

EMS

Eignungstest für das Medizinstudium in Österreich

Medizinische Universität Innsbruck
Medizinische Universität Wien
2008

Berichterstatte

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 bis 2008 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	5
2	BESCHREIBUNG DES EIGNUNGSTESTS	6
2.1	Aufbau des Tests	6
2.2	Berechnung der Werte	8
2.3	Mittlerer Rangplatz der Untertests	8
3	QUALITÄTSSICHERUNG	10
3.1	Prognosekraft als „ultimatives“ Kriterium	10
3.2	Genderunterschied in Prüfungen korrekt vorhergesagt.....	11
3.3	Umgang mit Genderunterschieden	11
3.4	Was ist Zuverlässigkeit, was muss zuverlässig sein?	13
3.5	Sind Multiple-Choice-Aufgaben geeignet?	13
3.6	Kann man den Test verkürzen?	14
3.7	Können Erfolgreiche häufiger angetreten sein?	14
3.8	„Verrechnungsfairness“ des EMS	14
3.9	Misst der Test überall das Gleiche?	16
3.10	Ist der EMS auf die Studienanforderungen bezogen?	16
3.11	EMS und Änderungen der Studienorganisation?.....	17
3.12	Soziale Kompetenzen berücksichtigen?	17
3.13	Eignung und Neigung	19
3.14	Perspektive.....	20
4	TESTANWENDUNG IN ÖSTERREICH 2008	21
4.1	Statistische Angaben zur Teilnahme	21
4.2	Verteilungsprüfung der einzelnen Punktwerte	24
4.3	Vergleich der Geschlechter	27
4.4	Vergleichbarkeit der Testorte	35
4.5	Vergleich nach Maturitätsländerquote	36
4.6	Vergleiche für Altersgruppen nach Geburtsjahr	36
4.7	Vergleich nach Disziplin	38
5	ERGEBNISSE ZUR TESTGÜTE	39
5.1	Zuverlässigkeit.....	39
5.2	Faktorielle Validität.....	39
5.3	Item-Trennschärfen	44
5.4	Konzentriertes und sorgfältiges Arbeiten	49
6	DIF-ANALYSEN FÜR UNTERGRUPPEN	52
6.1	DIF nach Geschlecht (AT, Humanmedizin)	53
6.1.1	Quantitative und formale Probleme	53
6.1.2	Schlauchfiguren	54
6.1.3	Textverständnis.....	55
6.1.4	Planen & Organisieren.....	56
6.1.5	Medizinisch-naturwissenschaftliches Grundverständnis	57
6.1.6	Figuren lernen.....	58

6.1.7	Fakten lernen.....	59
6.1.8	Muster zuordnen.....	60
6.1.9	Diagramme und Tabellen	61
7	BEISPIELAUFGABEN FÜR DIE UNTERTESTS	62
7.1	Untertest: Quantitative und formale Probleme.....	62
7.2	Untertest: Schlauchfiguren	62
7.3	Untertest: Textverständnis	63
7.4	Untertest: Planen und Organisieren	64
7.5	Untertest: Konzentriertes und sorgfältiges Arbeiten.....	66
7.6	Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	66
7.7	Untertest: Figuren lernen	67
7.8	Untertest: Fakten lernen	68
7.9	Untertest: Muster zuordnen	68
7.10	Untertest: Diagramme und Tabellen.....	69
8	LITERATUR.....	70
8.1	Originaltest zur Information und Vorbereitung	71
8.2	Information im Internet	71

1 Einleitung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2008 für das Medizinstudium und die Zulassung zum Studium der Human- und Zahnmedizin ab Wintersemester 2008/2009 an den Medizinischen Universitäten Innsbruck und Wien. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen in der Schweiz herangezogen. **Die Ergebnisse sind nicht repräsentativ und somit nicht geeignet 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, wenden die Medizinischen Universitäten in Wien und Innsbruck auf Basis der vom Nationalrat erlassenen Novelle zum Universitätsgesetz 2002 gemeinsam eine kapazitätsorientierte Studienplatzvergabe für ALLE StudienbewerberInnen an. Folgende Platzzahlen waren für die Studien „Humanmedizin“ und „Zahnmedizin“ verfügbar:

	Humanmedizin	Zahnmedizin	Gesamt
Med. Univ. Wien (MUW)	660	80	740
Med. Univ. Innsbruck (MUI)	360	40	400
Gesamt	1020	120	1140

Die Vergabe der Plätze erfolgte mittels eines erprobten und wissenschaftlich abgesicherten Eignungstests (EMS), der auf dem deutschen Test für Medizinische Studiengänge (TMS) basiert, in der Schweiz weiterentwickelt und seit 1998 angewendet wird. Der Eignungstest liefert einen Testwert, welcher nachweislich hoch mit der Studieneignung korreliert. Dieser wird für die Zulassung verwendet. Bei gleichem Testwert entscheidet der mittlere Rangplatz der Testleistungen in den einzelnen Untertests.

Rechtliche Grundlage für das Aufnahmeverfahren sind der §124b UG2002 und die in den Mitteilungsblättern der Medizinischen Universität Wien und der Medizinischen Universität Innsbruck veröffentlichten Verordnungen in der jeweils gültigen Fassung.

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 Beschreibung des Eignungstests

2.1 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 als faires und wissenschaftlich begründetes Zulassungsverfahren ist die Vorhersage des Studienerfolges. 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 2008 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 1: Struktur und Ablauf des EMS 2008

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 der Testentwicklung 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 1) 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.

2.2 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**). 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).

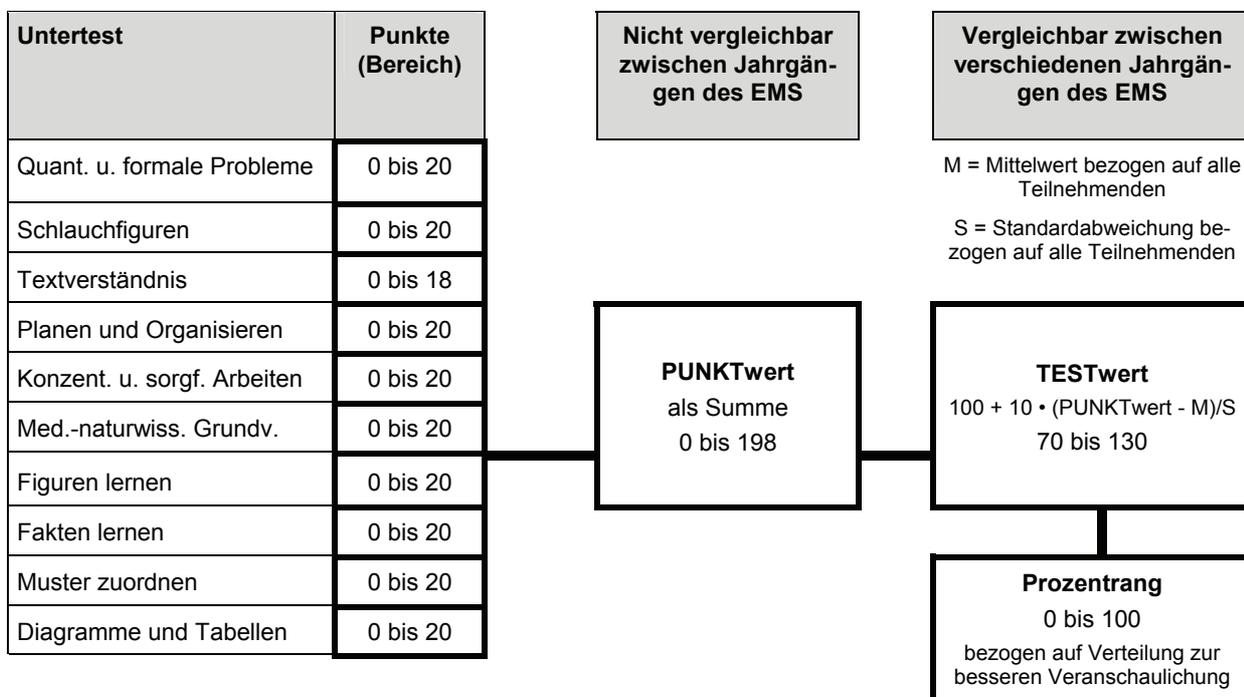


Tabelle 2: Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert (Maximalwerte 2008).

2.3 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 „Scheindif-

ferenzierung“ erfolgen, hinter der Wertestreuung 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 2008 zwischen 35 und 956. Niedrige Werte stehen für bessere Leistungen.
- Er korreliert mit dem Testwert mit -0.992 (Vorjahr -0.991), was praktisch für eine Äquivalenz spricht – siehe auch Abbildung 1.
- Er differenziert allerdings besser zwischen den Personen und lässt eine kapazitätsgenaue Auswahl zu.

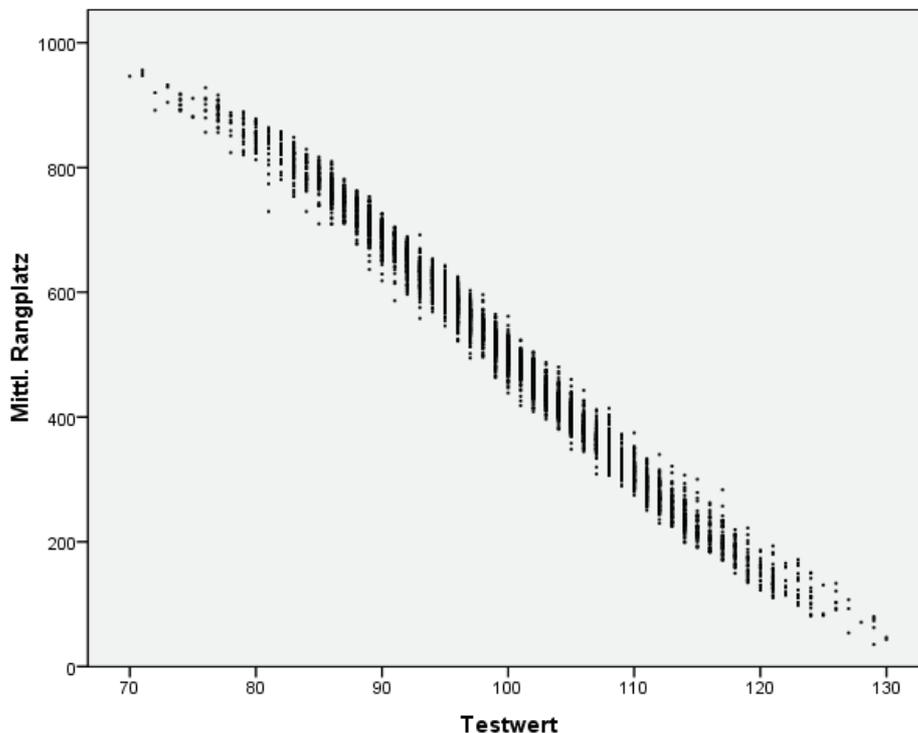


Abbildung 1: Beziehung zwischen Testwert und mittlerem Rangplatz für 2008.

3 Qualitätssicherung

Leistungsunterschiede zwischen Frauen und Männern in beiden Eignungstests für das Medizinstudium (Graz bzw. Wien/Innsbruck) sowie der Unterschied von Österreicherinnen und Österreichern gegenüber der Angehörigen der EU-Quote waren Gegenstand der politischen Diskussion. Auch die beteiligten Universitäten nahmen und nehmen diese Fragen ernst (siehe Berichte zum EMS-AT 2007 und 2008¹), um **testbedingte Benachteiligungen** in jedem Falle zu vermeiden. Bisher konnte eine testbedingte Benachteiligung, die auszugleichen wäre, nicht wissenschaftlich nachgewiesen werden.

3.1 Prognosekraft als „ultimatives“ Kriterium

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 z.B. losen. Der Test ist ein „wettbewerbsorientiertes Reihungsverfahren“, es sollen diejenigen bevorzugt einen Studienplatz erhalten, die in der vorgesehenen Zeit das Studium abschliessen und die beschränkten 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.

Weil 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 Mitarbeiter² 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 die Geschlechtergruppen unterschiedlich streng benotet werden (wofür es in Österreich Anzeichen gibt, vgl. Spiel 2007). 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 Varianzverringerung 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

Von Interesse ist auch, dass die Mittelwerte der **Nicht Erfolgreichen** in der SIP-1 (unter 67 Punkten) sich **im EMS** für Frauen (106.3) und Männer (106.7) **nicht signifikant** unterscheiden. Die „Erfolgsgrenzen“ sind somit vergleichbar, die „Hürde“ ist gleich. Es gibt keinen Hinweis auf systematische Unterschätzung des prognostizierten Studienerfolgs für Frauen durch den EMS.

¹<http://www.eignungstest-medizin.at>

² 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)

Auch in **Innsbruck** (Kraft 2008) schneiden die Studierenden, die durch den EMS ausgesucht worden sind, in der SIP-1 besser ab als die Studierenden, die ohne Selektion aufgenommen worden sind. Dieser Unterschied ist auf 3 Ebenen sichtbar: Mit EMS haben 75 % bestanden gegenüber 68 %; um zu bestehen haben sie im Mittel 1.78 Monate gebraucht gegenüber 3.5 Monaten; im Mittel erreichten sie (bei Bestehen) 114 richtige Antworten in der SIP-1 gegenüber 107. Zusätzlich ist auch der Prozentsatz der Studierenden, die nie zur SIP-1 antreten drastisch gesunken: Ohne EMS waren es 16.7 %; mit EMS nur mehr 4.4 %.¹

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%.

3.2 Genderunterschied in Prüfungen wird vorhergesagt.

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 unter den Frauen eine Gruppe geben, 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 politischem 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.

3.3 Umgang mit Genderunterschieden

In der Evaluationsstudie von Spiel u.M.³ werden „kontinuierliche Sozialisationsunterschiede“ zwischen Jungen und Mädchen in der Schule mit als Ursache genannt. Diese können bestehen

- in durch unterschiedliche Rollenbilder geprägtem Erziehungsverhalten, in welchem Mädchen beispielsweise mehr für Fleiss und Anpassung gelobt werden als Knaben und an Knaben häufig höhere Leistungsanforderungen gestellt werden. Dies konnte z.B. von Eder u. M. als eine grössere Notenstrenge gegenüber Knaben nachgewiesen werden (bei gleichen objektiven Leistungen erhalten Mädchen bessere Noten). Auch

¹ Persönliche Mitteilung von H.-G. Kraft (E-Mail vom 17.9. 2008).

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

³ Spiel, C. u.a. (2007). Evaluation der Eignungstests für das Medizinstudium. Evaluationsbericht im Auftrag des Bundesministeriums für Wissenschaft und Forschung.

Spiel u.M. fanden, dass Schulnoten die Testleistung bei Knaben gut vorhersagen, während dies bei Mädchen weniger deutlich ist. Daraus kann gefolgert werden, dass für die Benotung von Mädchen auch andere Kriterien eine Rolle spielen als die Leistung. Eine teilweise unrealistischere Leistungsrückmeldung für Mädchen suggeriert das Vorhandensein von mehr Wissen und Fähigkeiten auch auf mathematisch-naturwissenschaftlichem Gebiet, was die Studien- und Berufswahl beeinflussen kann.

- in einer speziell auf Mathematik und Naturwissenschaften bezogenen weniger förderlichen Bildungsumgebung für Mädchen, wobei vermittelt werden kann, dass Mathematik und Naturwissenschaften für Mädchen weniger berufsrelevant sind, was sich ebenfalls auf die Motivation zur Beschäftigung mit diesen Fächern auswirkt.
- in einem negativen Kreisprozess: Durch Vermittlung, dass Mathematik für Mädchen weniger berufsrelevant sei, wird weniger Vertrauen in die eigenen Fähigkeiten in Mathematik und Naturwissenschaften entwickelt. Dies führt zu geringerer Motivation und weniger Investment für diese Fächer sowie dann zu einem geringeren Können.

Die Sensibilisierung für diese Problematik durch die politische Diskussion im letzten Jahr und konkrete Massnahmen der Beratung und gezielten Vorbereitung auf den Test, die in Innsbruck und Wien stattfanden, haben den Unterschied um rund 25% reduzieren können. Damit ist das Problem aber nur teilweise gelöst und es bleibt wichtig, bei dieser Sensibilisierung auch die notwendige Nachhaltigkeit zu erreichen.

Langfristig müssen diese Probleme gelöst werden, indem die hier genannten Ursachen im Bildungssystem beseitigt werden. **Naturwissenschaftliche Orientierung muss auch für Mädchen eine (geforderte und geförderte) Selbstverständlichkeit werden, die Leistungsanforderungen an Jungen und Mädchen müssen gleich sein und es muss auch kommuniziert werden, dass das Medizinstudium vor allem mathematisch-naturwissenschaftliche Kenntnisse und Fähigkeiten erfordert.**

Sind Korrekturen des Zulassungsverfahrens möglich und begründet? Was sind die Fakten?

Es ist in den Evaluationsstudien nachgewiesen, dass für Männer und Frauen der gleiche Prognosezusammenhang EMS-Studienerfolg besteht, der EMS die Studieneignung von Frauen nicht unterschätzt. Der Leistungsunterschied beim EMS findet sich auch als Leistungsunterschied bei der Prüfung SIP-1, der Test sagt diesen korrekt voraus. Erst später holen Frauen auf und erreichen gleiche Leistungen. Dies ist ein Hinweis, dass unter den gleichen fördernden und fordernden Bedingungen des Studiums Frauen ebenso geeignet sind wie Männer und belegt die Bedeutung der Unterschiede bei der Bildungssozialisation.

Für die aktuellen Bewerbergenerationen muss deshalb überlegt werden, ob die genannten Unterschiede der Bildungssozialisation für Mädchen eine **objektive Benachteiligung** darstellen, die aus **politischen Gründen** ausgeglichen werden müsste, solange der Unterschied in Österreich beispielsweise signifikant grösser ist als derjenige in der Schweiz. Ein Teil des Unterschiedes wird weiter darauf zurückzuführen sein, dass die Bewerberquoten 60% Frauen und 40% Männer keine gleichen mittleren Leistungen erwarten lassen.

Will man eine insgesamt wenig attraktive Quotenregelung bei der Zulassung vermeiden, wäre eine gangbare Alternative, dass für Frauen und Männer die Standardisierung des Punktwertes zum Testwert nicht den Gesamt-Mittelwert und die Gesamt-Standardabweichung verwendet, sondern dies jeweils für beide Geschlechter gesondert berechnet wird. Bei dieser Lösung bleiben Verteilungsunterschiede erhalten, der mittlere Unterschied wird aber ausgeglichen. In der Abbildung 10 sind die Unterschiede dargestellt.

Das zu klärende rechtliche Problem dabei ist der sogenannte „**Paternostereffekt**“: Gleiche Testleistungen können dann zu Unterschieden bei der Zulassung führen. Da die Zahl der Zulassungen konstant bleibt, würde für alle Frauen, die dank des Bonus zugelassen werden, die gleiche Menge Männer mit besseren Testergebnissen nicht zugelassen. Die Interessenabwägung hat stattzufinden zwischen der Wertung der Benachteiligung als objektiv und der Tatsache, dass gleiche Testergebnisse auch eine gleiche Studienerfolgsprognose für Männer und Frauen bedeuten (Chancengleichheit). Hiergegen kann man einwenden, dass Frauen die Unterschiede spätestens nach dem ersten Jahr aufzuholen scheinen, wenn sie im Studium unter gleichen Bedingungen wie die Jungen gefordert werden.

3.4 Was ist Zuverlässigkeit, was muss zuverlässig sein?

Richtig ist, dass die verwendeten (!) Werte eines Tests zuverlässig (reliabel) sein müssen. Die für Eignungsdiagnostik massgebliche DIN 33430 bzw. gleichlautende ÖNORM D 4000 fordern je nach Verfahrensklasse 0.70 bis 0.85.

Beim EMS beträgt die Reliabilität des für die Zulassung verwendeten Testwertes 2008 jeweils **0.92 (für Schweiz und Österreich)**, berechnet mittels Testhalbierungsmethode auf Itemebene. Vergleichbare Werte wurden in allen Vorjahren erreicht. Der verwendete EMS-Testwert ist ohne Wenn und Aber ausreichend zuverlässig.

Die Reliabilitäten der einzelnen Aufgabengruppen würden nur Bedeutung erlangen, wenn diese Werte einzeln diagnostisch verwendet würden – etwa im Rahmen einer Bildungsberatung, um Stärken und Schwächen in Bereichen festzustellen. Dafür ist der EMS weder gedacht noch geeignet. Die etwas geringeren Reliabilitäten der Aufgabengruppen sind der „Preis“ für eine grössere Vielfalt, grössere Heterogenität innerhalb der einzelnen Untertests. Diese war den Konstrukteuren des TMS wichtig (mehr siehe unter: „*Verrechnungsfairness*“), um durch Vielfalt z.B. die Trainierbarkeit zu minimieren. Indem die Reliabilität des Testwertes „trotzdem“ sehr hoch ist und vor allem die Prognosegüte erwartet hoch ausfällt, wird dieses Vorgehen ebenfalls ohne Einschränkungen legitimiert¹.

3.5 Sind Multiple-Choice-Aufgaben geeignet?

Andere Antwortformate (z.B. mehrere richtige Lösungen pro Aufgabe) sind denkbar. Für die erprobten vorhandenen Aufgaben ist allerdings keine nachträgliche Änderung möglich, ohne eine erneute empirische Prüfung unter Ernstfallbedingungen notwendig zu machen, weil sich die Schwierigkeit drastisch ändert.

Es ist ein Charakteristikum des EMS, nur ausreichend vorerprobte Aufgaben zu verwenden (das könnte auch die festgestellten Unterschiede zum ähnlichen Untertest „Textverständnis“ in Graz erklären). Ein anderes Vorgehen wäre allerdings nicht problemlos: Es ist für die Rechtsfähigkeit des EMS bedeutsam, dass es pro Aufgabe tatsächlich eine richtige Lösung gibt, die sich ausreichend von allen anderen unterscheidet. Es dürfen z.B. keine sogenannten „Doppeldeutigkeiten“ auftreten, dass falsche Lösungen dennoch als vermeintlich richtig abgeleitet werden könnten. Bisher konnte in diesen Fällen immer der eindeutige empirische Beweis erbracht werden, dass die Leistungsbesten in einem Aufgabenbereich die richtige Lösung auch gehäuft wählen – die Trennschärfe der richtigen Lösung sich von der aller Falschlösungen deutlich genug unterscheidet. Da sich bei mehr richtigen Lösungen pro Aufgabe deren Einzelschwierigkeiten auch unterscheiden, würde dieser Unterschied graduell „aufgeweicht“ und die o.g. Beweisführung zumindest erschwert, wenn nicht unmöglich. Das wäre bei neuen Überlegungen (die möglich sind) zumindest zu bedenken.

