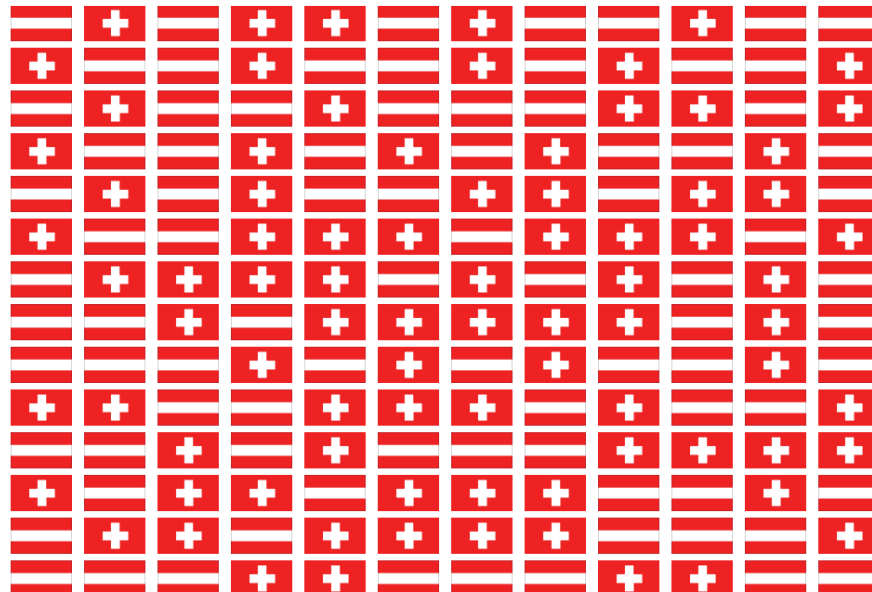


EMS

Eignungstest für das Medizinstudium in Österreich

Medizinische Universität Innsbruck
Medizinische Universität Wien
2006



Berichterstatter

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

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

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

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

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

Information Zulassungsverfahren: www.eignungstest-medizin.at

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

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

Inhaltsverzeichnis

1	EINLEITUNG	4
2	EIGNUNGSTEST UND NUMERUS CLAUSUS	5
3	BESCHREIBUNG DES EIGNUNGSTESTS.....	8
3.1	Aufbau des Tests	8
3.2	Berechnung der Werte	10
3.3	Mittlerer Rangplatz der Untertests	11
4	TESTANWENDUNG IN ÖSTERREICH 2006	12
4.1	Statistische Angaben zur Teilnahme	12
4.2	Verteilungsprüfung der einzelnen Punktwerte	15
5	ERGEBNISSE ZUR TESTGÜTE.....	18
5.1	Zuverlässigkeit.....	18
5.2	Faktorielle Validität.....	18
5.3	Item-Trennschärfen	23
5.4	Konzentriertes und sorgfältiges Arbeiten	28
6	FAIRNESS UND DIE VERGLEICHBARKEIT VON GRUPPEN	31
6.1	Vergleich der Geschlechter	33
6.2	Vergleichbarkeit der Testorte	51
6.3	Vergleich nach Maturitätsländerquote	53
6.4	Vergleiche für Altersgruppen nach Geburtsjahr	54
6.5	Vergleich nach Disziplin	56
7	VORBEREITUNG AUF DEN EMS	58
8	BEISPIELAUFGABEN FÜR DIE UNTERTESTS.....	62
8.1	Untertest: Quantitative und formale Probleme	62
8.2	Untertest: Schlauchfiguren.....	62
8.3	Untertest: Textverständnis	63
8.4	Untertest: Planen und Organisieren	64
8.5	Untertest: Konzentriertes und Sorgfältiges Arbeiten.....	66
8.6	Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	66
8.7	Untertest: Figuren lernen.....	67
8.8	Untertest: Fakten lernen	68
8.9	Untertest: Muster zuordnen	68
8.10	Untertest: Diagramme und Tabellen	69
9	WEITERE INFORMATIONEN ZU DEN BEWERBUNGEN	70
10	LITERATUR	74
10.1	Originaltest zur Information und Vorbereitung.....	75
10.2	Information im Internet.....	75

1 Einleitung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2006 für das Medizinstudium und die Zulassung zum Studium der Human- und Zahnmedizin ab Wintersemester 2006/2007 an den Medizinischen Universitäten Innsbruck und Wien. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen in der Schweiz herangezogen. **Die Ergebnisse sind nicht repräsentativ für Schlussfolgerungen hinsichtlich einer gesamthaften bzw. regionalen Evaluation von Bildungseinrichtungen und -wegen.**

Im Jahr 2006 wurde nach nur achtmonatiger Vorbereitungszeit an den Medizinischen Universitäten Wien und Innsbruck ein Eignungstest für das Medizinstudium als Kriterium für die Zulassung eingesetzt.

In Innsbruck hatten sich 2182 Personen auf 400 Studienplätze angemeldet, in Wien 3429 Personen auf 740 Studienplätze. Es war die Bereitstellung der Testlogistik für 5611 Personen notwendig.

Erfolgreich absolviert haben den Test in Österreich 3645 Personen (1415 in Innsbruck, 2230 in Wien). Die hohe Rückzugsrate von 35% ist u.a. darauf zurückzuführen, dass parallel das Aufnahmeverfahren für die Medizinische Universität Graz stattfand (Doppelanmeldungen) und dass die Anmeldung nicht wie in der Schweiz mit einer Kostenbeteiligung der Kandidaten verbunden war.

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

Allgemeine Abkürzungen (ohne statistische Prüfgrößen)

EU: Europäische Union (meint die Länderquote der EU)

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

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

HM: Humanmedizin

ZM: Zahnmedizin

m: Mittelwert

s: Standardabweichung

n: Personenanzahl

2 Eignungstest und Numerus Clausus

Nachdem auch in Österreich eine Zulassungsbegrenzung für das Fach Medizin eingeführt worden ist, erlaubt kein europäisches Industrieland mehr den Zugang zum Medizinstudium unbegrenzt. In der Studie von Ebach und Trost (1997) verblieben Österreich und Belgien als Länder mit freiem Zugang. Auch Belgien hat eine entsprechende Begrenzung einführen müssen. Die Attraktivität eines Medizinstudiums nimmt überall zu, kein europäisches Land scheint mehr in der Lage, die notwendigen Mittel für das vergleichsweise teure Medizinstudium unbegrenzt aufwenden zu können, um einen freien Hochschulzugang zu erhalten.

Wenn eine Zulassungsbeschränkung notwendig ist, bedarf es eines **fairen, rechtsgleichen und willkürfreien Kriteriums**, nach dem diese Zulassung erfolgt. Üblicherweise ist es die **Eignung zum Studium**.

Dieser Bezug ist dadurch gerechtfertigt, dass bei begrenzten Ressourcen diejenigen bevorzugt einen Studienplatz erhalten sollen, die beste Aussichten für einen erfolgreichen Abschluss in einer vertretbaren Studienzeit bieten und die Ressourcen dann auch für andere eher wieder freigeben.

Fairness muss dabei für den **Einzelfall** gelten: **Gleiche Eignung für das Studium muss zu gleichen Testwerten und damit gleichen Zulassungswahrscheinlichkeiten führen**. Dies ist gewährleistet, wenn gleiche Testwerte auch die gleiche Studienerfolgsprognose (Studiendauer, Noten) bedeuten. Für den TMS/EMS in Deutschland und der Schweiz ist dies mehrfach untersucht und bestätigt worden. Es darf dabei allerdings keine systematischen Benachteiligungen für bestimmte Gruppen geben, welche den erzielten Testwert und damit die Studieneignung unter- oder überschätzen würden.

Letzteres wäre beispielsweise der Fall wenn mathematisch-naturwissenschaftliches **Schulwissen** Gegenstand des Tests wäre. Hier hätten Personen einen Vorteil, die durch frühere Berufsausbildungen oder exzessive Trainingskurse einen Wissensvorsprung gegenüber anderen Personen haben. Wesentlich für die Fairness des EMS zur Erfassung der Studieneignung ist, dass kein spezielles Fachwissen abgefragt wird. Erfasst werden soll die Studierfähigkeit in dem Sinne, wie gut man sich **später im Studium** neues Wissen aneignen kann. So sind die "Spieße für alle gleich lang", wie man in der Schweiz sagt. Die Maturitäts- bzw. Abiturprüfung, an welcher bewiesen wurde, dass man über das notwendige Wissen verfügt, soll durch den EMS nicht abgewertet werden.

Die Aufgaben des EMS sind so konstruiert, dass notwendiges Fachwissen in der Aufgabe selbst mitgeteilt wird. Es kommt darauf an, eine Problemstellung zu erkennen, genau zu analysieren und aufgrund des mitgeteilten Wissens zu lösen. Dadurch ist auch die Trainierbarkeit des Tests über das empfohlene Vorgehen anhand des veröffentlichten Trainingsmaterials hinaus begrenzt, was die soziale Verträglichkeit gewährleistet.

Die ebenfalls häufig gewünschte Berufseignung („guter Arzt/gute Ärztin“) ist davon nicht unabhängig, hätte aber bei der **expliziten** Verwendung als Zulassungskriterium drei Probleme:

- Sie lässt sich wesentlich schwerer definieren und nicht mit ausreichender Zuverlässigkeit messen, sie ist für die verschiedenen Gebiete der Medizin nicht vergleichbar und alle heute bekannten Testmethoden sind hinsichtlich Trainierbarkeit oder Verfälschbarkeit in sozial erwünschter Richtung nicht ausreichend objektiv.
- Wichtige Grundlagen der Berufseignung werden erst im Studium gelegt und ausdifferenziert – es ist unklar, welche Voraussetzungen überhaupt schon vor einem Studium vorhanden sein können und sollten. Ggf. würden Personen mit Berufserfahrung aus medizinnahen Berufen hierdurch bevorteilt.
- Es scheitern vor allem Personen im Studium, indem sie die Prüfungen nicht bestehen. Insofern kann begründbar nur die Berechtigung abgeleitet werden, diesen Studienerfolg als Zulassungskriterium zu verwenden.

In der Schweiz kann man die letztgenannte Problematik am besten mit Zahlen verdeutlichen. Bekanntlich ist der NC in der Deutschschweiz seit 1998 für Humanmedizin eingeführt, in der französischsprachigen Schweiz gibt es bisher keinen NC. Bis zum Jahr 2002 war die Zahl der effektiv Abgewiesenen durch einen NC wegen der Rückzüge noch sehr gering, 90% der Bewerber erhielten ein Studienplatzangebot und die kapazitätsregulierende Wirkung des NC war am ehesten durch Dissuasion erklärbar.

Nach letzten Zahlen der Rektorenkonferenz der Schweizer Universitäten (CRUS) stieg in den **Universitäten mit NC** (Basel, Bern, Fribourg, Zürich) trotzdem die Erfolgsquote in Humanmedizin für die erste Vorprüfung von ca. 70% für Anfängerjahrgänge vor 1998 bis ca. 85% für den Anfängerjahrgang 2002, für die zweite Vorprüfung von ca. 65% auf ca. 75%. In den **Universitäten ohne NC** (Genf, Lausanne, Neuenburg) betrug die Erfolgsquote der ersten Vorprüfung vor 1998 ca. 55% – also schon niedriger als in den anderen Universitäten. Für den Anfängerjahrgang 2002 liegt sie immer noch bei 55%. Für die zweite Vorprüfung betrug sie 1998 50%, nach Schwankungen zwischen 40% und 55% für den Anfängerjahrgang 2001 45%. Den Universitäten mit NC gelingt es also summa summarum recht gut, durch die geringere Überlastung auch mehr Studierende zum Erfolg zu führen. Die hohe Zahl der Abbrecher bei den Universitäten ohne NC bleibt auch daher bedenklich, weil fast die Hälfte der Personen ihr Ziel nicht erreicht und dafür ein bis mehrere Lebensjahre investieren muss. Dass die ohnehin begrenzten Ausbildungskapazitäten dann ineffektiv eingesetzt worden sind, bleibt auch aus bildungsökonomischer Sicht bedenklich.

Die Information über die Studien- und Berufsanforderungen, die realistische Auseinandersetzung mit der Motivation zum Medizinberuf bleibt weiter wichtig – hier belegen Befragungsergebnisse aber, dass der heutige Stand der Informiertheit zumindest subjektiv als gar nicht so schlecht eingeschätzt wird.

Für das Zulassungskriterium bestehen hohe Anforderungen, um über persönliche Schicksale so fair wie möglich – nach Abwägung aller Interessen – zu entscheiden. Die „Eignung des Eignungstests“, ein solches Kriterium zu sein, kann man wie folgt zusammenfassen:

- Rechtliche Begründung der Ergebnisse durch Chancengleichheit und Willkürfreiheit (Forderung des Schweizer Bundesgerichtes für einen NC), indem alle Abläufe standardisiert sind und eine laufende Evaluation erfolgt;
- Prognose der Studieneignung (gesetzliche Grundlage des NC) ist ausreichend evaluiert, Studiendauer und Noten werden nachweislich gut vorhergesagt;
- Jährlich eine neue Testversion, die vorher nicht bekannt ist (keine Vorteile für irgendwelche Insider, was die Rechtsfähigkeit einschränken würde), dadurch sind auch wiederholte Antritte problemlos möglich;
- Prinzipiell wird kein spezielles Fachwissen vorausgesetzt, das empfohlene Informationskonzept und die offiziell angebotenen Vorbereitungsmöglichkeiten sind nachweislich genauso effektiv wie der Besuch kommerzieller Vorbereitungskurse (ist ebenfalls ausreichend evaluiert) – als Voraussetzung für die soziale Verträglichkeit.

Eine Metaanalyse zu eignungsdiagnostischen Verfahren im Rahmen der Studienzulassung konnte nachweisen, dass fachspezifische Studierfähigkeitstests im Mittel die gleichen Vorhersagewerte für Studienerfolg¹ wie die „anfallenden“ Maturitätsnoten erreichen. Es gibt keine besseren Prognosekriterien.

Drei wesentliche Nachteile der Schulnoten sprechen aber gegen ihre Verwendung als Zulassungskriterium:

- Regionale Besonderheiten des Notenmaßstabes (unterschiedliche Strenge der Beurteilung in verschiedenen Bundesländern durch Definition unterschiedlicher Anforderungen oder unterschiedlicher Maturitätsquoten). Ausgleichsverfahren

¹ Hell, Trapmann, Weigand Hirn & Schuler(2005): Die Validität von Prädiktoren des Studienerfolgs – eine Metaanalyse. Siehe: <http://www.studieneignung.de>

werden hierfür z.B. in Deutschland angewendet – sie setzen aber eine ausreichende Größe der Bundesländer bzw. Kandidatenzahl für solche statistischen Ausgleichsverfahren voraus.

- Subjektive Beurteilungseffekte und –fehler bestimmter Lehrpersonen bei der Beurteilung bestimmter Maturanden im Einzelfall. Diese Effekte sind nicht erkennbar und daher auch nicht ausgleichbar.
- Rückwirkungen auf die Notengebung durch das Wissen, dass sie als Zugangskriterium verwendet werden. Auch diese Effekte sind nicht ausgleichbar.

Die DIN-Norm 33430 für eignungsdiagnostische Verfahren ist für solche Tests anwendbar. In Österreich gilt eine ÖNORM D4000 mit vergleichbarem Inhalt und eine internationale Abstimmung hat ergeben, einen Diskussionsprozess einzuleiten, diese Norm zu einem ISO-Standard auszubauen. Wichtigste Anforderung ist die ausreichende empirische Überprüfung der Vorhersagefähigkeit für Studienerfolg – eine Absage an alle Verfahren, welche am „grünen Tisch“ ausgedacht worden sind und bereits in der Konzeption problematisch sind. Die „Not“, schnell Verfahren einführen zu müssen, hat hier nicht immer eine ausreichende Qualität gefördert.

Auch in Deutschland ist schließlich die Wiedereinführung des TMS geplant. Ab 2007 soll in Baden-Württemberg und einzelnen anderen Universitäten der Test verwendet werden, weitere Bundesländer prüfen den Einsatz. Die veränderte Verantwortlichkeit – Dezentralisierung und mehr Verantwortung für die Bundesländer bzw. Universitäten selbst bei insgesamt angespannter Haushaltslage – sind Faktoren, die den Entscheidungsprozess noch beeinflussen und erschweren.

3 Beschreibung des Eignungstests

3.1 Aufbau des Tests

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

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

- müssen sie jedes Jahr die Studieneignung gleich zuverlässig messen,
- muss das Schwierigkeitsspektrum aller Aufgaben annähernd vergleichbar sein,
- darf kein spezielles Fachwissen vorausgesetzt werden, um die Trainierbarkeit des Tests gering zu halten sowie die Fairness des Test nicht zu beeinträchtigen und es
- 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 2006 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 ab 2006

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 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 kapazitäts-entsprechendes Zulassungsverfahren möglich, die Rechtsfähigkeit des Verfahrens hat sich bei mehreren Überprüfungen (auch gerichtlich) bestätigt.

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

Die Erprobung neuer Aufgaben für sechs der neun Untertests (siehe Tabelle 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 Lösungsdeutigkeit und Schwierigkeit optimal sind.

3.2 Berechnung der Werte

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

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

Alle Punkte der Untertests werden zu einer Summe addiert (**Punktwert**, vgl. Abbildung 1). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Jahre vergleichbar ist. Deshalb findet eine Standardisierung auf den Mittelwert und die Standardabweichung der jeweiligen Testform statt. Dieser **Testwert** liegt zwischen 70 und 130 (der Mittelwert ist 100).

Beim „Planen und Organisieren“ wurden 2006 nur 18 Aufgaben gewertet, weil zwei Aufgaben nicht die notwendigen Kennwerte erreichten. Die Berechnung und Verwendung von Gütekriterien für die Itemselektion ist ein wichtiges Merkmal eines Tests. Insofern bevorzugen wir intern die Strategie, im Zweifel eine Aufgabe lieber auszuschließen und die Testgüte damit ausreichend hoch zu halten. Das Standardisierungsverfahren zur Bildung des Testwertes gewährleistet auch bei unterschiedlichen Aufgabenzahlen vergleichbare Testwerte zwischen den Jahren.

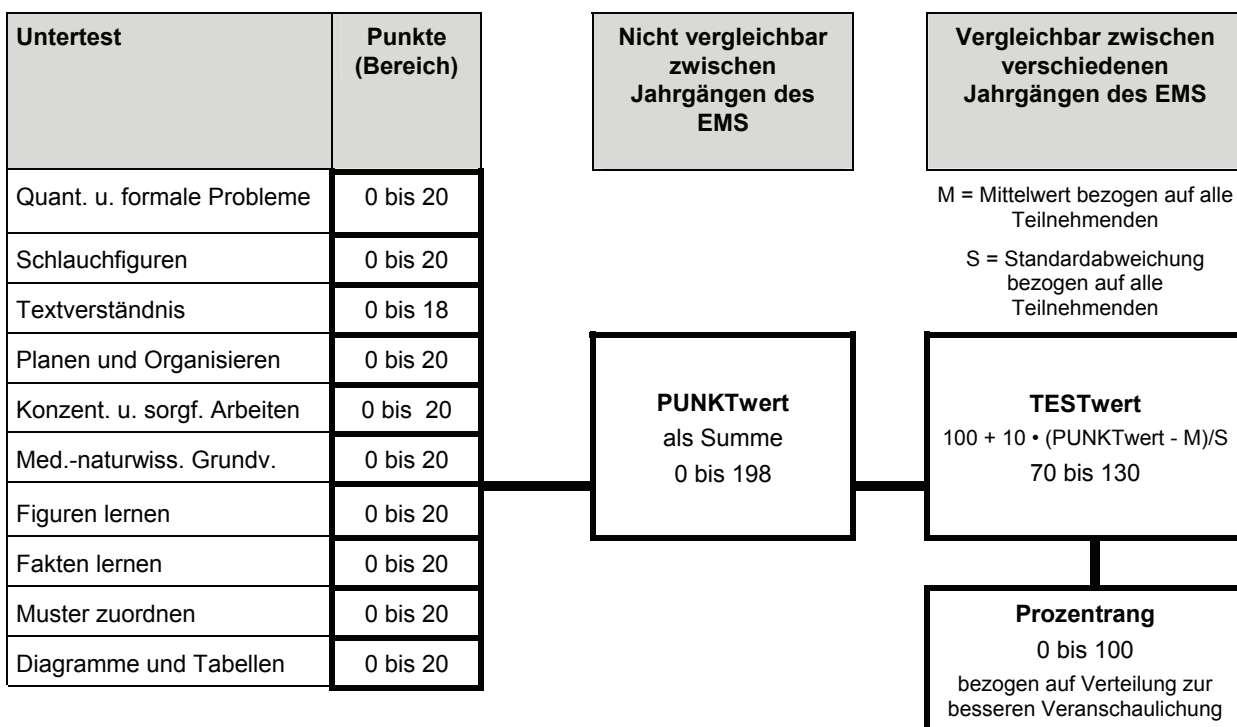


Abbildung 1: Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert (Maximalwerte 2006).

3.3 Mittlerer Rangplatz der Untertests

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

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

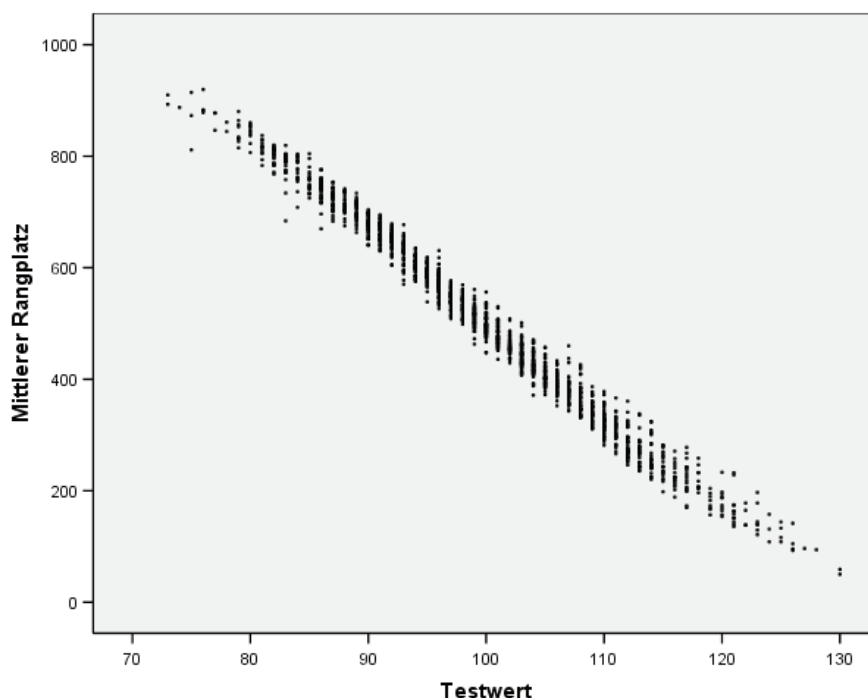
Bei **gleichem Testwert** werden die Personen deshalb in der Reihenfolge des **mittleren Rangplatzes aller Untertests** berücksichtigt und es werden nur so viele Personen zugelassen, wie freie Plätze tatsächlich vorhanden sind. In Österreich hat man sich diesem Verfahren angeschlossen.

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 2006 zwischen 50 und 920. Niedrige Werte stehen für bessere Leistungen.
- Er korreliert mit dem Testwert mit -0.991 , was praktisch für eine Äquivalenz spricht – siehe auch Abbildung 2.
- Er differenziert allerdings besser zwischen den Personen und lässt eine kapazitätsgenaue Auswahl zu.



Das seit 2005 angewandte Prinzip, bei gleichem Testwert den Rangplatz zu verwenden, kann daher auch für die Folgejahre beibehalten werden.

Abbildung 2: Beziehung zwischen Testwert und mittlerem Rangplatz für 2006.

4 Testanwendung in Österreich 2006

4.1 Statistische Angaben zur Teilnahme

				EU	nicht EU	Österreich	Gesamt	
Innsbruck	männlich	Altersklasse	bis 1982	39	1	26	66	
			1983-1985	237	1	45	283	
			nach 1985	113		197	310	
		Gesamt		389	2	268	659	
	weiblich	Altersklasse	bis 1982	41	1	21	63	
			1983-1985	193	1	33	227	
			nach 1985	204		262	466	
		Gesamt		438	2	316	756	
	Wien	männlich	Altersklasse	bis 1982	46	8	59	113
				1983-1985	189	18	105	312
nach 1985				78	13	430	521	
Gesamt			313	39	594	946		
weiblich		Altersklasse	bis 1982	39	18	56	113	
			1983-1985	189	17	76	282	
			nach 1985	159	16	714	889	
		Gesamt		387	51	846	1284	
Gesamt				1527	94	2024	3645	

Tabelle 2: Aufteilung der Teilnehmenden nach Testort, Geschlecht und Geburtsjahr

		Innsbruck	Wien	Gesamt
EU	Zahnmedizin	47	38	85
	Kombiniert	42	38	80
	Humanmedizin	738	624	1362
	Gesamt	827	700	1527
Nicht EU	Zahnmedizin	1	19	20
	Kombiniert		4	4
	Humanmedizin	3	67	70
	Gesamt	4	90	94
Österreich	Zahnmedizin	32	98	130
	Kombiniert	42	86	128
	Humanmedizin	510	1256	1766
	Gesamt	584	1440	2024
Gesamt		1415	2230	3645

Tabelle 3: Zulassungsquoten und Disziplinwahl

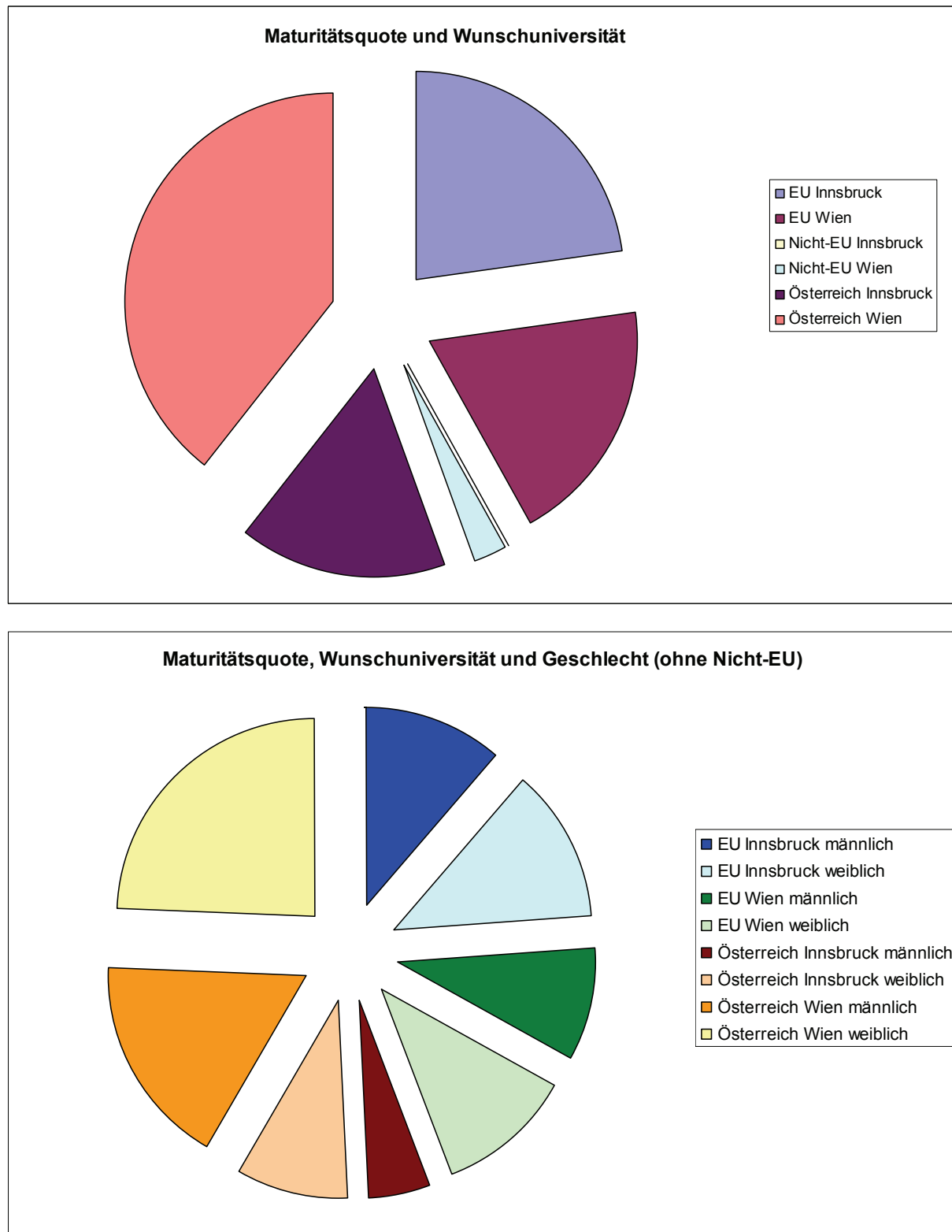


Abbildung 3: Grafische Veranschaulichung der Zusammensetzung der Bewerbungen

	Maturitätsjahr (Jahr des Abschlusses)																															
Geburtsjahr	1964	1965	1969	1974	1980	1981	1982	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Gesamt	
1944	1																												1	2		
1945			1																												1	
1947		1																													1	
1954				1																											1	
1957																										1					1	
1960																													1	1		
1961					1																								1	2		
1962					1	1																							1	3		
1963							1																								1	
1964																													1	1		
1965									1				1																		2	
1966								1	3																						4	
1967								1			1																				2	
1968										1	1		2																		4	
1969												1																	1	1	3	
1970													1																		1	
1971													1	2															2	5		
1972														4		1								1					1	7		
1973															1	1	1					1							4	3	11	
1974																	1		1		1								1	4		
1975																		1	2										1	2	6	
1976																			4	2										4	10	
1977																			2	5	5	1	1						1	5	20	
1978																			1	1	2	2	1	1			1	2	5	16		
1979																					4	5	5	2			2	1	2	9	30	
1980																						6	8	5	5	4	1		4	11	44	
1981																							5	7	13	9	3	3	4	13	57	
1982																							1	10	16	43	14	6	5	20	115	
1983																									7	43	79	23	10	18	180	
1984																								2	1	14	97	137	55	53	359	
1985																										2	17	180	280	86	565	
1986																											1	1	40	405	290	737
1987																												2	206	705	913	
1988																												2	11	520	533	
1989																														2	2	
1990																														1	1	
Gesamt	1	1	1	1	2	1	1	2	4	1	2	1	5	6	1	2	2	1	10	8	12	15	21	28	42	117	215	394	991	1757	3645	

Tabelle 4: Zuordnung der Kandidaten zu den Gruppen „Alter und Maturität“

4.2 Verteilungsprüfung der einzelnen Punktwerte

	Punktwert	Quantitative und formale Probleme	Schlauchfiguren	Textverständnis	Planen und Organisieren	Med.-naturwiss. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme und Tabellen	Konzentr. und sorgf. Arbeiten
Mittelwert	90.45	8.54	11.41	7.83	5.46	8.62	9.45	10.23	9.66	8.35	10.89
Stand.-abweich.	22.44	3.75	3.87	3.48	2.40	3.74	3.59	3.75	3.07	3.32	4.45
Median	89	8	11	8	5	8	9	10	10	8	11
Modalwert	84	8	12	6	6	8	9	10	9	7	11
Spannweite	141	20	20	18	18	20	20	20	20	20	20
25. Perzentil	74	6	9	5	4	6	7	7	8	6	8
75. Perzentil	105	11	14	10	7	11	12	13	12	11	14

Tabelle 5: Kennwerte des Punktwerts und für die Untertests

Maßgeblich ist der Punktwert. Er entspricht einer Normalverteilung, metrische Verfahren sind angemessen. Die einzelnen Untertests sind zu Vergleichszwecken mit angegeben. Der Test „Planen und Organisieren“ war im Jahre 2006 vergleichsweise sehr schwer. Er wurde im Unterschied zu anderen Untertests nicht unter Ernstfallbedingungen erprobt, sondern in einer Vergleichsstichprobe. Diese Stichprobe hatte bessere Ergebnisse als in der Anwendung unter Ernstfallbedingungen erzielt – im Jahre 2005 war dies nicht der Fall. Da der Test in Österreich vor allem im oberen Bewerberbereich differenzieren muss, wirkt sich dieser Effekt hier eher begünstigend aus.

