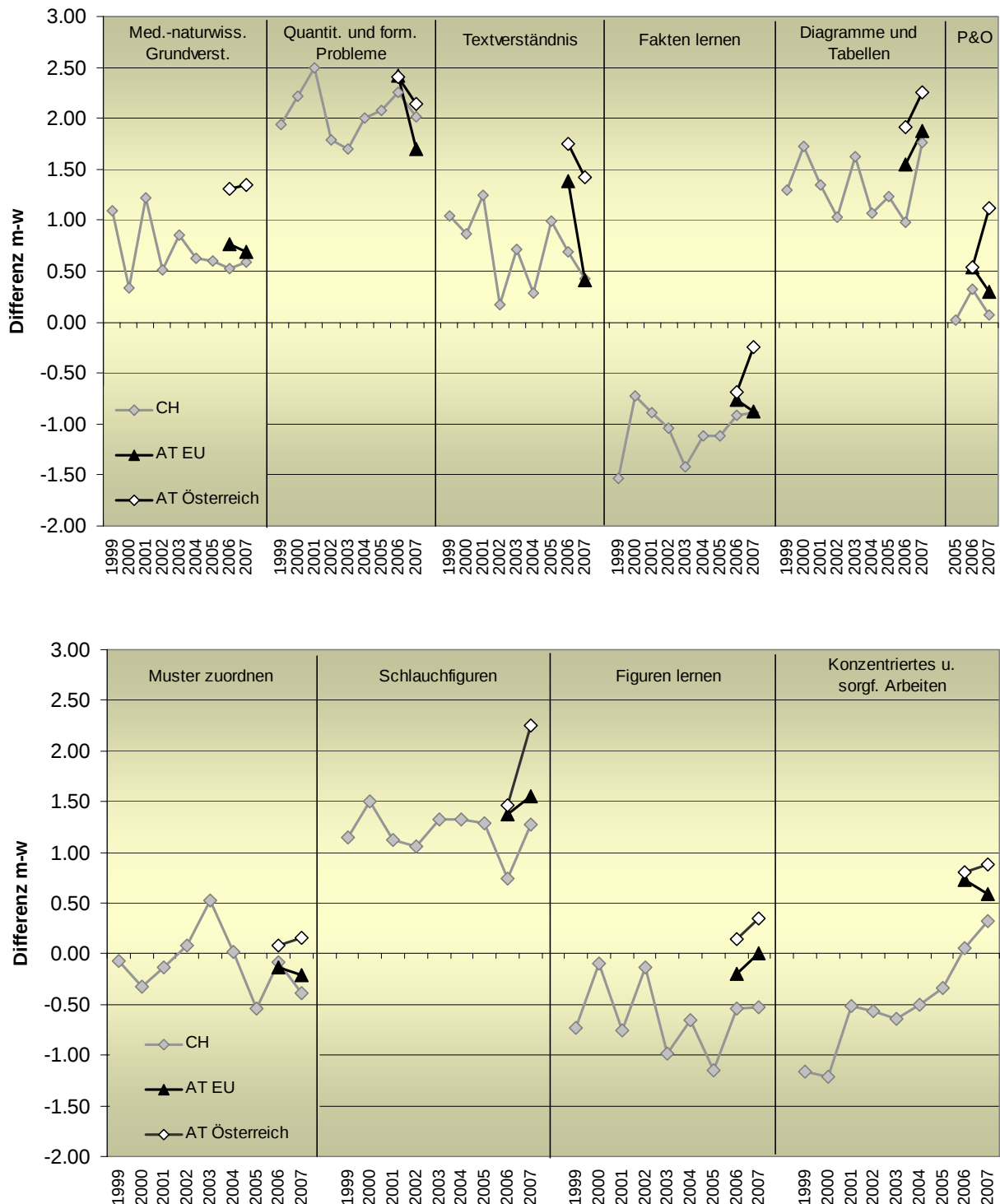


Abbildung 9: Geschlechterspezifische Unterschiede Schweiz 1999 bis 2007 und Vergleich mit den Ergebnissen in Österreich

Die Abbildung 10 verdeutlicht die Testwert-Unterschiede für unterschiedliche Kohorten in Österreich. Je höher der Balken, desto grösser die Differenz zwischen den Geschlechtern zugunsten der männlichen Bewerber. Der Unterschied ist 2007 praktisch identisch mit dem Jahrgang 2006. Allerdings wurde unter den EU-Kandidaten der Unterschied reduziert, während er unter AT-Bewerbern anstieg. Weniger Einfluss des Geschlechtes (mit einem Unterschied von 2 in etwa auf dem Niveau von der Schweiz oder Deutschland) gibt es in den

Gruppen der Zahnmedizin-Bewerbungen, der mittleren Altersgruppe und den älteren Personen mit später Maturität (2. Bildungsweg).

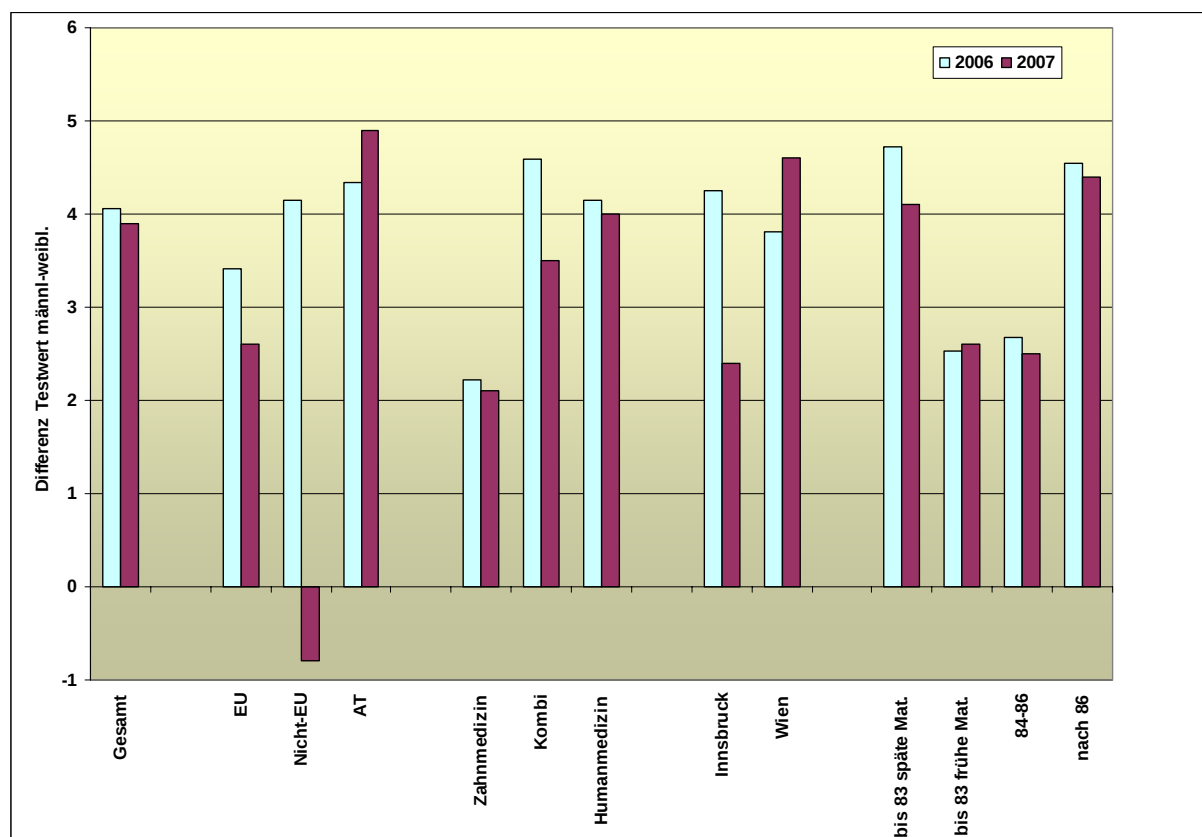


Abbildung 10: Geschlechterspezifische Unterschiede AT 2006 und 2007 nach Kohorten. Geburtsjahre sind für 2007 angegeben, 2006 um ein Jahr entsprechend versetzt.

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
EU	104.1	693	9.48	101.5	900	9.64	102.6	1593	9.66
nicht EU	91.2	51	11.70	92.0	56	10.79	91.6	107	11.19
Österreich	101.4	948	9.91	96.5	1292	8.96	98.5	2240	9.68
Total	102.2	1692	10.06	98.3	2248	9.65	100.0	3940	10.01

Tabelle 9: Testwert für Quoten und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
Humanmedizin	102.7	1475	10.06	98.7	1991	9.69	100.4	3466	10.04
Human- plus Zahn- medizin	100.6	119	9.67	97.1	119	8.53	98.9	238	9.26
Zahnmedizin	96.7	98	8.83	94.6	138	9.11	95.5	236	9.03

Tabelle 10: Testwert für Disziplin und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
Innsbruck	102.5	584	9.57	100.1	775	9.70	101.1	1359	9.71
Wien	102.0	1108	10.31	97.5	1473	9.51	99.4	2581	10.12

Tabelle 11: Testwert für Testort und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
geb. bis 1983, frühe Matura	103.3	361	10.21	100.6	339	10.16	102.0	700	10.27
geb. bis 1983, späte Matura	98.7	93	12.19	94.5	89	9.95	96.6	182	11.31
geb. 1984-1986	101.9	370	9.82	99.4	338	9.92	100.7	708	9.94
geb. nach 1986	102.2	868	9.77	97.8	1482	9.32	99.4	2350	9.72

Tabelle 12: Testwert für Alter, Maturität und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
Innsbruck EU	104.1	9.4	333	102.3	9.6	467	103.1	9.5	800
nicht EU	98.8	16.7	4	97.0	11.6	7	97.6	12.8	11
Österreich	100.3	9.3	247	96.6	8.8	301	98.3	9.2	548
Wien EU	104.0	9.6	360	100.6	9.7	433	102.1	9.8	793
nicht EU	90.6	11.2	47	91.3	10.6	49	91.0	10.8	96
Österreich	101.8	10.1	701	96.4	9.0	991	98.6	9.8	1692

Tabelle 13: Testwert für Testort und Quote

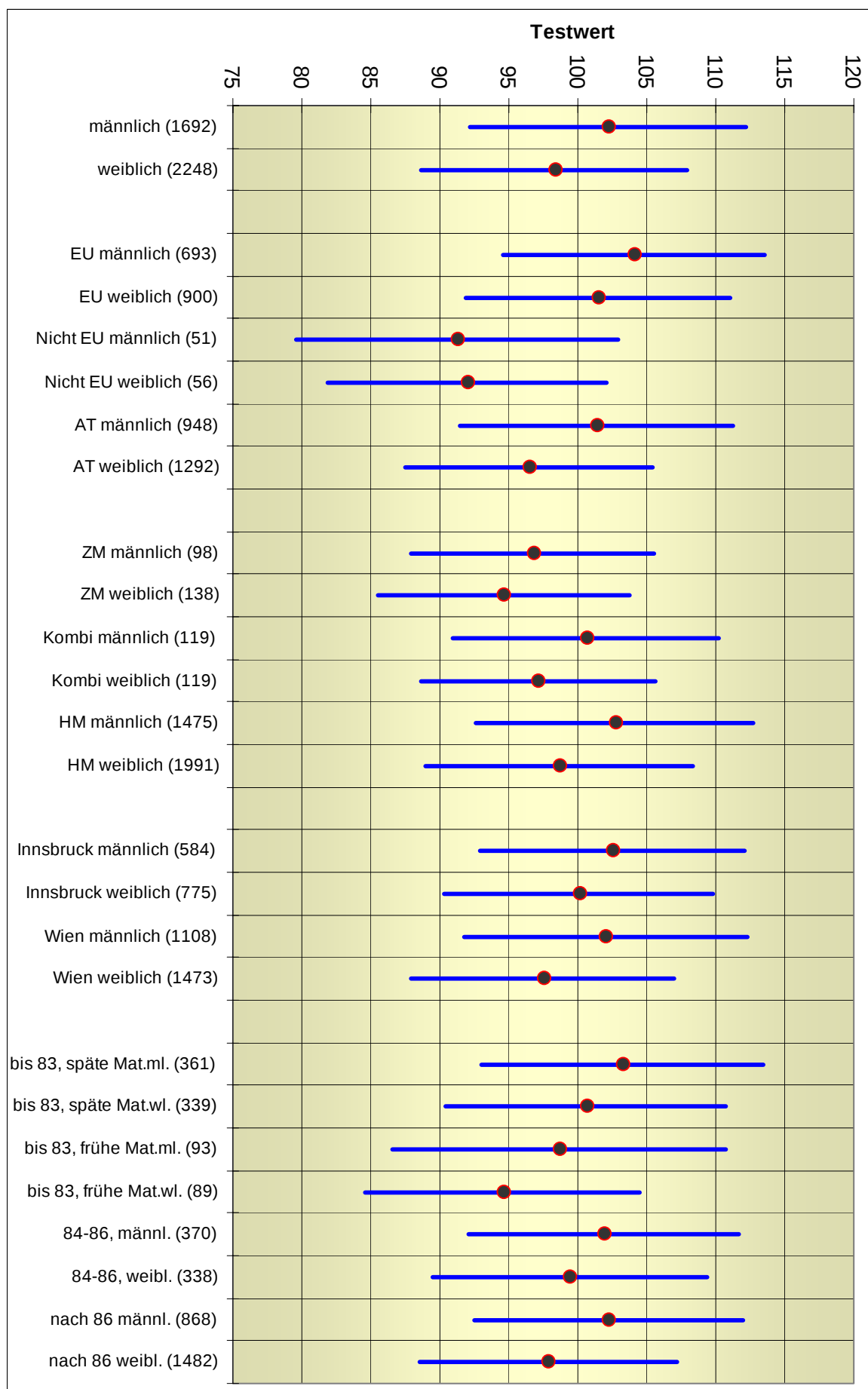


Abbildung 11: Mittelwerte und Standardabweichungen 2007 für verschiedene Teilgruppen

Entscheidung der Hypothesen aus 2006

Im Bericht 2006 wurden verschiedene Hypothesen bezüglich der Genderunterschiede formuliert. Aufgrund der Wiederholung 2007 und der Evaluation können nun folgende Bewertungen vorgenommen werden:

Die Unterschiede treten wieder genauso auf, die Prognosekraft ist vorhanden und für Männer und Frauen nicht unterschiedlich. Der EMS ist für die Vorhersage des Studienerfolges geeignet. Dies bedeutet nicht, dass die aktuellen Unterschiede zwischen den Geschlechtern hingenommen werden müssen. Entsprechend der vorgeschlagenen kurz- und mittelfristigen Massnahmen von Spiel u.M. (2007) sollte eine gezielte Förderung – offenbar vor allem im Bildungssystem – erfolgen.

Wirkung Verbesserung der Abnahmebedingungen 2007 gegenüber 2006	Zusammenhang Testergebnis 2006 und Prüfungserfolg dieser Kohorte	Ursachen und Optionen
Die Unterschiede beim Testwert treten bei Testabnahme 2007 wieder genauso auf. 2007: JA	Ausreichende Prognosekraft des Tests: Gleiche Erfolgswahrscheinlichkeit für Prüfung für Männer und Frauen 2007: JA SIP-1: gleiche Korrelation, gleiche „Bestehensgrenze“ in Testpunkten des EMS	Hinweise für „wahre“ Unterschiede zwischen Männern und Frauen (unterschiedliche Repräsentativität oder tatsächliche Unterschiede). Individuelle Chancengleichheit ist gewährleistet, es bestünde kein sachlich begründeter Korrekturbedarf beim EMS
	Ausreichende Prognosekraft des Tests, aber Erfolgswahrscheinlichkeit von Frauen wird systematisch unterschätzt bei gleichem Testwert.	Begründung für einen Korrekturfaktor für Frauen gegeben.
Die Unterschiede beim Testwert werden 2007 verringert gegenüber 2006 2007: NEIN	Ausreichende Prognosekraft des Tests: Gleiche Erfolgswahrscheinlichkeit für Männer und Frauen	Unterschiedliche Repräsentativität als Ursache für den Unterschied 2006 wahrscheinlich, Bewerberstruktur 2007 anders als 2006.
	Linear verschobene Erfolgswahrscheinlichkeit für Männer und Frauen	2006 haben sich die Testabnahmebedingungen stärker auf die Leistung der Frauen ausgewirkt. Erwartung, dass Zusammenhang Testergebnis 2007 und Prüfungserfolg ausgeglichen ist. Kein Handlungsbedarf.
	Nicht ausreichende Prognosekraft des Tests	Test „funktioniert“ in Österreich nicht, Zuschnitt des Tests auf Studienanforderungen in Österreich ist nicht mit Deutschland oder der Schweiz vergleichbar. Anderes Zulassungsverfahren oder Ergänzung ist zu prüfen.

4.4 Vergleichbarkeit der Testorte

Ob in beiden Testorten und allen Testlokalen vergleichbare Bedingungen geherrscht haben, kann post hoc aufgrund der Daten evaluiert werden. Die Personen wurden zufällig auf die Testlokale aufgeteilt, wobei zwischen Innsbruck und Wien ein Unterschied der Zusammensetzung deutlich wird: In Innsbruck bewerben sich relativ mehr Personen innerhalb der EU-Quote, was durch die Grenz Nähe zu Deutschland begründet scheint. In Wien bewerben sich über 90% aller Personen aus der Nicht-EU-Quote.

In Wien fand die Testabnahme in sieben, in Innsbruck in insgesamt sechs Lokalen, mit unterschiedlicher Größe, statt.

		Testort				Gesamt	
		Innsbruck		Wien			
		Anzahl	% von Testort	Anzahl	% von Testort	Anzahl	% von Testort
Land bzw. Quote	EU	800	58.9%	793	30.7%	1593	40.4%
	nicht EU	11	.8%	96	3.7%	107	2.7%
	Österreich	548	40.3%	1692	65.6%	2240	56.9%
Gesamt		1359	100.0%	2581	100.0%	3940	100.0%

Tabelle 14: Anteil Zulassungsquote pro Testlokal

Innerhalb der Testorte Wien und Innsbruck treten keine signifikanten Testwert-Unterschiede auf (dies gilt auch für alle einzelnen Untertests). Zwischen Innsbruck und Wien gibt es einen signifikanten Unterschied, der allerdings auf die unterschiedliche Zusammensetzung der Bewerberkohorten zurückzuführen ist: In einer zweifachen Varianzanalyse mit den Faktoren Testort und Länderquote wird der Einfluss des Testortes als nicht signifikant identifiziert.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Sektoren Innsbruck	<i>Zwischen Gruppen</i>	7259.564	9	806.618	1.555	.124
	<i>Innerhalb Gruppen</i>	699854.544	1349	518.795		
	<i>Total</i>	707114.107	1358			
Sektoren Wien	<i>Zwischen Gruppen</i>	6873.358	12	572.780	1.013	.433
	<i>Innerhalb Gruppen</i>	1451395.466	2568	565.185		
	<i>Total</i>	1458268.824	2580			

Tabelle 15: Varianzanalytische Prüfung der Homogenität für Punktwerte zwischen den Sektoren in Innsbruck und Wien, keine Unterschiede signifikant

	Testlokal	m	s	Levene's Test Varianz			t-test Mittelwert		
					F	Sig.	t	df	Sig.
Testwert	Innsbruck	101.1	9.71	<i>homogen</i>	1.826	.177	5.032	3938	<u>.000</u>
	Wien	99.4	10.12						

Tabelle 16: Vergleich der Mittelwerte des Testwertes für Wien und Innsbruck, Unterschied signifikant

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	23902.30	5	4780.46	50.73	.000
Konstanter Term	3283953.71	1	3283953.71	34846.47	.000
Testort	498.17	1	498.17	5.29	.022
Länderquote	15515.52	2	7757.76	82.32	.000
Ww. Testort x Länderquote	773.43	2	386.72	4.10	.017
Fehler	370742.68	3934	94.2406417		
Gesamt	39792845.00	3940			
Korrigierte Gesamtvariation	394644.98	3939			

Tabelle 17: Zweifache Varianzanalyse für Testwert nach Testort und Länderquote
(a): R-Quadrat = .061 (korrigiertes R-Quadrat = .059)

4.5 Vergleich nach Maturitätsländerquote

In Österreich gelten je nach Maturitätsland der Testteilnehmer unterschiedliche Zulassungsquoten, wobei die Kategorien „EU“, „nicht EU“ und „Österreich“ unterschieden werden (EU: 1593; Nicht-EU: 107; Österreich: 2240). Die Leistungen dieser drei Gruppen unterscheiden sich wie in Tabelle 27 dargestellt.