Die Literatur ist nach unserer Kenntnis zwar voll von Vermutungen, aber bis heute **eindeutige** empirische Belege dafür schuldig geblieben, dass unterschiedlich viele Lösungsoptionen pro Aufgabe deren **Zuverlässigkeit** tatsächlich verbessert. Alle grossen Studierfähigkeits-tests arbeiten, was die Zulässigkeit nur einer richtigen oder besten Lösung betrifft, offenbar nicht ohne Grund nach dem gleichen Prinzip wie der EMS.

Der Einfluss der **Ratewahrscheinlichkeit** wird im Übrigen durch die mehrfach gegebene Instruktion ausgeglichen, alle Fragen zu beantworten – also am Ende immer bei jenen zu raten, die unbeantwortet sind. Damit sind die Bedingungen für alle Personen gleich und echte Leistung wird als Differenz zur Ratewahrscheinlichkeit aufgefasst.

¹ In der Psychologie gilt: Ein nicht reliabler (zuverlässiger) Test kann nicht valide (gültig – hier für die Erfolgsprognose) sein. Andererseits muss ein ausreichend valider Test automatisch auch reliabel sein. Zu beachten ist auch, dass es keine Reliabilität „an sich“ gibt, sondern verschiedene Aspekte und Schätzmethoden existieren.

Ein häufiger Irrtum dabei ist, dass mit der zufälligen Lösungswahrscheinlichkeit auf Ebene der einzelnen Aufgaben argumentiert wird. Bei 5 Antwortstufen beträgt sie pro Aufgabe 20 %. Es werden aber 20 Aufgaben verwendet. Schon bei 2 Aufgaben ist die Wahrscheinlichkeit von 2 richtigen Lösungen nur noch 4%, bei 3 Aufgaben entsprechend 0.8% usw. bis zu 20 Aufgaben 1 E⁻¹² Prozent.

3.6 Kann man den Test verkürzen?

Eine Verkürzung ist 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 „Figurenlernen“ wird die Prognosegüte des Studienerfolges für die erste Vorprüfung nicht wesentlich geringer. Diese beiden Aufgaben wurden aber aus inhaltlichen Gründen im Test belassen, weil räumliches Denken bzw. räumliche Vorstellungen ggf. in späteren Studienabschnitten stärker gefordert werden.

3.7 Können Erfolgreiche häufiger angetreten sein?

Es wurde vermutet, dass der bisherige Leistungsunterschied der Deutschen durch eine größere Häufigkeit der Testteilnahmen erklärbar ist. In der Schweiz (EMS 1998 bis heute) werden praktisch keine EU-Bürger zum Studium und folglich zum EMS zugelassen. In Deutschland hat es den TMS letztmals 1996 gegeben – man durfte aber nur einmal am TMS überhaupt teilnehmen. 2007 im Mai hat es erstmals wieder einen TMS in Baden-Württemberg gegeben (auch jetzt darf wieder nur einmal am Test teilgenommen werden). Im Maximalfall könnten es also 2 vorherige Teilnahmen sein (2006 in Österreich, 2007 in Deutschland) – wobei es 2007 auch aus Österreich Testwiederholungen gibt. Es wäre möglich, dass die entsprechende Frage bei der Evaluation missverstanden wurde und sich beispielsweise auf alle Tests bezieht.

Die Empfehlung, eine veröffentlichte Originalversion unter Echtzeitbedingungen zu bearbeiten, wird unterstützt. In den ersten Jahren des EMS waren Personen, die das zweite Mal am EMS teilnahmen, auch deutlich erfolgreicher als beim ersten Mal. Dieser Unterschied hat abgenommen, seit „Probelaufe“ mit der Originalversion empfohlen und unterstützt werden.

3.8 „Verrechnungsfairness“ des EMS

In einigen Diskussionen wurde die „Verrechnungsfairness“ des EMS diskutiert. So wurde verschiedentlich gefordert, dass Testaufgaben nur immer eine einzige zu erfassende Fähigkeit „homogen“ erfassen dürfen und keine komplexen „lebensnahen“ Anforderungen, die mehrere Fähigkeiten in Kombination erfordern. Diese Kritik nimmt Bezug auf die sogenannte probabilistische Testtheorie (PTT) bzw. Rasch-Skalierung, für welche solche Konzepte ausschließlich gültig sind. Es ist **keinesfalls entschieden**, dass Konzepte dieser Theorie hinsichtlich der Skalierung für alle Tests gelten müssen. **„Verrechnungsfairness“ nach Rasch bedeutet nicht zwingend, dass im „normalen“ Sprachsinne Fairness verletzt wäre.**

Der EMS wurde nach der **klassischen Testtheorie** konstruiert. Entsprechend sind deren Bewertungskonzepte und Gütekriterien anzuwenden und eigentlich ist jeder Streit müssig, weil ein valides prognoserelevantes Instrument wie der EMS auch alle anderen Gütekriterien erfüllen muss – denn sonst könnte es nicht valide sein.

Der „Streit“ ist allerdings alt und wurde/wird teilweise etwas „fundamentalistisch“ geführt. Obwohl es die probabilistische Testtheorie schon seit 1960 gibt, wurden und werden die wichtigsten und anwendungshäufigsten Leistungstests der Psychologie nach wie vor auf der Basis der klassischen Testtheorie entwickelt. Auch wichtige amerikanische Studierfähigkeits-tests, die seit mehr als einem halben Jahrhundert zum Teil weltweit und mit Teilnehmerzahlen in Millionenhöhe verwendet und kontinuierlich von ausgewiesenen Testexperten weiterentwickelt werden, fassen bis heute auf der klassischen Testtheorie. Beide Strategien haben Vor- und Nachteile, beide können zu seriösen Tests führen. Für nach PTT konstruierte Tests muss man eine „latente Eigenschaft“ (Dimension) annehmen und durch eine Skala messen. Die Skala muss vergleichsweise restriktiven messtheoretischen Voraussetzungen folgen, was praktisch auch heisst: Homogen bezüglich der Aufgaben sein, um „eindimensional“ *die interessierende Kompetenz* zu messen. Das erreicht man praktisch nur, indem alle Aufgaben auch relativ gleichartig konstruiert sind.

Die Frage, ob man solche Tests verwenden soll, wurde bei der Konstruktion des TMS im damaligen „Beirat für psychologische und allgemeine Fragen der Testentwicklung und des Testeinsatzes bei der Hochschulzulassung“ ebenfalls diskutiert. Man hat sich bewusst für einen anderen Weg entschieden, komplexere und für ein Studium anforderungsnähere Tests mit mehr inhaltlicher Validität zu verwenden. Der Beirat empfahl mit Nachdruck, der TMS solle nicht homogene, **sondern komplexe Aufgaben enthalten und damit der Realität eines Studiums, in dem ganz verschiedene kognitive Funktionen gleichzeitig gefordert sind, möglichst nahe kommen**. Auf diese Weise lasse sich dank der Simulation solcher komplexer Anforderungssituationen eine höhere inhaltliche Validität erzielen. Eine Studienanforderung wie das Lesen von Tabellen, das Verstehen von Texten muss nicht psychologisch „eindimensional“, sondern kann wie im realen Leben komplex sein. Auch gibt es verschiedene Arten von Diagrammen oder Texten, die innerhalb der Aufgabengruppen variiert werden, um das vorherige Üben einzelner Prototypen zu erschweren.

Damit wird auch klar, warum die **Trennschärfen** der Aufgabengruppen tendenziell geringer sind als bei homogenen Tests, bei welchen das Ergebnis dieser Aufgabengruppe dann einzeln diagnostisch interpretiert werden soll. Eine Optimierung der Aufgabengruppe auf hohe Trennschärfe hätte die gewollte Aufgabenvielfalt verringert. In der jetzigen Höhe sind sie ein Optimum zwischen Aufgabenvielfalt und Sicherstellung, dass die richtige Lösung von den Leistungsbesten mit grösserer Wahrscheinlichkeit gewählt wird.

Dass die geprüfte Rasch-Skalierung nicht so eindeutig die „Verrechnungsfairness“ bzw. die Unterscheidung von „guten“ und „schlechten“ Personen aufklärt, zeigt eine Gegenüberstellung des tatsächlichen Beitrages der einzelnen Aufgabengruppen zur Unterscheidung der Personen nach Prüfungserfolg¹ und ihre Kategorisierung als weniger homogene Aufgaben. Die Prognosekraft (Validität) ist dabei ein externes Kriterium und entscheidet damit „objektiv“ über die Testgüte: Je höher das Gewicht in einer standardisierten Diskriminanzfunktion zur Trennung der Gruppen nach dem Bestehen der Prüfung, umso „wertvoller“ ist eine Aufgabengruppe. Die Aufgabengruppen mit dem höchsten Prognosebeitrag sind auch diejenigen, welche bezogen auf die Anforderungen „mehrdimensional“ und komplex sind.

¹ Vergleiche Bericht 7 des ZTD (2001). Je höher der Wert, desto höher ist das Gewicht eines Untertests für die Trennung von Erfolgreichen und nicht Erfolgreichen in einer Diskriminanzfunktion. Verwendet wurden die Daten der Studienanfänger 1998 und 1999 und das Bestehen der ersten Vorprüfung.

	Gewicht in einer standardisierten Diskriminanzfunktion zur Trennung der Gruppen nach Prüfungserfolg	Wenig Rasch-homogene Aufgabengruppen ⁸
Quantitative und formale Probleme	.695	X
Diagramme und Tabellen	.665	
Med.-naturwiss. Grundverständnis	.644	X
Textverständnis	.612	X
Muster zuordnen	.489	X
Konzentr. und sorgf. Arbeiten	.426	
Schlauchfiguren	.382	
Figuren lernen	.361	
Fakten lernen	.352	

Tabelle 3: Vergleich der Ergebnisse einer Diskriminanzanalyse (Wichtigkeit der Aufgabengruppe) und wenig rasch-homogene Aufgabengruppen der Evaluation von Spiel u.M.

3.9 Misst der Test überall das Gleiche?

Es wurde behauptet, dass eine Reihe von Aufgaben und Untertests bei Frauen und Männern sowie bei Deutschen und Österreichern nicht das Gleiche messen würde. Auch diese Annahme beruht letztendlich darauf, dass die Rasch-Skalierung für die gemessenen Merkmale angemessen wäre. Wir vermuten, dass dies ein Problem der Parameterschätzung im Rasch-Modell ist – weil es zwischen diesen Gruppen objektiv vergleichsweise grosse Verteilungsunterschiede bei den Ergebnissen in einzelnen Aufgabengruppen gibt.

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 im Bericht 2007 des EMS-AT 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.

3.10 Ist der EMS auf die Studienanforderungen bezogen?

Im November 2005 sind beide Medizinuniversitäten sehr kurzfristig an die Rektorenkonferenz der Schweizer Universitäten (CRUS) herantreten, ob die notwendige Beschränkung der Zulassung ab 2006 auf der Basis des Eignungstests für das Medizinstudium (EMS) erfolgen

⁸ Spiel, C. u.a. (2007). Evaluation der Eignungstests für das Medizinstudium. Evaluationsbericht im Auftrag des Bundesministeriums für Wissenschaft und Forschung.

kann. Es gab zwei wesentliche Argumente für alle Verantwortlichen, den EMS auch in Österreich einzusetzen:

- Der Test wurde in Deutschland und der Schweiz umfangreich positiv evaluiert. Die prognostische Validität war in beiden Ländern sehr gut, auch andere Anforderungen für ein solches Verfahren (elaboriertes Informationssystem mit Test-Info und im WEB, ausreichende Vorbereitungsmöglichkeiten stehen offiziell zur Verfügung) wurden erfüllt.
- Ein erster 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, 1998). 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 dann im EMS dazu.

TMS und EMS sind nicht statisch, die Übereinstimmung mit den Studienanforderungen wird fortlaufend beobachtet (u.a. Workshops mit den Fakultätsvertretern), Veränderungen und Erweiterungen werden vorgenommen. Richtig ist, dass sich die Medizinuniversitäten aus Österreich in diesen Prozess einbringen sollten, wenn der EMS dort weiter eingesetzt werden soll. Dies beinhaltet sowohl eine genauere Analyse der Passung der EMS-Anforderungen für die Studienanforderungen in Österreich, als auch das Einbringen neuer Anforderungen in die geplante Weiterentwicklung des EMS.

3.11 EMS und Änderungen der Studienorganisation?

Bei TMS und EMS wird **immer** angemerkt, dass es eine kontinuierliche Passung zwischen Test (als dem Vorhersageinstrument) und Studienerfolg (als dem Vorherzusagenden) gibt, der aktuelle Test erst einmal nur auf die Vorhersage des aktuellen Studienerfolgs ausgerichtet ist. Ändern sich die Erfolgskriterien, muss sich auch der Test ändern. Diese Entwicklung gab und gibt es aber. Schon in Deutschland wurde der Test mehrfach angepasst, es gab bekanntlich eine lange Evaluationsphase. In der Schweiz wurde der Test ebenfalls weiter angepasst – ein Untertest Planen und Organisieren ergänzt und einer (Konzentriertes und sorgfältiges Arbeiten) modifiziert. Weitere Veränderungen sind geplant, auch hinsichtlich der Erfassung sozialer Fähigkeiten.

In der Schweiz wurden die Studienreformen (problemorientierter Unterricht, verbesserte Betreuungsverhältnisse) ebenfalls umgesetzt und eine neueste Evaluation zeigt, dass auch unter diesen Bedingungen der aktuelle EMS den Studienerfolg sehr gut vorhersagt⁹.

3.12 Soziale Kompetenzen berücksichtigen?

In den öffentlichen Diskussionen wurde oft gefordert, soziale Kompetenzen stärker zu berücksichtigen weil dann ggf. der Leistungs-Unterschied zwischen Männern und Frauen eliminiert würde, weil Frauen sozial kompetenter seien als Männer.

- Hier kann es sich zunächst um einen einfachen Trugschluss handeln: Es gibt keine Hinweise, dass in entsprechend **objektiven Testverfahren** unter gleichen Bedingungen die soziale Kompetenz von Frauen wirklich besser wäre als die von Männern. Hier muss man überlegen, ob die oben dargestellten unterschiedlichen Sozialisationshintergründe und vor allem verschiedenen **Bewertungsmaßstäbe** für die Geschlechter für die Fehlerwartungen verantwortlich sind.

⁹ <http://www.unifr.ch/ztd/ems/emseval07.pdf>

- Selbst wenn sich die soziale Kompetenz unterscheiden würde, wäre ihr **Anteil** an den Studienanforderungen zu bestimmen. Medizin ist und bleibt ein mathematisch-naturwissenschaftlich anspruchsvolles Studium, die Eignung wird von diesen Fähigkeiten vor allem bestimmt. Soziale Kompetenz kann nur eine Facette der Eignung sein. Soziale Kompetenz in Verbindung mit fachlicher Inkompetenz wäre keine Kombination, die für hohe Studien- oder Berufseignung spricht.

Es wird richtig festgestellt, dass die Erfassung sozialer Kompetenzen (kommunikative und sozialkognitive Kompetenzen) im EMS noch nicht erfolgt. Eine Ursache ist, dass eine testpsychologische Erfassung bisher problematisch ist (vgl. Hänsgen 2007). Die meisten in standardisierten Tests verwendbaren Methoden beruhen auf Selbsteinschätzungen des eigenen Erlebens und Verhaltens (Sind Sie...? Neigen Sie zu...?). Die Antworten sind mehr oder weniger durchschaubar und können im Sinne der sozialen Erwünschtheit leicht beeinflusst und ggf. auch durch Trainingskurse erlernt werden – deshalb ist die Verwendung problematisch.

Für Interviews ist auch bekannt, dass ein optimales Verhalten trainiert werden kann. Die Prognosekraft von Interviews ist in der Metaanalyse von Schuler im Übrigen enttäuschend gering.

Die Bedeutung solcher Fähigkeiten steigt mit veränderten Studienanforderungen und es ist eine der laufenden Entwicklungsanstrengungen, zumindest die **Fähigkeiten** auf diesem Gebiet zu erfassen. Hier werden Situationen vorgegeben, die entsprechend einzuschätzen sind bzw. bei welchen eine optimale Entscheidung getroffen werden muss. Es gibt also auch hier ein „mehr oder weniger optimal“ der einzelnen Antwortalternativen, was als objektives Bewertungskriterium verwendet werden kann.

Aufgrund von Untersuchungen in Belgien¹⁰ setzen solche Tests aber offenbar aufwändige Videopräsentationen voraus.

Es wäre auch zu diskutieren, ob in Deutschland dies durch die Berücksichtigung anderer Kriterien wirklich „aufgefangen“ wurde. Es gab zu Zeiten des TMS verschiedene Zulassungsquoten (siehe Trost 1994, S. 139):

- ca. 10% der Plätze gingen als „Vorab-Quote an ausländische Bewerber, Härtefälle, Zweitstudienbewerber und Bewerber mit „besonderer Hochschulzugangsberechtigung“ oder „bevorzugt Zuzulassende“ (z.B. früher wegen Militärdienst zurückgestellt).
- Von den restlichen Plätzen wurden in der folgenden Reihenfolge vier Quoten bedient. Die Zahl der Bewerbungen um einen Studienplatz überstieg die Kapazität zu dieser Zeit um das Acht- bis Neunfache.
 - 45% nach einer Kombination Abiturnote (mit Ausgleich der Länderunterschiede) und TMS. Die Abiturnote allein benachteiligte, wenn man Schul- und Studienleistung in Beziehung setzt, auch dort die Männer wegen einer offenbar strengeren Benotung (siehe Trost 1996).
 - 10% nach dem Testergebnis im TMS
 - 20% nach der Wartezeit (Zahl der Bewerbungssemester), die Hartnäckigkeit belohnte, aber sicher auch von den finanziellen Möglichkeiten beeinflusst war, eine Überbrückungszeit mit einer anderen Ausbildung oder Tätigkeit einzulegen. Hier spielten Leistungskriterien dann keine Rolle und Beharrlichkeit ist nur ein marginaler Aspekt sozialer Kompetenzen.
 - 15% wurden aufgrund eines Interviews von den Universitäten ausgewählt, wobei maximal die dreifache Menge an Personen per Los (!) entsprechend der gewünsch-

¹⁰ Lievens verwendet eine Videopräsentation sozialer Situationen, die einzuschätzen sind. Die Prognosekraft für Studienerfolg ist insgesamt gering. Lievens und Sackett (2007) weisen zudem nach, dass eine Präsentation auf Papier noch geringere Prognosewerte aufweist. In Belgien wird der Test an einem Ort in einem Saal durchgeführt. Die Gewährleistung gleichguter Präsentationsbedingungen in den verschiedenen Testorten und Testlokalen wäre in der Schweiz und in Österreich zumindest extrem aufwändig und der Nutzen ist bisher wegen der geringen Prognosekraft im Verhältnis dazu zu gering.

ten Studienorte den Universitäten zugeteilt wurde. Es sind die niedrigen Prognosewerte für Studienerfolg von Interviews zu berücksichtigen, die in der Metaanalyse von Schuler lediglich bei Korrelationen um 0.10 bis 0.20 liegen.

Die Angehörigen der Wartezeit- und Auswahlgesprächsquote traten später zur ärztlichen Vorprüfung an und hatten – unabhängig vom Zeitpunkt – auch die geringsten Erfolgsraten (Trost 1994, S. 142 f). Die Leistungen dieser beiden Gruppen waren in schriftlichen und mündlichen Prüfungen verglichen mit den anderen Quoten am schlechtesten (S. 150).

Als die Schweiz ein eigenes Zulassungsverfahren konzipierte, wollte man auch aufgrund dieser Erfahrungen weder Interview noch Wartezeit berücksichtigen. Wegen der unterschiedlichen Notenmassstäbe in den (teilweise recht kleinen) Kantonen schien es auch praktisch unmöglich, die Maturitätsnoten zu berücksichtigen. Vor allem fürchtete man eine negative Rückwirkung auf die schulische Praxis der Notengebung, wenn von der Note die Zulassung zum Studium abhängt. Dies würde überall dort eine Rolle spielen, wo es keine zentralen und standardisierten Abiturprüfungen gibt.

3.13 Eignung und Neigung

Wenn mehr Ärzte gewünscht werden, sind vor allem die Kapazitäten der ausbildenden Kliniken begrenzend. Eine einfache Zulassung von mehr Personen würde nichts bringen, wenn diese im Verlauf des vorklinischen Studiums „herausgeprüft“ werden müssten oder Wartezeiten auf Übungen und Praktika einlegen müssten.

Die Erhöhung der Absolventenquote durch Reduktion der Abbrecherquote bleibt ein wichtiges Ziel. Jeder Studienabbruch bedeutet verschwendete Lebenszeit für die Studierenden und nicht effektiv eingesetzte Ressourcen für die Universität.

Nicht alle Entscheidungen sind von der tatsächlichen Eignung beeinflusst – eine Optimierung aller Entscheidungen (siehe Abbildung) scheint sinnvoll:

1. Wählen Personen ein Medizinstudium, die realistische Erwartungen bezüglich Studium und Beruf haben?
2. Sind die Personen für ein Medizinstudium geeignet?
3. Sind die Voraussetzungen vorhanden, das Studium erfolgreich (ohne Abbruch oder Wechsel) zu beenden?
4. Werden die Personen dann einen Medizinalberuf tatsächlich ergreifen?

Nur die Entscheidung (2) wird aktuell durch die Verwendung des EMS für die jeweiligen Universitäten optimiert. Leistungstests als „wettbewerbsorientierte Reihungsverfahren“ sind wie beschrieben nicht geeignet, solche Bereiche wie Interessen, soziale Kompetenzen oder Persönlichkeitsmerkmale zu erfassen.

