

# EMS • Eignungstest für das Medizinstudium in der Schweiz • 2001

*Bericht über die Durchführung und Ergebnisse 2001*

*✍ ✍ Vorhersage des Prüfungserfolges ✍ ✍*

*ab Seite 40*

*Trends und Schlussfolgerungen aus vier Jahren Testanwendung*

*avec un résumé en français*

**K.-D. Hänsgen und B. Spicher**

**Bericht 7 (2001)**

Hänsgen, Klaus-Dieter; Spicher, Benjamin (2001).

*EMS Eignungstest für das Medizinstudium in der Schweiz 2001*

*Bericht – Vorhersage Prüfungserfolg – Schlussfolgerungen*

Berichte des Zentrums für Testentwicklung, Nummer 7, im Auftrag der Schweizerischen Universitätskonferenz (SUK)

Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik, 2001

Unter Mitarbeit von Katharina Stress (Lektorat),

Mauro Bernasconi, Steve Fürst, Sébastien Simonet, Maurizio Strazzeri

Redaktion: Tanja Läng

*Die Mitglieder des Beirates „Eignungstest“ der Schweizerischen Universitätskonferenz zum Zeitpunkt der Testabnahme 2001:*

*Dr. N. Ischi (Vorsitzender, Bern); Prof. Dr. R. Bloch (Bern); Prof. Dr. P. Groscurth (Zürich); Prof. Dr. J.-P. Montani (Freiburg); Prof. Dr. K. Opwis (Basel); Prof. Dr. M. Perrez (Freiburg); Prof. Dr. U. Schallberger (Zürich); Prof. Dr. Th. Schläpfer (Bern); Prof. Dr. U. M. Spornitz (Basel); Prof. Dr. G. Trost (Bonn/Deutschland).*

---

© ZTD Zentrum für Testentwicklung und Diagnostik 2001  
am Departement für Psychologie der Universität Freiburg  
Direktor: Prof. Dr. K.-D. Hänsgen  
Route d'Englisberg 9, CH-1763 Granges-Paccot  
E-Mail: [ztd@unifr.ch](mailto:ztd@unifr.ch)  
Internet: <http://www.unifr.ch/ztd>

Druck: Cric-Print, Freiburg/Schweiz

## Inhaltsverzeichnis

<b>1</b>	<b>ZUSAMMENFASSUNG .....</b>	<b>5</b>
<b>2</b>	<b>RESUME .....</b>	<b>10</b>
<b>3</b>	<b>EINLEITUNG .....</b>	<b>15</b>
<b>4</b>	<b>NUMERUS CLAUSUS UND MEDIZINSTUDIUM.....</b>	<b>19</b>
<b>5</b>	<b>ANMELDUNG ZU MEDIZINSTUDIUM UND TEST 2001 .....</b>	<b>21</b>
5.1	Anmeldungen, Testantritte und Zulassungen .....	21
5.2	Grösse der Testlokale .....	26
5.3	Testorte und Wunschuniversitäten .....	26
5.4	Testorte und Wohnkantone .....	28
5.5	Testabsolvierung nach Alter und Geschlecht.....	30
5.6	Übernahme des Testergebnisses aus 2000 .....	33
5.7	Erneute Testteilnahme 2001.....	33
5.8	Sprachgruppen.....	34
5.9	Alter und Maturitätsjahr .....	36
<b>6</b>	<b>BESCHREIBUNG DES VERWENDETEN EIGNUNGSTESTS .....</b>	<b>37</b>
6.1	Aufbau des Tests .....	37
6.2	Berechnung der Werte .....	39
<b>7</b>	<b>EVALUATION: EIGNUNGSTEST UND PRÜFUNGSERFOLG .....</b>	<b>40</b>
7.1	Prüfungsverlauf für Testkohorten 1998 und 1999.....	42
7.1.1	Erste Vorprüfung und Prüfungsverlauf.....	42
7.1.2	Zweite Vorprüfung und Prüfungsverlauf.....	46
7.2	Prüfungserfolg und Testwert.....	47
7.2.1	Erste Vorprüfung und Prüfungserfolg gesamt .....	48
7.2.2	Effekt der Zulassung von der Warteliste .....	55
7.2.3	Erfolgsvorhersage für verschiedene Universitäten .....	56
7.2.4	Erfolgsvorhersage für Geschlechter.....	58
7.2.5	Erfolgsvorhersage für Alter und Maturitätsjahr.....	59
7.2.6	Erfolgsvorhersage für Sprachgruppen .....	60
7.2.7	Unterschiede zwischen Human- und Veterinärmedizin .....	61
7.2.8	Zweite Vorprüfung.....	64
7.3	Testwert und Prüfungsnoten (Erste Vorprüfung).....	66
7.4	Schlussfolgerungen zur Eignungsfeststellung .....	73

<b>8</b>	<b>TESTANWENDUNG IN DER SCHWEIZ 2001</b> .....	<b>75</b>
<b>8.1</b>	<b>Organisation der Testabnahme</b> .....	<b>75</b>
<b>8.2</b>	<b>Verteilungsprüfung</b> .....	<b>75</b>
<b>8.3</b>	<b>Vergleich der Testfassungen 1998 bis 2001</b> .....	<b>78</b>
<b>8.4</b>	<b>Äquivalenz der Sprachversionen</b> .....	<b>80</b>
8.4.1	Sprachvergleich für die Untertests .....	82
8.4.2	Darstellung des Korrekturverfahrens .....	85
8.4.3	Leistungen in identischen Testaufgaben 1998 – 2001 .....	97
<b>8.5</b>	<b>Vergleichbarkeit der Testlokale</b> .....	<b>98</b>
<b>8.6</b>	<b>Vergleich der Geschlechter</b> .....	<b>99</b>
<b>8.7</b>	<b>Vergleich der Kantone</b> .....	<b>105</b>
8.7.1	Vergleich 2001 innerhalb der Deutschschweiz .....	106
8.7.2	Zusammenfassende Analyse für Kantone 1998 bis 2001.....	110
<b>8.8</b>	<b>Vergleiche für Altersgruppen</b> .....	<b>114</b>
<b>8.9</b>	<b>Vergleiche nach Wunschuniversitäten</b> .....	<b>118</b>
<b>9</b>	<b>ERGEBNISSE ZUR TESTGÜTE</b> .....	<b>119</b>
<b>9.1</b>	<b>Zuverlässigkeit</b> .....	<b>119</b>
<b>9.2</b>	<b>Binnenstruktur</b> .....	<b>120</b>
<b>9.3</b>	<b>Item-Trennschärfen</b> .....	<b>125</b>
<b>9.4</b>	<b>Konzentriertes und sorgfältiges Arbeiten</b> .....	<b>129</b>
<b>10</b>	<b>LITERATUR</b> .....	<b>131</b>
<b>10.1</b>	<b>Originaltest zur Information und Vorbereitung</b> .....	<b>132</b>
<b>10.2</b>	<b>Frühere Berichte des ZTD</b> .....	<b>132</b>
<b>11</b>	<b>ANHANG ZUM BERICHT 2001</b> .....	<b>135</b>
<b>11.1</b>	<b>Beispielaufgaben für die Untertests</b> .....	<b>136</b>
11.1.1	Untertest: Muster zuordnen .....	136
11.1.2	Untertest: Medizinisch-naturwissenschaftliches Grundverständnis.....	136
11.1.3	Untertest: Schlauchfiguren .....	138
11.1.4	Untertest: Quantitative und formale Probleme.....	138
11.1.5	Untertest: Konzentriertes und sorgfältiges Arbeiten .....	139
11.1.6	Untertest: Figuren lernen.....	139
11.1.7	Untertest: Fakten lernen .....	140
11.1.8	Untertest: Textverständnis.....	140
11.1.9	Untertest: Diagramme und Tabellen .....	141
<b>11.2</b>	<b>Weitere Detailergebnisse</b> .....	<b>144</b>

# 1 Zusammenfassung

*Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest für das Medizinstudium in der Schweiz (EMS) und die Zulassung zum Studium der Human- und Veterinärmedizin ab Wintersemester 2001/2002. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen 1998 bis 2000 herangezogen. Ausserdem werden Ergebnisse zur Vorhersagbarkeit von Prüfungserfolg der Kohorten 1998 und 1999 vorgestellt.*

***Die Ergebnisse sind nicht repräsentativ für Schlussfolgerungen hinsichtlich einer gesamthaften Evaluation von Bildungseinrichtungen, Bildungswegen oder von Regionen.***

## Statistik

Die deutliche Überschreitung der Ausbildungskapazitäten bei der Anmeldung zum Medizinstudium machte in diesem Jahr einen Numerus Clausus (NC) für Personen notwendig, die sich für ein Studium **der Humanmedizin oder Veterinärmedizin** an den Universitäten **Basel, Bern, Freiburg oder Zürich** angemeldet hatten. Für die Zahnmedizin war kein NC notwendig, weil die Anfängerkapazität dort ausreichend war. Bereits in den Jahren 1999 und 2000 war dies so notwendig; 1998 war der NC nur für Humanmedizin gültig.

Bei einer Anfängerkapazität der vier Universitäten von insgesamt **748** Studienplätzen (Humanmedizin: 583, Veterinärmedizin: 165) lagen der Rektorenkonferenz der Schweizer Universitäten (CRUS) im Februar 2001 1191 Bewerbungen (Humanmedizin: 974, Veterinärmedizin: 217) für beide Disziplinen vor. In Bern wurde die Kapazität wegen der Doppelmaturitäten in diesem Jahr erhöht – dennoch wurde sie gesamthaft in Humanmedizin um 67%, in Veterinärmedizin um 32% überschritten.

Als disziplinspezifisches Zulassungskriterium wird seit 1998 der Eignungstest für das Medizinstudium in der Schweiz (EMS) verwendet. Insgesamt 931 Personen haben sich bis zur gesetzten Frist im Mai 2001 bei der CRUS zum Test angemeldet, tatsächlich angetreten sind am Testtag 880 Personen. 51 Personen haben ihre Testanmeldung zurückgezogen bzw. sind nicht zum Test erschienen. In diesem Jahr hat es keine Testabbrüche gegeben. **880 Personen haben den EMS vollständig und mit gültigem Ergebnis absolviert.**

## Organisation und Ablauf

Der Test fand am 6.7.2001 gleichzeitig an sieben Testorten in drei Sprachen (deutsch: Basel, Bern, Chur, St. Gallen, Zürich; französisch: Freiburg; italienisch: Bellinzona) statt. Die Bedingungen für eine Gleichbehandlung sind gegeben. Benachteiligungen bei der Testabnahme – beispielsweise durch einen irregulären Ablauf oder Pannen in einzelnen Testlokalen – traten nicht auf. Es wurde nur eine Verwarnung wegen regelwidrigem Verhalten während der Testabnahme ausgesprochen. Insgesamt spricht dies für einen ausserordentlich disziplinierten Testablauf. Das mit der Testdurchführung betraute Personal verfügt mittlerweile zum Teil über mehrjährige Erfahrungen und es konnte gewährleistet werden, dass in jedem Testlokal Testleiter<sup>1</sup> oder Stellvertreter bereits einmal einen EMS durchgeführt haben.

<sup>1</sup> Im Interesse der Flüssigkeit des Textes werden alle erwähnten Personengruppen nur in der männlichen Form genannt ("Testleiter", "Testteilnehmer", "Maturanden"), obwohl es sich selbstverständlich jeweils um Frauen und Männer handelt.

## Testgüte

Bekanntlich wird jedes Jahr eine neue Testform aus bereits umfangreich erprobten Aufgaben des deutschen Tests für Medizinische Studiengänge (TMS) zusammengestellt und in die drei Testsprachen adaptiert. Auch im vierten Testjahr sind die Testgütekriterien sowohl mit denen der Vorjahre, als auch mit den Werten aus Deutschland ausserordentlich gut vergleichbar. Sie belegen, dass die Testqualität konstant sehr hoch und dem Einsatzzweck angemessen ist: Die Reliabilität des Punktwertes liegt mit 0.91 nicht nur genauso hoch wie in den Vorjahren, sondern ist auch im Verhältnis zu anderen Eignungstests als hoch einzuschätzen. Die Zuverlässigkeitswerte der Einzeltests und des Testprofils insgesamt (0.80) rechtfertigen den Einsatz als Selektionsinstrument. Geringfügige, die Verwendbarkeit nicht einschränkende Abweichungen im Untertest Muster zuordnen sind nicht erklärbar (etwa aufgrund unterschiedlicher Druckqualität). Sie bedürfen zukünftig weiterer Aufmerksamkeit.

Die im Vergleich zu den Vorjahren nahezu identische Faktorenstruktur und Varianzaufklärung begründen auch die Übertragbarkeit der Ergebnisse zwischen den Jahren; es werden in jedem Jahr faktoriell vergleichbare Merkmale gemessen. Auch die beiden verwendeten Parallelformen, die pro Untertest eine unterschiedliche Abfolge der Items aufweisen, unterscheiden sich nicht, so dass deren Gleichwertigkeit gesichert ist.

## Gleichbehandlung der Sprachgruppen

In diesem Jahr treten wie in den Vorjahren Unterschiede der Testwertverteilung zwischen den Sprachgruppen auf. Es kann allerdings erneut als gesichert gelten, dass es keine testbedingten Benachteiligungen gibt und die Chancengleichheit gewährleistet ist. Unterschiede haben ihre Ursache nicht in der Übersetzung, sondern andere Faktoren sind dafür verantwortlich (z.B. eine unterschiedliche Zusammensetzung der Kohorten bzw. auch kulturelle Unterschiede). Bei allen Sprachvergleichen werden die italienisch- und die französischsprachige Gruppe jeweils separat mit der deutschsprachigen Gruppe verglichen.

Von den insgesamt 9 Untertests sind vier praktisch nicht sprachabhängig (*Muster zuordnen; Schlauchfiguren; Figuren lernen; Konzentriertes und sorgfältiges Arbeiten*) und fünf sprachabhängig (*Med.-naturwissenschaftliches Grundverständnis; Quantitative und formale Probleme; Textverständnis; Fakten lernen; Diagramme und Tabellen*). Unterschiede zur deutschsprachigen Gruppe treten sowohl in den nicht sprachabhängigen als auch in den sprachabhängigen Untertests auf – sie sind in den **nicht sprachabhängigen Untertests sogar höher**. Besonders in der italienischsprachigen Gruppe ist es 2001 zu einer Zunahme der Unterschiede in den nicht sprachabhängigen Untertests gekommen, während die Unterschiede in den sprachabhängigen Untertests über die Jahre etwa gleich blieben.

Entsprechend dem Beschluss der Schweizerischen Universitätskonferenz (SUK) wurde wiederum das in den Vorjahren bewährte Ausgleichsverfahren angewendet, um mögliche testbedingte Unterschiede bei einzelnen Aufgaben auszuschliessen (DIF-Methode<sup>1</sup>). In diesem Jahr weichen nur sehr wenige Items vom generellen Trend ab und mussten korrigiert werden (insgesamt 14 im Vergleich deutsch-französischsprachig und deutsch-italienischsprachig gegenüber 28 im Vorjahr und 25 im Jahre 1999). Dies kann auch für eine weiter verbesserte Übersetzungsqualität sprechen.

Die Unterschiede in den **nicht sprachabhängigen** Untertests (jeweils die deutschsprachige Gruppe verglichen mit der französisch- und der italienischsprachigen Gruppe) übersteigen

<sup>1</sup> Differential Item Functioning, siehe Seite 34

bereits vor der Korrektur diejenigen der **sprachabhängigen** Untertests. Die anschliessende Korrektur der Bewertungsskala ist bei besonders abweichenden Items nach wie vor sachlich begründet – der Korrekturunterschied bleibt aber im Verhältnis zum vorhandenen Unterschied zwischen den Sprachgruppen in den nicht sprachabhängigen Untertests aber gering.

Über die Ursachen dieser Unterschiede kann man aufgrund der vorliegenden Daten nur spekulieren. Die starke Variation der sozialen Zusammensetzung der französisch- und italienischsprachigen Gruppe zwischen den Jahren scheint ein wichtiger Faktor zu sein – möglicherweise hat sich noch nicht stabil herausgebildet, wer in der Westschweiz bzw. in Italien ein Studium aufnimmt und wer sich an einer der am NC beteiligten Universitäten bewirbt.

Aufgrund der **Prognosedaten zum Studienerfolg** lässt sich jetzt allerdings mit Sicherheit festhalten: gleich gute Testwerte stehen für gleiche Chancen, die Prüfung in **jeder der drei Sprachgruppen** zu bestehen. Der Test übervorteilt oder benachteiligt keine der Gruppen.

### Differenzierung nach Alter

Die Unterschiede beim Testwert nach Alter und Zeitpunkt des Ablegens der Maturität zeigen in diesem Jahr wiederum den gleichen Trend der Vorjahre: Jüngere Studienbewerber erreichen höhere Testwerte. Der Unterschied nach dem Zeitpunkt des Ablegens der Maturitätsprüfung ist innerhalb der ältesten Gruppe in diesem Jahr nicht signifikant – die Gruppengrößen sind allerdings geringer als in den Vorjahren.

Bei Unterteilung nach Geburtsjahren ist innerhalb der jüngsten Gruppe, die nach 12 bzw. 13 Jahren die Maturität im Testjahr abgelegt hat, ebenfalls kein Unterschied nachweisbar.

### Differenzierung nach Geschlecht

Der Anteil der Bewerberinnen für die Human- und Veterinärmedizin hat in den letzten drei Jahren zugenommen (1998: 55,2% - 2001: 61,5%). Die Erwartung identischer mittlerer Testwerte für beide Geschlechter ist allein schon deshalb gering. Eine weitere wichtige Aussage für einen Vergleich beider Geschlechter findet sich bei der Vorhersage des Prüfungserfolges. Es zeigt sich eindeutig, dass **gleiche Testwerte eine gleiche Bestehenswahrscheinlichkeit** der ersten Vorprüfung für beide Geschlechter aufweisen. Die Bestehenswahrscheinlichkeit für weibliche Studierende ist allerdings gesamthaft um 7% geringer als diejenige für ihre männlichen Kollegen.

Ein in diesem Jahr signifikanter Unterschied der Testwerte (2.3 Punkte) zwischen den Geschlechtern (1999: 1.5 Punkte; 2000: 1.4 Punkte) spricht keinesfalls gegen Chancengleichheit, weil sich die geschlechtsspezifische Wahrscheinlichkeit, aufgrund des gleichen Testwertes einen Studienplatz zu erhalten, nicht signifikant unterscheidet. Summa summarum werden Frauen im Zulassungsverfahren nach dem Kriterium des Testwerts also keinesfalls benachteiligt.

### Differenzierung nach Maturitätsquote und Kanton

Wiederum hat die Maturitätsquote allein keinen Einfluss auf den Testwert. Unterschiede zwischen den Kantonen (betrachtet wurden die Kantone mit ausreichender Personenzahl für statistische Analysen) bestätigen die langjährigen Trends. Neben der Maturitätsquote scheint der Anteil der Medizinstudierenden an der 19jährigen Wohnbevölkerung eine wichtige Grösse zu sein, die mit Unterschieden beim Testwert zusammenhängt.

## Evaluation: Prognose der Studieneignung

In diesem Jahr standen Prüfungsergebnisse der Testkohorten 1998 (erste und teilweise bereits zweite Vorprüfung) und 1999 (erste Vorprüfung) zur Verfügung, um die Vorhersagbarkeit des Prüfungserfolges aufgrund des Testwerts zu untersuchen.

„Eignung“ als Zulassungskriterium sollte auf „**Studieneignung**“ bezogen aufgefasst werden, weil „Berufseignung“ durch die vorliegenden gesetzlichen Regelungen und Entscheide des Bundesgerichtes nicht gerechtfertigt scheint.

Studieneignung lässt sich durch zwei messbare Kriterien erfassen: (1) die Absolvierung des Studiums in möglichst **kurzer (optimaler) Zeit** und (2) die Absolvierung des Studiums mit möglichst **guten Leistungen**. Ein NC-Zulassungskriterium ist dann begründet, (1) wenn es die **Studiendauer vorhersagbar** macht und (2) wenn es die **Studienleistungen vorhersagbar** macht - und die jeweils Besten dann grössere Chancen auf eine Zulassung zum Studium haben.

Weil bisher nur wenige Personen die zweite Vorprüfung absolviert haben (ein Teil der Kohorte 1998, der auch die erste Vorprüfung bereits nach dem ersten Jahr bestanden hat), lassen sich fundierte Aussagen bisher nur für die erste Vorprüfung treffen.

Die Befunde zeigen für beide bisher erfasste Jahrgänge und Prüfungen, dass der Eignungstest beide Kriterien sehr gut erfüllt:

- Personen, welche die **erste Vorprüfung im ersten Anlauf** bestehen, haben durchschnittlich einen signifikant (um 5 Punkte) besseren mittleren Testwert als alle anderen Gruppen (gebildet nach dem Abschneiden in dieser Prüfung).
- Dieser Unterschied (mindestens 5 Punkte in der ersten Vorprüfung) zeigt sich grundsätzlich in allen vier **beteiligten Universitäten**, wobei es Verschiebungen zwischen den Universitäten entsprechend der Zusammensetzung der Bewerber gibt.
- Dieser Unterschied zeigt sich identisch für **beide Geschlechter**: die Verteilungen der Testwerte für die Gruppen nach dem Prüfungsbestehen sind gleich, **gleiche Testwerte** weisen **unabhängig vom Geschlecht** auf die **gleiche Bestehenswahrscheinlichkeit** der Prüfung hin. Die Bestehenswahrscheinlichkeiten der Prüfung selber unterscheiden sich allerdings um 7% zugunsten der Männer.
- Dieser Unterschied zeigt sich gleichsinnig für alle **drei Sprachgruppen**: auch hier sind die Testwerte der Erfolgreichen vergleichbar. Die beim Testwert langjährig festgestellten Leistungsunterschiede zeigen sich in unterschiedlichen Bestehenswahrscheinlichkeiten der Prüfungen (79% deutsch-, 62% italienisch- und 47% französischsprachig). Auch hier sagt der Test also letztendlich eine vorhandene Realität unterschiedlicher Bestehenswahrscheinlichkeiten voraus – **bei gleichen Leistungen sind auch die Chancen gleich**.
- Dieser Unterschied zeigt sich gleichsinnig für **Veterinär- und für Humanmedizin**: Die Mittelwerte des Testwertes der Gruppen nach dem Prüfungsbestehen unterscheiden sich nicht. Das gleiche Testergebnis bedeutet in jeder Disziplin die gleiche Bestehenswahrscheinlichkeit der Prüfung. Da die mittleren Testwerte in der Veterinärmedizin geringer sind als in Humanmedizin, liegt die Bestehensquote in der Veterinärmedizin um rund 20% unter derjenigen der Humanmedizin.
- Personen, welche die **Prüfung nicht bestehen**, unterscheiden sich auch signifikant von Personengruppen, welche die Prüfung erst im zweiten Anlauf bestehen oder noch nicht zur Prüfung angetreten sind.

- Zwischen **Testwert und Bestehenswahrscheinlichkeit** der Prüfung besteht eine lineare Beziehung: je niedriger der Testwert, umso unwahrscheinlicher ist das Bestehen. Das Vorgehen, den Grenzwert der Zulassung kapazitätsentsprechend festzulegen, ist dadurch gerechtfertigt: für alle Personen unter diesem Grenzwert gilt eine geringere Bestehenswahrscheinlichkeit als für die Zugelassenen.
- In der **zweiten Vorprüfung** findet sich ebenfalls noch ein signifikanter Unterschied nach dem Bestehen der Prüfung (Personen, welche sie bestanden haben, sind rund 4 Punkte besser als diejenigen, die sie nicht bestanden haben). Zu beachten ist, dass Personen, die nach kürzestmöglicher Studienzeit bereits zur zweiten Vorprüfung antreten konnten, die Leistungsbesten sind.
- Auch die **Prüfungsnoten der ersten Vorprüfung** korrelieren mit den Testwerten signifikant - im Mittel um 0.50. Dies gilt für verschiedene Universitäten, Disziplinen und Notensysteme. Diese Korrelationen zeigen sich in den Universitäten, bezüglich Alter und Geschlecht sowie in den Sprachgruppen in gleicher Weise. Sie sind auch nicht spezifisch für die Art der Prüfung (Multiple-Choice oder Essay – schriftlich oder mündlich).

Der Test regelt die Zulassung nach der Eignung zum Studium der Human- und Veterinärmedizin, indem diejenigen bevorzugt einen Studienplatz erhalten, die das Studium in kürzerer Zeit (nachgewiesen für die erste und zweite Vorprüfung) und mit besseren Leistungen (nachgewiesen für die erste Vorprüfung) absolvieren. Es kann davon ausgegangen werden, dass gleiche Testwerte in jeder Teilgruppe (nach Sprache, Alter, Geschlecht, Disziplin, Universität) eine gleiche Bestehenswahrscheinlichkeit der Prüfung bedeuten. Wenn der Testwert für Eignung steht, dann wird die Forderung nach Chancengleichheit erfüllt: gleiche Chancen bei gleicher Eignung.

## Schlussfolgerungen

Der EMS erfasst nachweislich die Studieneignung, indem eine kürzere Studiendauer und bessere Studienleistungen mit einem höheren Testwert einhergehen. Gleichbehandlung und Fairness sind gewährleistet. Im Vergleich zu anderen möglichen Zulassungskriterien ist der Test auch ökonomisch eine der besten Lösungen.

Es kann gewährleistet werden, dass nunmehr über 4 Jahre immer wieder vergleichbare Testversionen zur Verfügung stehen. Somit ist die Übertragung des Testwertes ins Folgejahr gerechtfertigt.

Der Test erweist sich nach vier Jahren Einsatz in der Schweiz weiterhin als geeignetes Instrument für die Zulassung. Mögliche graduelle Änderungen der Studienanforderungen sollten genutzt werden, um den Test weiter daran anzupassen. Die Einbeziehung kommunikativer Fähigkeiten wäre ebenso denkbar wie beispielsweise die Überarbeitung des Aufmerksamkeits- und Gedächtnisteiles.

## 2 Résumé

*Le présent rapport expose les résultats du test d'aptitudes pour les études de médecine en Suisse (AMS), un test nécessaire pour être admis en médecine humaine et vétérinaire au semestre d'hiver 2001/2002 dans les Universités de Bâle, Berne, Fribourg ou Zurich. Nous avons utilisé des données comparatives des tests effectués de 1998 à 2000 afin d'éclairer certains points. Ce rapport fait également des pronostics quant aux chances de réussite des volées 1998 et 1999.*

***Les résultats de ce rapport ne sont pas suffisamment représentatifs pour pouvoir tirer des conclusions générales sur les établissements de formation, les filières d'études ou les régions.***

### Statistiques

Le nombre de personnes qui se sont inscrites pour commencer cette année des études **en médecine humaine ou vétérinaire à l'Université de Bâle, de Berne, de Fribourg ou de Zurich**, a nettement dépassé les capacités de formation. Il a donc été nécessaire d'appliquer un numerus clausus (NC), ce qui n'a pas été le cas pour la médecine dentaire dont les capacités d'accueil sont suffisantes. Nous retrouvons ici la situation de 1999 et de 2000, alors qu'en 1998, seule la médecine humaine avait requis un NC.

Tandis que les capacités en première année dans ces quatre universités sont de **748** places en tout (médecine humaine: 583, médecine vétérinaire: 165), la Conférence des Recteurs des Universités Suisses (CRUS) recensait, en février 2001, **1191** inscriptions pour les deux disciplines (médecine humaine: 974, médecine vétérinaire: 217). Les capacités avaient été rehaussées cette année à Berne en raison de la double maturité, mais cela n'a pas suffi à résorber le surcroît d'inscriptions qui dans l'ensemble dépassent de 67% les capacités en médecine humaine et de 32% les capacités en médecine vétérinaire.

Le test d'aptitudes pour les études de médecine en Suisse (AMS) sert depuis 1998 de critère d'admission dans cette discipline. En mai 2001, délai d'inscription, 931 personnes s'étaient annoncées auprès de la CRUS pour le passer: 880 d'entre elles se sont effectivement présentées le jour J, tandis que 51 autres retiraient leur candidature ou ne donnaient simplement pas signe de vie. Personne n'ayant abandonné cette année, **880 personnes donc ont effectué le test en entier en rendant des résultats valables.**

### Organisation et déroulement

Le test a eu lieu le 6 juillet 2001, simultanément dans sept villes réparties dans les trois régions linguistiques (français: Fribourg; italien: Bellinzone; allemand: Bâle, Berne, Coire, St-Gall, Zurich). Les conditions d'égalité de traitement sont remplies. Personne n'a été désavantagé au cours du test, en raison d'irrégularités ou de pannes... par exemple. Une seule personne a dû être avertie pour comportement non réglementaire. Les candidat-e-s se sont donc montré-e-s, dans l'ensemble, très discipliné-e-s. Parmi le personnel chargé de surveiller le déroulement du test, certains disposent désormais de plusieurs années d'expérience et partout, le responsable ou son suppléant avait déjà mené à bien un AMS.

## Qualité du test

Comme à l'accoutumée, un nouveau test a été préparé cette année à partir du vaste corpus d'exercices éprouvés que fournit le Test pour les études de médecine (TMS) de l'Allemagne et adapté dans les trois langues nationales. En cette quatrième édition, le test remplit les critères de qualité de manière tout à fait comparable à l'année précédente ou à l'Allemagne, preuve en est qu'il se maintient à un très bon niveau et qu'il est bien ciblé: sa fiabilité se situe à 0,91, soit au même niveau qu'en 2000, et peut même être considérée comme élevée par rapport à d'autres tests d'aptitudes. Les indices de fiabilité de chaque test ainsi que le profil général du test (0,80) en justifient l'utilisation comme instrument de sélection. Certains écarts minimes, qui ne portent pas préjudice à la validité du test, ont été constatés dans le sous-test Reconnaissance de fragments de figures et n'ont pas encore trouvé d'explication (comme étant p. ex. dus à une qualité d'impression différente). Ils devront retenir notre attention à l'avenir.

Chaque année certaines caractéristiques factorielles comparables sont mesurées. La structure factorielle et la distribution de la variance étant quasi identiques à celles des années passées, on peut transposer les résultats d'une année à l'autre. Aucune divergence n'a été enregistrée entre les deux formes parallèles (induisant chacune un développement différent) utilisées pour chaque sous-test, leur équivalence est ainsi garantie.

## Egalité de traitement entre groupes linguistiques

A l'instar des années passées, la répartition des résultats de 2001 diffère d'un groupe linguistique à l'autre. Toutefois, on peut à nouveau affirmer que ces écarts ne sont pas liés au test et donc que tous les groupes linguistiques ont été traités sur un pied d'égalité. Ces écarts ne sont pas dus à la traduction mais à d'autres facteurs (p. ex., composition des volées ou origine culturelle). Nous avons chaque fois comparé séparément le groupe francophone et le groupe italophone avec le groupe germanophone.

Dans quatre sous-tests, sur neuf, la langue ne joue pas de rôle particulier (*Reconnaissance de fragments de figures; Figures tubulaires; Travail avec soin et concentration; Mémorisation de figures*) alors qu'elle en joue un dans les cinq autres (*Compréhension de questions fondamentales de la médecine et des sciences naturelles; Problèmes quantitatifs et formels; Mémorisation de faits; Compréhension de textes; Diagrammes et tableaux*). Des différences avec le groupe germanophone apparaissent dans ces deux catégories et **sont même plus grandes dans les sous-tests à faible composante linguistique**. Cette année, c'est le cas surtout pour le groupe italophone, alors que les différences dans les sous-tests à forte composante linguistique restent à peu près stables au cours des années.

Conformément à la décision de la Conférence universitaire suisse (CUS), on a utilisé la même procédure de compensation que lors des éditions précédentes, afin d'éviter toute interférence de la formulation du test dans les exercices (méthode DIF<sup>1</sup>). Cette année, très peu d'items se sont écartés de la tendance générale et ont dû donc être corrigés (14 en tout dans les comparaisons allemand-français et allemand-italien contre 28 en 2000, et 25 en 1999), ce qui peut être aussi mis au compte d'une meilleure traduction.

Avant correctif, les sous-tests **à faible composante linguistique** (le groupe germanophone étant chaque fois comparé aux groupes francophone et italophone) présentent déjà de plus grandes divergences que les sous-tests **à forte composante linguistique**. Il demeure objectivement justifié d'apporter un correctif à l'échelle d'évaluation pour les items

---

<sup>1</sup> Differential Item Functioning, voire p. 34

présentant de gros écarts, son effet reste d'ailleurs presque insignifiant face à la différence observée entre les sous-tests en raison du rôle qu'y joue le langage.

Les données en présence ne permettent que de spéculer sur les causes de ces différences. Il semble que les très fortes variations dans la composition sociale des groupes tant francophone qu'italophone au cours des années jouent un rôle important: il est possible que les catégories de personnes qui commencent des études en Suisse romande ou en Italie, et celles qui se présentent à une université pratiquant le NC ne sont pas encore stabilisées.

Toutefois, nous pouvons désormais affirmer avec certitude, sur la base des **pronostics de réussite**, que de bons résultats donnent la même chance de réussir aux examens **quel que soit le groupe linguistique**. Le test n'avantage, ni ne désavantage aucun groupe linguistique.

### Différenciation en fonction de l'âge

Cette année encore, les résultats répartis en fonction de l'âge ou de l'époque où la maturité a été obtenue varient de la même manière que les années précédentes: plus on est jeune, meilleur est le résultat. Dans le groupe des plus âgés, l'époque où la maturité a été obtenue ne joue cette année pas de rôle significatif. Notons toutefois que les groupes sont plus restreints que les années précédentes.

Dans le classement par année de naissance, le groupe des plus jeunes, soit ceux qui, après 12 ou 13 années de scolarité, ont passé la maturité cette année, ne présente aucune différence significative.

### Différenciation en fonction du sexe

La proportion des candidates à la médecine humaine ou vétérinaire a augmenté au cours des trois dernières années (1998: 55,2% - 2001: 61,5%). On ne peut dès lors pas raisonnablement s'attendre à des moyennes identiques pour les deux sexes. Autre point important ici: les pronostics de réussite. Or, les **mêmes résultats** donnent clairement la **même probabilité de réussite** au premier propédeutique pour les deux sexes, même si globalement elle est de 7% inférieure chez les femmes que chez leurs homologues masculins.

Le fait que cette année les résultats se distinguent davantage entre les sexes (2,3 points contre 1,5 en 1999 et 1,4 en 2000) ne signifie aucunement un traitement inégal, parce qu'un même résultat offre quasi la même probabilité d'obtenir une place d'études qu'on soit femme ou homme. On peut donc conclure que, sous l'angle du résultat au test, les femmes ne sont en aucune manière défavorisées lors de la procédure d'admission.

### Différenciation par taux de maturité et par canton

Le taux de maturité n'explique toujours pas, à lui seul, les résultats. Les écarts entre les cantons (ayant un nombre suffisant de personnes pour une analyse statistique) confirment les tendances habituelles: outre le taux de maturité, un autre facteur explicatif semble être la proportion des étudiants en médecine dans la population résidante ayant 19 ans.

### Evaluation: pronostics quant à l'aptitude aux études

Nous disposons cette année des résultats d'examen des volées 1998 (1<sup>er</sup> propédeutique et parfois déjà 2<sup>e</sup> propédeutique) et 1999 (1<sup>er</sup> propédeutique) pour analyser dans quelle mesure le résultat du test permet de prévoir la réussite aux examens.

L'aptitude utilisée comme critère d'admission doit se référer aux études et non à l'exercice de la profession car cette dernière aptitude ne semble reposer sur aucun fondement légal, que ce soit au niveau des réglementations cantonales ou de la jurisprudence.

Cette aptitude peut être saisie grâce à deux critères mesurables: réussir ses études (1) le plus **rapidement (durée optimale)** et (2) le **mieux (qualité des résultats)** possible. Il s'ensuit que l'emploi d'un critère d'admission dans une procédure NC se justifie s'il rend **prévisible** tant (1) la **durée** que (2) les **résultats des études** et si les meilleurs d'après ces critères ont le plus de chance d'être admis aux études.

Etant donné que seul un nombre restreint de personnes ont passé le deuxième propédeutique (le groupe de la volée 1998 ayant passé le 1<sup>er</sup> propédeutique la première année déjà), les conclusions qu'on peut tirer ne concernent pour le moment que le premier propédeutique.

L'analyse des deux cohortes et des examens montre que le test d'aptitude remplit parfaitement les deux critères:

- Les personnes qui ont passé le **premier propédeutique du premier coup** ont obtenu au test un résultat nettement supérieur (de 5 points) en moyenne que celui des autres groupes (répartis en fonction de cet examen).
- Cet écart (au moins 5 points) existe en principe dans les quatre universités concernées, même si l'on constate des disparités entre elles correspondant à la composition des groupes.
- Cet écart est le même pour les deux sexes. Les résultats du test se répartissent de la même manière dans les groupes définis en fonction de l'examen: les **mêmes résultats** donnent la **même probabilité de réussite** que l'on soit **femme ou homme**, même si elle est 7% supérieure chez les hommes considérés globalement.
- Cet écart se manifeste de la même manière dans les **trois groupes linguistiques**: les résultats des meilleurs sont ici aussi comparables. Les différences constatées depuis longtemps au niveau des prestations au test se reflètent au niveau des pronostics de réussite aux examens (47% de francophones, 62% d'italophones et 79% de germanophones). Le test prédit donc bien, ici également, les chances de réussite: **à prestations identiques, chance identique.**
- Cet écart vaut aussi bien pour la **médecine humaine que pour la médecine vétérinaire**. Les moyennes du test sont les mêmes dans les groupes définis en fonction de l'examen: le même résultat au test signifie, dans chaque discipline, la même probabilité de réussite à l'examen. Mais, puisque les résultats au test de médecine vétérinaire sont moindres que ceux de la médecine humaine, le taux de réussite dans la première est environ 20% inférieur à celui de la deuxième.
- Les personnes qui ont **raté l'examen** divergent également des groupes de personnes qui n'ont réussi l'examen que lors de la deuxième tentative ou qui ne se sont pas encore présentées aux examens.
- Le **résultat au test** et la **probabilité de réussite à l'examen** sont en rapport linéaire strict: plus le test est bas, moins la réussite est probable. La procédure adoptée – fixer une limite d'admission en fonction des places disponibles – trouve ici sa justification: toutes les personnes se situant en deçà de cette limite ont moins de chance de réussir que celles qui sont admises.
- Au niveau du **deuxième propédeutique**, on trouve aussi un écart significatif en fonction de la réussite de l'examen (les personnes reçues ont environ 4 points de plus

que celles qui ont raté l'examen). Signalons que les personnes qui ont passé le deuxième propédeutique au plus tôt ont fourni les meilleures prestations au test.