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 bis zu 100mal geübt wurde und dann fast vollständig bearbeitet werden konnte. Dies gelingt durch den Wechsel der Anforderungen, die vorher nicht bekannt gegeben werden, nun nicht mehr.

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

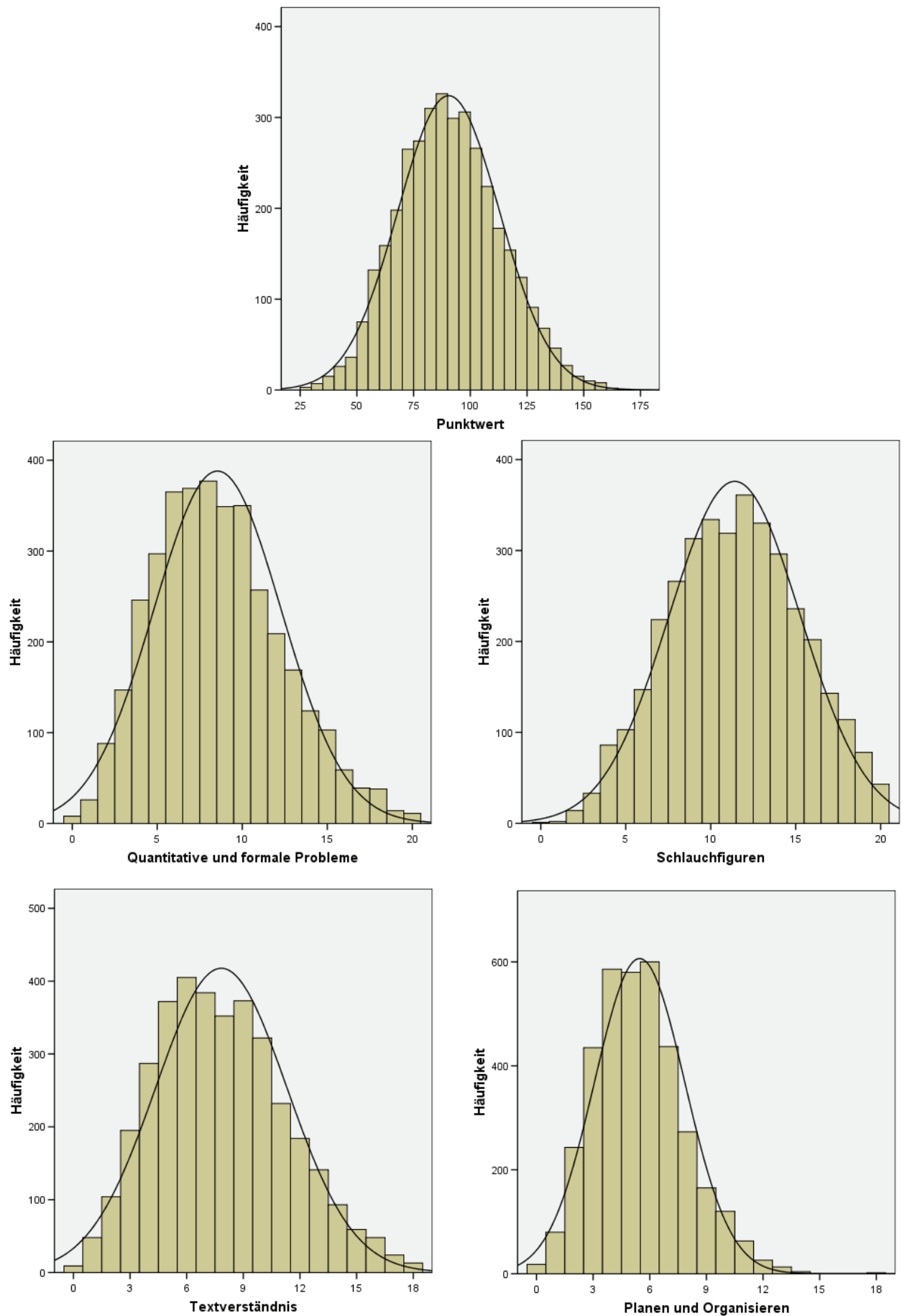


Abbildung 4: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (1)

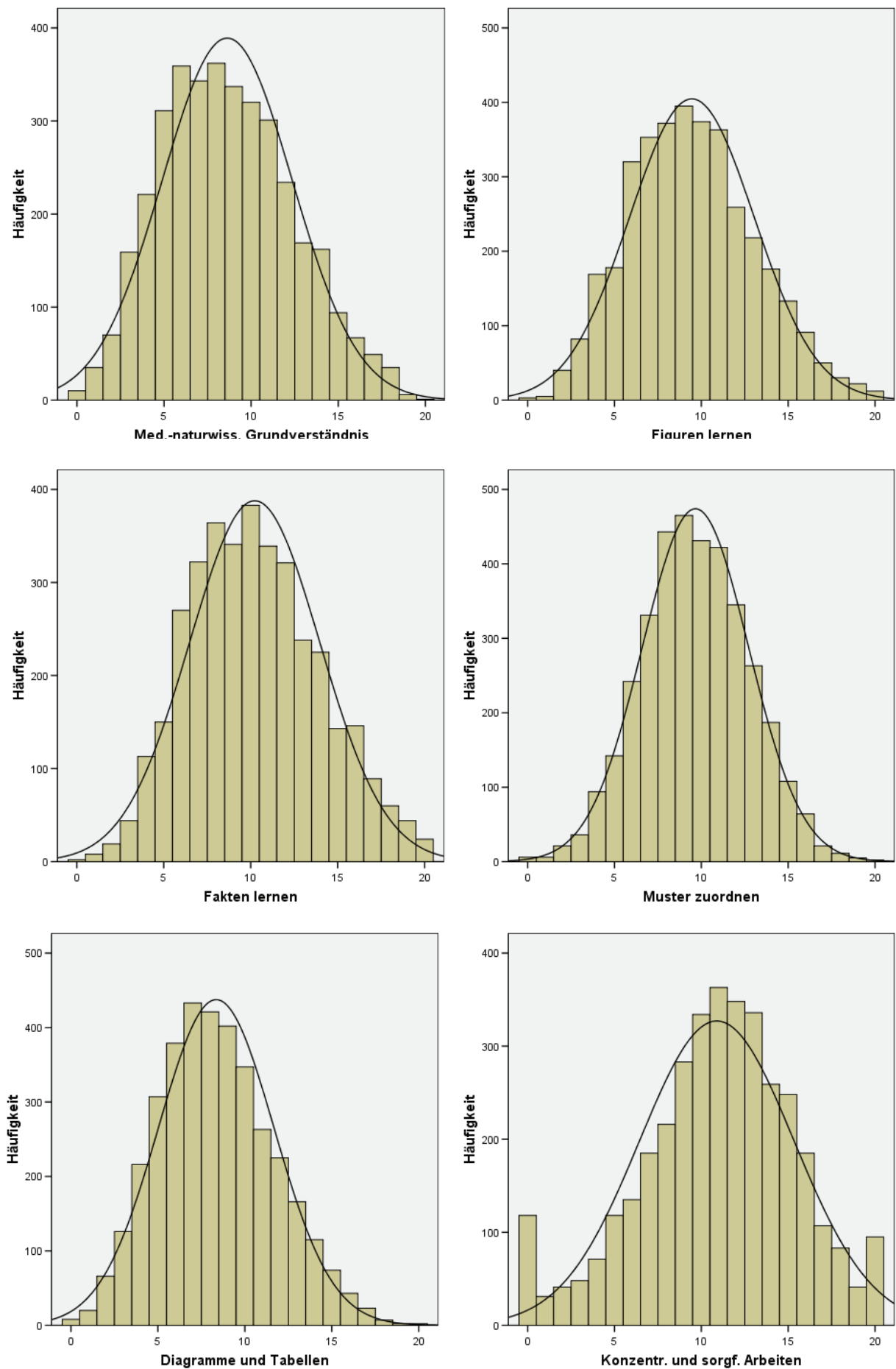


Abbildung 5: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (2)

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

Die Zuverlässigkeitswerte des Punktwertes liegen 2006 mit 0.90 bzw. 0.92 in beiden Ländern im gewohnt hohen Bereich der Vorjahre. Auch die Konsistenz des Testprofiles liegt um 0.80, dem aus dem Vorjahr gewohnten Wert. Hierbei ist zu beachten, dass zu hohe Werte für sehr gleichartige Untertests (mit der Frage, ob man einzelne weglassen kann) 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 Zusammenhanges 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 CH	2006 AT
Punktwert	.91 - .93	.92	.91	.90	.91	.91	.92	.91	.92	.90	.92

Tabelle 6: Reliabilität des Punktwertes (Split Half) für Deutschland (Bereich), Schweiz und Österreich

	Konsistenz des Testprofiles									
	1998	1999	2000	2001	2002	2003	2004	2005	2006 CH	2006 AT
Profil	.83	.81	.81	.80	.80	.81	.78	.82	.79	.83

Tabelle 7: Konsistenz des Testprofiles für Schweiz und Österreich

5.2 Faktorielle Validität

Aufgrund der Korrelationen zwischen den Untertests wurde mittels Strukturanalyse geprüft, ob die Struktur der Untertests für die Schweiz und Österreich mit der der Vorjahre vergleichbar ist. Dies ist 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 sowohl in allen Ländern und Jahren vergleichbar funktioniert.

	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		.20	.39	.17	.19	.32	.25	.19	.31	.10	.52	.55
Med.-naturw. Grundverst.	.25		.26	.53	.61	.20	.25	.57	.23	.31	.69	.73
Schlauchfiguren	.40	.35		.25	.26	.36	.27	.26	.36	.15	.60	.67
Quant. u. formale Probl.	.24	.54	.34		.52	.14	.17	.59	.22	.37	.66	.69
Textverständnis	.23	.64	.36	.53		.19	.20	.53	.21	.33	.67	.71
Figuren lernen	.33	.28	.41	.21	.26		.41	.17	.30	.06	.54	.58
Fakten lernen	.27	.31	.32	.20	.26	.41		.21	.28	.10	.55	.55
Diagramme und Tabellen	.26	.60	.36	.59	.57	.25	.25		.23	.40	.68	.72
Konzent. u. sorgfält. Arbeiten	.35	.32	.40	.30	.29	.30	.24	.32		.20	.59	.62
Planen und Organisieren	.14	.34	.18	.39	.34	.13	.14	.37	.22		.47	.47

Tabelle 8: 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					
2.91	2.96	29.12	29.64	29.12	29.64
2.35	2.44	23.49	24.44	52.61	54.08
Drei-Faktorenlösung					
2.91	2.93	29.05	29.34	29.05	29.34
1.81	1.78	18.13	17.81	47.18	47.15
1.40	1.54	13.97	15.36	61.15	62.51

Tabelle 9: Varianzanteile der einzelnen Faktorenlösungen Schweiz und Österreich 2006 (varimax-rotierte Lösungen)

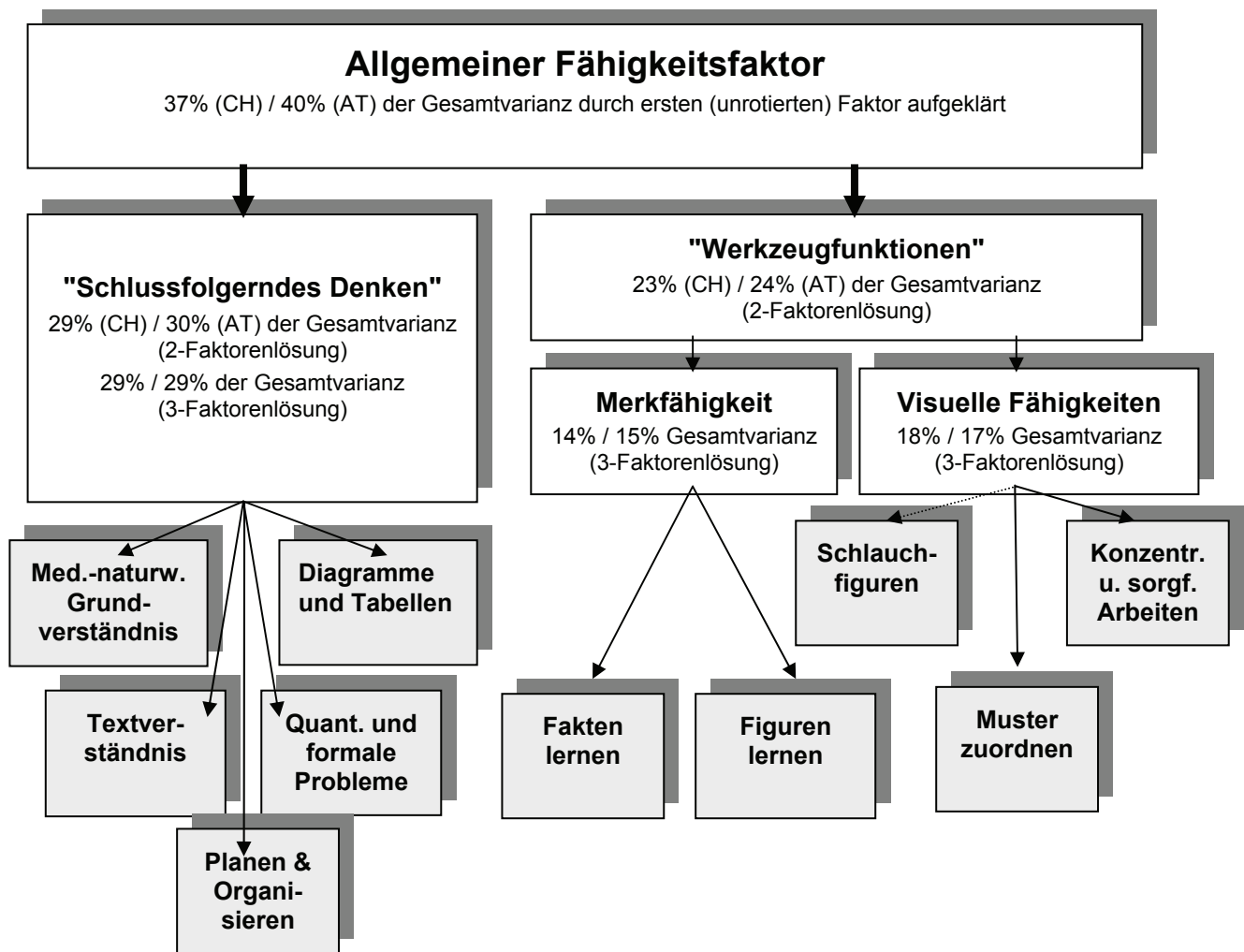


Abbildung 6: Struktur der Untertests des EMS, hierarchische Faktorenstruktur 2006, Werte Schweiz/Österreich der Varianzaufklärung, Untertests entsprechend der Hauptladungen zugeordnet

Die Struktur des EMS bleibt weiterhin sehr 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 „Textverständnis“ charakterisiert (Tabelle 10).

In der Zweifaktorenlösung stechen die Untertests „Diagramme und Tabellen“ sowie „quantitative und formale Probleme“ für den Faktor „schlussfolgerndes Denken“ und der Untertest „Figuren lernen“ als typischer Repräsentant für den Faktor „Werkzeugfunktionen“ hervor (Tabelle 11).

Die Dreifaktorenlösung (

Tabelle 12) teilt den Faktor „Werkzeugfunktionen“ in „Merkfähigkeit“ (repräsentiert durch „Fakten lernen“ und „Figuren lernen“) und „visuelle Fähigkeiten“ („Konzentriertes und sorgfältiges Arbeiten“, „Muster zuordnen“, „Schlauchfiguren“).

	1998	1999	2000	2001	2002	2003	2004	2005	2006 CH	2006 AT
Muster zuordnen	.56	.53	.57	.57	.71	.48	.45	.46	.47	.52
Med.-naturwiss. Grundverst.	.76	.76	.74	.74	.7	.74	.71	.79	.74	.77
Schlauchfiguren	.69	.64	.65	.62	.69	.61	.61	.58	.57	.64
Quant. und formale Probleme	.7	.7	.72	.66	.68	.73	.7	.75	.71	.72
Textverständnis	.72	.7	.71	.72	.62	.77	.7	.76	.72	.74
Figuren lernen	.62	.59	.53	.56	.62	.55	.54	.48	.47	.54
Fakten lernen	.53	.52	.55	.52	.57	.51	.51	.47	.48	.51
Diagramme und Tabellen	.72	.75	.71	.72	.56	.76	.74	.78	.74	.76
Konzentr. u. sorgf. Arbeiten	.47	.51	.55	.49	.54	.48	.44	.41	.52	.58
Planen und Organisieren								.68	.51	.50

Tabelle 10: Ladungen im ersten Faktor der unrotierten Lösung 1998 bis 2006 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor); 1998-2005 Schweiz und 2006 für beide Länder

	Faktor 1						Faktor 2						Kommunalitäten (h ²)					
	02	03	04	05	06 CH	06 AT	02	03	04	05	06 CH	06 AT	02	03	04	05	06 CH	06 AT
Diagramme und Tabellen	.77	.83	.79	.81	.80	.78	-.17	-.17					.63	.73	.66	.70	.66	.67
Med.-naturwiss. Grundverständnis	.77	.86	.80	.82	.77	.76	.21	-.11					.63	.76	.65	.71	.64	.66
Textverständnis	.82	.83	.70	.81	.77	.76	-.11	-.18					.68	.74	.63	.68	.62	.64
Quant. und formale Probleme	.79	.77	.70	.78	.79	.78	-.18	.20					.65	.65	.60	.64	.64	.64
Planen und Organisieren				.72	.61	.62										.54	.37	.39
Konzentr. u. sorgf. Arbeiten							.82	.79	.33	.51	.62	.57	.68	.62	.20	.28	.41	.40
Figuren lernen	.20						.70	.78	.74	.74	.73	.75	.52	.65	.56	.56	.54	.56
Muster zuordnen	.20						.63	.58	.70	.74	.66	.68	.43	.62	.47	.56	.45	.48
Fakten lernen		.26					.67	.50	.67	.62	.64	.64	.47	.78	.47	.40	.42	.42
Schlauchfiguren	.29	.26		.29			.60	.66	.67	.62	.68	.68	.45	.60	.52	.47	.50	.54

Tabelle 11: Ladungen und Kommunalitäten der Zwei-Faktorenlösung, varimaxrotiert; 2002-2005 Schweiz und 2006 für beide Länder

Faktor	Ergebnisse 2006								Deutschland		
	1		2		3		h ²		1	2	3
	2006 CH	2006 AT	2006 CH	2006 AT	2006 CH	2006 AT	2006 CH	2006 AT			
Diagramme und Tabellen	0.80	0.78					0.67	0.67	0.82		
Medizin.-naturwiss. Grundverständnis	0.79	0.77			0.25		0.68	0.69	0.81		0.2
Quantitat. u. formale Probleme	0.79	0.77					0.64	0.64	0.8	0.18	
Textverständnis	0.77	0.77					0.64	0.66	0.79		0.2
Planen und Organisieren	0.57	0.61	0.30		-0.30		0.51	0.42	Noch nicht enthalten		
Muster zuordnen			0.71	0.75			0.54	0.69		0.81	
Konzentr. u. sorgfält. Arbeiten		0.25	0.72	0.75			0.56	0.63		0.7	0.4
Schlauchfiguren		0.26	0.69	0.62	0.23	0.34	0.56	0.56	0.35	0.71	
Fakten lernen					0.77	0.83	0.66	0.73	0.21		0.87
Figuren lernen			0.36	0.37	0.71	0.70	0.63	0.63	0.13	0.47	0.64

Tabelle 12: Faktorenanalyse: Varimaxrotierte Drei-Faktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42).

5.3 Item-Trennschärfen

Die Item-Trennschärfen sind die Korrelationen des Punktwertes für den jeweiligen Untertest mit den zugeordneten Items. Positive Korrelationen weisen darauf hin, dass die Leistungsbesten im jeweiligen Untertest auch beim entsprechenden Item die richtige Lösung bevorzugt gewählt haben. Zu beachten ist, dass bei sehr leichten und sehr schweren 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 für die gewerteten Items. Die Trennschärfen 2006 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).

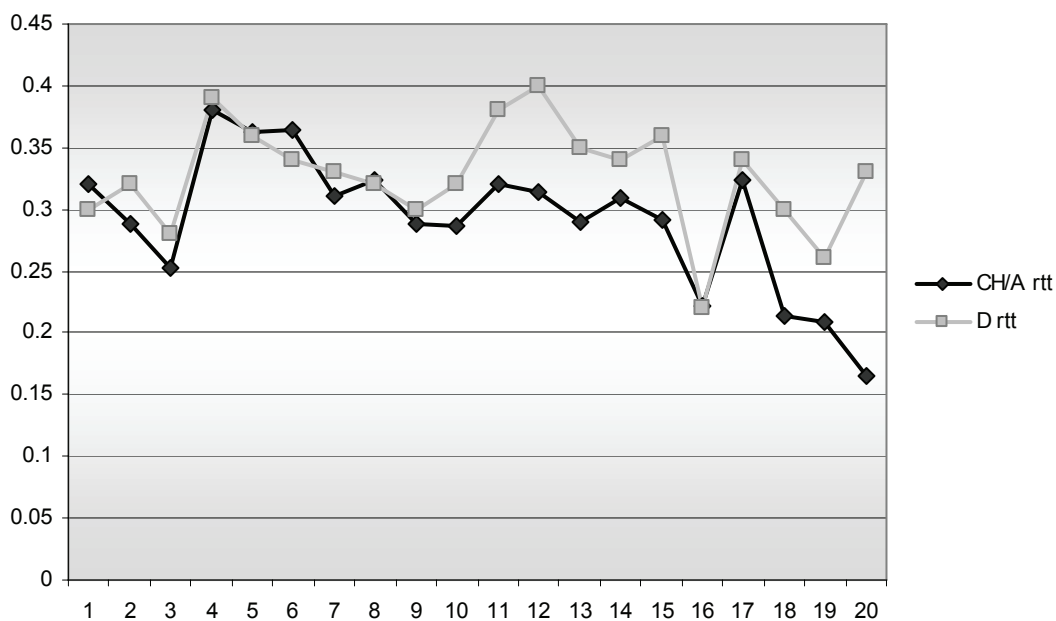


Abbildung 7: Trennschärfen für den Untertest „quantitative und formale Probleme“, alle Items wurden gewertet.

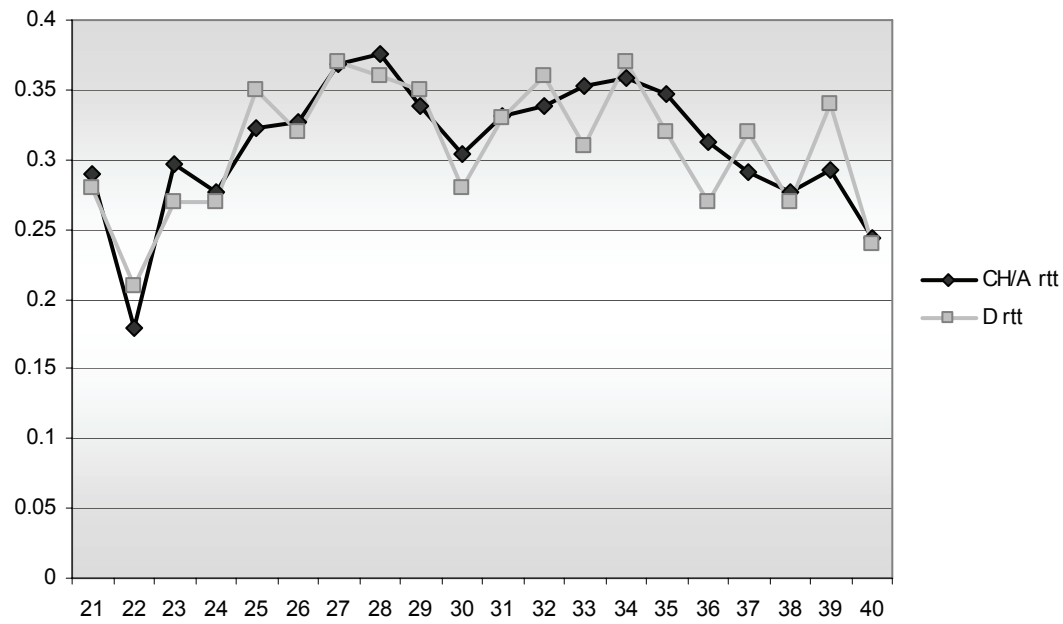


Abbildung 8: Trennschärpen für den Untertest „Schlauchfiguren“, alle Items wurden gewertet.