	Maturitäts- quote	Mittel- wert	Stan- dard- abw.	Quadrat- summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	EU	102.60	9.66	23017.96	2	11508.98	121.93	0.000	
	Nicht EU	91.65	11.19	371627.02	3937	94.39			
	Österreich	98.54	9.68	394644.98	3939				
Quantitative und formale Probleme	EU	9.21	3.69	1280.91	2	640.45	46.22	0.000	
	Nicht EU	7.21	4.15	54557.67	3937	13.86			
	Österreich	8.13	3.72	55838.58	3939				
Schlauchfiguren	EU	12.89	3.77	926.98	2	463.49	31.70	0.000	
	Nicht EU	10.80	4.20	57560.56	3937	14.62			
	Österreich	12.04	3.84	58487.55	3939				
Textverständnis	EU	9.30	3.33	3425.55	2	1712.78	162.84	0.000	
	Nicht EU	5.89	3.33	41409.22	3937	10.52			
	Österreich	7.56	3.18	44834.78	3939				
Planen und Organisieren	EU	9.70	3.50	2043.35	2	1021.67	83.20	0.000	
	Nicht EU	6.25	3.36	48345.80	3937	12.28			
	Österreich	8.55	3.51	50389.15	3939				
Med.-naturwiss. Grundverständnis	EU	10.41	3.37	2961.13	2	1480.57	133.05	0.000	
	Nicht EU	6.76	3.49	43811.47	3937	11.13			
	Österreich	8.88	3.31	46772.60	3939				
Figuren lernen	EU	10.43	3.57	564.41	2	282.20	22.68	0.000	
	Nicht EU	8.75	3.37	48982.35	3937	12.44			
	Österreich	9.79	3.50	49546.76	3939				
Fakten lernen	EU	10.32	3.80	475.77	2	237.89	17.08	0.000	
	Nicht EU	8.39	3.52	54836.71	3937	13.93			
	Österreich	9.87	3.69	55312.49	3939				
Muster zuordnen	EU	10.52	3.06	607.55	2	303.77	33.24	0.000	
	Nicht EU	9.05	3.35	35980.54	3937	9.14			
	Österreich	9.80	2.98	36588.08	3939				
Diagramme und Tabellen	EU	9.24	3.54	1978.80	2	989.40	79.74	0.000	
	Nicht EU	6.30	4.02	48851.36	3937	12.41			
	Österreich	7.98	3.49	50830.15	3939				
Konzentr. und sorgf. Arbeiten	EU	10.32	4.63	965.79	2	482.89	22.11	0.000	
	Nicht EU	7.23	5.12	85981.98	3937	21.84			
	Österreich	10.24	4.68	86947.77	3939				

Tabelle 18: Testwert und Punktwerte für die Maturitätsquoten – varianzanalytische Prüfung des Unterschiedes. Letzte Spalte: Schattierungen in der gleichen Spalte bedeuten, dass sich die grau markierten Gruppen NICHT voneinander unterscheiden.

4.6 Vergleiche für Altersgruppen nach Geburtsjahr

Es werden die nachfolgenden Gruppen nach dem Geburtsjahr verglichen, wobei die älteste Gruppe danach unterschieden wird, ob die Maturität früher oder später im Verlaufe des bisherigen Lebens gemacht wurde. Die Zahl der bis 1983 Geborenen ist 882, davon Maturität früh: 700, Maturität spät: 182, 1984-1986 geboren: 708, nach 1986 geboren: 2350.

	Altersgruppe	Mittelwert	Standard-abw.	Quadrat-summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	1948-1983, frühe Mat.	101.98	10.27	5849.51	3	1949.84	19.74	0.000	
	1948-1983, späte Mat.	96.65	11.31						
	1984-1986	100.71	9.94						
	1987-1990	99.45	9.72						
Quantitative und formale Probleme	1948-1983, frühe Mat.	9.11	3.83	411.36	3	137.12	9.74	0.000	
	1948-1983, späte Mat.	7.73	4.10						
	1984-1986	8.67	3.58						
	1987-1990	8.39	3.75						
Schlauchfiguren	1948-1983, frühe Mat.	12.90	3.82	488.00	3	162.67	11.04	0.000	
	1948-1983, späte Mat.	11.71	4.11						
	1984-1986	12.70	3.84						
	1987-1990	12.13	3.82						
Textverständnis	1948-1983, frühe Mat.	8.73	3.46	439.96	3	146.65	13.00	0.000	
	1948-1983, späte Mat.	7.53	3.54						
	1984-1986	8.55	3.30						
	1987-1990	8.02	3.33						
Planen und Organisieren	1948-1983, frühe Mat.	9.53	3.67	448.08	3	149.36	11.77	0.000	
	1948-1983, späte Mat.	7.92	3.51						
	1984-1986	8.96	3.48						
	1987-1990	8.86	3.56						
Med.-naturwiss. Grundverständnis	1948-1983, frühe Mat.	10.34	3.56	804.80	3	268.27	22.97	0.000	
	1948-1983, späte Mat.	9.13	3.83						
	1984-1986	9.61	3.39						
	1987-1990	9.15	3.35						
Figuren lernen	1948-1983, frühe Mat.	10.34	3.50	259.91	3	86.64	6.92	0.000	
	1948-1983, späte Mat.	9.35	3.54						
	1984-1986	10.32	3.70						
	1987-1990	9.88	3.50						
Fakten lernen	1948-1983, frühe Mat.	10.37	3.71	166.47	3	55.49	3.96	0.008	
	1948-1983, späte Mat.	9.42	3.82						
	1984-1986	9.88	3.82						
	1987-1990	9.99	3.72						
Muster zuordnen	1948-1983, frühe Mat.	10.23	3.00	165.58	3	55.19	5.96	0.000	
	1948-1983, späte Mat.	9.36	2.96						
	1984-1986	10.32	3.22						
	1987-1990	10.00	3.01						
Diagramme und Tabellen	1948-1983, frühe Mat.	9.15	3.71	630.51	3	210.17	16.48	0.000	
	1948-1983, späte Mat.	7.66	3.85						
	1984-1986	8.71	3.49						
	1987-1990	8.21	3.53						
Konzentr. und sorgf. Arbeiten	1948-1983, frühe Mat.	10.16	4.82	593.37	3	197.79	9.02	0.000	
	1948-1983, späte Mat.	8.46	4.89						
	1984-1986	10.19	4.83						
	1987-1990	101.98	10.27						

Tabelle 19: Testwert und Punktwerte für die Altersgruppen – varianzanalytische Prüfung des Unterschiedes (Gesamtstichprobe); Letzte Spalte: Schattierungen in der gleichen Spalte bedeuten, dass sich die grau markierten Gruppen NICHT voneinander unterscheiden.

Nicht immer sind die Jüngeren die Erfolgreicheren. In der EU-Gruppe sind es die Personen der mittleren Altersgruppe, in Österreich sogar die Ältesten mit früher Maturität. Wie in der Schweiz zeigt die Gruppe der Ältesten mit später Maturität (meist 2. Bildungsweg) geringere Testwerte. In den Evaluationsstudien der Schweiz konnte belegt werden, dass dies im Mittel auch mit einer geringeren Studieneignung verbunden war – gleiche Testwerte auch hier auf gleichen Studienerfolg hinweisen. Deshalb ist dies nicht als Benachteiligung zu werten, sondern eine korrekte Eignungsvorhersage des EMS.

		Innsbruck			Wien		
		n	m	s	n	m	s
EU	1948-1983, frühe Maturität	173	105.43	9.08	216	103.57	9.80
	1948-1983, späte Maturität	28	96.82	10.00	31	99.90	12.32
	1984-1986	252	102.83	9.12	232	102.75	9.27
	1987-1990	347	102.63	9.68	314	100.84	9.70
AT	1948-1983, frühe Maturität	60	100.33	9.46	219	99.73	9.79
	1948-1983, späte Maturität	24	95.96	8.80	81	97.38	11.87
	1984-1986	50	95.62	8.41	162	96.73	10.45
	1987-1990	414	98.41	9.22	1230	98.77	9.58

Abbildung 12: Altersunterschiede nach Universität

4.7 Vergleich nach Disziplin

Der mittlere Testwert für Zahnmedizin liegt in beiden Testorten tendenziell etwas tiefer als für Humanmedizin. Personen, die ein Kombinationsstudium wünschen, sind nicht geeigneter als Personen, die nur Humanmedizin studieren wollen.

		Innsbruck			Wien			Gesamtwert		
		n	m	s	n	m	s	n	m	s
Human- medizin	EU	714	103.50	9.51	696	102.63	9.61	1410	103.07	9.57
	nicht EU	11	97.64	12.83	74	91.14	11.01	85	91.98	11.39
	Österreich	480	98.45	9.31	1491	98.94	9.93	1971	98.82	9.79
	Gesamt	1205	101.43	9.78	2261	99.82	10.14	3466	100.38	10.04
Kombi- niert	EU	35	103.31	8.86	49	101.39	8.52	84	102.19	8.66
	nicht EU	-	-	-	5	94.60	13.07	5	94.60	13.07
	Österreich	36	98.31	8.60	113	96.75	9.11	149	97.13	8.99
	Gesamt	71	100.77	9.02	167	98.05	9.27	238	98.86	9.26
Zahn- medizin	EU	51	97.29	8.09	48	95.31	10.95	99	96.33	9.59
	nicht EU				17	89.18	9.76	17	89.18	9.76
	Österreich	32	95.28	7.85	88	95.82	8.29	120	95.68	8.15
	Gesamt	83	96.52	8.01	153	94.92	9.52	236	95.48	9.03

Abbildung 13: Testwert nach Wunschdisziplin und Universität

5 Ergebnisse zur Testgüte

5.1 Zuverlässigkeit

Die Reliabilität (Zuverlässigkeit) des Tests als ein Hauptgütekriterium kann anhand zweier Koeffizienten verglichen werden. Die Reliabilitätsschätzung nach der Testhalbierungsmethode (Teilung nach gerad- und ungeradzahligem Aufgaben) ist eine der gebräuchlichsten Zuverlässigkeitsschätzungen. Die internen Konsistenzen (Cronbach Alpha) schätzen die Messgenauigkeit anhand der Korrelationen jeder Aufgabe mit allen anderen des entsprechenden Untertests.

Entscheidend für die Beurteilung der Testgüte sind die Kennwerte des **Punktwertes**, welcher nach der Standardisierung als Testwert für die Zulassung verwendet wird.

Die Zuverlässigkeitswerte des Punktwertes liegen 2007 mit 0.90 bzw. 0.93 in beiden Ländern im Bereich der bisherigen Testdurchführungen. Auch die Konsistenz des Testprofils liegt mit über 0.80 im gewohnten Bereich. Hierbei ist zu beachten, dass zu hohe Werte für sehr gleichartige Untertests (mit der berechtigten Frage, ob man einzelne weglassen könnte) sprechen würden, sehr niedrige Werte für eine heterogene Testbatterie, die nicht ohne weiteres zu einem Testwert zusammengefasst werden dürfte. Der Bereich um 0.80 ist deshalb optimal, weil vergleichbare Werte auch in den Jahren mit einer erfolgreichen Evaluation des Zusammenhangs von Studienerfolg und Eignungstest gefunden worden sind. Dieser Wertebereich wird in den Folgejahren eingehalten und auch in Österreich erreicht. Es bestehen somit keine Einwände seitens der Zuverlässigkeit der Messung, den Punktwert bzw. Testwert für die Eignungsmessung zu verwenden.

	Reliabilität nach Testhalbierungsmethode										
	DE	2000	2001	2002	2003	2004	2005	2006 CH	2006 AT	2007 CH	2007 AT
Punktwert	.91 - .93	.90	.91	.91	.92	.91	.92	.90	.92	.90	.93

Tabelle 20: Reliabilität des Punktwertes (Split Half) für Deutschland (Bereich), Schweiz und Österreich

	Konsistenz des Testprofils									
	2000	2001	2002	2003	2004	2005	2006 CH	2006 AT	2007 CH	2007 AT
Profil	.81	.80	.80	.81	.78	.82	.79	.83	.80	.84

Tabelle 21: Konsistenz des Testprofils für die Schweiz und Österreich

5.2 Faktorielle Validität

Aufgrund der Korrelationen zwischen den Untertests wurde geprüft, ob die Struktur der Untertests für die Schweiz und Österreich mit jener der Vorjahre vergleichbar ist. Dies wäre ein Indiz, dass tatsächlich die gleichen Merkmale gemessen werden.

Die sehr gute Übereinstimmung der Faktorenstrukturen sowohl zwischen der Schweiz und Österreich als auch zu den Vorjahren zeigt, dass die gemessenen Merkmale identisch sind, der Test in allen Ländern und über die Jahre Vergleichbares misst.

	Muster zuordnen	Med.-naturw. Grundv.	Schlauchfiguren	Quant. u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Planen und Organisieren	Punkt-wert CH	Punkt-wert AT
Muster zuordnen		0.17	0.37	0.15	0.19	0.27	0.27	0.16	0.31	0.09	0.48	0.54
Med.-naturw. Grundverst.	0.22		0.28	0.52	0.59	0.20	0.21	0.58	0.17	0.43	0.67	0.70
Schlauchfiguren	0.40	0.33		0.35	0.28	0.35	0.32	0.32	0.32	0.20	0.63	0.68
Quant. u. formale Probl.	0.25	0.55	0.40		0.48	0.16	0.18	0.63	0.18	0.47	0.68	0.72
Textverständnis	0.23	0.63	0.36	0.51		0.22	0.24	0.52	0.16	0.40	0.66	0.69
Figuren lernen	0.36	0.25	0.42	0.25	0.27		0.40	0.19	0.29	0.15	0.54	0.59
Fakten lernen	0.28	0.30	0.34	0.28	0.31	0.41		0.19	0.24	0.19	0.55	0.58
Diagramme und Tabellen	0.25	0.60	0.39	0.64	0.54	0.26	0.27		0.17	0.47	0.69	0.73
Konzent. u. sorgfält. Arbeiten	0.35	0.20	0.38	0.27	0.21	0.33	0.25	0.24		0.15	0.54	0.57
Planen und Organisieren	0.20	0.48	0.30	0.49	0.45	0.23	0.27	0.53	0.22		0.58	0.64

Tabelle 22: Korrelationen zwischen Punktwerten der Untertests CH (über der Diagonale) und AT (unter der Diagonale) sowie mit dem Gesamtwert

Varimax-rotierte Lösungen					
Eigenwert		% Varianz		Kumuliert %	
CH	AT	CH	AT	CH	AT
Zwei-Faktorenlösung					
3.05	3.19	30.54	31.93	30.54	31.93
2.28	2.43	22.75	24.28	53.29	56.21
Drei-Faktorenlösung					
3.05	3.17	30.46	31.72	30.46	31.72
1.68	1.80	16.80	18.04	47.26	49.77
1.44	1.45	14.39	14.46	61.65	64.23

Tabelle 23: Varianzanteile der einzelnen Faktorenlösungen Schweiz und Österreich 2007 (varimax-rotierte Lösungen)

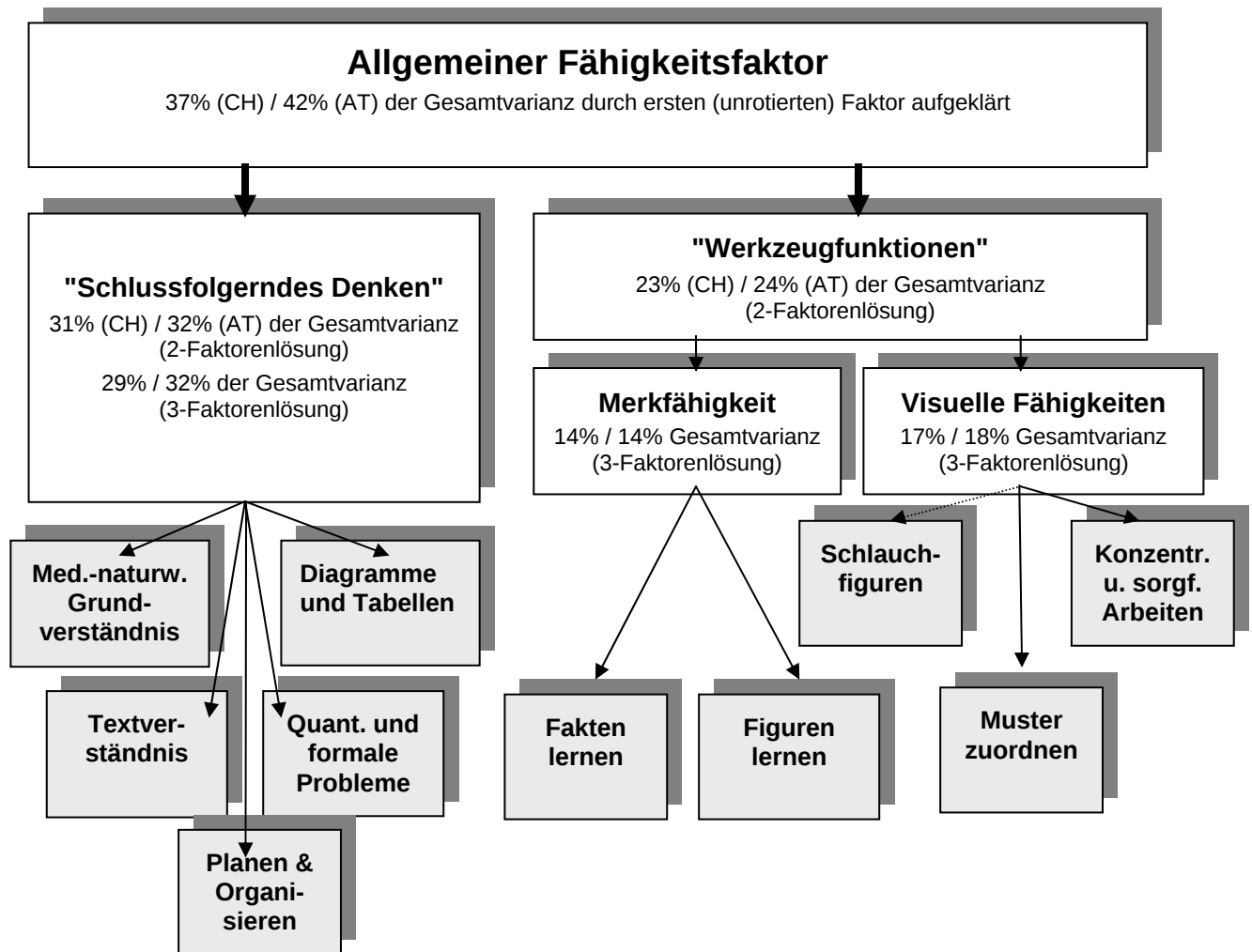


Abbildung 14: Struktur der Untertests des EMS, hierarchische Faktorenstruktur 2007, Werte Schweiz/Österreich der Varianzaufklärung, Untertests entsprechend der Hauptladungen zugeordnet.

Die Struktur des EMS bleibt weiterhin stabil – trotz Einführung eines neuen Untertests „Planen und Organisieren“ im Jahre 2005 und der laufenden Veränderung beim „Konzentrierten und sorgfältigen Arbeiten“.

Die Leistungen im gesamten Test werden in Form eines „allgemeinen Fähigkeitsfaktors“ am stärksten durch die Untertests „medizinisch-naturwissenschaftliches Grundverständnis“, „Diagramme und Tabellen“, „quantitative und formale Probleme“ sowie „Textverständnis“ charakterisiert (Tabelle 24).

In der Zweifaktorenlösung stechen die Untertests „Diagramme und Tabellen“ sowie „medizinisch-naturwissenschaftliches Grundverständnis“ für den Faktor „schlussfolgerndes Denken“ und der Untertest „Figuren lernen“ als typischer Repräsentant für den Faktor „Werkzeugfunktionen“ hervor (Tabelle 25).

Die Dreifaktorenlösung (Tabelle 26) teilt den Faktor „Werkzeugfunktionen“ in „Merkfähigkeit“ (repräsentiert durch „Fakten lernen“ und „Figuren lernen“) und „visuelle Fähigkeiten“ („Konzentriertes und sorgfältiges Arbeiten“, „Muster zuordnen“, „Schlauchfiguren“).