- On peut aussi établir un coefficient de corrélation, en moyenne de 0,50, entre les **notes du premier propédeutique** et les résultats du test dans plusieurs universités, disciplines ou systèmes de notes. On retrouve les mêmes corrélations au sein de chaque université, au niveau de l'âge, du sexe ou du groupe linguistique. Elles ne dépendent pas non plus du type d'examen (traditionnel ou questionnaire à choix multiple, oral ou écrit).

En bref, le test règle l'admission aux études en fonction de l'aptitude à étudier la médecine humaine ou vétérinaire, dans le sens où les personnes qui obtiennent en priorité une place sont celles qui achèvent leurs études le plus vite (avéré pour les 1<sup>er</sup> et 2<sup>e</sup> propédeutiques) et le mieux (avéré pour le 1<sup>er</sup> propédeutique). On peut dire que, dans chaque sous-groupe (défini par la langue, l'âge, le sexe, la discipline et l'université), un même résultat signifie un même pronostic quant aux chances de réussir l'examen et que, lorsque le résultat au test reflète l'aptitude, le principe de l'égalité de traitement est respecté: à aptitude égale, chance égale.

## Conclusions

Il est donc prouvé que l'AMS porte réellement sur l'aptitude aux études puisqu'un très bon résultat au test va de paire avec une durée d'études minimale et une place parmi les meilleurs aux examens. Egalité de traitement et équité sont garantis. Comparé à d'autres critères, ce test se défend aussi d'un point de vue économique.

De plus, on peut affirmer que depuis quatre ans les différentes éditions du test sont comparables, il est ainsi justifié de transposer les résultats du test sur l'année suivante.

Le test a fait ses preuves en Suisse: après avoir été appliqué durant quatre années, il s'avère un instrument de sélection approprié pour l'admission aux études de médecine. D'éventuelles modifications quant aux exigences pourraient être mises à profit afin de poursuivre son adaptation. On pourrait y intégrer les aptitudes en matière de communication ou retravailler par exemple la partie consacrée à l'attention et à la mémoire.

### 3 Einleitung

Ein Numerus Clausus (NC) wird für die Zulassung zum Studium der Medizin seit 1998 an den Universitäten Basel, Bern, Freiburg und Zürich angewendet. Jede Zulassungsbeschränkung erfordert Kriterien, nach denen die Auswahl der Personen erfolgt. Als Zulassungskriterium wird in der Schweiz ein Eignungstest (EMS) verwendet. Die recht intensiven politischen Diskussionen zum Zeitpunkt seiner Einführung werden nicht mehr geführt – NC und der Eignungstest sind in der öffentlichen Meinung faktisch eine „feste Einrichtung“ geworden. Volksentscheide in Zürich und Bern sind ohnehin deutlich pro NC ausgefallen – an der breiten öffentlichen Akzeptanz hat es nie gefehlt.

Die Entscheidung für einen NC ist eine politische Entscheidung, die unter Abwägung der verschiedenen Interessen getroffen wird. Die Schweiz gehört weiterhin zu den Ländern mit der grössten Ärztedichte. Die Gesundheitskosten wachsen und lassen sich nur schwer eindämmen – erst jüngst sind die Krankenkassen-Beiträge wieder drastisch gestiegen. Von dieser Seite gerät der NC nun unter Druck, auch berufspolitisch zu wirken und für eine Eindämmung der absehbaren und befürchteten „Ärztenschwemme“ zu sorgen. Es scheint daher wieder an der Zeit, darauf hinzuweisen, was seine Grundlagen sind – und was nicht.

Die Interkantonale Universitätsvereinbarung (IUV) vom 20.2.1997 fordert im Artikel 10 eine Gleichbehandlung bei Zulassungsbeschränkungen. Die gesetzlichen Grundlagen dafür sind auf der Ebene der Kantone geregelt und sagen aus, dass die Zulassung beschränkt werden kann, wenn Ausbildungs-Kapazitäten nicht ausreichen und keine andere Lösung gefunden werden kann. Es handelt sich also um eine bildungspolitische und keine berufspolitische Massnahme. Rückwirkungen auf den Arbeitsmarkt erfolgen allenfalls indirekt über eine langfristige Anpassung der Ausbildungskapazitäten an den Bedarf. Auch das Interesse für den Beruf spielt eine Rolle. Dieses wird beeinflusst durch die erlebten Aussichten, auf dem Gebiet der Medizin eine zukunftsichere und befriedigende Tätigkeit zu finden. Eine zu grosse „Dissuasion“ durch schlechtere Aussichten und Perspektiven würde ebenso zu Fehlregulationen führen wie eine zu grosse Idealisierung des Medizinerberufes. Ausserdem könnte eine Dissuasion die Geeigneten stärker abschrecken – etwa, wenn diese auch offener für andere Lebensperspektiven sind.

Der NC ist durch Entscheide des Bundesgerichts gestützt<sup>1</sup> und wird grundsätzlich gutgeheissen – aber gleichzeitig werden zwei Forderungen an das Zulassungsverfahren gestellt, die ausserordentlich bedeutsam scheinen: Es müssen Rechtsgleichheit und Willkürfreiheit gewährleistet sein<sup>2</sup>. Die Gleichbehandlung wird durch diese beiden Begriffe also weiter differenziert und in der Sache bestätigt.

Die Kantone sind zuständig für die gesetzlichen Grundlagen des Numerus Clausus. In diesen gesetzlichen Grundlagen für den NC der Kantone ist geregelt, dass bei der Zulassung von der Eignung auszugehen ist. Der Eignungsbegriff wird nicht weiter differenziert, um offen für das Einbeziehen aktueller Forschungsergebnisse oder internationaler Erfahrungen zu sein – schliesslich ist die Schweiz nicht das einzige Land, in welchem

---

<sup>1</sup> Urteile des Schweizerischen Bundesgerichtes 2P.171/1998/mks vom 3. März 1999 sowie 2P.228/1998 – 2P.332/1998/luc vom 20. August 1999.

<sup>2</sup> "Wie alle staatlichen Leistungen sind auch Studienplätze zwangsläufig ein beschränkt vorhandenes Gut. Verfassungsrechtlich besteht ein Anspruch auf eine willkürfreie und rechtsgleiche Regelung bei der Zulassung zu den vorhandenen Studienplätzen". Urteil vom 3. März 1999 S. 7.

Zulassungsbegrenzungen notwendig sind. Sie findet sich diesbezüglich in einer Reihe mit den meisten Industriestaaten wieder.

Auch in den bisher bekannten Entwürfen zum neuen Medizinalberufe-Gesetz ist der Eignungsbegriff weiterhin in der Diskussion als Zulassungskriterium. Obgleich dort der Versuch unternommen wird, den Eignungsbegriff weiter zu differenzieren (etwa als „psychosoziale Eignung“<sup>1</sup>), bleiben die gewählten Modelle grundsätzlich gültig.

Dass Eignung nicht weiter definiert ist, birgt allerdings auch Risiken für die Diskussion. Es werden alle denkbaren Wunschvorstellungen über „den guten Arzt/die gute Ärztin“ als Eignungskriterien genannt und damit als Zulassungskriterien zum Studium vorgeschlagen. Dies geschieht ohne Rücksicht darauf,

- (1) ob diese Kriterien bereits vor Beginn des Studiums vergleichbar für die verschiedenen Bewerbergruppen ausgeprägt sind und eingeschätzt werden können (oder beispielsweise Personen mit Berufserfahrung Vorteile hätten - was gegen die Gleichbehandlung sprechen würde);
- (2) ob die Vorhersagbarkeit der Entwicklung dieser Kriterien über einen so langen Zeitraum wie das Studium überhaupt gegeben ist (oder diese Entwicklung vor allem während des Studiums erst in Gang gesetzt wird);
- (3) ob überhaupt Methoden existieren, welche diese Kriterien mit ausreichender Genauigkeit messen können, was für die Rechtsfähigkeit der Zulassung eine unbedingte Voraussetzung ist (oder dass durch Training jede/r sehr leicht einen „guten Eindruck“ hinterlassen kann, ohne dass der sachlich gerechtfertigt wäre);
- (4) ob das Zulassungsverfahren auf der Basis dieser Kriterien in zeitlicher und finanzieller Hinsicht noch praktikabel ist.

Betrachten wir eine psychologische Definition von Eignung, so ist die „*hinreichende Übereinstimmung von individuellen Leistungsvoraussetzungen und aufgabenabhängigen Anforderungen bei konkreten Tätigkeiten*“<sup>2</sup> der gemeinsame Nenner aller Definitionen. Will man rechtsgleich und willkürfrei Personen nach unterschiedlicher Eignung auswählen, müssen die Anforderungen der Tätigkeiten und vor allem Erfolgskriterien für die Erfüllung der Anforderungen ausreichend definiert sein. Je globaler und undifferenzierter die Anforderungen sind (was bei der Definition „guter Ärzte“ nicht ausbliebe), umso ungenauer wird man Personen nach der erfolgreichen Erfüllung dieser Anforderung (was Eignung bestimmt) differenzieren können. Umso weniger begründet würde die Zulassung sein.

Eignung sollte auch nicht mit Neigung verwechselt werden. Der Begriff der Neigung wird eher mit dem Interesse, der Motiviertheit für den Medizinalberuf in Verbindung gebracht. Es wäre ideal, wenn die Informationen vor der Studienbewerbung so weit ausreichen würden, dass realistische Erwartungen in Bezug auf das Studium und die spätere berufliche Tätigkeit bestehen. Hier haben Universitäten und auch die Praxis gute Möglichkeiten, entsprechende Informationen zu liefern – bis hin zu „Schnupperpraktika“, wo sich die Bewerber und Bewerberinnen direkt vor Ort informieren und ihren Berufswunsch festigen, ihre Neigung zum Medizinalberuf erkennen und prüfen können. Es wäre allerdings vermessen, in solchen Praktika die Eignung zum Bestehen der Studienanforderungen oder

<sup>1</sup> „Psychosoziale Eignung“ meint alle psychischen und sozialen Prozesse und ihre Wechselwirkungen und ist daher als Begriff so global, wie wenn man nur von „Eignung“ spricht.

<sup>2</sup> Schorr, A. (Hrsg.) (1993). Handwörterbuch der Angewandten Psychologie. Stichwort „Eignungsdiagnostik“ (H.J. Liebel). Bonn: Deutscher Psychologen-Verlag.

generell für einen Medizinalberuf abzuschätzen. Interesse ist – wie in vielen anderen Bereichen, nicht mit Fähigkeit bzw. Eignung gleichzusetzen.

In den jetzt vorliegenden und nachfolgend dargestellten Ergebnissen zur Prognose der Studienergebnisse durch den Eignungstest kann schlüssig nachgewiesen werden, dass der EMS sowohl Studiendauer als auch Studienqualität sehr gut vorhersagen kann – beides sind wichtige Erfolgskriterien und damit Determinanten der Eignung. Es werden diejenigen bevorzugt zugelassen, die ihr Studium in optimaler Zeit mit besseren Leistungen abschliessen (bisher nachgewiesen bis zur 2. Vorprüfung).

Die Eignungsabklärung muss langfristig nicht auf den Leistungsaspekt wie im vorliegenden EMS begrenzt sein. Diese Begrenzung ist entstanden, weil die Leistungserfassung bisher die einzige Methode ist, welche alle Kriterien der Wissenschaftlichkeit und Ökonomie erfüllt. In diesem Zusammenhang wird häufig bemängelt, dass Merkmale der sozialen Kompetenz nicht als Eignungskriterien berücksichtigt werden.

Der wohl wichtigste Grund ist das Fehlen geeigneter Methoden. Lässt man soziales Verhalten in Situationen beobachten (etwa in Eignungsgesprächen), müssen die Beurteiler auf einen einheitlichen Masstab hin geschult werden. Ausserdem sind für jede Beurteilung mindestens 2-3 Urteiler notwendig, damit die Entscheidungen rechtsfähig sind. Das ist für alle Bewerberinnen und Bewerber um ein Medizinstudium nicht bezahlbar. Ein weiterer Nachteil ergibt sich aus der hochgradigen Trainierbarkeit von Sozialverhalten. Es gibt auf dem Gebiet der sogenannten Assessment-Center (einer verbreiteten Methode der Eignungsdiagnostik) einen „boomenden“ Trainingsmarkt, wo die Bewerberinnen und Bewerber entsprechend vorbereitet werden – zum Teil zu beträchtlichen Kosten. Das kann für die Bewerbung zum Medizinstudium nicht anstrebenswert sein.

Objektivere, ökonomischere und weniger trainierbare Methoden zur Erfassung von Aspekten sozialer Kompetenz als Bestandteile zukünftiger EMS-Versionen sind allerdings nicht gänzlich ausgeschlossen. Hier muss die Grundlagenforschung – beispielsweise zur sozialen Wahrnehmung – noch wichtige Vorleistungen erbringen. Dies sollte in Verbindung mit der Analyse der neuen Anforderungen an das Medizinstudium, die sich aus den Reformmodellen der einzelnen Universitäten ergeben, geschehen. Wichtig wäre, entsprechende Forschungen rechtzeitig anzuregen. Auch diese neuen Methoden müssen genauso intensiv entwickelt und überprüft werden wie die jetzt verwendeten Testteile.

Es ist legitim, auch andere Zulassungskriterien daraufhin zu überprüfen, ob sie die notwendigen Anforderungen erfüllen und dabei bezahlbar bleiben. Ein Teil der Wunschvorstellungen, Eignung auch auf Aspekte der beruflichen Eignung auszudehnen, wird an zu hohen Kosten oder fehlender Erfassbarkeit durch Methoden sowie mangelnder Rechtsgleichheit bzw. Willkürfreiheit scheitern. In der Tabelle 1 wurde der Versuch unternommen, die einzelnen Kriterien, die immer wieder in die Diskussion eingebracht werden, vergleichend zu bewerten. Letztlich wird deutlich, dass der jetzige EMS gegenüber allen anderen Zulassungskriterien in jedem einzelnen Bewertungsaspekt mit zu den Besten gehört. Die übrigen Kriterien haben Nachteile, welche die Praktikabilität einschränken.

	Eignungstest	Maturitätsnote	Aufnahmegespräche	Sozialpraktika	Intrauniversitäre Selektion
Erfasst Eignung?	Ja, wissenschaftlich nachgewiesene Vorhersage von Studiendauer und Studienleistungen	Möglich, nicht nachgewiesen (Benotungsunterschiede werden sich hier begrenzend auswirken)	Möglich, wenn standardisierte Beurteilung, mehrere Beurteiler und Training dieser Beurteiler	Nein, dient mehr dem Erkennen der Neigung. Betreuer würden durch Notwendigkeit der Eignungseinschätzung überfordert – Anforderung wie Aufnahmegespräche!	Ist <b>nicht</b> dem Test überlegen!
Kapazitätsentsprechende Zulassung?	Ja	Ja	Ja	Nein alle erfolgreichen Praktikanten wollen dann auch studieren	Weiter Überlastung der Kapazität im 1. Jahr, weiter grössere Abbrecherzahl nach 1 Jahr; Prüfung ist versteckter NC.
Rechtsgleichheit	Ja	Nein kantonale, schul- und lehrerbedingte Benotungsunterschiede	Nur bei Training der Beurteiler und Standardisierung der Bedingungen (aufwändig)	Nur wenn wirklich gleiche Praktikums- und Beurteilungbedingungen bestehen würden	Ja
Willkürfreiheit	Ja				Wenn Prüfungen vergleichbar sind
Ökonomie	Ja	„anfallend“	Nicht für alle realisierbar (dann zu teuer)	Nein (in Zürich erwies es sich als nicht durchführbar wg. fehlender Plätze)	Hohe Kosten im 1. Jahr
Negative Folgen?	Nein	Rückwirkung auf Benotungssystem der Maturität	„Trainingsmarkt“ würde wachsen	Geschultes Betreuungspersonal nötig (Kosten)	Versteckter NC

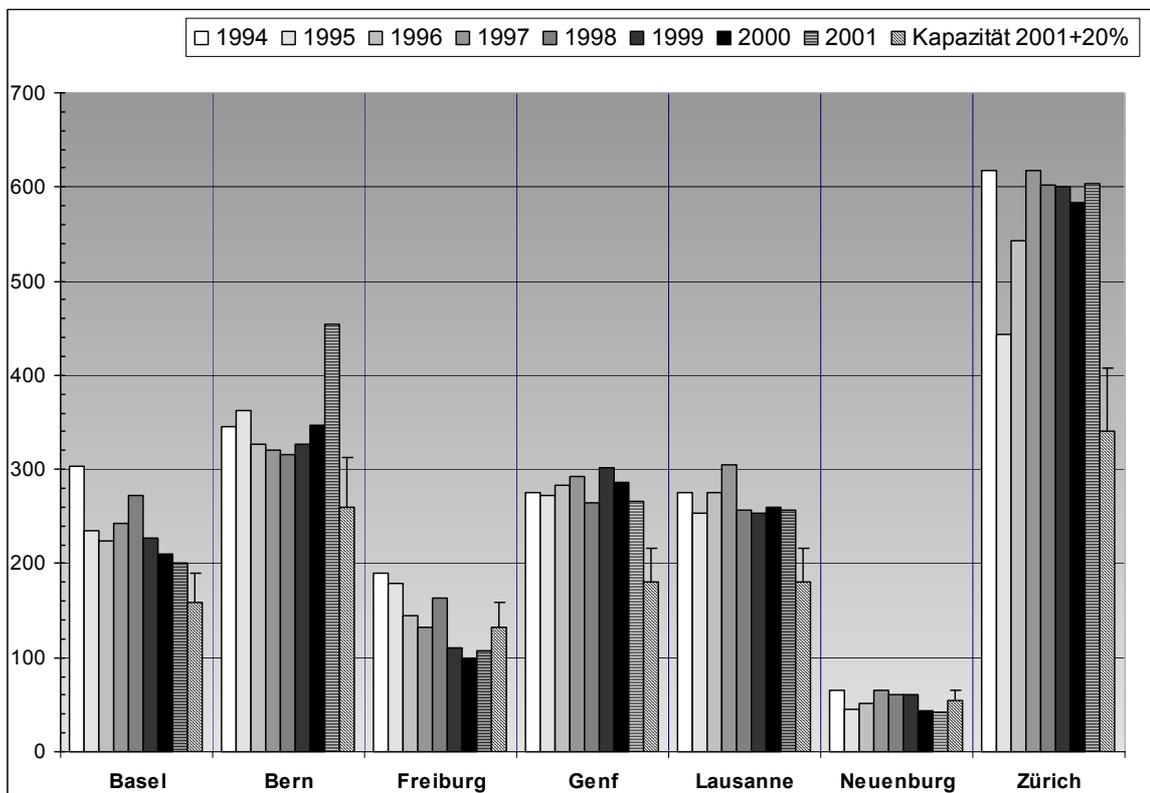
**Tabelle 1:** Mögliche Zulassungskriterien und deren Bewertung

Zusammenfassend kann man festhalten:

- Wenn eine Zulassung nach Eignung erfolgen soll, muss das „WOFÜR geeignet“ klar sein (also die Anforderungen müssen bekannt sein).
- Die Berufseignung ist bezüglich ihrer Anforderungen wenig eindeutig – es gibt auch verschiedene Disziplinen und Tätigkeiten innerhalb der Medizin, die sich stark unterscheiden. Die Beschränkung auf Studieneignung ist sinnvoll, denn...
- ... der Erfolg bei der Erfüllung der Anforderungen muss klar messbar sein (bei Studieneignung etwa die Studiendauer und die Studienleistungen). Berufserfolg ist kaum ausreichend definierbar.
- Es müssen Zulassungskriterien vorhanden sein, welche diesen Erfolg vorhersagbar machen und damit eine Unterscheidung der Personen nach der Eignung zulassen.
- Das Zulassungskriterium ist nur dann gerechtfertigt, wenn wissenschaftlich nachweisbar ist, dass es dafür sorgt, dass die Geeigneteren eine höhere Chance erhalten, zum Studium zugelassen zu werden.

## 4 Numerus Clausus und Medizinstudium

Die Abbildung 1 verdeutlicht den Anmeldeverlauf für das Medizinstudium von 1994 bis ins Jahr 2001. Als Kriterium wurden auch die Kapazitäten 2001 und das 20%-Überschreitungskriterium (als Entscheidungshilfe für die Festlegung eines Numerus Clausus) mit dargestellt. Zu beachten ist, dass 2001 die disziplinspezifischen Kapazitäten vor allem wegen der Doppelmaturitäten in Bern erhöht worden sind. Besonders in Bern findet sich im Jahr 2001 daher ein hoher Anstieg der Bewerbungszahlen. Aber auch die Zahl für Zürich bleibt unverändert hoch – 2002 werden bekanntlich dort Doppelmaturitäten erwartet.



**Abbildung 1:** Anmeldungen zum Medizinstudium 1994 bis 2001, Anfängerprognose und Kapazitäten 2001 (letzte Säule - mit Markierung des 20%-Überschreitungskriteriums für 2001 als Grenze für die Notwendigkeit einer Zulassungsbeschränkung)

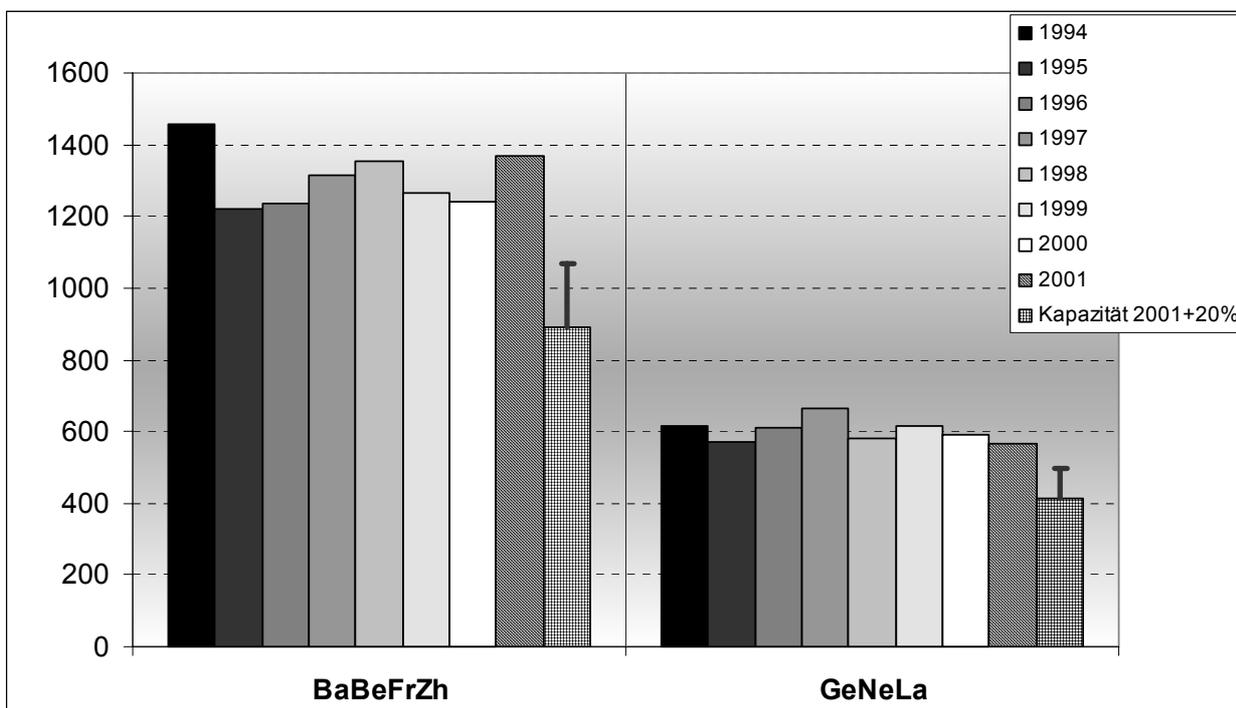
Zusätzliche Maturitäten infolge der Verkürzung der Maturitätszeiten von 13 auf 12 Jahre ergaben und ergeben sich infolge von Doppeljahrgängen:

1998: 780 (SZ, SG); 1999: 0; 2000: 60 (OW, AI); 2001: 1810 (BE, ZG, SH); 2002: 3550 (ZH, BE, LU, UR, NW, SO, NE); 2003: 670 (BE, GR) (Quelle: Bundesamt für Statistik. Bildung und Wissenschaft H 15. S. 39). Neuere Analysen<sup>1</sup> weisen allerdings darauf hin, dass der Zeitpunkt der Maturitätsprüfung für die Studienaufnahme entscheidend ist und sich diese Doppeljahrgänge auf mehrere Jahre verteilen – die Wartezeiten zum Teil auch bei der

<sup>1</sup> Bundesamt für Statistik (BfS) (2001). Doppelte Maturitätsjahrgänge und Zahl der Studienanfänger. Nur ein Fehlalarm? Vision - Das Schweizer Magazin für Wissenschaft und Innovation 2 S. 60-61.

Bewerbung bewusst so gelegt werden, dass ein Studienbeginn in den „überlasteten“ Anfangszeiten vermieden wird. Dies bedeutet, dass sich dieser grössere Ansturm auf mehrere Jahre verteilt und es nicht zu echten Verdoppelungen der Bewerberzahlen in den betroffenen Kantonen kommt. Die Zahlen aus 2001 für Bern sprechen dafür – wenn auch 2002 noch in Bern eine relative Zunahme gegenüber den Vorjahren vorhanden sein wird, wäre diese Theorie bestätigt.

Der Numerus Clausus hat wiederum nicht zu einer Abwanderung in die Westschweiz geführt (Abbildung 2). Die Bewerberzahl in der Westschweiz bleibt konstant – in den am NC beteiligten Universitäten ist der Anstieg vor allem auf die Situation in Bern zurückzuführen. Auch in der Westschweiz wäre das 20%-Kriterium erfüllt. Hier erfolgt die Selektion allerdings intrauniversitär im Verlauf des ersten Studienjahres.



**Abbildung 2:** Anmeldungen zum Medizinstudium 1994 bis 2001 nach Gruppen (**B**asel, **B**ern, **F**reiburg, **Z**ürich vs. **G**enf, **N**euenburg, **L**ausanne)

Für 2002 wäre die Vermutung aufgrund der Studie des Bundesamtes für Statistik, dass in Bern die Zahl der Bewerbungen weiter erhöht bliebe (als Nachwirkung der Doppelmaturität im Sinne der Verteilung der Studienaufnahmen auf die benachbarten Jahre) und dazu in Zürich wegen der aktuellen Doppelmaturitäten eine Zunahme der Bewerbungen vorhanden sein wird. In Zürich werden daher die Aufnahmekapazitäten erhöht, was einen Teil der Zunahme abfangen dürfte.

## 5 Anmeldung zu Medizinstudium und Test 2001

### 5.1 Anmeldungen, Testantritte und Zulassungen

Die Tabelle 2 fasst den jährlichen Anmeldeverlauf bis zur Testabsolvierung zusammen. Im Fach Humanmedizin ist der Bewerbungsindex seit 1998 ausserordentlich stabil (siehe auch Abbildung 3). Wie in den Vorjahren können etwa 80% der Testabsolventen einen Studienplatz erhalten. Seit dem Jahr 2000 hat auch die Zahl der Personen, die ihr Testergebnis aus dem Vorjahr übernehmen, eine beachtenswerte Grösse. Rückzüge in einem Jahr müssen also nicht bedeuten, dass kein Studium aufgenommen wird.

	Humanmedizin				Veterinärmedizin			Gesamt		
	1998	1999	2000	2001	1999	2000	2001	1999	2000	2001
<b>Kapazitäten</b>	<b>620</b>	<b>546</b>	<b>546</b>	<b>583</b>	<b>150</b>	<b>150</b>	<b>165</b>	<b>696</b>	<b>696</b>	<b>748</b>
Anmeldungen Mitte Februar d. J.	1020	845	844	<b>974</b>	231	213	<b>217</b>	1076	1057	<b>1191</b>
... in % zu Kapazität	186%	155%	154%	<b>167%</b>	153%	142%	<b>132%</b>	154%	152%	<b>159%</b>
Anmeldungen Test im Mai d. J.	797	717	678	<b>752</b>	194	176	<b>179</b>	911	854	<b>931</b>
...in % zum Februar	78,1	84,9	80,3	<b>77,7</b>	84,0	82,6	<b>82,5</b>	84,7	80,8	<b>78,1</b>
Übernahme Testergeb- nis aus dem Vorjahr	0	5	33	<b>27</b>	0	5	<b>5</b>	5	38	<b>32</b>
Total Bewerbungen	797	722	711	<b>779</b>	194	181	<b>184</b>	916	892	<b>963</b>
... in % zur Kapazität	129%	132%	130%	<b>134%</b>	129%	121%	<b>112%</b>	132%	128%	<b>129%</b>
Absolvierung Test	750	673	637	<b>709</b>	182	164	<b>171</b>	855	801	<b>880</b>
<b>Total der Bewerbungen im Juli</b>	<b>750</b>	<b>678</b>	<b>670</b>	<b>730</b>	<b>182</b>	<b>169</b>	<b>176</b>	<b>860</b>	<b>839</b>	<b>906</b>
Rückzugsquote zwischen Februar und Testtermin	26,5%	19,8%	20,6%	<b>25,1%</b>	21,2%	20,7%	<b>18,9%</b>	20,1%	20,6%	<b>24,0%</b>
Personen mit gültigem Testwert, denen ein Studienplatz zugewie- sen werden kann	82,7%	80,5%	81,4%	<b>79,9 %</b>	82,4%	88,7%	<b>93,7 %</b>	80,9%	82,9%	<b>82,6%</b>
Bewerbungsindex Bewerber pro Studienplatz	<b>1,21</b>	<b>1,24</b>	<b>1,23</b>	<b>1,25</b>	<b>1,21</b>	<b>1,13</b>	<b>1,06</b>	<b>1,23</b>	<b>1,21</b>	<b>1,21</b>

**Tabelle 2:** Disziplinspezifische Statistiken für die am NC beteiligten Universitäten Basel, Bern, Freiburg und Zürich; NC 1998 nur für Humanmedizin – 1999 bis 2001 für Human- und Veterinärmedizin

Für die Rückzüge im Jahr 2001 gäbe es die Erklärung, dass sich viele Personen bis zur Anmeldefrist im Februar noch alle Optionen offen halten wollten, es ihnen dann aber im Sinne der Umverteilung nicht attraktiv schien, im Jahr der Doppelmaturitäten ein Studium zu beginnen. Möglicherweise ist aus diesem Personenkreis mit mehr Bewerbungen im Folgejahr zu rechnen.

In Veterinärmedizin können von Jahr zu Jahr mehr Testabsolventen einen Studienplatz erhalten – wobei auch hier 2002 die Kapazität erhöht worden ist. Man könnte ableiten, dass diese Erhöhung relativ zu derjenigen in Humanmedizin zu deutlich erfolgte.

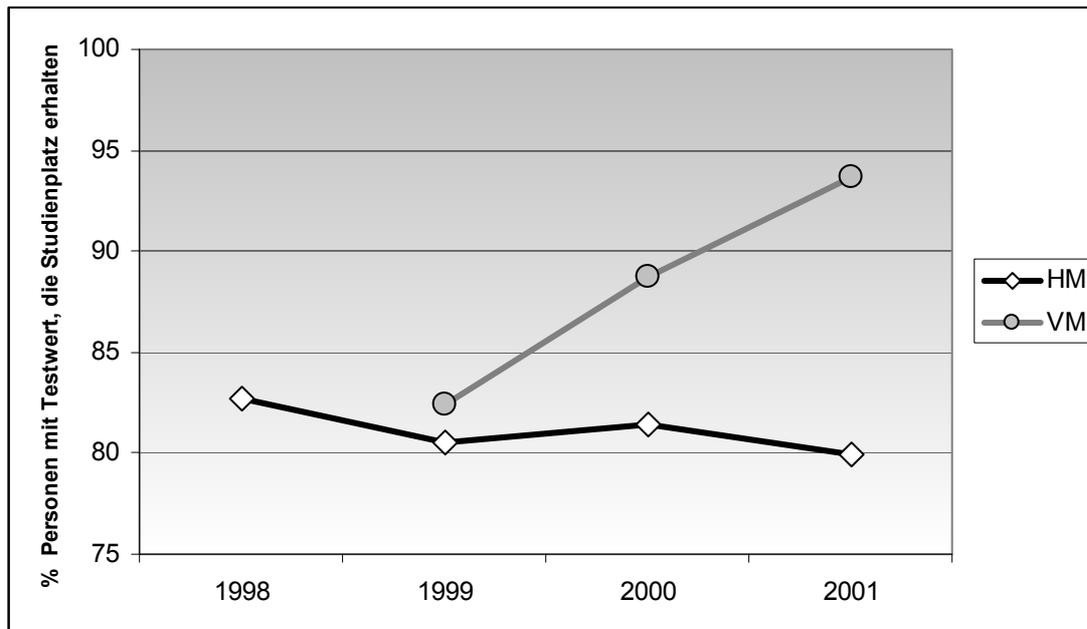


Abbildung 3: Prozent der Personen mit gültigem Testwert, denen ein Studienplatz zugewiesen werden kann – HM: Humanmedizin; VM: Veterinärmedizin 1998-2001

Die folgenden Abbildungen stellen Anmeldungen (Februar), Testanmeldungen und die Absolvierung des Tests im Verhältnis zur Kapazität für die bisherigen Testanwendungen dar. In Humanmedizin bleibt die Bewerberzahl bis zur Absolvierung des Tests die zweithöchste seit seiner Anwendung. Die Rückzüge im ersten Jahr 1998 sind mit denjenigen im Jahr 2001 vergleichbar – es hat aufgrund der Doppelmaturitäten kein stärkeres Rückzugsverhalten gegeben.

In Veterinärmedizin wird deutlich, dass das Rückzugsverhalten der letzten 3 Jahre durchaus vergleichbar ist. Es kam hier nicht zu einem nennenswerten Anstieg der Bewerberzahlen aufgrund der Doppelmaturitäten (die Situation ist vergleichbar mit dem Jahr 2000) – die Kapazitätserhöhung entspricht nicht diesem Anstieg. Zu beachten ist, dass die Bewerberinnen und Bewerber im Fach Veterinärmedizin aus mehr Herkunftskantonen als in der Humanmedizin stammen und somit der Anteil aus dem Kanton Bern geringer ist.