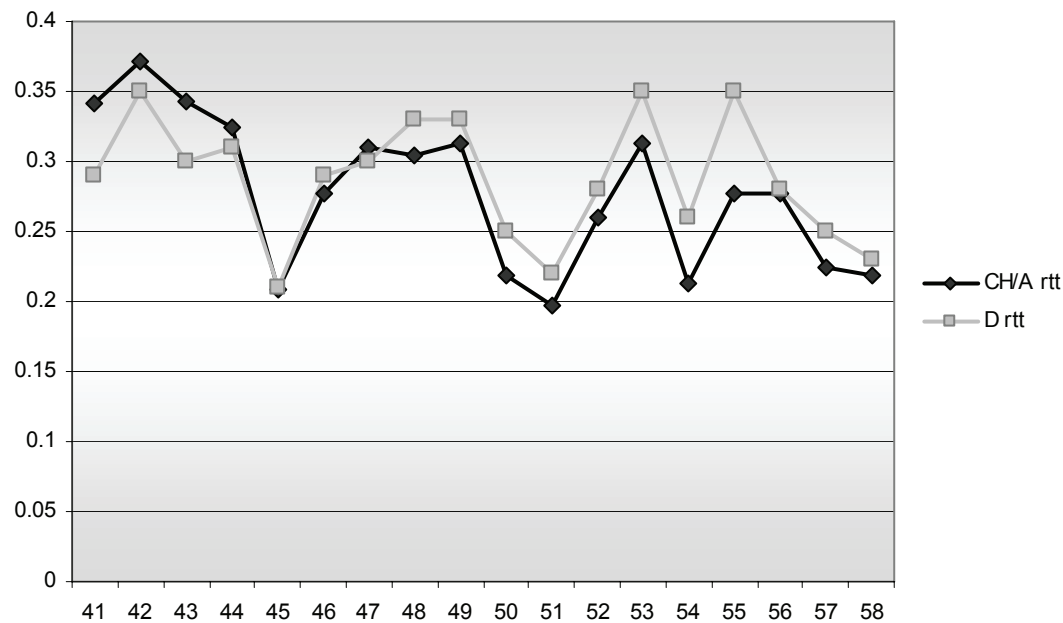


Abbildung 9: Trennschärpen für den Untertest „Textverständnis“, alle Items wurden gewertet.

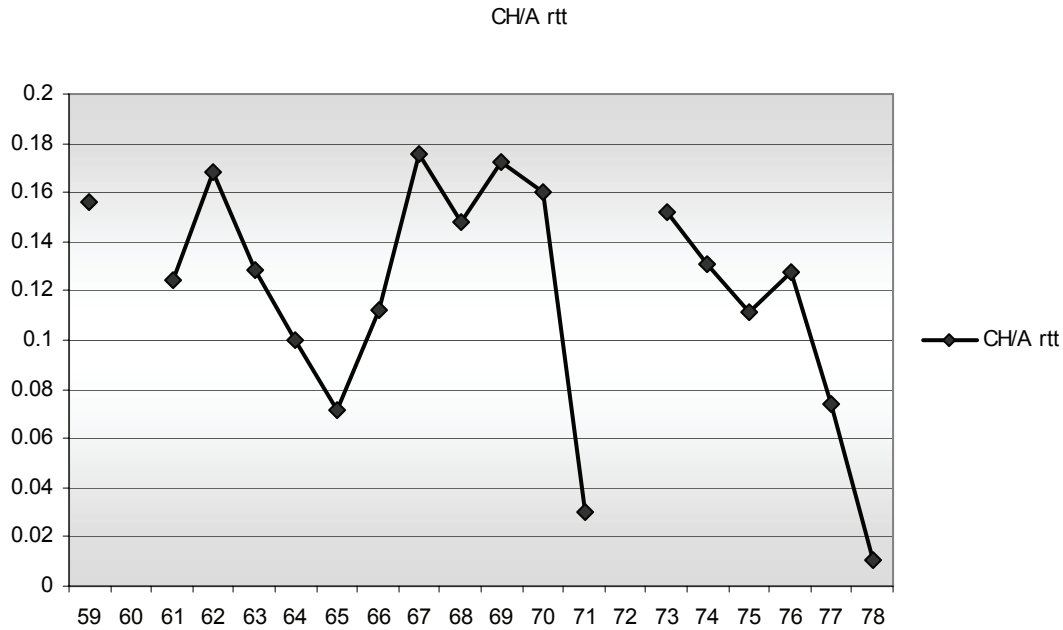


Abbildung 10: Trennschärfen für den Untertest „Planen und Organisieren“ (keine deutschen Vergleichswerte). Die Items 60 und 72 wurden für den Punktwert eliminiert, Item 71 und 78 belassen (sie sind sehr schwer und die geringe Trennschärfe ist vor allem auf die geringe Antwortvarianz „richtig-falsch“ zurückzuführen). Dieser Test wurde neu konstruiert und an einer vergleichbaren Stichprobe voreprobt, die sich nicht in einer Bewerbungssituation befand. Im Gegensatz zum Vorjahr erreichte die Erprobungsstichprobe bessere Werte als die Medizinbewerber.

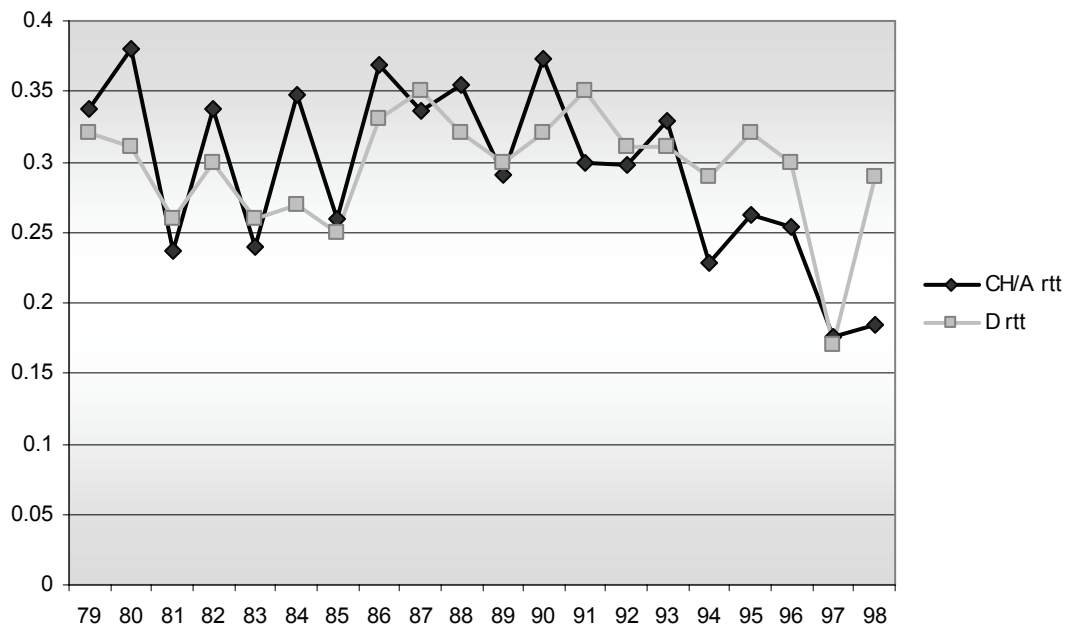


Abbildung 11: Trennschärfen für den Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“, alle Aufgaben wurden gewertet.

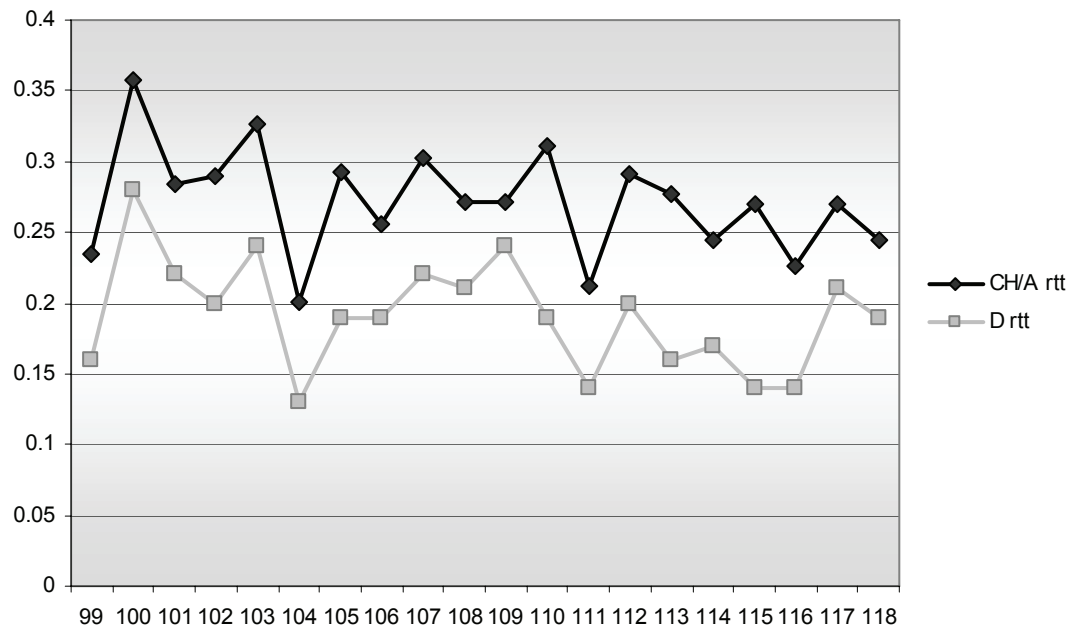


Abbildung 12: Trennschärfe für den Untertest „Figuren lernen“, alle Aufgaben wurden gewertet.

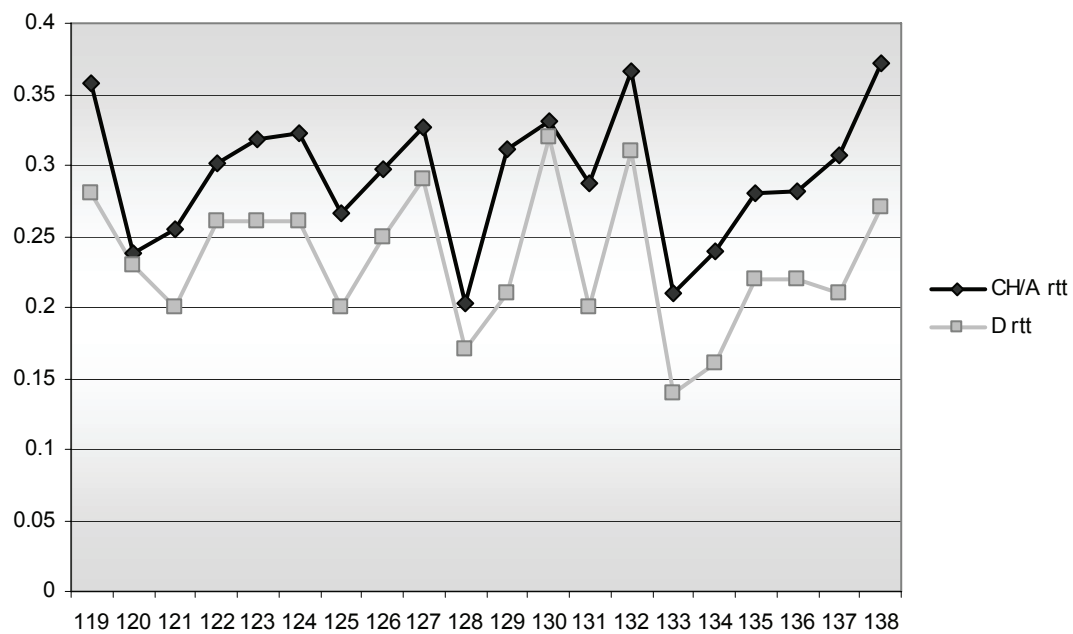


Abbildung 13: Trennschärfe für den Untertest „Fakten lernen“, alle Aufgaben wurden gewertet.

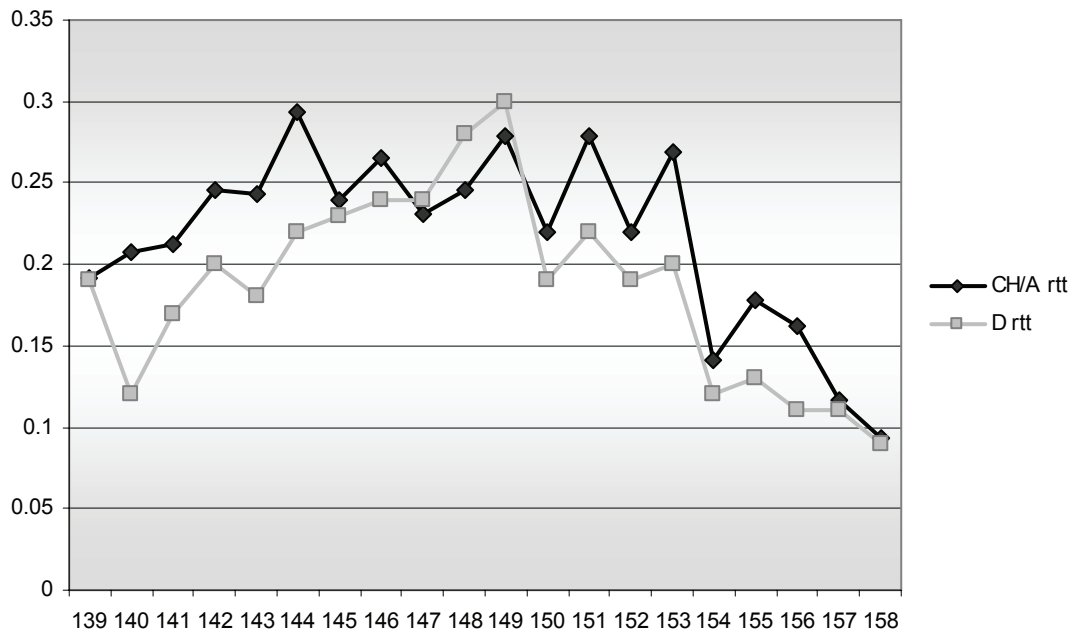


Abbildung 14: Trennschärpen für den Untertest „Muster zuordnen“, alle Aufgaben wurden gewertet.

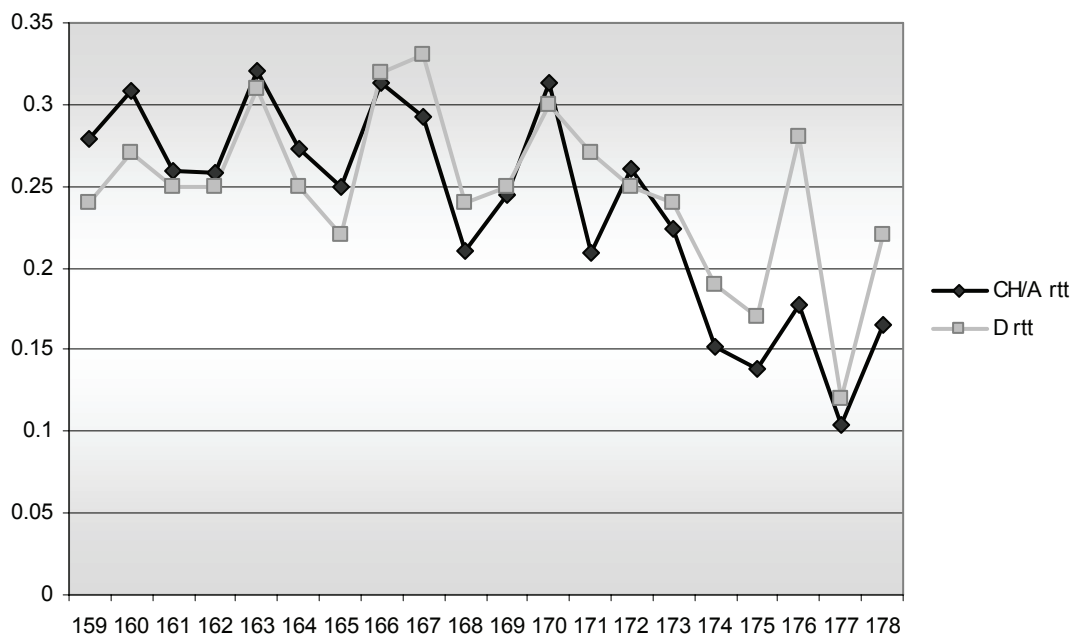


Abbildung 15: Trennschärpen für den Untertest „Diagramme und Tabellen“, alle Aufgaben wurden gewertet.

5.4 Konzentriertes und sorgfältiges Arbeiten

Jährlich wird eine neue, vorher nicht bekannte Anforderung (Instruktion, Zeichensatz) verwendet, um Trainingseffekte minimal zu halten.

Die Instruktion 2006 lautete: „Sie sehen nachfolgend ein Blatt mit 40 Zeilen, die aus je 40 Buchstaben b, d, p und q gebildet werden. Ihre Aufgabe ist es, zeilenweise jedes b durchzustreichen, wenn direkt davor ein d steht UND jedes p durchzustreichen, wenn direkt davor ein q steht.“ Durch zwei Zielzeichen wurde die Schwierigkeit in diesem Jahr weiter erhöht.

Als Parameter ergeben sich:

	Beschreibung	Wertebereich
Richtige	Target richtig markiert b nach d ODER p nach q	0...400
Fehler I	Target nicht markiert, „übersehen“	0...400
Fehler IIa	Nontarget markiert, „falscher Alarm“ b oder d markiert nach einem anderen Zeichen als dem richtigen Vorgänger	0...400
Fehler IIb	Distraktor markiert, „falscher Alarm“: p oder q markiert	0...800

Der Punktwert PW berechnet sich wie folgt:

$$PW = (Richtige - Fehler I - Fehler IIa - Fehler IIb) / K \rightarrow \text{im Intervall 0 bis 20}$$

$$K = \text{Maximum für Richtige} / \text{Maximum Punktwert (20)}$$

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, so dass sich die nachfolgende Umrechnungstabelle ergibt. Dieses Verfahren 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 (Abbildung 16).

		Anzahl	Mittelwert	Standard-abweichung	Minimum	Maximum
deutsch	Richtige	3645	174.5	46.78	0	395
	Fehler Gesamt	3645	38.9	37.46	0	465
	... Auslassungen	3645	34.2	33.28	0	301
	... Fehlmarkierungen	3645	4.7	9.31	0	230
	letztes bearbeitetes Zeichen	3643	834.1	198.40	30	1599

Tabelle 13: Statistiken für Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“

Punktwert 0-20	von Rohwert	bis Rohwert
0	Minimum	0
1	1	13
2	14	26
3	27	39
4	40	52
5	53	66
6	67	79
7	80	92
8	93	105
9	106	118
10	119	132
11	133	145
12	146	158
13	159	171
14	172	184
15	185	198
16	199	211
17	212	224
18	225	237
19	238	251
20	252	400

Tabelle 14: Umrechnung Rohwert in Punktwert „Konzentriertes und sorgfältiges Arbeiten“ 2006, AT

		Fehler	Auslassungen	Fehlmarkierungen
Richtige	Korrelation nach Pearson	-0.31(**)	-0.27(**)	-0.28(**)
	Signifikanz (2-seitig)	0.000	0.000	0.000
Fehler Gesamt	Korrelation nach Pearson		0.97(**)	0.55(**)
	Signifikanz (2-seitig)		0.000	0.000
Auslassungen	Korrelation nach Pearson	0.97(**)		0.34(**)
	Signifikanz (2-seitig)	0.000		0.000

Tabelle 15: Korrelationen der Parameter im Konzentrationstest; **Die Korrelation ist auf dem Niveau von 0,01(2-seitig) signifikant, n = 3645

Die Beziehung zwischen Mengen- und Sorgfaltsleistung ist mit -.31 in dieser Testform höher als in den Vorjahren (-0.10 im Jahr 2005).

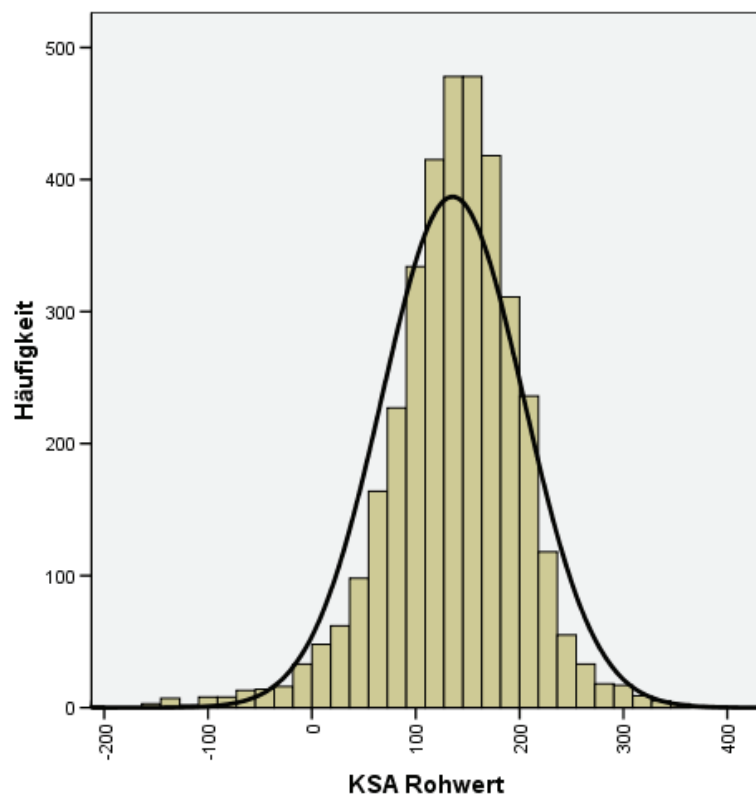


Abbildung 16: Verteilung des Rohwertes für „Konzentriertes und Sorgfältiges Arbeiten 2006“, negative Werte nicht Null gesetzt. 13 Ausreißer liegen unter -200 (nicht dargestellt).

6 Fairness und die Vergleichbarkeit von Gruppen

Fairness, Rechtsgleichheit und Willkürfreiheit sind wichtige Gütekriterien des EMS, die aufgrund der Konstruktionsprinzipien des Tests, der Durchführungs- und Auswertungs-Richtlinien und empirisch aufgrund der vorliegenden Daten beurteilt werden können.

Rechtsgleichheit und Willkürfreiheit wurden als juristische Begriffe in einem Schweizer Bundesgerichtsurteil zum NC als Anforderungen konkret benannt. Willkürfreiheit bedeutet übersetzt in die Terminologie der Testpsychologie vor allem eine hohe Objektivität. Durchführung, Auswertung und die Zulassungsentscheidung müssen frei von subjektiven Einflüssen sein. Die sehr strikte Standardisierung des Vorgehens in allen Phasen des EMS, die Schulung des beteiligten Personals, das Vorhandensein eines Testleiterhandbuches, in welchem Verhaltensweisen für die meisten (auch extremen) Situationen genau vorgeschrieben sind, gewährleisten diese Objektivität. Abweichungen vom Regelablauf und Besonderheiten können von den Testleitern in Protokollen vermerkt und hinsichtlich des Einflusses nachträglich bewertet werden. Die vorliegenden Protokolle aus Wien und Innsbruck weisen für 2006 auf keine einschränkenden Bedingungen hin.

Fairness (in etwa analog der Rechtsgleichheit) wurde und wird bei Eignungstests auch als Chancengleichheit definiert: **Bei gleicher Eignung müssen gleiche Chancen auf Zulassung bestehen und damit im Einzelfall praktisch gleiche Testwerte erzielt werden** (vgl. Trost u.a. 1998, Bartussek 1985, 1986).

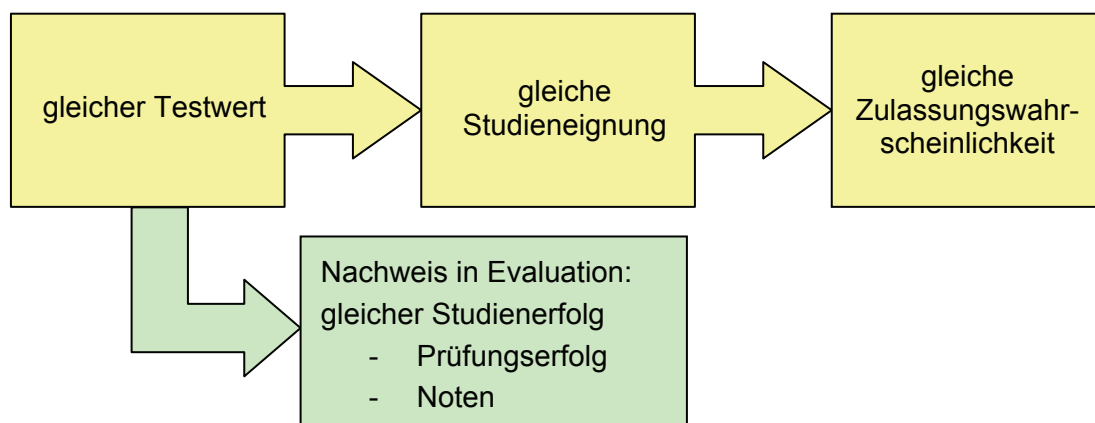


Abbildung 17: Individuelle Chancengleichheit – notwendige Zusammenhänge

Dies wurde in den Evaluationsstudien in Deutschland und der Schweiz vor allem dadurch geprüft, dass Personen mit gleichem Testwert sich nicht beim Studienerfolg unterscheiden – der gleiche Testwert die gleiche Erfolgsprognose erlaubt (s.u.).

Bei allen Evaluationen von Fairness ist der **Unterschied zwischen Einzelfall- und Mittelwertanalysen für die zugehörigen Gruppen** zu beachten. Häufig werden Gruppen (nach Alter, Geschlecht usw.) zusammengefasst und es wird erwartet, dass sich auch die **mittleren Testwerte** für die Gruppen nicht unterscheiden – bzw. bei Vorhandensein von Mittelwertunterschieden wird schnell eine Benachteiligung postuliert.

Fairness kann nicht bedeuten, dass alle Gruppen (z.B. nach Geschlecht, Alter, Vorbildungsgrad, Vorbildungsweg, Region) per Definitionem identische **mittlere Testwerte** erreichen müssen, das wäre in den meisten Fällen sogar ein Zufall. Zum einen kann es **tatsächliche Eignungsunterschiede** geben, zum anderen werden auch **Unterschiede der Repräsentativität** der Gruppen für die jeweilige Bevölkerungsschicht durch bestimmte Selbst-Selektionsprozesse den Vergleich beeinflussen (z.B. unterschiedlich motiviertes Interesse für bestimmte Studiengänge in den Gruppen). Auch **testbedingte Unterschiede** sind denkbar, indem bestimmte Gruppen tatsächlich durch den Test selbst begünstigt

werden. Tritt eine Mittelwertdifferenz auf, ist sehr sorgfältig zu prüfen, auf welchen Faktor bzw. welche Faktoren dieser Mittelwertunterschied zurückgeführt werden kann.

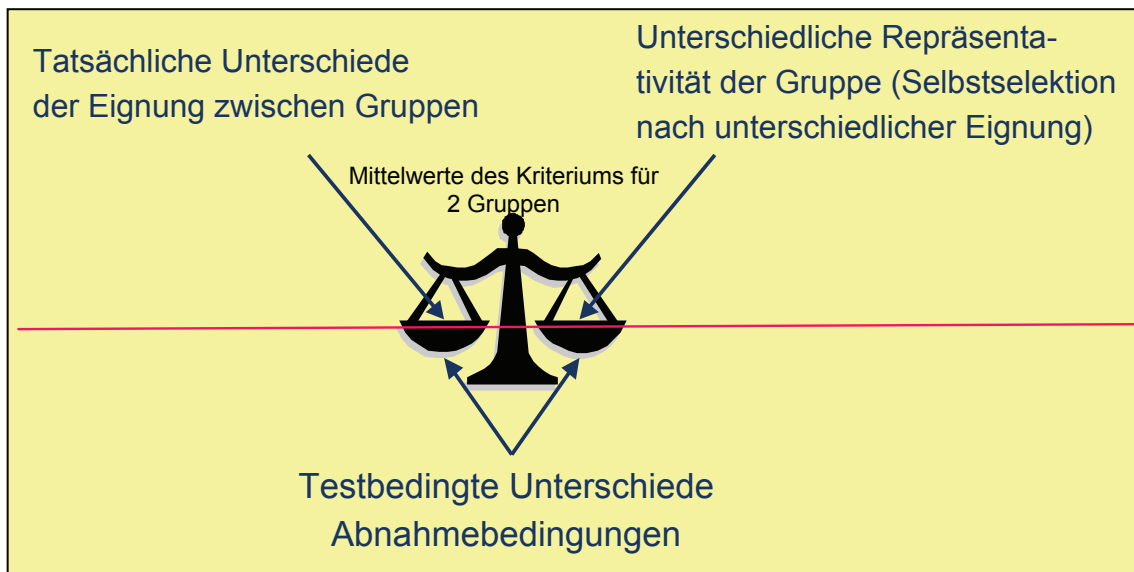


Abbildung 18: Quellen von Gruppenunterschieden (Mittelwerte) beim Zulassungskriterium

- **Tatsächliche mittlere Unterschiede der Eignung** kann es beispielsweise geben, wenn in einer Gruppe die Voraussetzungen weniger gut erfüllt werden als in anderen. Dies kann der Fall sein, wenn für ein anspruchsvolles Studium leistungsbezogene Zulassungskriterien für ältere Personen, die ihre Hochschulreife auf dem zweiten Bildungsweg erworben haben, mit derjenigen von Jüngeren verglichen werden. Hier ist ein geringerer Studienerfolg feststellbar – der Test darf ihn nicht nivellieren, sondern muss den Unterschied möglichst korrekt vorhersagen.
- Mittelwertunterschiede kann es auch geben, wenn sich aus einer **Gruppe bedeutend mehr Personen bewerben als aus einer anderen** – die Quoten nicht den Verhältnissen in der Normalbevölkerung entsprechen. **Selbst-Selektionsprozesse** können unterschiedlich Geeignete eine Studienwahl in den Gruppen treffen lassen. Dann wäre entscheidend, ob es die gleiche Relation zwischen Leistungsfähigkeit und Attraktivität des Bewerbungsziels in allen Gruppen gibt – was in der Regel selten der Fall sein dürfte. Im Fach Veterinärmedizin in der Schweiz bewerben sich beispielsweise 84% Frauen und 16% Männer – hier wäre es reiner Zufall, wenn die mittlere Eignung der Gruppen identisch wäre. Auch die Humanmedizin-Rate (60% Frauen, 40% Männer) spricht nicht unbedingt für die Erwartung gleicher Repräsentativität. Dabei ist es auch möglich, dass diese Selbst-Selektionsprozesse bereits in früheren Bildungsphasen erfolgen (z.B. wenn mehr Frauen die Maturitätsstufe absolvieren als Männer) und sich dann fortsetzen.
- **Testbedingte Unterschiede** würden dann entstehen, wenn eine Gruppe Vorteile bei der Lösung der Testaufgaben hätte, die nicht mit Studieneignung in Verbindung stehen. Hierfür wäre ein Beispiel ein Bezug auf solches Schulwissen, welches durch vorherige Berufsausbildungen oder exzessiven Trainingskursbesuch besser ausgeprägt sein kann. Die Personen würden dann unter vergleichbaren Studienbedingungen schlechter abschneiden als solche, die weniger Ausgangswissen haben, sich aber schneller neues Wissen im Studium aneignen können. Beim EMS wird davon ausgegangen, dass die Studienanforderungen für alle Personen gleich sind und es wurden solche Untertests gesucht, welche diese Anforderungen gut simulieren und vordergründig keine Gruppen benachteiligen.