	2000	2001	2002	2003	2004	2005	2006 CH	2006 AT	2007 CH	2007 AT
Muster zuordnen	.57	.57	.71	.48	.45	.46	.47	.52	.43	.51
Med.-naturwiss. Grundverst.	.74	.74	.70	.74	.71	.79	.74	.77	.74	.74
Schlauchfiguren	.65	.62	.69	.61	.61	.58	.57	.64	.60	.66
Quant. und formale Probleme	.72	.66	.68	.73	.70	.75	.71	.72	.73	.75
Textverständnis	.71	.72	.62	.77	.70	.76	.72	.74	.72	.73
Figuren lernen	.53	.56	.62	.55	.54	.48	.47	.54	.48	.55
Fakten lernen	.55	.52	.57	.51	.51	.47	.48	.51	.48	.55
Diagramme und Tabellen	.71	.72	.56	.76	.74	.78	.74	.76	.76	.77
Konzentr. u. sorgf. Arbeiten	.55	.49	.54	.48	.44	.41	.52	.58	.43	.49
Planen und Organisieren						.68	.51	.50	.62	.67

Tabelle 24: Ladungen im ersten Faktor der unrotierten Lösung 2000 bis 2007 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor); 2000-2005 Schweiz und 2006-2007 für beide Länder

	Faktor 1						Faktor 2						Kommunalitäten (h ²)					
	04	05	06 CH	06 AT	07 CH	07 AT	04	05	06 CH	06 AT	07 CH	07 AT	04	05	06 CH	06 AT	07 CH	07 AT
Diagramme und Tabellen	.79	.81	.80	.78	.82	.81							.66	.70	.66	.67	.69	.70
Med.-naturwiss. Grundverständnis	.80	.82	.77	.76	.79	.81							.65	.71	.64	.66	.64	.68
Textverständnis	.70	.81	.77	.76	.74	.76							.63	.68	.62	.64	.58	.62
Quant. und formale Probleme	.70	.78	.79	.78	.79	.77							.60	.64	.64	.64	.64	.64
Planen und Organisieren		.72	.61	.62	.69	.71								.54	.37	.39	.49	.53
Konzentr. u. sorgf. Arbeiten							.33	.51	.62	.57	.64	.67	.20	.28	.41	.40	.41	.46
Figuren lernen							.74	.74	.73	.75	.69	.73	.56	.56	.54	.56	.49	.56
Muster zuordnen							.70	.74	.66	.68	.67	.70	.47	.56	.45	.48	.46	.50
Fakten lernen							.67	.62	.64	.64	.64	.58	.47	.40	.42	.42	.44	.40
Schlauchfiguren		.29			.29	.33	.67	.62	.68	.68	.65	.66	.52	.47	.50	.54	.50	.54

Tabelle 25: Ladungen und Kommunalitäten der Zwei-Faktorenlösung, varimaxrotiert; 2004-2005 Schweiz und 2006-2007 für beide Länder

Faktor	Ergebnisse 2007								Deutschland		
	1		2		3		h ²		1	2	3
	2007 CH	2007 AT	2007 CH	2007 AT	2007 CH	2007 AT	2007 CH	2007 AT			
Diagramme und Tabellen	.82	.81					.67	.71	0.82		
Medizin.-naturwiss. Grundverständnis	.79	.81					.69	.69	0.81		0.2
Quantitat. u. formale Probleme	.79	.77					.64	.66	0.8	0.18	
Textverständnis	.73	.75					.66	.63	0.79		0.2
Planen und Organisieren	.69	.71					.42	.53	Noch nicht enthalten		
Muster zuordnen			.76	.71			.69	.56		0.81	
Konzentr. u. sorgfält. Arbeiten			.74	.81			.63	.67		0.7	0.4
Schlauchfiguren		.33	.61	.60	.28	.31	.56	.56	0.35	0.71	
Fakten lernen					.83	.85	.73	.78	0.21		0.87
Figuren lernen				.41	.76	.68	.63	.65	0.13	0.47	0.64

Tabelle 26: Faktorenanalyse: Varimaxrotierte Drei-Faktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42).

5.3 Item-Trennschärfen

Die Item-Trennschärfen sind die Korrelationen des Punktwertes für den jeweiligen Untertest mit den zugeordneten Items. Positive Korrelationen weisen darauf hin, dass die Leistungsbesten im jeweiligen Untertest auch beim entsprechenden Item die richtige Lösung bevorzugt gewählt haben. Zu beachten ist, dass bei sehr leichten und sehr schwierigen Items wegen der geringeren Antwortvarianz auch die Trennschärfe in der Regel geringer ausfallen wird. Negative Trennschärfen würden auf Items hinweisen, die missverständlich formuliert sind oder keine eindeutige Lösung haben – die Leistungsbesten eine andere als die vorgegebene Lösung gewählt haben. Entsprechend der festgelegten Auswertedirektive des EMS werden solche Items von der Auswertung nachträglich ausgeschlossen und nicht gewertet, um Risiken der Fehlbewertung auszuschließen.

Die nachfolgenden Abbildungen zeigen die Trennschärfen. Diese wurden für die Schweiz und Österreich gemeinsam berechnet, um den Test hinsichtlich der Vergleichbarkeit zu optimieren (die Unterschiede zwischen beiden Ländern sind marginal) und mit den Daten aus Deutschland für die entsprechenden Aufgaben verglichen (die Aufgaben von neun Untertests wurden bekanntlich bereits auch einmal in Deutschland eingesetzt).

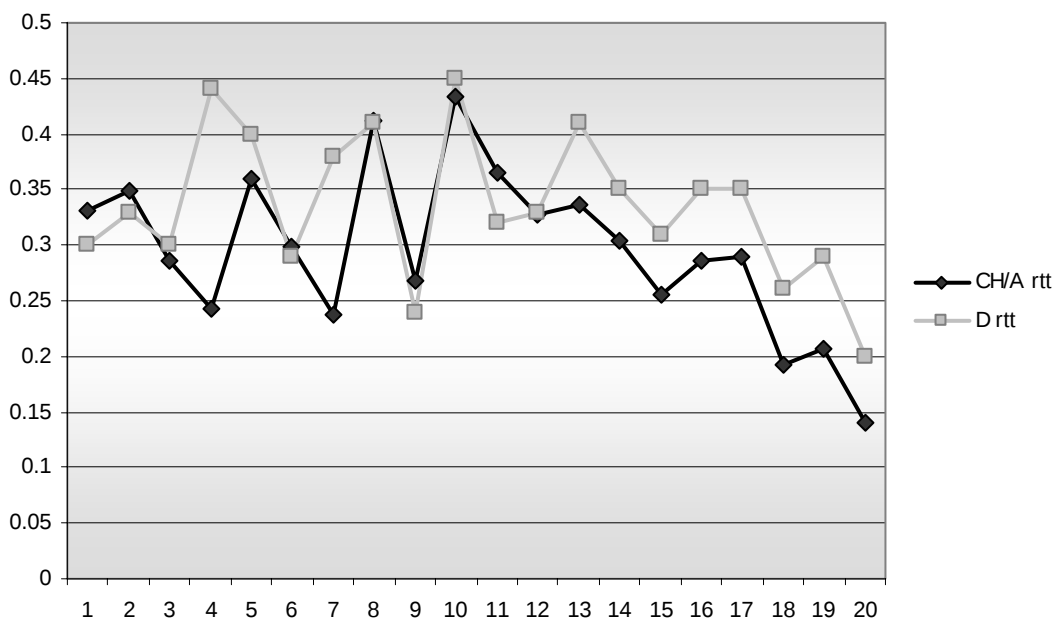


Abbildung 15: Trennschärfen für den Untertest „quantitative und formale Probleme“.

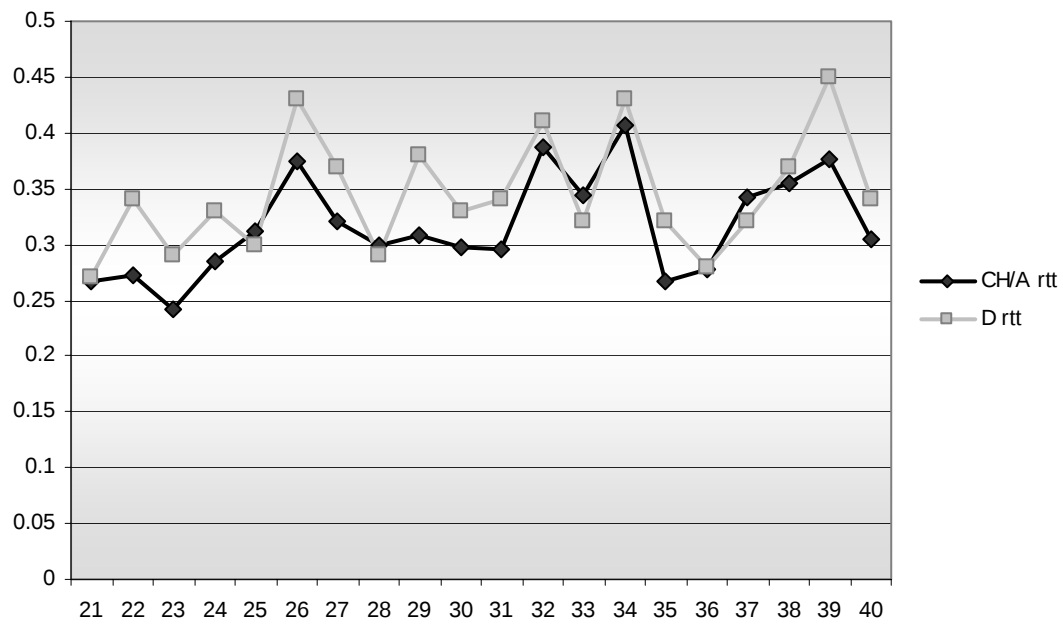


Abbildung 16: Trennschärpen für den Untertest „Schlauchfiguren“.

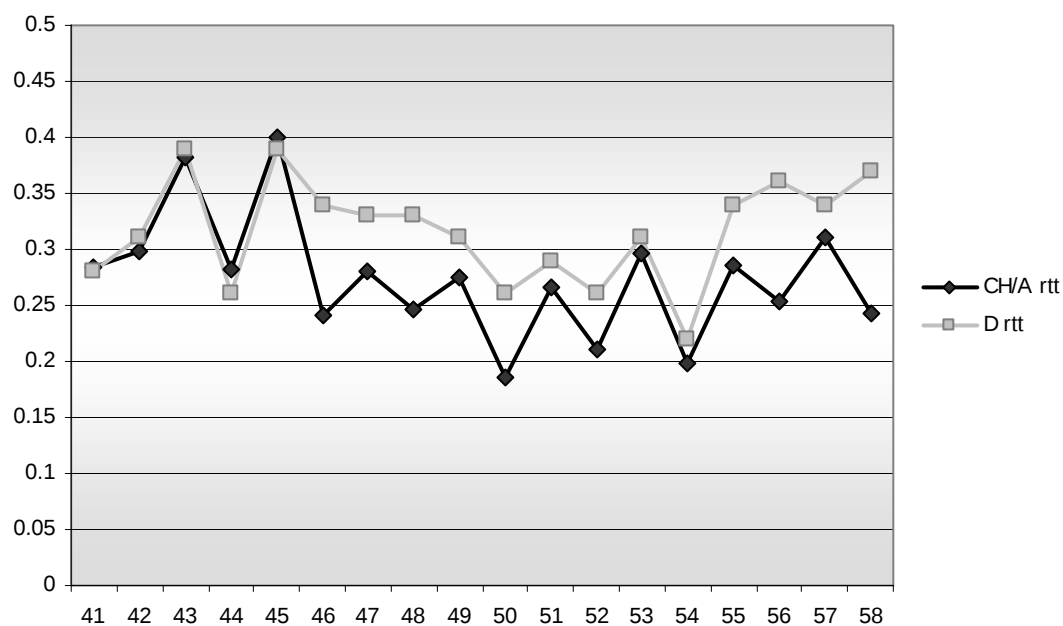


Abbildung 17: Trennschärpen für den Untertest „Textverständnis“.

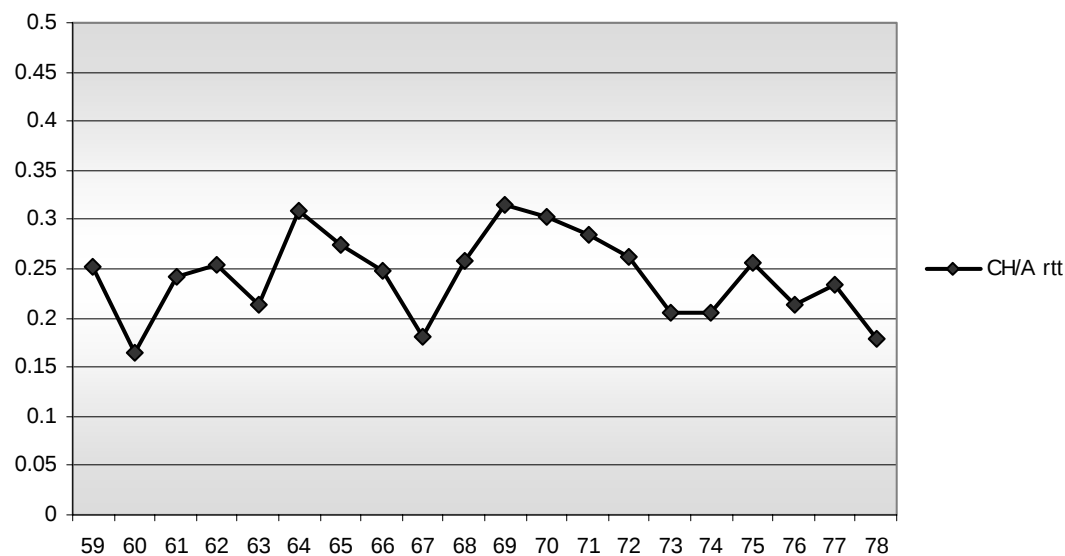


Abbildung 18: Trennschärfen für den Untertest „Planen und Organisieren“ (keine deutschen Vergleichswerte).

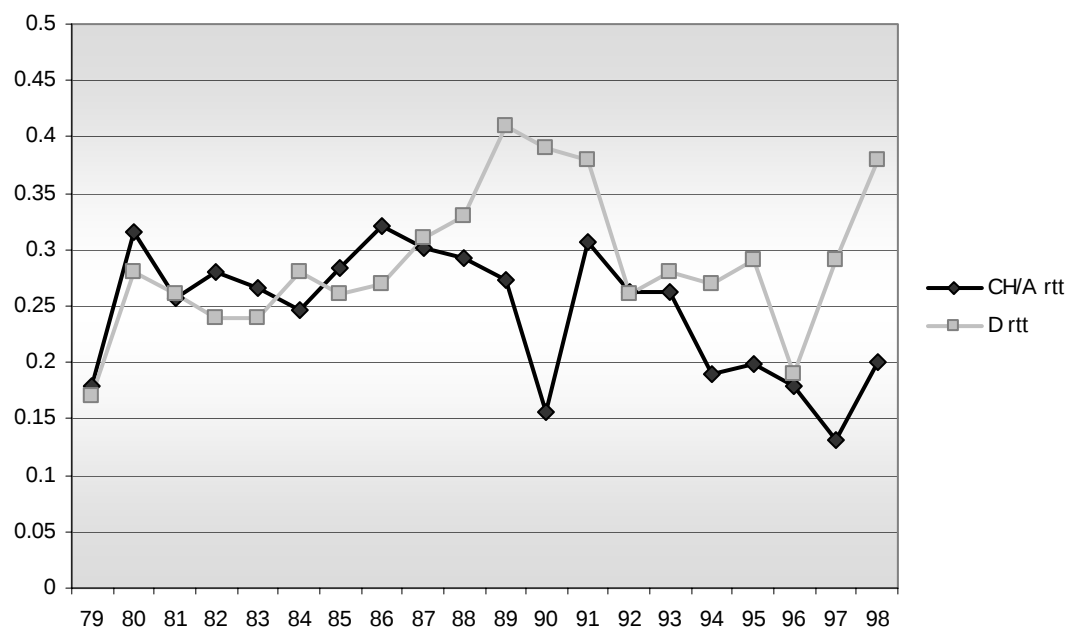


Abbildung 19: Trennschärfen für den Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“.

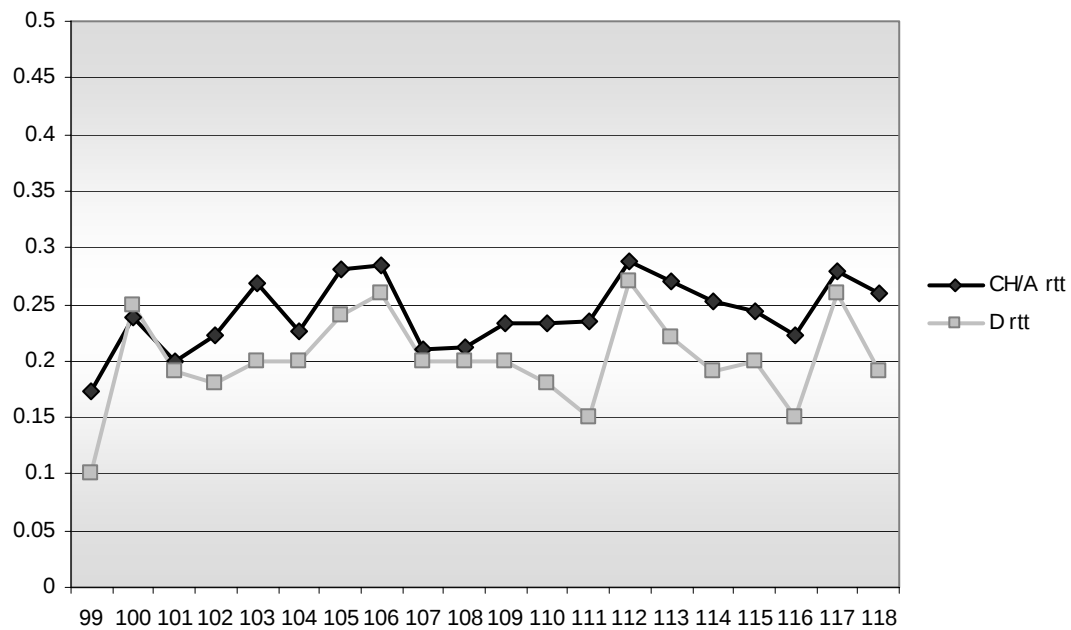


Abbildung 20: Trennschärpen für den Untertest „Figuren lernen“.

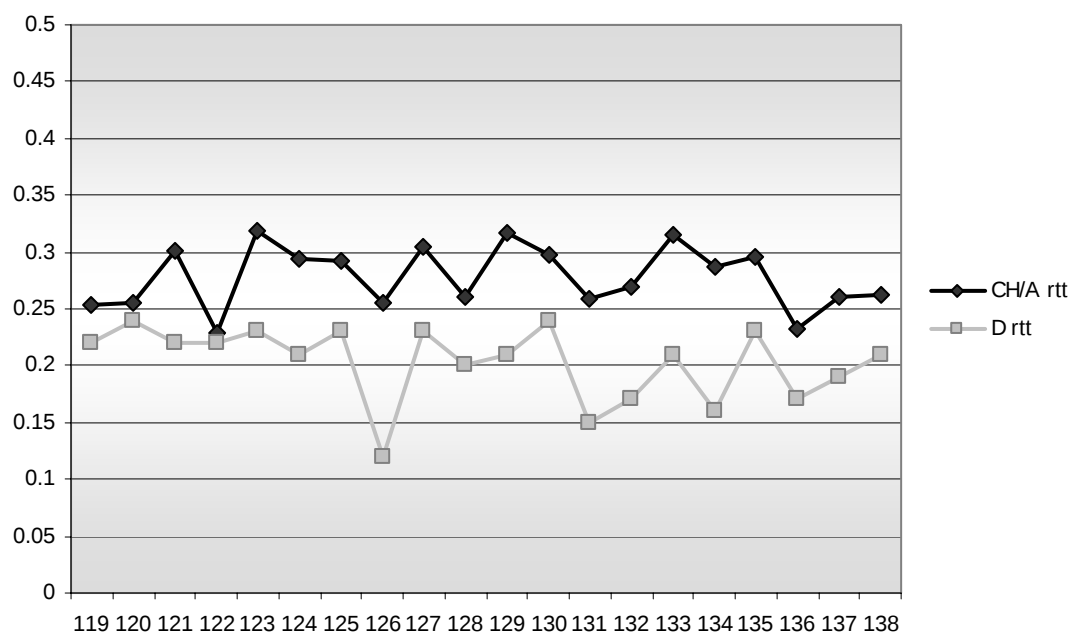


Abbildung 21: Trennschärpen für den Untertest „Fakten lernen“.