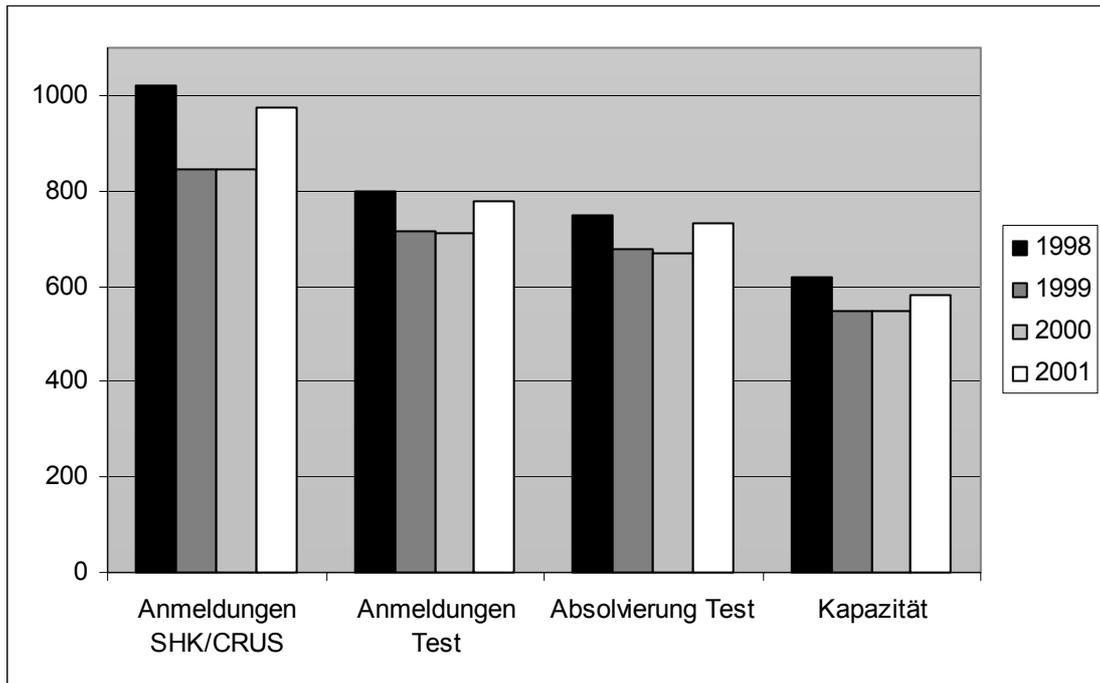


Abbildung 4: Anmeldeverlauf für Humanmedizin 1998 bis 2001 im Vergleich (mit NC)

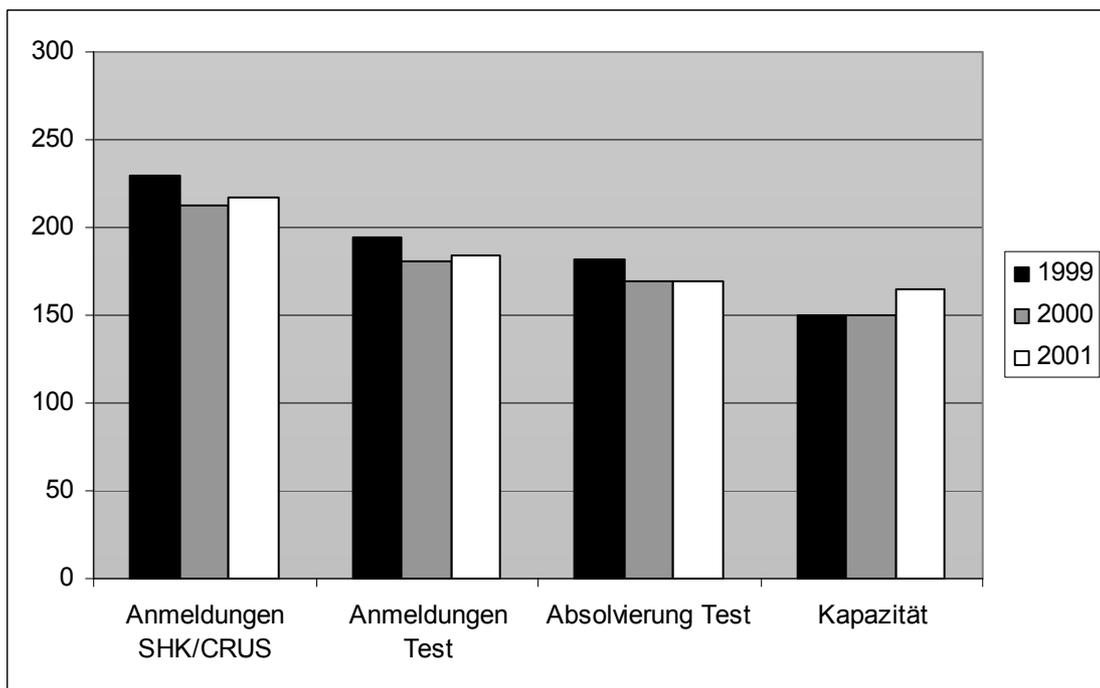
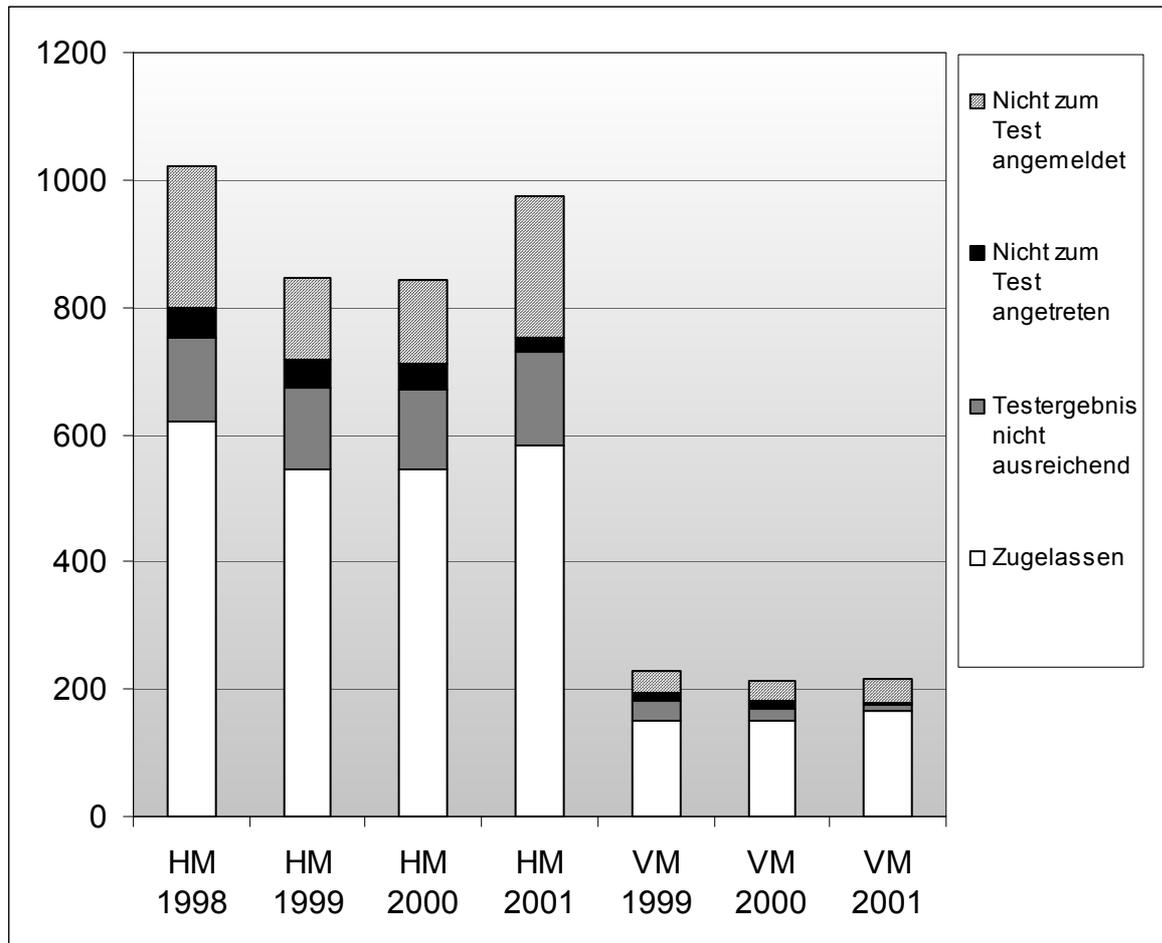


Abbildung 5: Anmeldeverlauf Veterinärmedizin 1999 bis 2001 im Vergleich (mit NC)



**Abbildung 6:** Kapazität, Anmeldungen und Testteilnahme für Human- (HM) und Veterinärmedizin (VM). Aufteilung der Voranmeldungen vom Februar des jeweiligen Jahres

In der Abbildung 6 werden die Anteile der einzelnen Gruppen bezogen auf die Gesamtzahlen der Anmeldung dargestellt. Deutlich wird, dass die Rückzüge zwischen Februar und der Testanmeldung zwischen den Jahren und innerhalb beider Disziplinen vergleichbar sind und die Rückzüge zwischen Testanmeldung und –absolvierung geringer geworden sind. Im Mai steht also bei mehr Personen der Entschluss fest, dass man sich für Medizin bewirbt oder nicht. Zwei Faktoren können dafür verantwortlich sein: Zum einen kann die Testangst durch die nun schon mehrjährige Bekanntheit des Tests abgebaut sein und die Zahl der „Panikabmeldungen“ zurückgehen. Zum anderen sind jetzt in mehr Kantonen die gesetzlichen Grundlagen vorhanden, Teilnahmegebühren für die Testabsolvierung zu erheben. Man wird sich nur dann anmelden und die Teilnahmegebühr zahlen, wenn eine ernstere Absicht zur Studienaufnahme besteht.

In der nachfolgenden Tabelle 3 und der Abbildung 7 (Quelle: CRUS) werden die Rückzüge zwischen Anmeldung im Februar und Testabsolvierung noch einmal nach Disziplinen und Universitäten differenziert. Es existieren bis zu diesem Zeitpunkt beträchtliche Unterschiede zwischen den Universitäten. Vor allem in Bern ist die Rückzugsquote gering – in Freiburg und im letzten Jahr in Zürich ist sie höher. In Zürich wären die Überlegungen der BfS-Studie (Seite 19) zu beachten: im Jahr vor der Doppel maturität sollen sich danach Personen bereits bewerben, die sonst gern noch ein Jahr gewartet hätten, um dem Jahr der Doppel maturität zu entgehen. Vielleicht ist es bei diesem Personenkreis beim Vorsatz geblieben und man hat sich zwar angemeldet – dann aber doch für das Warten entschieden.

	Studienbeginn 1999			Studienbeginn 2000			Studienbeginn 2001		
	Anmel- dungen am 15.2.	Pers. mit gültigem Testwert	Rückzüge in %	Anmel- dungen am 15.2.	Pers. mit gültigem Testwert	Rückzüge in %	Anmel- dungen am 15.2.	Pers. mit gültigem Testwert	Rückzüge in %
Basel HM	189	151 (2)	20.1	174	133 (9)	23.6	175	136 (7)	22.3
Bern HM	164	132 (2)	19.5	177	144 (5)	18.6	282	232 (5)	17.7
Freiburg HM	98	79 (0)	19.4	88	66 (2)	25.0	97	69 (1)	28.9
Zürich HM	394	316 (1)	19.8	405	327 (17)	19.3	420	299 (14)	28.8
<b>Total HM</b>	<b>845</b>	<b>678</b>	<b>19.8</b>	<b>844</b>	<b>670</b>	<b>20.6</b>	<b>974</b>	<b>736</b>	<b>24.4</b>
Bern VM	112	85 (0)	24.1	117	99 (0)	15.4	118	104 (4)	11.9
Zürich VM	119	97 (0)	18.5	96	70 (0)	21.7	99	72 (1)	27.3
<b>Total VM</b>	<b>231</b>	<b>182</b>	<b>21.2</b>	<b>213</b>	<b>169</b>	<b>20.7</b>	<b>217</b>	<b>176</b>	<b>18.9</b>

Tabelle 3: Anmeldeverlauf für Universitäten: Humanmedizin (HM) und Veterinärmedizin (VM). Personen mit gültigem Testwert - in Klammern: davon Übernahme aus dem Vorjahr

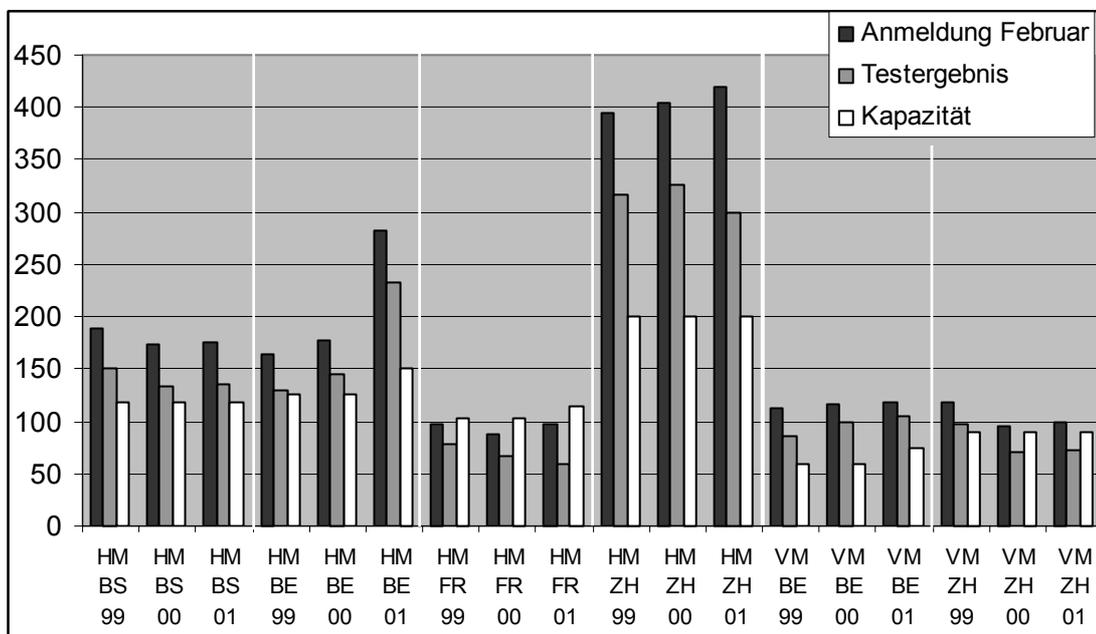


Abbildung 7: Kapazität, Anmeldungen und Testteilnahme für Fächer und Universitäten

## 5.2 Grösse der Testlokale

In diesem Jahr erfolgte die Testabnahme in 13 Testlokalen in 7 Testorten (diese wie seit 1998). In Bern waren erstmals 3 Testlokale notwendig, in Zürich dagegen nur vier. Die Grössen variierten im üblichen Rahmen. Die Zahl der Betreuer war wie bisher proportional zur Teilnehmerzahl vorgegeben. Es gibt keine Hinweise seitens des Betreuungspersonals oder von Kandidaten, dass unterschiedliche Raumgrössen sich auf die Ergebnisse auswirken. In den grossen Räumen ist der zur Verfügung stehende Platz teilweise grosszügiger (wenn es sich um grössere Hörsäle handelt).

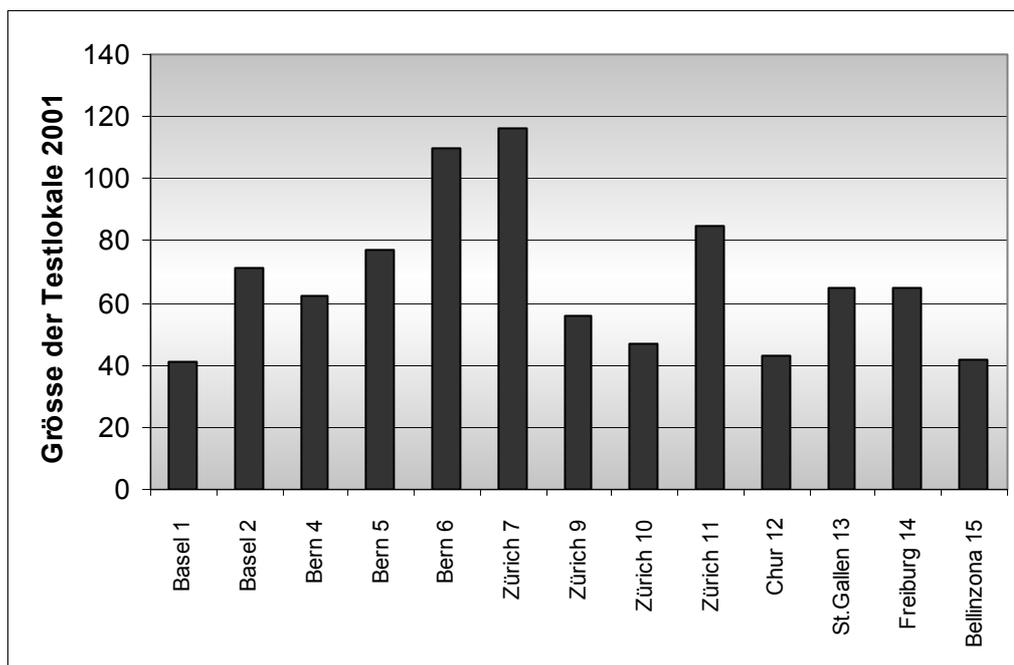


Abbildung 8: Grösse (Zahl der Personen) und Aufteilung der Testlokale 2001

## 5.3 Testorte und Wunschuniversitäten

Die Wahl des Testortes hat keinen Einfluss auf den zukünftigen Studienort. Es wurde empfohlen, den Testort zu wählen, der dem Wohnort am nächsten liegt. Längere Anreisen könnten sich negativ auf die Leistungsfähigkeit auswirken.

Wie in beiden Vorjahren zeigen sich Übereinstimmungen zwischen Test- und Wunsch-Studienort, da auch die Wahl der Universität häufig nach der Wohnortnähe erfolgt. Bekanntlich kann auch eine Umleitung vermieden werden, wenn der Wohnkanton mit dem Kanton der gewünschten Universität identisch ist. Diesen Vorteil will man offenbar nutzen.

Abbildung 9 macht durch die "Streckung" aller Universitäten auf 100% das Verhältnis der "lokalen" Bewerber zu den übrigen vergleichbar.

Testort	Vor Anmeldung Universität				Total
	Bern	Basel	Freiburg	Zürich	
Basel	10 8.9%	91 81.3%	3 2.7%	8 7.1%	112
Bellinzona	3 7.1%	4 9.5%	3 7.1%	32 76.2%	42
Bern	221 88.8%	8 3.2%	14 5.6%	6 2.4%	249
Chur	13 30.2%	10 23.3%	2 4.7%	18 41.9%	43
Freiburg	38 58.5%	1 1.5%	21 32.3%	5 7.7%	65
St. Gallen	13 20.0%	7 10.8%	10 15.4%	35 53.8%	65
Zürich	29 9.5%	8 2.6%	14 4.6%	253 83.2%	304
<b>Total</b>	<b>327</b> 37.2%	<b>129</b> 14.7%	<b>67</b> 7.6%	<b>357</b> 40.6%	<b>880</b>

Tabelle 4: Testorte und Wunschuniversität (erste Wahl) 2001

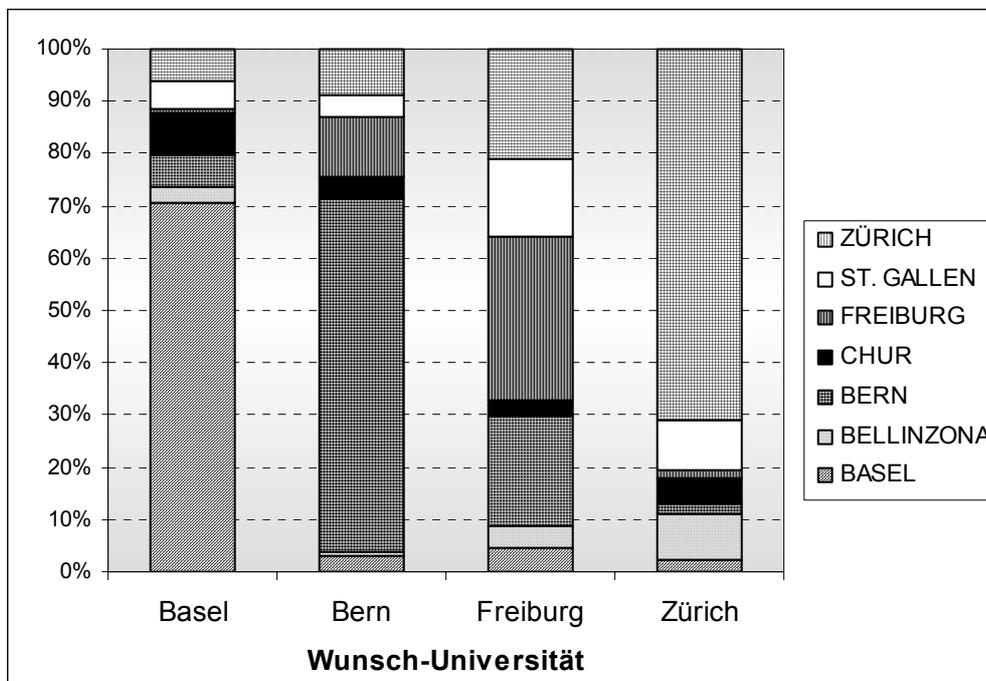


Abbildung 9: Aufteilung der Kandidaten nach Wunschuniversität auf die Testorte

## 5.4 Testorte und Wohnkantone

Nachfolgend sind die Testorte im Verhältnis zum Wohnkanton tabelliert. Angegeben sind auch die Vergleichswerte für 1999. In jenem Jahr gab es keine Doppeljahrgänge der Maturanden. Im Jahr 2000 sind Doppeljahrgänge in OW (40 zusätzliche Abschlüsse) und AI (20) vorhanden. Im Jahr 1998 gab es Doppeljahrgänge in SZ (160) und SG (620). Zu beachten ist, dass 1998 der NC nur für Humanmedizin galt.

Wohnkanton/ Wohnort	Testort 2001							Total			
	BASEL	BELLIN- ZONA	BERN	CHUR	FREI- BURG	ST. GALLEN	ZÜRICH	2001	2000	1999	1998 (HM)
AG	9		6				34	<b>49</b>	71	84	66
AI						1		<b>1</b>	<u>4</u>	0	2
AR						10		<b>10</b>	4	6	4
BE	1		177		5	1	2	<b>186</b>	114	113	100
BL	45							<b>45</b>	40	59	50
BS	36	1	1				1	<b>39</b>	28	39	40
FR			8		19			<b>27</b>	44	47	34
GE					10			<b>10</b>	11	13	1
GL							3	<b>3</b>	5	4	4
GR	2	2		32				<b>36</b>	19	29	19
JU					1			<b>1</b>	1	2	0
LU	1		21				25	<b>47</b>	41	44	43
NE					3			<b>3</b>	6	5	2
NW			1				2	<b>3</b>	2	3	6
OW	1		1			1		<b>3</b>	<u>9</u>	2	3
SG			1	8		32	7	<b>48</b>	55	61	<u>64</u>
SH							18	<b>18</b>	8	4	7
SO	9		13				3	<b>25</b>	20	28	26
SZ			1			1	13	<b>15</b>	18	18	<u>21</u>
TG						14	24	<b>38</b>	23	20	22
TI	1	37						<b>38</b>	27	29	30
UR	1						4	<b>5</b>	2	1	3
VD			1		22			<b>23</b>	17	11	3
VS	1		13		5			<b>19</b>	18	15	22
ZG			1				22	<b>23</b>	9	10	6
ZH		2	2	2			142	<b>148</b>	194	191	164
FL	1			1		4	1	<b>7</b>	4	2	3
Übriges Ausland	4		2			1	3	<b>10</b>	11	17	8
	112	42	249	43	65	65	304	<b>880</b>	801	855	750

Tabelle 5: Wohnkanton und Testorte, Vergleich zu 1998 bis 2000 (1998 nur Humanmedizin). Unterstrichen: Doppeljahrgänge der Maturanden in den Kantonen. Das Fürstentum Liechtenstein wird als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt

Wegen der geringen Absolutzahl ist in kleineren Kantonen mit Doppeljahrgängen allenfalls ein Trend zur Zunahme zu erkennen. Auch in den anderen Kantonen finden sich starke Fluktuationen der Bewerberzahlen zwischen den Jahren. In diesem Jahr ist die Zunahme der Bewerberzahlen in allen Kantonen mit Doppelmaturität allerdings recht deutlich.

Die Präferenz des gewünschten Studienortes in Abhängigkeit vom Wohnort wird in der Tabelle 6 dargestellt. Die Präferenzen scheinen vor allem geografisch motiviert zu sein. Zu beachten ist, dass Veterinärmedizin für die gesamte Schweiz seit 1999 nur in Bern und Zürich angeboten wird.

Wohnkanton/ Wohnort	Vor Anmeldung Universität 2001				Total 2001	Total 2000	Total 1999
	Basel	Bern	Freiburg	Zürich			
AG	11	8	2	28	49	71	84
AI	1				1	4	0
AR	1	2	5	2	10	4	6
BE	4	173	6	3	186	114	113
BL	37	2	3	3	45	40	59
BS	29	5		5	39	28	39
FR		3	23	1	27	44	47
GE	1	8		1	10	11	13
GL			1	2	3	5	4
GR	12	8		16	36	19	29
JU				1	1	1	2
LU	3	24	3	17	47	41	44
NE		3			3	6	5
NW	1	1		1	3	2	3
OW	1	2			3	9	2
SG	5	12	6	25	48	55	61
SH	1	6	1	10	18	8	4
SO	9	10		6	25	20	28
SZ		3	3	9	15	18	18
TG		3		35	38	23	20
TI	4	3	3	28	38	27	29
UR	1	1		3	5	2	1
VD		21	1	1	23	17	11
VS	1	14	3	1	19	18	15
ZG	1	2	2	18	23	9	10
ZH	2	9	5	132	148	194	191
FL	1	2		4	7	4	2
Übriges Ausland	3	2		5	10	11	17
<b>Total</b>	<b>129</b>	<b>327</b>	<b>67</b>	<b>357</b>	<b>880</b>	<b>801</b>	<b>855</b>

**Tabelle 6:** Gewünschte Studienorte nach Wohnkantonen, Vergleich zu 1999 und 2000. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt

Wohnkanton/-ort	Bern	Zürich	Gesamt
AG	1	6	7
AR	1	1	2
BE	<b>39</b>		39
BL	2	2	4
BS	3	4	7
FR	3		3
GE	8	1	9
GR	3	3	6
JU		1	1
LU	2	3	5
NE	3		3
SG	2	5	7
SH		3	3
SO	2	2	4
TG		8	8
TI	1	6	7
UR		1	1
VD	20	1	21
VS	7		7
ZG		4	4
ZH	1	<b>18</b>	19
FL	1	1	2
AUSLAND	1	1	2
<b>Total</b>	<b>100</b>	<b>71</b>	<b>171</b>

**Tabelle 7:** Bewerbungen 2001 für Veterinärmedizin nach Wohnkantonen. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt

Tabelle 7 geht der Frage nach, warum der Anstieg der Bewerberzahlen in Veterinärmedizin in Bern geringer ausfiel als in Humanmedizin. Während in Humanmedizin 173 von 327 Bewerbern aus dem Kanton Bern stammen (53%), sind es in Veterinärmedizin nur 39 von 100 (39%). Deshalb wirkt sich hier die Doppelmaturität weniger als in Humanmedizin aus. Die Bewerbungen in Zürich sind noch weniger auf den Kanton beschränkt. 132 von 357 (37%) Personen kommen in Humanmedizin aus dem eigenen Kanton – in Veterinärmedizin sind es 18 von 71 (25%) Personen. Vermutlich werden bei einer vergleichbaren Kapazitätserhöhung in Zürich 2002 wie in Bern 2001 die Plätze in Veterinärmedizin die Bewerberzahlen erneut kaum übersteigen.

## 5.5 Testabsolvierung nach Alter und Geschlecht

Die Zahl der Frauen überwiegt wie bisher auch im Jahr 2001 in den Fächern Human- und Veterinärmedizin die Zahl der Männer. Der relative Frauenanteil nimmt in diesem Jahr in beiden Disziplinen weiter deutlich zu. In Humanmedizin sind es vor allem die Jüngeren, wo nunmehr das Verhältnis 1/3 Männer zu 2/3 Frauen herrscht. Bei den Älteren (die allerdings weniger Personen umfassen), ist das Geschlechterverhältnis ausgeglichener. Eine mögliche Erklärung wäre, dass Männer eher eine Wartezeit zwischen Schulabschluss und Studienbeginn einlegen. Es können allerdings auch unterschiedlich wirkende Dissuasionseffekte dafür

verantwortlich sein (siehe Seite 74). In Veterinärmedizin ist das Verhältnis in jeder Altersgruppe 3/4 Frauen (und mehr) zu 1/4 Männer (und weniger).

	Geburtsjahre			Total
	1954 – 1977	1978 – 1980	1981 – 1985	
<b>Humanmedizin</b>				
<b>männlich</b>	26 51.0%	89 44.5%	158 34.5%	273 38.5% von Gesamt 1998 44.8%
<b>weiblich</b>	25 49.0%	111 55.5%	300 65.5%	436 61.5% von Gesamt 1998 55.2%
<b>Gesamt</b>	51 7.2% v. Gesamt 1998 10.4%	200 28.2% v. Gesamt 1998 29.5%	458 64.6% v. Gesamt 1998 60.1%	709
<b>Veterinärmedizin</b>				
<b>männlich</b>	4 25%	8 18.6%	23 20.5%	35 20.5% von Gesamt 1999 24.2%
<b>weiblich</b>	12 75.0%	35 81.4%	89 79.5%	136 79.5% von Gesamt 1999 75.8%
<b>Gesamt</b>	16 9.4% v. Gesamt 1999 9.9%	43 25.1% v. Gesamt 1999 26.9%	112 65.5% v. Gesamt 1999 63.2%	171

Tabelle 8: Testabsolventen nach Geburtsjahr und Geschlecht

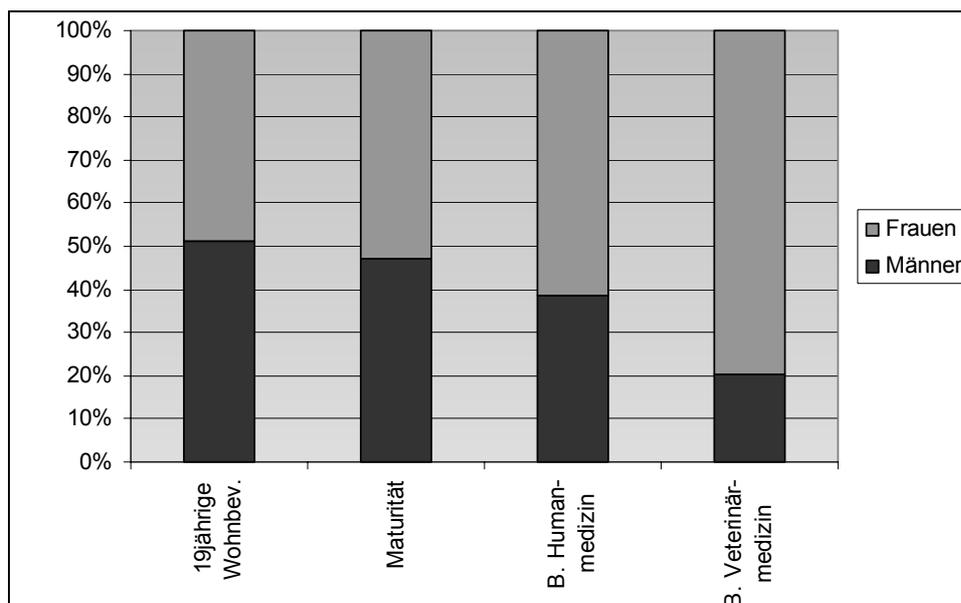


Abbildung 10: Geschlechteranteil an 19jähriger Wohnbevölkerung und Maturitäten (Stand 1999), sowie Bewerbungsverhältnisse Human- und Veterinärmedizin 2001

Die Abbildung 11 verdeutlicht, dass in Humanmedizin vor allem der Frauenanteil auch die Absolut-Unterschiede der Bewerberzahlen bedingt – die Zahl der Männer ist weiter geringfügig kleiner geworden. Die Anteile der Älteren sind 2001 mit 2000 vergleichbar (Abbildung 12), wo auch zwischen 1999 und 2000 ein Rückgang bei den Männern zu verzeichnen war.

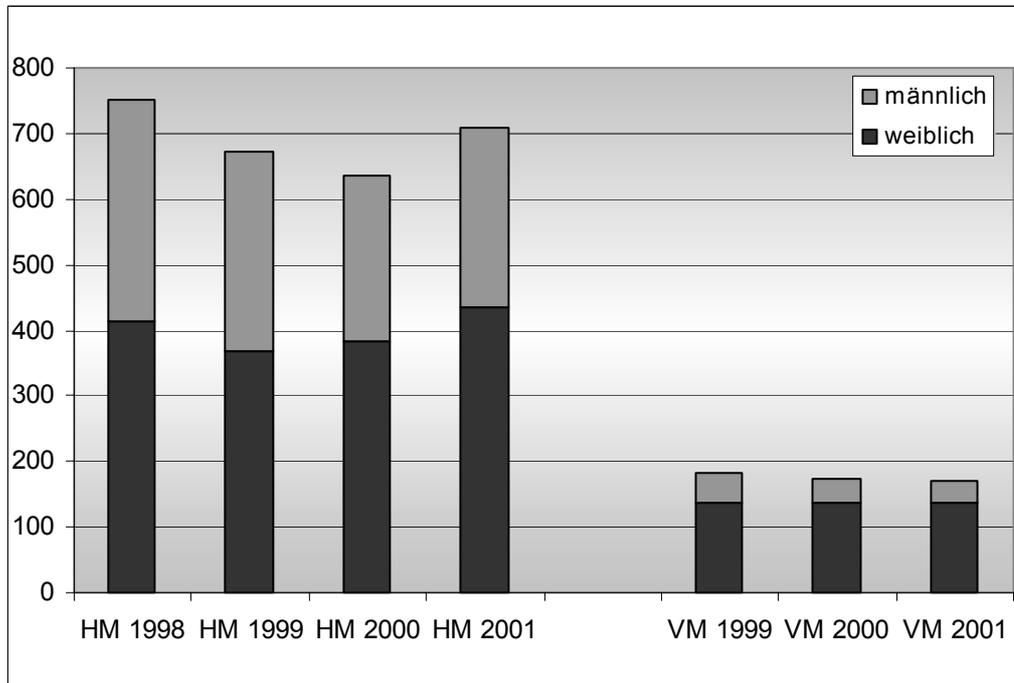


Abbildung 11: Anteil der Geschlechter bei der Bewerbung für Humanmedizin (HM) 1998 bis 2001 und Veterinärmedizin (VM) 1999 bis 2001

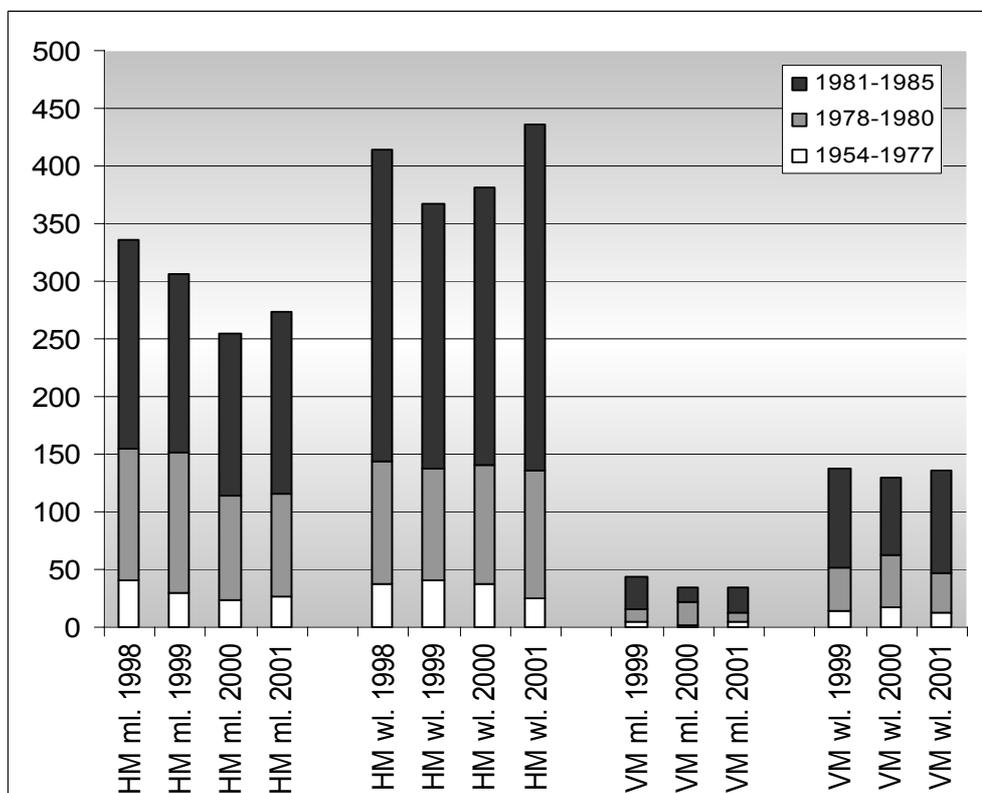


Abbildung 12: Zusammensetzung der Bewerbergruppen nach Geschlecht ml: männlich wl: weiblich und Alter (Geburtsjahre)

Insgesamt ist aufgrund der unterschiedlichen Zusammensetzung der Bewerbergruppe wiederum nicht damit zu rechnen, dass die Mittelwerte des Testwertes bei globaler Betrachtung für die Geschlechter gleich sind – unterschiedliche Repräsentativität für die Maturandengruppen wird die Ursache sein.

## 5.6 Übernahme des Testergebnisses aus 2000

Im Jahr 2001 übernehmen 32 Personen ihr Testergebnis aus dem Vorjahr. Es kann gezeigt werden, dass dieser Personenkreis im Mittel überdurchschnittliche Testwerte aufweist, was dem Vorjahresbefund entspricht. Waren 2001 noch 5,3% dieser Personen unter dem kapazitätsentsprechenden Zulassungskriterium, liegen im Jahr 2001 alle Personen über diesem Wert. Der kapazitätsabhängige kritische Testwert von 91 (Humanmedizin) bzw. 82 (Veterinärmedizin) wird von allen Personen erreicht. Offenbar ist es durch die jetzt vorhandenen Informationen leichter möglich, aufgrund des erreichten Testwertes die eigenen Chancen bei einer erneuten Bewerbung abzuschätzen.

Disziplin	N	Minimum	Maximum	Mittelwert	Stand.-abw.
Humanmedizin 2000	33	85	119	<b>102.85</b>	7.25
Humanmedizin 2001	27	<b>94</b>	119	<b>102.96</b>	7.28
Veterinärmedizin 2000	5	97	107	<b>102.00</b>	3.81
Veterinärmedizin 2001	5	<b>89</b>	112	<b>102.80</b>	9.52

Tabelle 9: Testwert für Personen mit Übernahme des Testergebnisses aus dem Vorjahr

## 5.7 Erneute Testteilnahme 2001

Neben der Übernahme des Testergebnisses aus dem Vorjahr ist es auch möglich, erneut zum Test anzutreten und dann dieses zweite Ergebnis für die Zulassung zu verwenden. Personen, die sich später als ein Jahr nach der ersten Testteilnahme noch einmal bewerben, müssen in jedem Falle noch einmal am Test teilnehmen.

Erste Testteilnahme	Zweite Testteilnahme		
	1999	2000	2001
1998	12	2	
1999		19	3
2000			12

Tabelle 10: Wiederholte Testteilnahme: Anzahl der Personen

Insgesamt 48 Personen haben seit 1998 zweimal am Test teilgenommen. Für fünf Personen beträgt der Abstand mehr als ein Jahr, 43 Personen haben auf die mögliche Anrechnung des Ergebnisses aus dem Vorjahr verzichtet. Betrag der mittlere Testwert bei der ersten Testung

87.8, ist er bei der zweiten Testung 97.0 – also einer mittleren Verbesserung um den Betrag einer ganzen Standardabweichung.

Der Testwert entscheidet im Falle einer Zulassung auch über die Universität, an welcher die Zulassung erfolgt. Deshalb haben offenbar auch Personen erneut am Test teilgenommen, deren Ergebnis für die eigentliche Zulassung gereicht hätte.