Üblicherweise ist der Nachweis der Fairness nur über Analysen der **Bewährung** der Zulassungsentscheidung zu erbringen: Ist tatsächlich gewährleistet, dass der gleiche

Testwert der Zulassung für den Einzelfall in allen Gruppen mit einer gleichen Erfolgswahrscheinlichkeit im Studium verbunden ist?

In unserem Fall soll der Test die Studieneignung vorhersagen. Studieneignung kann über Studiendauer und Notengebung objektiviert werden (in einer angemessenen Zeit das Studium mit guten Leistungen abschließen). Wenn nachgewiesen wird, dass in den zu vergleichenden Gruppen bei **gleichem Testwert die gleiche Bestehenswahrscheinlichkeit für Prüfungen und ein vergleichbares Notenniveau erreicht wird**, ist Chancengleichheit gewährleistet.

Diese Daten können natürlich nur retrospektiv gewonnen werden, indem der Studienverlauf für die Zugelassenen dann mit dem Testwert verglichen wird. Dies ist für die Situation in Österreich ebenfalls geplant. Ein Vorteil des EMS ist es, dass diese Analysen sowohl in Deutschland, als auch in der Schweiz bereits sehr ausführlich vorgenommen worden sind. Die Schweizer Ergebnisse sind mit Österreich besser vergleichbar, weil der EMS als ausschließliches Zulassungskriterium verwendet wurde.

In der Literatur findet sich hin und wieder die Empfehlung, quotenbezogen Subgruppen zu bilden (nach Alter, Geschlecht, Vorbildung o.ä.) und die Zulassung entsprechend der Anteile dieser Gruppen an der Gesamtbewerberzahl vorzunehmen. Dazu kann man entweder Mittelwertunterschiede ausgleichen („Boni“ für die schlechtere Gruppe geben) oder sich direkt an die Bewerbungsquoten halten und jede Gruppe proportional dazu zulassen.

Dieses Vorgehen hat Nachteile und kann nicht empfohlen werden, wenn man der bisherigen Argumentation folgt – es geht selten ohne Verletzung der Chancengleichheit auf individueller Ebene ab. So genannte „Paternostereffekte“ können die **individuelle Chancengleichheit** negativ beeinflussen: Bei jeder kapazitätsbezogenen Zulassung steht die Zahl der Zuzulassenden fest. Indem Personen bevorteilt werden, weil man vermeintliche Nachteile ausgleicht, wird dann immer genau die gleiche Personenzahl benachteiligt. Gleiche Testwerte bedeuten nicht mehr die gleiche Zulassungschance.

- Die Ausgangsmenge der Bewerbungen einer Gruppe beeinflusst deren Erfolg (diese muss nicht repräsentativ sein). Die Zulassung 2006 durch die politisch vorgegebenen Länderquoten bietet selbst ein Beispiel für die Problematik, hier für Humanmedizin: In der Nicht-EU-Quote (74 Bewerbungen, 33 Plätze) reicht für die Zulassung ein Testwert von 90, in der Österreich-Quote (1896 Bewerbungen, 495 Plätze) musste man dagegen einen Testwert von 105 erzielen. Nur 9 Personen aus der Nicht-EU-Quote hätten diesen Wert eigentlich erreicht, 24 Personen sind durch den „Paternoster“ der Länderquote in die Gruppe der Zugelassenen befördert worden.
- Die Rechtsgleichheit, bei gleicher Eignung gleiche Zulassungschancen zu erhalten, dürfte aus sachlogischen Gründen durch Verwendung der ursprünglichen Bewerberquoten kaum einzuhalten sein: Indem willkürliche Bewerberquoten übernommen werden, wird das zur Zulassung führende Leistungsniveau sich mit viel größerer Wahrscheinlichkeit unterscheiden. Männer und Frauen haben in der Normalbevölkerung einen Anteil von je 50%, bei den Bewerbungen rund 60% Frauen und 40% Männer. Welche Quoten soll man also zugrundelegen?

6.1 Vergleich der Geschlechter

Die Fairness gegenüber den Geschlechtern ist eine wichtige Voraussetzung, die bisher sowohl in Deutschland als auch in der Schweiz beim Einsatz des TMS bzw. EMS nachweislich gegeben war – dies stand auch immer sehr stark im Interesse der wissenschaftlichen Begleitforschung.

Fairness heißt wie oben begründet **nicht**, dass im **Mittel** gleiche Testwerte für Männer und Frauen erreicht werden müssen. Ein erster Grund ist, dass der Anteil nicht gleich ist: Nach wie vor bewerben sich mehr Frauen als Männer – beide Gruppen sind unterschiedlich repräsentativ für die entsprechende Wohnbevölkerung und eine Wertegleichheit wäre allein schon deshalb Zufall (auch wenn sich die unterschiedliche Repräsentativität schon in der

Maturaphase findet). Außerdem können unterschiedliche Selbstselektionsprozesse erfolgen, indem sich aus einer Gruppe nur die Leistungsbesten bewerben, in einer anderen diese Bewerbung unabhängig(er) von der eigenen Leistung erfolgt. Dennoch müssen die Differenzen analysiert werden, um – bei Vorhandensein – Hinweisen auf mögliche Benachteiligungen nachzugehen.

In der **Schweiz** betrug von 1998 bis 2005 der Unterschied beim Testwert für die zahlenmäßig große deutsche Sprachgruppe zwischen 1.4 und 2.2 Punkten zugunsten der Männer (1.6, 1.5, 1.4, 2.2, 1.3, 1.4, 2.0, 1.6). Der Unterschied von 1.9 im Jahr 2006 liegt im Rahmen der Vorjahre. Der Frauenanteil beträgt 2006 insgesamt 62.5%, in der Veterinärmedizin sogar 84%.

Eine Varianzanalyse nach Alter/Maturitätsgruppe, Disziplin (in der Schweiz sind Human-, Veterinär- und Zahnmedizin involviert, Veterinärmedizin wählen rund 80% Frauen und nur 20% Männer) und Geschlecht zeigt, dass der Effekt des Geschlechtes auf den Testwert statistisch nicht signifikant ist.

In früheren Evaluationsstudien (vgl. z.B. Hänsgen und Spicher 2002) konnte nachgewiesen werden, dass der geringe Werteunterschied beim Testwert auch mit einem entsprechenden Unterschied beim Studienerfolg (Bestehen der Prüfungen) verbunden ist – Frauen und Männer haben bei gleicher Eignung die gleichen Chancen, einen Studienplatz zu erreichen. In der zweiten Evaluationsstudie 2001 ergab sich folgender Zusammenhang für die erste medizinische Vorprüfung in der Schweiz:

Geschlecht	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
männlich	100 (20%)	96.8	7.0	407 (80%)	104.6	8.3
weiblich	176 (27%)	97.2	6.3	467 (73%)	104.8	7.9

Tabelle 16: Bestehen der 1. Vorprüfung nach Geschlecht – mittlerer Testwert der jeweiligen Gruppen. Die Erfolgsquote für Männer liegt um 7% besser als die der Frauen, die mittleren Testwerte der Erfolgreichen sind dagegen nahezu identisch.

In **Deutschland** betrug der Unterschied beim Testwert des TMS zwischen Männern und Frauen 1996 1.4, mit einem Effektstärkewert von .07 kann dieser als nicht bedeutsam angesehen werden. Im Jahre 1994 waren es 2.1 (Effektstärkewert .10 ist schwach bedeutsam). Der Unterschied ist mit den in der Schweiz ermittelten Unterschieden vergleichbar (Trost u. M. 1996, 1997). Der Anteil der Frauen bei den Bewerbungen pendelte ebenfalls um 60%.

Dabei ist auch festzuhalten, dass die 2006 verwendete Testfassung in Teilen bereits in Vorjahren sowohl in Deutschland als auch in der Schweiz eingesetzt worden ist, wo keine Hinweise auf Benachteiligungen gefunden wurden. Diese Vorbemerkungen sind notwendig, um die Ergebnisse in Österreich einordnen zu können.

Zusammenfassend bedeuten sie, dass es auch 2006 für die mit der Schweizer Fassung identische Testfassung, die in Österreich zum Einsatz kam, keine Hinweise auf bedeutsame Testwertunterschiede für die Geschlechter gibt. Auch in Deutschland wurde in dem Jahr des Testeinsatzes der jetzt genutzten Fassung kein Unterschied zwischen den Geschlechtern beim Testwert festgestellt, der sich von den hier dargestellten Trends unterscheidet.

Vorgreifend sei festgestellt, dass in Österreich sich im Verhältnis dazu größere mittlere Differenzen für die Geschlechtergruppen finden. Dabei wird berechtigterweise auch die Frage nach der „Genderfairness“ gestellt. Da ein genereller Benachteiligungseffekt für Frauen durch das Konzept des Tests aus den oben genannten Gründen ziemlich sicher ausgeschlossen werden kann, muss der Frage nachgegangen werden, warum in Österreich diese Differenzen größer sind. Wie oben beschrieben, können mehrere Faktoren zugleich für diese Unterschiede verantwortlich sein.

Zum gegenwärtigen Zeitpunkt stehen weitere soziodemografische Daten und Ergebnisse einer Befragung am Ende des EMS zur Verfügung, um die Systematik der Unterschiede weiter einzugrenzen. Die nachfolgenden Tabellen zeigen die Unterschiede beim Testwert zwischen den Geschlechtern

- gesamt und nach den 3 Zulassungsquoten EU, Nicht-EU und Österreich
- nach den gewünschten Disziplinen Zahn-, Humanmedizin und der Kombination
- nach den beiden Testorten Wien und Innsbruck
- nach den drei Altersgruppen (Geburtsjahre bis 1982, bis 1985 und nach 1985). Die älteste Gruppe ist zusätzlich danach geteilt, ob die Maturitätsprüfung früher oder später abgelegt wurde.
- nach drei Gruppen nach Durchschnittsnoten aus einer freiwilligen Befragung am Ende des Tests. Zum Zwecke der Vergleichbarkeit wurden nur Personen verwendet, die als theoretische Bestnote die Note 1 und einen Notendurchschnitt bis 3 angegeben haben (1 bis 1.5, 1.6 bis 2.2, über 2.2). Die nicht einbezogenen Personen haben andere theoretische Bestnoten angegeben oder Punktesysteme, die bis 10 reichen.
- nach der Antwort auf die Frage, ob der Test viel schwieriger als erwartet, schwieriger als erwartet, genauso wie erwartet oder leichter als erwartet erlebt wurde (freiwillige Befragung am Ende des Tests)
- nach der Antwort auf die Frage, ob die Tagesform viel schlechter, schlechter, genauso oder besser als sonst erlebt wurde (freiwillige Befragung am Ende des Tests).

	Personenzahl	Mittlerer Testwert	Standardabweichung
Männlich	1605	102.3	10.0
Weiblich	2040	98.2	9.6
Nach Quote			
EU männlich	702	104.1	9.3
EU weiblich	825	100.7	9.1
Nicht EU männlich	41	93.0	11.3
Nicht EU weiblich	53	88.8	11.5
AT männlich	862	101.2	10.1
AT weiblich	1162	96.9	9.4
Nach Disziplin			
ZM männlich	99	96.7	9.2
ZM weiblich	136	94.5	9.9
Kombi männlich	106	101.6	10.0
Kombi weiblich	106	97.0	9.2
HM männlich	1400	102.7	10.0
HM weiblich	1798	98.6	9.5
Nach Testort			
Innsbruck männlich	659	103.5	9.7
Innsbruck weiblich	756	99.2	9.5
Wien männlich	946	101.4	10.2
Wien weiblich	1284	97.6	9.6

Fortsetzung nächste Seite

	Personenzahl	Mittlerer Testwert	Standardabweichung
Nach Alter			
Bis 82, frühe Matur. männl.	112	102.3	9.7
Bis 82, frühe Matur. weibl.	106	97.6	10.6
bis 82, späte Matur. männl.	67	95.2	9.7
bis 82, späte Matur. weibl.	70	92.7	9.5
83-85, männl.	595	103.0	10.1
83-85, weibl.	509	100.3	10.0
nach 85 männl.	831	102.3	9.9
nach 85 weibl.	1355	97.7	9.2
Nach erfragtem Notendurchschnitt (Kommentar beachten)			
bessere Noten männl.	289	108.0	10.1
bessere Noten weibl.	582	103.4	8.7
mittlere Noten männl.	361	103.6	9.3
mittlere Noten weibl.	413	98.7	8.5
schlechtere Noten männl.	234	102.1	8.8
schlechtere Noten weibl.	179	95.9	7.7
Nach Vorbereitungsaufwand in Stunden			
Vorber. bis 8 h männl.	520	100.4	9.3
Vorber. bis 8 h weibl.	638	95.7	8.6
Vorber. 9-23 h männl.	473	103.6	9.9
Vorber. 9-23 h weibl.	598	99.3	9.1
Vorber. über 23 h männl.	490	103.8	10.0
Vorber. über 23 h weibl.	651	100.5	10.1
Nach erlebter Schwierigkeit			
viel schwieriger männl.	324	100.2	9.3
viel schwieriger weibl.	521	96.3	9.2
schwieriger männl.	789	102.6	9.9
schwieriger weibl.	941	98.9	9.5
wie erwartet männl.	374	103.4	10.2
wie erwartet weibl.	459	99.6	9.8
leichter männl.	31	103.4	12.7
leichter weibl.	26	101.0	11.7
Nach erlebter Tagesform			
Form viel schlechter männl.	128	98.6	9.9
Form viel schlechter weibl.	151	94.7	9.2
Form schlechter männl.	802	102.4	9.9
Form schlechter weibl.	1010	97.9	9.4
Form wie sonst männl.	533	103.1	10.0
Form wie sonst weibl.	723	99.7	9.6
Form besser männl.	48	103.1	9.8
Form besser weibl.	45	100.7	11.2

Tabelle 17: Geschlechtsspezifische Ergebnisse nach verschiedenen Untergruppen

In den nachfolgenden Abbildungen sind die Mittelwerte (Punkte) und Standardabweichungen $m+s$, $m-s$ grafisch dargestellt.

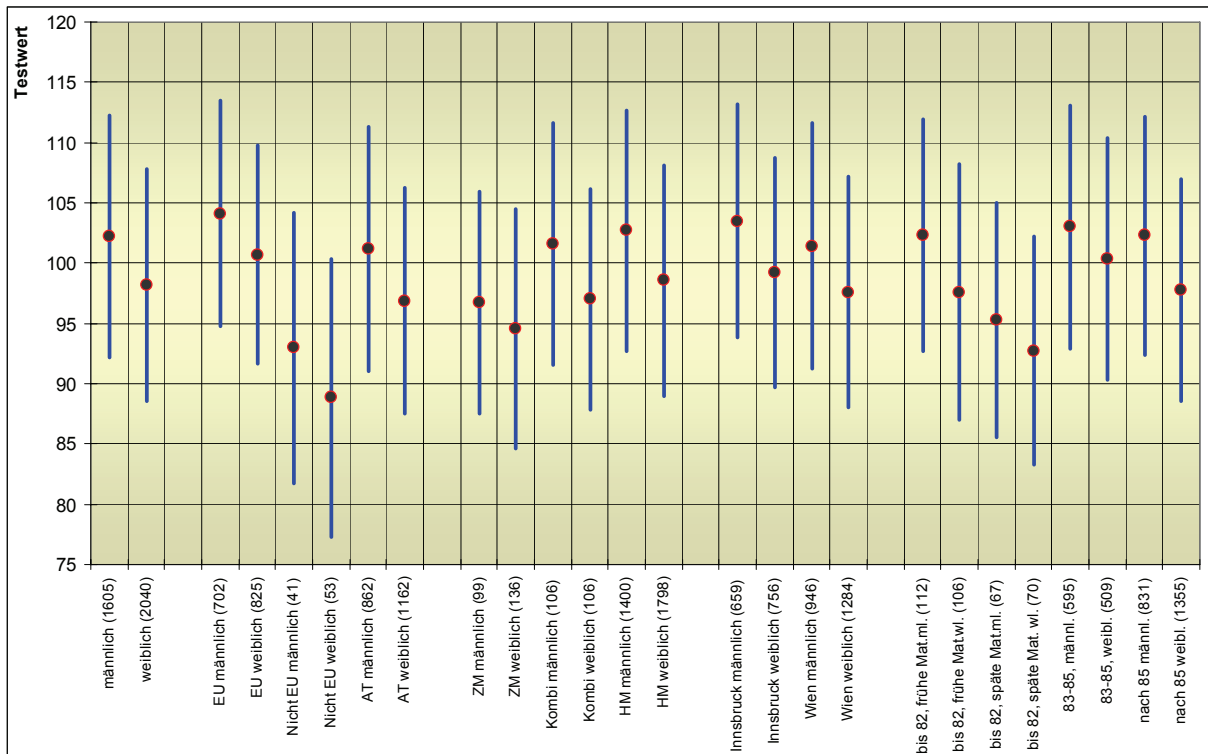


Abbildung 19: Mittelwert und Standardabweichung, geschlechtsspezifisch für verschiedene Gruppen

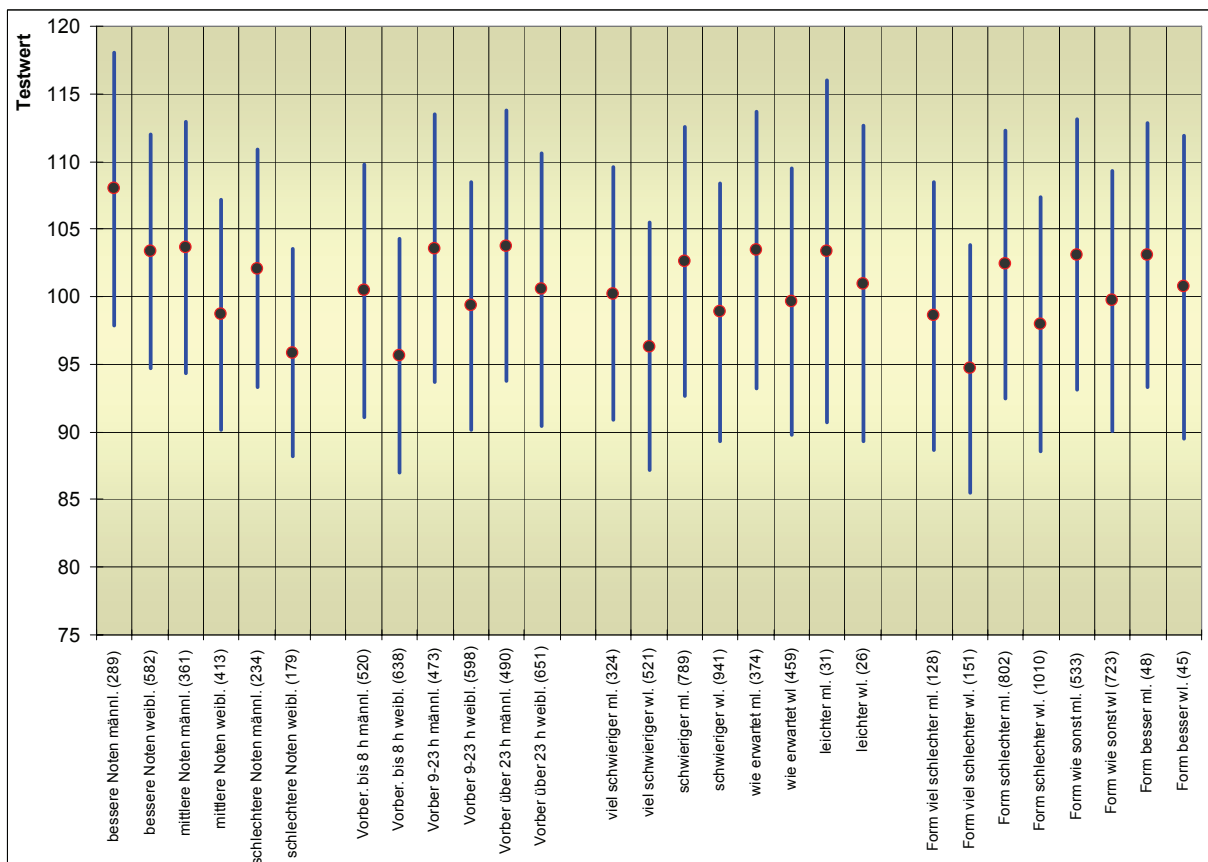


Abbildung 20: Mittelwert und Standardabweichung, geschlechtsspezifisch für verschiedene Gruppen

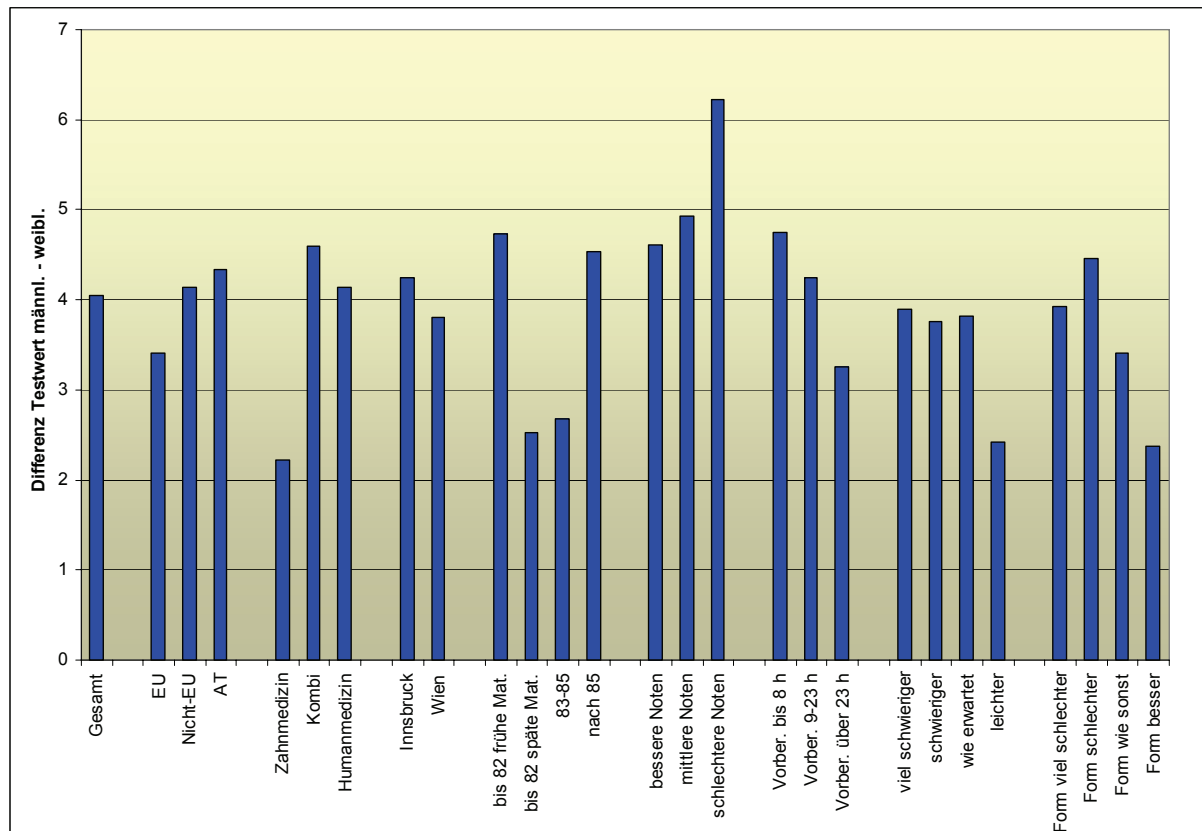


Abbildung 21: Differenz des Testwertes männlich-weiblich für verschiedene Gruppen

Die Abbildung 21 verdeutlicht die Testwert-Unterschiede, je höher der Balken, desto besser die Ergebnisse der Männer. Weniger Einfluss des Geschlechtes (mit einem Unterschied von 2 in etwa auf dem Niveau von Schweiz oder Deutschland) gibt es in den Gruppen der Zahnmedizin-Bewerbungen, der mittleren Altersgruppe und den älteren Personen mit später Maturität (2. Bildungsweg). Auch für Personen, die sich länger auf den Test vorbereitet haben, ist dies so. Erlebte Schwierigkeit und Tagesform haben ebenfalls einen Einfluss.

In einer mehrfachen Varianzanalyse des Testwertes mit Alter, Geschlecht, Länderquote und Disziplin bleibt der Alterseinfluss im Gegensatz zur Schweiz signifikant – für die Altersklassen ist der Effekt hier nicht signifikant.

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	42113.151(a)	51	825.748	9.209	.000
Konstanter Term	1749602.594	1	1749602.594	19511.099	.000
Alterklasse	228.762	2	114.381	1.276	.279
Geschlecht	728.058	1	728.058	8.119	.004
Länderquote	2226.655	2	1113.327	12.416	.000
Disziplin	980.409	2	490.205	5.467	.004
Fehler	322192.107	3593	89.672		
Gesamt	36803906.000	3645			
Korrigierte Gesamtvariation	364305.258	3644			

Tabelle 18: Dreifache Varianzanalyse des Testwertes nach Alter, Geschlecht und Disziplin

(a R-Quadrat = .116 (korrigiertes R-Quadrat = .103) Alle Wechselwirkungen sind nicht signifikant und werden nicht dargestellt.

Nachfolgend kann gezeigt werden, dass die geschlechtsspezifischen Unterschiede auf Untertestebene bis auf „Muster zuordnen“ und „Figuren lernen“ ebenfalls signifikant sind, im Trend den bisher bekannten Unterschieden in der Schweiz folgen, in vielen Fällen aber deutlicher ausgeprägt sind. 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. Sie können aufgrund dieser Ergebnisse nicht differenziert ausgewertet werden.

Untertest	Geschlecht	m	s	Levene's Test Varianz			t-Test Mittelwert		
					F	Sig.	t	df	Sig.
Testwert	männlich	102.25	10.03	homogen	3.089	.079	12.398	3643	.000
	weiblich	98.20	9.61						
Quantitative und formale Probleme	männlich	9.90	3.79	nicht homogen	33.472	.000	20.285	3219.27	.000
	weiblich	7.47	3.34						
Schlauchfiguren	männlich	12.23	3.76	homogen	0.773	.379	11.510	3643	.000
	weiblich	10.77	3.83						
Textverständnis	männlich	8.74	3.62	nicht homogen	36.475	.000	14.141	3214.75	.000
	weiblich	7.12	3.19						
Planen und Organisieren	männlich	5.77	2.44	homogen	3.342	.068	6.826	3643	.000
	weiblich	5.22	2.34						
Med.-naturwiss. Grundverständnis	männlich	9.24	3.82	nicht homogen	11.408	.001	8.895	3340.71	.000
	weiblich	8.14	3.60						
Figuren lernen	männlich	9.48	3.63	homogen	1.005	.316	0.442	3643	.659
	weiblich	9.43	3.56						
Fakten lernen	männlich	9.86	3.72	homogen	1.251	.263	-5.319	3643	.000
	weiblich	10.52	3.75						
Muster zuordnen	männlich	9.66	3.12	homogen	1.497	.221	0.044	3643	.965
	weiblich	9.66	3.02						
Diagramme und Tabellen	männlich	9.34	3.41	nicht homogen	28.921	.000	16.217	3236.33	.000
	weiblich	7.58	3.04						
Konzentr. und sorgf. Arbeiten	männlich	11.33	4.57	nicht homogen	7.320	.007	5.319	3347.43	.000
	weiblich	10.54	4.32						

Tabelle 19: Geschlechterbezogene Unterschiede in Untertests und Signifikanzprüfung mittels t-Test

Der Vergleich der Teilnehmer nach Zulassungsquoten in Österreich (mit Verlauf der Schweizer Ergebnisse über 8 Testdurchführungen) bezüglich Geschlechtsunterschiede auf Untertestebene zeigt, dass das Muster der in der Schweiz auftretenden, über die Jahre stabilen Geschlechtsunterschiede auf Untertestebene auch für die Ergebnisse aus Österreich gilt. Es können also in beiden Stichproben dieselben „Stärken und Schwächen“ beobachtet werden, wobei allerdings in der AT-Stichprobe die Unterschiede stärker ausgeprägt sind.