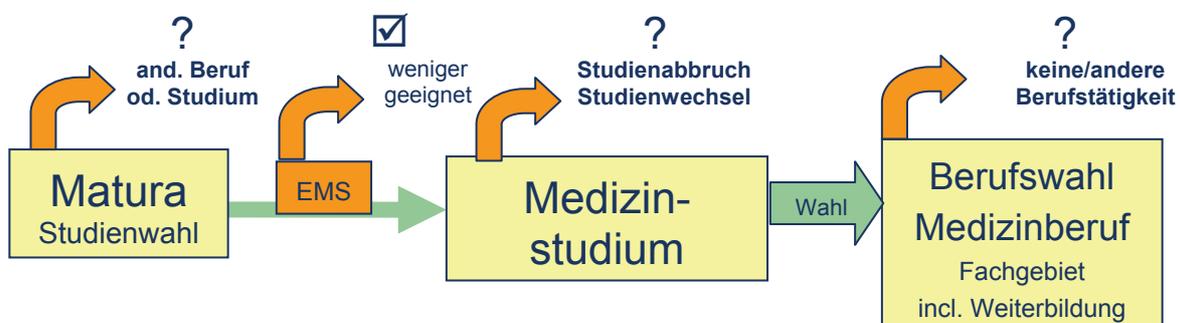


Abbildung 2: Auswahlprozesse im Rahmen der Medizinausbildung

In einigen Universitäten und Universitätsverbänden werden für andere Studienrichtungen (bisher nicht für Medizin, sondern Psychologie, Wirtschafts- und Rechtswissenschaften, technische Studienrichtungen) Portale für ein „**Self Assessment**“ angeboten. Hier besteht

kein direkter Zusammenhang zwischen Ergebnis und Zulassung, sondern die individuellen Entscheidungsprozesse der Personen werden unterstützt. Insofern sind Antworten nach der sozialen Erwünschtheit weniger relevant.

Für ein solches Self-Assessment-Portal bieten sich neben der **Informationsvermittlung** über Studium und Beruf folgende Assessment-Elemente an:

- Die eigenen **Interessen** werden erfasst und mit den durchschnittlichen Interessenprofilen der Studierenden einer Fachrichtung verglichen. (*Stimmen meine Interessen damit überein?*)
- Die eigenen **Vorstellungen** über die **Studianforderungen** und den zukünftigen **Beruf** werden erfasst und ebenfalls mit Durchschnittsprofilen erfolgreicher Studierender verglichen. (*Sind meine Erwartungen realistisch?*)
- **Soziale Kompetenzen**, andere Merkmale der **Persönlichkeit** („soft skills“) werden anhand von typischen Erlebens- und Verhaltensweisen erfragt und es erfolgt ebenfalls eine Rückmeldung auf der Basis des Vergleiches mit dem Durchschnitt (erfolgreicher) Studierender.
- Auch bestimmte **Fähigkeiten** („hard skills“) können auf dieser Basis durch Wissenstests oder Fähigkeitstests geprüft werden, sodass man vor dem EMS auch Rückmeldungen zur Studieneignung erhält. Dies trifft auch für soziale Fähigkeiten zu.
- Langfristig ist es auch möglich, die Vorbereitung auf den EMS (Beispielaufgaben) mit diesem Portal zu unterstützen.

Ziel ist eine **Rückmeldung an die Interessentinnen und Interessenten vor der eigentlichen Bewerbung**, welche ggf. vorhandene Defizite aufdeckt. Wenn die Ergebnisse vertraulich bleiben und die Zulassung nicht beeinflussen, sollte dies möglich sein. Die entsprechenden Schlussfolgerungen (mehr Information, Aufarbeitung der Defizite oder Wahl eines anderen Studiums) **kann und muss der Interessent allerdings selbst ziehen**. In einigen Universitäten ist die Teilnahme an einem solchen Assessment verpflichtend. Es ist allerdings auch möglich, dass sich das Interesse für Medizin verstärkt und sich noch mehr Personen bewerben (bei EMS-Einsatz dann aber unproblematisch).

3.14 Perspektive

Einen für die Zulassung geeigneten Test zu entwickeln und laufend an die sich verändernden Studienbedingungen anzupassen, ist eine grosse Herausforderung an die dafür notwendigen Ressourcen. Ein Land wie die Schweiz wäre aus Kostengründen gar nicht allein in der Lage gewesen, ein solches Projekt wie den EMS durchzuführen. Auch in Deutschland wurde der TMS seinerzeit mit aus Kostengründen eingestellt (und man hat ihn heute reaktiviert, weil ein faires und funktionierendes Zulassungskriterium nur auf der Basis der Abiturnoten und der anderen Kriterien suboptimal ist).

Es gäbe nun eine Chance, dass sich alle drei Länder (Deutschland, Schweiz und Österreich) so zusammentun, dass die weitere Testentwicklung gemeinsam erfolgt. Auf der Basis eines allgemeinen Rahmenkonzeptes kann man den Spezifika der Anforderungen der Länder bei Testkonzeption und Weiterentwicklung ausreichend gerecht werden (was die entsprechende Zusammenarbeit Deutschland-Schweiz belegt). Vergleicht man dies mit den Ressourcen entsprechender Institute im Ausland (Educational Testing Service bzw. AAMC in den USA, welche den dortigen Mediziner-Test betreuen), würde sich dann auch im deutschen Sprachraum eine entsprechende „kritische Masse“ ergeben, um jederzeit eine moderne Testentwicklung zu garantieren (auch mit der entsprechenden Grundlagenforschung) und die Kosten für alle Seiten zu minimieren. Dafür wären Strukturen einer paritätischen Zusammenarbeit zu finden, eine grundsätzliche Bereitschaft ist vorhanden.

4 Testanwendung in Österreich 2008

Im Jahr 2008 nahmen 4381 Personen am EMS teil. Gegenüber 2007 (3940) ist dies eine Steigerung um 11%. In der Österreich-Quote beträgt die Zunahme 8%, in der EU-Quote 14%. In der Nicht-EU-Quote hat die Zahl der Teilnahmen um 11% abgenommen.

4.1 Statistische Angaben zur Teilnahme

			EU	nicht EU	Österreich	Gesamt
Innsbruck	männlich	geb. vor 1985	53	4	25	82
		geb. 1985-1987	207	4	47	258
		geb. nach 1987	131	2	199	332
		Gesamt	391	10	271	672
	weiblich	geb. vor 1985	42	2	15	59
		geb. 1985-1987	192	3	53	248
		geb. nach 1987	236	4	308	548
		Gesamt	470	9	376	855
	Wien	männlich	geb. vor 1985	61	7	76
geb. 1985-1987			188	12	179	379
geb. nach 1987			129	14	543	686
		Gesamt	378	33	798	1209
weiblich		geb. vor 1985	64	6	106	176
		geb. 1985-1987	179	14	126	319
		geb. nach 1987	244	24	882	1150
		Gesamt	487	44	1114	1645
Gesamt			1726	96	2559	4381

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

		Innsbruck	Wien	Gesamt
EU	Humanmedizin	771	792	1563
	Human- und Zahnmedizin	38	23	61
	Zahnmedizin	52	50	102
	Gesamt	861	865	1726
Nicht EU	Humanmedizin	17	62	79
	Human- und Zahnmedizin	1	7	8
	Zahnmedizin	1	8	9
	Gesamt	19	77	96
Österreich	Humanmedizin	560	1707	2267
	Human- und Zahnmedizin	56	117	173
	Zahnmedizin	31	88	119
	Gesamt	647	1912	2559
Gesamt		1527	2854	4381

Tabelle 5: Zulassungsquoten und Disziplinwahl

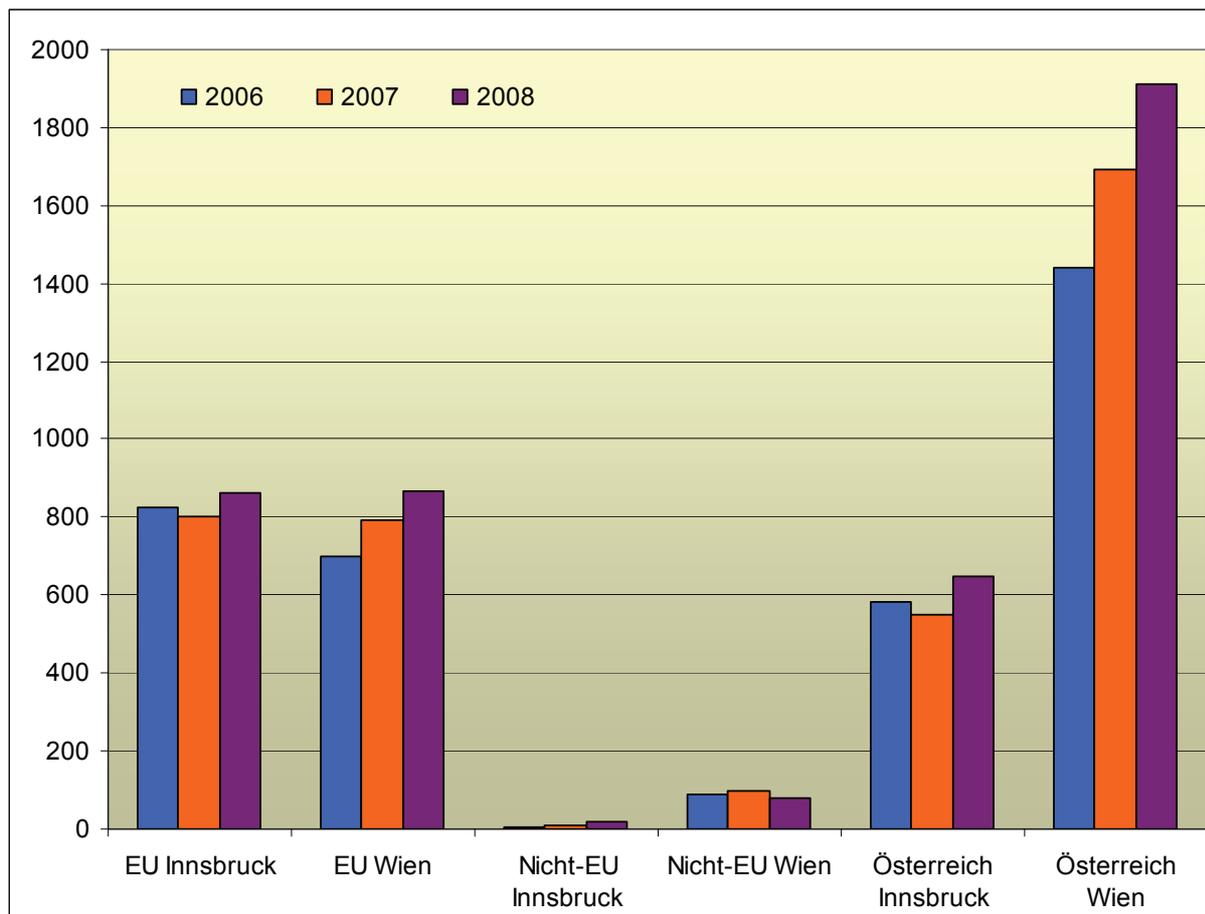


Abbildung 3: Quoten im Vergleich 2006, 2007 und 2008, Absolutzahlen

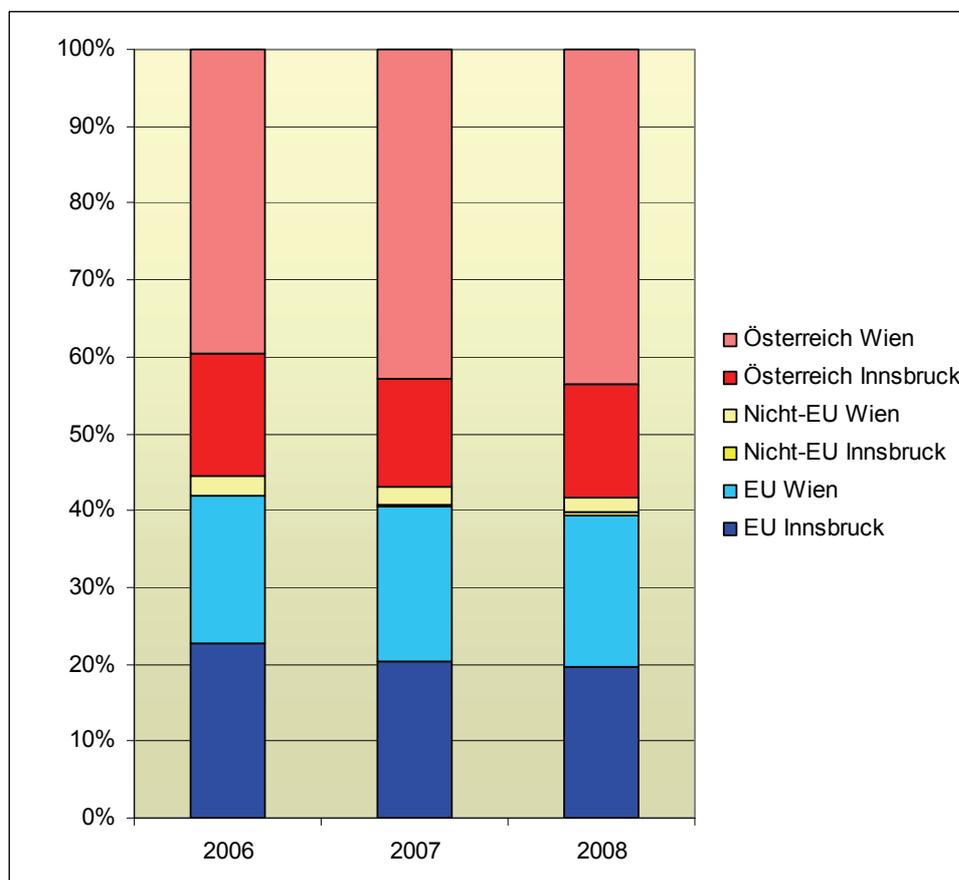


Abbildung 4: Quoten im Vergleich 2006, 2007 und 2008, prozentuale Aufteilung

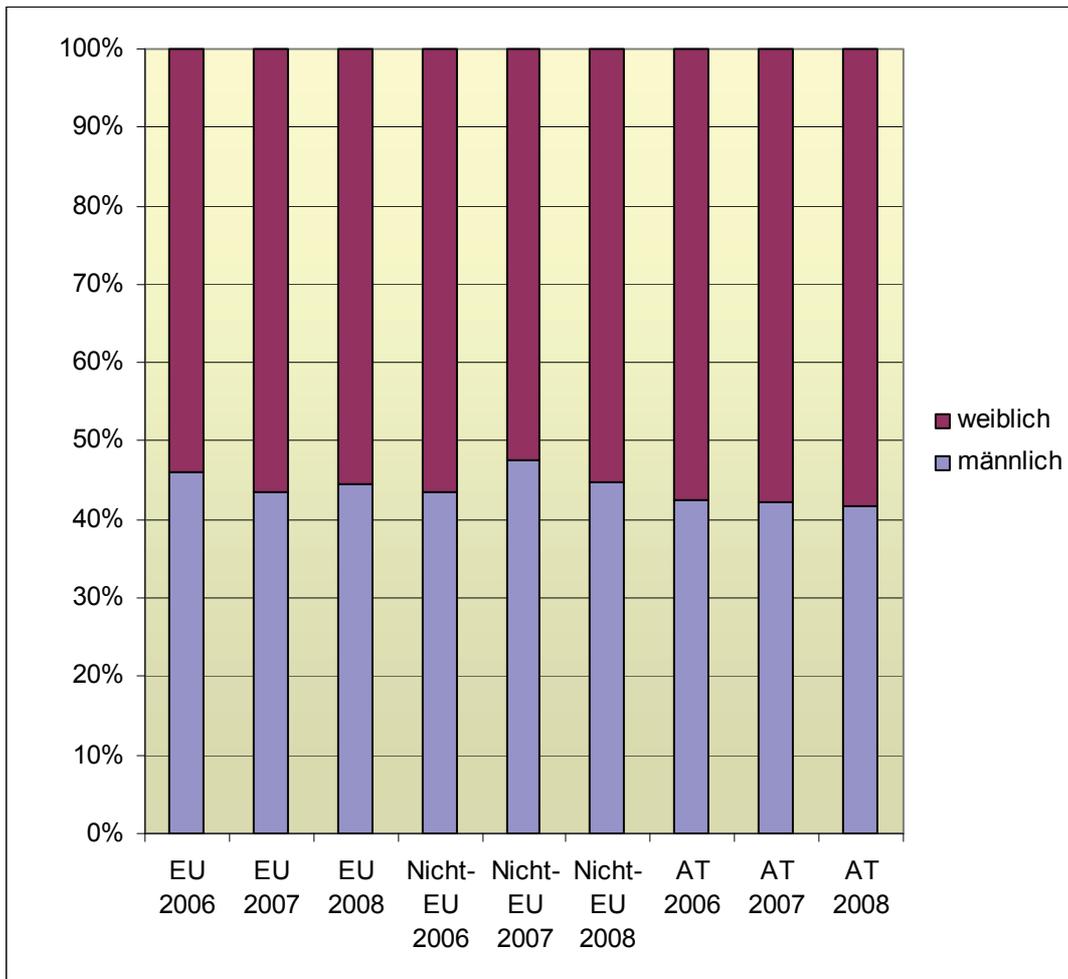


Abbildung 5: Geschlechterverhältnisse pro Quote für 2006 bis 2008

Die Zunahme der EU-Quote verteilt sich auf Innsbruck und Wien etwa gleich. Die Zunahme der Österreich-Quote ist in Wien am deutlichsten. Der Rückgang in der Nicht-EU-Quote zeigt sich vor allem in Wien. Insgesamt zeigen sich über alle drei Jahre der Anwendung des EMS allerdings recht konstante Verhältnisse der Bewerbungen (Abbildung 4).

Auch die Verhältnisse der Geschlechter sind zwischen 55 und 60% Frauen zu 45 und 40% Männer in etwa gleich.

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	104.40	8.66	12.80	7.82	8.99	10.30	12.25	12.17	11.92	8.49	10.99
Median	105	8	13	8	9	10	12	12	12	8	11
Modalwert	105	7	14	7	8	11	11	11	13	8	10
Stand.-abweich.	23.193	3.633	3.735	3.297	3.348	3.706	3.844	3.669	3.096	3.245	4.872
Spannweite	139	20	20	18	20	20	19	19	20	19	20
25. Perzentil	88	6	10	5	7	8	10	10	10	6	8
75. Perzentil	120	11	16	10	11	13	15	15	14	11	15

Tabelle 6: Kennwerte des Punktwerts und für die Untertests

Maßgeblich für die Zulassung ist der Testwert, welcher auf dem Punktwert basiert. Dieser ist normalverteilt, metrische statistische Prüf-Verfahren sind also angemessen. Die einzelnen Untertests sind zu Vergleichszwecken mit angegeben.

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.

In den Untertests „Schlauchfiguren“ und „Muster zuordnen“ wurden tendenziell höhere Punktwerte erreicht, die Untertests „Quantitative und formale Probleme“, „Textverständnis“, „Planen und Organisieren“ sowie „Diagramme und Tabellen“ gehörten 2008 tendenziell zu den schwierigeren Elementen.

Der Punktwert wird nach der Formel von Seite 8 in den Testwert umgerechnet. Dieser hat dann den Mittelwert 100 und die Standardabweichung 10.

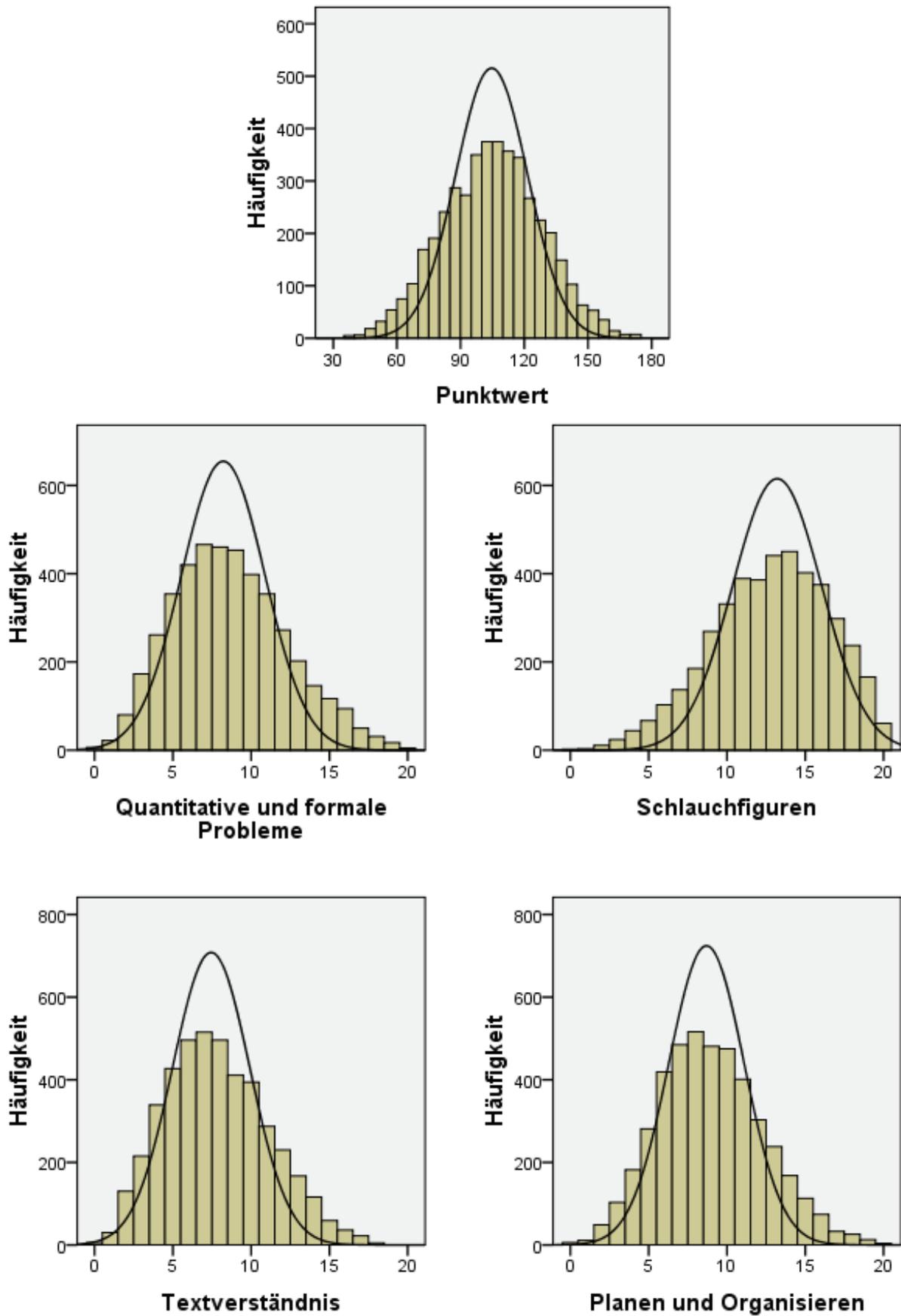


Abbildung 6: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (1)