Die Tatsache, dass bei wiederholter Testung bessere Ergebnisse auftreten, spricht für die Empfehlung im Rahmen der Vorbereitung, einmal einen Probelauf unter „ernstfallnahen“ Bedingungen durchzuführen. Ein solcher wird in vielen Kantonen unter geeigneten Bedingungen angeboten. Es wäre denkbar, dass diejenigen Personen erneut zum Test antreten, die sich aus der Erfahrung der ersten Teilnahme dann auf die zweite Testteilnahme besser vorbereiten.

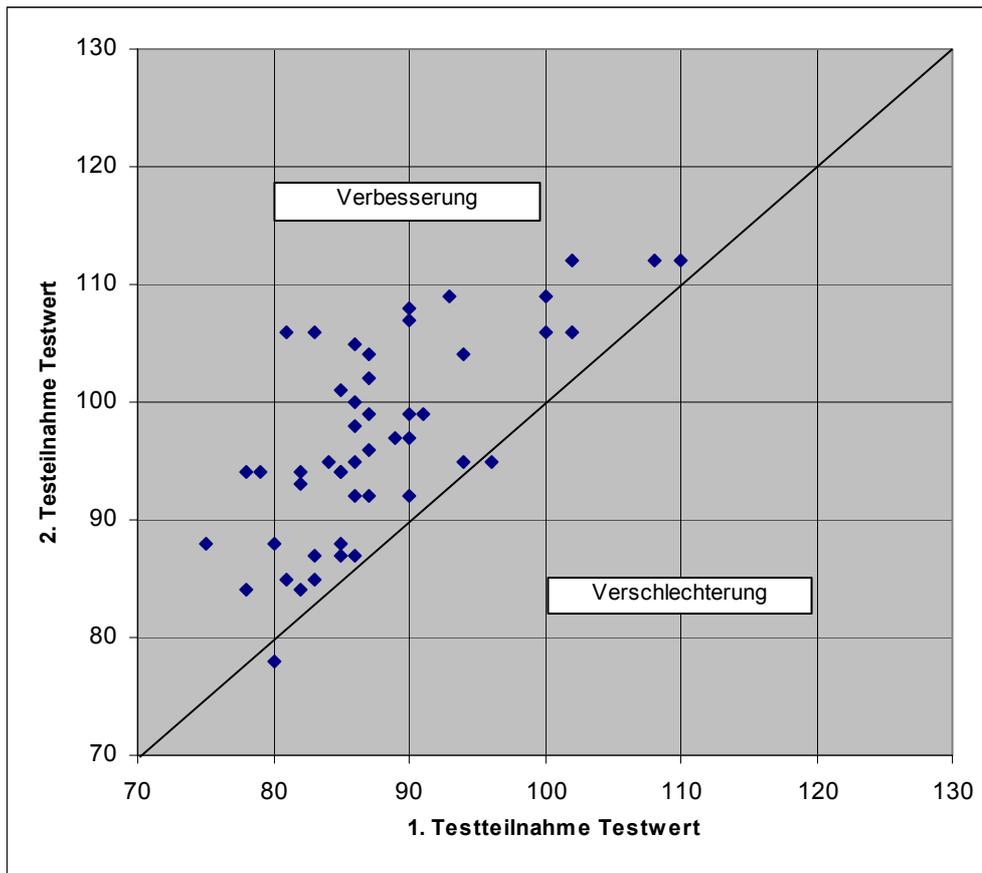


Abbildung 13: Testwerte für Personen, die zweimal am Test teilnahmen

## 5.8 Sprachgruppen

Die Zusammensetzung der Sprachgruppen unterlag in den letzten Jahren einer beständigen Dynamik: für die französisch- und italienischsprachigen Personen existieren Alternativen in der Romandie bzw. für die italienischsprachigen in Italien.

Die französischsprachige Gruppe ist in ihrer Grösse etwa gleich geblieben, während die italienischsprachige Gruppe etwas mehr Personen umfasst. Insgesamt bleiben diese beiden Gruppen aber weiter gering – entsprechende Vorsicht ist bei der Bewertung aller statistischen Vergleiche geboten.

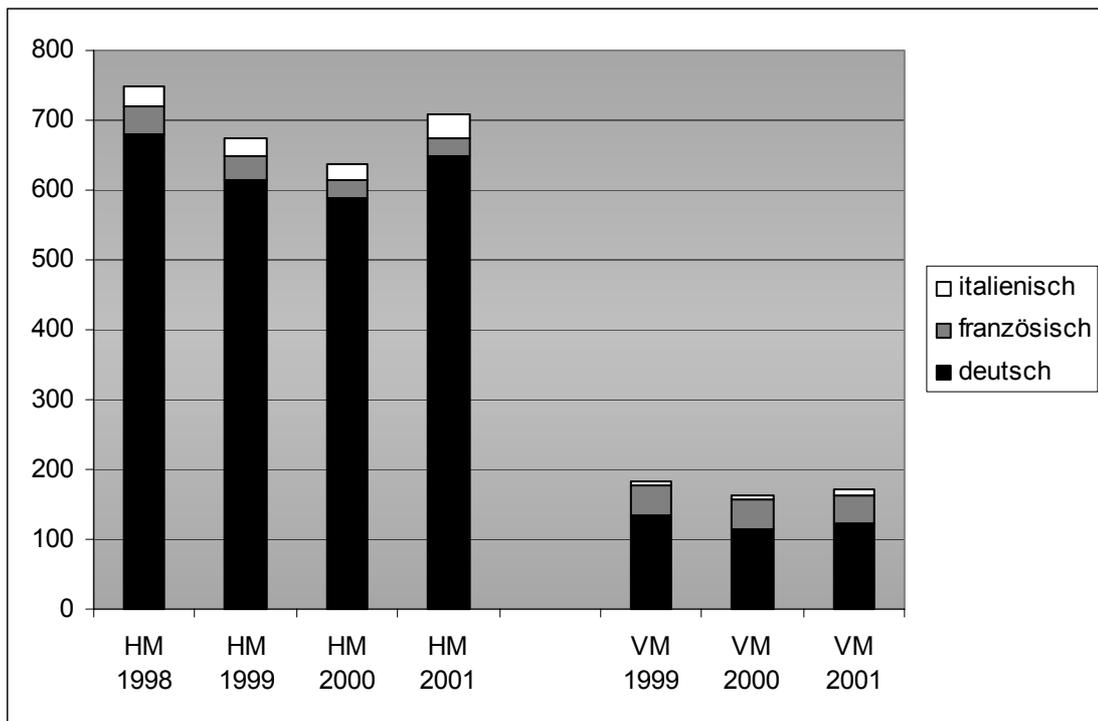


Abbildung 14: Vergleich der Sprachgruppen 1998 bis 2001

Disziplin x Sprache x Geschlecht			Klassen Geburtsjahr		
			1954-1977	1978-1980	1981-1985
Humanmedizin	deutsch	männlich	24 9.6%	85 34.0%	141 56.4%
		weiblich	24 6.0%	107 26.8%	268 67.2%
	französisch	männlich	2 28.6%	2 28.6%	3 42.9%
		weiblich	1 5.6%	3 16.7%	14 77.8%
	italienisch	männlich		2 12.5%	14 87.5%
		weiblich		1 5.3%	18 94.7%
Veterinärmedizin	deutsch	männlich	4 16.0%	7 28.0%	14 56.0%
		weiblich	10 10.1%	31 31.3%	58 58.6%
	französisch	männlich		1 12.5%	7 87.5%
		weiblich	2 6.3%	4 12.5%	26 81.3%
	italienisch	männlich			2 100.0%
		weiblich			5 100.0%

Tabelle 11: Sprachgruppen nach Alter und Geschlecht aufgeschlüsselt

## 5.9 Alter und Maturitätsjahr

In diesem Jahr tritt wegen der Doppelmaturität eine Häufung der jüngeren Jahrgänge auf. Wir teilen später die Personen analog der Vorjahre zu Vergleichszwecken in drei Altersgruppen auf (dickere Linien). Die älteste Gruppe wird wiederum danach unterschieden, ob die Maturitätsprüfung unmittelbar nach der Mittelschule absolviert wurde, oder noch eine Wartezeit dazwischen bestand.

Geburts- jahr	Maturitätsjahr														Total
	73	81	86	87	90	92	94	95	96	97	98	99	00	01	
1954	1														1
1961													1	2	3
1962													1		1
1963		1													1
1966														1	1
1967			1											1	2
1968													1	1	2
1969				1											1
1970									1	1					2
1971														1	1
1972					1									2	3
1973						2								2	4
1974						1	1						1	6	9
1975										1				3	4
1976							1	1		1	2		2	6	13
1977										4	5	2	1	4	16
1978									1	1	7	6	6	7	28
1979											9	15	16	20	60
1980											1	7	46	99	153
1981												2	51	278	331
1982													7	192	199
1983														38	38
1984														1	1
1985														1	1
	1	1	1	1	1	3	2	1	2	8	24	32	133	665	875

**Tabelle 12:** Geburtsjahr und Maturitätsjahr, die dicken Linien zeigen die Aufteilung der Gruppen für die späteren Gruppenvergleiche

## 6 Beschreibung des verwendeten Eignungstests

### 6.1 Aufbau des Tests

Der EMS besteht aus neun verschiedenen Untertests, die zu einem Gesamtwert verrechnet werden. Die Tabelle 13 gibt einen Überblick über den Testaufbau, die Aufgabenzahl und die Dauer der einzelnen Untertests. Aufgabenbeispiele finden Sie im Anhang ab Seite 136.

Bezeichnung der Untertests	Geprüfte Fähigkeiten	Zahl der Aufgaben	Bearbeitungszeit (in Minuten)
Muster zuordnen	Differenzierte visuelle Wahrnehmung	24 (20)*	22
Medizinisch-naturwissenschaftliches Grundverständnis	Verständnis für medizinisch-naturwissenschaftliche Problemstellungen	24 (20)*	60
Schlauchfiguren	Räumliches Vorstellungsvermögen	24 (20)*	15
Quantitative und formale Probleme	Quantitatives Problemlösen in medizinisch-naturwissenschaftlichen Kontexten	24 (20)*	60
Konzentriertes und sorgfältiges Arbeiten	Konzentrationsfähigkeit, Aufmerksamkeit	1200 Zeichen 20 Punkte	8
<b>MITTAGSPAUSE 60 Minuten</b>			
Lernphase zu den Gedächtnistests: Figuren lernen Fakten lernen			4 6
Textverständnis	Verständnis und Interpretation medizinischer und naturwissensch. Texte	24 (18)*	60
Reproduktionsphase Gedächtnistests: Figuren lernen Fakten lernen	Behalten von figuralem Material Behalten von verbalem Material	20 20	5 7
Diagramme und Tabellen	Interpretation von Diagrammen und Tabellen	24 (20)*	60
<b>Gesamttest</b>		<b>204 (178)*</b>	<b>5 Std. 7 Min.</b>

**Tabelle 13:** Struktur und Ablauf des EMS. \*: Aufgabenzahl und in Klammern die Zahl der gewerteten Aufgaben pro Untertest und für den gesamten Test, da zusätzliche (nicht gewertete) Einstreuaufgaben verwendet worden sind, Erläuterungen siehe Text

Beim EMS handelt es sich um eine Adaptation der deutschen Tests für Medizinische Studiengänge (TMS). Die Struktur des TMS hat sich in Deutschland im Rahmen von 12 Testeinsätzen bewährt (1986 zwei Testtermine und von 1987 bis 1996 je ein Termin pro Jahr). Beim EMS der Schweiz sind Struktur und Testlänge mit denen des deutschen TMS identisch, um dort langjährig erprobte Abläufe und damit auch die Güte-Kennwerte nicht zu verändern.

**Seine Vorteile lassen sich folgendermassen zusammenfassen:**

- Auswahl der Testanforderungen aus einer grösseren Menge möglicher Studienanforderungen aufgrund einer Anforderungsanalyse;
- Wissenschaftliche Überprüfung der Prognosekraft dieser Anforderungen für den Studienerfolg, was die Zulassung nach der Eignung zum Studium erlaubt;
- Konstruktion der Aufgaben durch Experten UND anschliessende empirische Überprüfung, dass die Aufgaben alle Gütekriterien erfüllen, die gewünschte Fähigkeit tatsächlich messen und optimal „schwer“ 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<sup>1</sup>);
- Ökonomisches und genau kapazitätsentsprechendes 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. Auch an dieser Aufgabenentwicklung nahmen zahlreiche Lehrbeauftragte und Experten teil. Die Aufgaben müssen sehr hohe Qualitätsstandards erfüllen, u.a. (1) müssen sie jedes Jahr die Studieneignung gleich zuverlässig messen, (2) muss das Schwierigkeitsspektrum aller Aufgaben annähernd vergleichbar sein, (3) darf kein spezielles Fachwissen vorausgesetzt werden, um die Trainierbarkeit des Tests gering zu halten, (4) muss eine eindeutige richtige Lösung existieren.

Die Erprobung neuer Aufgaben für sechs der neun Untertests (siehe Tabelle 13) erfolgte in Deutschland im Rahmen sogenannter „**Einstreuaufgaben**“. Nur bei ausreichender Bewährung wurden solche Aufgaben in nachfolgenden Testversionen für die Werteberechnung verwendet. Vier neue Aufgaben pro Untertest wurden in jeder Testform probeweise mitbearbeitet – ihr Ergebnis wurde jedoch nicht gezählt. Da in Deutschland acht verschiedene Testformen bei jedem Durchführungstermin zusammengestellt worden sind, konnten jeweils 32 neue Aufgaben pro Untertest an ausreichend grossen Stichproben erprobt werden. Maximal 20 davon wurden pro Jahr gebraucht – dieser Überschuss war nach Aussagen der Entwickler auch notwendig, da nicht alle Aufgaben die Kriterien zufriedenstellend erfüllten. Es versteht sich, dass die Einstreuaufgaben nicht besonders gekennzeichnet worden sind – jede Aufgabe des Tests konnte eine solche sein.

<sup>1</sup> Zum Trainingsaspekt siehe Bericht 6 des ZTD, sowie die Argumentationen unter [www.unifr.ch/ztd/ems/](http://www.unifr.ch/ztd/ems/)

## 6.2 Berechnung der Werte

Alle Untertests, ausser dem Konzentrierten und sorgfältigen Arbeiten, liefern eine Summe ("Punkte") richtig gelöster Aufgaben zwischen 0 und 20 bzw. 18 (bei Textverständnis). Summiert werden die gewerteten Aufgaben, nicht die Einstreuaufgaben.

Beim Test Konzentriertes und sorgfältiges Arbeiten müssen insgesamt 1200 Zeichen der Reihe nach bearbeitet werden – 600 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. Sechshundert Richtige wären das Maximum und entsprechen 20 Punkten.

Alle Punkte der Untertests werden zu einer Summe addiert (Punktwert, vgl. Abbildung 15). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Geburtsjahre vergleichbar ist. Auch die Eichung verschiedener äquivalenter Sprachformen ist noch notwendig (s.u.). 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) und kann in einen Prozentrangwert umgerechnet werden. Prozentränge lassen sich am einfachsten veranschaulichen indem angegeben wird: x Prozent aller Teilnehmenden haben einen besseren oder schlechteren Testwert erreicht als die entsprechende Person.

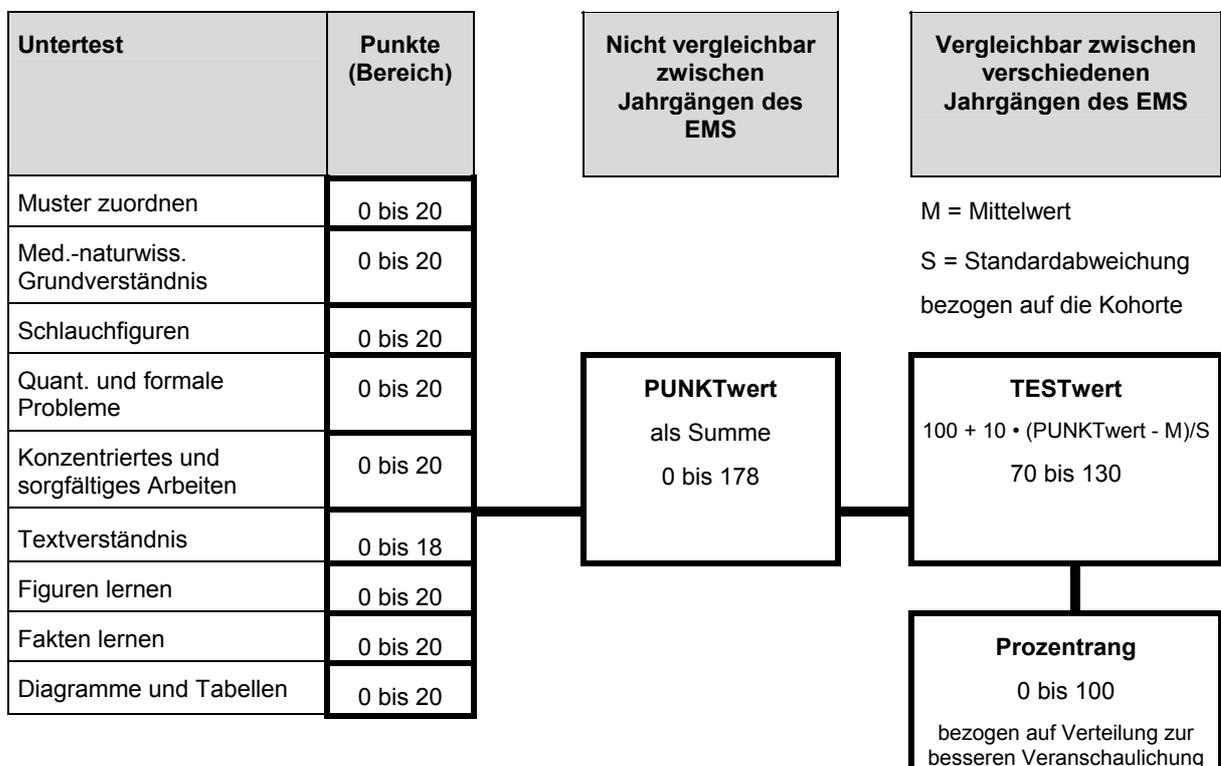


Abbildung 15: Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert

## 7 Evaluation: Eignungstest und Prüfungserfolg

Wenn auf politischer Ebene der Entscheid eines Numerus Clausus getroffen wird, muss ein **Zulassungskriterium** den Zugang zum Studium regeln. Ablehnungen müssen, wenn sie unvermeidbar sind, ausreichend begründet sein. Ein solches Zulassungskriterium muss Chancengleichheit gewährleisten, objektiv, fair und wissenschaftlich fundiert sein.

„**Eignung**“, welche entsprechend der gesetzlichen Grundlagen der Kantone als Zulassungskriterium dienen soll, muss in objektive, messbare und überprüfbare Erfolgskriterien umgesetzt werden. Eine Person ist dann geeigneter, wenn sie diese Erfolgskriterien besser erfüllt. Wenn sich der Numerus Clausus als **bildungspolitische** Massnahme begreift (und nicht als berufspolitisches Planungsinstrument den Arbeitsmarkt der Medizin regeln soll), muss die **Studieneignung** die Zulassung leiten.

Es bieten sich zwei **Erfolgskriterien** für Studieneignung an: Studienplätze bevorzugt an diejenigen zu vergeben, die

- (1) in möglichst kurzer bzw. **optimaler Zeit** das Studium absolvieren, sowie
- (2) das Studium in **guter Qualität** bzw. mit auf hohem Niveau erworbenen Fähigkeiten und Fertigkeiten absolvieren.

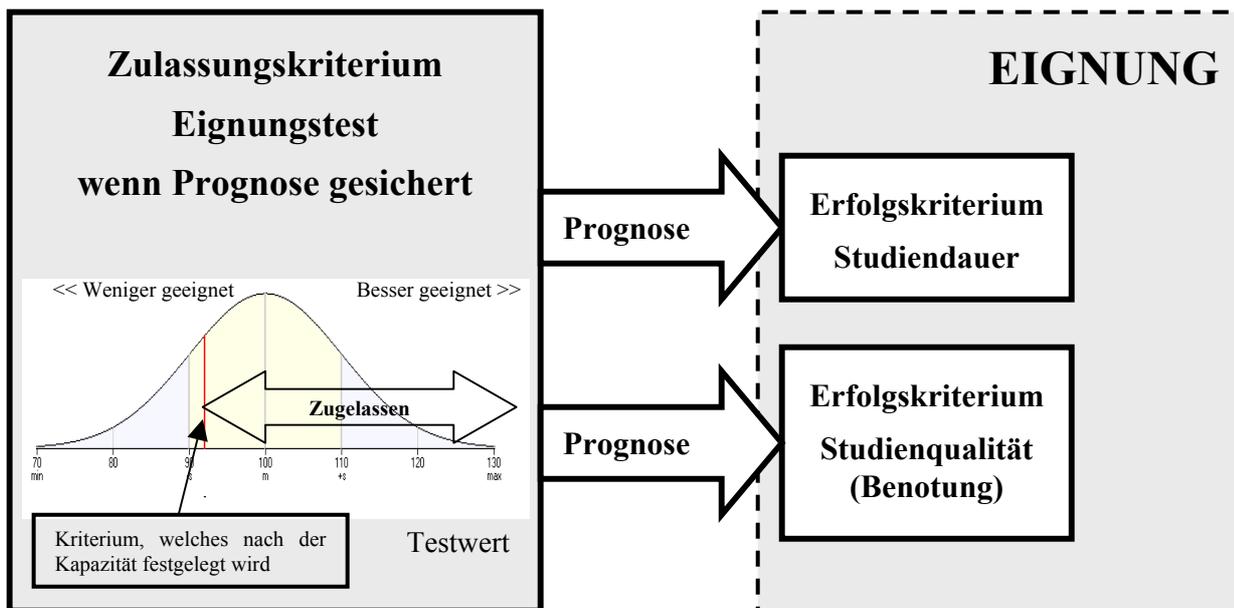


Abbildung 16: Zusammenhang Eignung – Erfolgskriterium – Zulassungskriterium

In der Abbildung 16 wird der Zusammenhang noch einmal verdeutlicht: Eignung wird anhand zweier Erfolgskriterien definiert (Studiendauer und -qualität). Indem nachgewiesen wird, dass das Zulassungskriterium diese beiden Erfolgskriterien vorhersagt (also Eignung tatsächlich misst), können Geeignere tatsächlich bevorzugt zugelassen werden.

Durch Vermeidung einer verlängerten Studiendauer oder gar von Studienabbrüchen werden die begrenzt vorhandenen Ausbildungsressourcen nicht mehr als notwendig beansprucht und somit früher wieder für Nachfolgende frei. Im Medizinstudium ergeben sich Begrenzungen beispielsweise durch eine nicht beliebig erweiterbare Patientenzahl an den Universitätsspitalern und vergleichsweise teuer ausgestattete Übungs- und Praktikumsplätze. Man muss aber nicht nur diese „volkswirtschaftliche“ Perspektive sehen,

denn auch für die Studierenden besteht der Nutzen darin, dass weniger verlorene Lebenszeit auf dem Weg zum Beruf entsteht, wenn weniger Abbrüche nach einem oder sogar zwei Jahren auftreten (vor dem Numerus Clausus brechen etwa 1/3 der Personen ab!).

**Gesucht ist daher ein Zulassungskriterium, welches Studiendauer und Studienqualität möglichst vor Antritt des Studiums optimal vorhersagbar macht.**

Intrauniversitäre Selektionen sind in der Regel teuer und es bleiben in der Anfangszeit weiterhin Ausbildungs-Engpässe erhalten. Nachteilige Wirkungen auf die Persönlichkeitsentwicklung können erwartet werden, weil man mit den Mitstudierenden in dieser Zeit konkurrenziert, statt das für den späteren Beruf notwendige Kooperationsverhalten zu erlernen.

Der Eignungstest EMS soll diese Studieneignung vorhersagen. Wenn er im Vergleich zu anderen Kriterien ausserdem konkurrenzlos geringe Kosten, hohe Chancengleichheit und Fairness gewährleistet, ist seine Wahl gut begründet.

Eine wichtige Frage der Evaluation des EMS als Zulassungskriterium ist daher die Analyse der Beziehungen

- (1) zwischen Testleistung und dem Bestehen der Prüfungen mit direktem Einfluss auf die Studiendauer und
- (2) zwischen Testleistung und den Prüfungsnoten als Ausdruck der Studienqualität.

Bereits im Jahre 2000 wurde eine erste Evaluation der Testkohorte 1998 vorgelegt (Hänsgen, Spicher 2000), wo nach einem Jahr Studiendauer gute Vorhersagewerte ermittelt werden konnten.

Für die Analyse des Zusammenhanges von Testergebnis und Prüfungserfolg stehen in diesem Jahr die Kohorten der Testteilnahme 1998 und 1999 zur Verfügung. 1998 war der Numerus Clausus für Humanmedizin, 1999 für Human- und Veterinärmedizin an den Universitäten Basel, Bern, Freiburg und Zürich notwendig. In der Tabelle 14 wird dargestellt, wie viele Testteilnehmer jeweils ein Studium begonnen haben. Die Analyse von Prüfungsnoten wird allerdings dadurch erschwert, dass zwischen den Universitäten und auch innerhalb der Jahrgänge verschiedene Prüfungsarten und Notensysteme zur Anwendung kommen. Im Rahmen der Studienreform wird teilweise noch experimentiert und Noten werden durch Kreditpunkte abgelöst, die nur noch zwischen Bestehen und Nichtbestehen differenzieren.

		Studienbeginn Medizin			Bisher kein Studienbeginn Medizin*	Testteilnahme gesamt
		1998	1999	Gesamt		
Test-jahr	1998	597	3	600	150	750
	1999		666	666	189	855
Total		597	669	1266	339	1605

Tabelle 14: Test 1998 und 1999 und Studienbeginn; \*: Ablehnungen und Rückzüge

Die 1998 Immatrikulierten haben bisher vier mögliche Prüfungstermine wahrnehmen können (Sommer und Herbst 1999, Sommer und Herbst 2000). Wurde die erste Vorprüfung 1999 bestanden, kann im Jahre 2000 bereits die zweite Vorprüfung abgelegt worden sein.

Eine zweijährige Studiendauer bis zur zweiten Vorprüfung wäre damit das mögliche Optimum. Die 1999 Immatrikulierten konnten bisher nur im Sommer oder Herbst 2000 die erste Vorprüfung ablegen.

Die Prüfungsdaten des Jahres 2000 wurden wiederum im Bundesamt für Gesundheit (BAG) erfasst<sup>1</sup>. Zur Quantifizierung der Studiendauer nummerieren wir die möglichen Prüfungstermine für jeden Studierenden vom Studienbeginn an fortlaufend. Erste und zweite Prüfungssession liegen damit im Sommer bzw. Herbst nach dem ersten Studienjahr, dritte und vierte Prüfungssession im Sommer bzw. Herbst nach dem zweiten Studienjahr.

## 7.1 Prüfungsverlauf für Testkohorten 1998 und 1999

Die Studienzulassung erfolgt praktisch in mehreren Schritten. Nicht ausgenutzte Plätze werden in zwei Runden an Personen vergeben, die zunächst aufgrund des Testwertes nicht zugelassen werden können. Testteilnehmende, die sich nicht immatrikuliert haben, werden danach unterschieden, ob der Testwert unter- oder oberhalb des Testwertes liegt, welcher der genauen Ausnutzung der Kapazität entspricht. Von Interesse sind dabei nur die Ergebnisse der Personen, die sich nicht immatrikuliert haben, vom Testwert her aber klar über dem Kapazitäts-Kriterium liegen. Sind dies gehäuft Geeignete, weniger Geeignete oder ist dies bezüglich der Eignung gleichmässig verteilt? Nachfolgend werden einige Statistiken über den Studienverlauf dargestellt, die jetzt und vor allem zukünftig Aufschluss über veränderte Studienverläufe unter den Bedingungen eines NC geben.

### 7.1.1 Erste Vorprüfung und Prüfungsverlauf

Abbildung 17 verdeutlicht, dass die Verhältnisse der beiden Testkohorten nach jeweils einem Studienjahr vergleichbar sind. Innerhalb des NC-Zeitraumes treten bisher keine Veränderungen auf. Für die Zeit vor dem NC fehlen entsprechende Vergleichsdaten.

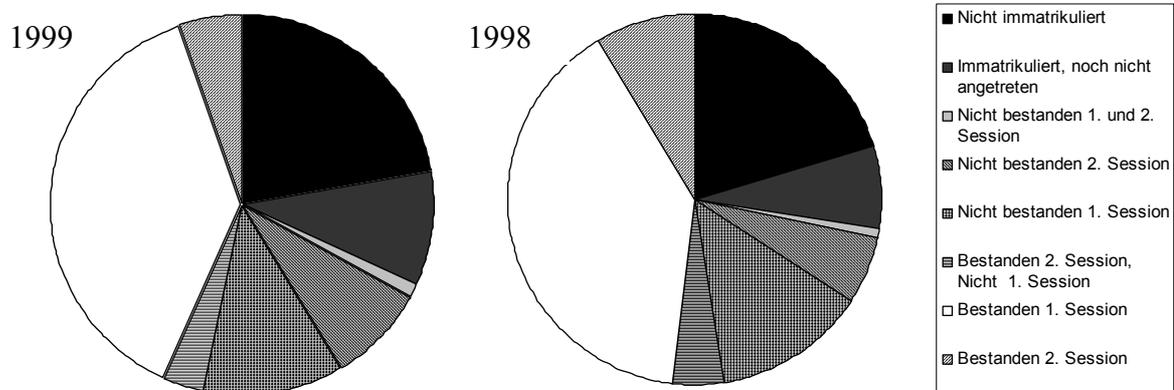
Es wäre damit zu rechnen, dass sich die Zahl der bestandenen Prüfungen in der ersten und zweiten Session langfristig erhöht. Dabei ist auch der vergleichsweise milde Numerus Clausus in der Schweiz zu beachten. Eine anfängliche Ablehnungsrate von 20% ist im internationalen Vergleich bereits recht niedrig. Rückzüge und die nachfolgende erneute Verteilung der Plätze an Personen mit niedrigeren Testwerten führen dazu, dass auch weniger Geeignete zum Studium zugelassen werden – dann zumeist nicht an ihrer Wunschuniversität. Mit anderen Worten: der NC führt nicht unbedingt und überall zu einer veränderten Zusammensetzung der Studierenden – seine Wirkung besteht vor allem in der kapazitätsentsprechenden Auslastung der Universitäten und dem Abbau der Überlastung.

Diese Überlastungen können im übrigen auch nicht sofort abgebaut werden – Repetenten aus den Vorjahren beanspruchen Kapazitäten noch 1 bis 2 Jahre länger, wenn die hier festgestellten Prüfungsverläufe auch für frühere Zeiten gelten. Hoffnung auf Erhöhung der

---

<sup>1</sup> Wir danken Herrn Megert, Frau Hirvi sowie Herrn Neuhaus vom BAG herzlich für die Zusammenarbeit. Das BAG erhielt die Immatrikulationsnummern der Studienanfänger von der Rektorenkonferenz der Schweizer Universitäten (CRUS). Prüfungserfolgsdaten wurden in elektronischer Form zur Verfügung gestellt, Noten von den Prüfungsbogen erfasst. Dabei waren EMS-Ergebnisse nicht bekannt. Erst durch die CRUS wurden die Prüfungsergebnisse mit den Testergebnissen verbunden – ein Original befindet sich zu Nachweiszwecken dort. Die Evaluationsberechnungen wurden dann am ZTD mit einer Kopie dieser Datenbank vorgenommen, was eine ausreichende Objektivität der Evaluation gewährleisten sollte.

Bestehensquoten besteht allerdings auch, weil die kapazitätsentsprechende Auslastung der Universitäten nunmehr auch Reformstudiengänge erlaubt, die zu einer neuen Studienqualität beitragen können.



**Abbildung 17:** Vergleich der Testkohorten 1998 und 1999 nach jeweils einem Studienjahr

Die Tabelle 15 stellt das Prüfungsverhalten beider Testkohorten für die erste Vorprüfung differenziert zusammen. Für die Kohorte 1998 wurde der Stand 1999 noch einmal mit dargestellt, um einen Vergleich zur Kohorte 1999 nach ebenfalls einem Jahr Studium zu ermöglichen. Die Ergebnisse rechtfertigen eine gemeinsame Analyse beider Kohorten für das erste Jahr.

Eine Zusammenfassung des Verlaufes der ersten Vorprüfung ist nach verschiedenen Kriterien möglich. Die Tabelle 16 zeigt, dass die Zahl der Prüfungsanläufe nach jeweils einem Jahr in etwa vergleichbar ist – in der Testkohorte 1998 nach 2 Jahren mehr Personen im zweiten Anlauf bestanden oder zweimal nicht bestanden haben – also ein Jahr länger für ihr Studium benötigt haben. Noch deutlicher wird dies in Tabelle 17, wo die immatrikulierten Personen danach zusammengefasst werden, wie lange das Studium minimal bis zur ersten Vorprüfung dauert. Bei rund 14 Prozent der Kohorte 1998 ist sicher, dass mehr als ein Jahr Studienverlängerung bis zum Bestehen der ersten Vorprüfung notwendig ist (siehe dazu auch die Abbildung 18).

Auch nach 2 Jahren haben „nur“ 63% der Personen die erste Vorprüfung erfolgreich abgelegt – 52% waren es schon nach einem Jahr. Zweimal nicht bestanden haben 6% nach 2 Jahren – etwa 1/3 der Zugelassenen verbleiben nach 2 Jahren als „offen“, was die erste Vorprüfung betrifft. Es ist möglich, dass einige davon nicht mehr immatrikuliert sind, einige aber noch später antreten werden.

Prüfungsverlauf	Test 1998				Test 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht immatrikuliert, TW unter Kriterium	107	14.3%	107	14.3%	88	10.3%
Nicht immatrikuliert, TW über Kriterium	43	5.7%	46	6.1%	101	11.8%
Immatrikuliert, noch nicht angetreten	33	4.4%	54	7.2%	83	9.7%
Nicht bestanden 3. und 4. Session	2	0.3%	Zeitlich nicht möglich			
Nicht bestanden 2. und 4. Session	2	0.3%				
Nicht bestanden 2. und 3. Session	13	1.7%				
Nicht bestanden 1. und 4. Session	5	0.7%				
Nicht bestanden 1. und 3. Session	18	2.4%				
Nicht bestanden 1. und 2. Session	6	0.8%	6	0.8%	11	1.3%
Nicht bestanden 4. Session	2	0.3%	Zeitlich nicht möglich			
Nicht bestanden 3. Session	7	0.9%				
Nicht bestanden 2. Session	14	1.9%	44	5.9%	71	8.3%
Nicht bestanden 1. Session	24	3.2%	100	13.3%	101	11.8%
Bestanden 2. Session, Nicht 1. Session	32	4.3%	32	4.3%	30	3.5%
Bestanden 3. Session, Nicht 1. Session	43	5.7%	Zeitlich nicht möglich			
Bestanden 4. Session, Nicht 1. Session	9	1.2%				
Bestanden 3. Session, Nicht 2. Session	15	2.0%				
Bestanden 4. Session, Nicht 2. Session	1	0.1%				
Bestanden 4. Session, Nicht 3. Session	1	0.1%				
Bestanden 1. Session	302	40.3%	297	39.6%	325	38.0%
Bestanden 2. Session	64	8.5%	64	8.5%	45	5.3%
Bestanden 3. Session	6	0.8%	Zeitlich nicht möglich			
Bestanden 4. Session	1	0.1%				
Total	750		750		855	

**Tabelle 15:** 1. Vorprüfung für Testkohorte 1998 (Stand 1999 und 2000) sowie Testkohorte 1999 (Stand 2000). Sessions wurden pro Studierenden vom Studienbeginn an durchnummeriert (siehe Text)

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht immatrikuliert, TW unter Kriterium	107	14.3%	107	14.3%	88	10.3%
Nicht immatrikuliert, TW über Kriterium	43	5.7%	46	6.1%	101	11.8%
Immatrikuliert, noch nicht angetreten	33	4.4%	54	7.2%	83	9.7%
Zweimal Nicht bestanden	46	6.1%	6	0.8%	11	1.3%
Einmal Nicht bestanden	47	6.3%	144	19.2%	172	20.1%
Zweiter Anlauf Bestanden	101	13.5%	32	4.3%	30	3.5%
Sofort Bestanden	373	49.7%	361	48.1%	370	43.3%
Total	750	100.0%	750	100.0%	855	100.0%

Tabelle 16: Zusammenfassung 1. Vorprüfung nach der Zahl der Prüfungsanläufe

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Studienverlängerung mehr als 1 Jahr	82	13.67%				
Studienverlängerung mindestens 1 Jahr	120	20.00%	204	34.17%	266	39.94%
„minimale“ Studienzeit	398	66.33%	393	65.83%	400	60.06%
Total	600		597		666	

Tabelle 17: Zusammenfassung 1. Vorprüfung nach dem Einfluss auf die Studienzeit

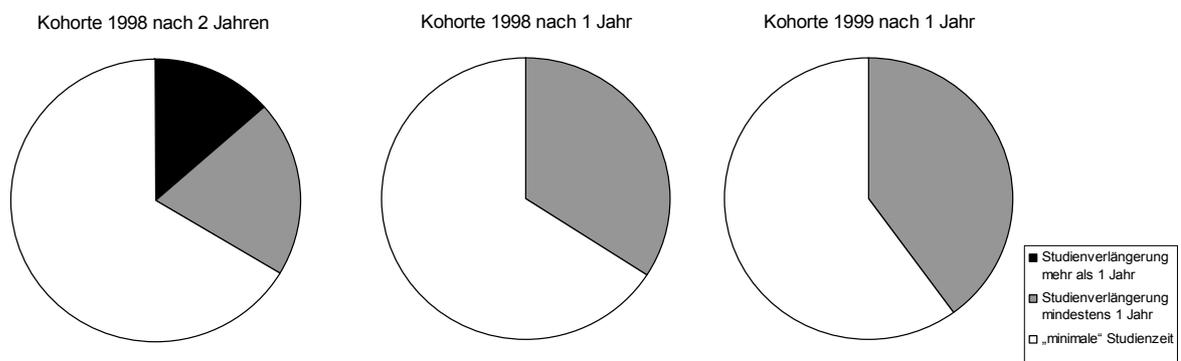


Abbildung 18: Vergleich der Test-Kohorten nach dem Einfluss auf die Studienzeit

In der Tabelle 18 werden die Personen nach dem Prüfungserfolg zusammengefasst. Wenn zwei Prüfungsanläufe unternommen wurden, zählt das jeweils letzte Resultat. Im Herbst 2000 haben geringfügig weniger Personen die erste Vorprüfung bestanden als im Herbst

1999. Zu beachten ist, dass in der zweiten Test-Kohorte Veterinärmedizin mit enthalten ist – 1998 bezog sich der NC nur auf Humanmedizin. Später wird gezeigt, dass die Bestehensquote in Veterinärmedizin geringer ist als in Humanmedizin (siehe Seite 61). Hier könnte sich allerdings auch andeuten, dass die erste Vorprüfung vom Charakter her immer noch als ein „versteckter“ Numerus Clausus aufgefasst wird und eine auf die Verteilung der Leistungen orientierte Benotungsstrategie vorherrscht. Durch eine Homogenisierung des Leistungsniveaus als Folge des Numerus Clausus – oder auch als Folge der verbesserten Studienbedingungen (weil die Kapazitäten nicht mehr so überschritten sind) – wird die Differenzierungsfähigkeit zwischen Personen geringer und das gewählte Bestehenskriterium für Prüfungen lässt bei strengen Massstäben mehr Personen die Prüfung nicht bestehen.