Für die Nicht-EU-Kohorte ist anzumerken, dass für die sprachabhängigen Untertests ein Bodeneffekt zu beobachten ist. Insbesondere bei den Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“, „Textverständnis“ und „Diagramme und Tabellen“

wurden von diesen Teilnehmern insgesamt geringe Ergebnisse erzielt, wodurch auch keine auffälligen Geschlechtsunterschiede mehr auftreten.

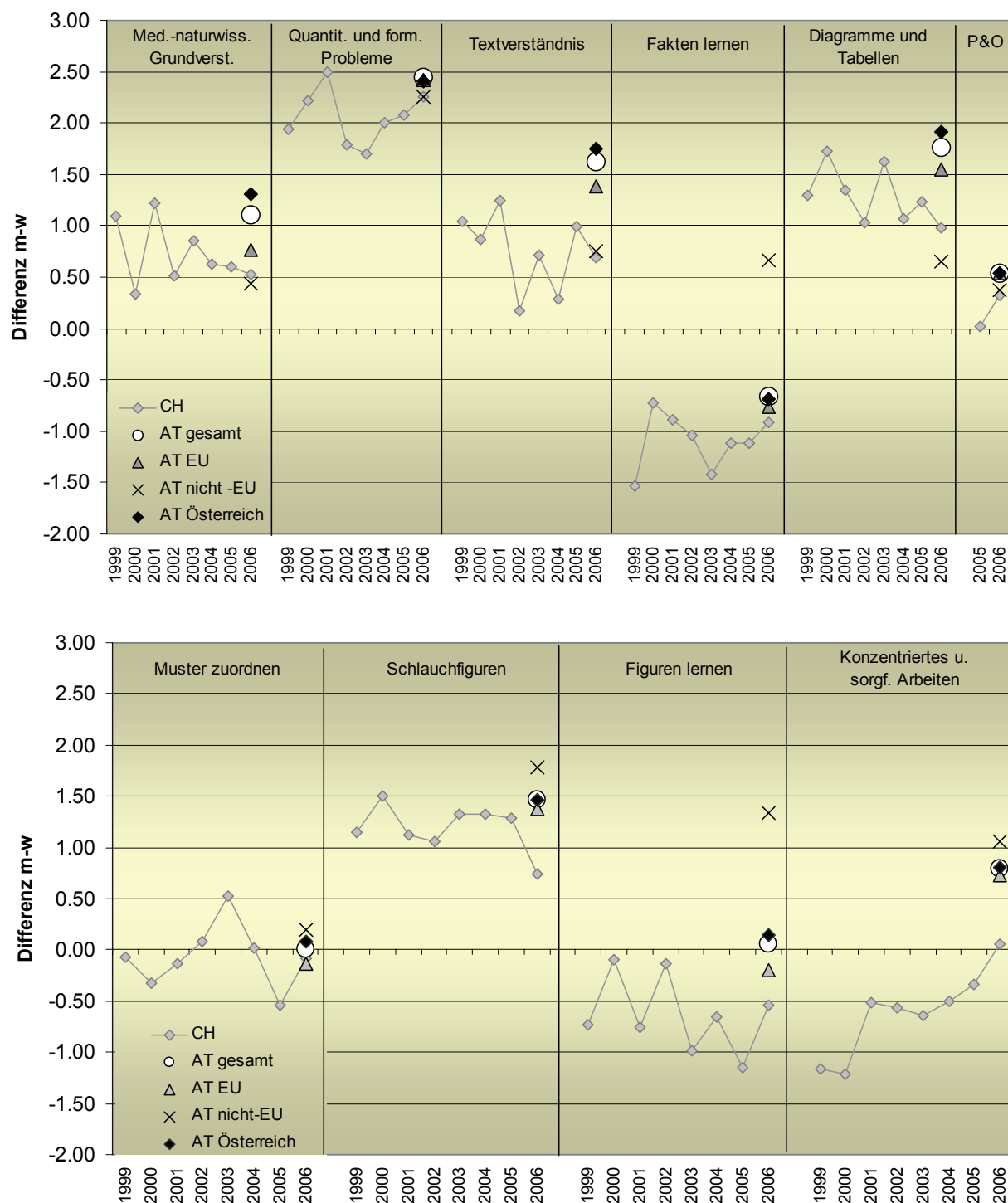
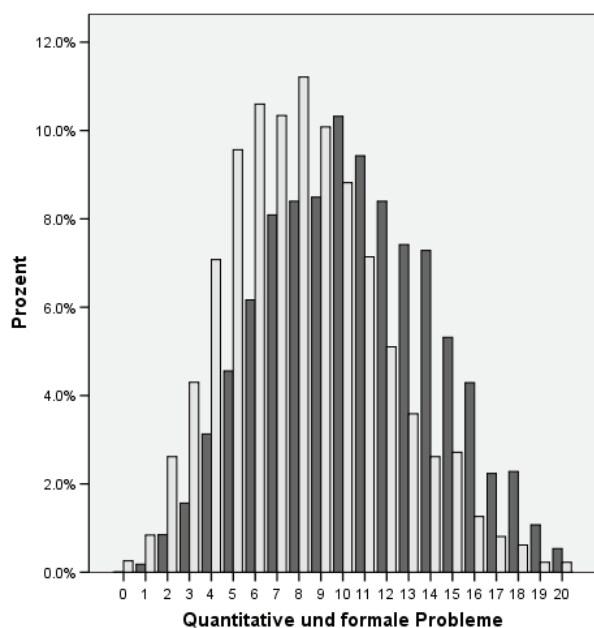


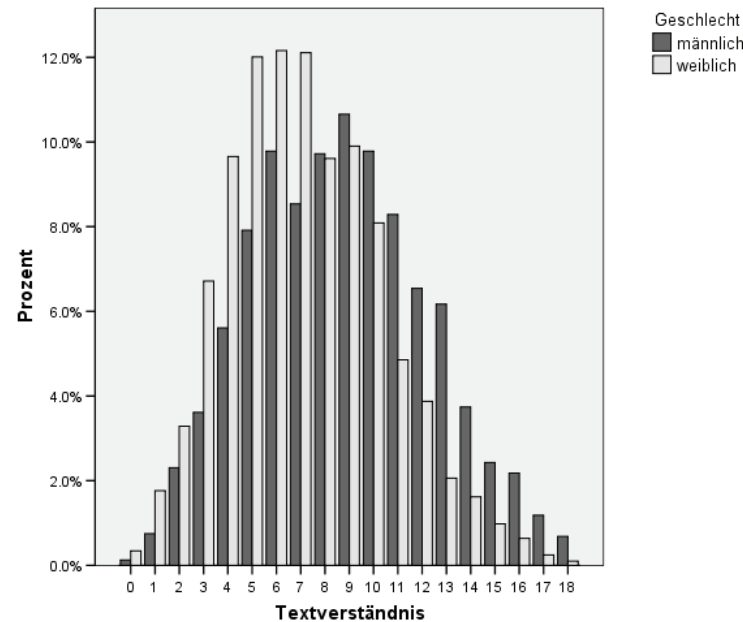
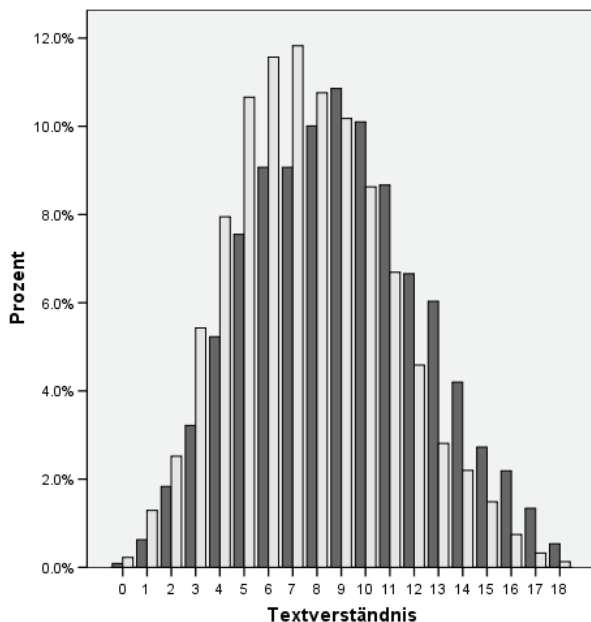
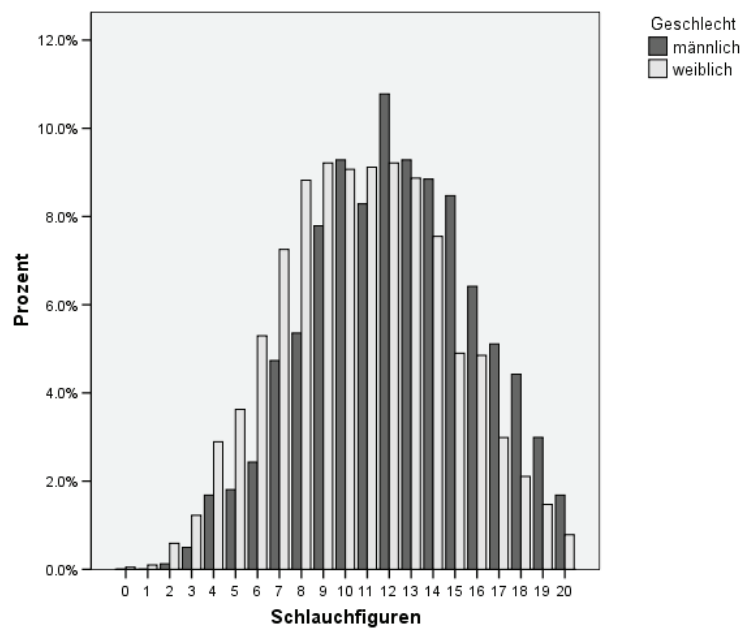
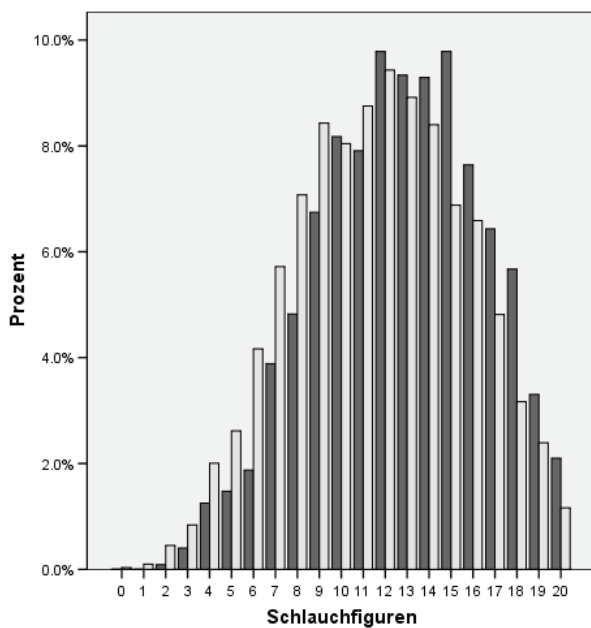
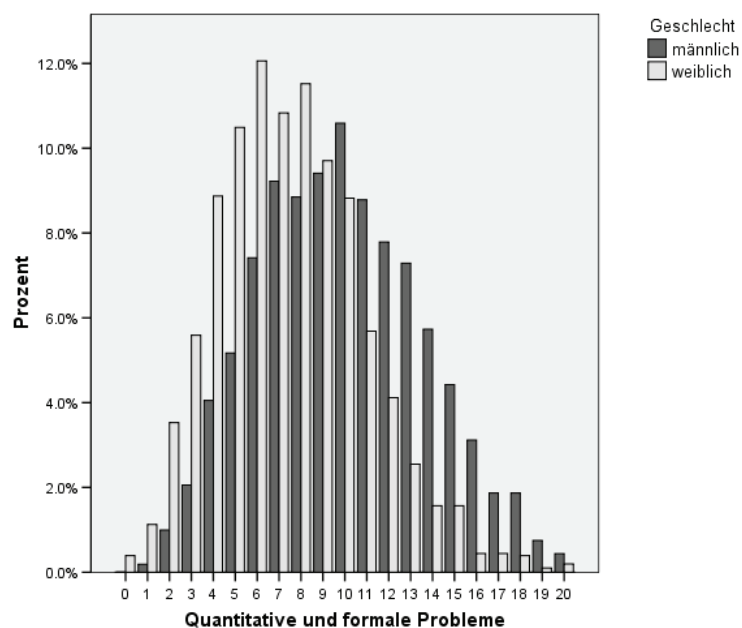
Abbildung 22: Geschlechterspezifische Unterschiede Schweiz 1999 bis 2006 und Vergleich mit den Ergebnissen in Österreich

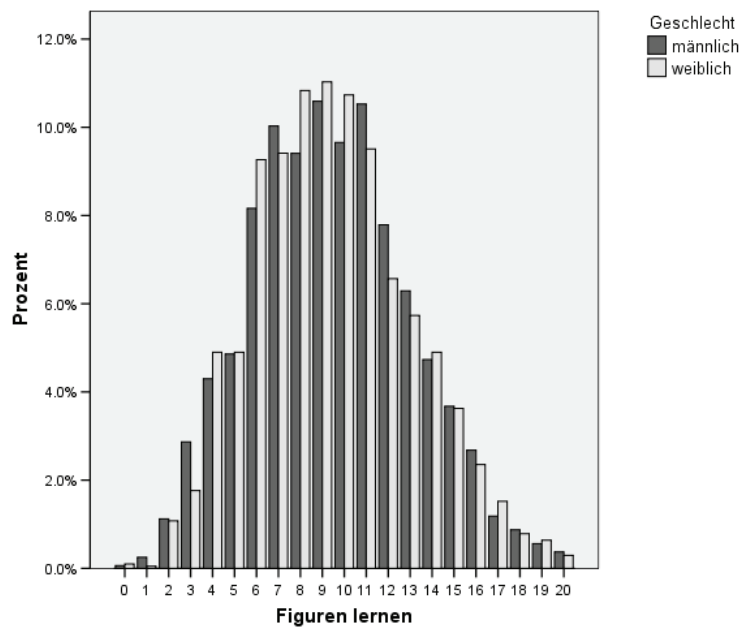
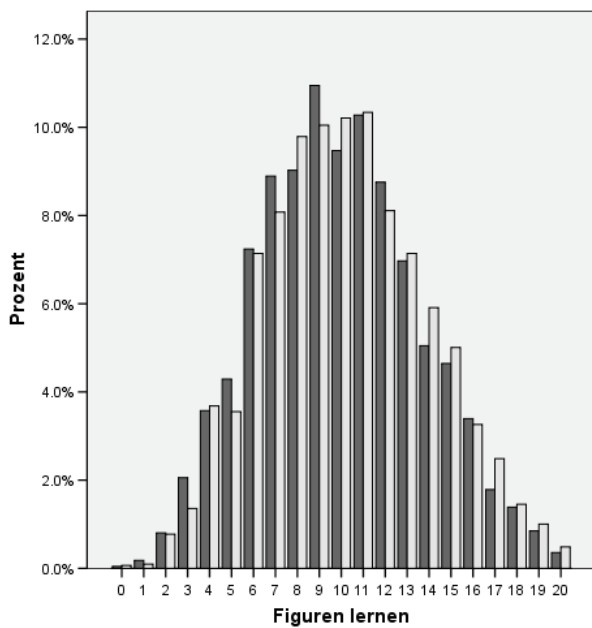
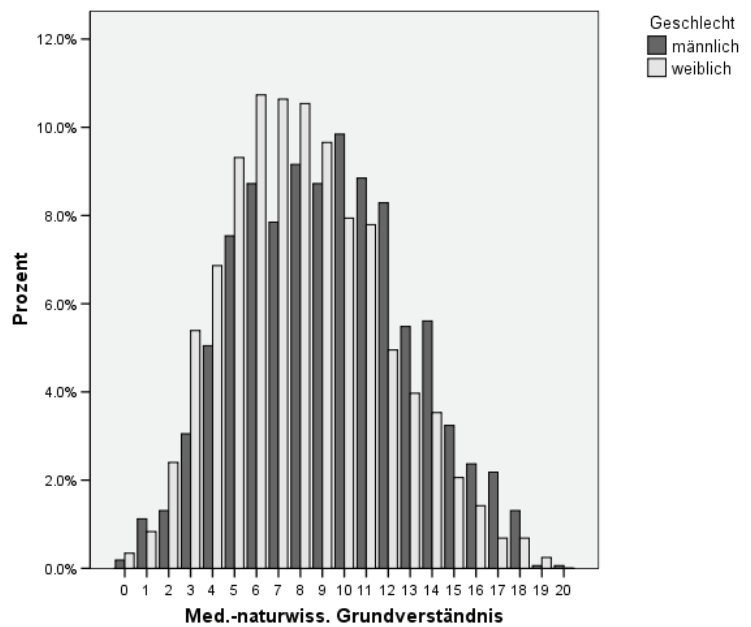
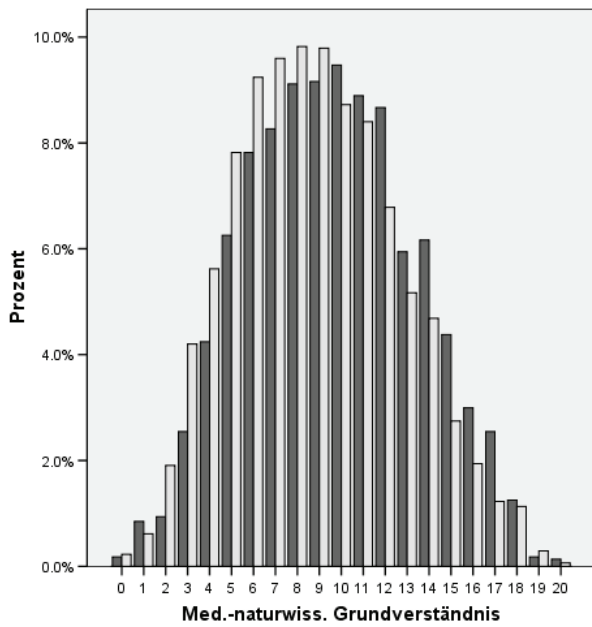
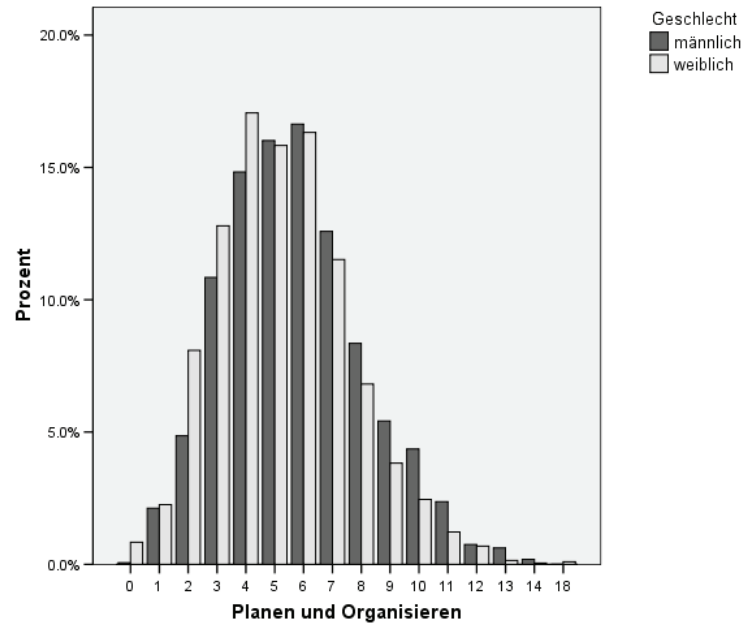
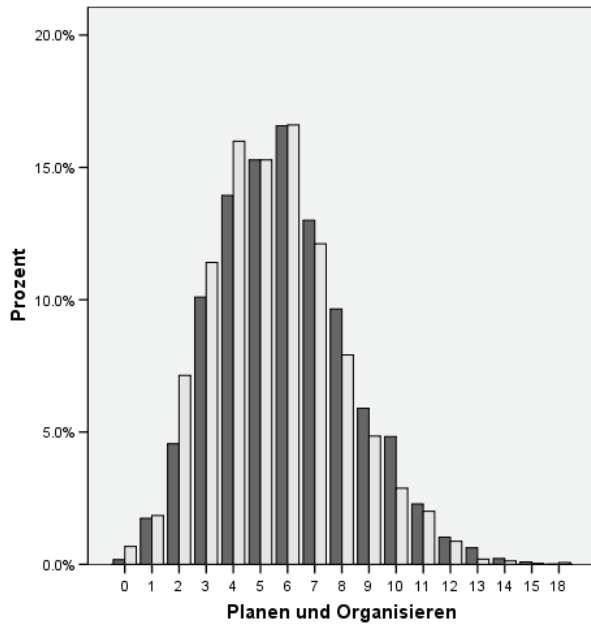
In den nachfolgend dargestellten Häufigkeitsverteilungen (**links** jeweils die Gesamtstichprobe, **rechts** nur die Österreich-Quote) wird deutlich, dass die Verteilungen für Männer und Frauen bei Vorhandensein deutlicherer Unterschiede gesamthaft verschoben sind.

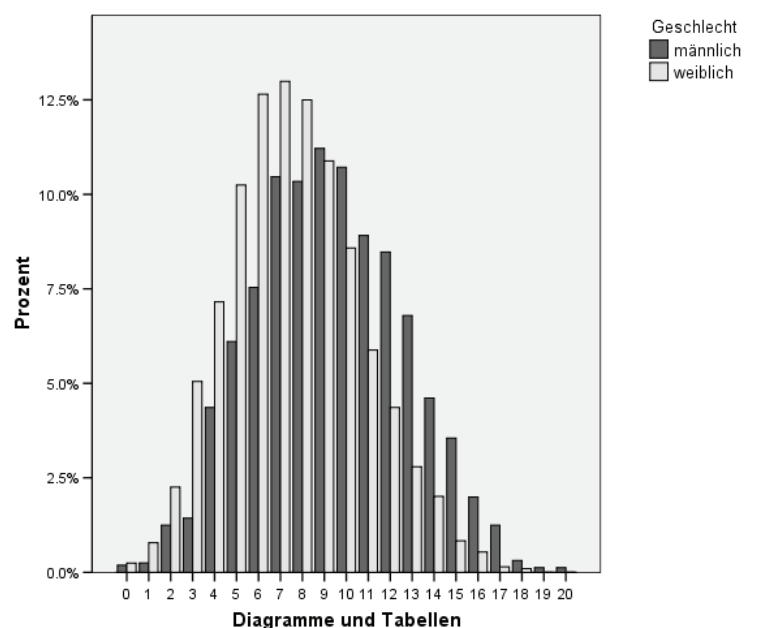
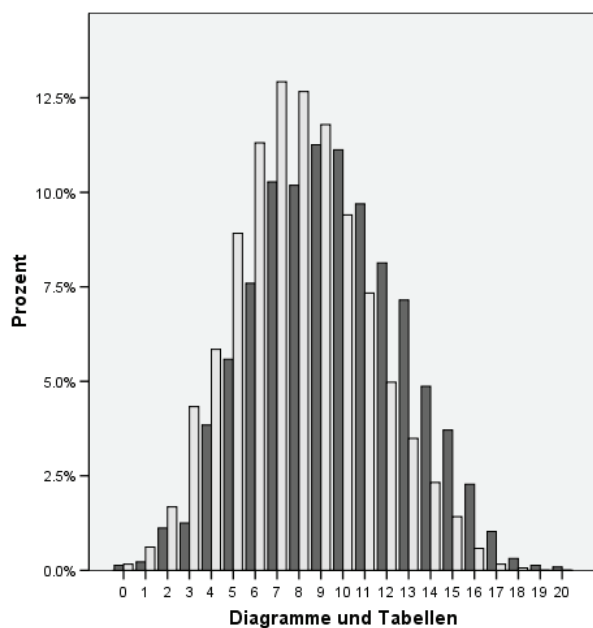
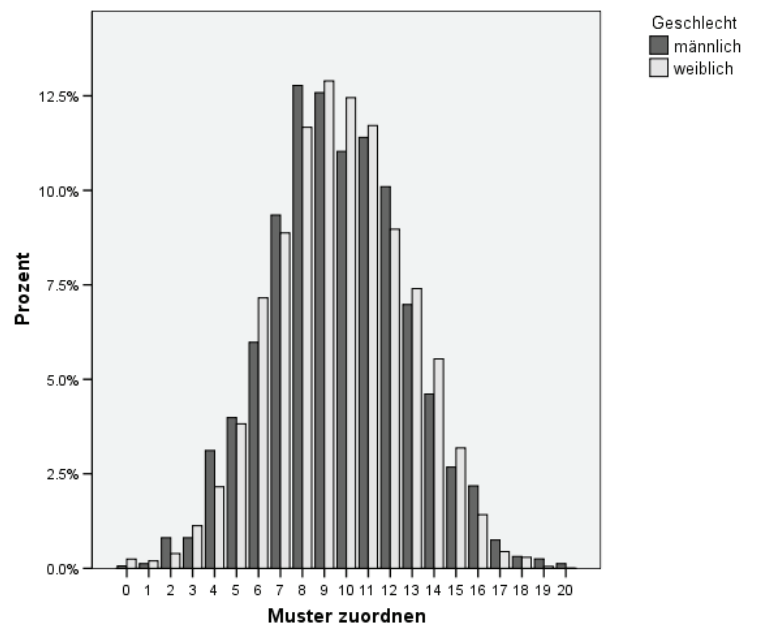
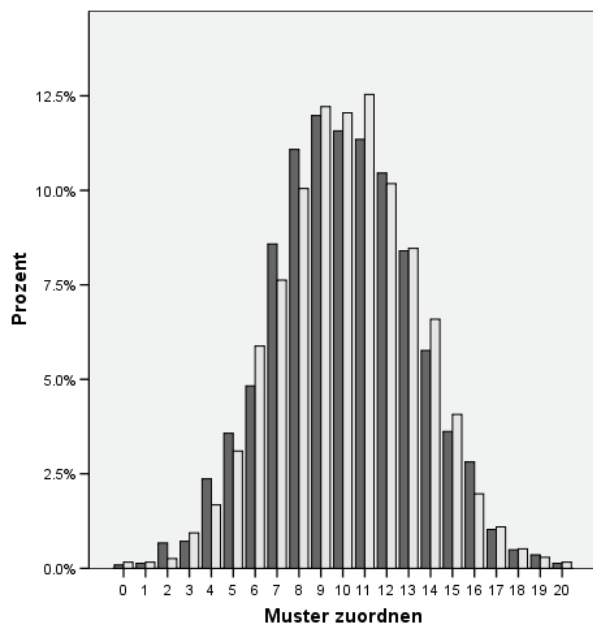
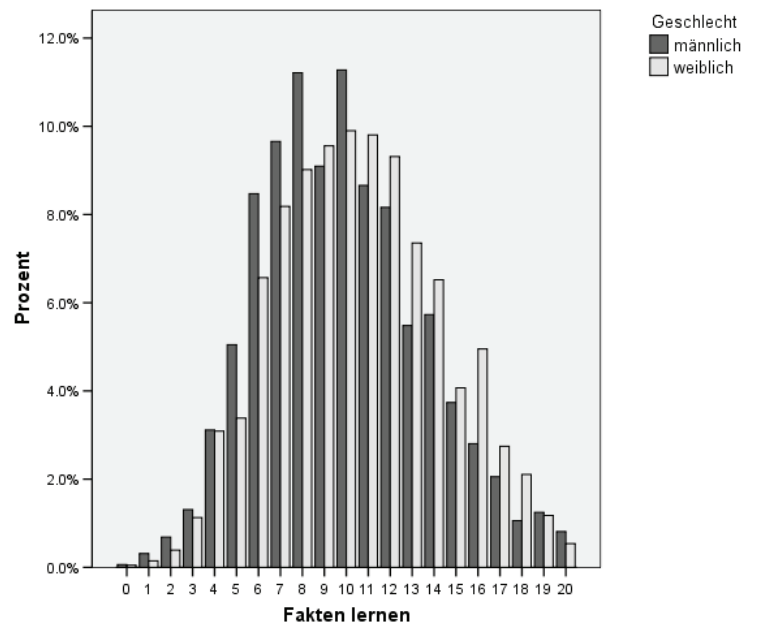
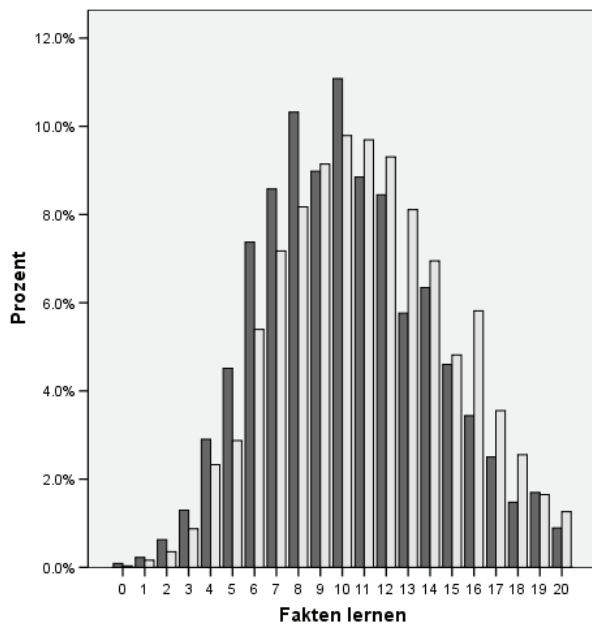
Gesamtstichprobe