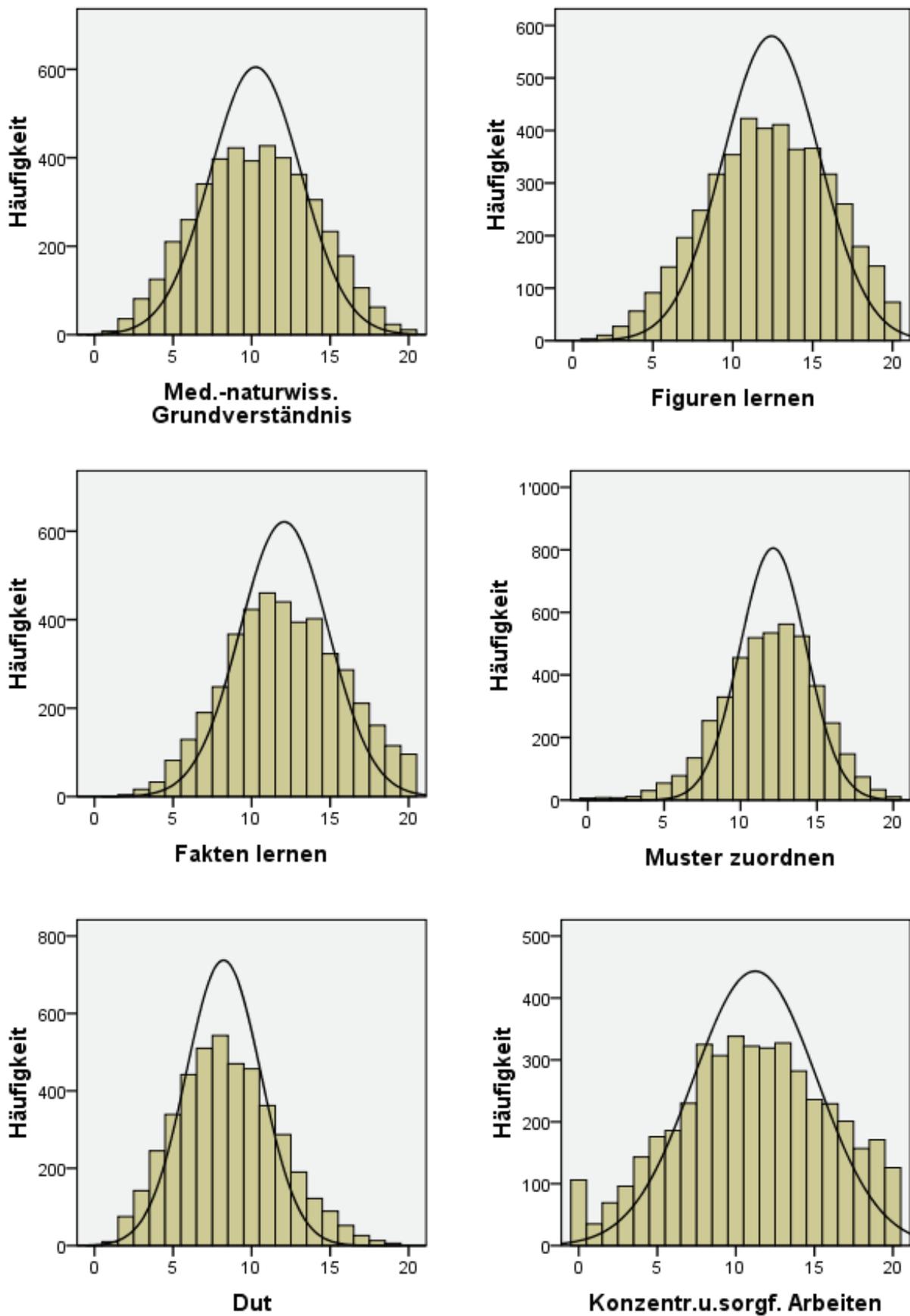


Abbildung 7: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (2)

4.3 Vergleich der Geschlechter

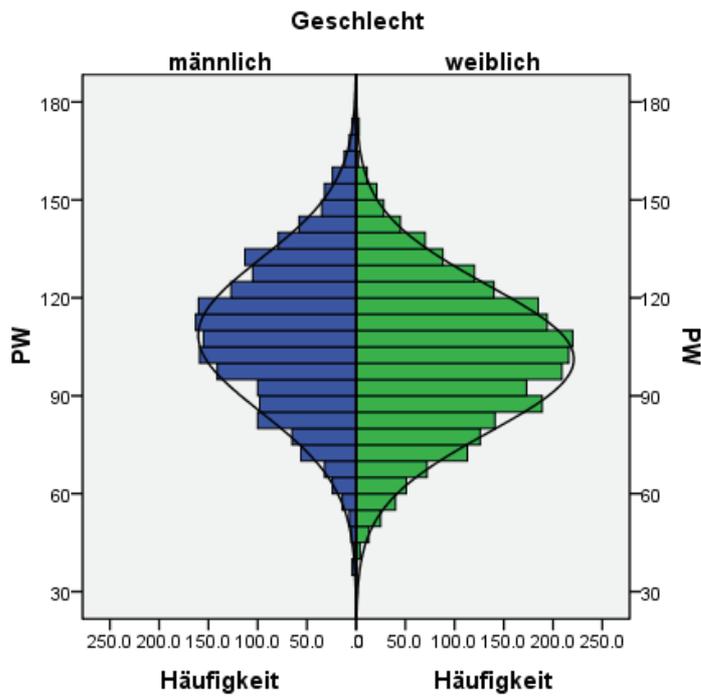


Abbildung 8: Häufigkeitsverteilungen des Punktwertes für Männer und Frauen, Gesamt

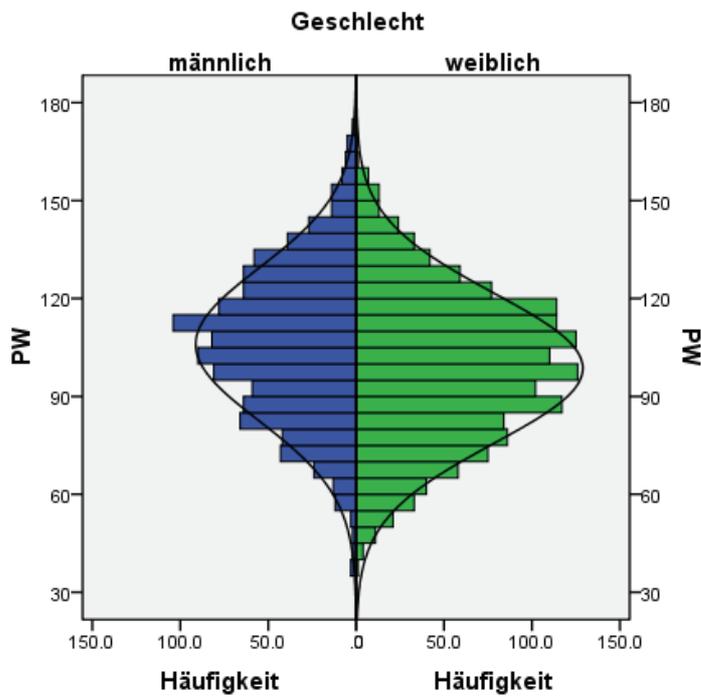


Abbildung 9: Häufigkeitsverteilungen des Punktwertes für Männer und Frauen, nur Österreichquote

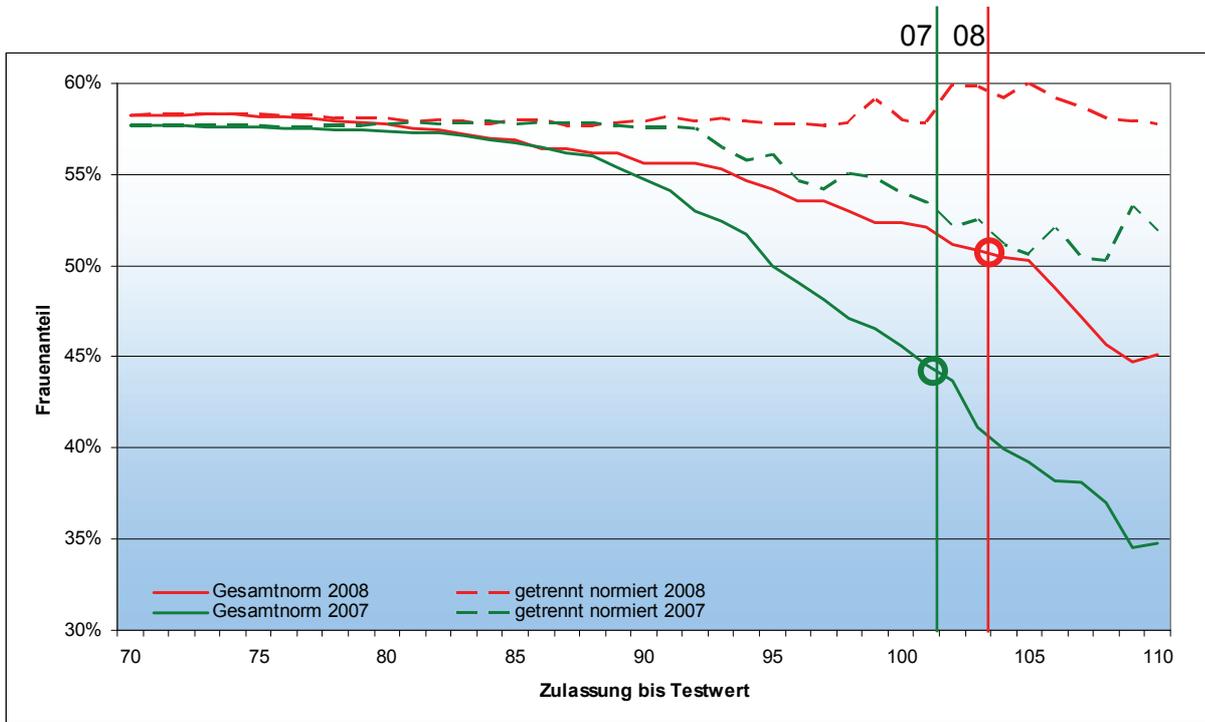


Abbildung 10: Geschlechterverhältnis bei Zulassung mit einem bestimmten Testwert für die **Österreicher-Quote**. Die durchgezogenen Linien basieren auf dem gemeinsam standardisierten Testwert, die gestrichelten Linien auf dem separat für Männer und Frauen standardisierten Testwert. Senkrechte Linien: Zulassungsgrenzwerte für das jeweilige Jahr.

Die Verteilung der Testergebnisse für Männer und Frauen zeigt auch 2008 eine leichte Verschiebung zu besseren Werten für die Männer. In Abbildung 10 wird dargestellt, wie hoch der Anteil der Frauen wäre, wenn der Zulassungsgrenzwert beim jeweiligen Testwert (einschliesslich) liegen würde. Beim niedrigsten Testwert (70) entspricht dies dem Verhältnis der Bewerbungen. Die **durchgezogenen** Linien entsprechen den verwendeten Testwerten, die für Männer und Frauen auf **den gleichen Mittelwert und die gleiche Standardabweichung** bezogen sind. Es fällt auf, dass vor allem im Bereich der sehr hohen Testleistungen der Männeranteil höher ist als die tatsächlichen Bewerbungsverhältnisse betragen. 2008 ist der Effekt allerdings viel weniger deutlich als 2007 – es haben sich mehr Frauen beworben, die sehr gute Leistungen erreicht haben.

Wenn man wie beschrieben für Männer und Frauen **getrennte Standardisierungen** vornehmen würde (gestrichelte Linien – siehe auch Seite 11), entspräche 2008 das Geschlechterverhältnis der Zugelassenen bei jedem Testwert der Bewerberquote – im Bereich der sehr guten Leistungen wäre die Frauenquote sogar etwas grösser. 2007 hätte diese unterschiedliche Standardisierung nicht diese Auswirkungen gehabt, die Verteilungen der Testwerte unterschieden sich deutlicher, weil offenbar im höchsten Leistungsbereich Frauen deutlich unterrepräsentiert waren. Diese Angleichungen 2008 sind auch ein Zeichen, dass die eingeleiteten kurzfristigen Massnahmen sich tatsächlich ausgleichend ausgewirkt haben und der eingeschlagene Weg richtig ist.

Für 2009 wären zwei Szenarien denkbar:

1. Erneut gleiche Standardisierung für Männer und Frauen (dann gibt es kein Fairnessproblem in dem Sinne, dass gleiche Testwerte nicht gleiche Zulassungschancen bedeuten)
2. Separate Standardisierung für Männer und Frauen – begründet mit der kurzfristig nicht änderbaren bildungssoziologischen Benachteiligung für Frauen als politische Entscheidung

Tabelle 7 beschreibt die Geschlechtsunterschiede auf Untertestebene. Mit Ausnahme der Untertests „Figuren lernen“, „Fakten lernen“ und „Konzentriertes und sorgfältiges Arbeiten“ erzielen die Bewerberinnen signifikant geringere Ergebnisse als die Bewerber (Tabelle 8).

			Testwert	Quant. u. form. Probl.	Schlauchfig.	Textverständnis	Planen und Organisieren	Med.-naturw. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme u. Tabellen	Konz. u. sorgf. Arbeiten
EU	männl.	m	104.0	10.3	14.4	9.3	9.7	11.9	12.3	12.4	12.3	10.0	10.8
		s	9.4	3.5	3.2	3.3	3.3	3.5	3.8	3.7	3.0	3.2	4.7
	weibl.	m	100.6	8.5	12.4	8.0	9.2	10.7	12.6	12.8	12.1	8.5	11.0
		s	9.0	3.3	3.5	3.1	3.1	3.4	3.7	3.5	2.9	3.0	4.7
nicht EU	männl.	m	91.5	8.2	11.3	5.5	6.2	7.0	11.0	9.0	10.4	6.8	9.3
		s	11.0	4.3	4.7	3.4	2.8	4.0	4.1	4.1	2.7	3.2	5.3
	weibl.	m	95.1	8.1	11.4	5.8	6.1	7.3	12.4	12.0	12.6	6.3	11.1
		s	10.7	4.3	4.0	3.0	2.2	3.3	4.4	4.1	3.7	3.0	5.8
AT	männl.	m	100.6	9.2	13.6	8.2	9.2	10.5	11.8	11.6	12.0	8.9	10.8
		s	10.1	3.7	3.5	3.2	3.5	3.7	3.8	3.7	3.1	3.3	4.9
	weibl.	m	97.5	7.6	11.8	6.8	8.6	9.2	12.4	12.2	11.6	7.5	11.2
		s	9.9	3.4	3.8	3.1	3.3	3.6	3.9	3.6	3.2	3.0	5.0
Ge- samt	männl.	m	101.8	9.6	13.9	8.6	9.3	11.0	12.0	11.9	12.1	9.3	10.8
		s	10.1	3.7	3.5	3.4	3.4	3.7	3.8	3.7	3.1	3.3	4.8
	weibl.	m	98.6	7.9	12.0	7.2	8.8	9.7	12.4	12.4	11.8	7.8	11.2
		s	9.7	3.4	3.7	3.1	3.3	3.6	3.9	3.6	3.1	3.0	4.9

Tabelle 7: Geschlechterbezogene Unterschiede in Untertests

	Levene-Test Varianzgleichheit		T-Test für die Mittelwertgleichheit				
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz
Testwert	1.187	.276	10.413	4379	.000	3.14376	.30191
Quantitative und formale Probleme	17.763	.000	15.642	3862.010	.000	1.707	.109
Schlauchfiguren	15.109	.000	16.955	4189.945	.000	1.854	.109
Textverständnis	16.180	.000	13.693	3892.812	.000	1.362	.099
Planen und Organisieren	4.920	.027	5.492	3935.591	.000	.563	.103
Med.-naturwiss. Grundverständnis	4.266	.039	11.604	3946.010	.000	1.301	.112
Figuren lernen	2.152	.142	-3.547	4379	.000	-.416	.117
Fakten lernen	1.743	.187	-4.758	4379	.000	-.532	.112
Muster zuordnen	.000	.982	3.279	4379	.001	.310	.094
Diagramme und Tabellen	16.981	.000	15.310	3854.247	.000	1.495	.098
Konzentr. u. sorgf. Arbeiten	.981	.322	-2.461	4081.477	.014	-.365	.148

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 11). 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.

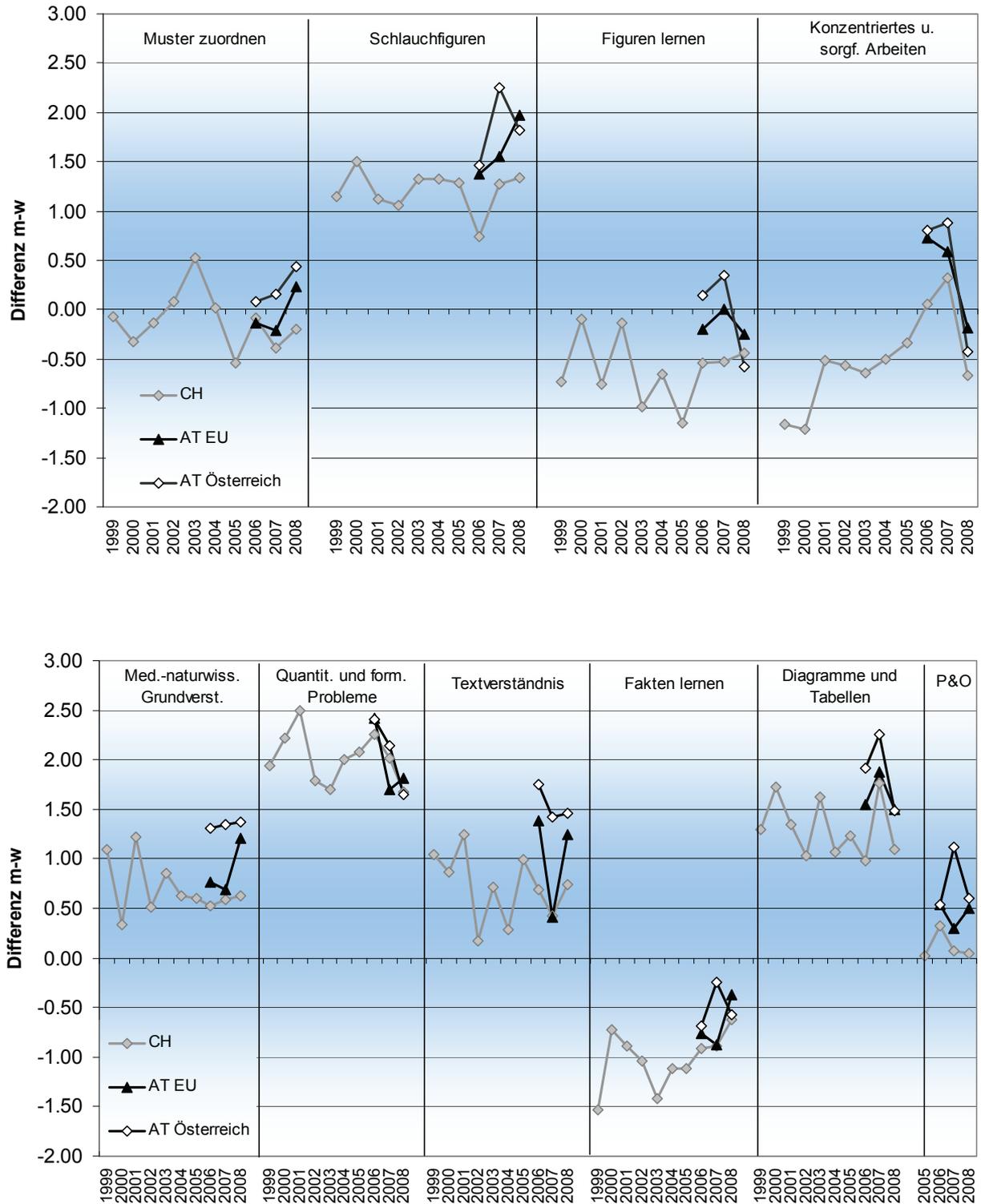


Abbildung 11: Geschlechterspezifische Unterschiede Schweiz 1999 bis 2008 und Vergleich mit den Ergebnissen in Österreich

Vergleicht man die Kurven AT und AT-EU wird deutlich, dass sich die Unterschiede 2008 massiv reduzieren. In sechs der zehn Untertests ist auch eine Annäherung der Kurven AT und CH zu beobachten (Schlauchfiguren, Figuren lernen, konzentriertes und sorgfältiges Ar-

beiten, Quantitative und formale Probleme, Textverständnis, Fakten lernen), in den übrigen Untertests bleibt diese Differenz in etwa konstant.

Die Abbildung 12 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 war 2007 praktisch identisch mit dem Jahrgang 2006. Allerdings wurde unter den EU-Kandidaten der Unterschied reduziert, während er unter AT-Bewerbern anstieg. 2008 ist insgesamt eine Reduktion der Differenz zu beobachten, wobei diese vor allem auf die Leistungen der AT-BewerberInnen zurückzuführen ist (die Gruppe Nicht-EU ist zu klein). Am geringsten ist der Geschlechtsunterschied in der Gruppe der kombinierten Disziplinen und den älteren Personen mit früher Maturität (2. Bildungsweg).