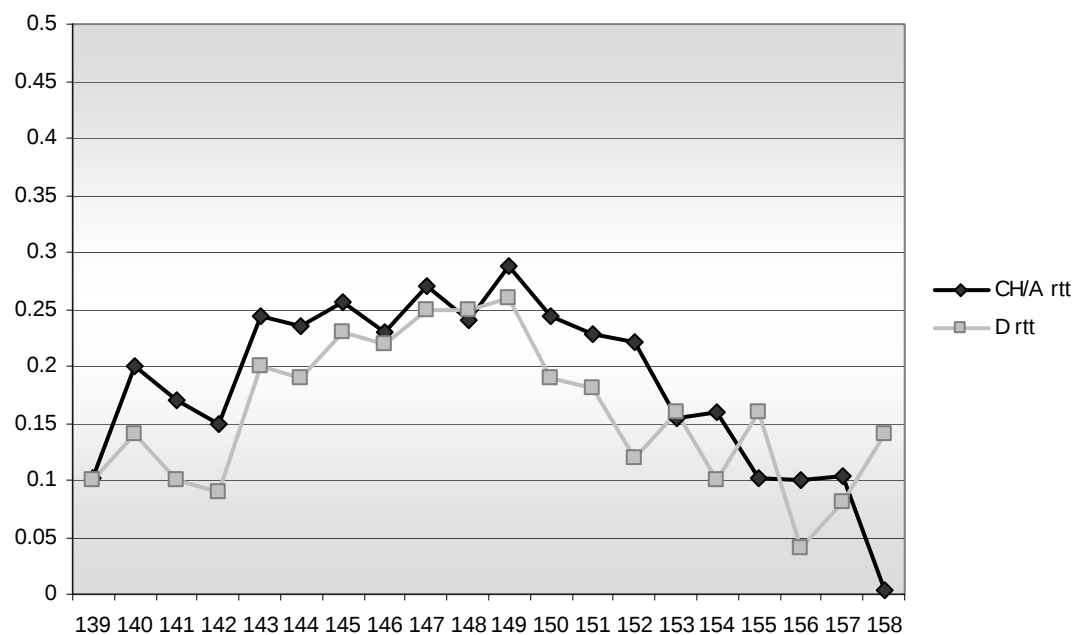


Abbildung 22: Trennschärfen für den Untertest „Muster zuordnen“.

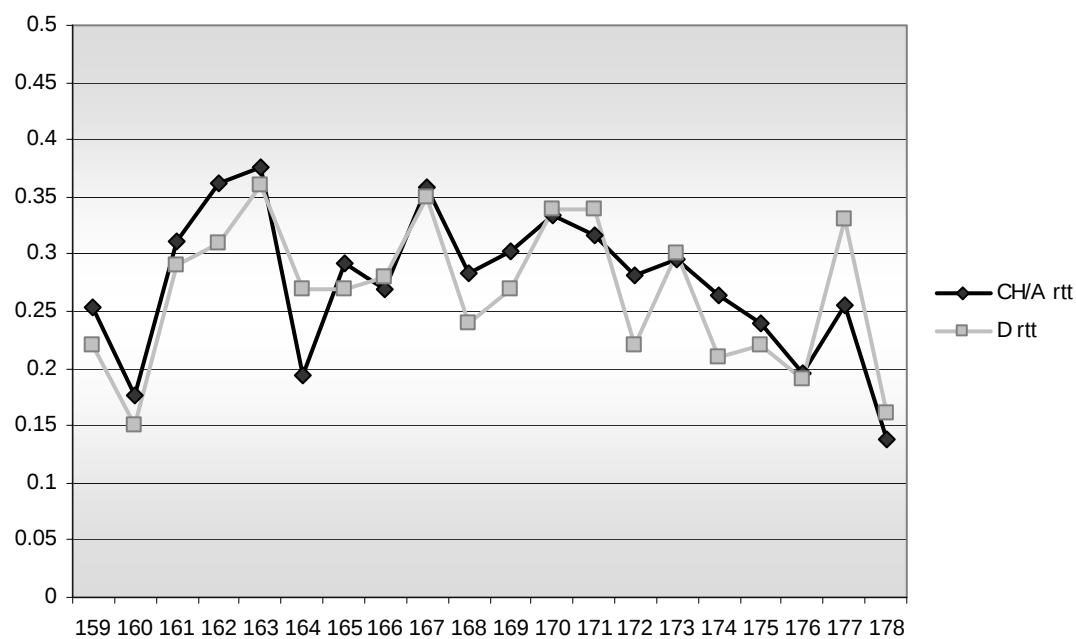


Abbildung 23: Trennschärfen für den Untertest „Diagramme und Tabellen“.

5.4 Konzentriertes und sorgfältiges Arbeiten

Dieser Untertest wurde 2004 revidiert und jährlich mit einer variierenden Anforderung (Zeichen sowie Durchstreichregel) durchgeführt, die vorher nicht bekannt ist. Es hatte sich früher gezeigt, dass durch exzessives Üben (40 bis teilweise 80mal) bei diesem Test ein so hoher Automatisierungsgrad erreicht wurde, dass nahezu alle Zeichen bearbeitet werden konnten (vgl. Bericht 9 des ZTD und Vorbereitungsreport 2003). Dieses Ergebnis ist nicht im Sinne der Messintention dieses Tests: Es soll kein „Fleißtest“ sein, sondern tatsächlich Konzentration und Sorgfalt gemessen werden.

Aus der Literatur ist bekannt, dass hochspezialisierte Automatisierungen stark bedingungsabhängig sind – ändert sich etwas, müssen diese Automatismen neu antrainiert werden (wie etwa bei der Fließbandarbeit). Indem nun jedes Jahr Zeichen und Regel vorher nicht bekannt sein werden, wirken sich Lerneffekte durch exzessives Üben nicht mehr so deutlich aus. In den Jahren 2004 bis 2007 findet sich eine glockenförmige Verteilung des Punktwertes – im Unterschied zu stark rechtsschiefen Verteilungen in den Vorjahren.

Ab 2006 wurde auch in der Test-Info eine Form dieses Untertests bereitgestellt, die formal der neuen Version entspricht. Die trotzdem weiter empfohlene Vorbereitung mit der veröffentlichten Originalversion soll dem Zweck dienen, Erfahrungen zu sammeln, wie man 8 Minuten zusammenhängend konzentriert arbeiten kann und wie man die Markierungen vornehmen muss, damit sie regelkonform sind.

Die Zahl der Zeichen auf dem Arbeitsbogen wurde von 1200 auf 1600 erhöht. Dies war möglich, weil die neue Einlesetechnik keine Markierungszonen um das eigentliche Zeichen mehr verlangt – dadurch wurde der Bogen insgesamt sogar übersichtlicher. „Deckeneffekte“, d.h. dass alle Zeichen richtig gelöst werden, sind dadurch unwahrscheinlich und der Test differenziert über den gesamten Leistungsbereich gut. Die Sorgfalt beim Markieren (nicht in Nachbarzeichen markieren, deutlich markieren) gehört bei diesem Test mit zur Anforderung – entsprechende Regelverstöße werden als Fehler gewertet.

2007 bestand der Untertest aus zwei graphischen, sehr ähnlichen Zeichen (Kreise mit Öffnungen bei „6 Uhr“, beziehungsweise „7.30 Uhr“) – nennen wir sie „a“ und „b“. Es sollte jedes „b“ markiert werden, wenn direkt danach ein „a“ folgte.

	Beschreibung	Wertebereich
Richtige	Target richtig markiert (b vor a)	0...400
Fehler I	Target nicht markiert, „übersehen“	0...400
Fehler IIa	Nontarget markiert, „falscher Alarm“: (b markiert vor b)	0...400
Fehler IIb	Distraktor markiert, „falscher Alarm“: a markiert	0...800

Der Punktwert (PW) berechnet sich wie folgt:

PW = Richtige – Fehler I – Fehler IIa – Fehler IIb

Der Rohwert Richtige minus Fehler beträgt in Österreich im Mittel 259.9 mit einer Standardabweichung von 73.1. Werden mehr Fehler als Richtige markiert, entstehen negative Werte, die auf Null gesetzt werden.

Die Standardisierung zum Punktwert von 0 bis 20 erfolgt nach folgender Regel: Aus der Verteilung der Rohwerte erhalten die unteren 2.5% der Personen den Punktwert 0, die oberen 2.5% den Punktwert 20. Der Wertebereich für die mittleren 95% der Personen wird in 19 gleiche Abschnitte geteilt, sodass sich die nachfolgende Umrechnungstabelle ergibt. Dieses Verfahren gewährleistet, dass auch unterschiedlich schwierige Konzentrationstests zu einer ver-

gleichbaren Bepunktung über die einzelnen Jahre führen und der differenzierte Wertebereich nicht durch Ausreißer nach oben und unten eingeschränkt wird. Die Leistung der mittleren 95% der Personen entspricht einer Normalverteilung.

		Anzahl	Mittelwert	Standard-abweichung	Minimum	Maximum
A	Richtige	3940	273.10	62.94	0	400
	Fehler Gesamt	3940	13.24	29.97	0	661
	... Auslassungen	3940	6.37	16.25	0	329
	... Fehlmarkierungen	3940	6.87	16.49	0	367
	letztes bearbeitetes Zeichen	3940	1116.15	252.10	408	1599

Tabelle 27: Statistiken für Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“

Punktwert 0-20	von Rohwert	bis Rohwert
0	Minimum	103
1	104	112
2	113	121
3	122	129
4	130	138
5	139	147
6	148	156
7	157	165
8	166	173
9	174	182
10	183	191
11	192	200
12	201	208
13	209	217
14	218	226
15	227	235
16	236	244
17	245	252
18	253	261
19	262	270
20	271	400

Tabelle 28: Umrechnung Rohwert in Punktwert „Konzentriertes und sorgfältiges Arbeiten“ 2007, AT

		Fehler	Auslassungen	Fehl-markierungen
Richtige	Korrelation nach Pearson	-0.13(**)	-0.13(**)	-0.11(**)
	Signifikanz (2-seitig)	0.000	0.000	0.000
Fehler Gesamt	Korrelation nach Pearson		0.91(**)	0.92(**)
	Signifikanz (2-seitig)		0.000	0.000
Auslassungen	Korrelation nach Pearson	0.91(**)		0.67(**)
	Signifikanz (2-seitig)	0.000		0.000

Tabelle 29: Korrelationen der Parameter im Konzentrationstest; **Die Korrelation ist auf dem Niveau von 0,01(2-seitig) signifikant, n = 3940

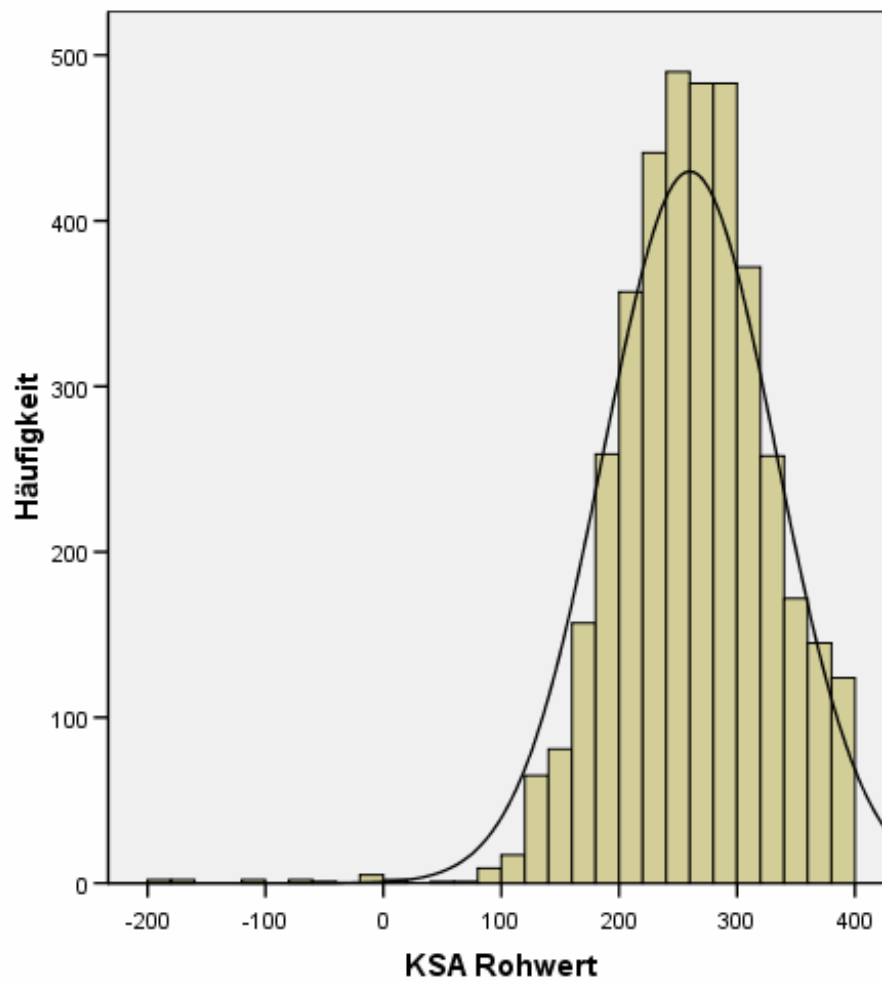


Abbildung 24: Verteilung des Rohwertes für „Konzentriertes und sorgfältiges Arbeiten 2007“, negative Werte nicht Null gesetzt. 9 Ausreißer liegen unter -200 (nicht dargestellt).

6 DIF-Analysen für Untergruppen

Eine Möglichkeit zur Identifikation von DIF (Differential Item functioning) ist die Methode „Delta-Plot“. Bei diesem Verfahren werden die zu vergleichenden Item-Schwierigkeiten z-standardisiert und anschliessend in „ Δ -Werte“ transformiert. Die Transformation erfolgt über die Formel:

$$\Delta = 13 - 4z.$$

Hohe Werte stehen für „schwierige“ (von weniger Probanden gelöste) Items.

Die aus den Delta-Werten abgeleitete Regressionsgerade $Y = AX + B$ beschreibt die Beziehung zwischen den interessierenden Sprachversionen. Eine graphische Darstellung der Delta-Werte (Delta-Plot) würde im Idealfall eine ellipsenförmige Anordnung von Punkten entlang der Diagonalen ergeben. Dies würde bedeuten, dass sowohl die Itemschwierigkeiten wie auch deren Reihenfolge in beiden Sprachversionen vergleichbar sind.

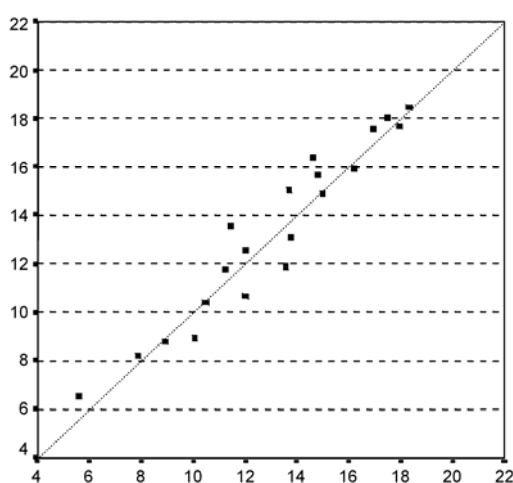


Abbildung 25: Beispiel für einen Delta-Plot.

Die Diagonale wird in den folgenden Abbildungen durch eine unterbrochene Linie dargestellt. Ein Abweichen der Werte von dieser Diagonalen ergibt eine Regressionsgerade, welche sich sowohl in Steigung wie Achsenschnittpunkt von der Diagonalen unterscheiden kann. Solche Verschiebungen der Regressionsgeraden stehen für systematische Unterschiede in den Untertests, deren Ursachen eher nicht testbedingt, sondern in realen Gegebenheiten zu vermuten sind. Um die Regressionsgerade gruppierte Items folgen also diesen Gegebenheiten und sind demzufolge nicht DIF-auffällig.

Starke Abweichungen einzelner Punkte von der Regressionsgeraden weisen hingegen auf Items hin, welche zusätzlich zu einer eventuellen systematischen Verschiebung spezielle Eigenschaften aufweisen. Bei einer parallelen Verschiebung der Regressionsgeraden

den bleibt trotz einer Veränderung der absoluten Schwierigkeiten die „Rangfolge“ der Aufgaben erhalten. Von der Regressionsgeraden abweichende Items stimmen aber in der untersuchten Gruppe bezüglich der „Schwierigkeitshierarchie“ nicht mit der Referenzgruppe überein (sie sind im Vergleich zu den anderen Items zu schwer oder zu leicht ausgefallen). Die Ursache solcher Differenzen kann auch testbedingt sein. Denkbar wäre dann, dass in der Fragestellung Konzepte enthalten sein könnten, welche in den Gruppen unterschiedliche Bekanntheitsgrade aufweisen.

Regressionsgeraden werden in der Folge als durchgezogene Linien dargestellt. Massgeblich für die Entscheidung, ob bei einem Item DIF vorliegt, ist die Distanz des entsprechenden Punktes von dieser Geraden. Die Distanz D wird nach der Formel

$$D_i = \frac{AX_i - Y_i + B}{\sqrt{A^2 + 1}}$$

berechnet, wobei unter A die Steigung und B der Achsenschnittpunkt der Regressionsgeraden zu verstehen ist, X_i bezeichnet den Delta-Wert der Referenzgruppe, Y_i denjenigen der zu vergleichenden Gruppe. In der Folge sollen innerhalb der auszugleichenden Untertests kritische Items identifiziert werden. Da für die Behandlung der Sprachgruppen in der Schweiz keine symmetrische DIF-Behandlung verwendet wird, ist die Richtung der Abweichung der Regressionsgeraden von der Diagonalen massgebend. Es sind also jene Items auffällig, **welche mindestens 1.5 Punkte** (Longford, Holland & Thayer, 1993) von der Regressionsgeraden in entgegengesetzter Richtung zur Diagonalen abweichen.

6.1 DIF nach Quoten (EU vs. AT), Humanmedizin

Für die Analyse zwischen der EU- und der AT-Kohorte werden zwei (von 178) Items identifiziert, die DIF aufweisen. Beide sind im Untertest „Textverständnis“ zu finden. Die übrigen Items weisen (wenn überhaupt) eine systematische Verschiebung auf, die sich nicht auf die Rangfolge der Schwierigkeiten auswirkt.