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht bestanden	93	16.4%	150	27.6%	183	31.4%
Bestanden	474	83.6%	393	72.4%	400	68.6%
Total	567	100.0%	543	100.0%	583	100.0%

Tabelle 18: Zusammenfassung 1. Vorprüfung nach dem Erfolg

### 7.1.2 Zweite Vorprüfung und Prüfungsverlauf

Die zweite Vorprüfung kann nur von Personen der Testkohorte 1998 abgelegt worden sein – entweder im Sommer oder Herbst 2000; die Tabelle 19 fasst den Prüfungsverlauf differenziert zusammen.

Prüfungsverlauf	Testkohorte 1998	
Nicht immatrikuliert, unter Zulassungskriterium	107	14.3%
Nicht immatrikuliert, über Zulassungskriterium	43	5.7%
Wegen Nichtbestehen 1. Prüfung nicht möglich	126	16.8%
Zeitlich unmöglich (1. Vorprüfung zu spät bestanden)	79	10.5%
Noch nicht angetreten (aber möglich)	49	6.5%
Einmal Nicht bestanden 4. Session	41	5.5%
Einmal Nicht bestanden 3. Session	11	1.5%
Bestanden 3. Session	110	14.7%
Bestanden 4. Session	182	24.3%
Bestanden 4. Session, Nicht 3. Session	2	0.3%
Total	<b>750</b>	

Tabelle 19: 2. Vorprüfung für Testkohorte 1998, Sessionen vom Beginn des Studiums an fortlaufend durchnummeriert

346 Personen sind zur 2. Vorprüfung angetreten – lediglich 15% haben sie nicht bestanden (Tabelle 20). Weitere 49 Personen hätten zu dieser Prüfung antreten können, haben dies aber nicht getan. Diese Prüfung hat damit nicht mehr den Charakter eines „versteckten“ NC. Bisher konnten nur Personen die 2. Vorprüfung ablegen, welche die erste Vorprüfung in minimal möglicher Zeit bestanden haben (die also ohnehin die geeignetsten wären). Es ist hier also noch kein repräsentativer Leistungsquerschnitt vorhanden.

Prüfungsverlauf	Testkohorte 1998	
Nicht bestanden	52	15.0%
Bestanden	294	85.0%
Total	346	100.0%

Tabelle 20: Antritte 2. Vorprüfung nach Bestehen

Der Zusammenhang zwischen erster und zweiter Vorprüfung wird in der Tabelle 21 dargestellt.

Testkohorte 1998	2. Vorprüfung							Total
	Zeitlich nicht möglich	Noch nicht angetreten	1 x Nicht bestanden 4. Session	1 x Nicht bestanden 3. Session	Bestanden 3. Session	Bestanden 4. Session	Bestanden 4. Session, Nicht bestanden 3. Session	
Bestanden 1. Session	3*	25	17	10	<u>105</u>	<u>40</u>	2	302
Bestanden 2. Session		21	10		<u>1</u>	<u>32</u>		64
Bestanden 3. - 4. Session	7							7
Bestanden 2. Session, Nicht bestanden 1. Session		3	14	1	<u>4</u>	<u>10</u>		32
Bestanden 3. - 4. Session, Nicht bestanden 1. - 3. Session	69							69
	79	49	41	11	<u>110</u>	<u>182</u>	2	474

Tabelle 21: 2. Vorprüfung für Personen der Testkohorte 1998, welche die 1. Vorprüfung bestanden haben. \*: Studienbeginn erst 1999. Unterstrichen: minimale Studiendauer

## 7.2 Prüfungserfolg und Testwert

Die Vorhersage des Prüfungserfolges aufgrund des Testwertes wird als wichtigstes Evaluationskriterium betrachtet. Dies hat vor allem Einfluss auf die Studiendauer. Bekanntlich ist der EMS ein Test, der auf die Vorhersage der Studierfähigkeit hin ausgerichtet wurde. Er ist kein Wissenstest – es wird vielmehr geprüft, inwieweit die Erarbeitung von Wissen (z.B. aus Texten), das Erkennen und die Anwendung logischer Regeln oder die Konzentrations- und Merkfähigkeit gelingt. Für die Lösung der Aufgaben notwendiges Wissen wird jeweils bei den einzelnen Aufgaben mit dargestellt.

Bei den Vorprüfungen handelt es sich dagegen um den tatsächlich im Studium erworbenen Wissensstand (Fähigkeiten und später auch Fertigkeiten). Der Test ist dann ein gutes Zulassungskriterium, wenn dieser Wissenserwerb von den geprüften kognitiven Fähigkeiten so stark abhängt, dass eine Verwendung als Zulassungskriterium begründet ist. Es ist klar, dass auch andere Komponenten den Prüfungserfolg mit bestimmen (Motivation, aufgewendete Zeitressourcen und soziale Situation, Lernbedingungen usw.).

### 7.2.1 Erste Vorprüfung und Prüfungserfolg gesamt

Die Testwerte der einzelnen Gruppen nach dem Prüfungsverlauf sind in Tabelle 22 und als Boxplots in der Abbildung 19 dargestellt. In beiden Jahrgängen ergeben sich vergleichbare Testleistungen für die Gruppen. Unterschiedliche Zeiträume (ein oder zwei Jahre bis zur Prüfung) haben dabei nur einen geringen Einfluss. Auch in der dritten und vierten Prüfungssession hängt der Prüfungserfolg vom Testwert ab.

Die Personengruppe „Prüfung sofort bestanden“ unterscheidet sich von allen anderen signifikant, deren mittlerer Testwert liegt eine halbe Standardabweichung über dem aller anderen Gruppen (multipler Mittelwertvergleich für alle Personen, Tabelle 23). Personen, die im zweiten Anlauf bestanden haben, noch nicht angetreten sind sowie sich nicht immatrikuliert haben (aber über dem Kapazitäts-Kriterium liegen) bilden die zweite Gruppe mit einem mittleren Testwert um 100. Personen, die nicht bestanden haben, unterscheiden sich wiederum deutlich von dieser Gruppe – der Mittelwert liegt deutlich unter 100.

	Test 1998 (nach 2 Jahren)			Test 1999 (nach 1 Jahr)			Gesamt		
	n	m	s	n	m	s	n	m	s
Nicht immatrikuliert, TW unter Kriterium	107	<b>83.91</b>	4.639	88	<b>83.18</b>	4.372	195	<b>83.58</b>	4.523
Nicht immatrikuliert, TW über Kriterium	43	<b>99.77</b>	6.982	101	<b>100.92</b>	6.431	144	<b>100.58</b>	6.597
Noch nicht angetreten	33	<b>99.61</b>	6.955	83	<b>98.27</b>	7.727	116	<b>98.65</b>	7.510
Zweimal Nicht bestanden	46	<b>97.17</b>	6.441	11	<b>95.91</b>	7.148	57	<b>96.93</b>	6.535
Einmal Nicht bestanden	47	<b>96.34</b>	5.446	172	<b>97.31</b>	6.855	219	<b>97.10</b>	6.578
Zweiter Anlauf Bestanden	101	<b>99.94</b>	5.407	30	<b>101.67</b>	7.862	131	<b>100.34</b>	6.067
Sofort Bestanden	373	<b>105.54</b>	7.721	370	<b>105.43</b>	8.560	743	<b>105.48</b>	8.144
Total	750	<b>100.02</b>	9.965	855	<b>100.02</b>	10.018	1605	<b>100.02</b>	9.990

Tabelle 22: Testwert Mittelwert (m) und Standardabweichung (s) für Gruppen nach der ersten Vorprüfung

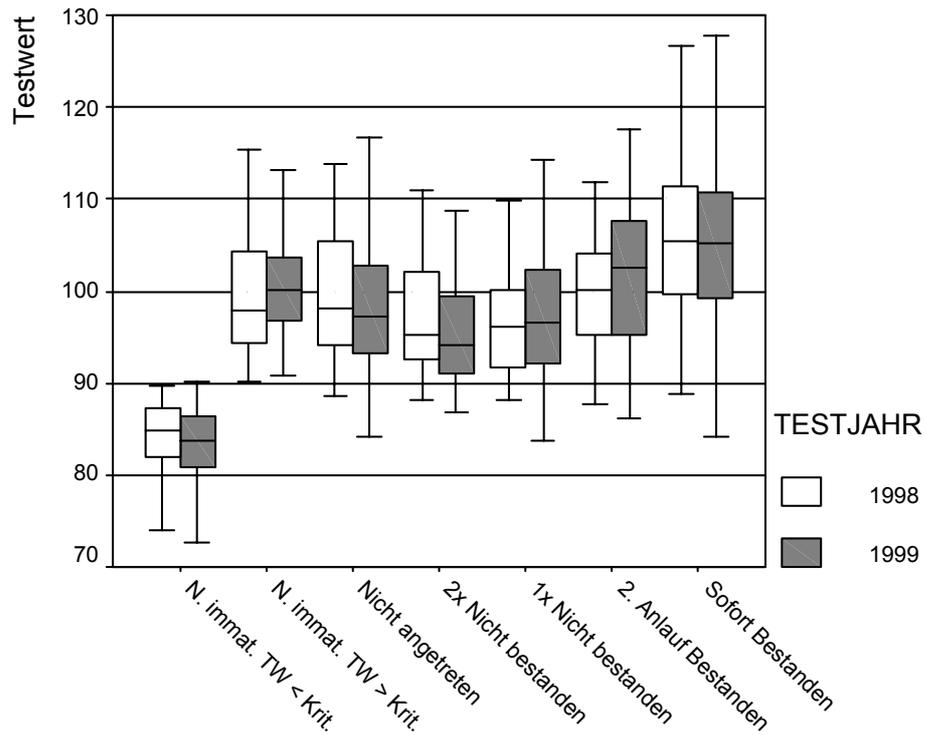


Abbildung 19: Boxplots für Testwerte (Median und 1. bis 4. Quartil) nach dem Prüfungsverhalten für die Kohorten 1998 und 1999

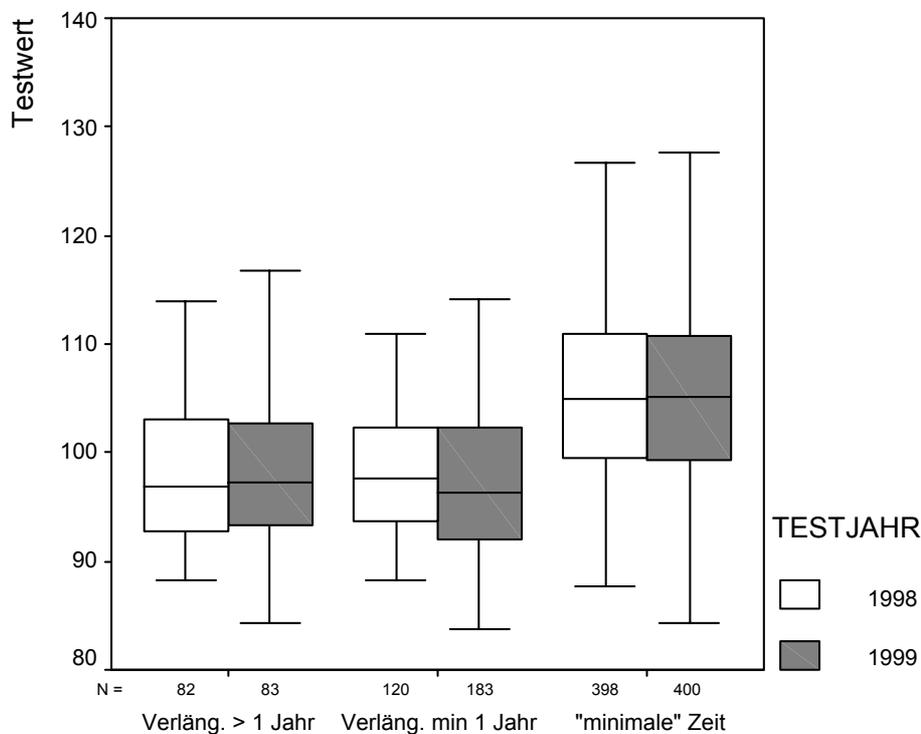
Erfolg 1. Vorprüfung nach Prüfungsanläufen	n	Subset for alpha = .05			
		1	2	3	4
Nicht immatrikuliert, TW unter Kriterium	195	83.607			
Zweimal Nicht bestanden	57		96.875		
Einmal Nicht bestanden	219		97.086		
Noch nicht angetreten	116		98.568	98.568	
Zweiter Anlauf Bestanden	131			100.261	
Nicht immatrikuliert, TW über Kriterium	144			100.547	
Sofort Bestanden	743				105.474
Signifikanz		1.000	.457	.263	1.000

Tabelle 23: Multipler Mittelwertvergleich des Testwertes (Tukey) für das Prüfungsverhalten beider Kohorten zusammengefasst

Gruppiert man die immatrikulierten Personen nach der sich ergebenden Studienverlängerung gegenüber dem optimalen Studienablauf (Tabelle 24 bzw. Abbildung 20), ergibt sich wiederum eine klare Trennung der Personengruppe mit einem Jahr Studiendauer bis zur ersten Vorprüfung von allen anderen (Mittelwertprüfung siehe Tabelle 25). Die beiden Gruppen mit notwendiger Verlängerung unterscheiden sich allerdings nicht voneinander.

Prüfungsverlauf	1998			1999			Total		
	n	m	s	n	m	s	n	m	s
Studienverlängerung mehr als 1 Jahr	82	<b>98.1</b>	6.8	83*	<b>98.2</b>	7.7	165*	<b>98.2</b>	7.2
Studienverlängerung mindestens 1 Jahr	120	<b>98.4</b>	5.6	183	<b>97.2</b>	6.8	303	<b>97.7</b>	6.4
"minimale" Studienzeit	398	<b>105.2</b>	7.8	400	<b>105.1</b>	8.5	798	<b>105.2</b>	8.2
	600	<b>102.9</b>	8.0	666	<b>102.1</b>	8.8	1266	<b>102.5</b>	8.4

**Tabelle 24:** Mittelwert (m) und Streuung (s) des Testwertes für die Studiendauer,  
\*: Personen, die nach einem Jahr noch nie zur Prüfung angetreten sind



**Abbildung 20:** Boxplots für die Zusammenfassung der Personen nach der Studiendauer für beide Kohorten

Prüfungsverlauf	n	Subset alpha = .05	
		1	2
Studienverlängerung mindestens 1 Jahr	303	97.6	
Studienverlängerung mehr als 1 Jahr	165	98.1	
"minimale" Studienzeit	798		105.1
Signifikanz		.728	1.000

**Tabelle 25:** Multipler Mittelwertvergleich des Testwertes nach Studiendauer für beide Kohorten

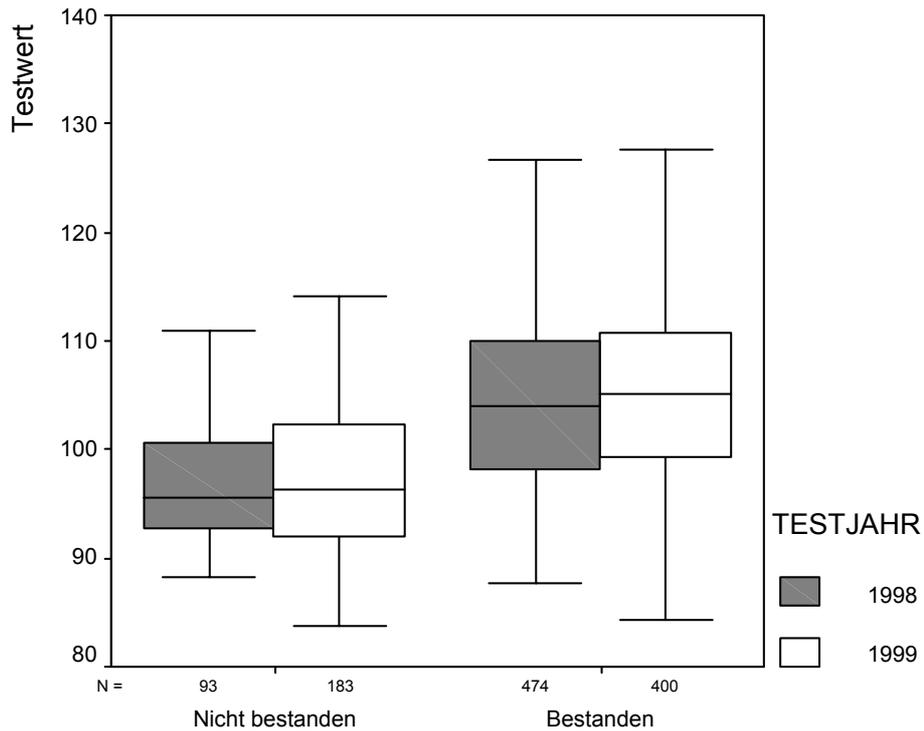
Die Testwerte derjenigen Personen, die zur ersten Vorprüfung angetreten sind, wurden nach dem Bestehen dieser Prüfung unterschieden und in der Tabelle 26 dargestellt. In beiden Jahren ergeben sich etwa gleiche Unterschiede (siehe auch Boxplot in Abbildung 21) und die Unterschiede sind statistisch signifikant (siehe Tabelle 27).

Testjahr		n	m	s
1998	Nicht bestanden	93	<b>96.75</b>	5.941
	Bestanden	474	<b>104.35</b>	7.638
1999	Nicht bestanden	183	<b>97.22</b>	6.860
	Bestanden	400	<b>105.15</b>	8.558

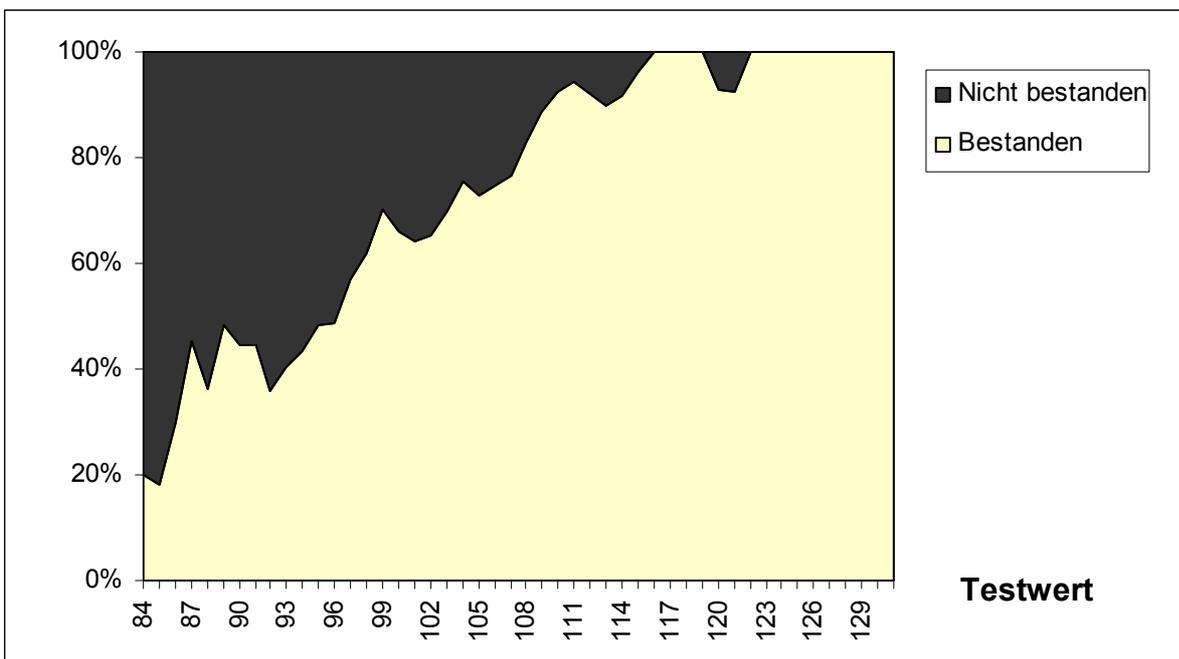
**Tabelle 26:** Mittelwert (m) und Streuung (s) des Testwertes für den Studienerfolg

Test-jahr	Varianzhomogenität	Levene's Test Varianzhomogenität		t-test Mittelwertgleichheit		
		F	Sig.	t	df	Sig. (2s)
1998	Varianzen inhomogen	9.573	.002	-10.714	158.107	.000
1999	Varianzen inhomogen	9.731	.002	-11.937	433.237	.000

**Tabelle 27:** Mittelwertvergleich (t-Test) des Testwertes nach dem Bestehen der ersten Vorprüfung



**Abbildung 21:** Boxplots des Testwertes für beide Kohorten nach dem Bestehen der ersten Vorprüfung



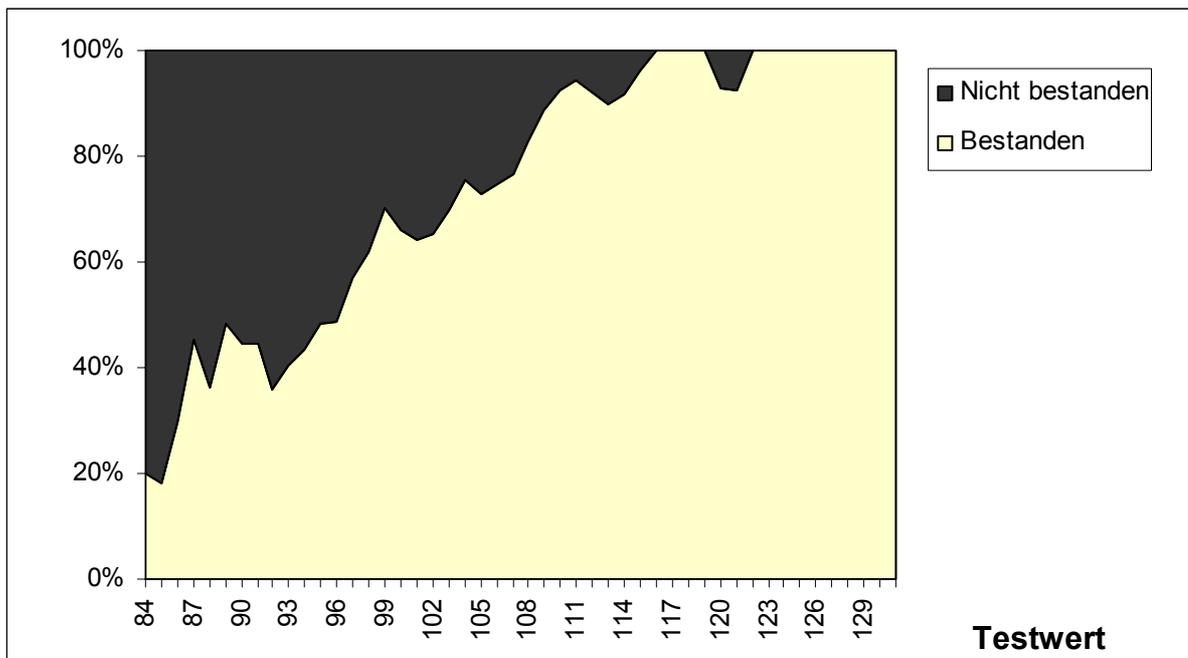
**Abbildung 22:** Wahrscheinlichkeit des Bestehens der 1. Vorprüfung pro Testwert für Kohorte 1998 nach 2 Jahren

Die Bestehenswahrscheinlichkeiten der Prüfung pro individuellem Testwert wurden berechnet und in Abbildung 22 (Kohorte 1998 nach 2 Jahren) und Abbildung 23 (Kohorte 1999 nach einem Jahr) dargestellt. Wiederum ergeben sich im unteren Bereich der Testwerte lineare Zusammenhänge zwischen Testwert und Bestehenswahrscheinlichkeit. In der Kohorte 1998 bleibt dieser Zusammenhang auch nach zwei Jahren Analysezeitraum deutlich

(Bestimmtheitsmass 0.83). Für die Kohorte 1999 nach einem Jahr Prüfungszeitraum ist das Bestimmtheitsmass mit 0.94 höher. Dies bedeutet, dass tatsächlich mit sinkendem Testwert die Wahrscheinlichkeit monoton abnimmt, eine Prüfung zu bestehen. Dies bedeutet allerdings auch, dass unterhalb eines bestimmten Testwertes die Wahrscheinlichkeit, eine Prüfung zu bestehen, recht gering wird. Durch die erneute Zuweisung von durch Rückzügen freiwerdender Studienplätze an diesen Personenkreis wird die Gesamt-Erfolgsrate des Jahrganges dann nachweislich verringert.

	Quadratsumme	Df	MQ	F	Sig.
Regression	.869	1	.869	122.630	.000
Residual	.170	24	.007		
Total	1.039	25			

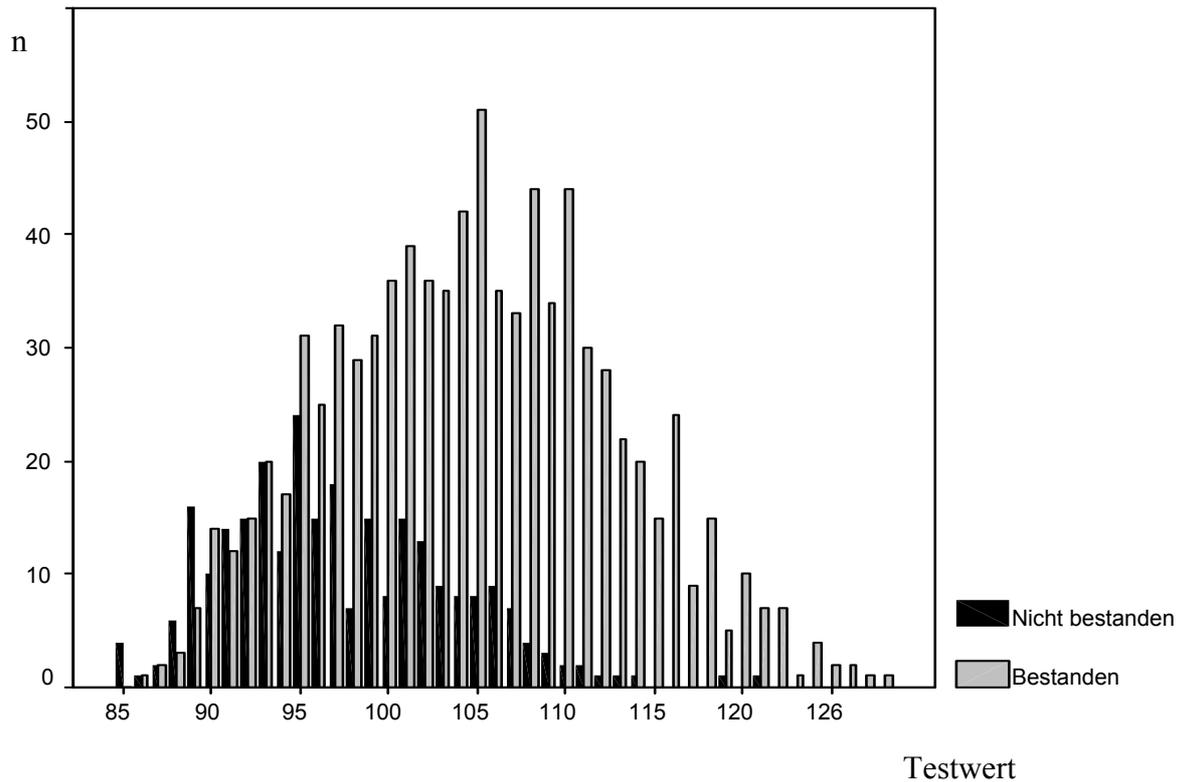
**Tabelle 28:** Varianzanalyse der Regression: linearer Teil zwischen Testwert 88 und 113 Bestimmtheitsmass (adjustiert):  $r^2 = 0.83$



**Abbildung 23:** Wahrscheinlichkeit des Bestehens der 1. Vorprüfung pro Testwert für Kohorte 1999 nach 1 Jahr

	Quadratsumme	Df	MQ	F	Sig.
Regression	1.650	1	1.650	569.948	.000
Residual	.090	31	.003		
Total	1.740	32			

**Tabelle 29:** Varianzanalyse der Regression linearer Teil zwischen Testwert 84 und 116: Bestimmtheitsmass (adjustiert):  $r^2 = 0.94$



**Abbildung 24:** Häufigkeitsverteilung für den Testwert nach dem Bestehen der ersten Vorprüfung

In der Abbildung 24 sind die Häufigkeitsverteilungen nach dem Bestehen der Prüfung dargestellt. Beide Verteilungen sind deutlich unterschieden. Mittels Diskriminanzanalyse wurde die Trennbarkeit der beiden Gruppen geprüft. Bei 77.5% der Personen lässt sich allein aufgrund des Testwertes korrekt vorhersagen, ob die Prüfung bestanden wird oder nicht. In der Tabelle 31 werden die Ergebnisse einer Diskriminanzanalyse vorgestellt, die auf gewichteten Koeffizienten für die einzelnen Untertests beruht. Es wäre denkbar, dass eine differenziertere Verrechnung der Untertestwerte eine bessere Trennung erzielt. Es zeigt sich aber, dass dann nur 79.1% richtig klassifiziert werden – der Gewinn gegenüber der Verwendung des Testwertes als Summe der Untertestwerte von 1.6% zu vernachlässigen ist. Der Testwert ist in der aktuell verwendeten Form damit ein für die Erfolgsprognose valider Koeffizient. Die Gewichte der einzelnen Untertests entsprechen dabei in etwa der Reihenfolge der Korrelation mit den Noten, wobei die sehr studienanforderungsnahen Anforderungen ebenfalls besser abschneiden als Konzentration, räumliches Vorstellen oder Merkfähigkeit.

Tatsächlicher Prüfungserfolg	Vorhergesagt aufgrund des Testwertes		Total
	Nicht bestanden	Bestanden	
Nicht bestanden	<b>68</b>	208	276
Bestanden	50	<b>822</b>	872

**Tabelle 30:** Optimale Klassifikation nach dem Bestehen der ersten Vorprüfung aufgrund des Testwertes: 77.5% der Personen werden korrekt klassifiziert

Untertests	Standardisierte Koeffizienten der Diskriminanzfunktion	Strukturmatrix
Quantitative und formale Probleme	.415	.695
Diagramme und Tabellen	.255	.665
Med.-naturw. Grundverständnis	.187	.644
Textverständnis	.181	.612
Muster zuordnen	.245	.489
Konzentr. und sorgf. Arbeiten	.354	.426
Schlauchfiguren	-.069	.382
Figuren lernen	.066	.361
Fakten lernen	.122	.352

Tabelle 31: Optimale Klassifikation nach dem Bestehen der ersten Vorprüfung aufgrund einer Diskriminanzfunktion der einzelnen Untertestwerte: 79.1% werden korrekt klassifiziert

## 7.2.2 Effekt der Zulassung von der Warteliste

Die durch Rückzüge freiwerdenden Plätze werden bekanntlich an Personen vergeben, deren Testwert unterhalb des kapazitätsentsprechenden Testwerts (nachfolgend als Kriterium bezeichnet) liegt, die zunächst also eine Warteliste bilden. Dieses Kriterium wird nur nach der Kapazität festgelegt – es trennt keinesfalls Geeignete von Ungeeigneten. Stünden weniger Plätze zur Verfügung, würde das Kriterium höher festgelegt. Bei mehr Plätzen wäre es entsprechend niedriger. Die nachfolgende Statistik bezieht beide Kohorten ein.

Erfolg 1. Vorprüfung				
		Nicht bestanden	Bestanden	
Testwert unter Kriterium	Anzahl	38	25	63
	Prozentverhältnis	<b>60.3%</b>	<b>39.7%</b>	
Testwert über Kriterium	Anzahl	238	849	1087
	Prozentverhältnis	<b>21.9%</b>	<b>78.1%</b>	
Erfolg 2. Vorprüfung				
		Nicht bestanden	Bestanden	
Testwert unter Kriterium	Anzahl	<b>1</b>	<b>1</b>	<b>2 (!)</b>
	Prozentverhältnis	50.0%	50.0%	
Testwert über Kriterium	Anzahl	<b>51</b>	<b>293</b>	<b>344</b>
	Prozentverhältnis	14.8%	85.2%	

Tabelle 32: Erfolg in beiden Vorprüfungen und kapazitätsentsprechender Testwert (Kriterium), Erläuterungen siehe Text

Aus Tabelle 32 wird ersichtlich, dass 63 Personen zur ersten Vorprüfung angetreten sind, die das Kriterium eigentlich nicht erreicht haben und über die Warteliste zugelassen worden sind. Deren Erfolgsaussicht zum Bestehen dieser Prüfung beträgt nur 40% - im Vergleich zu 78% für Personen mit einem Testwert über dem Kriterium. Die zweite Vorprüfung erreichten bisher sogar nur zwei Personen mit einem Testwert unterhalb des Kriteriums. In der Tabelle 33 sind Mittelwerte und Standardabweichungen des Testwertes für die einzelnen Gruppen angegeben. Personengruppen, deren Testwert unter dem Kriterium liegt – die aber dennoch zugelassen wurden, weisen deshalb eine sehr geringe Variation der Testwerte auf. Die Erfolgreichen unterscheidet von den Nicht Erfolgreichen auch hier eine halbe Standardabweichung – der wesentlichere Unterschied ist allerdings die schon erwähnte Halbierung der Erfolgswahrscheinlichkeit.

Diese Befunde zeigen das Risiko auf, welches eine Kapazitätserhöhung bei gleichbleibendem Bewerbungsverhalten birgt: die Studiendauer wird länger und die Erfolgsaussichten werden geringer. Vermutlich wird auch die Abbrecherzahl steigen. Mehr Aufmerksamkeit sollte den Fragen gewidmet werden, wie sich mehr Geeignete für ein Medizinstudium bewerben, ob die Dissuasionsfaktoren möglicherweise die „Falschen“ abschrecken und selektiv wirken. Mit dem Eignungstest steht ein Instrument zur kapazitätsentsprechenden Zulassung zur Verfügung. Die Zugelassenen haben gute Chancen, das Studium mit Erfolg abzuschließen. Unterstellt man, dass steigende Maturandenzahlen eigentlich auch ein steigendes Interesse für das Medizinstudium hervorbringen müssten, bleiben nur Dissuasionsfaktoren als eine Ursache für das hohe, aber gleichbleibende Interesse für das Fach Medizin. Ein Teil davon wird in beruflichen Faktoren liegen. Dissuasion wurde in der Zeit vor Einführung des NC sogar gefördert und auch der Test wirkte anfangs dissuasiv. Vielleicht wäre jetzt die Zeit für eine Überprüfung dieses Herangehens. Wenn mehr Personen den Tests als Chance sehen würden, ihre Eignung festzustellen (und ihre Neigung zu diesem Studium prüfen), würden mehr Geeignete das Studium beginnen.

			n	m	s
Testwert unter Kriterium	Vergleichswert alle Personen		274	<b>84.9</b>	4.5
	Erste Vorprüfung	Nicht angetreten	16	<b>88.3</b>	1.7
		Nicht bestanden	38	<b>87.8</b>	2.0
		Bestanden	25	<b>88.8</b>	1.5
Testwert über Kriterium	Erste Vorprüfung	Nicht bestanden	238	<b>98.5</b>	5.8
		Bestanden	849	<b>105.2</b>	7.7

**Tabelle 33:** Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Gruppen nach dem Erfolg in der ersten Vorprüfung und dem Erreichen des Kriteriums

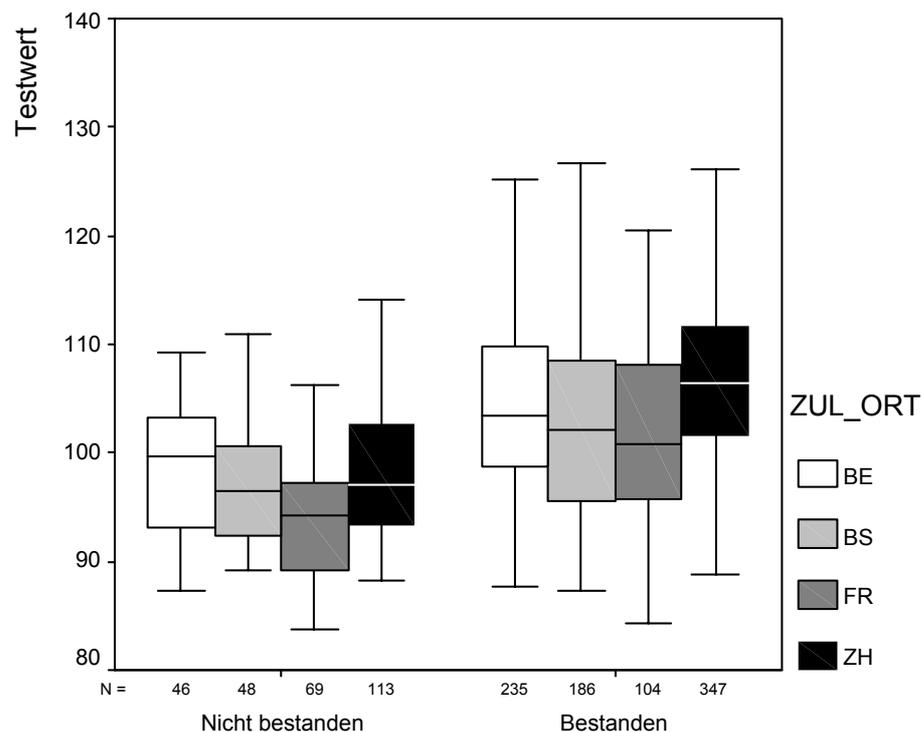
### 7.2.3 Erfolgsvorhersage für verschiedene Universitäten

Die Vorhersagekraft des Testwertes für das Bestehen der Prüfung ist nicht spezifisch für einzelne Universitäten. Aus der Tabelle 34 wird ersichtlich, dass in allen Universitäten der Unterschied mindestens 5 Testwert-Punkte zwischen Bestehen und Nichtbestehen beträgt. Die Testwert-Niveaus zwischen den Universitäten unterscheiden sich weniger stark als

zwischen den beiden Gruppen nach dem Prüfungsbestehen (siehe Abbildung 25). Unterschiede finden sich allerdings in den Bestehensquoten – die liegen aber offenbar nicht (nur) in unterschiedlicher „Strenge“ bei den Prüfungen begründet, sondern auch in Leistungsunterschieden der Personen (Unterschiedliche „Bewerberprofile“ der Universitäten und Folgen der Umleitungen sind zu beachten).