Nur Österreich







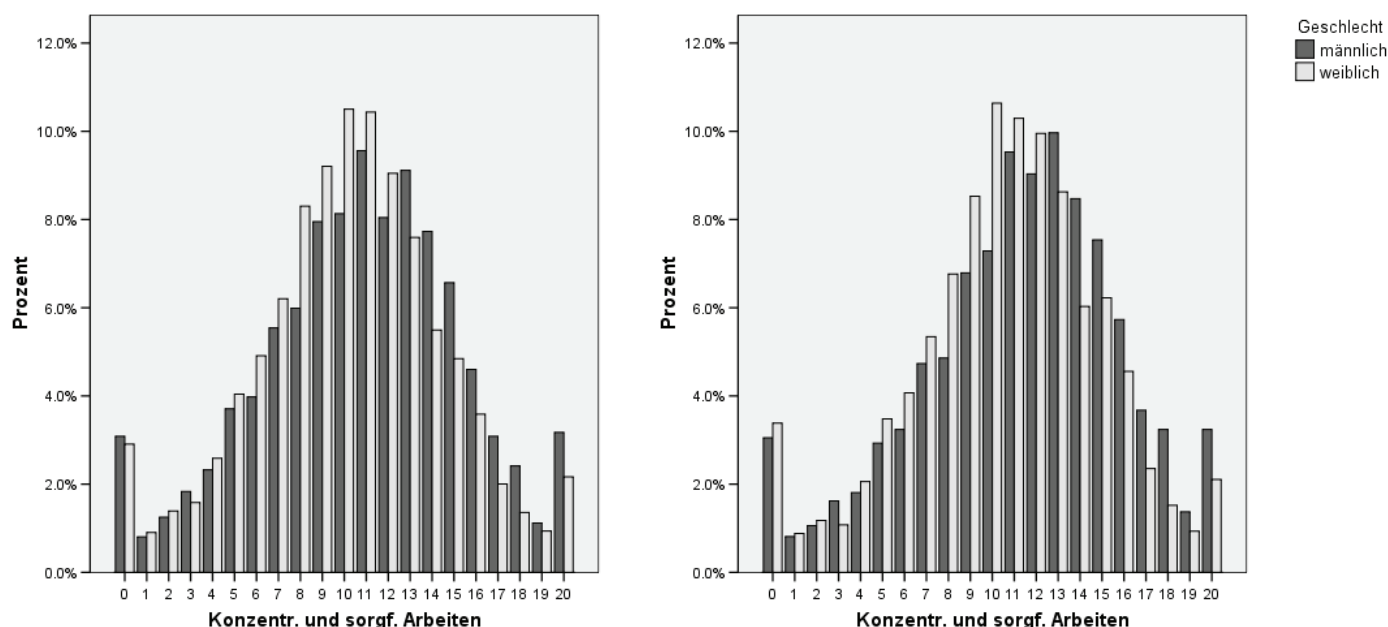


Abbildung 23: Häufigkeitsverteilungen für Punktwerte der Untertests, links Gesamtstichprobe, rechts Österreich-Quote.

In zwei zusätzlichen Analysen werden die Untertests zusammengefasst:

- nach der **Sprachabhängigkeit**: Als sprachabhängige Untertests gelten „Quantitative und formale Probleme“, „Textverständnis“, „Planen und Organisieren“, „Medizinisch-naturwissenschaftliches Grundverständnis“, „Diagramme und Tabellen“ und „Fakten lernen“; als gering sprachabhängig „Schlauchfiguren“, „Figuren lernen“, „Muster zuordnen“ und „Konzentriertes und sorgfältiges Arbeiten“.
- nach **Vormittags- oder Nachmittagsteil**. Der Test soll gewollt auch eine hohe Beanspruchung simulieren. Wenn sich hier Unterschiede tageszeitspezifisch zeigen, kann es unterschiedliche Belastungsverläufe geben.

	Personenzahl	Mittl. Testwert	Standardabweichung
Sprachabhängige Untertests			
EU sprachabhängig ml	702	105.4	9.7
EU sprachabhängig wl.	825	100.3	8.6
AT sprachabhängig ml.	862	101.9	10.2
AT sprachabhängig wl.	1162	95.9	8.7
Nicht/wenig sprachabhängige Untertests			
EU wenig sprachabhängig ml.	702	101.8	9.7
EU wenig sprachabhängig wl.	825	101.0	9.6
AT wenig sprachabhängig ml.	862	100.1	10.1
AT wenig sprachabhängig wl.	1162	98.7	9.7
Vormittagsteil			
EU Vormittag ml.	702	104.9	9.4
EU Vormittag wl.	825	99.8	9.0
AT Vormittag ml.	862	102.0	10.0
AT Vormittag wl.	1162	96.4	9.1
Nachmittagsteil			
EU Nachmittag ml.	702	102.7	9.5
EU Nachmittag wl.	825	101.6	9.4
AT Nachmittag ml.	862	100.2	10.2
AT Nachmittag wl.	1162	97.8	9.6

Tabelle 20: Sprachabhängigkeit und Durchführungszeit geschlechtsspezifisch für die Gruppen EU und Österreich

Für die vier Teil-Testwerte (sprachabhängig, wenig sprachabhängig, vormittags, nachmittags) erfolgte, wie für den Testwert auf Mittelwert und Standardabweichung aller Teilnehmer, eine Standardisierung, so dass auch für die Teilwerte der Mittelwert 100 und die Standardabweichung 10 ist.

Vergleicht man die Gruppe der sprachabhängigen gegenüber den weniger sprachabhängigen Untertests (Punktwerte), zeigt sich, dass der deutlich größere Anteil der Unterschiede auf die sprachabhängigen Untertests zurückzuführen ist. Hierbei muss man allerdings beachten, dass die sprachabhängigen Tests hauptsächlich den Faktor „Schlussfolgerndes Denken“ repräsentieren – die Unterschiede nicht nur in reinen Sprachkompetenzunterschieden liegen müssen.

Die Unterschiede scheinen am Vormittag deutlicher als am Nachmittag; sie „nivellieren“ sich aus. Dies könnte dafür sprechen, dass am Nachmittag die Belastung für beide Geschlechter vergleichbar wurde. Die „initiale“ Belastung der Frauen war am Vormittag größer als die der Männer.

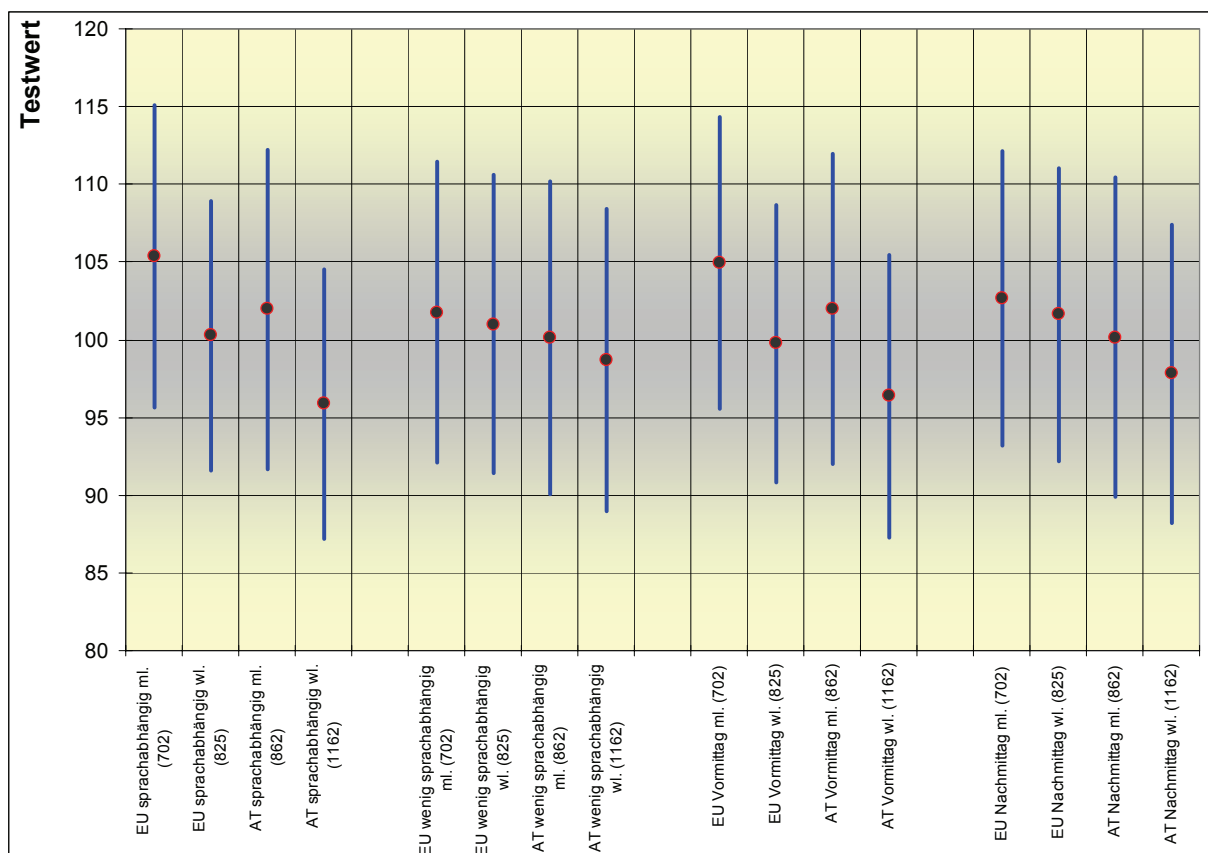


Abbildung 24: Sprachabhängigkeit und Durchführungszeit geschlechtsspezifisch: Mittelwert und Standardabweichungen m+s, m-s

Eine weitere Hypothese war, dass die Geschlechter unterschiedliche Strategien anwenden könnten oder durch die Testabnahmebedingungen unterschiedlich beeindruckt worden sind. Bekanntlich können selten alle Aufgaben eines Untertests bearbeitet werden. Es wird u.a. in der Test Info empfohlen, die nicht bearbeiteten Aufgaben dennoch durch Raten zu beantworten. Diese liegen meist am Ende der Untertests.

Auf den ersten Blick werden durch Frauen im Mittel etwas mehr Aufgaben dennoch nicht beantwortet (EU 4.5 zu 3.0, AT 4.6 zu 4.1 jeweils Frauen zu Männern). Der Unterschied für EU ist statistisch signifikant, für Österreich allerdings nicht. Ein Unterschied von 2 Aufgaben würde bei 5 Antwortstufen aufgrund der Ratewahrscheinlichkeit allerdings nur einen mittleren Unterschied von 0.4 Punkten bedeuten.

Fasst man jeweils die ersten 10 Aufgaben eines Untertests zusammen und vergleicht man sie mit den restlichen 8 bis 10 Aufgaben, ist der Unterschied in beiden Zusammenfassungen vorhanden – also auch in der ersten Hälfte der Aufgaben – und wesentlich deutlicher als durch einfache Ratewahrscheinlichkeit erklärbar.

			n	m	s
EU	Items 1-10	männlich	702	53.46	10.91
		weiblich	825	50.21	10.68
	Items 11+	männlich	702	35.32	9.34
		weiblich	825	31.47	8.88
Österreich	Items 1-10	männlich	862	49.97	11.77
		weiblich	1162	44.79	11.69
	Items 11+	männlich	862	32.48	9.94
		weiblich	1162	28.68	8.40

Tabelle 21: Unterschiede für Teil-Testwerte erste 10 Items und restliche Items zusammengefasst nach Geschlecht.

In der Schweiz wird das so genannte DIF-Verfahren zur Identifikation von Items angewendet, die in den drei Sprachgruppen unterschiedlich funktionieren (siehe Hänsgen und Spicher 2006).

Diese Methodik wurde ursprünglich verwendet, um sogenannte „kulturfaire“ Tests zu konstruieren und Items mit einem „Bias“ auszuschließen. Sie kann hier auch zur Identifikation biasbehafteter Einzel-Items benutzt werden, die von Männern und Frauen unterschiedlich aufgefasst werden.

Tatsächliche Leistungsunterschiede äußern sich eher in systematischen Unterschieden über alle Aufgaben eines Untertests. Einzelne Items, die einen Bias für eine bestimmte Gruppe aufweisen, weichen als „Ausreißer“ von diesem Trend ab. Grundidee des Verfahrens ist der Vergleich der Rangfolge der Aufgabenschwierigkeiten. Steht eine Aufgabe in Gruppe A am unteren, in Gruppe B aber am oberen Ende der Schwierigkeitshierarchie, so spricht dies dafür, dass Gruppe B durch diese eine Aufgabe in irgendeiner Art benachteiligt war. Items, die sich zwischen den Geschlechtern systematisch unterscheiden, sollten ebenfalls einen DIF-Effekt zeigen.

Vergleicht man die Leistungen der männlichen und der weiblichen Kandidierenden mittels DIF-Methode, so weisen lediglich drei Aufgaben auffällige Werte zugunsten der Männer auf. Die betroffenen Aufgaben und das Ausmaß der Differenz (in Punkten):

Aufgabe	Untertest	p _{männlich}	p _{weiblich}	p _{erwartet} *	Differenz zu real / erwartet
4	Quantitative und formale Probleme	.61	.39	.47	.22 / .08
29	Schlauchfiguren	.70	.54	.62	.16 / .08
133	Fakten lernen	.47	.41	.50	.05 / .09

* Durch Regression ermittelt

Insgesamt könnten also nur maximal 0.25 Punkte (theoretischer Erwartungswert) Unterschied zwischen männlichen und weiblichen Teilnehmern auf geschlechtsspezifische Eigenheiten des Tests zurückzuführen sein. Im Vergleich dazu wurde in der Schweizerischen Stichprobe nur eine Aufgabe (Nr. 9) gefunden, welche DIF aufweist. Dort wurden allerdings die weiblichen Kandidatinnen leicht bevorteilt (0.02 Punkte).

Eine inhaltliche Analyse der betroffenen Aufgaben ergibt keine stichhaltigen Hinweise auf mögliche Gründe. Beim Fakten Lernen, bei der Aufgabe 133 geht es um die richtige Zuordnung eines Berufes zu einem 40jährigen Patienten, die Schlauchfiguren der Aufgabe 29 unterscheiden sich nicht wirklich von den anderen – außer dass die Schwierigkeit erst auf den zweiten Blick erkennbar ist und Aufgabe 4 ist eine vergleichsweise einfache Proportionalrelation, die sich mit Blutzuckerkonzentrationen beschäftigt.

Zusammenfassende Diskussion zum Vergleich der Geschlechter:

Die vorhandenen Testwert-Unterschiede für Männer und Frauen beim EMS in Österreich sind signifikant in allen Untergruppen und betragen durchschnittlich 40% der Standardabweichung. Es erschienen 44% Männer und 56% Frauen zum Test – aufgrund des Testwertes zugelassen werden 55% Männer und 45% Frauen. Die „Verschiebung“ um 11% entspricht total immerhin einer Personenzahl von 126 Personen.

Für diesen Effekt können mehrere Faktoren ggf. zugleich verantwortlich sein:

Dass dieser Effekt **testbedingt** (Konzept) ist, kann zumindest als unwahrscheinlich gelten.

- In der Schweiz wurde 2006 der identische Test zur gleichen Zeit eingesetzt, hier ist der Einfluss des Geschlechts auf den Testwert bei gleichzeitiger Berücksichtigung von Alter und Disziplin statistisch nicht signifikant.
- Dies war ebenfalls so, als Teile dieses Tests erstmalig in einem der Vorjahre eingesetzt wurden.
- Auch in Deutschland gab es allgemein und konkret beim Einsatz dieser Testform keine Differenzen in diesem Ausmaß.
- Bei der Konstruktion des TMS und aller Weiterentwicklungen beim EMS wurde von den allgemeinen Studienanforderungen des Medizinstudiums ausgegangen, die für beide Geschlechter gleichermaßen gelten. Laut Auskunft des ITB Bonn als Autor des deutschen TMS wurde bei der 1. Revision auch mit darauf geachtet, dass der Testwert ausgewogen gegenüber Geschlechtseinflüssen ist (es wurden ursprünglich mehr Untertests erprobt).

Tatsächlich vorhandene Unterschiede: Es kann weitgehend ausgeschlossen werden, dass die durch den EMS gemessenen Merkmale der Eignung „genuin“ ein bestimmtes Geschlecht bevorzugen, indem beispielsweise Frauen grundsätzlich nicht über gleiche mathematisch-naturwissenschaftliche Kompetenz verfügen können. Moderne Konzepte zum Gender-Mainstreaming gehen davon aus, dass es sich bei solchen Vermutungen um Stereotypisierungen (sozial gefestigte, undifferenzierte und verallgemeinernde Wahrnehmungsmuster)¹ handelt. Dies wird beispielsweise auch dadurch belegt, dass bei den Unterschieden in den Basiskompetenzen der Pisa-Studie 2003 für Österreich in Mathematik, den Naturwissenschaften und beim Problemlösen keine signifikanten Geschlechterunterschiede auftreten, die Mädchen in den letzten beiden Bereichen im Mittel sogar 3 Punkte mehr erreichten als die Jungen (Prenzel u.a. 2003).

Man ist sich wohl einig, dass sich keine Unterschiede zwischen den Geschlechtern in der allgemeinen Intelligenzhöhe gibt, es aber trotzdem Unterschiede hinsichtlich der einzelnen kognitiven Leistungen geben kann. So würden Frauen im Durchschnitt über etwas bessere verbale Fähigkeiten – Männer über bessere räumlich-visuelle Vorstellung und quantitativ-mathematisches Denken verfügen (zusammenfassende Darstellung von Dlugosch 2005). Für viele dieser Untersuchungen und auch die Befunde anderer Eignungstests, wie SAT, GRE, GMAT, TMS usw. gilt aber, dass es sich um keine repräsentativen Stichproben handelt, sondern diverse Selektionsprozesse bereits vorab wirken. Insofern ist der oben zitierte Befund der PISA-Studie ggf. höher zu bewerten.

Die DIF-Analyse zeigt, dass es auch keine **einzelnen** Items gibt, die einen „Item-Bias“ aufweisen, der auf unterschiedliches „Funktionieren“ der Items für beide Geschlechter hindeutet. Die Unterschiede betreffen die gemessenen Fähigkeiten generell.

Die unterschiedliche **Repräsentativität** der Medizinbewerbenden für die jeweiligen Kohorten (Wohnbevölkerung oder Maturanden insgesamt) kann aufgrund fehlender Daten nicht

¹ <http://www.gender-mainstreaming.net/gm/Wissensnetz/instrumente-und-arbeitshilfen,did=16604.html>

abschließend beantwortet werden. Unterschiedliche **Selbst-Selektionsprozesse** bei der Studienwahl bleiben wahrscheinlich. Hier könnte es aber einen Unterschied zwischen den Quoten geben: Entsprechend der Befunde beim TMS in Deutschland werden Frauen durch die Abiturnote gegenüber Männern leicht bevorteilt (z.B. Trost u.a. 1997). Beim deutschen NC werden bisher vor allem die Abiturnoten berücksichtigt und es liegt nahe, dass bei der Zulassung in Deutschland Frauen bereits stärker berücksichtigt wurden und gleich geeignete Männer nicht zugelassen worden sind, die sich nun in Österreich beworben haben.

Es gibt allerdings Hinweise, dass sich die Fähigkeiten zwischen Männern und Frauen im Medizinstudium im Mittel in Österreich tatsächlich unterscheiden: Haidinger, Frischenschlager und Mitterauer (2006) untersuchten 2002/2003 und 2003/2004 den Zusammenhang zwischen Prüfungserfolg und mehreren Prognosekriterien für die Medizinische Universität Wien. Einer der vier gefundenen Erfolgsprädiktoren in beiden Jahrgängen war das Geschlecht, Männer waren erfolgreicher. Diese Ergebnisse wurden in der politischen Diskussion sehr schnell als „Diskriminierung“ von Frauen verallgemeinert¹. **Sollte es diesen Unterschied bezüglich der Fähigkeiten aber tatsächlich geben, müsste auch der Test diesen Unterschied ebenso zeigen, wenn er den Studienerfolg unverzerrt prognostiziert. Insofern würden sich Test- und Prüfungsergebnisse damit auch gegenseitig bestätigen.**

Bevor man diese Schlussfolgerung zieht, muss aber noch eine andere sehr wahrscheinliche Quelle von Unterschieden analysiert werden: Der EMS wurde bekanntlich sehr kurzfristig in Österreich eingeführt. Ein weiterer Faktor können die spezifischen **Abnahmebedingungen** des Tests sein. In der Schweiz (und auch seinerzeit in Deutschland) fand die Testabnahme dezentral statt, die Testlokale waren am Testtag bequem ohne notwendige Übernachtung erreichbar. Die Testlokale waren auch wesentlich kleiner und die Randbedingungen (Einlass, andere Kontrollen) konnten individueller gestaltet werden. Auch die individuelle Vorbereitung war langfristiger möglich. Es gab keine zeitlichen Verzögerungen, die ebenfalls als Stressfaktoren zu werten sein können.

Viele Personen mussten in Österreich vorab anreisen und übernachten (teilweise Jugendherbergen mit Mehrbettzimmern oder Camping), was ein zusätzlicher Stressfaktor sein kann. Beim Einlass und den Kontrollen entstand Stau und eine Verzögerung, was ebenfalls Stress bewirken kann. Die Abhängigkeit der Geschlechterunterschiede von der erlebten Tagesform und die Angleichung der Leistungen am Nachmittag wären Indizien, dass Frauen stärker durch die Testsituation beeindruckt waren als Männer. Eine Literaturrecherche zu dieser unterschiedlichen Stressempfindlichkeit von Männern und Frauen zeigt, dass Frauen unterschiedlich und teilweise nachhaltiger auf Stress reagieren. Insofern könnte dies die Unterschiede erklären helfen. Ein Beleg dafür könnte sein, dass die Unterschiede beim Vormittagsteil größer waren – sich am Nachmittag eher nivellierten.

Belastung in der Testsituation (Zeitdauer, Ablauf) gehört allerdings gewollt mit zum Testkonzept und ist kein wirklicher „Störfaktor“. In Medizinstudium und -beruf sind entsprechende Anforderungen später ebenfalls vorhanden. Was zu diskutieren wäre, ist die geschlechtsspezifisch unterschiedliche Wirkung dieses Belastungsfaktors in Österreich.

Indizien für die differenzielle Wirkung der Abnahmebedingungen wären auch, dass der Geschlechtsunterschied beim Testwert geringer ist für Personen, die sich länger auf den Test vorbereitet haben und ihn weniger schwer erlebten sowie eine bessere Tagesform angaben.

¹ z.B. siehe <http://beschraenkt.at/beschraenkt/aktuelles/114123926385>

Schlussfolgerungen für zukünftige Testabnahmen und Zulassungsverfahren

Wenn die festgestellten Geschlechterunterschiede bei der Testabnahme in Österreich **keine tatsächlich unterschiedlichen** individuellen Eignungen ausdrücken würden, die durch unterschiedliche Selbst-Selektionsprozesse oder tatsächliche Eignungsunterschiede zustande kämen, wären sie im Ausmaß zu hoch.

Aufgrund der Befunde aus der Schweiz und Deutschland gibt es aktuell keinen Grund für die Annahme, dass das **Testkonzept** für Österreich verändert werden müsste. Hier wird man mehr Gewissheit erlangen, wenn die Korrelationen zwischen Testergebnis und Prüfungserfolg vorliegen, um die individuelle Vorhersagekraft des Testergebnisses für die Prüfungsleistung zu evaluieren. Dann wird man auch Antwort auf die Frage erhalten, ob es eine systematische Unterschätzung der Eignung für Frauen bei der Testabnahme 2006 gegeben hat. Der Test ist auf die Studienanforderungen in Deutschland und der Schweiz zugeschnitten (und wird auch hinsichtlich der Übereinstimmung laufend aktualisiert). Sollte die Prognosekraft in Österreich geringer sein, müssten dann auch die Studien- resp. Prüfungsanforderungen genauer verglichen werden.

Hinweisen auf unterschiedliche Selbst-Selektionsprozesse kann durch die Befragung nach dem EMS 2007 nachgegangen werden.

Wenn die **Bedingungen der Testabnahme** einen Einfluss haben, sollten folgende Maßnahmen für die Testabnahme 2007 stressreduzierend wirken:

1. Ausreichende Information über den Ablauf und dass der Ablauf auch vom Konzept her eine Belastungssituation darstellt (gehört zur Studieneignung für Medizin mit dazu). Eventuell wurde dies wegen der Kürze der Zeit zwischen Testankündigung und -durchführung nicht ausreichend klar.
2. Hinweis auf die Notwendigkeit der Vorbereitung auf den EMS anhand der veröffentlichten Materialien.
3. Hinweis, sich, so notwendig, ruhige Übernachtungsmöglichkeiten zu suchen und ausreichend zu schlafen.
4. Optimierung aller Abläufe am Testtag hinsichtlich weniger Stressinduktion.

Eine dezentrale Testabnahme in Österreich in der Form, dass jedes Testlokal am Morgen mit öffentlichen Verkehrsmitteln erreichbar ist, bleibt eine Option, um die Abnahmebedingungen anzugleichen.

Sollte die Prognosekraft des EMS in Österreich für Studienerfolg nicht wie in Deutschland und der Schweiz ausreichend hoch sein, müssen die Studienanforderungen verglichen und ggf. andere Konzepte für die Zulassung geprüft und verwendet werden.

Ist die Prognosekraft insgesamt gegeben, tritt aber der Geschlechterunterschied auch unter den optimierten Bedingungen 2007 erneut auf und entspricht **keinen tatsächlichen** individuellen Eignungsunterschieden, verbliebe die Möglichkeit, dann **begründet** eine Korrektur einzuführen. Dies müsste keine direkte Quotierung sein – es wäre auch denkbar, bei der Berechnung der Testwerte aus den Punktwerten für Männer und Frauen getrennte Mittelwerte und Standardabweichungen zu verwenden. Männer und Frauen hätten dann je den Mittelwert 100 und die Standardabweichung 10.

Abschließende Zusammenfassung der Optionen:

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

6.2 Vergleichbarkeit der Testorte

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

In Wien fand die Testabnahme in einem einzigen großen Lokal statt, in Innsbruck in insgesamt 6 Lokalen mit unterschiedlicher Größe.