6.1.1 Quantitative und formale Probleme

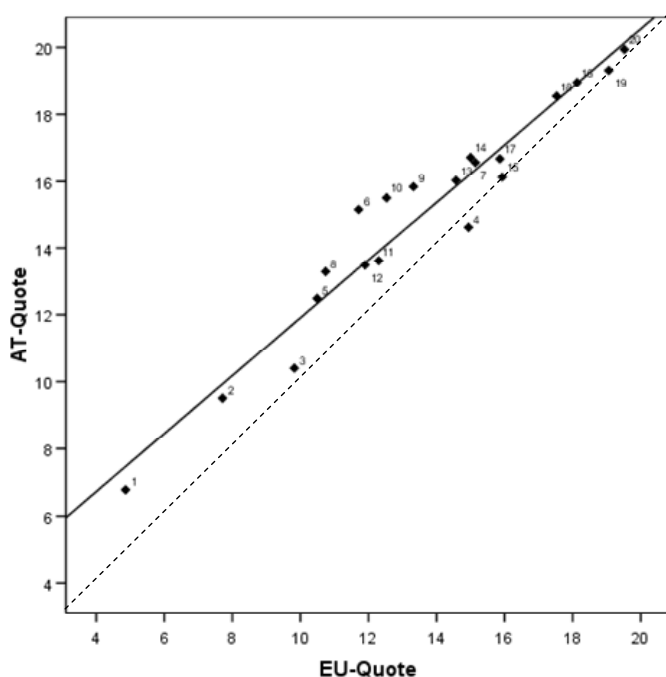


Abbildung 26: Delta-Plot für „Quantitative und formale Probleme“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
1	0.82	0.74	4.86	6.78	0.52	
2	0.70	0.63	7.71	9.51	0.32	
3	0.62	0.60	9.82	10.40	1.02	
4	0.41	0.43	14.94	14.62	1.18	
5	0.59	0.51	10.49	12.48	-0.11	
6	0.54	0.40	11.71	15.15	-1.34	
7	0.40	0.35	15.14	16.57	-0.17	
8	0.58	0.48	10.74	13.28	-0.56	
9	0.48	0.38	13.32	15.84	-0.80	
10	0.51	0.39	12.53	15.50	-1.06	
11	0.52	0.47	12.30	13.61	0.22	
12	0.53	0.47	11.90	13.48	0.05	
13	0.43	0.37	14.57	16.03	-0.13	
14	0.41	0.34	15.00	16.72	-0.37	
15	0.37	0.36	15.93	16.13	0.68	
16	0.28	0.25	18.13	18.96	-0.02	
17	0.38	0.34	15.86	16.68	0.22	
18	0.31	0.27	17.53	18.54	-0.10	
19	0.25	0.24	19.06	19.32	0.31	
20	0.23	0.21	19.52	19.94	0.14	

Tabelle 30: DIF-Analyse „Quantitative und formale Probleme“ (EU vs. AT, Humanmedizin)

6.1.2 Schlauchfiguren

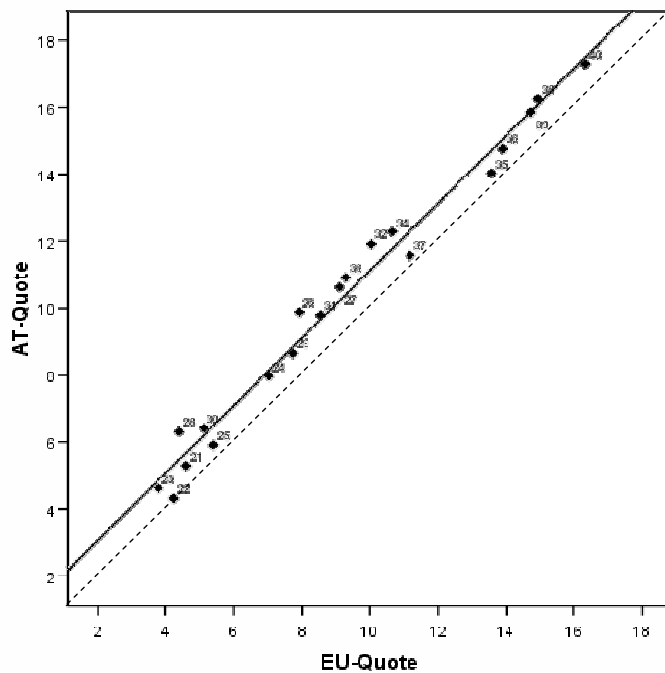


Abbildung 27: Delta-Plot für „Schlauchfiguren“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
21	0.83	0.80	4.58	5.30	0.26	
22	0.84	0.84	4.23	4.32	0.70	
23	0.86	0.83	3.77	4.62	0.16	
24	0.73	0.69	7.04	7.97	0.12	
25	0.80	0.78	5.41	5.91	0.42	
26	0.70	0.62	7.91	9.89	-0.62	
27	0.65	0.59	9.10	10.63	-0.29	
28	0.84	0.76	4.39	6.31	-0.59	
29	0.70	0.67	7.73	8.65	0.13	
30	0.81	0.76	5.11	6.41	-0.15	
31	0.67	0.62	8.56	9.80	-0.09	
32	0.61	0.53	10.05	11.91	-0.52	
33	0.64	0.57	9.28	10.92	-0.37	
34	0.59	0.52	10.65	12.28	-0.36	
35	0.47	0.45	13.57	14.04	0.47	
36	0.45	0.42	13.89	14.76	0.19	
37	0.56	0.55	11.18	11.58	0.51	
38	0.41	0.36	14.94	16.25	-0.12	
39	0.42	0.37	14.73	15.86	0.01	
40	0.36	0.32	16.33	17.30	0.13	

Tabelle 31: DIF-Analyse „Schlauchfiguren“ (EU vs. AT, Humanmedizin)

6.1.3 Textverständnis

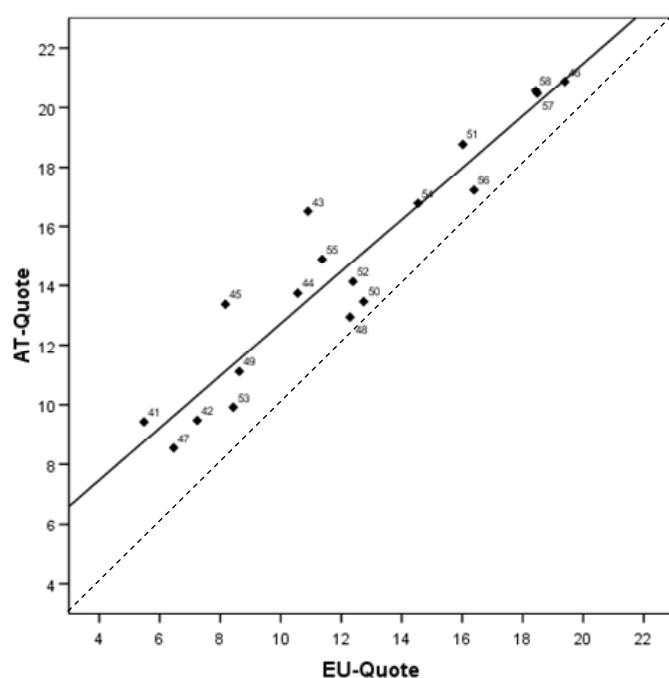


Abbildung 28: Delta-Plot für „Textverständnis“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
41	0.79	0.63	5.48	9.43	-0.50	
42	0.72	0.63	7.24	9.48	0.62	
43	0.58	0.35	10.90	16.51	-2.25	.23
44	0.59	0.46	10.56	13.74	-0.39	
45	0.69	0.48	8.17	13.37	-1.69	.21
46	0.23	0.17	19.39	20.88	0.05	
47	0.75	0.67	6.46	8.55	0.81	
48	0.52	0.49	12.29	12.94	1.34	
49	0.67	0.57	8.63	11.11	0.31	
50	0.50	0.47	12.74	13.46	1.26	
51	0.37	0.26	16.02	18.77	-0.58	
52	0.51	0.44	12.39	14.13	0.52	
53	0.67	0.61	8.43	9.92	1.08	
54	0.43	0.34	14.54	16.78	-0.06	
55	0.56	0.41	11.37	14.88	-0.72	
56	0.35	0.32	16.39	17.23	0.81	
57	0.27	0.19	18.48	20.48	-0.25	
58	0.27	0.19	18.43	20.57	-0.35	

Tabelle 32: DIF-Analyse „Textverständnis“ (EU vs. AT, Humanmedizin)

6.1.4 Planen & Organisieren

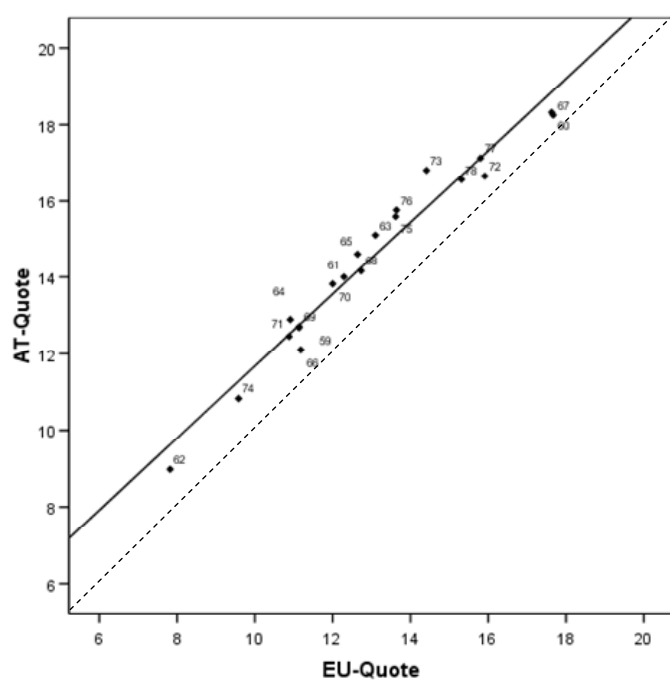


Abbildung 29: Delta-Plot für „Planen & Organisieren“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
59	0.57	0.50	11.13	12.68	0.04	
60	0.30	0.28	17.63	18.32	0.39	
61	0.52	0.45	12.29	14.00	-0.13	
62	0.70	0.65	7.82	8.98	0.47	
63	0.49	0.41	13.10	15.10	-0.37	
64	0.57	0.49	10.91	12.89	-0.26	
65	0.50	0.43	12.64	14.60	-0.32	
66	0.56	0.53	11.18	12.10	0.50	
67	0.30	0.28	17.67	18.25	0.46	
68	0.50	0.44	12.73	14.16	0.06	
69	0.57	0.50	11.14	12.70	0.04	
70	0.53	0.46	12.00	13.82	-0.19	
71	0.58	0.51	10.88	12.45	0.04	
72	0.37	0.34	15.91	16.66	0.42	
73	0.43	0.34	14.41	16.81	-0.71	
74	0.63	0.58	9.58	10.83	0.33	
75	0.47	0.39	13.62	15.59	-0.37	
76	0.46	0.38	13.64	15.76	-0.48	
77	0.38	0.32	15.80	17.13	0.00	
78	0.40	0.35	15.31	16.57	0.08	

Tabelle 33: DIF-Analyse „Planen & Organisieren“ (EU vs. AT, Humanmedizin)

6.1.5 Medizinisch-naturwissenschaftliches Grundverständnis

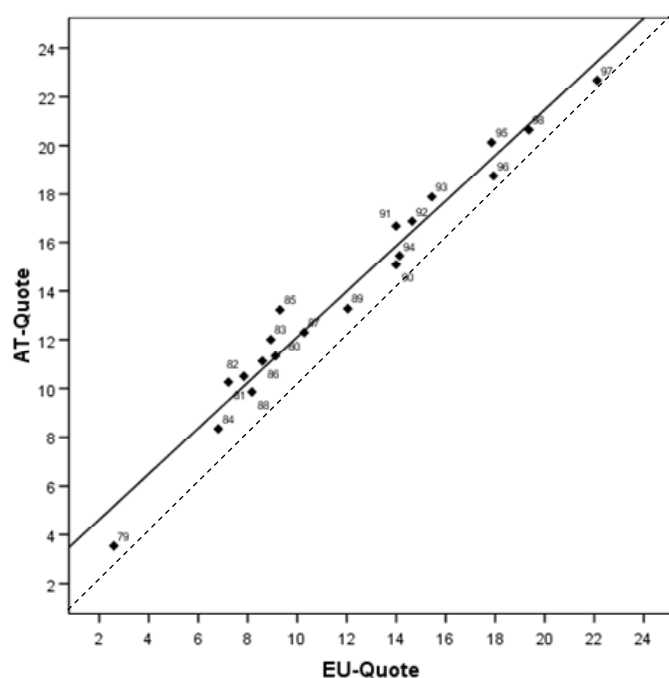


Abbildung 30: Delta-Plot für „Medizinisch-naturwissenschaftliches Grundverständnis“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
79	0.91	0.87	2.59	3.54	1.20	
80	0.65	0.56	9.12	11.33	-0.03	
81	0.72	0.60	7.22	10.26	-0.55	
82	0.70	0.59	7.84	10.50	-0.30	
83	0.65	0.53	8.93	12.01	-0.66	
84	0.74	0.68	6.81	8.35	0.57	
85	0.64	0.48	9.30	13.22	-1.29	
86	0.67	0.57	8.59	11.12	-0.24	
87	0.60	0.52	10.28	12.30	0.05	
88	0.69	0.62	8.17	9.86	0.39	
89	0.53	0.48	12.04	13.27	0.55	
90	0.45	0.41	13.99	15.08	0.56	
91	0.45	0.34	13.99	16.68	-0.61	
92	0.42	0.33	14.64	16.88	-0.31	
93	0.39	0.29	15.44	17.89	-0.51	
94	0.44	0.39	14.13	15.45	0.39	
95	0.30	0.20	17.85	20.13	-0.50	
96	0.29	0.26	17.93	18.76	0.57	
97	0.12	0.10	22.12	22.68	0.56	
98	0.23	0.18	19.36	20.65	0.16	

Tabelle 34: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“ (EU vs. AT, Humanmedizin)

6.1.6 Figuren lernen

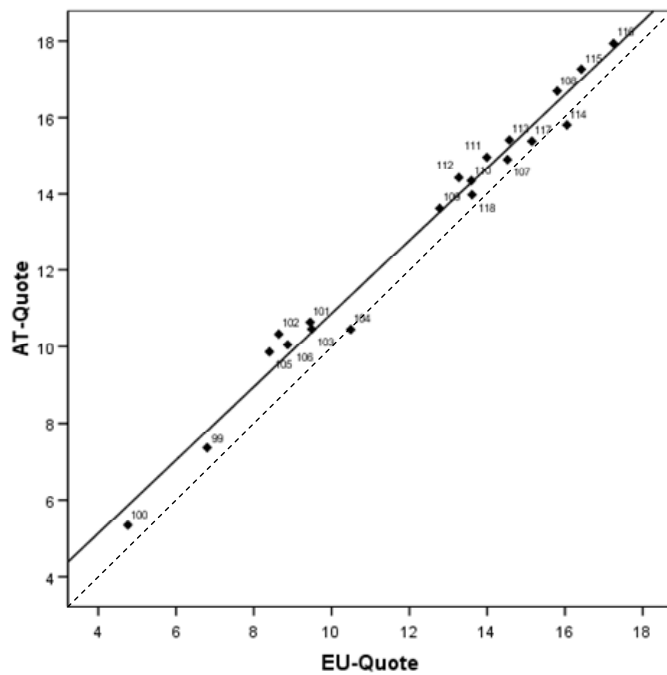


Abbildung 31: Delta-Plot für „Figuren lernen“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
99	0.74	0.72	6.80	7.36	0.32	
100	0.82	0.80	4.76	5.34	0.37	
101	0.63	0.59	9.45	10.64	-0.22	
102	0.67	0.60	8.64	10.33	-0.55	
103	0.63	0.59	9.49	10.46	-0.06	
104	0.59	0.59	10.49	10.45	0.64	
105	0.68	0.62	8.40	9.86	-0.38	
106	0.66	0.61	8.87	10.05	-0.19	
107	0.43	0.41	14.52	14.91	0.20	
108	0.38	0.34	15.80	16.69	-0.20	
109	0.50	0.47	12.78	13.62	-0.07	
110	0.47	0.44	13.59	14.35	-0.04	
111	0.45	0.41	13.99	14.97	-0.21	
112	0.48	0.43	13.27	14.43	-0.32	
113	0.43	0.39	14.57	15.42	-0.13	
114	0.37	0.38	16.05	15.81	0.60	
115	0.35	0.32	16.42	17.27	-0.20	
116	0.32	0.29	17.25	17.94	-0.11	
117	0.40	0.39	15.15	15.39	0.29	
118	0.47	0.45	13.61	13.98	0.24	

Tabelle 35: DIF-Analyse „Figuren lernen“ (EU vs. AT, Humanmedizin)

6.1.7 Fakten lernen

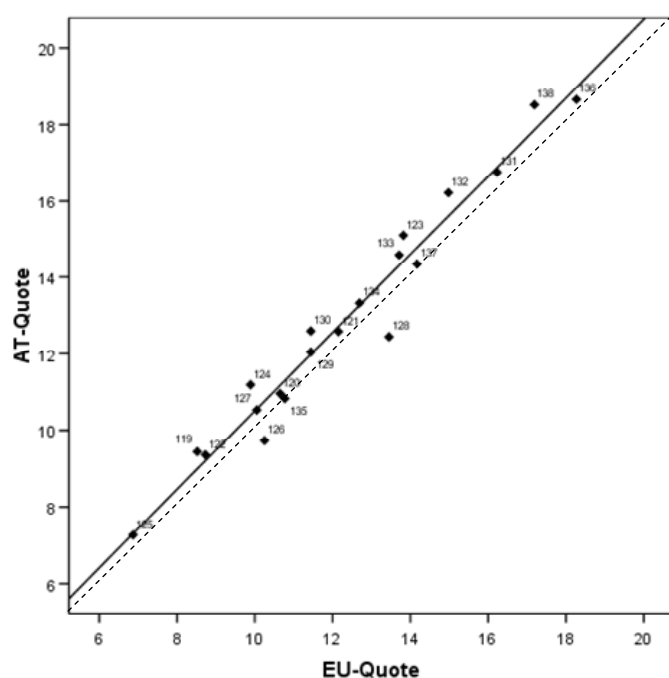


Abbildung 32: Delta-Plot für „Fakten lernen“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
119	0.67	0.63	8.52	9.44	-0.32	
120	0.59	0.57	10.65	10.96	0.15	
121	0.52	0.51	12.15	12.58	0.09	
122	0.66	0.64	8.73	9.36	-0.10	
123	0.46	0.41	13.82	15.10	-0.48	
124	0.62	0.56	9.89	11.19	-0.56	
125	0.74	0.72	6.87	7.27	0.02	
126	0.60	0.62	10.25	9.73	0.71	
127	0.61	0.59	10.05	10.53	0.02	
128	0.47	0.51	13.45	12.44	1.11	
129	0.55	0.53	11.44	12.04	-0.04	
130	0.55	0.51	11.44	12.59	-0.43	
131	0.36	0.34	16.23	16.76	0.09	
132	0.41	0.36	14.98	16.22	-0.43	
133	0.46	0.43	13.71	14.58	-0.19	
134	0.50	0.48	12.69	13.32	-0.04	
135	0.58	0.58	10.77	10.83	0.33	
136	0.28	0.26	18.27	18.66	0.22	
137	0.44	0.44	14.17	14.34	0.30	
138	0.32	0.27	17.19	18.52	-0.45	

Tabelle 36: DIF-Analyse „Fakten lernen“ (EU vs. AT, Humanmedizin)

6.1.8 Muster zuordnen

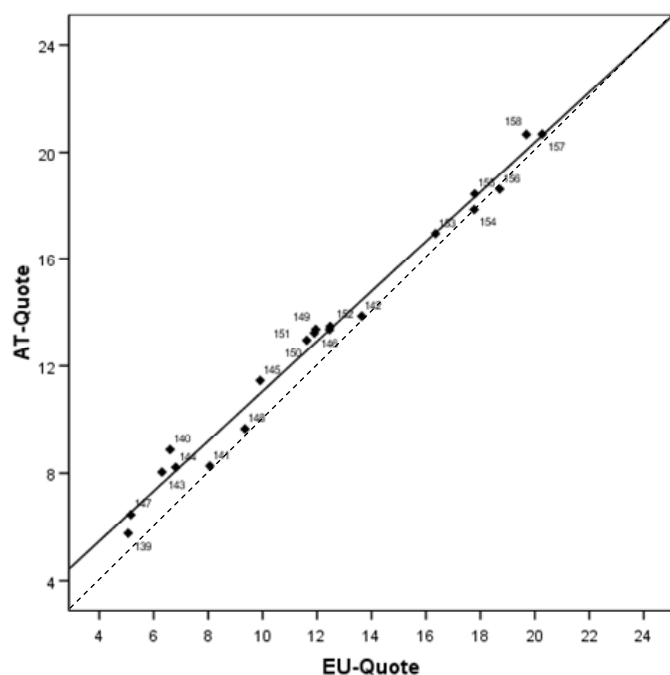


Abbildung 33: Delta-Plot für „Muster zuordnen“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
139	0.81	0.78	5.06	5.76	0.51	
140	0.75	0.66	6.60	8.88	-0.72	
141	0.69	0.68	8.06	8.27	0.72	
142	0.46	0.46	13.64	13.86	0.43	
143	0.76	0.69	6.30	8.04	-0.31	
144	0.74	0.68	6.81	8.22	-0.10	
145	0.61	0.55	9.91	11.46	-0.35	
146	0.51	0.48	12.46	13.36	0.00	
147	0.81	0.75	5.16	6.46	0.06	
148	0.64	0.63	9.35	9.66	0.58	
149	0.53	0.48	11.95	13.37	-0.36	
150	0.55	0.49	11.62	12.96	-0.29	
151	0.53	0.48	11.90	13.23	-0.30	
152	0.51	0.47	12.48	13.48	-0.08	
153	0.36	0.33	16.35	16.95	0.02	
154	0.30	0.29	17.77	17.85	0.33	
155	0.30	0.27	17.79	18.44	-0.09	
156	0.26	0.26	18.71	18.62	0.40	
157	0.20	0.18	20.27	20.68	-0.04	
158	0.22	0.18	19.69	20.67	-0.42	