Universität	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
Bern	46 (16%)	<b>98.0</b>	6.2	235 (84%)	<b>104.2</b>	7.8
Basel	48 (21%)	<b>97.1</b>	5.4	186 (79%)	<b>102.6</b>	8.2
Freiburg	69 (40%)	<b>93.8</b>	6.2	104 (60%)	<b>102.3</b>	9.0
Zürich	113 (25%)	<b>98.6</b>	6.7	347 (75%)	<b>106.8</b>	7.4

**Tabelle 34:** Mittelwerte (m) und Standardabweichungen (s) des Testwertes für die Universitäten – beide Kohorten zusammengefasst



**Abbildung 25:** Boxplots des Testwertes für die verschiedenen Universitäten nach dem Bestehen der Prüfung

Die besondere Situation von Freiburg wird ebenfalls sichtbar: Bekanntlich ist die Rate der umgeleiteten Studierenden dort am höchsten. Würde man dort die Kapazitäten weiter erhöhen, führte dies vermutlich zur Zulassung von noch mehr Personen, die geringere Aussichten auf das Bestehen der Prüfung haben.

## 7.2.4 Erfolgsvorhersage für Geschlechter

Auch bezüglich der beiden Geschlechter findet sich der gleiche deutliche Unterschied des Testwertes für die beiden Gruppen nach dem Bestehen der Prüfung (Tabelle 35 und Abbildung 26). Bekanntlich waren 1998 und 1999 keine signifikanten Unterschiede der Testwerte für beide Geschlechter nachweisbar, wobei tendenziell die Testwerte der Männer immer ca. 1.5 Punkte besser waren. Für die beiden Gruppen, welche die Prüfung bestanden haben, sind die Testwerte hier nun nahezu identisch – die Gruppen, welche die Prüfung nicht bestanden haben, unterscheiden sich nur um 0.4 Punkte.

Dies ist ein Beleg, dass mögliche Testwertunterschiede zwischen den Geschlechtern keinesfalls als Chancenunterschiede zu interpretieren wären, sondern sehr wahrscheinlich die unterschiedliche Repräsentativität mit einem Überwiegen weiblicher Studierender dafür verantwortlich sein wird. Die testwertabhängigen „Chancen“, die erste Vorprüfung zu bestehen, sind praktisch bei gleicher Eignung identisch. Dieser Befund entkräftet jegliche Befürchtungen, dass es zu Ungleichbehandlungen bezüglich der Geschlechter kommt.

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

Tabelle 35: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für die Geschlechter – beide Kohorten zusammengefasst

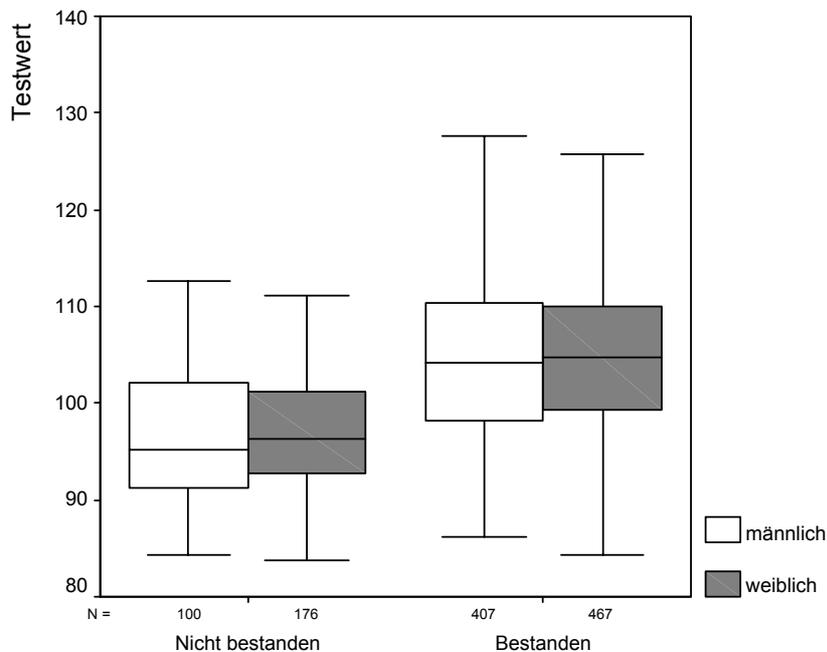


Abbildung 26: Boxplots des Testwertes für die Geschlechter nach dem Bestehen der Prüfung

### 7.2.5 Erfolgsvorhersage für Alter und Maturitätsjahr

In allen Testjahren ergab sich bisher der Befund, dass die Testwerte von Personen, die älter sind und die Maturitätsprüfung spät abgelegt haben, geringer sind als von Gleichaltrigen, welche die Maturität zum frühestmöglichen Zeitpunkt abgelegt haben. Nachfolgend werden zwei Gruppen „älterer Personen“ (für Geburtsjahre vor 1975) gebildet, die nach dem Ablegen der Maturität nach 1995 („späte Maturität“) oder vor 1996 („frühe Maturität“) unterschieden werden. Diese werden mit den „jüngeren Personen“ verglichen.

In allen drei Gruppen treten wiederum deutliche Testwertunterschiede hinsichtlich des Bestehens der Prüfung auf. Bei der älteren Gruppe mit später Maturitätsprüfung (die tendenziell geringere Testwerte haben) fällt die hohe Bestehensrate von 84% ins Auge – die Testwerte dieser Personen liegen dagegen mit 101.1 im Mittel etwas geringer als für die anderen beiden Gruppen.

Alter und Maturitätsabschluss	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
geb. vor 1975, Matur vor 1996	15 (44%)	<b>96.5</b>	4.7	19 (56%)	<b>105.7</b>	8.1
geb. vor 1975, Matur nach 1995	7 (16%)	<b>94.9</b>	5.1	37 (84%)	<b>101.1</b>	6.0
geb. nach 1974	254 (24%)	<b>97.1</b>	6.7	816 (76%)	<b>104.8</b>	8.1

Tabelle 36: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Gruppen nach Alter und Maturitätsjahr – beide Kohorten zusammengefasst

Die Testwerte unterscheiden in allen drei Gruppen Erfolgreiche von weniger Erfolgreichen deutlich (mindestens  $\frac{1}{2}$  Standardabweichung). In der älteren Gruppe mit später Maturität können mit anderen Gruppen vergleichbare Erfolge allerdings auch mit niedrigeren Testwerten erreicht werden. Besonders im Boxplot (Abbildung 27) wird deutlich, dass für die ältere Gruppe mit später Maturität die Verteilungen beider Teilgruppen nach dem Prüfungsbestehen in Richtung niedriger Testwerte verschoben sind. Dies könnte ein Beleg dafür sein, dass die grössere „Lebenserfahrung“ dann für einen Bonus sorgt, wenn die „Lernerfahrung“ durch eine unlängst abgelegte Maturitätsprüfung aufgefrischt worden ist.

Die ältere Gruppe mit früherer Maturität hat bezogen auf die Testwerte identische Chancen wie die jüngere Gruppe, eine Prüfung zu bestehen. Diese Gruppe ist im biografischen Ablauf auch bis zum Ablegen der Maturität mit der „jüngeren Gruppe“ identisch – es wird nach der Maturität jedoch kein oder ein anderes Studium aufgenommen. Die Gruppe liegt bezüglich des Testwertes in allen bisher analysierten Jahren immer niedriger als die Personengruppe, die sofort das Studium aufnimmt. Die Chancengleichheit ist erfüllt: gleiche Testwerte bedeuten nachweislich für beide Gruppen altersunabhängig eine gleiche Eignung im Studium. Daher ist es sehr wahrscheinlich, dass dieser Eignungsunterschied schon nach dem Zeitpunkt der Maturitätsprüfung bestand. Wenn die Personen entsprechende Erfahrungen in der Schule machten (wie leicht oder schwer diese fiel), kann dies auch von einer frühen Studienaufnahme abgehalten haben.

Vorgreifend können wir feststellen, dass für die Gruppe der Älteren mit später Maturitätsprüfung der **einzige Beleg vorliegt**, dass im Vergleich zu anderen Gruppen niedrigere Testwerte eine höhere Bestehenswahrscheinlichkeit der Prüfung bedeuten. Auch hier unterscheidet sich der Testwert innerhalb der Personengruppe zwischen Erfolgreichen

und weniger Erfolgreichen. Es handelt sich um eine lineare Verschiebung um 4 bis 5 Punkte nach unten. Eventuelle Konsequenzen für das Zulassungsverfahren wären zu bedenken. Personen des „zweiten Bildungsweges“ (um die es sich vor allem handelt) unterscheiden sich im „biografischen Verlauf“ des Lebens von denen, welche die Maturitätsprüfung auf dem ersten Bildungsweg abgelegt haben. Es wäre denkbar, dies im Sinne einer Punktegutschrift zu berücksichtigen, die abhängig vom Alter bei Ablegen der Maturitätsprüfung gegeben wird.

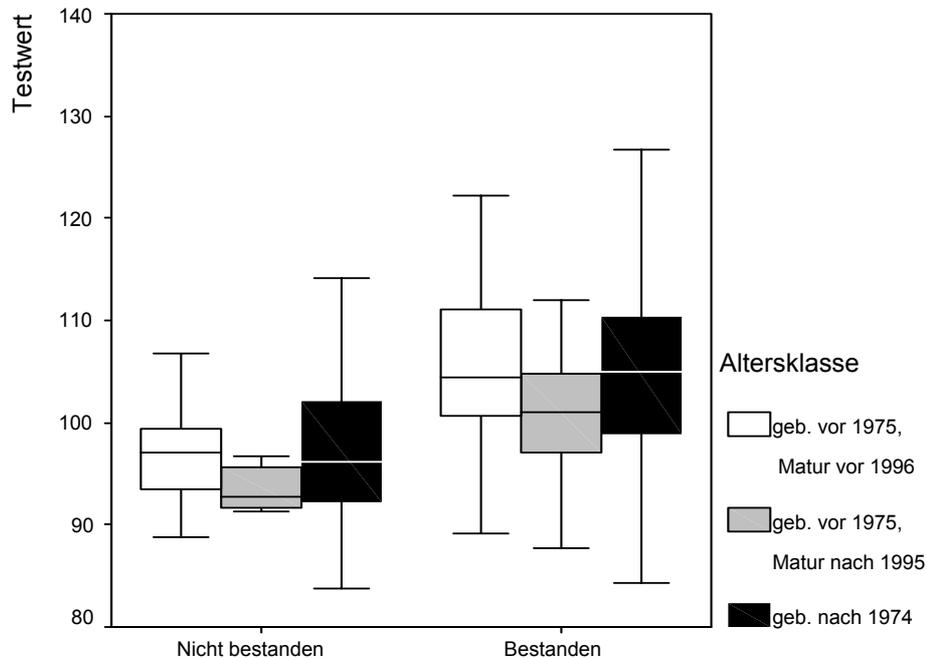


Abbildung 27: Boxplots des Testwertes für die Gruppen nach Alter und Maturitätsjahr nach dem Bestehen der Prüfung

## 7.2.6 Erfolgsvorhersage für Sprachgruppen

Auch für die drei Sprachgruppen treten vergleichbare Unterschiede nach dem Prüfungserfolg auf, wobei er in der italienischen Sprachgruppe tendenziell etwas weniger deutlich ist (sie ist allerdings auch die zahlenmässig kleinste Gruppe). Dieser Befund ist bedeutsam, weil er auch zeigt, dass es keine unterschiedliche Eignung bei gleichem Testwert zwischen den Sprachgruppen gibt. Die Chancengleichheit zwischen den Sprachgruppen ist gegeben.

Sprache	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
deutsch	216 (21%)	<b>97.1</b>	6.3	806 (79%)	<b>104.8</b>	8.1
französisch	41 (53%)	<b>95.9</b>	7.5	37 (47%)	<b>104.1</b>	7.3
italienisch	19 (38%)	<b>98.3</b>	6.7	31 (62%)	<b>102.7</b>	7.7

Tabelle 37: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Sprachgruppen – beide Kohorten zusammengefasst

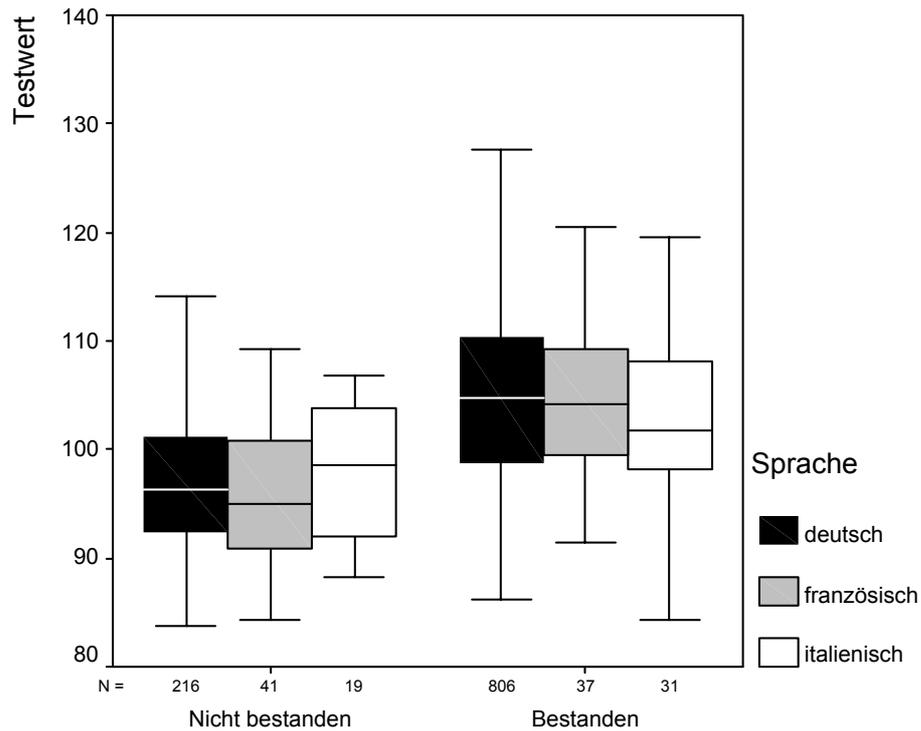


Abbildung 28: Boxplots des Testwertes für Sprachgruppen nach Bestehen der Prüfung

### 7.2.7 Unterschiede zwischen Human- und Veterinärmedizin

Einbezogen wurde der Testjahrgang 1999, wo erstmals für Human- und Veterinärmedizin ein Numerus Clausus zur Anwendung kam. Berücksichtigt sind in diesen Statistiken diejenigen Personen, die 1999 tatsächlich immatrikuliert worden sind. Tabelle 38 zeigt zunächst, dass in Veterinärmedizin weniger Personen die erste Vorprüfung bestehen als in Humanmedizin. Ist die Prüfung dort strenger – oder unterscheiden sich die Personen hinsichtlich ihrer Eignung?

Prüfungsverlauf	Humanmedizin		Veterinärmedizin	
	Anzahl	Prozent	Anzahl	Prozent
Noch nicht angetreten	65	12.4%	18	12.8%
2 x Nicht bestanden 1. und 2. Session	10	1.9%	1	0.7%
1 x Nicht bestanden 2. Session	54	10.3%	17	12.1%
1 x Nicht bestanden 1. Session	60	11.4%	41	29.1%
Bestanden 1. Session	272	51.8%	53	37.6%
Bestanden 2. Session	40	7.6%	5	3.5%
Bestanden 2. Session, Nicht bestanden 1. Session	24	4.6%	6	4.3%
Total	525		141	

Tabelle 38: Prüfungsverlauf für 1999 immatrikulierte Personen nach Disziplin

Für Eignungsunterschiede spricht, dass sich die mittleren Testwerte für Personen, welche die Prüfung bestehen, nicht signifikant zwischen Human- und Veterinärmedizin unterscheiden. Gleiches gilt für die mittleren Testwerte der weniger Erfolgreichen<sup>1</sup>. Bei gleichem Testwert (sprich gleicher Eignung) ist der gleiche Prüfungserfolg wahrscheinlich. Die Zusammensetzung der Studierenden in Veterinärmedizin unterscheidet sich von derjenigen in Humanmedizin: die Testwerte sind im Mittel geringer.

	Humanmedizin			Veterinärmedizin		
	n	Testwert		n	Testwert	
		Mittelwert	Stand.abw.		Mittelwert	Stand.abw.
Nicht bestanden	124	96.9	7.2	59	97.9	6.0
Bestanden	336	105.3	8.8	64	104.4	6.8

Tabelle 39: Testwert für 1999 immatrikulierte Personen nach Disziplin und Prüfungserfolg

	Humanmedizin			Veterinärmedizin		
	n	Testwert		n	Testwert	
		Mittelwert	Stand.abw.		Mittelwert	Stand.abw.
Noch nicht angetreten	65	98.7	8.1	18	96.5	5.8
Zweimal Nicht bestanden	10	96.0	7.4	1	94.2	-
Einmal Nicht bestanden	114	97.0	7.2	58	97.9	6.1
Zweiter Anlauf Bestanden	24	101.2	8.0	6	103.2	7.4
Sofort Bestanden	312	105.6	8.8	58	104.5	6.8

Tabelle 40: Testwert für 1999 immatrikulierte Personen nach Disziplin und Prüfungsverlauf

<sup>1</sup> t-Test Vergleich Human- zu Veterinärmedizin für Bestanden:  $t = 0,37$ ,  $p \leq 0.67$ ; für Nicht bestanden:  $t = 0.98$ ,  $p \leq 0.32$ .

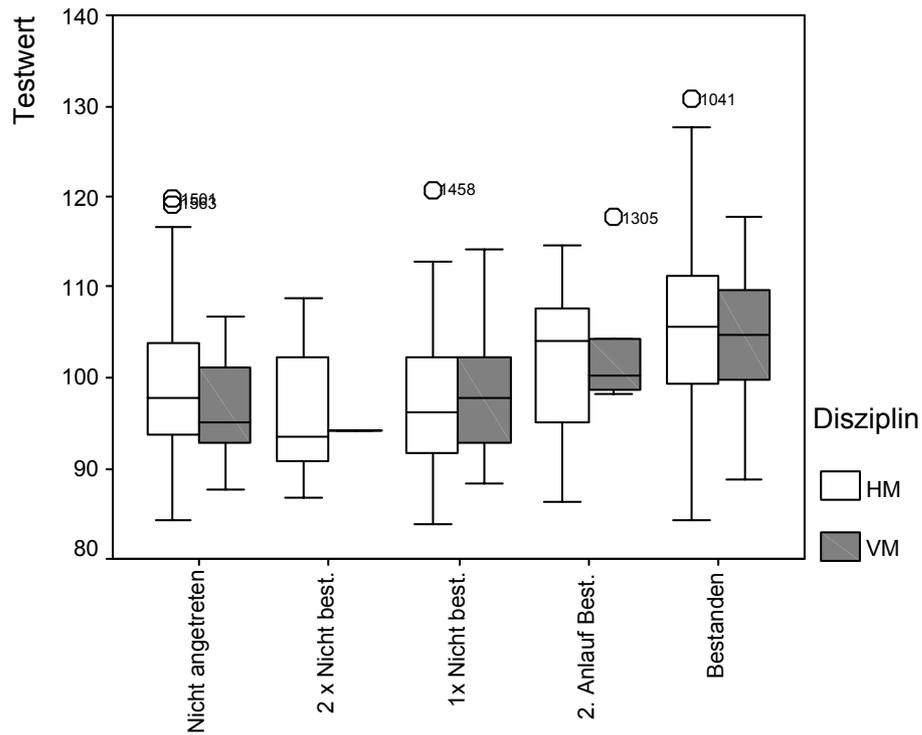


Abbildung 29: Testwertverteilung (Median und Quartile) für 1999 immatrikulierte Personen nach Disziplin und Prüfungsverlauf

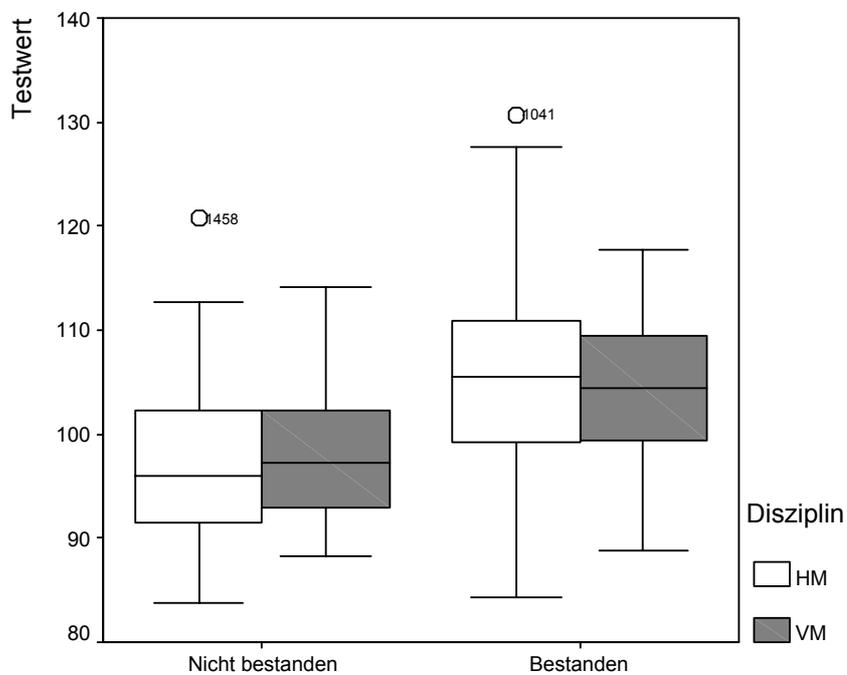


Abbildung 30: Testwertverteilung (Median und Quartile) für 1999 immatrikulierte Personen nach Disziplin und Prüfungserfolg

## 7.2.8 Zweite Vorprüfung

Die zweite Vorprüfung kann bisher nur durch die Personen der Kohorte 1998 absolviert worden sein. Dabei sind nur Personen angetreten, welche die erste Vorprüfung bestanden haben. Die Zusammensetzung ist daher nicht mit der Situation vor der ersten Vorprüfung zu vergleichen. Personen mit geringeren Testwerten sind in dieser Gruppe noch unterrepräsentiert und durch diese spezielle „Homogenisierung“ ist die Aussagefähigkeit eingeschränkt. Erst in den Folgejahren werden mehr Personen mit geringeren Testwerten über Prüfungsergebnisse für die 2. Vorprüfung verfügen. Deshalb hat auch die Gruppe der Personen, welche die Prüfung nicht bestanden hat, mit im Mittel 102.1 einen vergleichsweise hohen Testwert. Dennoch unterscheiden sich die Testwerte für Erfolgreiche und nicht Erfolgreiche signifikant (Tabelle 41).

Prüfungserfolg	n	m	s		
Nicht bestanden	52	102.15	6.623		
Bestanden	294	105.99	7.605		
Varianzgleichheit angenommen	Levene's Test Varianz		t-test Mittelwert		
	F	Sig.	t	df	Sig. (2-seitig)
	1.374	.242	-3.411	344	.001

**Tabelle 41:** Mittelwert (m) und Standardabweichungen (s) des Testwertes für 2. Vorprüfung nach Bestehen, sowie Ergebnisse des Mittelwertvergleiches (t-Test)

Prüfungsverlauf	n	m	s
Wegen Nichtbestehen 1. Vorprüfung noch nicht möglich	126	97.5	6.3
Zeitlich noch nicht möglich (1. Vorprüfung in 3. oder 4. Session)	79	100.0	5.4
Noch nicht angetreten, wäre aber möglich gewesen	49	103.9	8.6
1 x Nicht bestanden 4. Session	41	102.5	6.6
1 x Nicht bestanden 3. Session	11	101.0	6.7
Bestanden 4. Session, Nicht bestanden 3. Session	2	100.5	13.4
Bestanden 3. Session	110	105.8	7.8
Bestanden 4. Session	182	106.2	7.4

**Tabelle 42:** Mittelwert (m) und Standardabweichung (s) für die Gruppen nach dem Erfolg in der zweiten Vorprüfung

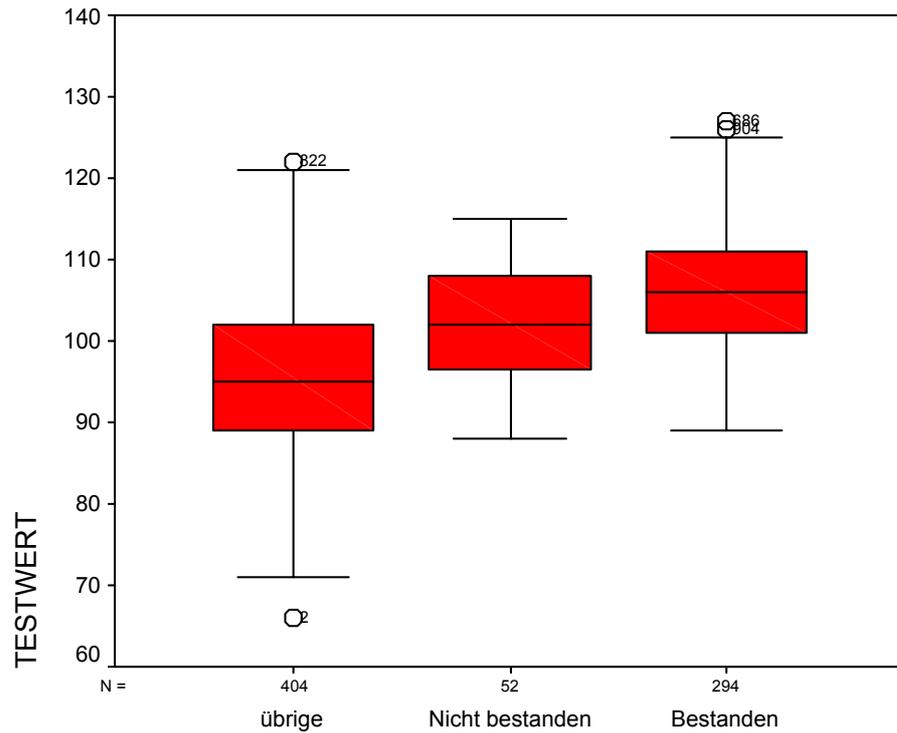


Abbildung 31: Boxplots des Testwertes für die Gruppen nach dem Bestehen der 2. Vorprüfung

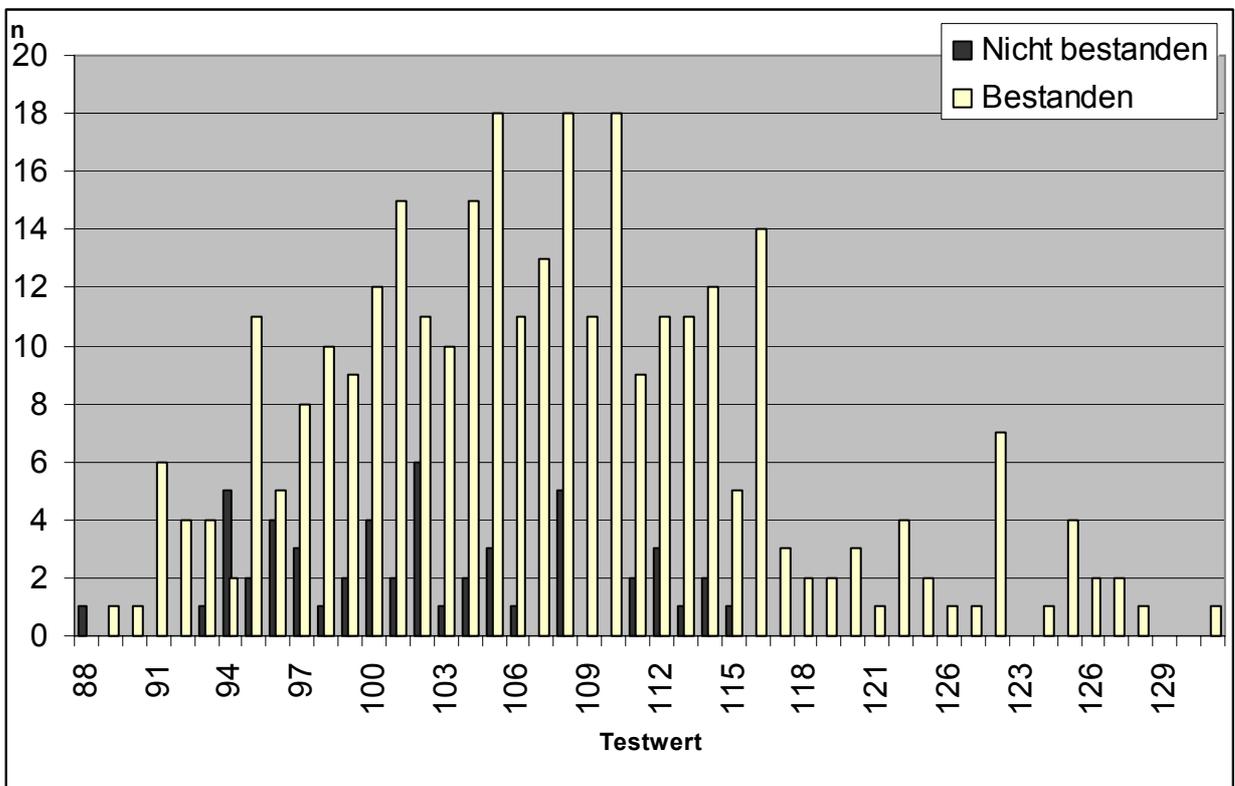


Abbildung 32: Häufigkeitsverteilung des Bestehens der 2. Vorprüfung nach Testwerten

Dass die Prognosekraft des Testwertes für die 2. Vorprüfung keinesfalls geringer sein wird als für die erste Vorprüfung belegt ein Vergleich der punktbiserialen Korrelationen (Assoziationskoeffizient Eta) zwischen Testwert und dem dichotomen Bestehenskriterium. Für die erste Vorprüfung beträgt dieser Wert 0.42 und für die zweite Vorprüfung 0.40. Rechnet man bei der zweiten Vorprüfung alle Personen, die schon zur Prüfung hätten antreten können als „nicht bestanden“ dazu, beträgt Eta sogar 0.44.

### 7.3 Testwert und Prüfungsnoten (Erste Vorprüfung)

Neben der Vorhersage der Studiendauer ist die Studienleistung das zweite wichtige Evaluationskriterium. Dessen Untersuchung wird dadurch erschwert, weil sich die Bewertungssysteme zwischen den Universitäten und zwischen den Jahren unterscheiden. Durch die Vergabe von (gleichen) Kreditpunkten im Falle des Bestehens der Prüfung wird auf die Notendifferenzierung z.T. bewusst verzichtet (Bern). Nachfolgend werden diejenigen Noten analysiert, die eine leistungsabhängige Streuung aufweisen und an einer ausreichenden Stichprobe gewonnen werden konnten.

Bezüglich der Vorhersage der Noten der zweiten Vorprüfung aufgrund des Testwertes sind noch keine statistisch gesicherten Aussagen möglich. Zum einen ist die Personengruppe mit vorliegenden Noten eine „homogenisierte“ Auswahl der Leistungsbesten der Kohorte 1998 (mit geringerer Stichprobengrösse), wo die Notenstreuung noch zu gering ist. In den Folgejahren werden mehr Personen die zweite Vorprüfung bestehen und es wird wahrscheinlich auch hier zu Notenverteilungen mit mehr Variation kommen.

Die erste Vorprüfung wurde in Bern 1999 (teilweise), sowie in beiden Jahren in Basel, Freiburg und Zürich in vier Fächern abgelegt. Erfasst wurden jeweils die letzten Noten, wenn eine Prüfung zweimal abgelegt wurde. In der Tabelle 43 wird deutlich, dass die Noten ausreichend streuen. Verwendet wurden auch die leistungsabhängig differenzierten Punktwerte in der ersten Vorprüfung nach dem Examenstyp „Reform 2“<sup>1</sup>.

Note	Physik und Physiologie (a)	Chemie und Biochemie (b)	Allgemeine und Humanbiologie I (c)	Allgemeine und Humanbiologie II (d)
1	4	7	3	2
2	47	70	32	45
3	192	161	107	118
4	354	390	244	209
5	270	256	353	376
6	56	40	185	174

Tabelle 43: Notenverteilung 1. Vorprüfung gesamt (ohne Bern Reform 1999 und 2000)

<sup>1</sup> Gemäss Verordnung vom 1.11.99 für die medizinische Fakultät Bern. Wir danken Herrn Rainer Hofer vom Institut für Aus-, Weiter- und Fortbildung (IAWF) der Universität Bern für die Bereitstellung dieser Daten.

	Bern				Basel				Freiburg				Zürich			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1		1	1	1	2				2	6		1			2	
2	6	6	9	4	6	11	2	6	17	27	8	14	18	26	13	21
3	5	9	8	6	55	28	17	31	39	31	34	23	93	93	48	57
4	17	11	12	11	87	96	48	44	46	52	73	54	203	230	110	100
5	29	28	18	26	65	81	122	103	46	38	44	57	130	108	169	190
6	7	9	16	16	17	16	43	48	19	15	10	20	13		116	90
Total	64	64	64	64	232	232	232	232	169	169	169	169	457	457	458	458
m	4.9	4.9	5.0	5.2	4.4	4.6	5.1	5.0	4.6	4.4	4.5	4.8	4.3	4.3	5.2	5.3
s	0.6	0.7	0.9	0.7	0.8	0.7	0.7	0.7	1.0	0.9	0.8	0.7	0.7	0.6	0.7	0.7

**Tabelle 44:** Notenverteilung 1. Vorprüfung nach Universitäten (ohne Bern Reform 1999 und 2000), Fächerbezeichnung siehe Kopf der Tabelle 43

	n	Minimum	Maximum	m	s
Praktische Fertigkeiten (mündliche Prüfung)	111	7	40	27.5	7.3
Multiple-Choice-Prüfung theoretische Fertigkeiten	111	62	173	121.8	25.6

**Tabelle 45:** Mittelwerte und Standardabweichungen erste Vorprüfung Bern 2000 (Reform 2)

Die Korrelationen des Testwertes beider untersuchten Jahrgänge mit den Noten der ersten Vorprüfung (nach 4 Sessionen) zeigt die Tabelle 46, jene nach 2 Sessionen für den Jahrgang 1999 die Tabelle 47. Die Korrelationen mit dem Testwert sind für alle Fächer und den Durchschnitt hoch. Nach 2 Sessionen sind die Korrelationen deshalb etwas höher, weil hier noch mehr schlechtere Noten der nicht bestandenen Prüfungen eingehen – dieser Anteil verringert sich nach 4 Sessionen und es kommt zu einer Homogenisierung der Notenvariation, die korrelationsmildernd wirkt. Die Ergebnisse nach 2 Sessionen sind mit den Befunden der Kohorte 1998 nach einem Jahr (Hänsgen und Spicher, 2000) voll vergleichbar. Die Untertests korrelieren wie erwartet unterschiedlich mit den Noten. Es sei daran erinnert, dass das Weglassen jedes Untertests die Zuverlässigkeit des Testwertes mindert und seine hohe Güte sich aus allen verwendeten Untertests ableitet.

Auch die Kreditpunkte der Prüfung in Bern korrelieren sehr hoch mit dem Testwert und den einzelnen Testvariablen. Das ist auch auf die höhere Variation und damit auf die bessere Differenzierung der Bewertung zwischen den Personen zurückzuführen.