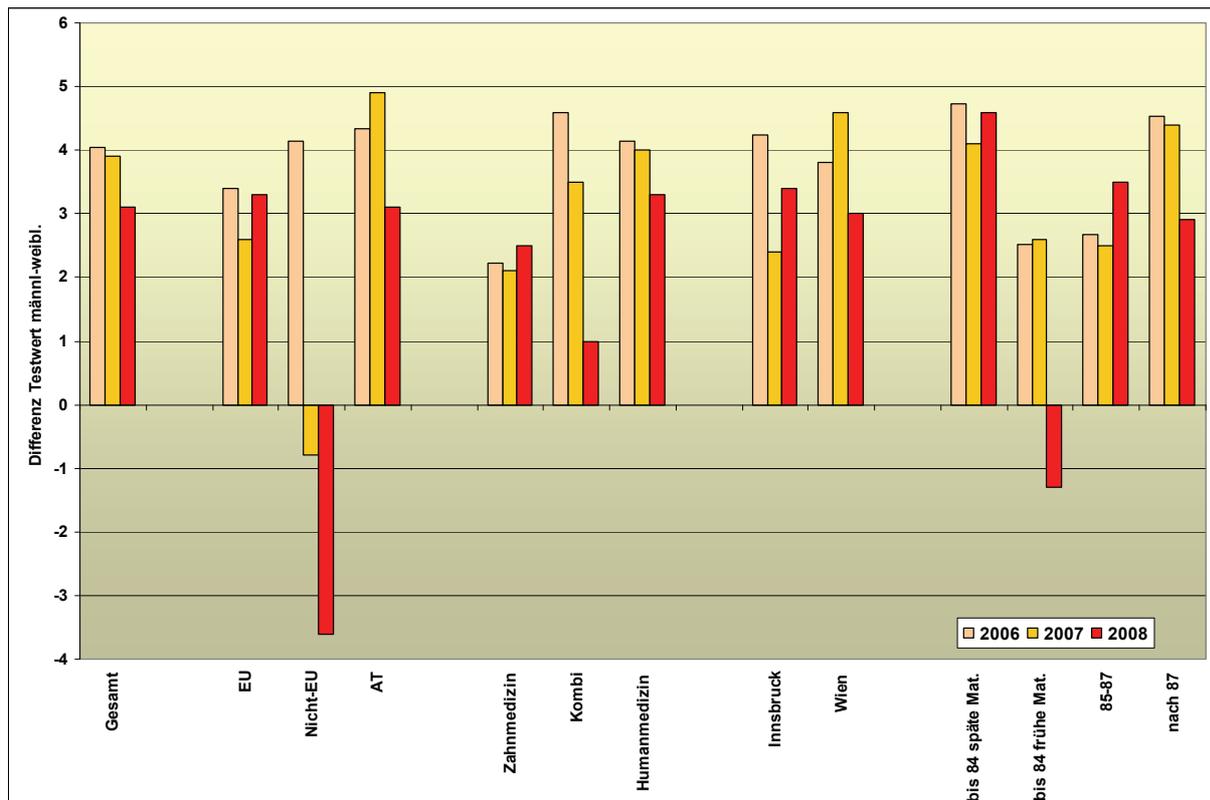


Abbildung 12: Geschlechterspezifische Unterschiede AT 2006 bis 2008 nach Kohorten. Geburtsjahre sind für 2008 angegeben, Vorjahre entsprechend versetzt

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
EU	103.95	769	9.45	100.60	957	8.98	102.10	1726	9.34
nicht EU	91.49	43	10.95	95.08	53	10.74	93.47	96	10.93
Österreich	100.64	1069	10.08	97.51	1490	9.95	98.82	2559	10.12
Total	101.78	1881	10.10	98.64	2500	9.73	99.99	4381	10.01

Tabelle 9: Testwert für Quoten und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
Humanmedizin	102.12	1681	10.04	98.83	2228	9.74	100.25	3909	10.00
Human- plus Zahnmedizin	100.29	107	9.90	99.29	135	9.98	99.73	242	9.94
Zahnmedizin	97.44	93	10.30	94.88	137	8.54	95.92	230	9.36

Tabelle 10: Testwert für Disziplin und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
Innsbruck	102.44	672	9.74	99.03	855	9.39	100.53	1527	9.70
Wien	101.42	1209	10.27	98.44	1645	9.90	99.70	2854	10.17

Tabelle 11: Testwert für Testort und Geschlecht

	Geschlecht								
	männlich			weiblich			Total		
	m	n	s	m	n	s	m	n	s
geb. bis 1984, späte Matura	99.88	185	11.85	95.29	181	10.52	97.61	366	11.43
geb. bis 1984, frühe Matura	96.02	41	10.86	97.33	54	10.21	96.77	95	10.46
geb. 1985-1987	102.74	637	10.05	99.24	567	9.45	101.09	1204	9.92
geb. nach 1987	101.77	1018	9.63	98.84	1698	9.66	99.94	2716	9.75

Tabelle 12: Testwert für Alter, Maturität und Geschlecht

		Geschlecht								
		männlich			weiblich			Total		
		m	n	s	m	n	s	m	n	s
Innsbruck	EU	104.57	391	9.26	100.76	470	8.89	102.49	861	9.25
	nicht EU	97.40	10	12.31	99.67	9	13.67	98.47	19	12.66
	Österreich	99.56	271	9.56	96.86	376	9.47	97.99	647	9.59
Wien	EU	103.31	378	9.61	100.45	487	9.08	101.70	865	9.42
	nicht EU	89.70	33	10.03	94.14	44	9.97	92.23	77	10.17
	Österreich	101.01	798	10.23	97.73	1114	10.10	99.09	1912	10.28

Tabelle 13: Testwert für Testort und Quote

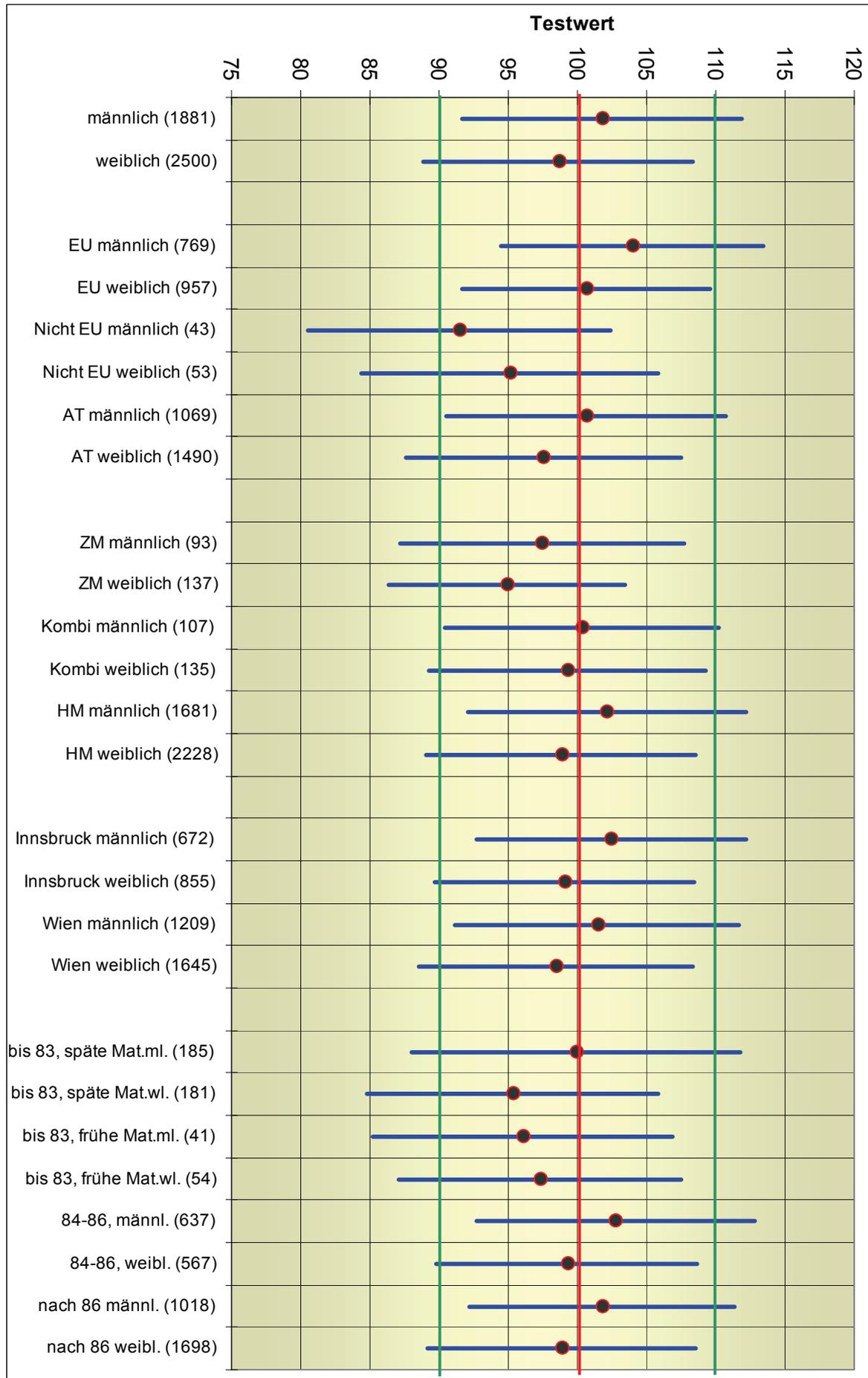


Abbildung 13: Mittelwerte und Standardabweichungen 2008 für verschiedene Teilgruppen

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 Grenznähe zu Deutschland begründet scheint. In Wien bewerben sich demgegenüber bedeutend mehr 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	% Gesamt
Land bzw. Quote	EU	861	56.4%	865	30.3%	1726	39.4%
	nicht EU	19	1.2%	77	2.7%	96	2.2%
	Österreich	647	42.4%	1912	67.0%	2559	58.4%
Gesamt		1527	100%	2854	100%	4381	100%

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 primär auf die unterschiedliche Zusammensetzung der Bewerberkohorten zurückzuführen ist: In einer zweifachen Varianzanalyse mit den Faktoren Testort und Länderquote wird der Einfluss des Testortes als auf 5%-Niveau signifikant identifiziert, die Effektgröße ist mit .001 jedoch vernachlässigbar.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Sektoren Innsbruck	Zwischen Gruppen	1474.632	11	134.057	1.469	.139
	Innerhalb Gruppen	57959.312	635	91.275		
	Total	59433.944	646			
Sektoren Wien	Zwischen Gruppen	2183.615	14	155.973	1.481	.110
	Innerhalb Gruppen	199749.439	1897	105.298		
	Total	201933.054	1911			

Tabelle 15: Varianzanalytische Prüfung der Homogenität für Punktwerte zwischen den Sektoren in Innsbruck und Wien (Quote „Österreich“), keine Unterschiede signifikant

	Testlokal	m	s	Levene's Test Varianz		t-test Mittelwert			
				F	Sig.	t	df	Sig.	
Testwert	Innsbruck	100.5	9.7	homogen	2.78	.095	2.63	4379	<u>.008</u>
	Wien	99.7	10.2						

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

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: 1726; Nicht-EU: 96; Österreich: 2559). Die Leistungen dieser drei Gruppen unterscheiden sich wie in Tabelle 17 dargestellt.

	Maturitäts- quote	Mittel- wert	Stan- dard- abw.	Quadrat- summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	EU	102.1	9.3	15262.7	2	7631.3	78.8	.000	■ ■ ■
	Nicht EU	93.5	10.9	423769.9	4378	96.8			■ ■ ■
	Österreich	98.8	10.1	439032.6	4380				■ ■ ■
Quantitative und formale Probleme	EU	9.3	3.5	1166.2	2	583.1	45.1	.000	■ ■ ■
	Nicht EU	8.1	4.3	56648.8	4378	12.9			■ ■ ■
	Österreich	8.3	3.6	57815.0	4380				■ ■ ■
Schlauchfiguren	EU	13.3	3.5	788.6	2	394.3	28.6	.000	■ ■ ■
	Nicht EU	11.4	4.3	60308.8	4378	13.8			■ ■ ■
	Österreich	12.5	3.8	61097.4	4380				■ ■ ■
Textverständnis	EU	8.6	3.2	1904.9	2	952.5	91.2	.000	■ ■ ■
	Nicht EU	5.7	3.1	45703.1	4378	10.4			■ ■ ■
	Österreich	7.4	3.2	47608.0	4380				■ ■ ■
Planen und Organisieren	EU	9.4	3.2	1207.4	2	603.7	55.2	.000	■ ■ ■
	Nicht EU	6.1	2.5	47874.3	4378	10.9			■ ■ ■
	Österreich	8.8	3.4	49081.7	4380				■ ■ ■
Med.-naturwiss. Grundverständnis	EU	11.3	3.5	3366.1	2	1683.1	129.8	.000	■ ■ ■
	Nicht EU	7.2	3.6	56778.1	4378	13.0			■ ■ ■
	Österreich	9.8	3.7	60144.3	4380				■ ■ ■
Figuren lernen	EU	12.4	3.7	119.6	2	59.8	4.1	.017	■ ■ ■
	Nicht EU	11.7	4.3	64603.2	4378	14.8			■ ■ ■
	Österreich	12.1	3.9	64722.8	4380				■ ■ ■
Fakten lernen	EU	12.6	3.6	673.1	2	336.6	25.3	.000	■ ■ ■
	Nicht EU	10.7	4.3	58297.8	4378	13.3			■ ■ ■
	Österreich	11.9	3.6	58970.9	4380				■ ■ ■
Muster zuordnen	EU	12.2	3.0	224.1	2	112.0	11.7	.000	■ ■ ■
	Nicht EU	11.6	3.4	41760.1	4378	9.5			■ ■ ■
	Österreich	11.7	3.2	41984.1	4380				■ ■ ■
Diagramme und Tabellen	EU	9.2	3.2	1682.8	2	841.4	82.9	.000	■ ■ ■
	Nicht EU	6.5	3.1	44435.9	4378	10.1			■ ■ ■
	Österreich	8.1	3.2	46118.7	4380				■ ■ ■
Konzentr. und sorgf. Arbeiten	EU	10.9	4.7	60.8	2	30.4	1.3	.278	■ ■ ■
	Nicht EU	10.3	5.6	103907.1	4378	23.7			■ ■ ■
	Österreich	11.1	4.9	103967.9	4380				■ ■ ■

Tabelle 17: 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 1984 Geborenen ist 461, davon Maturität früh: 95, Maturität spät: 366, 1985-1987 geboren: 1204, nach 1987 geboren: 2716.

	Altersgruppe	Mittelwert	Standard-abw.	Quadrat-summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	Bis 1984, späte Mat.	97.6	11.4	4535.8	3	1511.9	15.2	.000	
	Bis 1984, frühe Mat.	96.8	10.5						
	1985-1987	101.1	9.9						
	Nach 1987	99.9	9.8						
Quantitative und formale Probleme	Bis 1984, späte Mat.	8.2	4.0	103.0	3	34.3	2.6	.050	
	Bis 1984, frühe Mat.	8.8	4.1						
	1985-1987	8.8	3.7						
	Nach 1987	8.6	3.6						
Schlauchfiguren	Bis 1984, späte Mat.	12.5	4.0	436.5	3	145.5	10.5	.000	
	Bis 1984, frühe Mat.	12.3	4.6						
	1985-1987	13.3	3.6						
	Nach 1987	12.6	3.7						
Textverständnis	Bis 1984, späte Mat.	7.7	3.5	219.1	3	73.0	6.7	.000	
	Bis 1984, frühe Mat.	7.5	3.3						
	1985-1987	8.2	3.3						
	Nach 1987	7.7	3.2						
Planen und Organisieren	Bis 1984, späte Mat.	8.3	3.4	279.8	3	93.3	8.4	.000	
	Bis 1984, frühe Mat.	8.2	3.5						
	1985-1987	9.0	3.3						
	Nach 1987	9.1	3.3						
Med.-naturwiss. Grundverständnis	Bis 1984, späte Mat.	10.2	4.0	468.0	3	156.0	11.4	.000	
	Bis 1984, frühe Mat.	10.7	3.8						
	1985-1987	10.8	3.8						
	Nach 1987	10.1	3.6						
Figuren lernen	Bis 1984, späte Mat.	11.5	4.0	408.6	3	136.2	9.3	.000	
	Bis 1984, frühe Mat.	10.9	3.7						
	1985-1987	12.4	3.9						
	Nach 1987	12.3	3.8						
Fakten lernen	Bis 1984, späte Mat.	11.5	3.6	561.4	3	187.1	14.0	.000	
	Bis 1984, frühe Mat.	10.4	3.5						
	1985-1987	12.5	3.7						
	Nach 1987	12.2	3.6						
Muster zuordnen	Bis 1984, späte Mat.	11.3	3.4	358.6	3	119.5	12.6	.000	
	Bis 1984, frühe Mat.	10.7	3.0						
	1985-1987	12.2	3.0						
	Nach 1987	11.9	3.1						
Diagramme und Tabellen	Bis 1984, späte Mat.	8.1	3.5	168.6	3	56.2	5.4	.001	
	Bis 1984, frühe Mat.	8.2	3.5						
	1985-1987	8.8	3.3						
	Nach 1987	8.4	3.2						
Konzentr. und sorgf. Arbeiten	Bis 1984, späte Mat.	9.5	5.0	1137.6	3	379.2	16.1	.000	
	Bis 1984, frühe Mat.	9.4	5.3						
	1985-1987	11.1	4.9						
	Nach 1987	11.2	4.8						

Tabelle 18: 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.

In der EU-Gruppe schneiden die Personen der mittleren Altersgruppe am besten ab, in der AT-Gruppe sind es die jüngsten Anwärter.

		Innsbruck			Wien		
		m	n	s	m	n	s
EU	Bis 1984, späte Maturität	101.3	81	9.5	100.2	102	11.2
	Bis 1984, frühe Maturität	99.4	14	9.0	96.4	23	7.8
	1985-1987	103.2	399	9.1	103.2	367	9.1
	Nach 1987	102.1	367	9.3	100.9	373	9.0
AT	Bis 1984, späte Maturität	96.2	40	10.1	94.8	129	11.8
	Bis 1984, frühe Maturität				97.5	53	11.0
	1985-1987	95.7	100	8.5	98.2	305	10.5
	Nach 1987	98.6	507	9.7	99.7	1425	9.9

Abbildung 14: Altersunterschiede im Testwert nach Zulassungsquote und 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, schneiden in der Quote „Österreich“ gegenüber den anderen Kandidaten am besten ab. In der Quote „EU“ erzielen hingegen jene Personen, die sich für ein Studium der Humanmedizin bewerben, die höchsten Ergebnisse.

		Innsbruck			Wien			Gesamtwert		
		m	n	s	m	n	s	m	n	s
Human- medizin	EU	102.8	771	9.3	102.2	792	9.3	102.5	1563	9.3
	nicht EU	98.7	17	13.3	92.0	62	10.1	93.5	79	11.1
	Österreich	97.9	560	9.5	99.3	1707	10.2	98.9	2267	10.1
	Gesamt	100.7	1348	9.8	100.0	2561	10.1	100.2	3909	10.0
Kombi- niert	EU	101.1	38	10.2	98.7	23	9.6	100.2	61	9.9
	nicht EU	102.0	1	-	95.7	7	8.7	96.5	8	8.4
	Österreich	100.1	56	10.2	99.6	117	10.0	99.7	173	10.0
	Gesamt	100.5	95	10.1	99.2	147	9.8	99.7	242	9.9
Zahn- medizin	EU	98.3	52	6.4	95.3	50	8.7	96.8	102	7.7
	nicht EU	91.0	1	-	90.9	8	12.4	90.9	9	11.6
	Österreich	96.5	31	9.5	95.2	88	10.6	95.5	119	10.3
	Gesamt	97.6	84	7.7	95.0	146	10.1	95.9	230	9.4

Abbildung 15: Testwert nach Wunschdisziplin und Zulassungsquote 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 2008 mit 0.92 in beiden Ländern im oberen 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	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
DE / CH	.91 - .93	.92	.91	.90	.91	.91	.92	.91	.92	.90	.90	.92
AT										.92	.93	.92

Tabelle 19: Reliabilität des Punktwertes (Split Half) für Deutschland (Bereich), Schweiz und Österreich

	Konsistenz des Testprofils											
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	
CH	.83	.81	.81	.80	.80	.81	.78	.82	.79	.80	.82	
AT										.83	.84	.83

Tabelle 20: 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.-naturwis. 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		0.23	0.42	0.23	0.19	0.34	0.32	0.27	0.34	0.21	0.56	0.58
Med.-naturw. Grundverst.	0.27		0.32	0.53	0.60	0.26	0.27	0.61	0.20	0.47	0.70	0.73
Schlauchfiguren	0.43	0.38		0.32	0.28	0.40	0.33	0.32	0.33	0.25	0.63	0.67
Quant. u. formale Probl.	0.26	0.56	0.39		0.46	0.20	0.22	0.54	0.19	0.44	0.65	0.68
Textverständnis	0.21	0.61	0.32	0.49		0.23	0.27	0.52	0.16	0.44	0.64	0.64
Figuren lernen	0.37	0.28	0.41	0.23	0.19		0.47	0.26	0.35	0.21	0.61	0.60
Fakten lernen	0.31	0.33	0.31	0.24	0.26	0.44		0.24	0.29	0.23	0.59	0.58
Diagramme und Tabellen	0.31	0.62	0.39	0.58	0.54	0.26	0.29		0.21	0.50	0.69	0.72
Konzent. u. sorgfält. Arbeiten	0.37	0.19	0.33	0.20	0.16	0.33	0.27	0.20		0.20	0.58	0.55
Planen und Organisieren	0.24	0.50	0.31	0.46	0.43	0.22	0.25	0.53	0.21		0.62	0.63

Tabelle 21: 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					
3.1	3.2	30.5	31.8	30.5	31.8
2.5	2.4	24.6	24.3	55.1	56.2
Drei-Faktorenlösung					
3.0	3.2	30.3	31.6	30.3	31.6
1.8	1.8	18.1	18.0	48.4	49.7
1.5	1.5	14.6	14.5	63	64.2

Tabelle 22: Varianzanteile der einzelnen Faktorenlösungen Schweiz und Österreich 2008 (varimax-rotierte Lösungen)

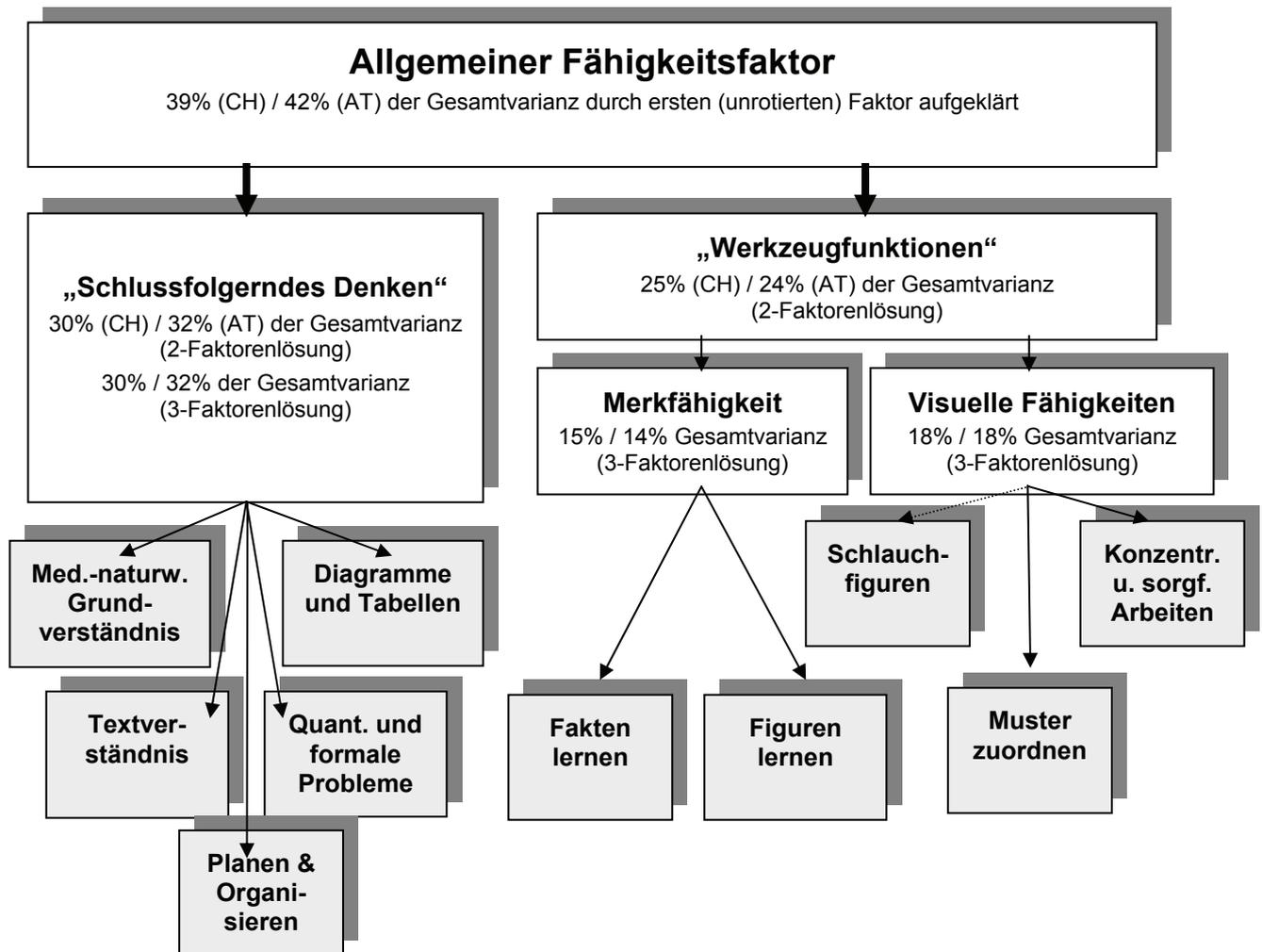


Abbildung 16: Struktur der Untertests des EMS, hierarchische Faktorenstruktur 2008, Werte Schweiz/Österreich der Varianzaufklärung, Untertests entsprechend der Hauptladungen zugeordnet.