Tabelle 37: DIF-Analyse „Muster zuordnen“ (EU vs. AT, Humanmedizin)

6.1.9 Diagramme und Tabellen

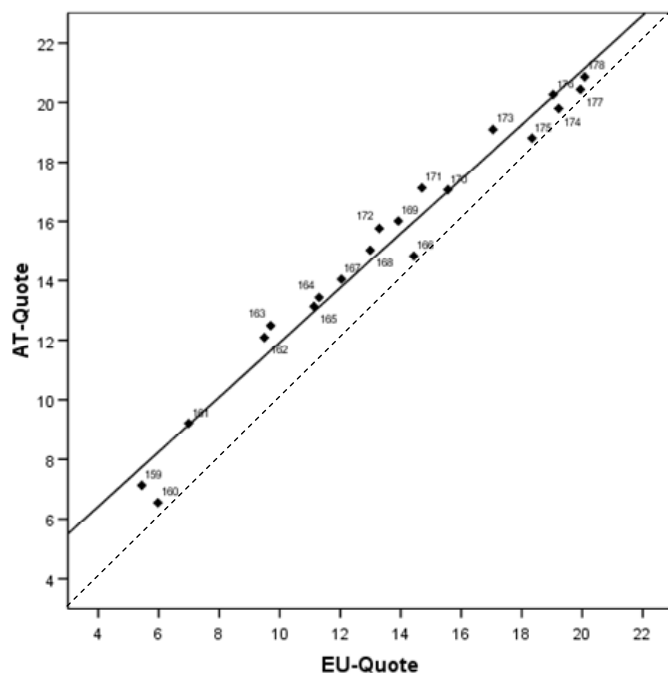


Abbildung 34: Delta-Plot für „Diagramme und Tabellen“ (EU vs. AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	EU	AT	EU	AT		
159	0.80	0.73	5.44	7.13	0.45	
160	0.77	0.75	5.97	6.55	1.23	
161	0.73	0.64	6.99	9.21	-0.03	
162	0.63	0.53	9.49	12.09	-0.47	
163	0.62	0.51	9.70	12.49	-0.62	
164	0.56	0.47	11.30	13.43	-0.24	
165	0.57	0.49	11.13	13.13	-0.13	
166	0.43	0.42	14.43	14.82	0.86	
167	0.53	0.45	12.04	14.04	-0.18	
168	0.49	0.41	12.99	15.02	-0.26	
169	0.45	0.37	13.92	16.01	-0.37	
170	0.39	0.33	15.56	17.07	-0.04	
171	0.42	0.32	14.70	17.13	-0.67	
172	0.48	0.38	13.29	15.76	-0.61	
173	0.33	0.24	17.05	19.10	-0.52	
174	0.24	0.22	19.22	19.80	0.42	
175	0.28	0.26	18.34	18.81	0.56	
176	0.25	0.20	19.04	20.26	-0.03	
177	0.21	0.19	19.94	20.43	0.45	
178	0.20	0.17	20.08	20.88	0.21	

Tabelle 38: DIF-Analyse „Diagramme und Tabellen“ (EU vs. AT, Humanmedizin)

6.2 DIF nach Geschlecht (AT, Humanmedizin)

Auch für den Gendervergleich bezüglich allfälliger testbedingter Unterschiede auf Itemebene weisen insgesamt nur zwei Items (von 178) auffällige DIF-Werte auf. Beide sind im Untertest „Quantitative und formale Probleme“ zu finden. Bei einem Ausgleich würden hier den Teilnehmerinnen maximal .46 Punkte gutgeschrieben (bei einseitiger Anwendung des DIF Verfahrens und wenn beide Items nicht korrekt beantwortet wurden). Da es sich den Gender-

vergleichen – anders als bei den Sprachgruppen in der Schweiz – nicht um Minoritäten handelt, müsste allerdings auch ein symmetrischer DIF-Ausgleich in Erwägung gezogen werden (es gibt keine plausible Begründung, die eine oder andere Stichprobe als Referenzstichprobe zu betrachten). In diesem Fall würden sich leichte Vor- und Nachteile auf Itemebene praktisch die Waage halten (Items 8 und 10 vs. Items 4 und 6).

6.2.1 Quantitative und formale Probleme

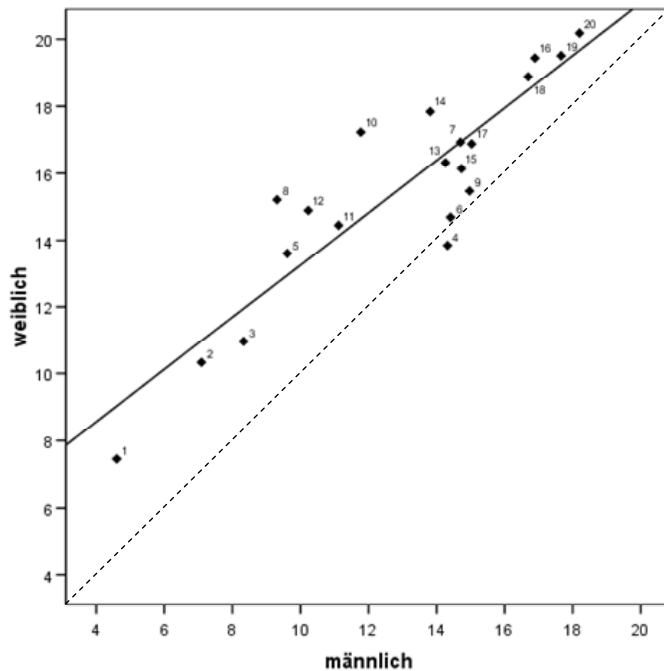


Abbildung 35: Delta-Plot für „Quantitative und formale Probleme“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
1	0.81	0.69	4.60	7.45	1.25	
2	0.71	0.58	7.09	10.33	0.51	
3	0.66	0.55	8.33	10.96	0.78	
4	0.41	0.43	14.32	13.85	2.19	
5	0.61	0.44	9.61	13.61	-0.52	
6	0.41	0.40	14.41	14.69	1.58	
7	0.40	0.31	14.70	16.93	0.01	
8	0.62	0.38	9.31	15.21	-1.97	.24
9	0.39	0.37	14.97	15.47	1.31	
10	0.52	0.30	11.77	17.23	-2.04	.22
11	0.54	0.41	11.12	14.45	-0.26	
12	0.58	0.39	10.23	14.89	-1.14	
13	0.42	0.33	14.26	16.32	0.21	
14	0.43	0.27	13.81	17.84	-1.26	
15	0.40	0.34	14.73	16.14	0.64	
16	0.31	0.21	16.89	19.44	-0.63	
17	0.39	0.31	15.03	16.88	0.24	
18	0.32	0.23	16.69	18.88	-0.31	
19	0.28	0.20	17.66	19.51	-0.20	
20	0.26	0.18	18.20	20.18	-0.40	

Tabelle 39: DIF-Analyse „Quantitative und formale Probleme“ (m vs. w, AT, Humanmedizin)

6.2.2 Schlauchfiguren

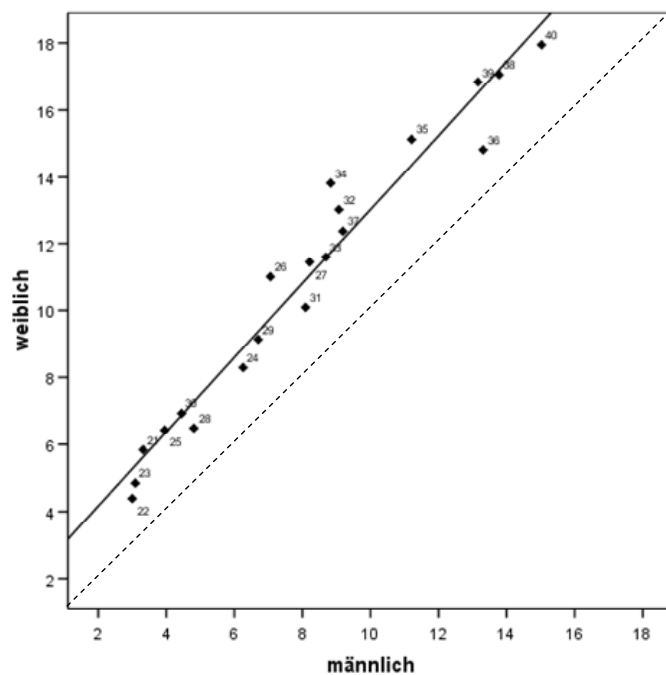


Abbildung 36: Delta-Plot für „Schlauchfiguren“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
21	0.86	0.76	3.32	5.84	-0.14	
22	0.87	0.82	3.00	4.39	0.59	
23	0.87	0.80	3.09	4.85	0.35	
24	0.74	0.66	6.26	8.29	0.39	
25	0.83	0.73	3.95	6.43	-0.07	
26	0.71	0.55	7.06	11.00	-0.83	
27	0.66	0.53	8.21	11.44	-0.26	
28	0.80	0.73	4.81	6.49	0.52	
29	0.72	0.62	6.70	9.14	0.16	
30	0.81	0.71	4.45	6.93	-0.03	
31	0.67	0.59	8.09	10.09	0.55	
32	0.63	0.47	9.07	13.02	-0.69	
33	0.64	0.52	8.69	11.59	-0.01	
34	0.64	0.43	8.83	13.82	-1.40	
35	0.54	0.38	11.21	15.13	-0.52	
36	0.45	0.39	13.31	14.82	1.25	
37	0.62	0.49	9.19	12.37	-0.16	
38	0.44	0.30	13.78	17.06	0.10	
39	0.46	0.31	13.16	16.84	-0.21	
40	0.39	0.27	15.03	17.95	0.43	

Tabelle 40: DIF-Analyse „Schlauchfiguren“ (m vs. w, AT, Humanmedizin)

6.2.3 Textverständnis

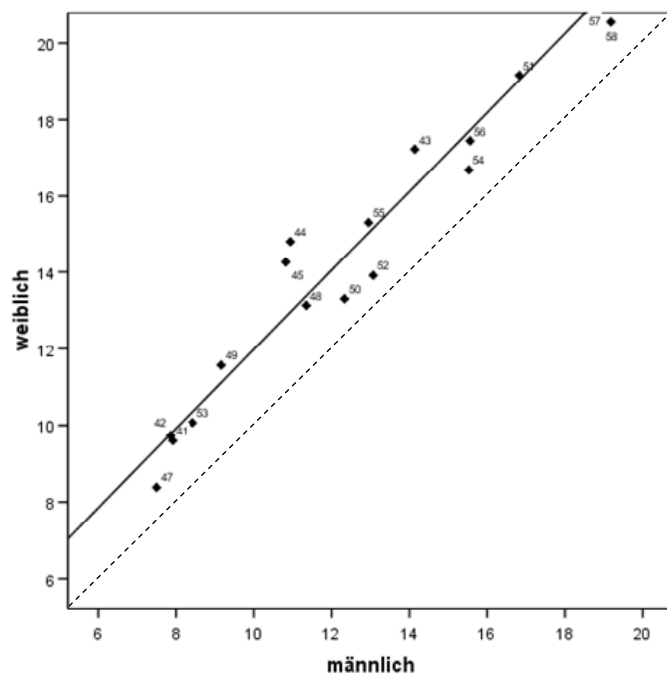


Abbildung 37: Delta-Plot für „Textverständnis“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
41	0.67	0.61	7.92	9.60	0.16	
42	0.68	0.60	7.86	9.73	0.03	
43	0.42	0.30	14.14	17.23	-0.68	
44	0.55	0.39	10.94	14.80	-1.29	
45	0.56	0.42	10.82	14.26	-0.99	
46	0.25	0.11	18.29	21.74	-0.83	
47	0.69	0.65	7.50	8.38	0.71	
48	0.53	0.46	11.35	13.13	0.17	
49	0.62	0.53	9.16	11.57	-0.32	
50	0.49	0.46	12.33	13.30	0.76	
51	0.31	0.22	16.83	19.16	-0.08	
52	0.46	0.43	13.07	13.91	0.87	
53	0.65	0.59	8.42	10.07	0.19	
54	0.36	0.32	15.53	16.69	0.70	
55	0.47	0.37	12.95	15.30	-0.18	
56	0.36	0.29	15.56	17.45	0.20	
57	0.24	0.15	18.52	20.87	-0.06	
58	0.22	0.16	19.18	20.55	0.64	

Tabelle 41: DIF-Analyse „Textverständnis“ (m vs. w, AT, Humanmedizin)

6.2.4 Planen & Organisieren

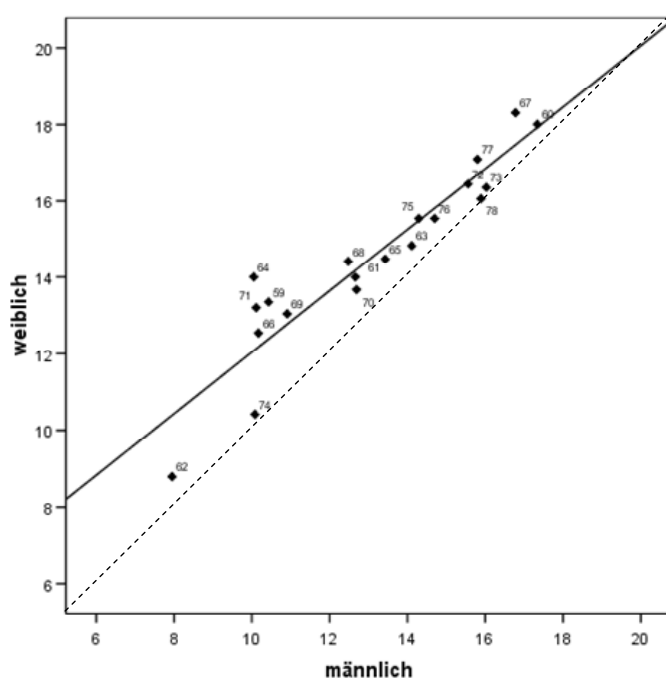


Abbildung 38: Delta-Plot für „Planen & Organisieren“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
59	0.57	0.45	10.43	13.35	-0.75	
60	0.29	0.26	17.34	18.01	-0.06	
61	0.48	0.43	12.66	14.00	0.14	
62	0.67	0.64	7.95	8.79	1.25	
63	0.42	0.39	14.11	14.82	0.40	
64	0.59	0.43	10.05	14.00	-1.50	
65	0.45	0.41	13.43	14.47	0.25	
66	0.58	0.49	10.17	12.54	-0.29	
67	0.31	0.25	16.78	18.31	-0.65	
68	0.49	0.41	12.48	14.41	-0.30	
69	0.55	0.47	10.91	13.04	-0.21	
70	0.48	0.44	12.69	13.67	0.41	
71	0.58	0.46	10.11	13.20	-0.83	
72	0.36	0.33	15.56	16.45	0.04	
73	0.34	0.33	16.03	16.36	0.41	
74	0.59	0.57	10.08	10.42	1.31	
75	0.42	0.36	14.29	15.54	-0.05	
76	0.40	0.36	14.70	15.54	0.21	
77	0.35	0.30	15.80	17.10	-0.32	
78	0.35	0.34	15.89	16.06	0.55	

Tabelle 42: DIF-Analyse „Planen & Organisieren“ (m vs. w, AT, Humanmedizin)

6.2.5 Medizinisch-naturwissenschaftliches Grundverständnis

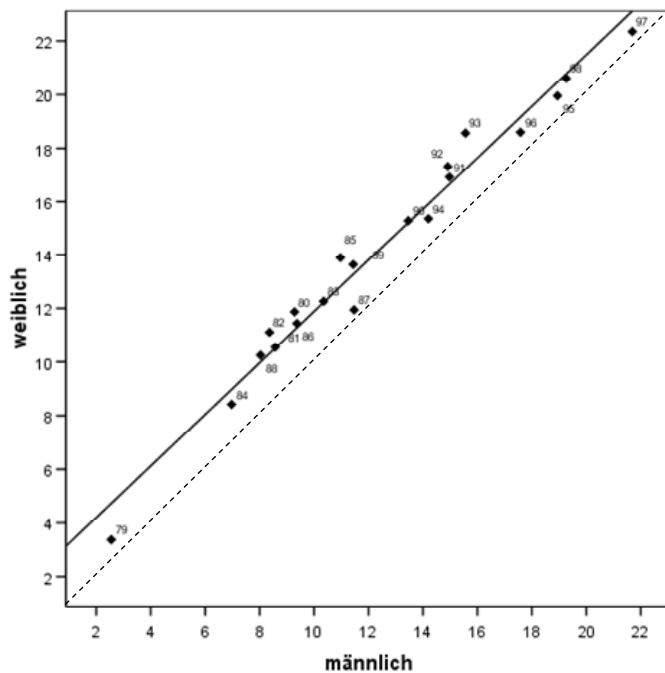


Abbildung 39: Delta-Plot für „Medizinisch-naturwissenschaftliches Grundverständnis“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
79	0.89	0.86	2.55	3.37	0.98	
80	0.62	0.51	9.28	11.87	-0.49	
81	0.65	0.57	8.57	10.55	-0.03	
82	0.66	0.54	8.36	11.11	-0.58	
83	0.58	0.50	10.35	12.26	-0.03	
84	0.71	0.65	6.97	8.42	0.40	
85	0.55	0.43	10.97	13.89	-0.78	
86	0.61	0.53	9.37	11.44	-0.12	
87	0.53	0.51	11.47	11.94	0.98	
88	0.67	0.58	8.03	10.25	-0.18	
89	0.53	0.44	11.44	13.63	-0.26	
90	0.45	0.38	13.46	15.28	-0.05	
91	0.39	0.31	14.97	16.93	-0.20	
92	0.39	0.29	14.91	17.32	-0.52	
93	0.36	0.24	15.56	18.57	-0.97	
94	0.42	0.37	14.20	15.36	0.40	
95	0.23	0.19	18.94	19.96	0.36	
96	0.28	0.24	17.58	18.60	0.41	
97	0.12	0.09	21.69	22.35	0.55	
98	0.21	0.16	19.26	20.61	0.12	

Tabelle 43: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“ (m vs. w, AT, Humanmedizin)

6.2.6 Figuren lernen

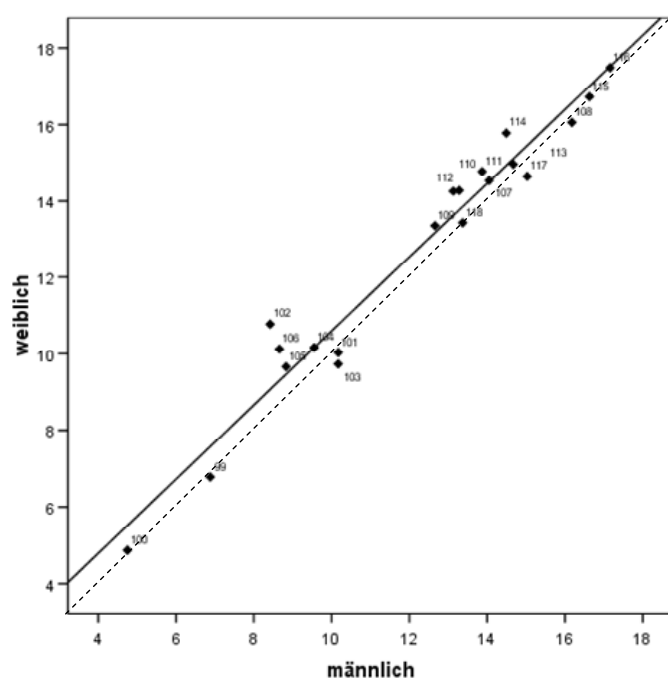


Abbildung 40: Delta-Plot für „Figuren lernen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
99	0.72	0.72	6.88	6.78	0.57	
100	0.80	0.80	4.75	4.87	0.46	
101	0.58	0.59	10.17	10.03	0.52	
102	0.65	0.56	8.42	10.77	-1.22	
103	0.58	0.60	10.17	9.73	0.74	
104	0.61	0.58	9.55	10.16	0.00	
105	0.64	0.60	8.83	9.66	-0.14	
106	0.64	0.58	8.66	10.12	-0.59	
107	0.42	0.41	14.05	14.54	-0.02	
108	0.34	0.34	16.18	16.06	0.37	
109	0.48	0.45	12.66	13.35	-0.13	
110	0.46	0.42	13.13	14.26	-0.46	
111	0.43	0.40	13.87	14.78	-0.32	
112	0.46	0.42	13.28	14.28	-0.37	
113	0.40	0.39	14.67	14.97	0.10	
114	0.41	0.36	14.49	15.78	-0.60	
115	0.32	0.32	16.63	16.73	0.20	
116	0.30	0.29	17.16	17.49	0.02	
117	0.39	0.40	15.03	14.65	0.58	
118	0.45	0.45	13.37	13.43	0.30	