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Durchschnitts- note
	N = 509	N = 509	N = 509	N = 509	N = 509
<b>Testwert</b>	<b>.432</b>	<b>.380</b>	<b>.400</b>	<b>.449</b>	<b>.495</b>
	.000	.000	.000	.000	.000
Muster zuordnen	.116	.160	.157	.176	.182
	.009	.000	.000	.000	.000
Med.-naturw. Grundverständnis	.306	.284	.287	.336	.362
	.000	.000	.000	.000	.000
Schlauchfiguren	.195	.152	.157	.207	.212
	.000	.001	.000	.000	.000
Quantitative und formale Probleme	.434	.371	.321	.359	.441
	.000	.000	.000	.000	.000
Textverständnis	.331	.240	.304	.352	.367
	.000	.000	.000	.000	.000
Figuren lernen	.159	.146	.172	.199	.203
	.000	.001	.000	.000	.000
Fakten lernen	.150	.178	.241	.220	.238
	.001	.000	.000	.000	.000
Diagramme und Tabellen	.316	.249	.239	.297	.328
	.000	.000	.000	.000	.000
Konzentr. und sorgf. Arbeiten	.105	.191	.216	.214	.218
	.018	.000	.000	.000	.000

**Tabelle 46:** Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung Kohorte 1998 nach 4 Prüfungssessionen (1999 und 2000), ohne Bern „Reform 2“

	Mündliche Prüfung gesamt (Bern)  N = 111	Multiple Choice Prüfung gesamt (Bern)  N = 111	Physik und Physiologie  N = 414	Chemie und Biochemie  N = 415	Allgemeine und Human- biologie I  N = 415	Allgemeine und Human- biologie II  N = 415	Durch- schnitts- note  N = 414
<b>Testwert</b>	<b>.548</b>	<b>.577</b>	<b>.507</b>	<b>.491</b>	<b>.500</b>	<b>.478</b>	<b>.554</b>
	.000	.000	.000	.000	.000	.000	.000
Muster zuordnen	.335	.322	.198	.197	.230	.212	.233
	.000	.001	.000	.000	.000	.000	.000
Med.-naturw. Grundverständnis	.386	.349	.410	.371	.413	.370	.438
	.000	.000	.000	.000	.000	.000	.000
Schlauchfiguren	.348	.392	.232	.197	.259	.256	.269
	.000	.000	.000	.000	.000	.000	.000
Quantitative und formale Probleme	.454	.494	.507	.464	.403	.357	.484
	.000	.000	.000	.000	.000	.000	.000
Textverständnis	.261	.309	.360	.338	.363	.348	.396
	.006	.001	.000	.000	.000	.000	.000
Figuren lernen	.293	.299	.202	.201	.213	.235	.241
	.002	.001	.000	.000	.000	.000	.000
Fakten lernen	.120	.138	.214	.230	.247	.235	.259
	.209	.147	.000	.000	.000	.000	.000
Diagramme und Tabellen	.484	.517	.457	.442	.375	.362	.454
	.000	.000	.000	.000	.000	.000	.000
Konzentr. und sorgf. Arbeiten	.278	.284	.189	.259	.244	.245	.266
	.003	.003	.000	.000	.000	.000	.000

**Tabelle 47:** Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung Kohorte 1999 nach 2 Prüfungssessionen 2000 (Bern „Reform 2“ und zusammengefasst Basel, Zürich, Freiburg)

In den beiden nachfolgenden Tabellen werden Human- und Veterinärmedizin verglichen. Festzuhalten ist eine hohe Korrelation der Noten mit den Testergebnissen, wobei sie in der Veterinärmedizin tendenziell sogar etwas höher sind. Die Prognosegüte des Tests für die Studienleistungen ist in beiden Disziplinen gegeben.

	Physik und Physiologie N = 346	Chemie und Biochemie N = 346	Allgemeine und Humanbiologie I N = 346	Allgemeine und Humanbiologie II N = 346	Durchschnitts- note N = 346
<b>Testwert</b>	<b>.495</b>	<b>.474</b>	<b>.482</b>	<b>.444</b>	<b>.533</b>
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Muster zuordnen	.194	.193	.210	.204	.224
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Med.-naturw. Grundverständnis	.407	.361	.395	.340	.422
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Schlauchfiguren	.233	.171	.235	.213	.245
	<i>.000</i>	<i>.001</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Quantitative und formale Probleme	.489	.458	.388	.341	.471
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Textverständnis	.342	.326	.359	.343	.387
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Figuren lernen	.228	.231	.238	.255	.270
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Fakten lernen	.220	.227	.253	.221	.258
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Diagramme und Tabellen	.470	.445	.380	.347	.457
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Konzentr. und sorgf. Arbeiten	.173	.237	.237	.213	.245
	<i>.001</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>

Tabelle 48: Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung für **Humanmedizin**, Kohorte 1999, ohne Bern „Reform 2“

	Physik und Physiologie N = 68	Chemie und Biochemie N = 68	Allgemeine und Humanbiologie I N = 68	Allgemeine und Humanbiologie II N = 68	Durchschnitts- note N = 68
<b>Testwert</b>	<b>.551</b>	<b>.569</b>	<b>.595</b>	<b>.660</b>	<b>.661</b>
	.000	.000	.000	.000	.000
Muster zuordnen	.210	.213	.314	.250	.279
	.086	.081	.009	.040	.021
Med.-naturw. Grundverständnis	.414	.420	.481	.518	.512
	.000	.000	.000	.000	.000
Schlauchfiguren	.171	.277	.328	.433	.341
	.164	.022	.006	.000	.004
Quantitative und formale Probleme	.573	.460	.441	.397	.515
	.000	.000	.000	.001	.000
Textverständnis	.448	.404	.377	.363	.439
	.000	.001	.002	.002	.000
Figuren lernen	.126	.105	.159	.205	.168
	.305	.394	.195	.093	.171
Fakten lernen	.141	.202	.203	.271	.228
	.252	.099	.097	.025	.061
Diagramme und Tabellen	.327	.357	.327	.388	.387
	.006	.003	.007	.001	.001
Konzentr. und sorgf. Arbeiten	.215	.325	.250	.382	.324
	.078	.007	.040	.001	.007

**Tabelle 49:** Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung für **Veterinärmedizin**, Kohorte 1999, ohne Bern „Reform 2“

Weitere Teilgruppen werden in der Tabelle 50 verglichen. Die Prognosegüte des Testwertes für die Studiennoten ist nicht spezifisch für Universitäten (wo es auch Variationen der Prüfungsarten Schriftlich Essay, Schriftlich Multiple-Choice und Mündlich gibt, die keinen Einfluss auf die Prognosegüte haben – vergleiche dazu Hänsgen und Spicher 2000). Auch bezüglich der Geschlechter, der Sprachgruppen und der Altersgruppen gibt es keine Hinweise auf Unterschiede hinsichtlich der Prognosegüte des Tests.

	n	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Human- biologie I	Allgemeine und Human- biologie II	Durch- schnittsnote
<b>Studienort</b>						
Bern (Teil 1999)	64	.636	.509	.479	.580	.585
		.000	.000	.000	.000	.000
Basel	233	.394	.350	.304	.377	.410
		.000	.000	.000	.000	.000
Freiburg	169	.499	.527	.381	.389	.533
		.000	.000	.000	.000	.000
Zürich	458	.516	.495	.506	.501	.577
		.000	.000	.000	.000	.000
<b>Alter und Maturität</b>						
geb. vor 1975, Matur vor 1996	30	.501	.579	.520	.426	.590
		.005	.001	.003	.019	.001
geb. vor 1975, Matur nach 1995	34	.407	.341	.450	.312	.465
		.017	.048	.007	.068	.005
geb. nach 1974	859	.472	.436	.447	.462	.521
		.000	.000	.000	.000	.000
<b>Geschlecht</b>						
männlich	411	.458	.403	.427	.407	.489
		.000	.000	.000	.000	.000
weiblich	512	.497	.473	.466	.501	.557
		.000	.000	.000	.000	.000
<b>Sprache</b>						
deutsch	829	.472	.433	.446	.460	.524
		.000	.000	.000	.000	.000
französisch	52	.571	.549	.464	.501	.596
		.000	.000	.001	.000	.000
italienisch	42	.407	.420	.408	.496	.474
		.000	.000	.000	.000	.000

**Tabelle 50:** Korrelation der Noten für die 1. Vorprüfung mit dem Testwert für verschiedene Teilgruppen (kursiv: das Signifikanzniveau)

## 7.4 Schlussfolgerungen zur Eignungsfeststellung

Die Ergebnisse der vorliegenden Evaluation lassen es zu, für den Zeitraum bis zur zweiten ärztlichen Vorprüfung gesicherte Erkenntnisse zum Zusammenhang zwischen Zulassung aufgrund des Eignungstests und der Eignung zu gewinnen:

- Der Eignungstest für das Medizinstudium in der Schweiz (EMS) erfasst nachweislich Studieneignung, indem (1) Personen mit höheren Testwerten die Prüfungen mit höherer Wahrscheinlichkeit und zeitlich eher bestehen als die mit niedrigeren und (2) Personen mit höheren Testwerten auch eine höhere Studienqualität erreichen, die sich in besseren Prüfungsnoten ausdrückt.
- Der Zusammenhang zwischen Testergebnis und Studieneignung kann auf einem statistisch bedeutsamen Niveau festgestellt werden, welches die Verwendung des Testwertes als Zulassungskriterium rechtfertigt.
- Obwohl die Ergebnisse der zweiten Vorprüfung erst für wenige Personen vorliegen gibt es ausreichend Hinweise, dass die Vorhersagekraft des Testwertes für beide Vorprüfungen gleichermaßen hoch ist.
- Die Universitäten sind jetzt in der Lage, eine kapazitätsentsprechende Zulassung vorzunehmen und dadurch die Studienbedingungen zu verbessern. Das ist der wichtigste Vorteil des EMS, der seine Anwendung rechtfertigt. Ausserdem kann mit seiner Hilfe fair entschieden werden, wer an der „Wunschuniversität“ studieren kann und wer eine Umleitung in Kauf nehmen muss.
- Was sich offenbar nicht ändert, ist die Zahl der Geeigneten, die sich für ein Medizinstudium anmelden. Der Numerus Clausus ist mit rund 20% abgewiesener Bewerbungen sehr mild. Von den 80% Zugelassenen tritt ein Teil das Studium nicht an – Umleitungen von der Wunsch- an eine andere Universität sind eine wichtige Ursache. Dies führt nun dazu, dass die freiwerdenden Studienplätze an weniger Geeignete aus den 20% der Personen mit den geringeren Testwerten vergeben werden und das Niveau der Eignung der Studierenden gegenüber der Situation ohne Numerus Clausus praktisch unverändert bleibt. Die Zulassung von weniger Geeigneten, nur um Kapazitäten auszunutzen, hat nachweislich wenig Erfolgsaussicht: diese Personen haben um die Hälfte verringerte Erfolgchancen bei der ersten Vorprüfung, bei der zweiten Vorprüfung sind bisher sogar nur zwei Personen aus der „Warteliste“ angelangt.
- Das jetzt praktizierte Verfahren könnte ohne jegliche Systemveränderung im schlimmsten Falle dazu führen, dass die Studiendauern sich nur langsam verkürzen, die Abbrecherzahlen und Notenverteilungen sich nur unwesentlich gegenüber dem Zeitraum vor Einführung des NC verändern. Vor allem die bisher immer auch gewollten Dissuasionseffekte durch Test und Information über Studien- und Berufsbedingungen wären hinsichtlich möglicher selektiv-negativer Wirkungen zu hinterfragen. Welche Möglichkeiten zur Systemoptimierung gäbe es, um den relativen Anteil der Geeigneten im Medizinstudium zu erhöhen und dadurch auch den Einsatz der Ausbildungsressourcen zu optimieren?
  - o **Einflüsse auf das Bewerbungsverhalten**, indem vermehrt Geeignete angesprochen werden. Zu analysieren ist dabei, ob durch die „gewollten“ Dissuasionsfaktoren NC, Eignungstest und der geschmälernten Berufsperspektive („Ärztenschwemme“, Kostenregulierung u.a.) selektiv die Geeigneteren stärker abgeschreckt werden, sich zu bewerben. Möglicherweise muss das „Dissuasionsparadigma“ generell revidiert

werden. Dazu gehört auch die Klärung der Frage, warum der Männeranteil bei den Bewerbungen so viel geringer ist und von Jahr zu Jahr weiter abnimmt<sup>1</sup>.

- Personen, die sich trotz Zulassung nicht immatrikulieren, weisen ebenfalls nachweislich geringere Testwerte auf. Da ein grösserer Teil davon die Umgeleiteten betrifft und diejenigen mit niedrigeren Testwerten auch eher umgeleitet werden, ist dies nicht verwunderlich.
- **Verzicht, Personen zum Studium zuzulassen**, die nachweislich ein „Mindest-Niveau“ der **Eignung nicht erreichen** – wo die Wahrscheinlichkeit, Prüfungen zu bestehen, von vornherein als gering eingeschätzt werden muss. Dabei würde man in Kauf nehmen müssen, die festgelegten Kapazitäten zu unterschreiten (die aktuellen Kapazitäten der ersten beiden Jahre beinhalten ihrerseits noch Schätzungen von Rückzügen, Abbrüchen und Verzögerungen, die auf den Zahlen der Situation vor dem NC beruhen).
- Breite **Einführung der Reformstudiengänge**, um den Geeigneten dann solche Studienbedingungen zu verschaffen, die eine Optimierung von Studiendauer und Studienqualität erlauben, wodurch ein Studienabbruch dann einmal die Ausnahme würde.

Eine zu klärende Frage für das Zulassungsverfahren selbst wäre die Einführung einer Bonifikation für Personen, die älter sind UND die Maturität später abgelegt haben. Dies ist die einzige Teilgruppe, für die nachgewiesen werden kann, dass sich die Basis für die Vorhersage des Prüfungsbestehens aufgrund der Testwerte linear in Richtung niedrigerer Werte verschiebt. Eine mögliche Lösung wäre, den Mittelwertunterschied der Erfolgreichen als maximale Gutschrift proportional zur Differenz zwischen Lebensalter und Maturitätsalter für die Gruppe zu vergeben, die mehr als 6 Jahre älter sind als die jüngsten normalerweise erwarteten Bewerberinnen und Bewerber.

*Beispiel:  $104.8 - 101.1 = 3.7$  Punkte Differenz zwischen den erfolgreichen Jüngeren und Älteren mit später Maturität. Maturitätsalter liegt zwischen 18 und 34 Jahre. Gutschrift 3.7 Punkte bei 34 Jahren, 0 Punkte bei 23 Jahren – dazwischen ab 24 Jahren entsprechend einer linearen Funktion.*

---

<sup>1</sup> Diese Analyse ist ohne Auswirkung auf die Chancengleichheit zu verstehen. Ein wichtiger und ermutigender Befund dieser Evaluation ist, dass Männer und Frauen bei gleichem Testwert und damit gleicher Eignung nahezu identische Chancen im Studium haben, Prüfungen in optimalen Zeiträumen zu bestehen und gute Noten zu erreichen. Es geht nicht darum, den absoluten Frauenanteil bei den Bewerbungen zu verringern, sondern im Gegenteil sollte es weiterhin ein Ziel sein, mehr Frauen zum Erwerb der Hochschulreife und der Aufnahme eines Studiums zu motivieren und die entsprechenden Bedingungen dafür zu schaffen.

Es geht um die Frage, warum der Männeranteil absolut und relativ beständig abnimmt. Uns scheint die Vermutung nicht ganz abwegig zu sein, dass – aus einem alten Rollenverständnis heraus – Männer durch den Test, die Möglichkeit einer Ablehnung und die Rückmeldung des Testergebnisses mit einer objektiven Vergleichsmöglichkeit zu anderen Personen stärker abgeschreckt werden als Frauen, wo diese Meidung von Konkurrenzsituationen im Leistungsbereich und „Versagensängste“ nicht so stark aus einem traditionellen Rollenverständnis begründet wären (obgleich sie natürlich auftreten können). Träfe das zu, wäre der stärkere Rückzug der Männer tatsächlich ein Hinweis, dass noch viel für die Überwindung der traditionellen Rollenverständnisse getan werden muss und ist insofern auch bedenklich.

## 8 Testanwendung in der Schweiz 2001

### 8.1 Organisation der Testabnahme

Im vierten Jahr des Testeinsatzes konnte auf erfahrene Strukturen zur Testorganisation und Abnahme zurückgegriffen werden. Pro Testort war ein Koordinator für die Organisationsfragen zuständig. Bezüglich aller Testlokale lagen Erfahrungen vor, wie das Verhältnis von Grösse und Betreuungsstrukturen optimal gestaltet sein muss. Eine Testleiterin/ein Testleiter war verantwortlich – eine Stellvertretung und je nach Grösse des Testlokals entsprechendes Aufsichtspersonal sorgten für gleiche Bedingungen an allen Testorten. Eine eintägige Schulung aller Testleiterinnen und Testleiter diente dem gleichen Zweck. Hier wurde vor allem festgelegt, auf alle vorhersehbaren Ereignisse vergleichbar zu reagieren. Auch die vorzulesenden Instruktionen sind vereinheitlicht, damit in jedem Testlokal tatsächlich von gleichen Bedingungen ausgegangen werden kann.

In diesem Jahr sind keine Testabbrüche aufgetreten. Alle Personen, die den Test begannen, konnten ihn auch mit einem gültigen Testergebnis beenden. Alle Testabläufe wurden genauestens protokolliert – beispielsweise auch die genaue Einhaltung der Zeiten. Aus den Protokollen ergeben sich keine Hinweise darauf, dass irreguläre Bedingungen vorhanden gewesen wären. Die Testabnahme kann daher in allen Testlokalen als uneingeschränkt gültig angesehen werden.

### 8.2 Verteilungsprüfung

Der Testwert ist wiederum normalverteilt, eine parametrische Prüfung ist zulässig. Die Punktwerte der Einzeltests sind, trotz der optischen Ähnlichkeit mit der eingezeichneten Normalverteilung in den Abbildungen 15 und 16 von der Normalverteilung unterschiedlich.

	Testwert	Musterzuordnen	Med.-naturw. Grundv.	Schlauchfiguren	Quant.u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konz. u. sorgf. Arbeiten
Mittelwert	100.03	10.14	10.07	13.27	10.91	8.52	10.60	10.78	10.27	13.07
Stand.-abweich.	9.985	2.812	3.466	3.743	3.912	3.462	3.469	3.654	3.134	3.702
extremste Differenz	.037	.074	.069	.096	.062	.079	.075	.088	.073	.096
Kolmog.-Smirnov Z	1.089	2.187	2.049	2.849	1.846	2.344	2.227	2.618	2.159	2.835
Asymp. Sig.	.186	.000	.000	.000	.002	.000	.000	.000	.000	.000

**Tabelle 51:** Verteilungsprüfung auf Normalverteilung (Kolmogorov-Smirnov Test)

In den Untertests Schlauchfiguren und Konzentriertes und sorgfältiges Arbeiten sind die Verteilungen deutlich rechtsschief, treten gehäuft hohe bzw. volle Punktzahlen auf. Dennoch ist die Differenzierung im unteren Leistungsbereich noch ausreichend, sodass keine Testveränderung (z.B. durch Zeitverkürzung) notwendig wird.

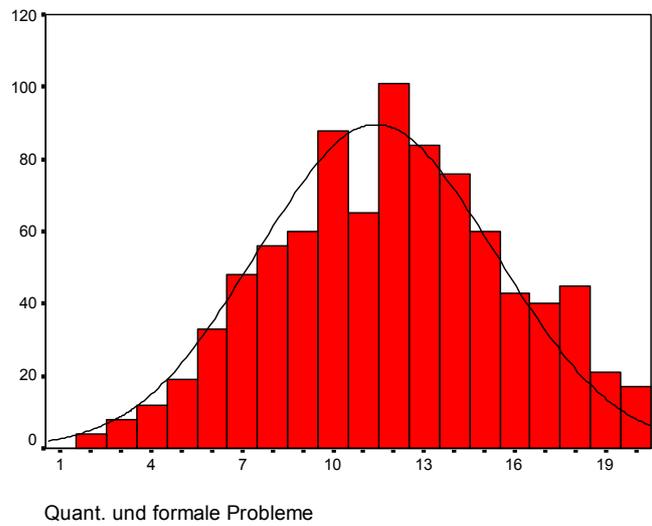
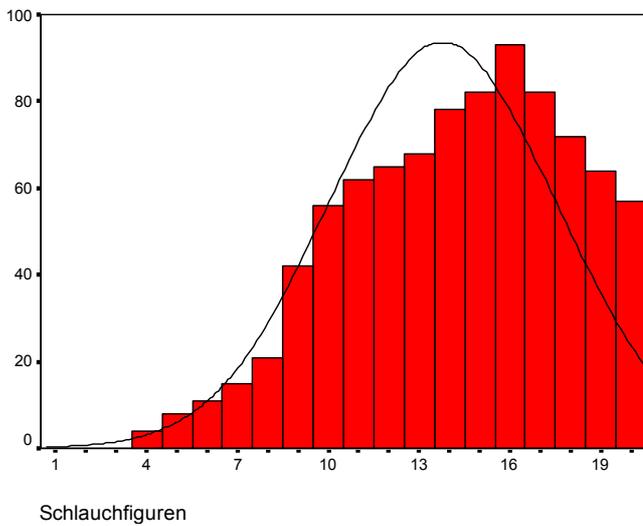
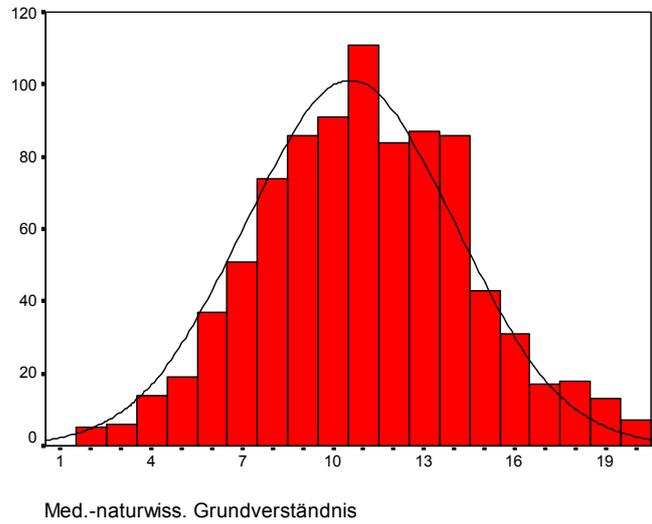
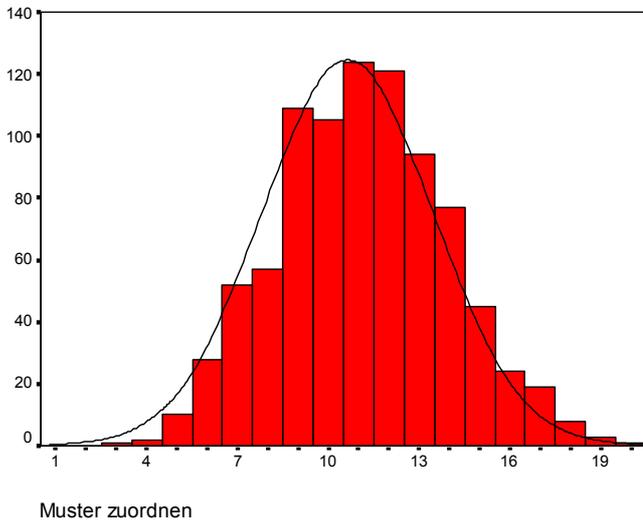
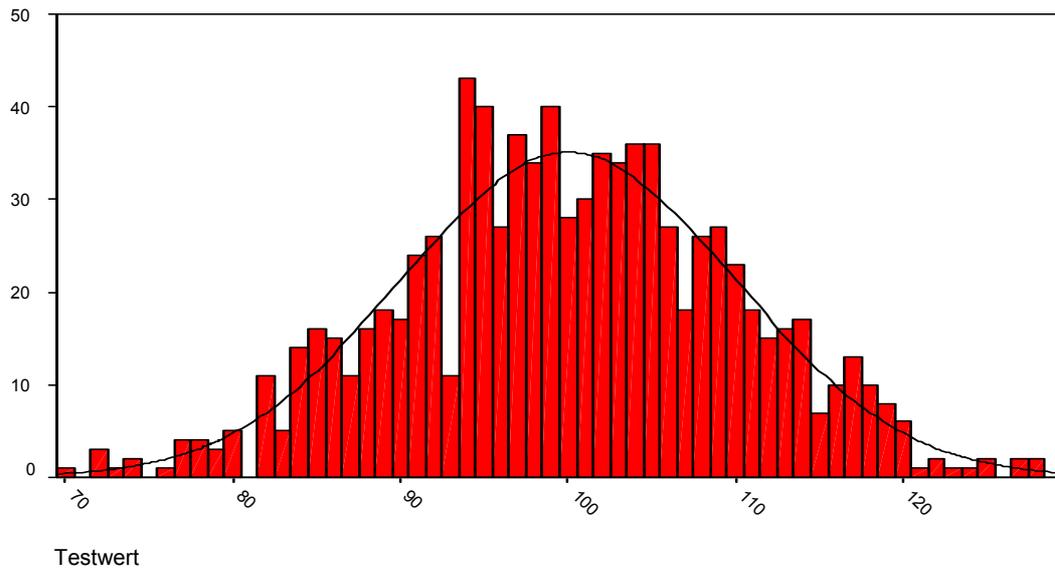
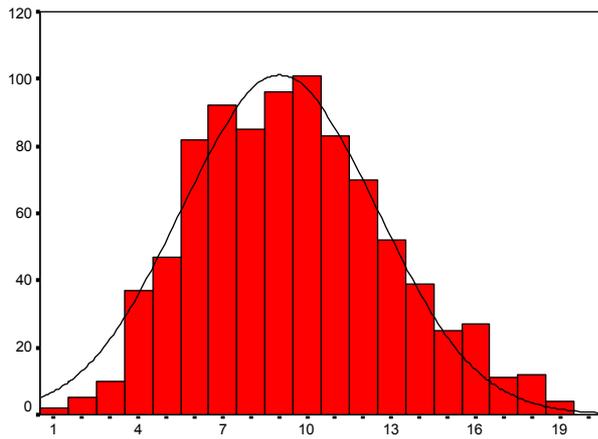
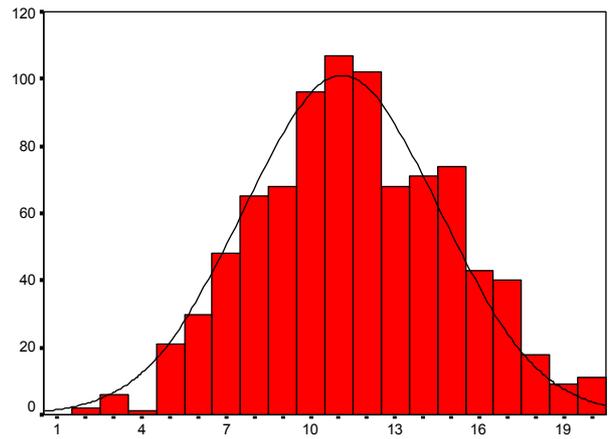


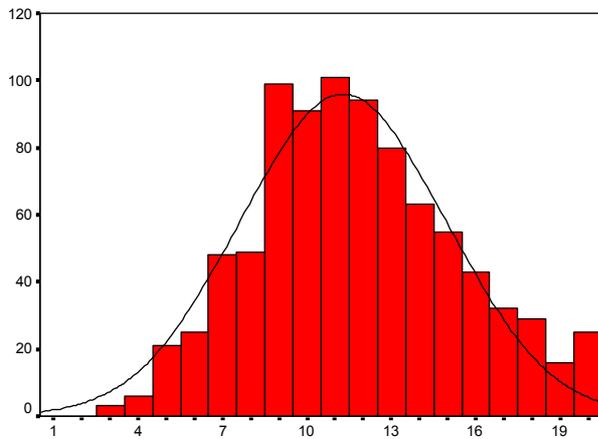
Abbildung 33: Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 1-4



Textverständnis



Figuren lernen



Fakten lernen

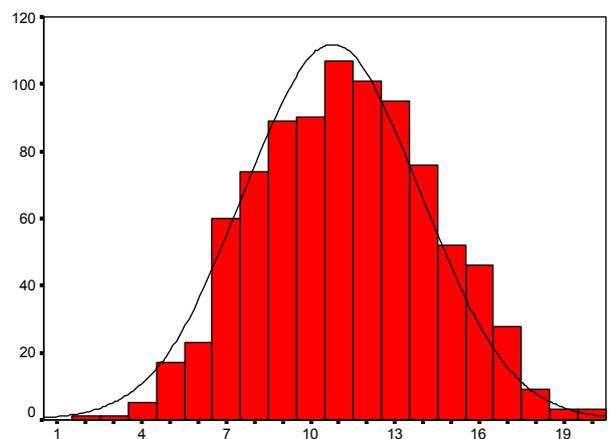
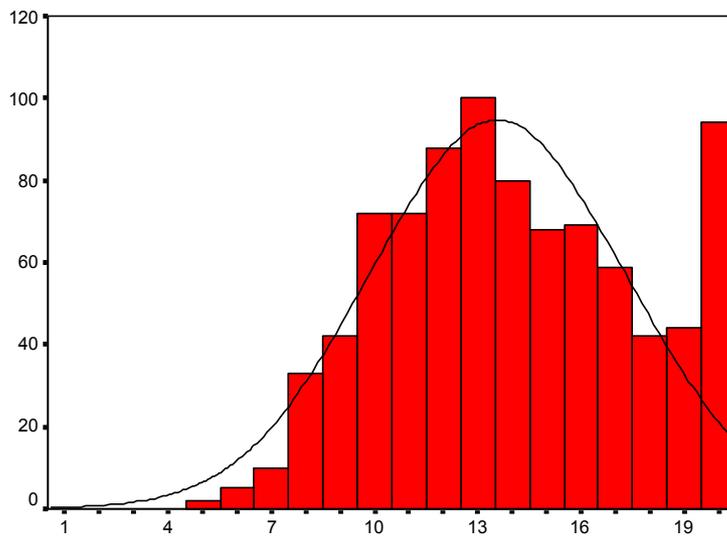


Diagramme und Tabellen



Konzentr. u. sorgf. Arbeiten

Abbildung 34: Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 5-9

### 8.3 Vergleich der Testfassungen 1998 bis 2001

Durch die Transformation der Punktwerte in eine Skala unter Berücksichtigung des jeweiligen Mittelwertes und der Standardabweichungen pro Jahr ist gewährleistet, dass die **Testwerte** aus verschiedenen Jahren miteinander verglichen werden können. Da jedes Jahr vergleichbare Gütekriterien ermittelt werden konnten, wird dies zusätzlich legitimiert.

Jeder **Testwert** bedeutet in jedem Jahr also das Gleiche und er kann deshalb ohne Benachteiligung ins Folgejahr übernommen werden, wenn ein späterer Studienantritt gewünscht wird.

Die "absolute" Schwierigkeit als **Punktwert** unterliegt andererseits über die Jahre geringfügigen Schwankungen. Tabelle 52 zeigt den Vergleich für die bisherigen vier Testdurchführungen in der Schweiz mit den entsprechenden geschätzten Punktwerten aus Deutschland. Zur Schätzung wurden die aus Deutschland berichteten Item-Schwierigkeiten herangezogen. Für die Berechnung der Schweizer Punktwerte wurden nur die Ergebnisse der deutschsprachigen Gruppe berücksichtigt.

	D 2001	CH 2001 Gesamt	D 2000	CH 2000 Gesamt	D 1999	CH 1999 Gesamt	D 1998	CH 1998 nur HM
Muster zuordnen	9.7	10.1	9.3	10.3	11.0	11.6	12.2	13.2
Med.-naturwiss. Grundverst.	10.7	10.1	10.8	10.7	11.2	11.3	11.6	12.5
Schlauchfiguren	12.2	13.3	12.3	13.2	11.9	12.9	11.8	13.3
Quant. und form. Probleme.	9.8	10.9	10.2	11.8	9.7	11.6	9.9	11.9
Textverständnis	8.6	8.5	9.3	9.0	8.9	8.9	10.2	10.3
Figuren lernen	9.2	10.6	9.3	10.2	11.8	12.7	11.6	12.6
Fakten lernen	9.3	10.8	8.9	10.1	11.0	11.8	11.2	11.6
Diagramme und Tabellen	10.0	10.3	9.7	10.4	10.3	10.5	11.0	11.4
<b>Total</b> (ohne Konzentriertes und sorgfältiges Arbeiten)	<b>79.5</b>	<b>84.6</b>	<b>79.8</b>	<b>85.7</b>	<b>85.8</b>	<b>91.3</b>	<b>89.5</b>	<b>96.8</b>

Tabelle 52: Mittelwerte (Punkte) der Untertests 1998 bis 2001 deutsche Sprachgruppe Schweiz und Schätzung aufgrund der Item-Schwierigkeitswerte aus Deutschland (HM: Humanmedizin)

Über die vier Durchführungen kann eine Verringerung der absoluten Punktzahlen beobachtet werden. Da diese Tendenz mit den deutschen Kennwerten übereinstimmt, ist eine leichte Zunahme der absoluten Schwierigkeit des Tests zu vermuten. Weiterhin ist die Streuung der Punktwerte ausreichend, um die notwendige Personenunterscheidung zu erreichen.

Seit Beginn der Testdurchführung werden im Untertest Schlauchfiguren durchschnittlich die höchsten Punktzahlen erreicht. Beim Untertest Textverständnis handelt es sich in der Schweiz um den mit Abstand schwierigsten Untertest. In Deutschland zeigt sich

diesbezüglich allerdings ein ähnliches Bild. Dort wurden einzig im Test aus dem Jahr 2000 vergleichbar geringe Punktzahlen noch in den Untertests Figuren lernen und Fakten lernen erreicht.

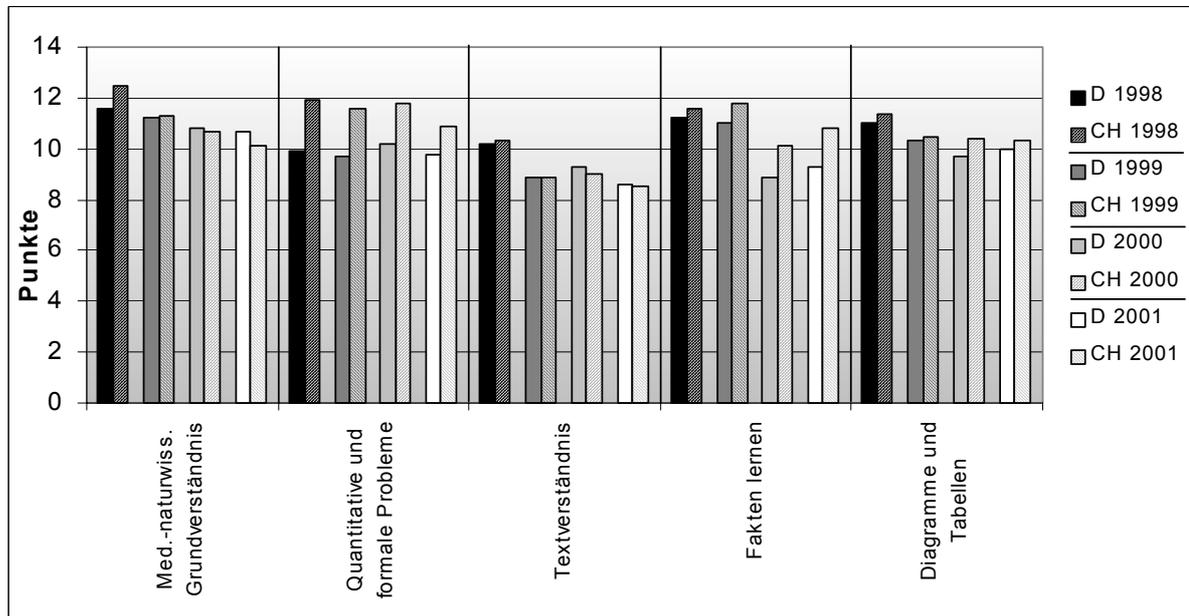


Abbildung 35: Mittelwerte eher sprachabhängiger Untertests Schweiz (deutschsprachig) und Vergleichswerte Deutschland 1998 bis 2001 – jeweils Humanmedizin

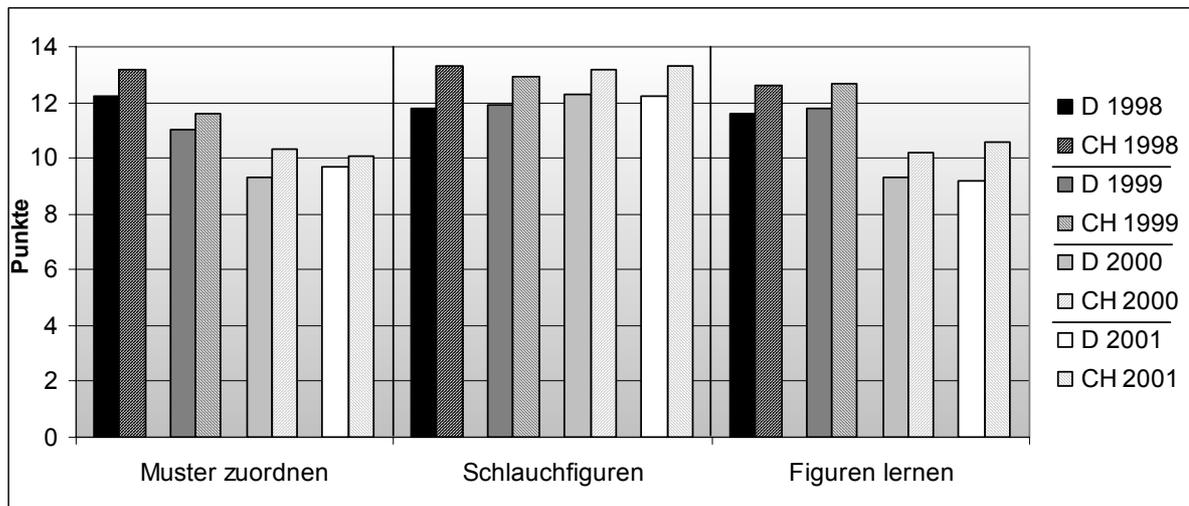


Abbildung 36: Mittelwerte wenig sprachabhängiger Untertests Schweiz (deutschsprachig) und Vergleichswerte Deutschland 1998 bis 2001 – jeweils Humanmedizin

Bezogen auf den Vergleich Schweiz - Deutschland setzen sich die Trends in Textverständnis und Medizinisch-naturwissenschaftliches Grundverständnis fort (Abbildung 35). Es handelt sich hier erneut um jene Untertests, in welchen die Schweizer Stichprobe schlechter abschneidet als die deutschen Vergleichsdaten. Die besseren Leistungen der Schweizer Stichprobe in den anderen Untertests sind auch durch die höhere Bewerberquoten in Deutschland begründbar. Das schlechtere Abschneiden in den beiden angesprochenen Tests könnte allerdings auch eine Bestätigung der von Oswald (1999) vertretenen These sein, dass die Fähigkeiten, Textaufgaben zu lösen – d.h. in Texten elementare logische Zusammenhänge zu erkennen – relativ schlecht entwickelt seien.