Die Struktur des EMS bleibt 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“ sowie „quantitative und formale Probleme“ charakterisiert (Tabelle 21).

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 24).

Die Dreifaktorenlösung (Tabelle 25) teilt den Faktor „Werkzeugfunktionen“ in „Merkfähigkeit“ (repräsentiert durch „Fakten lernen“ und „Figuren lernen“) und „visuelle Fähigkeiten“ („Konzentriertes und sorgfältiges Arbeiten“, „Muster zuordnen“, „Schlauchfiguren“).

	2002	2003	2004	2005	2006 CH	2006 AT	2007 CH	2007 AT	2008 CH	2008 AT
Muster zuordnen	.71	.48	.45	.46	.47	.52	.43	.51	.53	.56
Med.-naturwiss. Grundverst.	.70	.74	.71	.79	.74	.77	.74	.74	.75	.77
Schlauchfiguren	.69	.61	.61	.58	.57	.64	.60	.66	.61	.66
Quant. und formale Probleme	.68	.73	.70	.75	.71	.72	.73	.75	.68	.72
Textverständnis	.62	.77	.70	.76	.72	.74	.72	.73	.69	.69
Figuren lernen	.62	.55	.54	.48	.47	.54	.48	.55	.56	.55
Fakten lernen	.57	.51	.51	.47	.48	.51	.48	.55	.55	.55
Diagramme und Tabellen	.56	.76	.74	.78	.74	.76	.76	.77	.74	.77
Konzentr. u. sorgf. Arbeiten	.54	.48	.44	.41	.52	.58	.43	.49	.48	.46
Planen und Organisieren				.68	.51	.50	.62	.67	.65	.67

Tabelle 23: Ladungen im ersten Faktor der unrotierten Lösung 2002 bis 2008 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor); 2002-2005 Schweiz und 2006-2008 für beide Länder

	Faktor 1						Faktor 2						Kommunalitäten (h ²)					
	06 CH	06 AT	07 CH	07 AT	08 CH	08 AT	06 CH	06 AT	07 CH	07 AT	08 CH	08 AT	06 CH	06 AT	07 CH	07 AT	08 CH	08 AT
Diagramme und Tabellen	.80	.78	.82	.81	.79	.80					.20	.21	.66	.67	.69	.70	.66	.69
Med.-naturwiss. Grundverständnis	.77	.76	.79	.81	.81	.81						.21	.64	.66	.64	.68	.69	.70
Textverständnis	.77	.76	.74	.76	.76	.77						.11	.62	.64	.58	.62	.60	.61
Quant. und formale Probleme	.79	.78	.79	.77	.74	.76						.19	.64	.64	.64	.64	.58	.61
Planen und Organisieren	.61	.62	.69	.71	.70	.70						.18	.37	.39	.49	.53	.52	.52
Konzentr. u. sorgf. Arbeiten							.62	.57	.64	.67	.66	.68	.41	.40	.41	.46	.45	.47
Figuren lernen							.73	.75	.69	.73	.74	.75	.54	.56	.49	.56	.57	.57
Muster zuordnen					.15	.17	.66	.68	.67	.70	.68	.70	.45	.48	.46	.50	.48	.52
Fakten lernen					.18	.22	.64	.64	.64	.58	.66	.61	.42	.42	.44	.40	.47	.43
Schlauchfiguren			.29	.33	.27	.36	.68	.68	.65	.66	.65	.61	.50	.54	.50	.54	.50	.51

Tabelle 24: Ladungen und Kommunalitäten der Zwei-Faktorenlösung, varimax-rotiert 2006-2008

Faktor	Ergebnisse 2008								Deutschland		
	1		2		3		h ²		1	2	3
	2008 CH	2008 AT	2008 CH	2008 AT	2008 CH	2008 AT	2008 CH	2008 AT			
Diagramme und Tabellen	.79	.80	.21	.19			.67	.69	0.82		
Medizin.-naturwiss. Grundverständnis	.81	.81				.22	.69	.71	0.81		0.2
Quantitat. u. formale Probleme	.74	.76	.22	.22			.59	.62	0.8	0.18	
Textverständnis	.76	.77			.23		.63	.62	0.79		0.2
Planen und Organisieren	.70	.70		.17			.52	.52	Noch nicht enthalten		
Muster zuordnen		.17	.78	.74		.18	.65	.61		0.81	
Konzentr. u. sorgfält. Arbeiten			.69	.78	.21		.52	.63		0.7	0.4
Schlauchfiguren	.26	.36	.67	.59	.21	.24	.56	.54	0.35	0.71	
Fakten lernen		.21			.84	.86	.77	.79	0.21		0.87
Figuren lernen			.35	.39	.74	.72	.69	.68	0.13	0.47	0.64

Tabelle 25: Faktorenanalyse: Varimaxrotierte Drei-Faktorenlösung, Österreich, 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).

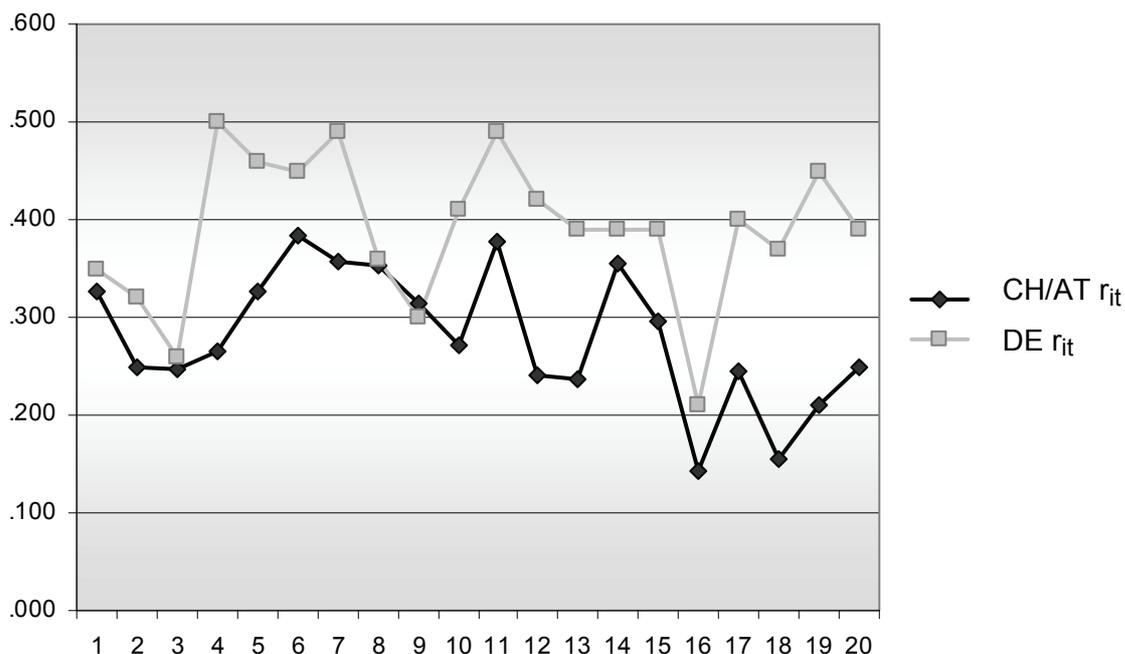


Abbildung 17: Trennschärfen für den Untertest „quantitative und formale Probleme“.

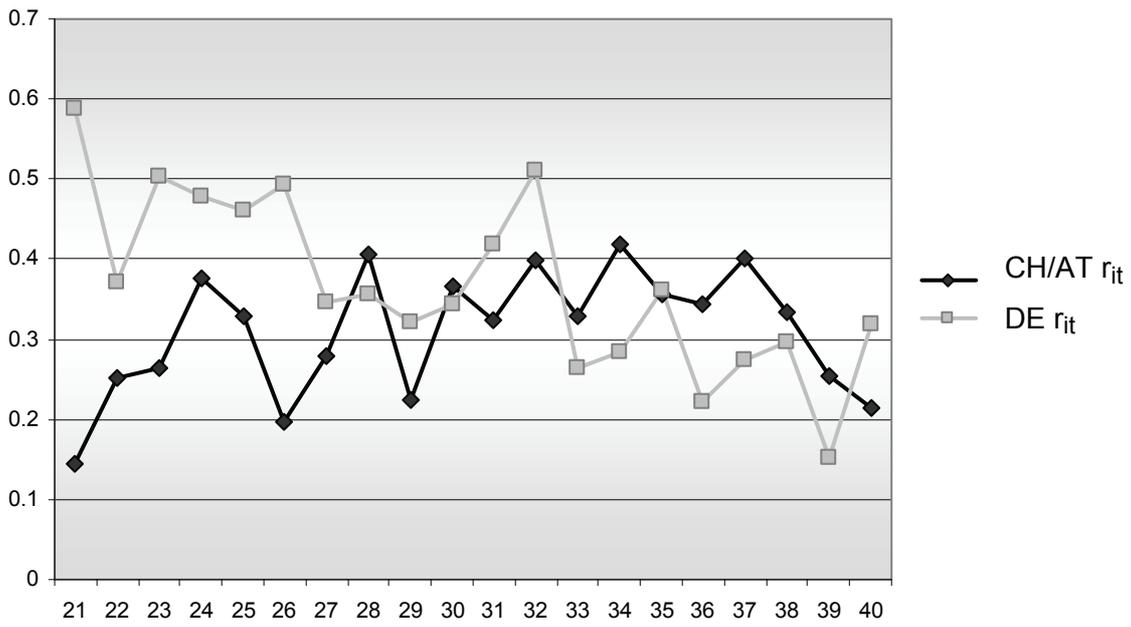


Abbildung 18: Trennschärpen für den Untertest „Schlauchfiguren“.

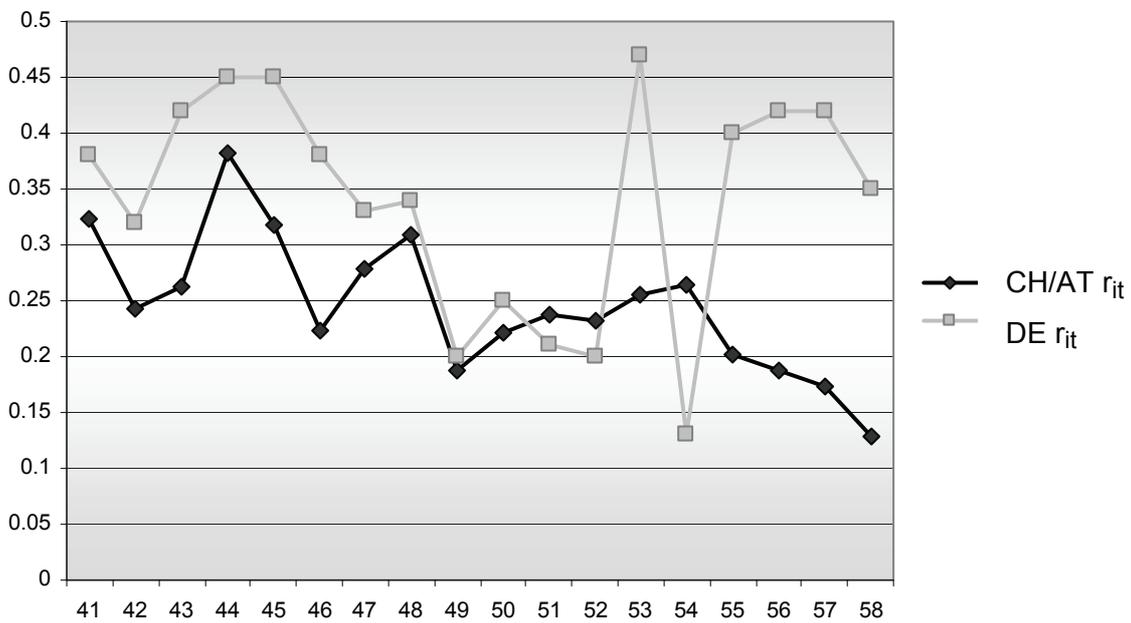


Abbildung 19: Trennschärpen für den Untertest „Textverständnis“.

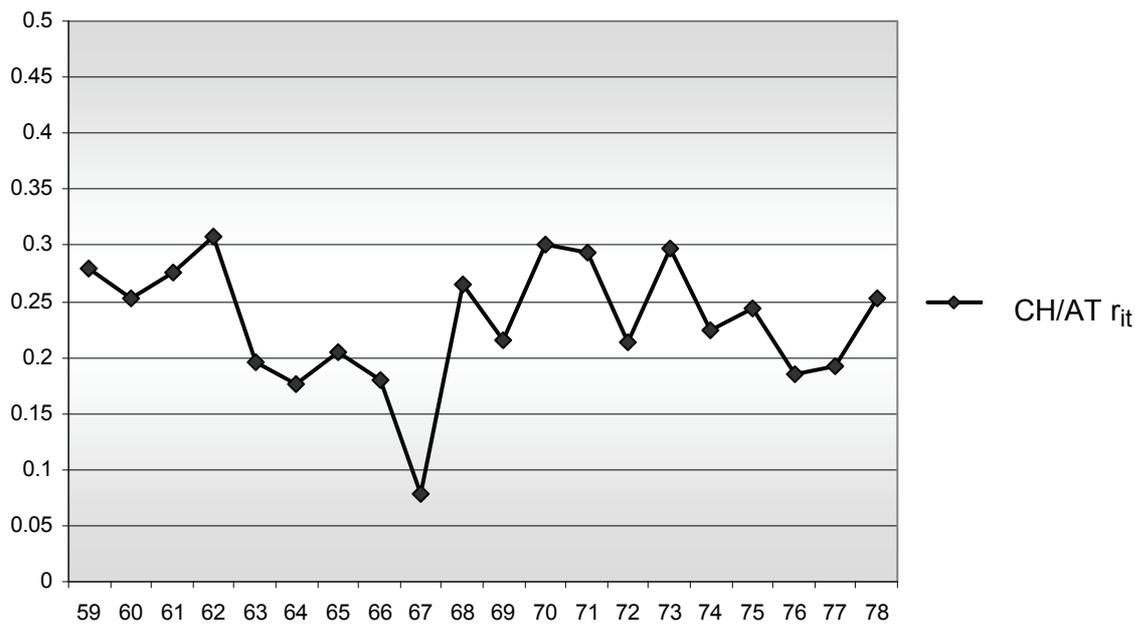


Abbildung 20: Trennschärpen für den Untertest „Planen und Organisieren“ (keine deutschen Vergleichswerte).

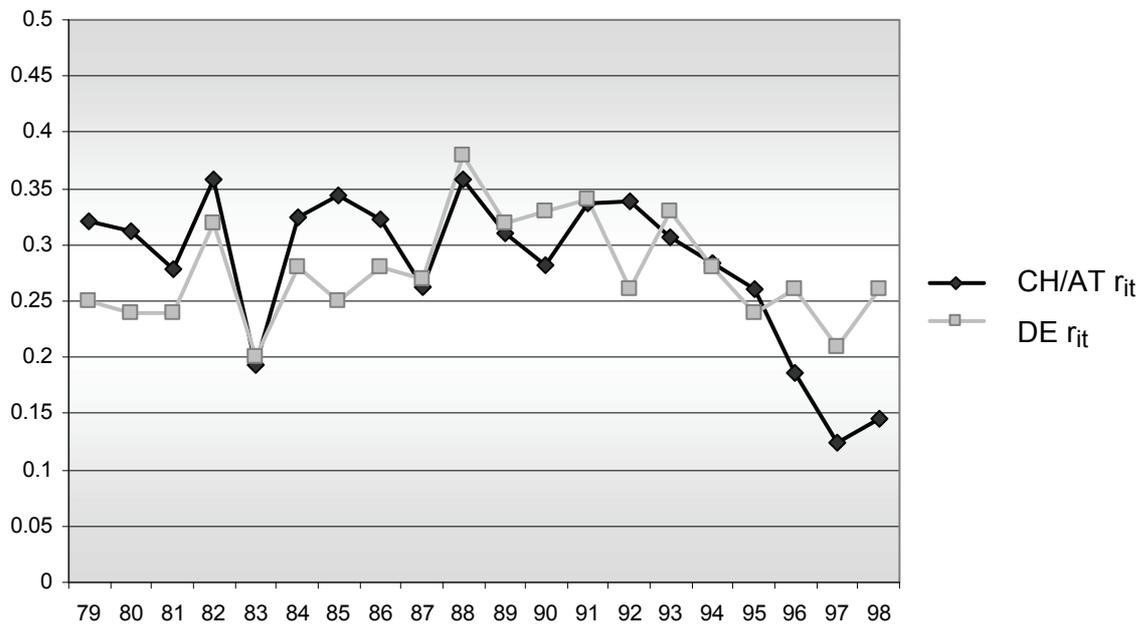


Abbildung 21: Trennschärpen für den Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“.

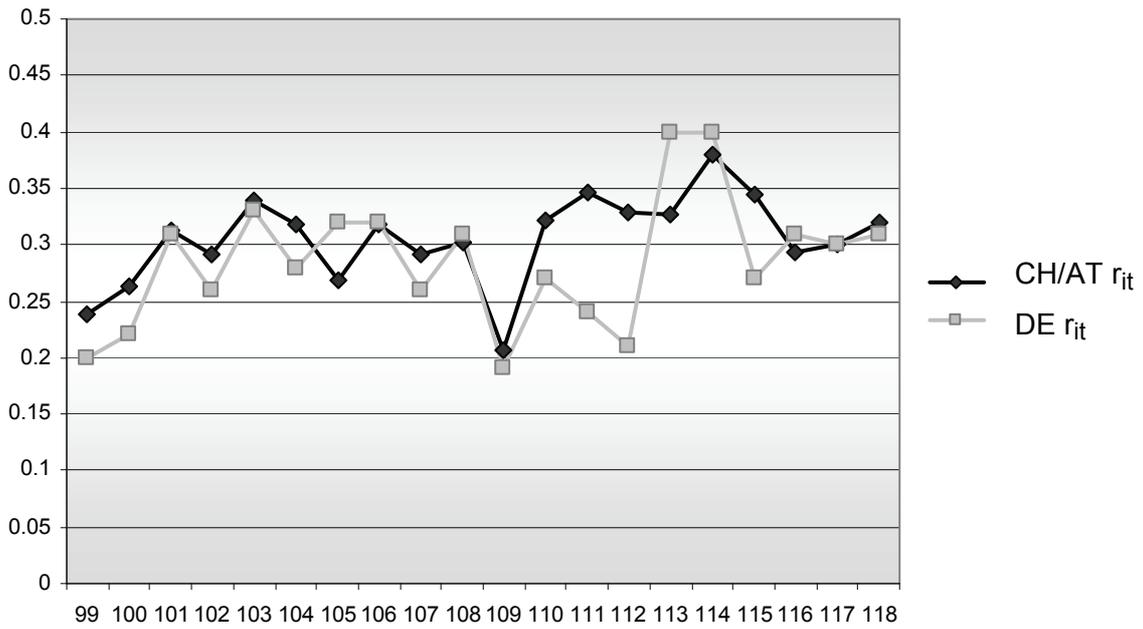


Abbildung 22: Trennschärfen für den Untertest „Figuren lernen“.

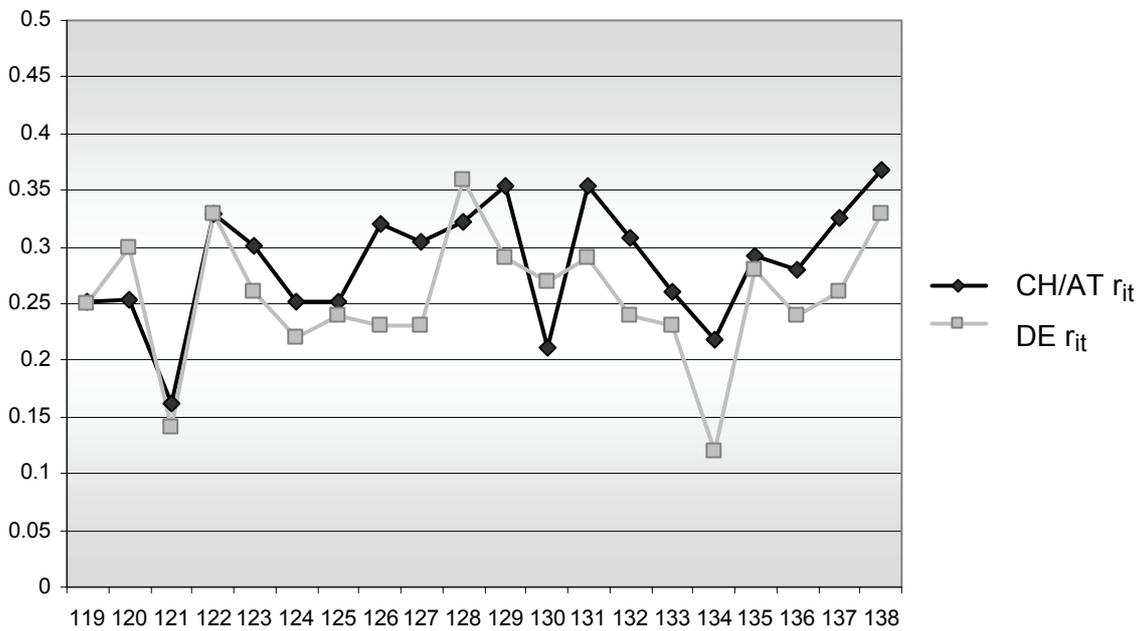


Abbildung 23: Trennschärfen für den Untertest „Fakten lernen“.