Tabelle 44: DIF-Analyse „Figuren lernen“ (m vs. w, AT, Humanmedizin)

6.2.7 Fakten lernen

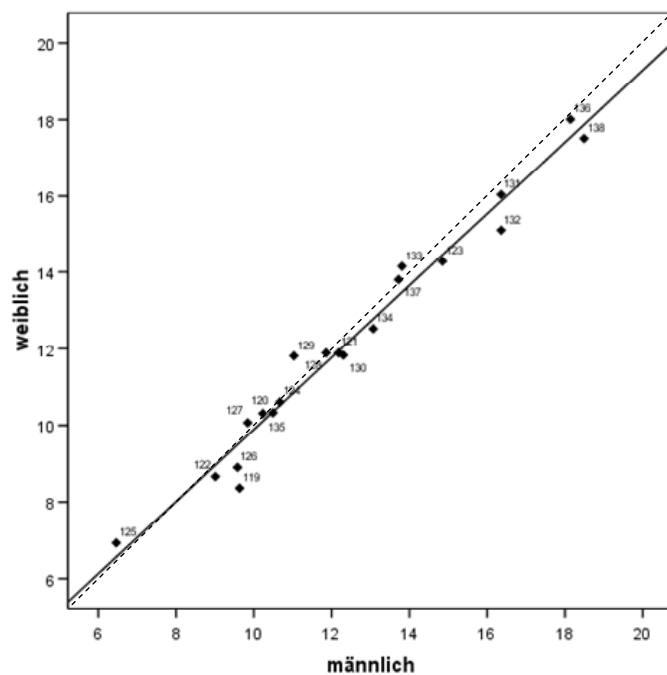


Abbildung 41: Delta-Plot für „Fakten lernen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
119	0.60	0.66	9.63	8.36	0.87	
120	0.57	0.58	10.49	10.33	0.02	
121	0.50	0.51	12.18	11.89	0.04	
122	0.63	0.64	9.01	8.66	0.22	
123	0.39	0.42	14.85	14.28	0.12	
124	0.56	0.56	10.67	10.61	-0.07	
125	0.73	0.71	6.46	6.93	-0.26	
126	0.61	0.63	9.58	8.90	0.43	
127	0.60	0.59	9.84	10.07	-0.24	
128	0.51	0.51	11.86	11.89	-0.19	
129	0.55	0.52	11.03	11.81	-0.69	
130	0.50	0.51	12.30	11.83	0.16	
131	0.33	0.34	16.36	16.04	-0.12	
132	0.33	0.38	16.36	15.10	0.56	
133	0.43	0.42	13.81	14.15	-0.49	
134	0.46	0.49	13.07	12.52	0.19	
135	0.58	0.58	10.23	10.31	-0.15	
136	0.26	0.26	18.14	18.01	-0.34	
137	0.44	0.43	13.72	13.80	-0.30	
138	0.24	0.28	18.49	17.51	0.27	

Tabelle 45: DIF-Analyse „Fakten lernen“ (m vs. w, AT, Humanmedizin)

6.2.8 Muster zuordnen

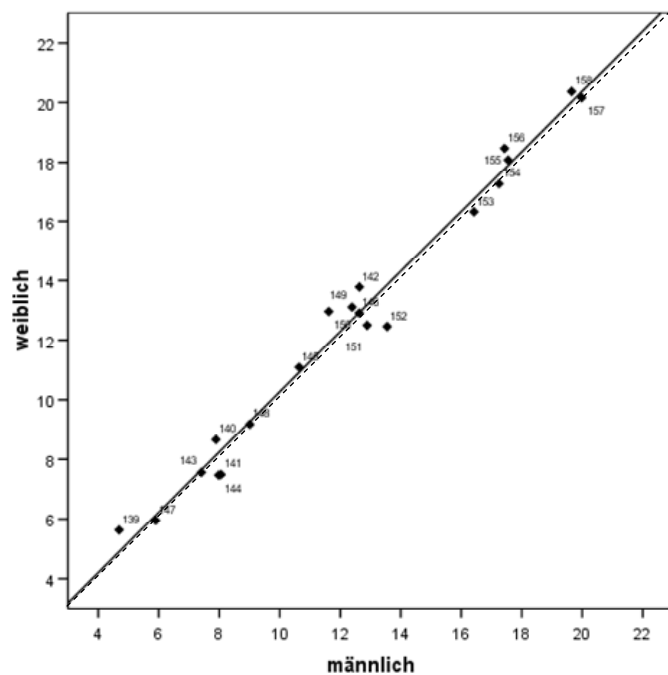


Abbildung 42: Delta-Plot für „Muster zuordnen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
139	0.80	0.77	4.69	5.63	-0.51	
140	0.67	0.64	7.89	8.66	-0.37	
141	0.67	0.69	8.06	7.49	0.58	
142	0.48	0.44	12.63	13.78	-0.60	
143	0.69	0.69	7.41	7.56	0.07	
144	0.67	0.69	7.98	7.47	0.53	
145	0.56	0.54	10.64	11.09	-0.12	
146	0.48	0.47	12.63	12.91	0.01	
147	0.76	0.75	5.90	5.95	0.12	
148	0.63	0.62	9.01	9.18	0.06	
149	0.49	0.46	12.39	13.11	-0.30	
150	0.52	0.47	11.62	12.96	-0.74	
151	0.47	0.49	12.89	12.50	0.49	
152	0.45	0.49	13.55	12.46	0.98	
153	0.33	0.33	16.42	16.32	0.31	
154	0.30	0.29	17.25	17.27	0.23	
155	0.28	0.26	17.55	18.08	-0.13	
156	0.29	0.25	17.43	18.47	-0.48	
157	0.18	0.18	19.98	20.16	0.14	
158	0.20	0.17	19.65	20.37	-0.25	

Tabelle 46: DIF-Analyse „Muster zuordnen“ (m vs. w, AT, Humanmedizin)

6.2.9 Diagramme und Tabellen

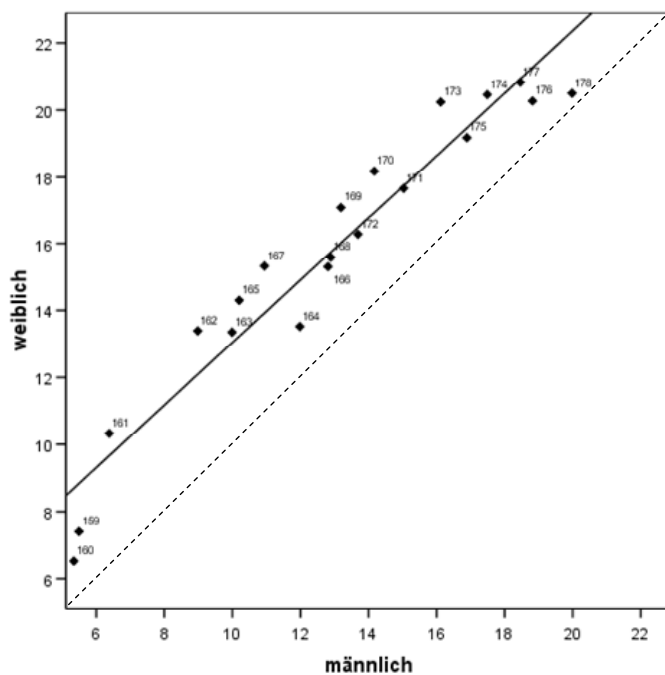


Abbildung 43: Delta-Plot für „Diagramme und Tabellen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
159	0.77	0.69	5.49	7.40	1.04	
160	0.78	0.73	5.34	6.52	1.59	
161	0.74	0.58	6.38	10.33	-0.49	
162	0.63	0.45	8.98	13.39	-0.95	
163	0.59	0.45	9.99	13.35	-0.23	
164	0.51	0.45	11.98	13.52	1.00	
165	0.58	0.41	10.20	14.30	-0.79	
166	0.48	0.37	12.81	15.30	0.26	
167	0.55	0.37	10.94	15.32	-1.03	
168	0.47	0.36	12.89	15.58	0.12	
169	0.46	0.30	13.19	17.08	-0.78	
170	0.42	0.26	14.17	18.18	-0.92	
171	0.39	0.28	15.03	17.66	0.05	
172	0.44	0.33	13.69	16.28	0.15	
173	0.34	0.17	16.12	20.24	-1.09	
174	0.29	0.17	17.49	20.46	-0.32	
175	0.31	0.22	16.89	19.18	0.21	
176	0.23	0.17	18.82	20.27	0.73	
177	0.25	0.15	18.46	20.83	0.08	
178	0.18	0.16	19.98	20.50	1.35	

Tabelle 47: DIF-Analyse „Diagramme und Tabellen“ (m vs. w, AT, Humanmedizin)

7 Beispielaufgaben für die Untertests

Nachfolgend wird pro Untertest eine Beispielaufgabe dargestellt. So können lediglich die Prinzipien der Aufgabenstruktur verdeutlicht werden – die Aufgaben unterscheiden sich innerhalb jedes Untertests bezüglich des Schwierigkeitsgrades und der Anforderung.

7.1 Untertest: Quantitative und formale Probleme

Mit Hilfe dieses Untertests wird die Fähigkeit überprüft, im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen mit Zahlen, Größen, Einheiten und Formeln richtig umzugehen. Diese Anforderung dürfte für mehrere Fächer des Grundlagenstudiums der Medizin bedeutsam sein.

Zum Beispiel:

Eine Broteinheit (BE) ist definiert als diejenige Nahrungsmenge in Gramm, die 12 Gramm Kohlenhydrate enthält. Bei der Verbrennung von 1 g Kohlenhydraten im Organismus werden 16 Kilojoule (kJ) an Energie frei. Ein Patient, der auf Diät gesetzt ist, soll pro Tag 4800 kJ zu sich nehmen, ein Fünftel davon in Kohlehydraten.

Wie viele sind dies täglich?

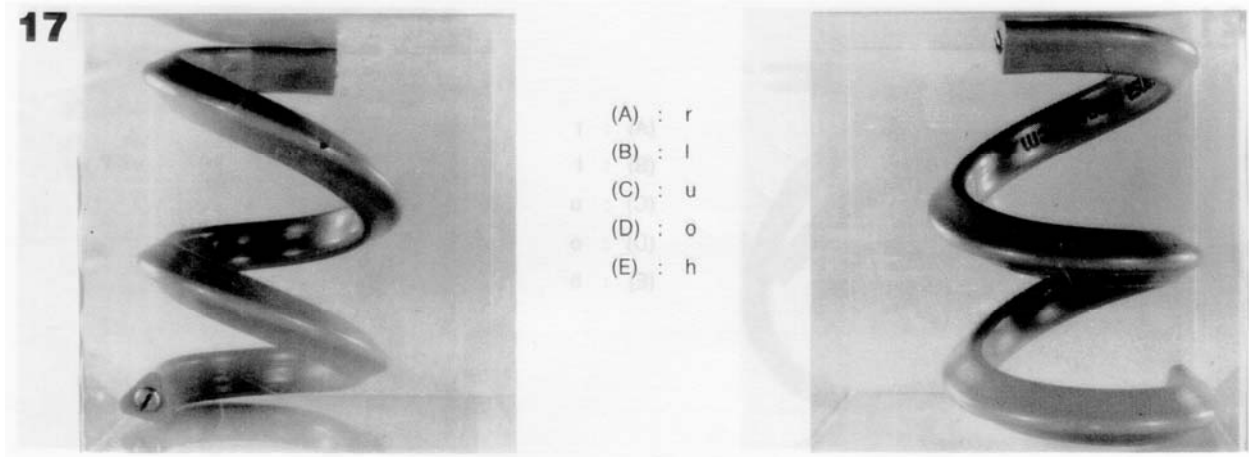
- (A) 60 BE
- (B) 25 BE
- (C) 6 BE
- (D) 5 BE
- (E) 0,5 BE

Bei solchen Fragen werden die Kenntnisse der Mittelstufen-Mathematik, nicht jedoch Lerninhalte vorausgesetzt. Der Patient soll ein Fünftel von 4800 kJ in Kohlehydraten zu sich nehmen, das sind also 960 kJ. Dividiert man diese Zahl durch 16, so erhält man die Anzahl g Kohlehydrate, nämlich 60 g, die es braucht, damit 960 kJ an Energie frei werden. Umgerechnet in Broteinheiten müssen die 60 g Kohlehydrate noch einmal durch 12 dividiert werden und das gibt 5 BE. Somit ist bei dieser Frage die Antwort (D) richtig.

7.2 Untertest: Schlauchfiguren

Die folgenden Aufgaben prüfen das räumliche Vorstellungsvermögen – eine Funktion, die beispielsweise für das Verständnis von Röntgenbildern wichtig ist. Während des Studiums werden zahlreiche eigentlich dreidimensional zu betrachtende Strukturen und Vorgänge in zweidimensionalen Abbildungen vermittelt.

Jede Aufgabe besteht aus zwei Abbildungen eines durchsichtigen Würfels, in dem sich ein, zwei oder drei Kabel befinden. Die erste Abbildung (links) zeigt stets die Vorderansicht des Würfels; auf dem rechten Bild daneben, in welchem derselbe Würfel noch einmal abgebildet ist, soll die Testteilnehmerin oder der Testteilnehmer herausfinden, ob die Abbildung die Ansicht von rechts (r), links (l), unten (u), oben (o) oder von hinten (h) zeigt.



Hier sehen Sie den Würfel von vorne!

Hier sehen Sie den Würfel von ? (hinten!)

7.3 Untertest: Textverständnis

Mit Hilfe dieses Untertests wird die Fähigkeit geprüft, umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten. Die Texte sind inhaltlich und grammatikalisch anspruchsvoll – sie können unter Nutzung von Notizen und Unterstreichungen erarbeitet werden. Die Abfrage erfolgt wiederum über die Auswahl einer richtigen oder falschen Aussage aus fünf vorgegebenen Aussagen. Diese Texte waren vor allem beim Übersetzen anspruchsvoll – zur Schwierigkeit gehören nicht nur die Inhalte, sondern auch die Satzstruktur.

Ein Beispiel:

Zu den Aufgaben der Schilddrüse gehören Bildung, Speicherung und Freisetzung der jodhaltigen Hormone Trijodthyronin (T_3) und Thyroxin (T_4). In der Schilddrüse befinden sich zahlreiche Hohlräume, Follikel genannt, deren Wände von einer Schicht so genannter Epithelzellen gebildet werden. Diese Follikel sind mit einer Substanz gefüllt, in der die Hormone T_3 und T_4 als inaktive Speicherformen enthalten sind. Beim Menschen ist in den Follikeln so viel T_3 und T_4 gespeichert, dass der Organismus damit für etwa 10 Monate versorgt werden kann.

Das für die Hormonbildung erforderliche Jod entstammt der Nahrung und wird von den Epithelzellen als Jodid aus dem Blut aufgenommen. Die Jodidaufnahme erfolgt an der äußeren Zellmembran der Epithelzellen durch eine sogenannte Jodpumpe. Diese wird durch ein Hormon aus der Hirnanhangsdrüse, das TSH, stimuliert und kann pharmakologisch durch die Gabe von Perchlorat gehemmt werden. Ferner gibt es erbliche Schilddrüsenerkrankungen, bei deren Vorliegen die Jodpumpe nicht funktioniert.

Bei Gesunden wird das in die Epithelzellen aufgenommene Jodid im nächsten Schritt unter dem Einfluss eines Enzyms in freies Jod umgewandelt und in die Follikel abgegeben. Die Aktivität dieses Enzyms kann ebenfalls pharmakologisch gehemmt werden.

Die letzten Schritte der Hormonbildung finden in den Follikeln, also außerhalb der einzelnen Epithelzellen, statt. In dort vorhandene so genannte Tyrosin-Reste (des Thyreoglobulins) wird zunächst ein Jodatom eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schließlich T_4 , während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T_3 hervorgeht. T_3 und T_4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt.

Diese Freisetzung von T_3 und T_4 ins Blut (Sekretion) wird über die Hirnanhangsdrüse und den Hypothalamus, einen Teil des Zwischenhirns, gesteuert: Das erwähnte Hormon TSH stimuliert ausser der Bildung auch die Sekretion von T_3 und T_4 ; es ist hinsichtlich seiner eigenen Sekretionsrate jedoch abhängig von der Stimulation durch das hypothalamische Hormon TRH. Die TRH-Sekretion wiederum wird z.B. durch Kälte stimuliert, während Wärme hemmend wirken kann. Neben diesen übergeordneten Steuerungsmechanismen existiert noch ein sogenannter Rückkoppelungsmechanismus: Eine hohe Konzentration von T_3 und T_4 im Blut hemmt die TSH- und die TRH-Sekretion, eine niedrige

Konzentration stimuliert sie. Bei den an der Steuerung der Schilddrüsenhormon-Sekretion beteiligten Arealen von Hirnanhangsdrüse und Hypothalamus können krankheitsbedingte Störungen auftreten, die zu einer Über- oder Unterfunktion der Schilddrüse führen.

Eine der Hauptwirkungen von T_3 und T_4 ist die Beeinflussung des Energieumsatzes durch eine Steigerung des Sauerstoffverbrauchs in stoffwechselaktiven Organen. Entsprechend senkt eine zu niedrige Konzentration der beiden Hormone im Blut (Hypothyreose) den Energieumsatz bzw. die Stoffwechselaktivität unter den normalen Wert, während bei einer zu hohen Konzentration (Hyperthyreose) die Stoffwechselaktivität gesteigert wird. Die Hormone T_3 und T_4 können ebenso wie TSH und TRH für diagnostische und therapeutische Zwecke synthetisch hergestellt werden.

Auf einen solchen Text folgen Fragen, die sich ausschließlich auf im Text vorhandene Inhalte beziehen; eine Frage mit niedrigem Schwierigkeitsgrad ist zum Beispiel so formuliert:

Welcher der folgenden Vorgänge gehört nicht zu den im Text beschriebenen Schritten, die zur Bildung von T_3 führen?

- (A) Transport von Jod aus den Epithelzellen in die Follikel
- (B) Umwandlung von Jod in Jodid in den Follikeln
- (C) Transport von Jodid aus dem Blut in die Epithelzellen
- (D) Verknüpfung von MIT- und DIT-Resten in den Follikeln
- (E) Verknüpfung von Jod und Tyrosin-Resten in den Follikeln

Für die Beantwortung dieser Frage ist das Verständnis der im obigen Text unterstrichenen Stellen wichtig (im Original sind selbstverständlich keine Hervorhebungen). Der Text sagt nichts über eine Umwandlung von Jod in Jodid in den Follikeln aus, und auch der umgekehrte Prozess, die Umwandlung von Jodid in Jod, findet nicht in den Follikeln statt, sondern in den Epithelzellen. Somit gehört der Vorgang (B) nicht zu den vom Text beschriebenen Schritten, die zur Bildung von T_3 führen.

7.4 Untertest: Planen und Organisieren

Der Untertest „Planen und Organisieren“ ist erstmals Bestandteil des EMS und prüft Fähigkeiten, die für eine effiziente Selbstorganisation im Studium wichtig sind. Er umfasst verschiedene von einander unabhängige Szenarien, auf die sich jeweils mehrere Aufgaben beziehen. Auch bei diesem Untertest wird kein Wissen vorausgesetzt. Es werden Aufgabenstellungen in Form so genannter Szenarien vorgegeben. Zu diesen Szenarien werden dann genau wie bei anderen Untertests einige Fragen gestellt. Bitte beachten Sie, dass im späteren „richtigen“ EMS sehr vielfältige Szenarien zur Anwendung kommen, die auch ganz anderen Schemen folgen können. Das nachfolgende Szenario soll das Prinzip verdeutlichen.