		Testort				Gesamt	
		Innsbruck		Wien			
		Anzahl	% von Testort	Anzahl	% von Testort	Anzahl	% von Testort
Land bzw. Quote	EU	827	58.4%	700	31.4%	1527	41.9%
	nicht EU	4	.3%	90	4.0%	94	2.6%
	Österreich	584	41.3%	1440	64.6%	2024	55.5%
Gesamt		1415	100.0%	2230	100.0%	3645	100.0%

Tabelle 22: Anteil Zulassungsquote pro Testlokal

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

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Sektoren Innsbruck	<i>Zwischen Gruppen</i>	733.565	11	66.688	0.688	.751
	<i>Innerhalb Gruppen</i>	136058.830	1403	96.977		
	<i>Total</i>	136792.396	1414			
Sektoren Wien	<i>Zwischen Gruppen</i>	1050.051	11	95.459	0.950	.491
	<i>Innerhalb Gruppen</i>	222984.820	2218	100.534		
	<i>Total</i>	224034.870	2229			

Tabelle 23: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den Sektoren in Innsbruck und Wien, keine Unterschiede signifikant

	Testlokal	m	s	Levene's Test Varianz			t-test Mittelwert		
					F	Sig.	t	df	Sig.
Testwert	Innsbruck	101.21	9.836	<i>homogen</i>	0.116	.733	5.926	3643	<u>.000</u>
	Wien	99.21	10.025						

Tabelle 24: Vergleich der Mittelwerte des Testwertes für Wien und Innsbruck, Unterschied signifikant

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Korrigiertes Modell	20230.998(a)	5	4046.200	42.793	.000
Konstanter Term	1288721.207	1	1288721.207	13629.780	.000
Testort	72.420	1	72.420	.766	.382
Länderquote	10494.332	2	5247.166	55.495	.000
Ww. Testort * Länderquote	283.486	2	141.743	1.499	.223
Fehler	344074.260	3639	94.552		
Gesamt	36803906.000	3645			
Korrigierte Gesamtvariation	364305.258	3644			

Tabelle 25: Zweifache Varianzanalyse für Testwert nach Testort und Länderquote (a): R-Quadrat = .056 (korrigiertes R-Quadrat = .054). Effekt Testort ist nicht signifikant, nur die Länderquote klärt die Unterschiede auf.

Testlokale zusammengefasst in Innsbruck		Anzahl	Mittelwert	Standardabweichung
H2 oben	Männlich	168	102.1	9.8
	Weiblich	175	98.7	10.8
H3 unten	Männlich	86	104.6	9.5
	Weiblich	112	99.4	9.1
H3 oben	Männlich	90	104.7	9.7
	Weiblich	99	99.8	8.7
H2a	Männlich	126	103.0	9.5
	Weiblich	124	98.3	9.0
H4	Männlich	168	103.6	9.8
	Weiblich	201	100.3	9.4
H5	Männlich	21	106.0	9.7
	Weiblich	45	97.5	9.2

Tabelle 26: Mittlere Testwerte für die einzelnen Testlokale in Innsbruck

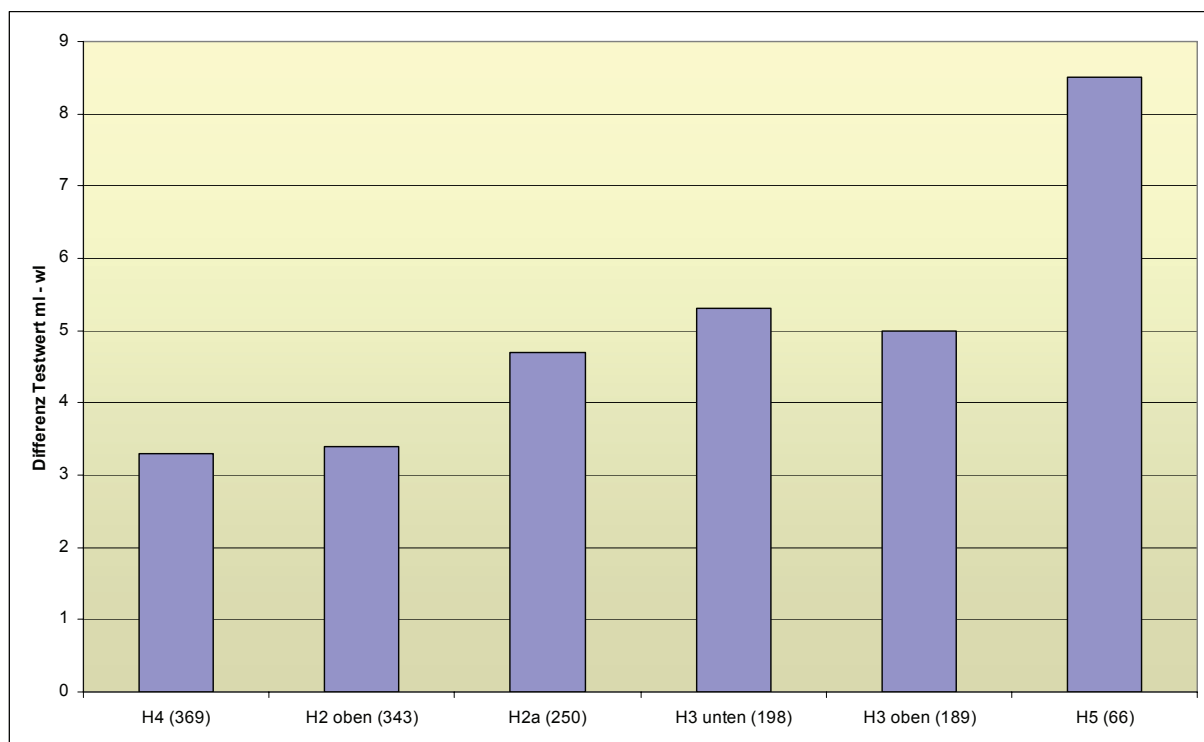


Abbildung 25: Differenz des Testwertes männlich-weiblich für die Testlokale

In Innsbruck wurden mehrere Säle verwendet, die sich nach der Größe unterschieden. Vorgreifend auf die Vergleiche für Frauen und Männer werden die Testlokale in Innsbruck nach der Größe zusammengefasst – der Geschlechterunterschied beim Testwert sinkt nicht in den kleineren Testlokalen, sondern wird dort tendenziell sogar größer. Dies spräche dagegen, dass in großen Testlokalen stressinduzierende Faktoren wirken, die die Leistungsfähigkeit von Frauen stärker beeinträchtigen.

6.3 Vergleich nach Maturitätsländerquote

In Österreich gelten je nach Maturitätsland der Testteilnehmer unterschiedliche Zulassungsquoten, wobei die Kategorien „EU“, „nicht EU“ und „Österreich“ unterschieden werden (EU: 1527; Nicht-EU: 94; Österreich: 2024). Die Leistungen dieser drei Gruppen unterscheiden sich wie in Tabelle 27 **Fehler! Verweisquelle konnte nicht gefunden werden.**dargestellt.

	Maturitäts- quote	Mittel- wert	Standard- abw.	Quadrat- summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	EU	102.26	9.35	19432.83	2	9716.42	102.61	.000	
	Nicht EU	90.62	11.53	344872.43	3642	94.69			
	Österreich	98.71	9.92	364305.26	3644				
Quantitative und formale Probleme	EU	9.15	3.63	1171.19	2	585.59	42.67	.000	
	Nicht EU	6.68	4.10	49983.93	3642	13.72			
	Österreich	8.17	3.74	51155.12	3644				
Schlauchfiguren	EU	11.92	3.82	832.68	2	416.34	28.26	.000	
	Nicht EU	9.80	4.57	53658.04	3642	14.73			
	Österreich	11.10	3.82	54490.72	3644				
Textverständnis	EU	8.87	3.37	3371.39	2	1685.69	150.51	.000	
	Nicht EU	4.72	2.65	40790.53	3642	11.20			
	Österreich	7.20	3.36	44161.91	3644				
Planen und Organisieren	EU	5.71	2.48	344.39	2	172.20	30.44	.000	
	Nicht EU	3.94	2.06	20601.59	3642	5.66			
	Österreich	5.34	2.31	20945.99	3644				
Med.-naturwiss. Grundverständnis	EU	9.46	3.61	2381.49	2	1190.75	89.38	.000	
	Nicht EU	5.68	3.34	48518.84	3642	13.32			
	Österreich	8.13	3.70	50900.33	3644				
Figuren lernen	EU	9.79	3.58	401.17	2	200.59	15.66	.000	
	Nicht EU	8.15	3.96	46636.10	3642	12.81			
	Österreich	9.26	3.56	47037.27	3644				
Fakten lernen	EU	10.61	3.67	859.52	2	429.76	31.08	.000	
	Nicht EU	7.74	3.89	50365.48	3642	13.83			
	Österreich	10.06	3.75	51225.00	3644				
Muster zuordnen	EU	9.93	2.99	289.73	2	144.87	15.52	.000	
	Nicht EU	8.50	3.49	34002.66	3642	9.34			
	Österreich	9.50	3.08	34292.39	3644				
Diagramme und Tabellen	EU	9.01	3.26	1555.23	2	777.62	73.19	.000	
	Nicht EU	5.83	2.87	38696.46	3642	10.63			
	Österreich	7.97	3.27	40251.69	3644				
Konzentr. und sorgf. Arbeiten	EU	11.14	4.32	708.30	2	354.15	18.08	.000	
	Nicht EU	8.35	5.56	71342.81	3642	19.59			
	Österreich	10.82	4.44	72051.11	3644				

Tabelle 27: 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.

6.4 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 1982 Geborenen ist 355, davon Maturität früh: 218, Maturität spät: 137, 1983-1985 geboren: 1104, nach 1985 geboren: 2186.

	Maturitätsquote	Mittelwert	Standard-abw.	Quadrat-summe	df	MQ	F	Sig.	Homogene Gruppen
Testwert	1944-1982, frühe Mat.	100.02	10.39	9145.71	3	3048.57	31.25	0.000	
	1944-1982, späte Mat.	93.95	9.63						
	1983-1985	101.78	10.14						
	1986-1989	99.46	9.70						
Quantitative und formale Probleme	1944-1982, frühe Mat.	8.89	4.13	655.68	3	218.56	15.76	0.000	
	1944-1982, späte Mat.	6.73	3.28						
	1983-1985	8.91	3.83						
	1986-1989	8.43	3.65						
Schlauchfiguren	1944-1982, frühe Mat.	10.99	4.12	927.03	3	309.01	21.01	0.000	
	1944-1982, späte Mat.	10.08	3.88						
	1983-1985	12.11	3.84						
	1986-1989	11.19	3.80						
Textverständnis	1944-1982, frühe Mat.	8.82	3.65	1452.45	3	484.15	41.27	0.000	
	1944-1982, späte Mat.	6.62	3.19						
	1983-1985	8.61	3.55						
	1986-1989	7.42	3.35						
Planen und Organisieren	1944-1982, frühe Mat.	5.45	2.42	92.97	3	30.99	5.41	0.001	
	1944-1982, späte Mat.	4.94	2.10						
	1983-1985	5.66	2.51						
	1986-1989	5.39	2.35						
Med.-naturwiss. Grundverständnis	1944-1982, frühe Mat.	9.06	3.95	691.13	3	230.38	16.71	0.000	
	1944-1982, späte Mat.	7.01	3.39						
	1983-1985	9.08	3.80						
	1986-1989	8.45	3.67						
Figuren lernen	1944-1982, frühe Mat.	9.00	3.29	242.68	3	80.89	6.29	0.000	
	1944-1982, späte Mat.	8.66	3.54						
	1983-1985	9.76	3.72						
	1986-1989	9.39	3.55						
Fakten lernen	1944-1982, frühe Mat.	9.63	3.76	653.61	3	217.87	15.69	0.000	
	1944-1982, späte Mat.	8.32	3.37						
	1983-1985	10.49	3.74						
	1986-1989	10.28	3.74						
Muster zuordnen	1944-1982, frühe Mat.	9.31	3.05	257.21	3	85.74	9.17	0.000	
	1944-1982, späte Mat.	8.71	3.16						
	1983-1985	9.96	3.16						
	1986-1989	9.60	3.00						
Diagramme und Tabellen	1944-1982, frühe Mat.	8.94	3.68	845.40	3	281.80	26.04	0.000	
	1944-1982, späte Mat.	6.42	2.66						
	1983-1985	8.79	3.33						
	1986-1989	8.20	3.26						
Konzentr. und sorgf. Arbeiten	1944-1982, frühe Mat.	10.40	4.68	429.08	3	143.03	7.27	0.000	
	1944-1982, späte Mat.	9.38	4.71						
	1983-1985	11.13	4.55						
	1986-1989	10.91	4.33						

Tabelle 28: 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.

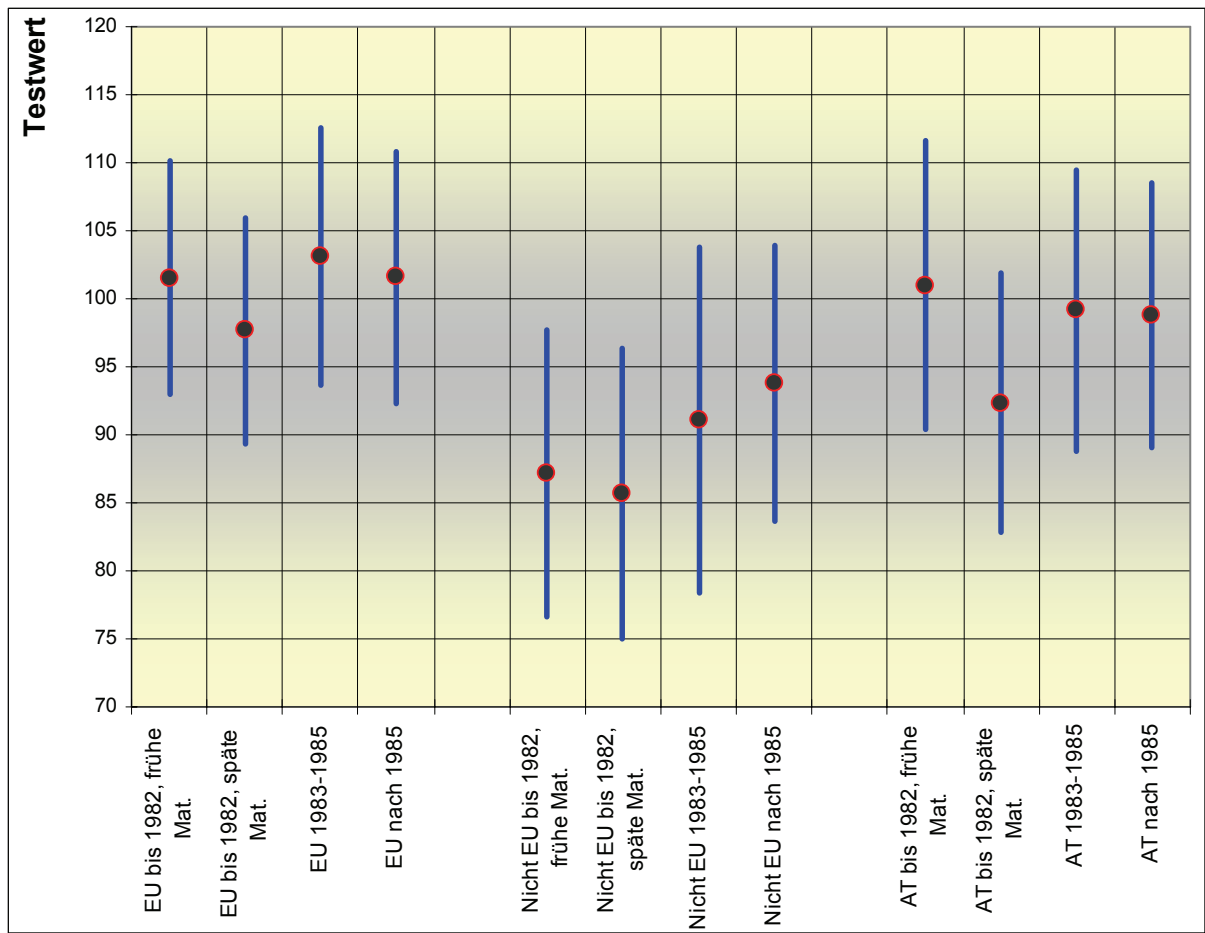


Abbildung 26: Vergleiche für Altersgruppen (Geburtsjahre) nach Maturitätsquoten

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

			Alter und Maturitätsjahr				Gesamt
			1961-1982, frühe Maturität	1961-1982, späte Maturität	1983-1985	1986-1989	
Wunsch- universität	Innsbruck	Anzahl	79	50	510	776	1415
		% von Wunschuniv.	5.6%	3.5%	36.0%	54.8%	
	Wien	Anzahl	139	87	594	1410	2230
		% von Wunschuniv.	6.2%	3.9%	26.6%	63.2%	
Gesamt		Anzahl	218	137	1104	2186	3645
		% von Wunschuniv.	6.0%	3.8%	30.3%	60.0%	

Tabelle 29: Vergleich der Kandidaten nach Wunschuniversität: Alter und Maturitätsjahr

		Innsbruck			Wien		
		n	m	s	n	m	s
EU	bis 1982, frühe Maturität	54	102.6	8.6	60	100.6	8.4
	bis 1982, späte Maturität	26	95.4	7.9	25	100.0	8.2
	1983-1985	430	104.1	9.6	378	102.0	9.3
	nach 1985	317	101.9	8.9	237	101.1	9.7
AT	bis 1982, frühe Maturität	25	98.8	10.7	59	101.9	10.6
	bis 1982, späte Maturität	22	93.0	9.6	56	92.1	9.5
	1983-1985	78	100.1	11.8	181	98.7	9.6
	nach 1985	459	99.0	9.6	1144	98.8	9.8

Abbildung 27: Altersunterschiede nach Universität

6.5 Vergleich nach Disziplin

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

		Innsbruck			Wien			Gesamtwert		
		n	m	s	n	m	s	n	m	s
Zahn- medizin	EU	47	99.4	8.3	38	95.6	8.4	85	97.7	8.5
	nicht EU	1	108.0	.	19	86.4	7.1	20	87.5	8.5
	Österreich	32	95.5	8.3	98	95.2	10.5	130	95.2	9.9
	Gesamt	80	97.9	8.5	155	94.2	10.0	235	95.5	9.7
Kom- biniert	EU	42	101.7	8.9	38	99.1	10.0	80	100.5	9.5
	nicht EU	0			4	93.3	8.7	4	93.3	8.7
	Österreich	42	98.6	11.6	86	98.8	9.3	128	98.7	10.1
	Gesamt	84	100.2	10.4	128	98.7	9.5	212	99.3	9.9
Human- medizin	EU	738	103.2	9.4	624	102.0	9.2	1362	102.6	9.3
	nicht EU	3	88.3	11.7	67	91.5	12.4	70	91.4	12.4
	Österreich	510	99.1	9.9	1256	98.9	9.8	1766	99.0	9.9
	Gesamt	1251	101.5	9.8	1947	99.6	10.0	3198	100.4	9.9

Abbildung 28: Testwert nach Wunschdisziplin und Universität

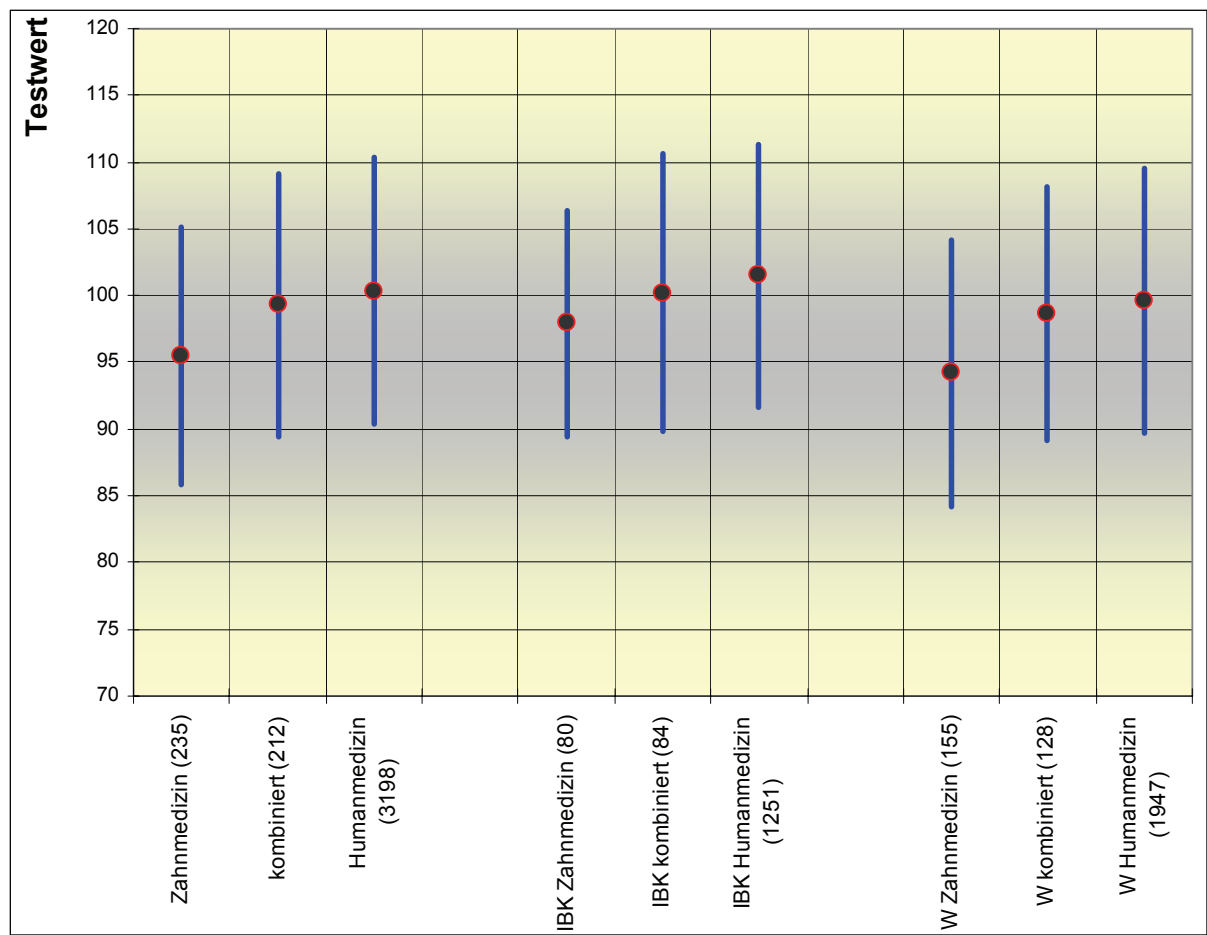


Abbildung 29: Testwert nach Disziplin Gesamt und getrennt für Innsbruck (IBK) und Wien (W)

7 Vorbereitung auf den EMS

Ein wesentliches Merkmal des EMS ist es, dass **kein spezifisches Fachwissen** vorausgesetzt wird und dass die empfohlene Vorbereitung (Studieren der Test-Info, Bearbeiten einer veröffentlichten Originalversion) hinreichend ist, das optimale Leistungsniveau zu erreichen. Darauf wurde bereits bei der Aufgabenkonstruktion geachtet. Wo nötig, wird das für die Lösung notwendige Wissen in der Aufgabe selbst mit dargestellt. Der Eignungstest soll die **zukünftige Studierfähigkeit** prüfen und nicht das heute vorhandene Wissen, was auch aufgrund der Vorerfahrungen und bereits absolvierten Ausbildungen unterschiedlich ist. Würde man diesem Wissen einen zu großen Stellenwert einräumen, wäre die Chancengleichheit deutlich infrage gestellt. Weil z.B. nichts über die Zeitdauer und Intensität bekannt ist, in der vorab bereits Fachwissen erworben wurde, hätten exzessiv vortrainierte Personen größere Chancen auf den Erwerb eines Studienplatzes. Dieser erworbene Wissensstand würde nicht automatisch bedeuten, dass dann unter den Bedingungen des Studiums mit gleicher Effektivität Wissen erworben werden kann.

Der notwendige und wirksame Aufwand ist zeitlich und finanziell vertretbar und vereinbart sich gut mit einer sozialen Verträglichkeit des EMS. Ein Ergebnis der bisherigen drei Vorbereitungsreports der Schweiz¹ war, dass Vorbereitung auf den EMS notwendig ist (Kennenlernen der Instruktionen der Untertests und Vertrautmachen mit den Anforderungen, um während der Abnahme keine zusätzliche Zeit für die Orientierung zu verlieren). Hat man das individuelle Leistungsniveau aber erreicht, bringt das Training keinen zusätzlichen Effekt mehr. Dies wurde bereits in Deutschland sowie in der Schweiz vor allem für die Wirksamkeit kommerzieller Kurse nachgewiesen.

Es konnte und kann wohl nicht verhindert werden, dass in einer für die weitere Zukunft so bedeutsamen Entscheidungssituation eine erhöhte Empfänglichkeit gegenüber allen Aussagen besteht, den Test erfolgreich zu absolvieren und einen Studienplatz zu erhalten. Einige kommerzielle Trainingsanbieter gehen in ihren Aussagen und Versprechungen sehr weit und neigen dabei auch zu nachweislichen Falschaussagen². Erschwerend wirkte sich dabei aus, dass ÖH-nahe Internetseiten der beteiligten Universitäten Werbung solcher Anbieter direkt übernommen und damit „officialisiert“ haben.

Wie sehen die Fakten aus? In Österreich wurde am Ende des Tests eine Befragung durchgeführt, die u.a. auch Stundenangaben für die Vorbereitung erfasste. Aufgrund der starken Beanspruchung durch den Test und die Freiwilligkeit des Ausfüllens müssen die Ergebnisse mit der nötigen Vorsicht interpretiert werden, dennoch sind die Ergebnisse von 3370 Personen verwendbar (114 fehlen oder sind logisch unplausibel).

	Gemeinsame Vorbereitung				Stunden Vorbereitung in Kursen				Stunden Selbstvorbereitung			
	n	m	Median	s	n	m	Median	s	n	m	Median	s
EU	365	9.8	4.0	22.8	186	32.5	36.0	18.2	1356	24.4	12.0	40.2
nicht EU	23	19.0	10.0	26.9	17	23.4	20.0	17.5	64	42.5	20.0	64.9
AT	635	8.0	3.0	20.8	315	33.9	40.0	17.8	1822	18.5	10.0	42.7
Total	1023	8.9	3.0	21.7	518	33.1	40.0	18.0	3242	21.5	10.0	42.5

Tabelle 30: Vorbereitungszeiten (gemittelt für Angaben ab 1 Stunde) für Art der Vorbereitung und Länderquote

¹ Vorbereitungsreports 2003, 2004 und 2005, siehe www.unifr.ch/ztd/ems/berichte/welcome_b.htm

² Die Auseinandersetzung dazu wird seit Beginn des EMS in der Schweiz geführt, vergleiche www.unifr.ch/ztd/ems, wo auch Argumente der Trainingsanbieter kommentiert werden.

Bezüglich aller drei Vorbereitungsarten gibt es keinen bedeutsamen Unterschied zwischen Männern und Frauen – dies kann nicht die Ursache der oben diskutierten Unterschiede sein.

	Gemeinsame Vorbereitung			Stunden Vorbereitung in Kursen			Stunden Selbstvorbereitung		
	n	m	s	n	m	s	n	m	s
männlich	426	9.7	25.0	215	33.4	17.5	1415	21.3	44.1
weiblich	597	8.3	19.0	303	32.8	18.3	1827	21.6	41.1
Gesamt	1023	8.9	21.7	518	33.1	18.0	3242	21.5	42.5

Tabelle 31: Vorbereitungszeiten nach Geschlecht

Wie in der Schweiz gibt es auch keinen eindeutigen Zusammenhang zwischen Testwert und Vorbereitungsaufwand. Zu wenig Aufwand ist mit geringeren Testwerten verbunden. Mit steigendem Aufwand werden die Testwerte aber nicht linear besser. Auch die Unterschiede nach den Schulnoten (die statistisch gesehen auch die Studieneignung schätzen), werden durch die Vorbereitungsintensität nicht nivelliert.

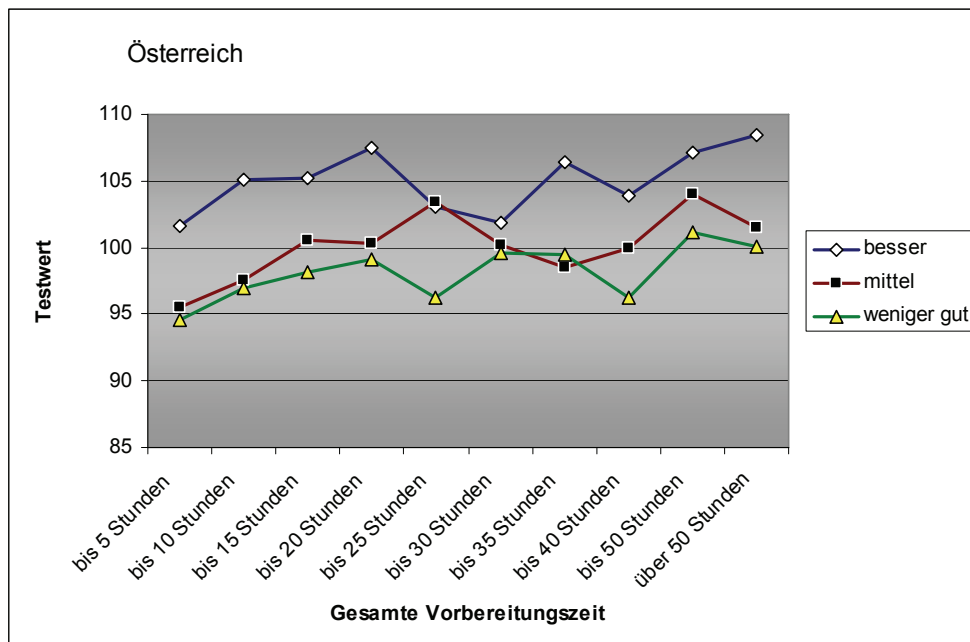


Abbildung 30: Testwert für Gesamtzubereitungszeit nach Notendurchschnitten; besser: 1 bis 1.5, mittel: 1.6 bis 2.2, weniger gut: über 2.2

Wie in der Schweiz zeigt sich, dass der Vorbereitungstyp auch keinen Einfluss auf das Ergebnis hat: Es wurden drei Personengruppen gebildet (1) nur selbständige Vorbereitung, (2) gemeinsame Vorbereitung und (3) Kursbesuch. Die Gruppe 2 kann dabei zusätzlich selbständige Vorbereitung gewählt haben, die Gruppe 3 selbständige oder gemeinsame Vorbereitung.