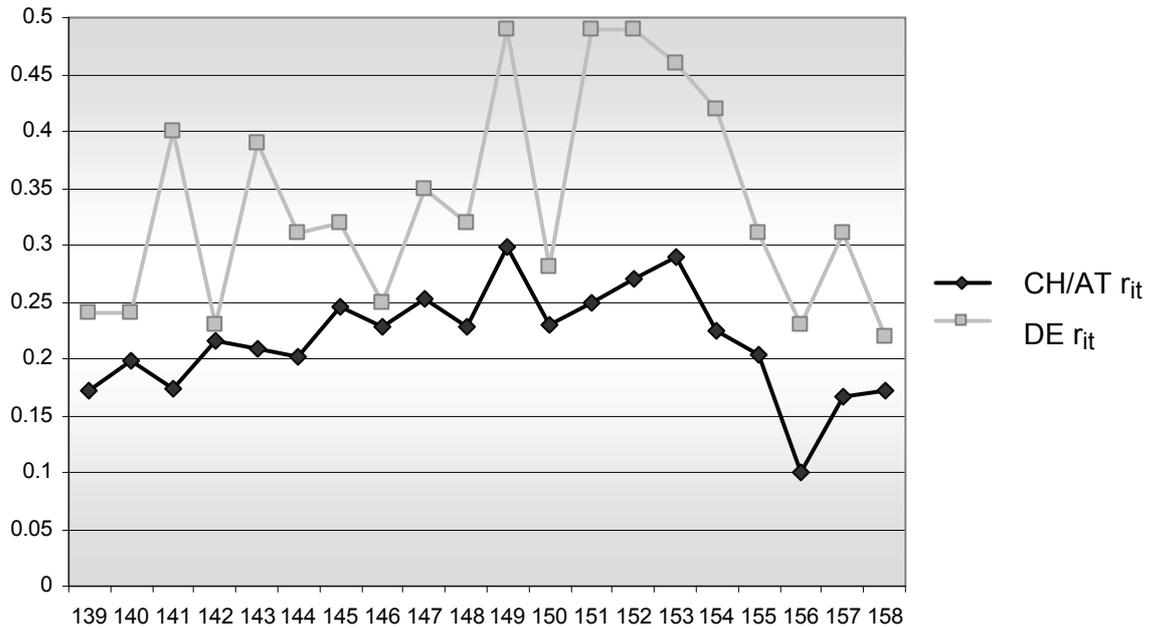


Abbildung 24: Trennschärfen für den Untertest „Muster zuordnen“.

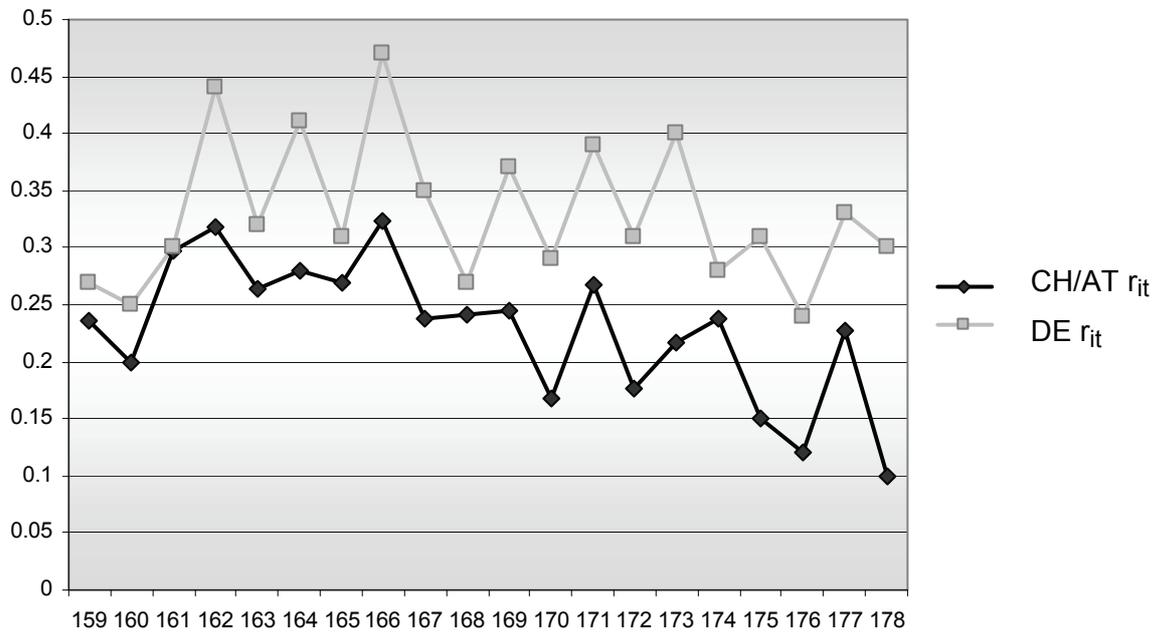


Abbildung 25: Trennschärfen für den Untertest „Diagramme und Tabellen“.

5.4 Konzentriertes und sorgfältiges Arbeiten

Dieser Untertest wurde 2004 revidiert und seither 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). Da nun jedes Jahr Zeichen und Regel vorher nicht bekannt sind, wirken sich Lerneffekte durch exzessives Üben nicht mehr so deutlich aus. In den Jahren 2004 bis 2008 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.

2008 bestand der Untertest aus den Zeichen „f“ und „t“. Es sollte jedes „f“ markiert werden, wenn direkt danach ein „t“ folgte.

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

Der Punktwert (PW) berechnet sich wie folgt:

$$\text{PW} = \text{Richtige} - \text{Fehler I} - \text{Fehler IIa} - \text{Fehler IIb}$$

Der Rohwert Richtige minus Fehler beträgt in Österreich im Mittel 294.8 mit einer Standardabweichung von 70.0. Damit war die Version 2008 verglichen mit den bisherigen Instruktionen die leichteste Variante, wobei die Differenzierungsfähigkeit dadurch nicht beeinträchtigt war.

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 Ver-

fahren gewährleistet, dass auch unterschiedlich schwierige Konzentrationstests zu einer vergleichbaren 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
Richtige	4381	306.3	56.4	0	400
Fehler Gesamt	4381	11.5	32.3	0	793
... Auslassungen	4381	5.1	17.5	0	399
... Fehlmarkierungen	4381	6.5	17.0	0	396
letztes bearbeitetes Zeichen	4381	1243.6	220.8	559	1600

Tabelle 26: Statistiken für Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“

Punktwert 0-20	von Rohwert	bis Rohwert
0	Minimum	179
1	180	190
2	191	201
3	202	213
4	214	224
5	225	235
6	236	246
7	247	257
8	258	269
9	270	280
10	281	291
11	292	302
12	303	314
13	315	325
14	326	336
15	337	347
16	348	358
17	359	370
18	371	381
19	382	392
20	393	400

Tabelle 27: Umrechnung Rohwert in Punktwert „Konzentriertes und sorgfältiges Arbeiten“ 2008, AT

		Fehler	Auslassungen	Fehl-markierungen
Richtige	Korrelation nach Pearson	-0.19(**)	-0.21(**)	-0.13(**)
	Signifikanz (2-seitig)	0.000	0.000	0.000
Fehler Gesamt	Korrelation nach Pearson		0.94(**)	0.93(**)
	Signifikanz (2-seitig)		0.000	0.000
Auslassungen	Korrelation nach Pearson	0.94(**)		0.75(**)
	Signifikanz (2-seitig)	0.000		0.000

Tabelle 28: Korrelationen der Parameter im Konzentrationstest; **Die Korrelation ist auf dem Niveau von 0.01(2-seitig) signifikant, n = 4381

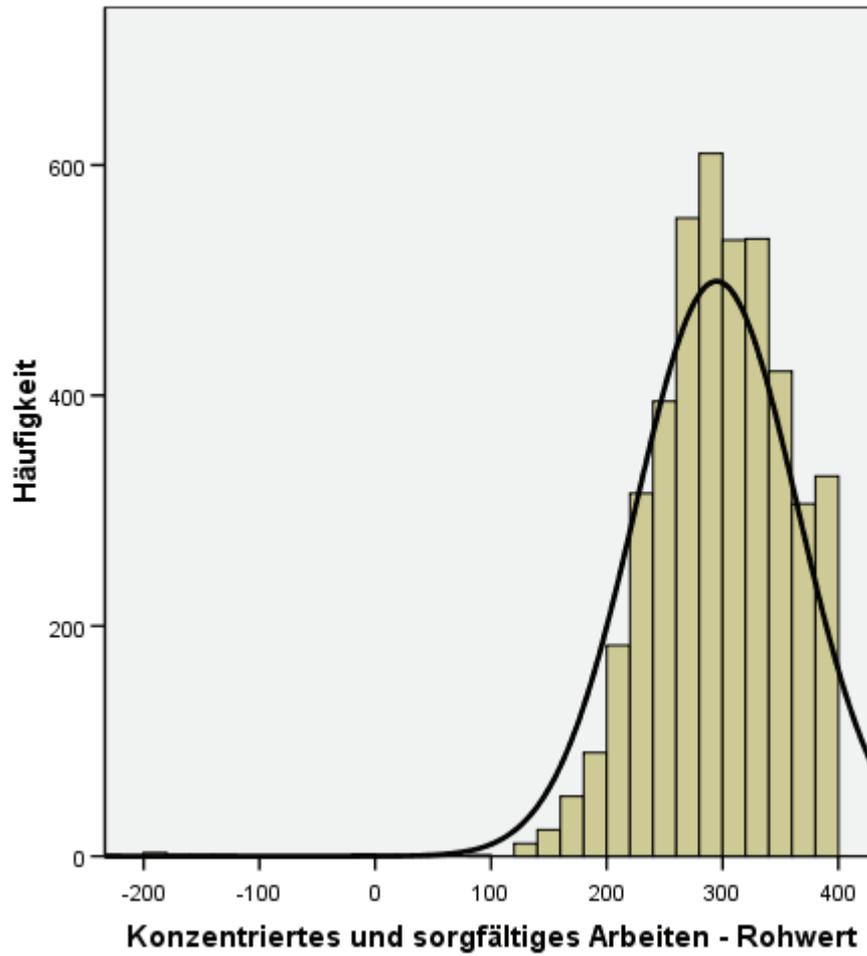


Abbildung 26: Verteilung des Rohwertes für „Konzentriertes und sorgfältiges Arbeiten 2008“, negative Werte nicht Null gesetzt. 8 Ausreisser liegen unter -200 (nicht dargestellt).

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 Gruppen. 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 Gruppen vergleichbar sind.

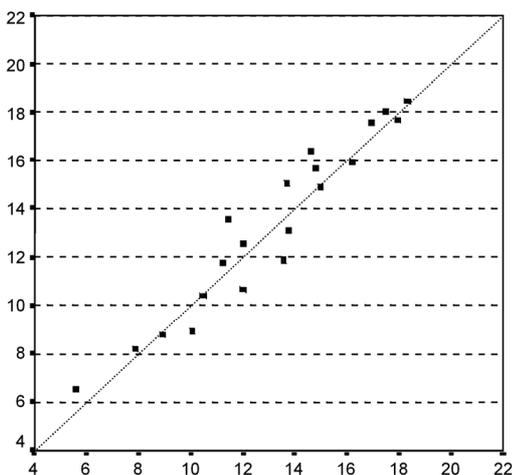


Abbildung 27: Beispiel für einen Delta-Plot.

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 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

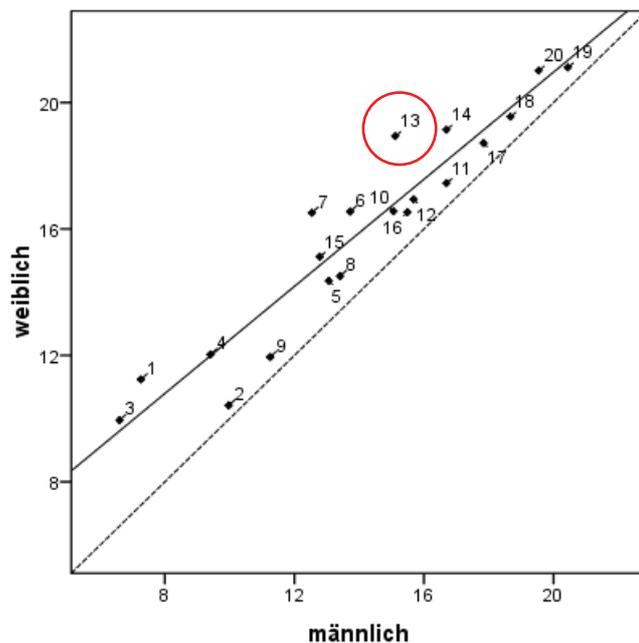
$$D_i = \frac{AX_i - Y_i + B}{\sqrt{A^2 + 1}}$$

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. Es werden jene Items als auffällig betrachtet, **welche mindestens 1.5 Punkte** (Longford, Holland & Thayer, 1993) von der Regressionsgeraden in entgegengesetzter Richtung zur Diagonalen abweichen.

6.1 DIF nach Geschlecht (AT, Humanmedizin)

Für den Gendervergleich bezüglich allfälliger testbedingter Unterschiede auf Itemebene weisen insgesamt nur drei Items von 178 auffällige DIF-Werte auf. Sie sind in den Untertests „Quantitative und formale Probleme“, „Planen und Organisieren“ sowie „Schlauchfiguren“ zu finden. Insbesondere das letztgenannte Ergebnis zeigt, dass DIF wohl Hinweise zu inhaltlich bedingten Unterschieden geben, jedoch nicht ausschliesslich dadurch erklärt werden kann. Beim Untertest „Schlauchfiguren“ handelt es sich um „sinnfreies“ Material. Es finden sich damit keine Hinweise, dass bei einzelnen Themen oder Untertests Frauen in bestimmten Items systematisch schlechtere Leistungen erreichen als Männer.

6.1.1 Quantitative und formale Probleme



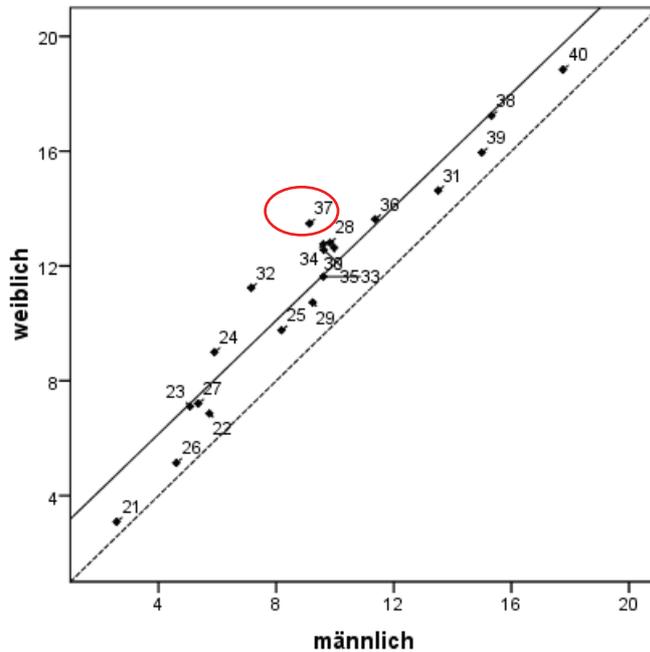
Im Untertest „quantitative und formale Probleme“ fällt ein Item (13) durch DIF auf. Der Unterschied bezüglich Itemschwierigkeit beträgt .17.

Abbildung 28: 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	0.77	0.59	7.26	11.24	-0.81	
2	0.65	0.63	9.97	10.42	1.56	
3	0.80	0.65	6.60	9.95	-0.26	
4	0.67	0.55	9.41	12.03	-0.04	
5	0.51	0.45	13.06	14.36	0.55	
6	0.48	0.35	13.72	16.55	-0.69	
7	0.53	0.35	12.54	16.51	-1.43	
8	0.49	0.44	13.41	14.51	0.66	
9	0.59	0.56	11.25	11.95	1.22	
10	0.42	0.35	15.06	16.56	0.16	
11	0.35	0.31	16.69	17.45	0.54	
12	0.39	0.33	15.68	16.94	0.28	
13	0.42	0.25	15.11	18.94	-1.62	0.17
14	0.35	0.24	16.69	19.14	-0.75	
15	0.52	0.42	12.78	15.12	-0.21	
16	0.40	0.35	15.49	16.53	0.46	
17	0.29	0.25	17.84	18.72	0.32	
18	0.26	0.22	18.67	19.55	0.22	
19	0.18	0.15	20.44	21.11	0.17	
20	0.22	0.15	19.54	21.01	-0.33	

Tabelle 29: DIF-Analyse „Quantitative und formale Probleme“ (m vs. w, AT, Humanmedizin)

6.1.2 Schlauchfiguren



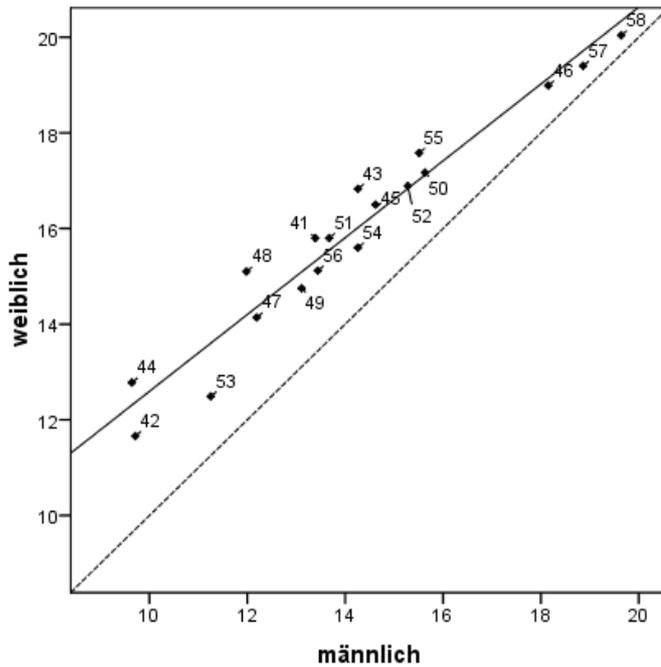
Der Untertest „Schlauchfiguren“ weist eine parallele Verschiebung der Itemschwierigkeiten auf, wobei Item 37 in den kritischen DIF-Bereich fällt. Da es sich bei diesem Untertest um abstraktes, „sinnfreies“ Material handelt, scheinen keine inhaltlichen Gründe zu diesem Ergebnis beizutragen. Bei diesem Untertest ist 2008 in der Quote AT die grösste Differenz zwischen den Geschlechtern festzustellen.

Abbildung 29: Delta-Plot für „Schlauchfiguren“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
21	0.98	0.95	2.57	3.09	1.17	
22	0.84	0.79	5.73	6.86	0.71	
23	0.87	0.77	5.07	7.11	0.07	
24	0.83	0.69	5.90	9.00	-0.69	
25	0.73	0.66	8.18	9.76	0.37	
26	0.89	0.86	4.60	5.14	1.14	
27	0.85	0.77	5.35	7.21	0.20	
28	0.65	0.52	9.83	12.81	-0.64	
29	0.68	0.61	9.24	10.73	0.43	
30	0.65	0.53	9.97	12.64	-0.42	
31	0.49	0.44	13.51	14.63	0.65	
32	0.77	0.59	7.15	11.24	-1.40	
33	0.66	0.57	9.60	11.63	0.04	
34	0.66	0.52	9.60	12.76	-0.77	
35	0.66	0.53	9.62	12.56	-0.61	
36	0.58	0.48	11.36	13.63	-0.14	
37	0.68	0.49	9.13	13.48	-1.61	0.19
38	0.41	0.32	15.32	17.23	0.08	
39	0.42	0.38	14.99	15.95	0.76	
40	0.30	0.25	17.75	18.84	0.64	

Tabelle 30: DIF-Analyse „Schlauchfiguren“ (m vs. w, AT, Humanmedizin)

6.1.3 Textverständnis



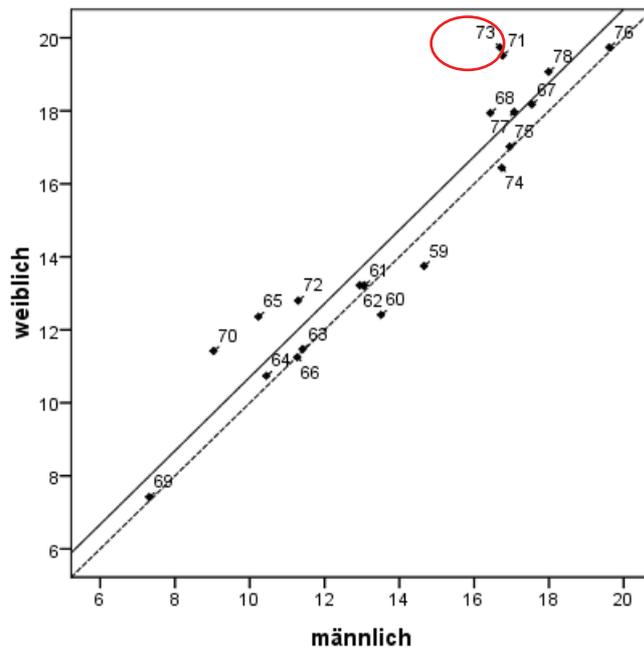
Im Untertest „Textverständnis“ ist kein DIF-Item zu registrieren. Die Lösungswahrscheinlichkeiten für die beiden Geschlechter gleichen sich mit zunehmender Schwierigkeit der Aufgaben an.