Szenario: Semesterarbeit

Sie haben die Aufgabe, im kommenden Wintersemester eine Semesterarbeit zu einem vorgegebenen Thema zu schreiben. Nach gründlichem Studium der einschlägigen Literatur sollen Sie in Ihrer Arbeit zunächst einen Überblick über die wichtigsten vorliegenden Ergebnisse zum Thema geben und dann einen eigenen Lösungsansatz entwickeln.

Die folgenden Termine liegen bereits fest:

- 11.10., 17.00 Uhr: Sie erhalten das Thema Ihrer Semesterarbeit, Literaturhinweise und weitere Instruktionen.
- 17.10.: Beginn der Vorlesungszeit
- 19.12. bis 1.1.: Weihnachtsferien (vorlesungsfreie Zeit)
- 11.2.: Ende der Vorlesungszeit
- 7.3., 8.00 Uhr: Abgabe der Semesterarbeit

- Während der Vorlesungswochen: Jede Woche von Montag bis Donnerstag Lehrveranstaltungen; an diesen Tagen können Sie sich nicht der Semesterarbeit widmen.
- In der ersten Hälfte der Weihnachtsferien (16. bis 24.12.): Urlaubsreise
- Zwei volle Wochen im unmittelbaren Anschluss an die Vorlesungszeit: Prüfungsvorbereitung (1 Woche) und Teilnahme an mündlichen Prüfungen (1 Woche)

Die folgenden Arbeitsschritte bzw. -phasen müssen Sie einplanen (sie sind hier nicht in chronologischer Abfolge aufgeführt):

- Zwei Treffen mit dem Betreuer Ihrer Arbeit: Eines unmittelbar vor dem Schreiben der Semesterarbeit, eines vor der abschließenden Überarbeitung. Dauer: Je ein voller Arbeitstag. (Der Betreuer ist nur dienstags und freitags verfügbar, und zwar von der dritten bis zur neunten Vorlesungswoche sowie nach dem Ende der Vorlesungszeit.)
- Literatursuche und -beschaffung (Dauer: 5 Tage)
- Erstellung des Konzepts der Semesterarbeit; die Literatur muss hier bereits verarbeitet sein (Dauer: 3 Tage)
- Korrekturlesen durch einen Freund (Überprüfung auf Tippfehler, unklare Formulierungen etc.) (Dauer: 3 Tage)
- Lesen und Zusammenfassen der Literatur (Dauer: 10 Tage)
- Schreiben der Semesterarbeit (Dauer: 18 Tage)
- Abschließendes Überarbeiten der Semesterarbeit (Dauer: 4 Tage)
- Kopieren der Arbeit (unmittelbar vor der Abgabe) (Dauer: 1 Tag)
- Ausführen der Korrekturvorschläge Ihres Freundes, bevor Sie die Arbeit mit dem Betreuer durchsprechen (Dauer: 1 Tag)
- Reserve für Unvorhergesehenes (Dauer: 1 Tag)

Eine Arbeitsphase muss stets beendet sein, bevor die nächste beginnt (Ausnahme: Korrekturlesen).

Ferner müssen Sie einplanen:

- An den Wochenenden arbeiten Sie – wie auch Ihr Freund – nur samstags (im Urlaub natürlich gar nicht).
- Am zweiten Weihnachtstag arbeiten Sie nicht.

Welche der folgenden Aussagen über die Semesterarbeit trifft bzw. treffen zu?

- I. Noch vor den Weihnachtsferien können Sie mit dem Schreiben der Semesterarbeit beginnen.
- II. Insgesamt werden Sie sich in 18 Wochen jeweils mindestens einen Tag lang mit der Semesterarbeit beschäftigen.

- (A) Nur Aussage I trifft zu.
(B) Nur Aussage II trifft zu.
(C) Beide Aussagen treffen zu.
(D) Keine der beiden Aussagen trifft zu.

Welche der folgenden Aussagen über die Einbeziehung Ihres Freundes trifft bzw. treffen zu?

- I. Der günstigste Zeitraum, in dem Ihr Freund Ihre Arbeit Korrektur lesen kann, ist die erste Hälfte der letzten Vorlesungswoche.
 - II. Angenommen, Ihr Freund teilt Ihnen mit, er habe leider nur in der Zeit vom 25. bis zum 27. Februar Gelegenheit zum Korrekturlesen; dann nützt Ihnen sein Angebot für den ursprünglich geplanten Zweck dieses Korrekturvorgangs nicht mehr.
- (A) Nur Aussage I trifft zu.
 (B) Nur Aussage II trifft zu.
 (C) Beide Aussagen treffen zu.
 (D) Keine der beiden Aussagen trifft zu.

7.5 Untertest: Konzentriertes und Sorgfältiges Arbeiten

Nachfolgend eine Beispielinstruktion aus der Test Info:

Mit diesem Test soll Ihre Fähigkeit, rasch, sorgfältig und konzentriert zu arbeiten erfasst werden. Sie sehen nachfolgend ein Blatt mit 40 Zeilen, die aus je 40 Buchstaben u und m gebildet werden.

**Ihre Aufgabe ist es, zeilenweise jedes u zu markieren,
 VOR dem in der Zeile unmittelbar ein m steht:**

m ✗

Sie dürfen kein u markieren, vor dem kein m steht. Sie dürfen natürlich auch kein m markieren. Beides wären Fehler. Nachfolgend sehen Sie ein richtig bearbeitetes Beispiel:

m ✗ m ✗ m ✗ u u m m ✗ m ✗ u u u m ✗ m m

Es werden Zeichen vorgegeben und bestimmte Zeichen sind zu markieren. Dies kann ein Merkmal eines einzelnen Zeichens sein (ein b mit 2 Strichen) oder eine Zeichenfolge (wenn ein p auf ein q folgt). Bei diesem Test hatte sich gezeigt, dass exzessives Üben zu verbesserten Leistungen führt. Da ein „Fleißtest“ nicht intendiert ist, werden seit 2004 die Zeichen und die Regel vor dem Test nicht mehr bekannt gegeben. Übungseffekte haben dadurch nachweislich einen geringen Einfluss.

7.6 Untertest: Medizinisch-naturwissenschaftliches Grundverständnis

Hier wird das Verständnis für Fragen der Medizin und der Naturwissenschaften geprüft. Der Text könnte so in einem Lehrbuch stehen. Wichtig für das Verständnis dieser Textpassage ist, ob daraus bestimmte logische Schlüsse gezogen werden können. Alle Fakten, die für die Beantwortung der Aufgabe notwendig sind, stehen im Text – spezielles medizinisches Vorwissen ist nicht erforderlich. Dieses wichtige Prinzip findet sich bei allen Untertests und ist verantwortlich für die geringe Trainierbarkeit der Aufgabenlösung.

Im Kindesalter kann das Zentrum für Sprache, Spracherwerb und Sprachverständnis noch in der linken oder in der rechten Hälfte (Hemisphäre) des Gehirns in einem umschriebenen Hirnrindengebiet

(sog. Sprachregion) angelegt werden. Spätestens im zwölften Lebensjahr sind die sprachlichen Fähigkeiten jedoch fest in einer der beiden Hemisphären verankert, und zwar bei den Rechtshändern in der Regel links, bei den Linkshändern in der Mehrzahl ebenfalls links, zum Teil aber auch rechts; die korrespondierende Region der Gegenseite hat zu diesem Zeitpunkt bereits andere Funktionen fest übernommen. Welche der nachfolgenden Aussagen lässt bzw. lassen sich aus diesen Informationen ableiten?

Bei irreversiblen Hirnrindenverletzungen im Bereich der so genannten Sprachregion der linken Hemisphäre ...

- I. kommt es bei erwachsenen Linkshändern in der Regel zu keinen wesentlichen Sprachstörungen.
 - II. kommt es bei einem Vorschulkind in der Regel zu einer bleibenden Unfähigkeit, die Muttersprache wieder zu erlernen.
 - III. ist bei zwanzigjährigen Rechtshändern die Fähigkeit, eine Sprache zu erlernen, in der Regel verloren gegangen.
- (A) Nur Ausfall I ist zu erwarten.
 - (B) Nur Ausfall II ist zu erwarten.
 - (C) Nur Ausfall III ist zu erwarten.
 - (D) Nur die Ausfälle I und III sind zu erwarten.
 - (E) Nur die Ausfälle II und III sind zu erwarten.

Bei diesem Aufgabentyp folgen nach der Schilderung des Sachverhalts in der Regel drei oder fünf Aussagen in Form von Behauptungen. Die Testperson muss sich dabei entscheiden, ob sich die Aussagen aus den im Aufgabentext enthaltenen Informationen ableiten lassen. Dazu sind keine speziellen Sachkenntnisse erforderlich. Die korrekte Beurteilung der einzelnen Aussagen setzt das Verstehen des Sachverhalts voraus sowie die Fähigkeit, Schlussfolgerungen aus den im Text enthaltenen Informationen zu ziehen. Konkret lässt sich die Aufgabe, unter Berücksichtigung des unterstrichenen Textes, folgendermaßen lösen:

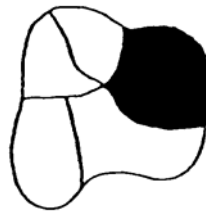
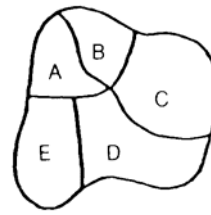
- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind immer noch die Möglichkeit, die Muttersprache wieder zu erlernen. Die Sprachregion wird dann in der rechten Hälfte der Hemisphäre angelegt. Somit ist Aussage II ebenfalls falsch.
- III. Da spätestens im zwölften Lebensjahr die Sprachregion bei Rechtshändern in der Regel fest in der linken Hälfte des Gehirns liegt, ist bei zwanzigjährigen Rechtshändern zu erwarten, dass sie die Fähigkeit, eine Sprache zu erlernen, verloren haben. Die Aussage III ist darum richtig.

Nach dieser Analyse des Textes ist es offensichtlich, dass die Antwort (C) richtig ist.

7.7 Untertest: Figuren lernen

Für beide nachfolgenden Gedächtnistests wird nach der Mittagspause das Material zum Einprägen ausgeteilt. Vor der Abfrage des Gelernten wird der Untertest „Textverständnis“ bearbeitet, damit liegt die Zeit des Behaltens der gelernten Inhalte über einer Stunde. Gedächtnisleistungen sind wichtige Voraussetzungen für Studienerfolg.

Der Untertest „Figuren lernen“ prüft, wie gut man sich Einzelheiten von Gegenständen einprägen und merken kann.

Gezeigte Figur
zum EinprägenGezeigte Figur
beim Abfragen

Die Testperson hat vier Minuten Zeit, um sich 20 solcher Figuren einschließlich der Lage der schwarzen Flächen einzuprägen. Nach ca. einer Stunde muss sie angeben können, welcher Teil der Abbildung geschwärzt war, und dies direkt auf dem Antwortbogen eintragen. Die Lösung ist natürlich C.

7.8 Untertest: Fakten lernen

Analog dem Prinzip beim „Figuren lernen“ sollen hier Fakten eingeprägt und behalten werden, die ebenfalls nach der gleichen Zwischenzeit abgefragt werden. Dabei werden 15 Patienten vorgestellt, von denen jeweils der Name, die Altersgruppe, Beruf und Geschlecht, ein weiteres Beschreibungsmerkmal (z.B. Familienstand) sowie die Diagnose erfahren wird. Ein Beispiel für eine derartige Fallbeschreibung ist:

Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

Eine Frage zum obigen Beispiel könnte z.B. lauten:

Der Patient mit dem Schädelbasisbruch ist von Beruf ...

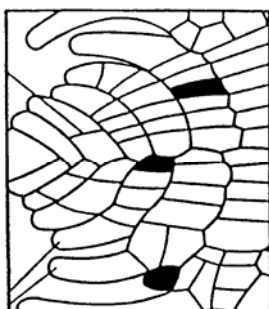
- (A) Installateur
- (B) Lehrer
- (C) Dachdecker
- (D) Handelsvertreter
- (E) Physiker

7.9 Untertest: Muster zuordnen

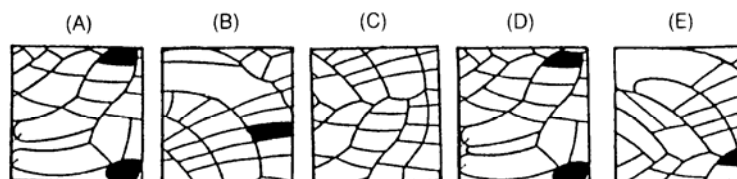
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wieder zu erkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner beliebigen Stelle deckungsgleich und vollständig auf das Muster gelegt werden kann.

Ein Beispiel dazu:

Muster



Musterausschnitte



In den meisten Aufgaben dieser Art heben sich die vier nicht deckungsgleichen Musterausschnitte dadurch vom Muster ab, dass Details entweder hinzugefügt oder weggelassen sind. Zugleich stellt dieser Untertest Anforderungen an die Schnelligkeit der Bearbeitung.

In durchschnittlich 55 Sekunden je Aufgabe muss die Testperson die richtige Lösung herausgefunden haben, dass beispielsweise in der obigen Aufgabe nur der Musterausschnitt (A) deckungsgleich mit einem Teil des Musters ist, und zwar in dessen unterem Bereich, etwa in der Mitte.

7.10 Untertest: Diagramme und Tabellen

Mit dieser Aufgabengruppe wird die Fähigkeit geprüft, Diagramme und Tabellen richtig zu analysieren und zu interpretieren. In dieser Form werden während des Studiums zahlreiche Zusammenhänge vermittelt. Eine Aufgabe dazu:

Die folgende Tabelle beschreibt die Zusammensetzung und den Energiegehalt von vier verschiedenen Milcharten. Unter Energiegehalt der Milch verstehen wir dabei die Energiemenge, gemessen in Kilojoule (kJ), welche 100 Gramm (g) Milch dem Organismus ihres Konsumenten liefern können.

Milchart	Eiweiss	Fett	Milchzucker	Salze	Energiegehalt
menschliche Muttermilch	1,2 g	4,0 g	7,0 g	0,25 g	294 kJ
Vollmilch	3,5 g	3,5 g	4,5 g	0,75 g	273 kJ
Magermilch	3,3 g	0,5 g	4,5 g	0,75 g	160 kJ
Buttermilch	3,0 g	0,5 g	3,0 g	0,55 g	110 kJ

Welche Aussage lässt sich aus den gegebenen Informationen nicht ableiten?

- (A) Menschliche Muttermilch enthält mehr als doppelt soviel Milchzucker wie Buttermilch.
- (B) Vollmilch enthält im Vergleich zur menschlichen Muttermilch etwa die dreifache Menge an Salzen und Eiweiss.
- (C) Zur Aufnahme der gleichen Energiemenge muss ein Säugling fast dreimal soviel Buttermilch wie Muttermilch trinken.
- (D) Der Unterschied zwischen Magermilch und Vollmilch ist bei der Mehrzahl der aufgeführten Merkmale geringer als der Unterschied zwischen Magermilch und Buttermilch.
- (E) Der Eiweissgehalt der Milch ist für den Energiegehalt von entscheidender Bedeutung.

Wie bei den Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“ und „Textverständnis“ sind auch hier zur Lösung dieser Aufgabe keine speziellen naturwissenschaftlichen, medizinischen oder statistischen Kenntnisse erforderlich. Die richtige Lösung lässt sich allein aus der jeweils graphisch oder tabellarisch dargebotenen Information und dem zugehörigen Aufgabentext ableiten. Aus den angegebenen Werten ist kein systematischer Zusammenhang zwischen Eiweiß- und Energiegehalt ableitbar, so dass die Aussage (E) nicht abgeleitet werden kann.

Diese Beispielaufgaben aus den zehn Untertests zeigen, dass es hier um Problemstellungen geht, die auch aus einem Lehrbuch des Grundstudiums Medizin stammen könnten. In den Aufgabenstellungen sind alle Informationen enthalten, die man zum Lösen benötigt. Das Problem ist zunächst zu erkennen, die Information genau zu analysieren und eine Lösung zu finden.

8 Literatur

- Deidesheimer Kreis (1997). Hochschulzulassung und Studieneignungstests: studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus Clausus und andere Studiengänge. Göttingen, Zürich: Vandenhoeck und Ruprecht.
- Dlugosch, S. (1995). Prognose von Studienerfolg. Aachen: Shaker Verlag.
- Ebach, J., Trost, G. (1997). Admission to Medical Schools in Europe. Lengerich: Pabst.
- Frischenschlager O., Mitterauer L., Haidinger G (2005): Leistungsfaktoren als potenzielle Auswahlkriterien im Medizinstudium. E-ZfHD und Zeitschrift für Hochschuldidaktik, Heft 6, Dezember 2005.
- Hängsen K-D, Spicher B. (2002). Numerus Clausus: Finden wir mit dem «Eignungstest für das Medizinstudium» die Geeigneten? Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri - 2002; 83(31):1653-1660.
<http://www.saez.ch/pdf/2002/2002-31/2002-31-842.PDF>
- Hängsen K-D, Spicher B. (2002). Numerus Clausus: Numerus Clausus: le « test d'aptitudes pour les études de médecine » (AMS) permet-il de trouver les personnes les plus aptes? Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri - 2002; 83 (47) 2562 – 2569. <http://www.saez.ch/pdf/2002/2002-47/2002-47-1144.PDF>
- Hängsen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus Clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 1: Erfahrungen mit dem EMS als Zulassungskriterium. Schweizerische Ärztezeitung Heft 12 S. 666 – 672.
- Hängsen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus Clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 2: EMS und Chancengleichheit. Schweizerische Ärztezeitung Heft 13 S. 723-730.
- Hängsen, K.-D., Hofer, R. & Ruefli, D. (1995a). Un test d'aptitudes aux études de médecine est-il faisable en Suisse? Bulletin des médecins suisses, 7, S. 267 - 274.
- Hängsen, K.-D., Hofer, R. & Ruefli, D. (1995b). Der Eignungstest für das Medizinstudium in der Schweiz. Schweizerische Ärztezeitung, 37, S. 1476 - 1496.
- Haidinger G., Frischenschlager O., Mitterauer L. (2006): Reliability of predictors of study success in medicine. Wiener medizinische Wochenschrift, (in press), Zusammenfassung siehe <http://www.springerlink.com/content/bm5854nq41533t14/>
- Oswald, U. (1999). Der Eignungstest 1998 für das Medizinstudium. Schweizerische Ärztezeitung 80, S. 1313 – 1317.

- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff H.-G., Rost, J., Schiefele U. (Hrsg.) PISA-Konsortium Deutschland. PISA 2003. Ergebnisse des zweiten internationalen Vergleichs. Zusammenfassung. Forschungsbericht.
- Trost, G. (Hrsg.) (1994). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1995). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (19. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1996). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (20. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1997). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (21. Arbeitsbericht). Bonn: ITB.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H.-U. (1998). Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse. Bonn: ITB.

8.1 Originaltest zur Information und Vorbereitung

- Institut für Test- und Begabungsforschung (Hrsg.). (1995). Test für medizinische Studiengänge (Aktualisierte Originalversion 2). Herausgegeben im Auftrag der Kultusminister der Länder der BRD. 4. Auflage. Göttingen: Hogrefe.
- Centre pour le développement de tests et le diagnostic, Université de Fribourg (Suisse) en collaboration avec l'Institut für Test- und Begabungsforschung, Bonn, Allemagne (Editeur). (1996). Le test d'aptitudes pour les études de médecine. Adaptation française de la version originale dans son intégralité. Göttingen: Hogrefe.
- Centre pour le développement de tests et le diagnostic, Università di Friburgo (Svizzera) in collaborazione con l'Institut für Test- und Begabungsforschung, Bonn, Germania (Editore). (1996). Il test attitudinale per lo studio della medicina. Adattamento italiano di una versione originale completa. Göttingen: Hogrefe.

8.2 Information im Internet

Die genannten Beiträge des ZTD und weitere Informationen können Sie abrufen über:

www.unifr.ch/ztd/ems/

Informationen zum Zulassungsverfahren in Österreich können Sie abrufen über:

www.eignungstest-medizin.at/