Es unterscheidet sich vor allem die Vorbereitungszeit zwischen den drei Personengruppen, Personen mit Kursbesuch bereiten sich deutlich länger vor. Dem steht aber kein entsprechender Unterschied beim Testwert gegenüber – alle Gruppen bleiben vergleichbar.

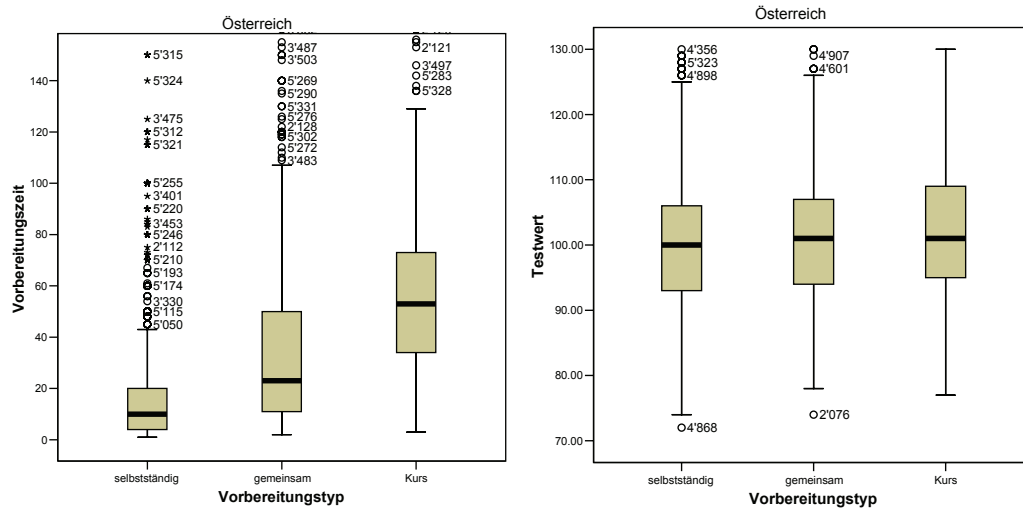


Abbildung 31: Vorbereitungszeiten und Einfluss auf den Testwert

Deutlicher wird dies in einer Regressionsschätzung (nur Österreich und mehr als 0 Stunden in die Korrelation einbezogen, für die Personen mit kommerzieller Vorbereitung). Es findet sich kein signifikanter Einfluss der Intensität der kommerziellen Vorbereitungskurse auf den Testwert, nur ein signifikanter Einfluss der zusätzlichen selbstständigen Vorbereitung.

Modell 1	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
	B	Standardfehler	Beta		
(Konstante)	99.312	1.023		97.094	.000
Stunden gemeinsame Vorbereitung	.021	.031	.042	.683	.495
Stunden Vorbereitung in Kursen	.051	.031	.101	1.677	.095
Stunden Selbstvorbereitung	.057	.020	.176	2.793	.006

Tabelle 32: Regression Testwert und der Stundeneinsatz für die drei Vorbereitungsarten

Auch die Korrelationen der einzelnen Untertests sind nicht substanziell genug, um einen Zusammenhang von Vorbereitungsaufwand und Testwert zu begründen.

		Stunden Selbstvorbe- reitung	Stunden gemeinsame Vorbereitung	Stunden kommerzielle Vorbereitung
Quantitative und formale Probleme	Korrelation nach Pearson	-.045	.001	.007
	<i>Signifikanz (2-seitig)</i>	.010	.985	.876
Schlauchfiguren	Korrelation nach Pearson	.111	.014	.134
	<i>Signifikanz (2-seitig)</i>	.000	.659	.002
Textverständnis	Korrelation nach Pearson	-.017	-.023	-.048
	<i>Signifikanz (2-seitig)</i>	.341	.461	.281
Planen und Organisieren	Korrelation nach Pearson	-.050	.017	.017
	<i>Signifikanz (2-seitig)</i>	.005	.596	.702
Medizinisch- naturwissenschaftliches Grundverständnis	Korrelation nach Pearson	-.007	-.011	-.009
	<i>Signifikanz (2-seitig)</i>	.710	.717	.842
Figuren lernen	Korrelation nach Pearson	.139	.042	.167
	<i>Signifikanz (2-seitig)</i>	.000	.177	.000
Fakten lernen	Korrelation nach Pearson	.087	-.046	.070
	<i>Signifikanz (2-seitig)</i>	.000	.143	.111
Muster zuordnen	Korrelation nach Pearson	.089	-.001	.075
	<i>Signifikanz (2-seitig)</i>	.000	.982	.090
Diagramme und Tabelle	Korrelation nach Pearson	-.032	-.032	-.039
	<i>Signifikanz (2-seitig)</i>	.066	.312	.380
Testwert Österreich	Korrelation nach Pearson	.061	-.001	.069
	<i>Signifikanz (2-seitig)</i>	.001	.981	.118
N		3242	1023	518

Tabelle 33: Korrelation der Untertestleistungen mit den Vorbereitungsstunden

8 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.

8.1 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?

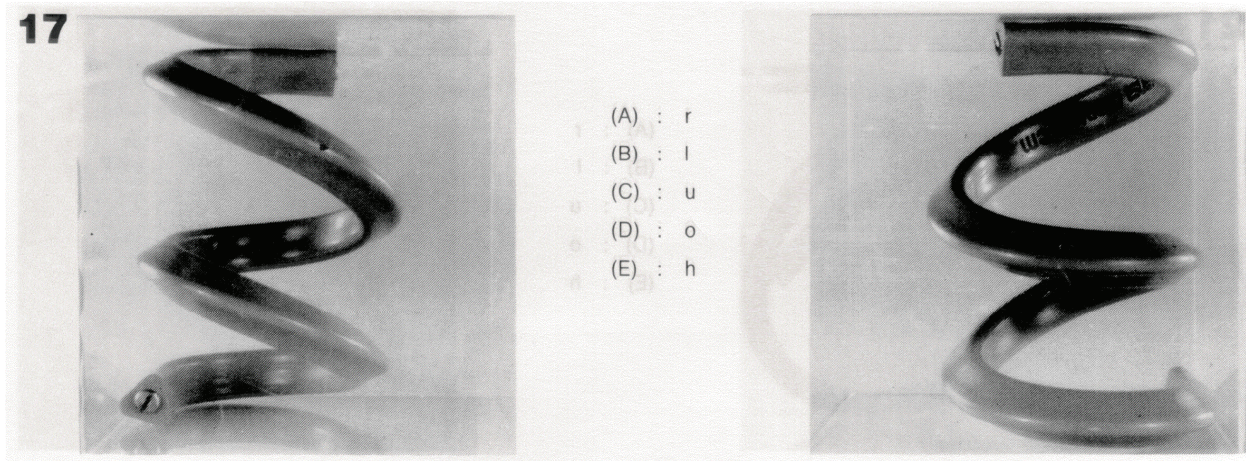
- (A) 60 BE
- (B) 25 BE
- (C) 6 BE
- (D) 5 BE
- (E) 0,5 BE

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.

8.2 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!)

8.3 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.

8.4 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.

8.5 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.

8.6 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.

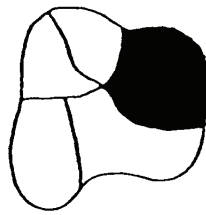
Nach dieser Analyse des Textes ist es offensichtlich, dass die Antwort (C) richtig ist.

8.7 Untertest: Figuren lernen

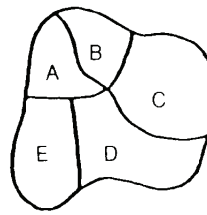
Für beide nachfolgenden Gedächtnistests wird nach der Mittagspause das Material zum Einprägen ausgeteilt. Vor der Abfrage des Gelernten wird der Untertest „Textverständnis“ bearbeitet, damit liegt die Zeit des Behaltens der gelernten Inhalte über einer Stunde. Gedächtnisleistungen sind wichtige Voraussetzungen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

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

Gezeigte Figur zum Einprägen



Gezeigte 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.

8.8 Untertest: Fakten lernen

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

Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

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

Der Patient mit dem Schädelbasisbruch ist von Beruf ...

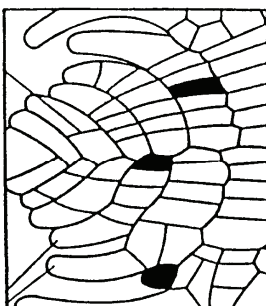
- (A) Installateur
- (B) Lehrer
- (C) Dachdecker
- (D) Handelsvertreter
- (E) Physiker

8.9 Untertest: Muster zuordnen

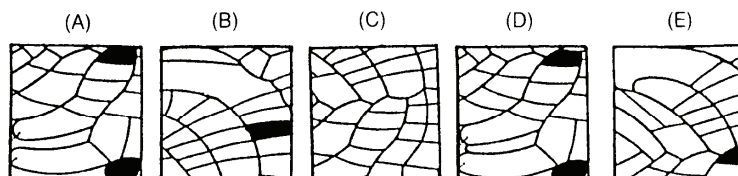
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wieder zu erkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

8.10 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.

9 Weitere Informationen zu den Bewerbungen

Neben der Eignung zum Studium ist auch die Neigung (Studienmotivation) eine wichtige Erfolgskomponente. Sie kann nicht als Selektionskriterium verwendet werden (nicht valide genug messbar, leicht verfälschbar, trainierbar), die Auseinandersetzung mit den Studienanforderungen bleibt für die Einzelnen dennoch wichtig. In der Befragung wurden daher auch diese Bereiche abgebildet. Neben den Österreichischen Ergebnissen sind auch diejenigen der Schweiz hier enthalten. Im direkten Vergleich werden Schweiz, Österreich und Deutschland sowie alle anderen Länder zusammengefasst, die Zahl der gültigen Angaben ist jeweils aufgeführt.

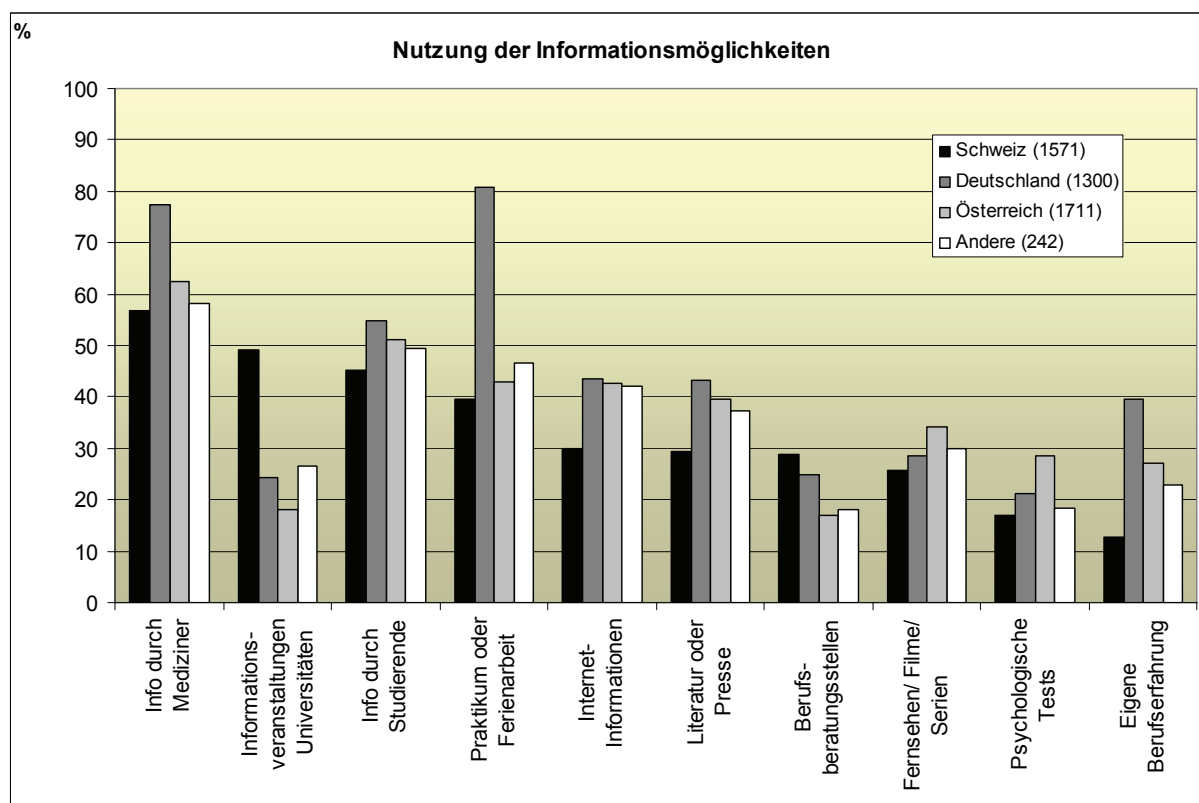


Abbildung 32: Nutzung der Informationsmöglichkeiten für Studienwahl

Die deutschen Bewerber scheinen erfahrener, haben überwiegend Praktika oder Ferienarbeit geleistet sowie mehr eigene nutzbare Berufserfahrung. In der Schweiz werden die Informationsveranstaltungen der Universitäten (Ganztagsveranstaltungen) deutlich mehr genutzt als in den anderen Ländern.

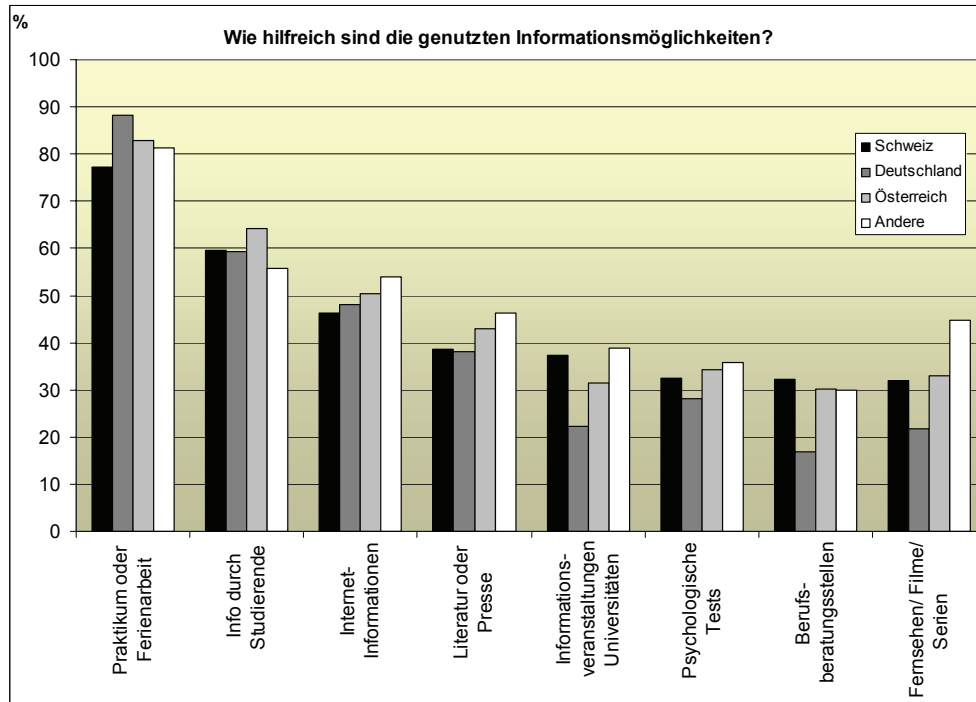


Abbildung 33: Einschätzung des Nutzens der Informationsmöglichkeiten

Praktika und Ferienarbeit sowie Informationen durch Studierende werden am hilfreichsten eingeschätzt – auch das Internet wird noch um 50% hilfreich erlebt. Trotz der höheren Nutzungsfrequenz in der Schweiz sind Berufsberatungen und Informationsveranstaltungen der Universitäten etwa gleich hilfreich wie in den anderen Ländern.

Die Informationsmöglichkeiten über den Wunschberuf werden in der Mehrzahl als ausreichend bis sehr gut eingeschätzt – ein dringlicher Nachholbedarf scheint auf diesem Sektor nicht zu bestehen.

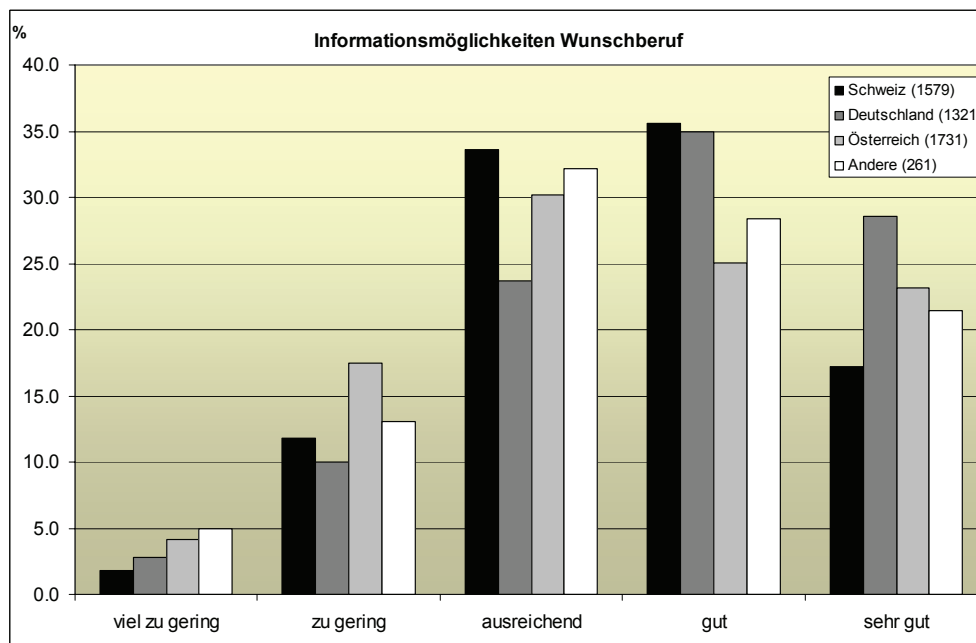


Abbildung 34: Einschätzung der Informationsmöglichkeiten über den Wunschberuf

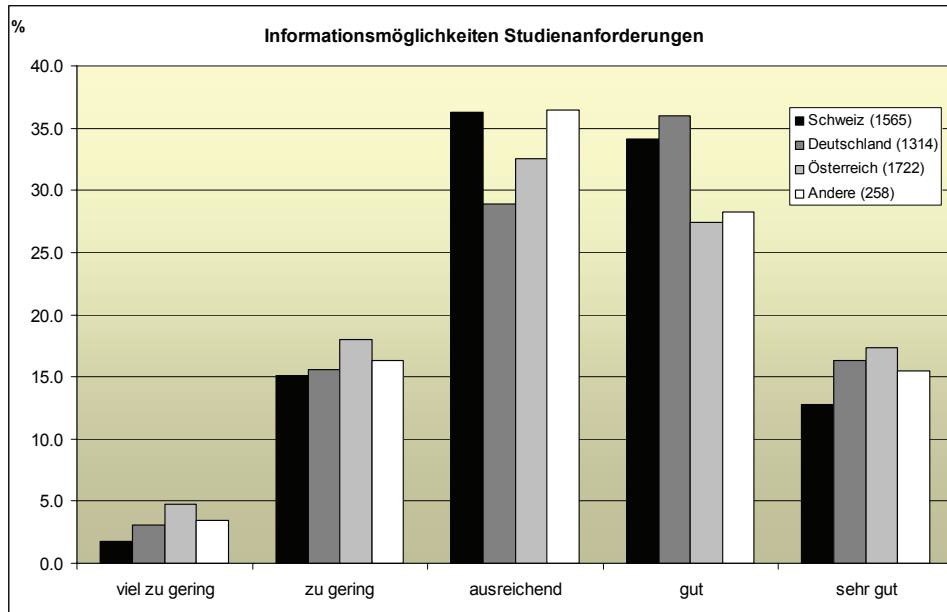


Abbildung 35: Einschätzung der Informationsmöglichkeiten über die Studienanforderungen

Auch bezüglich der Informationsmöglichkeiten über die Studienanforderungen besteht kein dringlicher Handlungsbedarf – sie werden überwiegend als ausreichend bis sehr gut erlebt.

Es wurde auch gefragt, welches Patientel, Fachgebiet und welche Einrichtung zukünftig angestrebt wird. Zu beachten ist, dass sich hier ein Grossteil der Personen noch nicht festgelegt hat. Sinn der Befragung ist es, etwas über die initiale Berufsmotivation zu erfahren und dies im Studium dann entsprechend zu berücksichtigen.

Bei der Orientierung bezüglich des späteren Patientels besteht eine Übereinstimmung zwischen den Ländern und auch zum Vorjahr der Schweiz. Die Hälfte bleibt unentschieden, Kinder und Erwachsene halten sich in etwa die Waage. Es gibt zu diesem Zeitpunkt auch einen geringen Prozentsatz von Personen, die keinen Patientenkontakt anstreben.

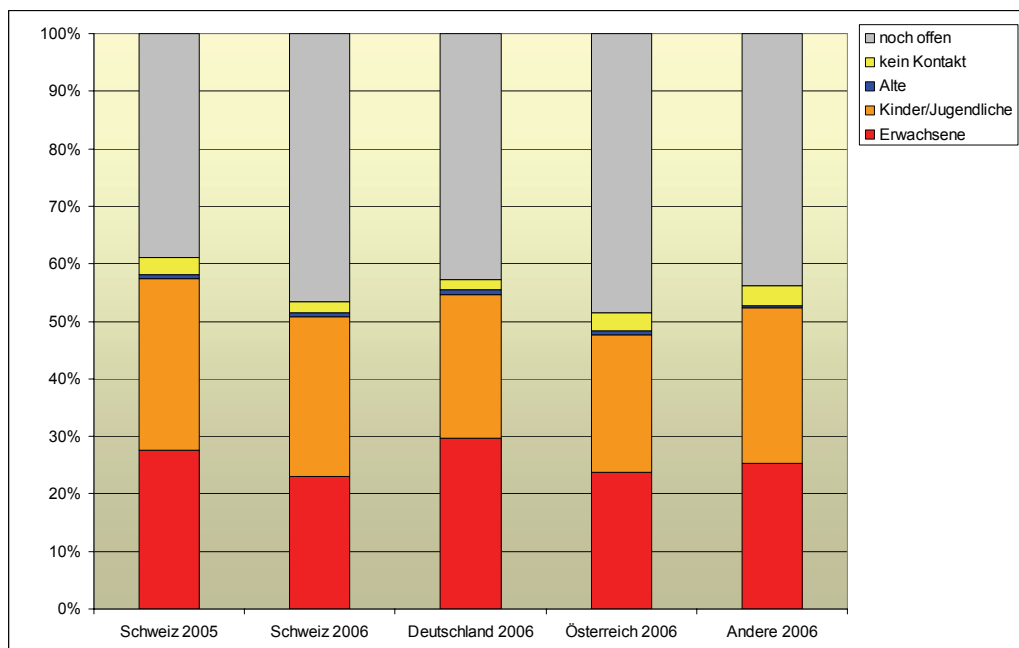


Abbildung 36: Gewünschte zukünftige berufliche Orientierung, Patientel

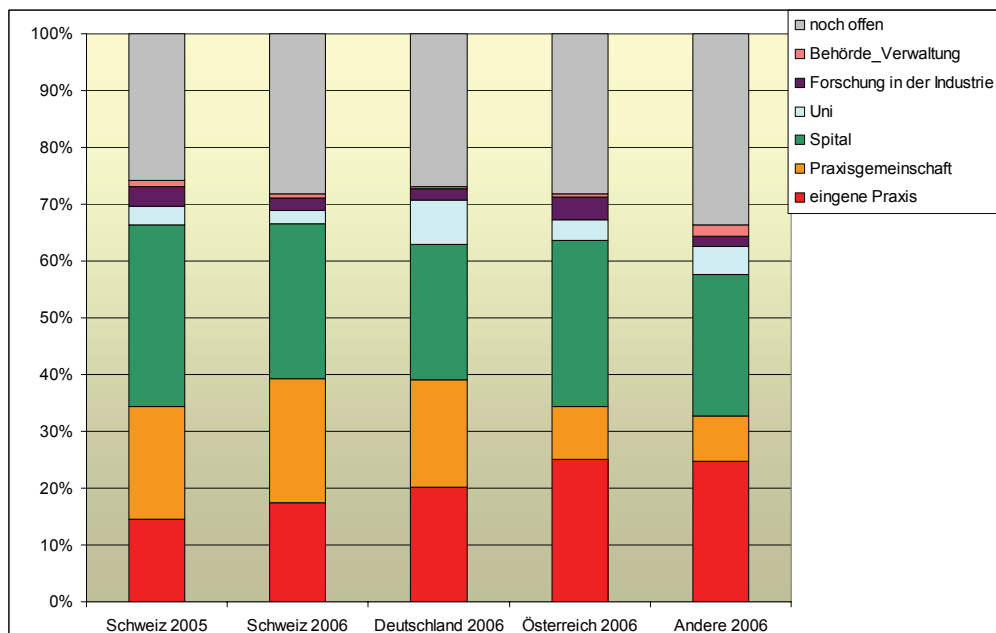


Abbildung 37: Gewünschte zukünftige berufliche Orientierung, Einrichtung

Eigene Praxis oder Praxisgemeinschaft überwiegen als angestrebtes Arbeitsgebiet. Auch Spitalarzt ist überall ungefähr gleich attraktiv.

In der Schweiz hat 2006 – wohl nicht zuletzt durch die öffentlichen Diskussionen – das Interesse an einer Tätigkeit als Hausarzt zugenommen. Die Perspektive „Spezialarzt“ bleibt weiter die anstrebenswerteste.

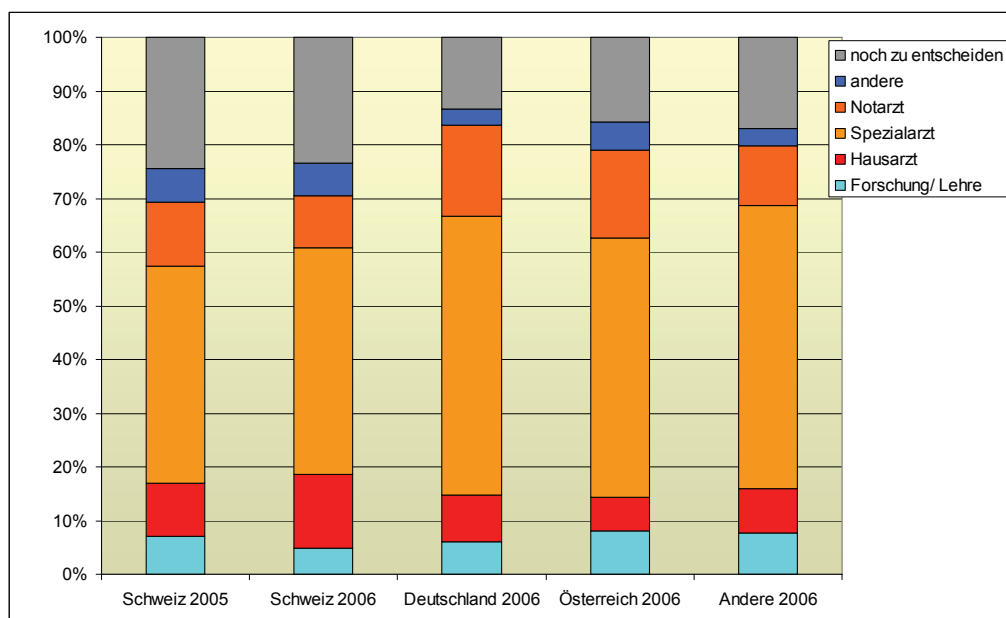


Abbildung 38: Gewünschte zukünftige berufliche Orientierung, Gebiet

10 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.
- Hänsen 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>
- Hänsen 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>
- Hänsen, 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.
- Hänsen, 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.
- Hänsen, 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.
- Hänsen, 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.

10.1 Originaltest zur Information und Vorbereitung

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

10.2 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/