Abbildung 30: Delta-Plot für „Textverständnis“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
41	0.49	0.39	13.39	15.80	-0.37	
42	0.66	0.57	9.71	11.66	0.55	
43	0.45	0.34	14.26	16.83	-0.64	
44	0.66	0.52	9.64	12.78	-0.37	
45	0.44	0.35	14.62	16.50	-0.15	
46	0.28	0.24	18.15	18.99	0.12	
47	0.55	0.46	12.19	14.14	0.17	
48	0.56	0.42	11.98	15.10	-0.72	
49	0.51	0.43	13.11	14.75	0.27	
50	0.39	0.32	15.63	17.17	-0.04	
51	0.48	0.39	13.67	15.80	-0.20	
52	0.41	0.34	15.28	16.89	-0.04	
53	0.59	0.53	11.25	12.49	0.86	
54	0.45	0.39	14.26	15.60	0.33	
55	0.40	0.31	15.51	17.58	-0.43	
56	0.49	0.42	13.44	15.12	0.18	
57	0.25	0.22	18.86	19.40	0.25	
58	0.21	0.20	19.64	20.04	0.23	

Tabelle 31: DIF-Analyse „Textverständnis“ (m vs. w, AT, Humanmedizin)

6.1.4 Planen & Organisieren



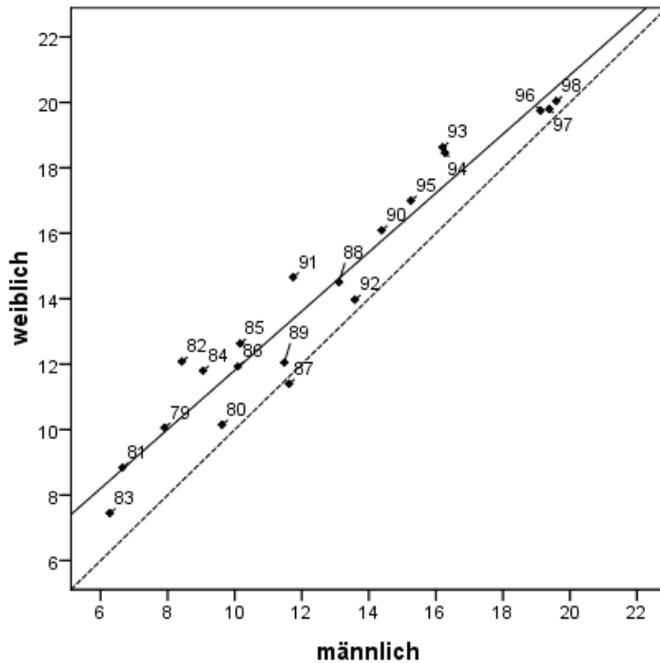
Der Untertest „Planen und Organisieren“ weist eine geringe Parallelverschiebung auf. Ein Item (73) liegt über der kritischen Grenze und weist somit DIF auf. Allerdings fallen auch einige Aufgaben durch eine für Frauen geringere Schwierigkeit auf (59, 60, 74).

Abbildung 31: Delta-Plot für „Planen & Organisieren“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
59	0.44	0.48	14.66	13.75	1.17	
60	0.49	0.54	13.51	12.41	1.29	
61	0.51	0.50	13.06	13.22	0.40	
62	0.51	0.50	12.94	13.22	0.32	
63	0.58	0.58	11.41	11.47	0.46	
64	0.62	0.61	10.44	10.74	0.29	
65	0.63	0.54	10.23	12.36	-1.00	
66	0.59	0.59	11.27	11.25	0.51	
67	0.31	0.28	17.54	18.18	0.09	
68	0.36	0.29	16.43	17.94	-0.53	
69	0.77	0.76	7.31	7.42	0.40	
70	0.69	0.58	9.03	11.42	-1.19	
71	0.34	0.22	16.76	19.52	-1.41	
72	0.59	0.52	11.29	12.80	-0.56	
73	0.35	0.21	16.69	19.74	-1.61	0.14
74	0.34	0.36	16.74	16.44	0.74	
75	0.33	0.33	16.95	17.02	0.48	
76	0.21	0.21	19.61	19.74	0.46	
77	0.33	0.29	17.07	17.97	-0.10	
78	0.29	0.24	17.99	19.07	-0.22	

Tabelle 32: DIF-Analyse „Planen & Organisieren“ (m vs. w, AT, Humanmedizin)

6.1.5 Medizinisch-naturwissenschaftliches Grundverständnis



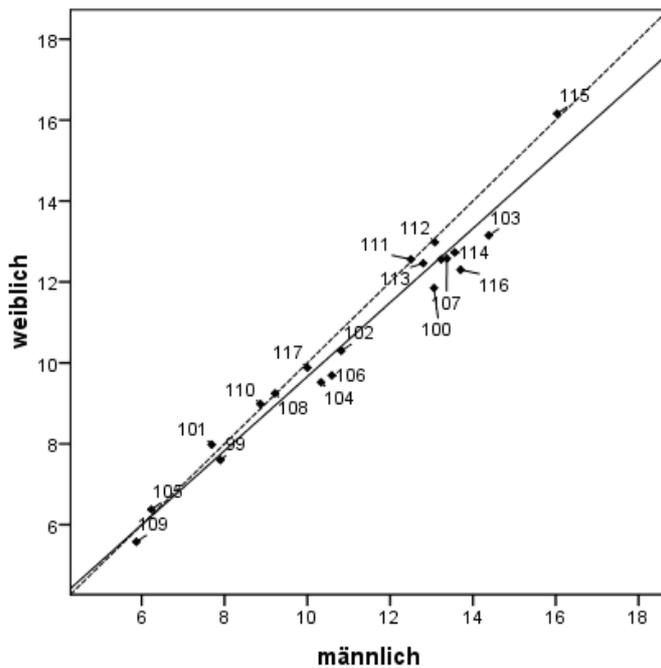
Im Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“ ist kein DIF-Item zu registrieren. Die Lösungswahrscheinlichkeiten für die beiden Geschlechter gleichen sich mit zunehmender Schwierigkeit der Aufgaben an.

Abbildung 32: 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	0.74	0.64	7.90	10.06	-0.11	
80	0.66	0.64	9.62	10.15	0.98	
81	0.79	0.70	6.65	8.84	-0.04	
82	0.72	0.55	8.42	12.08	-1.26	
83	0.81	0.76	6.27	7.45	0.74	
84	0.69	0.56	9.05	11.80	-0.62	
85	0.64	0.53	10.16	12.63	-0.49	
86	0.64	0.56	10.09	11.93	-0.03	
87	0.57	0.58	11.62	11.40	1.39	
88	0.51	0.44	13.11	14.51	0.08	
89	0.58	0.55	11.48	12.05	0.82	
90	0.45	0.37	14.38	16.09	-0.24	
91	0.57	0.44	11.74	14.66	-0.95	
92	0.48	0.47	13.58	13.97	0.80	
93	0.37	0.26	16.20	18.63	-0.91	
94	0.36	0.27	16.27	18.46	-0.74	
95	0.41	0.33	15.25	17.00	-0.33	
96	0.24	0.21	19.12	19.75	0.22	
97	0.23	0.21	19.38	19.80	0.35	
98	0.22	0.20	19.59	20.04	0.32	

Tabelle 33: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“ (m vs. w, AT, Humanmedizin)

6.1.6 Figuren lernen



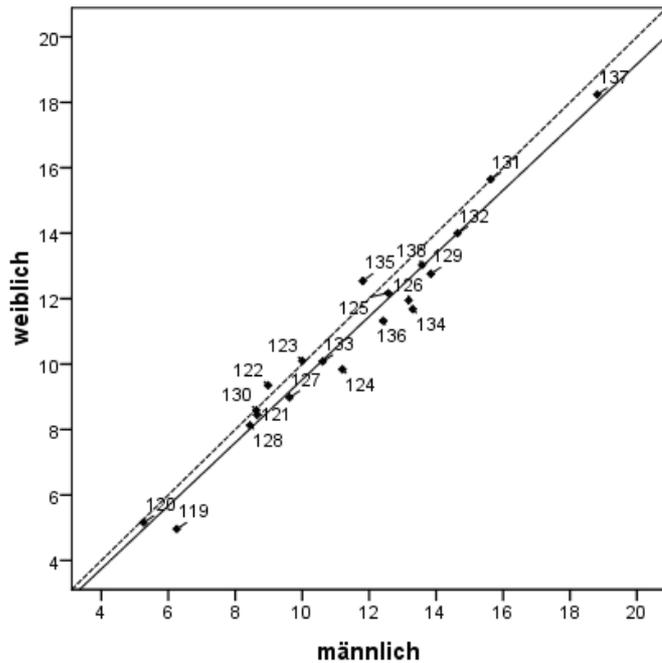
Im Untertest „Figuren lernen“ erzielen die Kandidatinnen bessere Ergebnisse als die österreichischen Kandidaten.

Abbildung 33: Delta-Plot für „Figuren lernen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
99	0.74	0.75	7.90	7.60	0.10	
100	0.51	0.56	13.06	11.85	0.46	
101	0.75	0.74	7.69	7.98	-0.32	
102	0.61	0.63	10.82	10.30	0.08	
103	0.45	0.50	14.38	13.15	0.38	
104	0.63	0.67	10.33	9.52	0.33	
105	0.81	0.81	6.23	6.38	-0.13	
106	0.62	0.66	10.59	9.69	0.38	
107	0.49	0.53	13.37	12.58	0.12	
108	0.68	0.68	9.22	9.25	-0.22	
109	0.83	0.84	5.87	5.58	0.22	
110	0.70	0.69	8.87	8.98	-0.26	
111	0.53	0.53	12.50	12.56	-0.45	
112	0.51	0.51	13.08	12.98	-0.37	
113	0.52	0.53	12.80	12.46	-0.17	
114	0.49	0.52	13.56	12.73	0.14	
115	0.38	0.37	16.03	16.16	-0.72	
116	0.48	0.54	13.70	12.30	0.55	
117	0.64	0.65	10.00	9.88	-0.16	
118	0.50	0.53	13.23	12.56	0.04	

Tabelle 34: DIF-Analyse „Figuren lernen“ (m vs. w, AT, Humanmedizin)

6.1.7 Fakten lernen



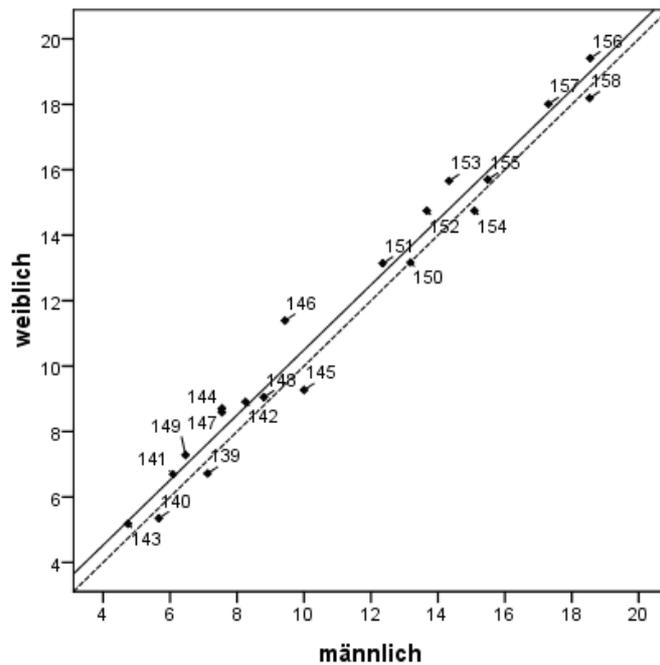
Im Untertest „Fakten lernen“ erzielen die Kandidatinnen bessere Ergebnisse als die österreichischen Kandidaten.

Abbildung 34: Delta-Plot für „Fakten lernen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ-Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
119	0.81	0.87	6.25	4.96	0.69	
120	0.86	0.86	5.26	5.16	-0.15	
121	0.70	0.71	8.65	8.45	-0.16	
122	0.69	0.67	8.98	9.35	-0.58	
123	0.64	0.64	10.00	10.10	-0.41	
124	0.59	0.65	11.20	9.84	0.60	
125	0.53	0.55	12.57	12.17	-0.12	
126	0.50	0.56	13.18	11.95	0.46	
127	0.66	0.69	9.62	8.98	0.13	
128	0.71	0.73	8.44	8.13	-0.08	
129	0.47	0.52	13.84	12.76	0.34	
130	0.71	0.71	8.63	8.59	-0.28	
131	0.39	0.39	15.63	15.65	-0.50	
132	0.44	0.47	14.64	14.00	0.00	
133	0.62	0.64	10.61	10.08	0.02	
134	0.50	0.57	13.30	11.68	0.74	
135	0.56	0.53	11.81	12.54	-0.91	
136	0.54	0.59	12.42	11.32	0.39	
137	0.25	0.28	18.81	18.24	-0.16	
138	0.48	0.51	13.58	13.03	-0.04	

Tabelle 35: DIF-Analyse „Fakten lernen“ (m vs. w, AT, Humanmedizin)

6.1.8 Muster zuordnen



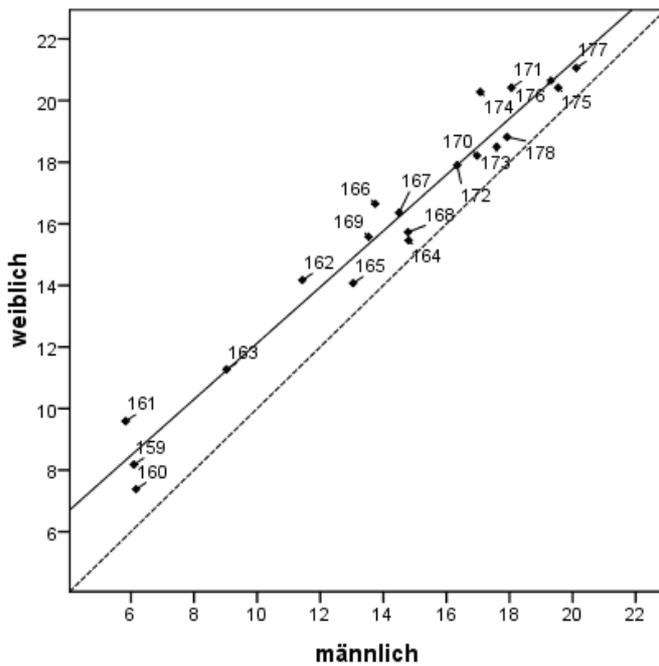
Der Untertest „Muster zuordnen“ weist eine leichte Parallelverschiebung hinsichtlich der Itemschwierigkeiten auf. Kein Item liegt im auffälligen DIF-Bereich.

Abbildung 35: Delta-Plot für „Muster zuordnen“ (m vs. w, AT, Humanmedizin)

Item	Schwierigkeit		Δ -Wert		Distanz	Differenz p falls DIF
	m	w	m	w		
139	0.77	0.79	7.12	6.72	0.65	
140	0.84	0.85	5.66	5.35	0.59	
141	0.82	0.79	6.09	6.69	-0.06	
142	0.72	0.69	8.25	8.91	-0.11	
143	0.88	0.86	4.74	5.18	0.06	
144	0.75	0.71	7.55	8.59	-0.38	
145	0.64	0.68	10.00	9.27	0.87	
146	0.67	0.58	9.43	11.39	-1.03	
147	0.75	0.70	7.55	8.71	-0.46	
148	0.70	0.69	8.80	9.05	0.18	
149	0.80	0.77	6.46	7.28	-0.22	
150	0.50	0.50	13.18	13.17	0.34	
151	0.54	0.50	12.35	13.14	-0.21	
152	0.48	0.43	13.67	14.75	-0.43	
153	0.45	0.39	14.33	15.66	-0.61	
154	0.42	0.43	15.09	14.75	0.57	
155	0.40	0.39	15.49	15.70	0.18	
156	0.26	0.22	18.55	19.41	-0.30	
157	0.32	0.29	17.30	18.01	-0.18	
158	0.26	0.28	18.53	18.19	0.55	

Tabelle 36: DIF-Analyse „Muster zuordnen“ (m vs. w, AT, Humanmedizin)

6.1.9 Diagramme und Tabellen



Insgesamt ist im Untertest „Diagramme und Tabellen“ eine recht deutliche Verschiebung der Itemschwierigkeiten zu beobachten, allerdings ohne, dass ein einzelnes Item auffällige, auf DIF hinweisende Werte aufweisen würde.

Abbildung 36: 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	0.82	0.73	6.09	8.18	0.28	
160	0.82	0.76	6.16	7.38	0.91	
161	0.83	0.66	5.83	9.59	-0.94	
162	0.58	0.46	11.43	14.17	-0.55	
163	0.69	0.59	9.03	11.27	-0.02	
164	0.43	0.40	14.80	15.46	0.77	
165	0.51	0.46	13.04	14.07	0.61	
166	0.48	0.35	13.74	16.65	-0.82	
167	0.44	0.36	14.50	16.36	-0.10	
168	0.43	0.39	14.78	15.73	0.55	
169	0.49	0.40	13.53	15.58	-0.17	
170	0.33	0.28	16.97	18.21	0.20	
171	0.28	0.18	18.06	20.42	-0.70	
172	0.36	0.29	16.34	17.90	0.00	
173	0.31	0.26	17.59	18.50	0.40	
174	0.33	0.19	17.07	20.28	-1.27	
175	0.22	0.18	19.54	20.42	0.30	
176	0.23	0.17	19.31	20.65	-0.04	
177	0.19	0.15	20.11	21.06	0.20	
178	0.29	0.25	17.92	18.82	0.38	

Tabelle 37: 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?

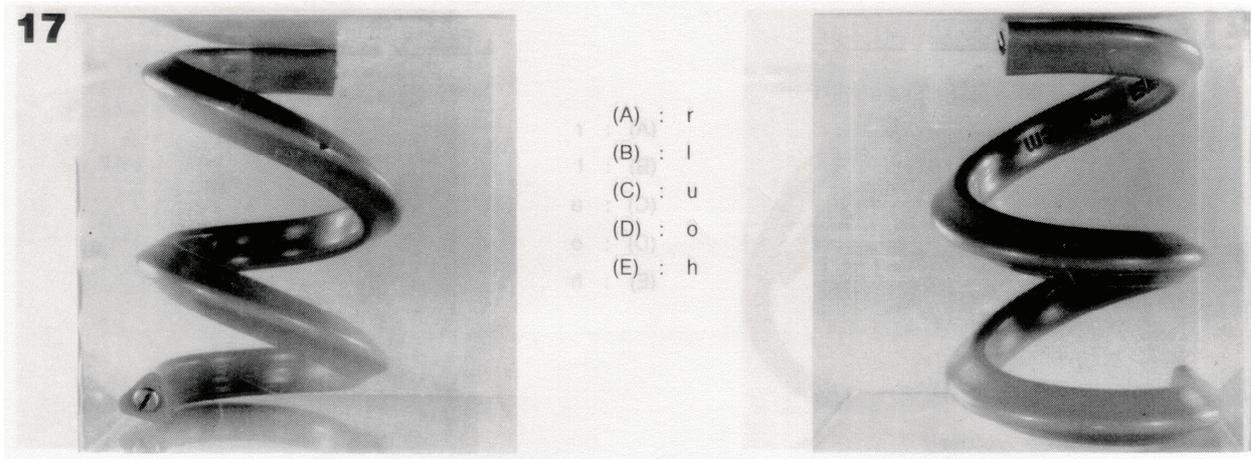
- (A) 60 BE
- (B) 25 BE
- (C) 6 BE
- (D) 5 BE
- (E) 0,5 BE

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 Jodatombau 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:

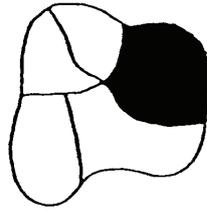
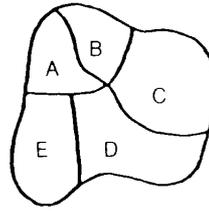
- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind 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 eingepägt 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...

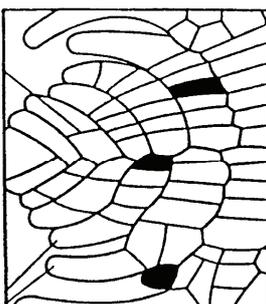
- (A) Installateur
- (B) Lehrer
- (C) Dachdecker
- (D) Handelsvertreter
- (E) Physiker

7.9 Untertest: Muster zuordnen

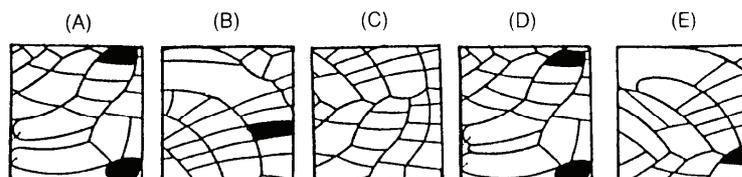
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wieder zu erkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner 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.
- Hänsgen 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>
- Hänsgen 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>
- Hänsgen, 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.
- Hänsgen, 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.
- Hänsgen, 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.
- Hänsgen, 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 Friborgo (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/

**Eignungstest Medizinstudium
EMS-AT**

suchen...

- Allgemeine Informationen
 - » Quotenregelung
- Ablauf und Termine
 - » Voranmeldung
 - » Persönliche Anmeldung
 - » Testvorbereitung
 - » Testteilnahme
 - » Ergebnisfeststellung
 - » Ergebnisabfrage
- Weitere Informationen / Links
- Kontakt
- Impressum

EMS-AT

INFORMATION-S-HOMEPAGE FÜR STUDIENWERBERINNEN, DIE IM STUDIENJAHR 2009/10 MIT DEM STUDIUM HUMAN- ODER ZAHNMEDIZIN AN DEN MEDIZINISCHEN UNIVERSITÄTEN WIEN ODER INNSBRUCK BEGINNEN WOLLEN

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, werden die Medizinischen Universitäten in Wien und Innsbruck auf Basis der vom Nationalrat erlassenen Novelle zum Universitätsgesetz 2002 gemeinsam eine kapazitätsorientierte Studienplatzvergabe für ALLE StudienwerberInnen durchführen. Folgende Platzzahlen werden für die Studien "Humanmedizin" und "Zahnmedizin" verfügbar sein:

	Humanmedizin	Zahnmedizin	Gesamt
Med. Univ. Wien (MUW)	660	80	740
Med. Univ. Innsbruck (MUI)	360	40	400
Gesamt	1020	120	1140

Die Vergabe der Plätze erfolgt mittels eines erprobten und wissenschaftlich abgesicherten Eignungstests (EMS), der in Deutschland entwickelt, in der Schweiz weiterentwickelt und seit 1998 angewendet wird. Der Eignungstest liefert einen Testwert, welcher nachweislich hoch mit der Studieneignung korreliert. Dieser wird für die Zulassung verwendet und bildet das Zulassungskriterium.

Rechtliche Grundlage für das Aufnahmeverfahren sind der §124b UG2002 und die in den Mitteilungsblättern der [Medizinischen Universität Wien](#) und der [Medizinischen Universität Innsbruck](#) veröffentlichten Verordnungen in der jeweils gültigen Fassung.