Unter den wenig sprachabhängigen Tests (Abbildung 36) bleiben die Differenzen zu den deutschen Referenzkennwerten bis auf den Untertest Muster zuordnen recht konstant.

## 8.4 Äquivalenz der Sprachversionen

Der Test wurde nach den gleichen Kriterien wie in den Vorjahren von der deutschen Vorlage adaptiert. Eine ausführliche Beschreibung des Verfahrens kann den Berichten 4 (1998) und 5 (1999) entnommen werden. Zielsetzung waren möglichst gut vergleichbare Aufgabenschwierigkeiten zwischen den Sprachversionen.

Chancengleichheit als Gleichbehandlung der Sprachgruppen muss gewährleistet sein, die Lösungswahrscheinlichkeit einer Aufgabe muss bei gleicher Fähigkeit tatsächlich auch vergleichbar sein.

Diese Gleichheit ist nicht allein durch Übersetzung zu beeinflussen, sondern hängt von mehreren Faktoren ab. Bei allen Adaptationen muss mit solchen Unterschieden gerechnet werden. Die Abbildung 37 fasst zusammen, welche Ursachen für Unterschiede in den drei Sprachgruppen verantwortlich sein können:

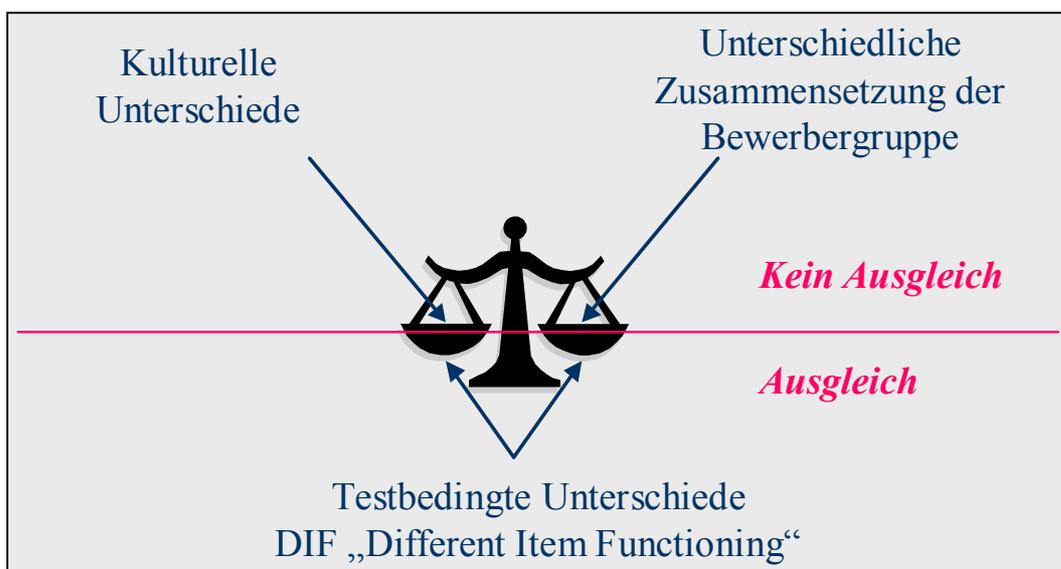


Abbildung 37: Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen

Bei den ersten beiden Ursachen handelt es sich um „echte“ Personen-Unterschiede, die nicht auf den Test oder seine Adaptation zurückgeführt werden können. Sie dürfen dementsprechend nicht ausgeglichen werden.

Testbedingte Unterschiede werden durch die Adaptation verursacht und sollen möglichst vollständig ausgeglichen werden. Beispiele wären eine Vereinfachung oder Erschwerung der Aufgabe bei der Übersetzung (Wortwahl, Satzgliederung o.ä.).

Wie bereits in den Vorjahren wird dazu das DIF-Verfahren<sup>1</sup> angewendet – eine ausführliche Beschreibung kann beispielsweise in Bericht 5 (1999) nachgelesen werden.

<sup>1</sup> DIF als Abkürzung für Differential Item Functioning

**Kulturelle Unterschiede zwischen den Sprachgruppen**

- Mögliche generelle Fähigkeitsunterschiede oder unterschiedliche Fähigkeitsprofile (anderes Verhältnis der einzelnen Fähigkeiten zueinander) im Vergleich der Kulturen können vorhanden sein, die beispielsweise auf Unterschiede im Schulsystem zurückzuführen sein können. Auch die Strategien, wie entsprechende Aufgaben gelöst werden, können sich unterscheiden und sich auf die Ergebnisse auswirken.
- Unterschiede in Leistungs-Voraussetzungen (z.B. Motivation, Belastbarkeit, Ausdauer) können zu unterschiedlichen Resultaten beitragen.

**Unterschiedliche Repräsentativität der Stichproben für die Sprachgruppen**

- Die jeweils untersuchten Stichproben können verschiedene Ausschnitte aus der jeweiligen Sprachgruppe sein. Die geringe Zahl der französisch- und italienischsprachigen Teilnehmer macht dies wahrscheinlich – die Auswahl kann die "Spitze" oder das "Ende" der Leistungsrangreihe aller Maturanden der jeweiligen Sprachgruppe überrepräsentieren, etwa durch unterschiedliches "Wahlverhalten" aufgrund der vorhandenen Alternativen für Studienorte.

**Testbedingte Unterschiede**

- Die Testaufgaben können nach der Übertragung eine unterschiedliche Aufgabenschwierigkeit aufweisen, indem durch Satzstellung, Wortwahl, Kompliziertheit des Satzes etc. ein Unterschied auftritt. Die Lösungsschwierigkeit eines Items wird unter anderem auch von der Formulierung und dem Satzaufbau einer Fragestellung beeinflusst. Bereits geringe Änderungen innerhalb einer Sprache können zu unterschiedlichen Schwierigkeiten führen. Allerdings sind diese Differenzen nicht vorherzusehen, sondern können erst empirisch nachgewiesen werden.

Für das Jahr 2001 ist der Beschluss der Schweizerischen Hochschulkonferenz aus dem Jahre 1999 weiter bindend. Er beinhaltet,

1. einen Ausgleich nur bei sprachabhängigen Tests vorzunehmen. Dies sind fünf von neun Untertests. Der Sprachausgleich beschränkt sich demnach auf die eher sprachabhängigen Untertests Medizinisch-naturwissenschaftliches Grundverständnis, Quantitative und formale Probleme, Textverständnis, Fakten lernen sowie Diagramme und Tabellen;
2. den Ausgleich nur vorzunehmen, wenn Mittelwertunterschiede im entsprechenden Untertest vorhanden sind. Geprüft werden dazu die Abweichungen der jeweiligen Zielsprache von der deutschen Sprachgruppe.

Adaptationsbedingte Effekte sind praktisch nur in den fünf sprachabhängigen Untertests (Textverständnis, Diagramme und Tabellen, Fakten lernen, Quantitative und formale Probleme, Medizinisch-naturwissenschaftliches Grundverständnis) zu erwarten, während die vier weitgehend sprachunabhängigen Untertests (Muster zuordnen, Figuren lernen, Schlauchfiguren, Konzentriertes und sorgfältiges Arbeiten) davon kaum betroffen sind.

Spricht man von DIF, so muss dieser klar von systematischen Mittelwertsdifferenzen, welche alle Aufgaben eines Untertests gleichermassen betreffen, abgegrenzt werden. Die Ergebnisse des letzten Jahres haben gezeigt, dass sich die Sprachgruppen in manchen Untertestscores signifikant unterscheiden. Differenzen kommen nicht nur durch einzelne Items zustande, sondern auch durch eine systematische Verschiebung der Lösungsschwierigkeit aller Items eines Untertests. Es ist unwahrscheinlich, dass diese systematischen Unterschiede testbedingt sind. Testbedingte Unterschiede als sprachliche Besonderheiten zeigen sich in spezifischen Abweichungen einzelner Items. Systematische

Unterschiede sind dagegen mit grösserer Wahrscheinlichkeit den Faktoren „Kulturunterschiede“ und „unterschiedliche Repräsentativität“ zuzurechnen.

Es ergibt sich für das Ausgleichsverfahren die nachfolgende Zielstellung:

***Das Risiko, dass testbedingte Unterschiede die Sprachunterschiede systematisch beeinflussen, soll verringert werden. Andererseits soll kein Ausgleich von Unterschieden aufgrund unterschiedlicher Zusammensetzung der Stichproben und der kulturellen Unterschiede erfolgen.***

#### 8.4.1 Sprachvergleich für die Untertests

Wie in den Vorjahren treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Untertests gleichermassen auf. 1998 und 1999 erzielte die italienischsprachige Gruppe noch den besten Wert bei Quantitative und formale Probleme, was seit der Durchführung 2000 nicht mehr der Fall ist. In diesem Jahr hat die italienischsprachige Gruppe im Untertest Fakten lernen erstmals gleiche Werte erreicht wie die deutschsprachige Gruppe (Abbildung 38).

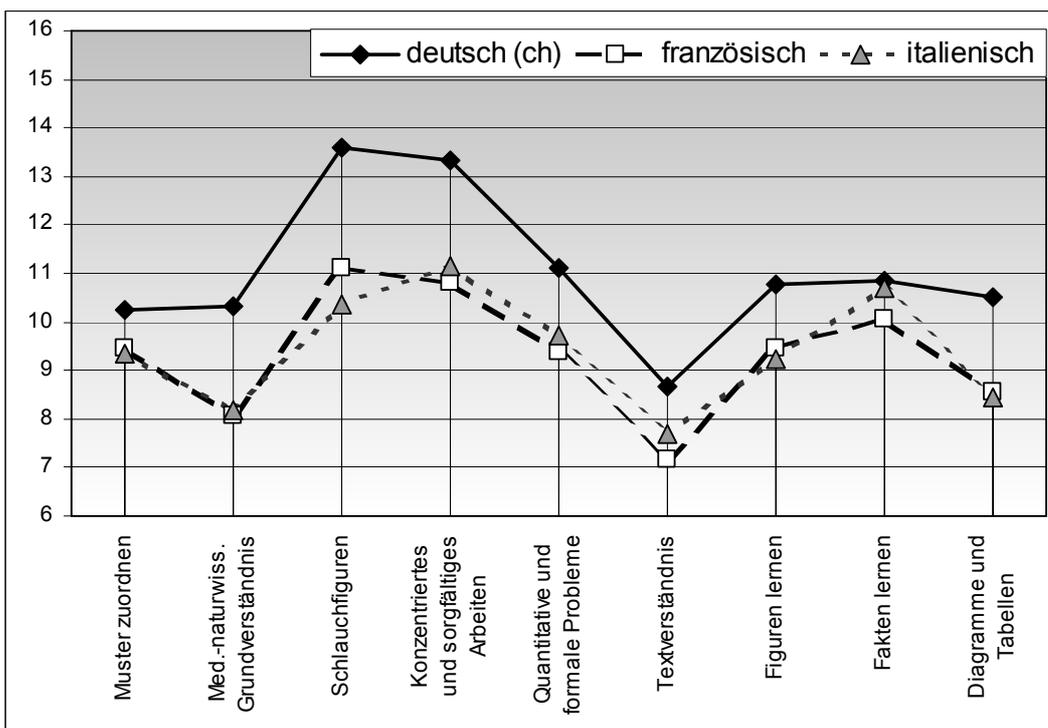


Abbildung 38: Mittelwerte der Untertests für die Sprachgruppen 2001 (unkorrigiert)

Die Abbildung 39 zeigt einen Vergleich zwischen deutsch- und französischsprachigen Testteilnehmern über die vergangenen Jahre. Angegeben sind die Punktdifferenzen (Mittelwerte), wobei positive Werte höhere Punktzahlen für die deutschsprachige Gruppe bedeuten.

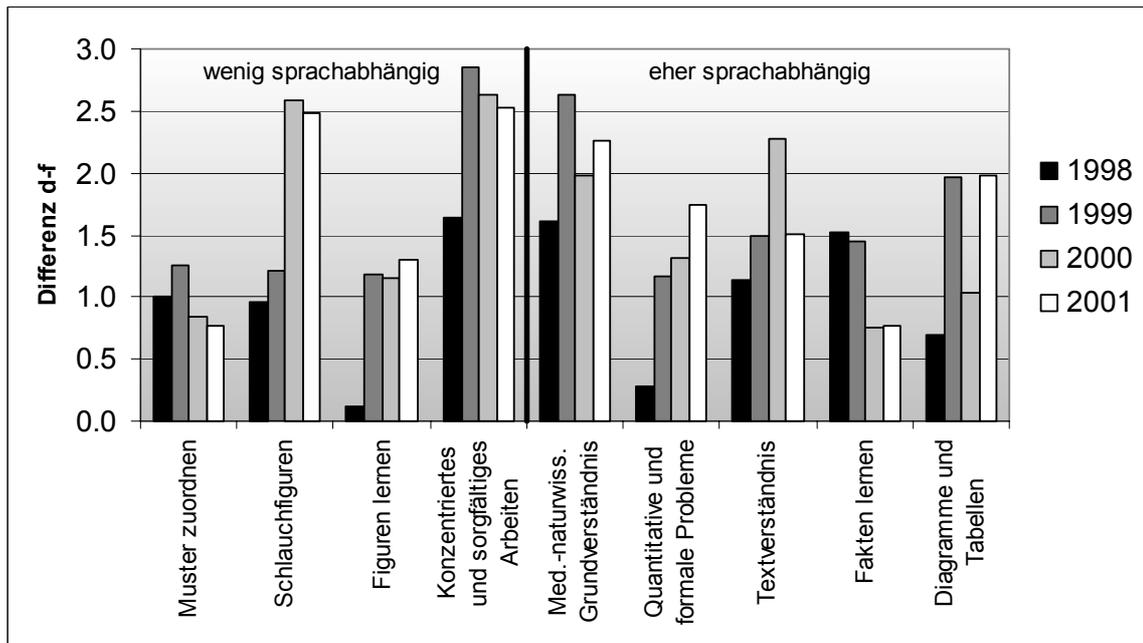


Abbildung 39: Mittelwertsdifferenzen deutsch-französisch (unkorrigiert), 1998 bis 2001

Es zeigen sich in fast allen Untertests über die Jahre zum Teil erhebliche Schwankungen. Einerseits handelt es sich bei den Untertests mit den kleinsten Differenzen (Muster zuordnen und Figuren lernen) um nicht-sprachabhängige Untertests, andererseits sind in den ebenfalls sprachunabhängigen Untertests Schlauchfiguren und Konzentriertes und sorgfältiges Arbeiten die grössten Unterschiede zu beobachten.

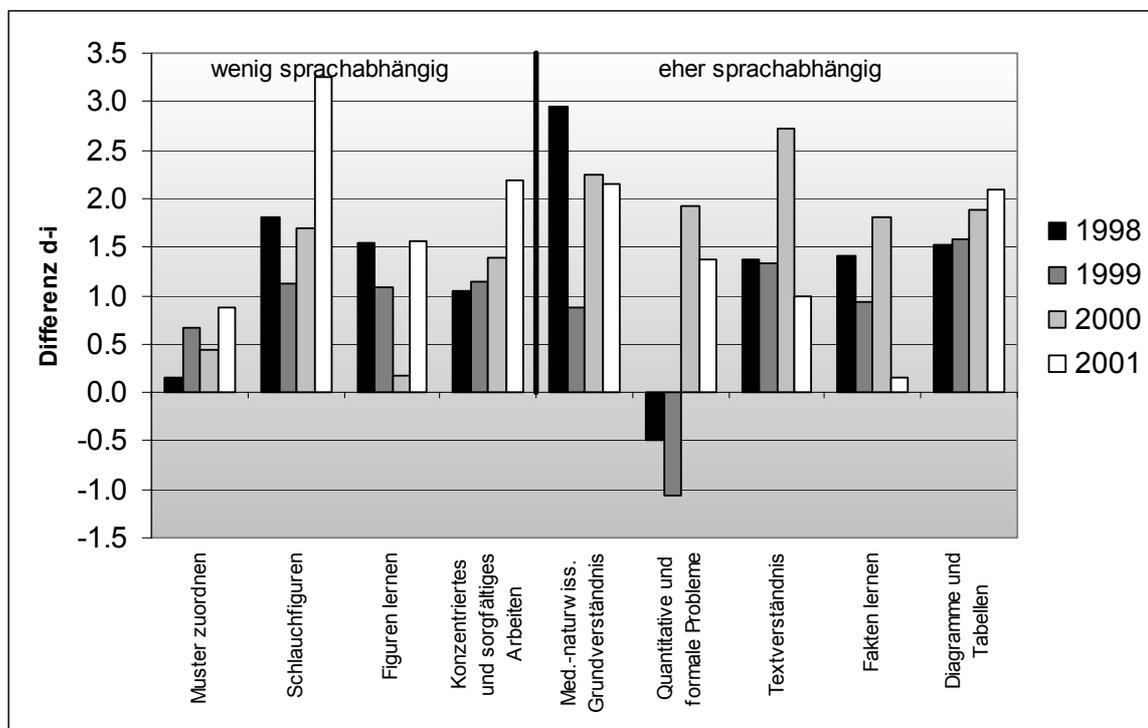


Abbildung 40: Mittelwertsdifferenzen deutsch-italienisch (unkorrigiert), 1998 bis 2001

Abbildung 40 zeigt den entsprechenden Vergleich für die italienischsprachige Gruppe. Auffällig ist die Veränderung im Untertest Quantitative und formale Probleme (1998 und 1999 vs. 2000 und 2001), sowie die hohe Differenz im Untertest Schlauchfiguren.

Andererseits wiederholt sich die grosse Differenz im Untertest Textverständnis aus dem Jahr 2000 in diesem Jahr nicht. Im Untertest Fakten lernen besteht erstmals kaum ein Unterschied zur deutschsprachigen Gruppe.

Weil die Differenzen und ihre Schwankungen zwischen den einzelnen Jahren sowohl in sprachabhängigen wie auch weniger sprachabhängigen Untertests auftreten, sind grössere adaptionsbedingte Textunterschiede als massgebliche Ursache dafür sehr unwahrscheinlich. Durch das Korrekturverfahren werden vorhandene testbedingte Differenzen zwischen den Sprachversionen ausgeglichen – die Erwartung, dass danach alle Sprachgruppen die gleichen Mittelwerte erreichen, ist dennoch gering. Wegen des hierarchischen Faktormodells mit hohem Generalfaktoranteil an der Varianzaufklärung (siehe Seite 120) werden die Unterschiede wahrscheinlich mindestens so hoch wie in den weniger sprachabhängigen Untertests bleiben.

		n	Mittelwert	Stand.abw.	Minimum	Maximum
Testwert ** <i>D &gt; (F,I)</i>	deutsch	773	100.89	9.86	72	128
	französisch	65	93.58	8.37	70	116
	italienisch	42	94.17	9.10	79	117
Muster zuordnen ** <i>homogen</i>	deutsch	773	10.23	2.83	2	19
	französisch	65	9.46	2.74	4	15
	italienisch	42	9.36	2.38	3	14
Med.-naturwiss. Grundverständnis ** <i>D &gt; (F,I)</i>	deutsch	773	10.34	3.44	1	19
	französisch	65	8.23	3.10	1	15
	italienisch	42	8.60	2.84	4	15
Schlauchfiguren ** <i>D &gt; (F,I)</i>	deutsch	773	13.61	3.67	3	20
	französisch	65	11.12	3.44	4	18
	italienisch	42	10.36	3.16	4	16
Quantitative und formale Probleme ** <i>D &gt; F</i>	deutsch	773	11.11	3.86	1	20
	französisch	65	9.50	4.07	1	19
	italienisch	42	10.10	3.80	2	19
Textverständnis ** <i>D &gt; F</i>	deutsch	773	8.68	3.46	0	18
	französisch	65	7.64	2.96	1	13
	italienisch	42	7.69	3.47	2	18
Figuren lernen ** <i>D &gt; I</i>	deutsch	773	10.77	3.44	1	20
	französisch	65	9.48	3.47	2	19
	italienisch	42	9.21	3.42	4	18
Fakten lernen <i>homogen</i>	deutsch	773	10.85	3.71	2	20
	französisch	65	10.08	3.23	2	18
	italienisch	42	10.69	3.19	5	17
Diagramme und Tabellen ** <i>D &gt; (F,I)</i>	deutsch	773	10.52	3.07	3	20
	französisch	65	8.70	2.73	2	14
	italienisch	42	9.01	3.23	2	18
Konzentriertes und sorgfältiges Arbeiten ** <i>D &gt; (F,I)</i>	deutsch	773	13.36	3.72	4	20
	französisch	65	10.82	2.88	5	18
	italienisch	42	11.14	2.75	6	18

**Tabelle 53:** Testwert und Punktwerte der Untertests für die Sprachgruppen. \*\*: signifikanter Einfluss des Faktors "Sprache" bei varianzanalytischer Prüfung; daneben sind signifikante Unterschiede bei multiplem Mittelwertvergleich angegeben. In französischer und italienischer Sprachgruppe korrigierte Punktwerte. Statistische Tests siehe Anhang

Die Mittelwertsdifferenzen für jeden Untertest **nach** dem Sprachausgleich sind der Tabelle 53 zu entnehmen. Die varianzanalytische Prüfung zeigt einzig für den Untertest Fakten lernen keinen signifikanten Einfluss des Faktors Sprache. Die Ergebnisse multipler Mittelwertvergleiche können ebenfalls der ersten Spalte entnommen werden.

## 8.4.2 Darstellung des Korrekturverfahrens

Die Identifikation und der Ausgleich testbedingter Unterschiede geschieht wie folgt:

- 1) Sprachausgleiche werden nur in den sprachabhängigen Untertests Medizinisch-naturwissenschaftliches Grundverständnis, Quantitative und formale Probleme, Textverständnis, Fakten lernen sowie Diagramme und Tabellen vorgenommen. Dass Unterschiede in den nichtsprachlichen Tests testbedingt sind (und ausgeglichen werden müssen), ist unwahrscheinlich. Die Aufgaben selber erfordern praktisch keine Sprachkompetenzen und die sprachspezifischen Anforderungen der Instruktion scheinen so einfach, dass Verständnisunterschiede unwahrscheinlich sind.
- 2) Damit ein Sprachausgleich in einem bestimmten sprachabhängigen Untertest erfolgt, müssen sich die entsprechenden mittleren Punktzahlen signifikant unterscheiden (Vergleich deutsch - französisch, beziehungsweise deutsch – italienisch mittels t-Test). Wenn keine Mittelwert- und Varianzunterschiede vorhanden sind, trägt der Untertest auch nicht zu Mittelwert-Unterschieden beim Testwert bei. Eine Korrektur allfälliger DIF-Items (die vorhanden sein können), würde zusätzliche Unterschiede produzieren.
- 3) Es werden jene Items ausgeglichen, welche über eine systematische Differenz hinaus DIF (Differential Item Functioning) aufweisen. Für diese Items ist der Verdacht am ehesten begründet, dass es sich um testbedingte Differenzen handelt. Der Ausgleich wird nicht symmetrisch vorgenommen, da es darum geht, eine Benachteiligung der Vergleichsgruppe gegenüber der Referenzgruppe zu vermeiden.
- 4) Als Sprachausgleich wird den französisch- und italienischsprachigen Probanden die Differenz zur entsprechenden Itemschwierigkeit in der deutschsprachigen Version gutgeschrieben, falls sie das Item nicht gelöst haben. Um jede Benachteiligung auszuschliessen, wird nicht nur die Differenz zur Regressionsgeraden ausgeglichen. Die Lage dieser Regressionsgeraden wird auch von allen DIF-Items mitbestimmt. Man kann das so interpretieren, dass für DIF-Items bei Nichtlösung zumindest der Schwierigkeitsunterschied zur Referenzgruppe ausgeglichen wird.

Eine generelle Gutschrift des Mittelwertunterschiedes pro Untertest für die französisch- und italienischsprachige Gruppe könnte für einzelne Personen zu mehr Punkten pro Untertest führen als es theoretisch gibt, und sich auf die Gewichtung der Untertests negativ auswirken. Aus diesem Grund wird nicht mehr als ein Punkt pro Aufgabe vergeben, und nur Personen welche das betreffende Item nicht gelöst haben erhalten den Bonus.

### 8.4.2.1 Welche Untertests ausgleichen?

Die folgenden Tabellen zeigen die Sprachvergleiche für die eher sprachabhängigen Untertests. Referenz ist die deutschsprachige Testversion. Für die französischsprachige Sprachform (Tabelle 54) zeigen sich in vier der fünf sprachabhängigen Untertests signifikante Differenzen. Der Untertest Fakten lernen muss nicht ausgeglichen werden.

Untertest	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	773	10.34	3.44	<b>.000</b>	Ja
	f	65	8.08	3.11		
Quantitative und formale Probleme	d	773	11.11	3.86	<b>.001</b>	Ja
	f	65	9.37	4.11		
Textverständnis	d	773	8.68	3.46	<b>.001</b>	Ja
	f	65	7.17	3.09		
Fakten lernen	d	773	10.85	3.71	.105	Nein
	f	65	10.08	3.23		
Diagramme und Tabellen	d	773	10.52	3.07	<b>.000</b>	Ja
	f	65	8.54	2.75		

Tabelle 54: Signifikanzprüfung deutsche vs. französische Sprachform

Die italienische Sprachform (Tabelle 55) weist in drei sprachabhängigen Untertests signifikante Unterschiede auf. Hier müssen Textverständnis und Fakten lernen nicht ausgeglichen werden.

Untertest	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	773	10.34	3.44	<b>.000</b>	Ja
	i	42	8.19	2.97		
Quantitative und formale Probleme	d	773	11.11	3.86	<b>.026</b>	Ja
	i	42	9.74	3.90		
Textverständnis	d	773	8.68	3.46	.073	Nein
	i	42	7.69	3.47		
Fakten lernen	d	773	10.85	3.71	.788	Nein
	i	42	10.69	3.19		
Diagramme und Tabellen	d	773	10.52	3.07	<b>.000</b>	Ja
	i	42	8.43	3.41		

Tabelle 55: Signifikanzprüfung deutsche vs. italienische Sprachform

Mit der Bestimmung der auszugleichenden Untertests kann noch nichts darüber ausgesagt werden, ob es sich bei den Differenzen um reale oder um adaptationsbedingte Unterschiede handelt. Zu diesem Zweck ist eine Analyse der Items für die betreffenden Untertests notwendig.

#### **8.4.2.2 Identifikation von DIF-Items und Bestimmung des Korrekturwertes**

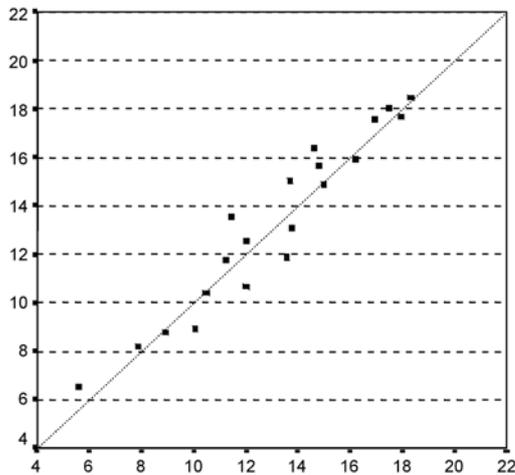
Eine Möglichkeit zur Identifikation von DIF, auch bei kleinen Stichprobengrößen, ist die Methode "Delta-Plot". Bei diesem Verfahren werden die zu vergleichenden Itemschwierigkeiten z-standardisiert und anschliessend in " $\Delta$ -Werte" transformiert. Die Transformation erfolgt über die Formel

$$\Delta = 13 - 4z$$

Dies bedeutet, dass  $\Delta$  einen Mittelwert von 13 und eine Standardabweichung von 4 aufweist. Hohe Werte stehen für "schwierige" (von weniger Probanden gelöste) Items.

Die aus den Delta-Werten abgeleitete Regressionsgerade  $Y = AX + B$  beschreibt die Beziehung zwischen den interessierenden Sprachversionen. Eine graphische Darstellung der Delta-Werte (Delta-Plot) würde im Idealfall eine ellipsenförmige Anordnung von Punkten

entlang der Diagonalen ergeben (Abbildung 41). Dies würde bedeuten, dass sowohl die Itemschwierigkeiten wie auch deren Reihenfolge in beiden Sprachversionen vergleichbar sind.



**Abbildung 41.** Beispiel für einen Delta-Plot

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

Auffällige Abweichungen einzelner Punkte von dieser Regressionsgeraden hingegen weisen auf Items hin, welche zusätzlich zu einer eventuellen systematischen Verschiebung spezielle Eigenschaften aufweisen. Bei einer parallelen Verschiebung der Regressionsgeraden bleibt trotz einer Veränderung der absoluten Schwierigkeiten die „Schwierigkeits-Rangfolge“ der Items erhalten. Von der Regressionsgeraden abweichende Items stimmen aber in der untersuchten Sprachgruppe bezüglich der "Schwierigkeitshierarchie" nicht mit der Referenzgruppe überein (sie sind im Vergleich zu den anderen Items zu schwer oder zu leicht ausgefallen). Die Ursache solcher Differenzen kann eher testbedingt, also beispielsweise in der Übersetzung vermutet werden. Betroffene Items müssen nicht in jedem Fall als "schlecht" übersetzt betrachtet werden. Denkbar ist etwa auch, dass in der Fragestellung Konzepte enthalten sein könnten, welche in den Sprachgruppen unterschiedliche Bekanntheitsgrade aufweisen.

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

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

berechnet, wobei unter A die Steigung und B der Achsenschnittpunkt der Regressionsgeraden zu verstehen ist,  $X_i$  bezeichnet den Delta-Wert der Referenzgruppe,  $Y_i$  denjenigen der zu vergleichenden Gruppe.

In der Folge werden die auszugleichenden Untertests auf Itemebene analysiert. Als "kritische" Distanz zur Regressionsgeraden wird üblicherweise ein DIF-Wert von 1.5 angenommen (Longford, Holland & Thayer, 1993).

In der Folge sollen innerhalb der auszugleichenden Untertests die kritischen Items identifiziert werden. Da, wie oben erwähnt, keine symmetrische DIF-Behandlung verwendet wird, ist die Richtung der Abweichung der Regressionsgeraden von der Diagonalen massgebend. Es werden also jene Items ausgeglichen, welche mindestens 1.5 Punkte von der Regressionsgeraden in entgegengesetzter Richtung zur Diagonalen abweichen. Als Sprachausgleich wird der benachteiligten Sprachgruppe die Differenz der betreffenden

Itemschwierigkeit zur deutschen Itemschwierigkeit gutgeschrieben (also die Differenz der Lösungswahrscheinlichkeit). Allerdings betrifft dies nur Personen, welche das fragliche Item nicht korrekt beantwortet haben.

### 8.4.2.3 Medizinisch-naturwissenschaftliches Grundverständnis (MNGV)

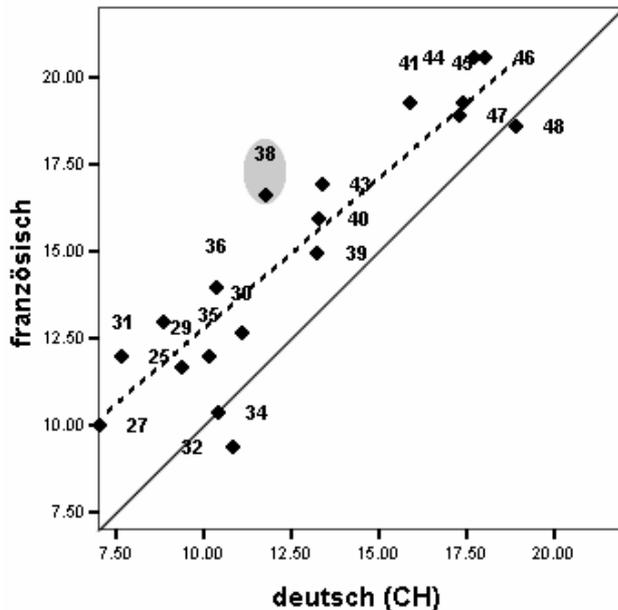


Abbildung 42: Delta-Plot für MNGV (deutsch-französisch)

Abbildung 42 zeigt den Delta-Plot für den Untertest MNGV für den Vergleich zwischen deutsch- und französischsprachiger Version, die Abbildung 43 denjenigen zwischen deutsch- und italienischsprachiger Version.

Die französische Fassung zeigt eine deutliche Parallelverschiebung der Schwierigkeiten. Die Items 48, 34 und 32 liegen fast auf der Diagonalen und sind damit eigentlich durch die Übersetzung im Vergleich mit den anderen Items zu leicht geraten. Da kein symmetrisches Vorgehen bei der Korrektur angewendet

wird, bleiben diese Effekte erhalten. Andererseits liegt das Item 38 mehr als 1.5 Punkte in Richtung einer Benachteiligung der französischen Sprachgruppe von der Regressionsgeraden entfernt und wird somit ausgeglichen.

Die italienischsprachige Version (Abbildung 43) weist im Vergleich zur deutschsprachigen Fassung keine klare Parallelverschiebung auf, die Steigung der Regressionsgeraden beträgt hier 0.68. Vor allem einige „einfache“ Aufgaben wurden möglicherweise in der italienischsprachigen Version durch die Adaptation „erschwert“. Im einzelnen betrifft dies hier die Items 32 und 31, welche durch das Ausgleichsverfahren bis auf die Diagonale „herunterkorrigiert“ werden.

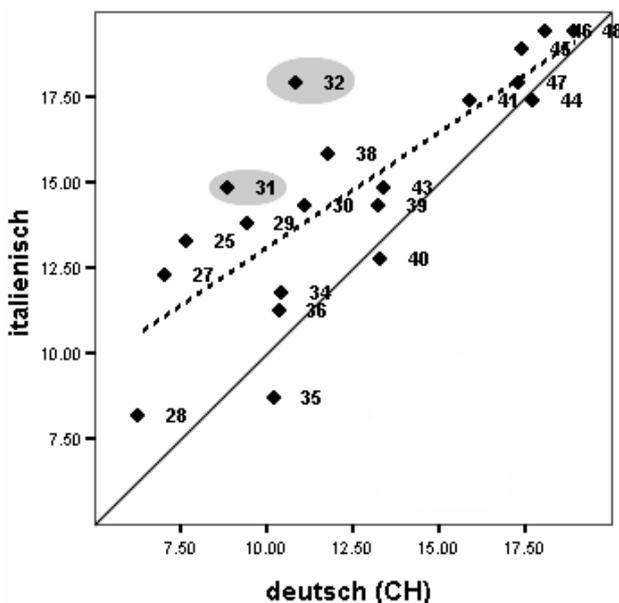


Abbildung 43: Delta-Plot für MNGV (deutsch-italienisch)

Die genauen Werte für beide Sprachvergleiche sind aus Tabelle 56 ersichtlich. Die ersten drei Spalten zeigen die Itemschwierigkeiten, danach folgen die in Delta-Werte transformierten Angaben. DIF weisen jene Items auf, bei welchen die absolute Differenz der Deltawerte 1.5 oder mehr beträgt. Da kein symmetrisches Verfahren gewählt wurde, sind nur negative Werte relevant. Der Bonus für die französische oder italienische Sprachgruppe berechnet sich aus der Differenz zwischen den ursprünglichen Itemschwierigkeiten.

Maximal können in diesem Untertest den französisch- oder italienischsprachigen Teilnehmern, welche die entsprechenden Items nicht gelöst haben, 0.23 (f) respektive 0.61 (i) Punkte gutgeschrieben werden.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
25	0.74	0.54	0.48	7.67	12.00	13.33	-0.95	-1.50		
27	0.77	0.63	0.52	7.03	10.01	12.31	0.13	-1.01		
28	0.81	0.58	0.71	6.25	11.01	8.22	-1.12	1.95		
29	0.66	0.55	0.45	9.42	11.67	13.84	0.45	-0.94		
30	0.58	0.51	0.43	11.11	12.66	14.36	0.81	-0.41		
31	0.68	0.49	0.40	8.86	12.99	14.87	-0.92	<b>-2.10</b>		0.28
32	0.59	0.66	0.26	10.84	9.35	17.93	3.12	<b>-3.54</b>		0.33
34	0.61	0.62	0.55	10.42	10.34	11.80	2.10	1.31		
35	0.62	0.54	0.69	10.20	12.00	8.73	0.71	3.73		
36	0.61	0.45	0.57	10.36	13.98	11.29	-0.68	1.71		
38	0.55	0.32	0.36	11.78	16.62	15.89	<b>-1.76</b>	-1.31	0.23	
39	0.48	0.40	0.43	13.25	14.97	14.36	0.46	0.78		
40	0.48	0.35	0.50	13.31	15.96	12.82	-0.26	2.08		
41	0.36	0.20	0.29	15.89	19.26	17.42	-1.07	-0.28		
43	0.47	0.31	0.40	13.39	16.95	14.87	-0.95	0.44		
44	0.27	0.14	0.29	17.72	20.58	17.42	-0.87	0.74		
45	0.29	0.20	0.21	17.42	19.26	18.96	-0.07	-0.70		
46	0.26	0.14	0.19	18.06	20.58	19.47	-0.65	-0.76		
47	0.29	0.22	0.26	17.31	18.93	17.93	0.11	0.09		
48	0.22	0.23	0.19	18.92	18.60	19.47	1.41	-0.28		

Tabelle 56: DIF-Analyse Medizinisch-naturwissenschaftliches Grundverständnis

#### 8.4.2.4 Quantitative und formale Probleme

Die folgenden Abbildungen zeigen die Delta-Plots für den Untertest Quantitative und formale Probleme. Auch hier werden sowohl die französischsprachige wie auch die italienische Version ausgeglichen.

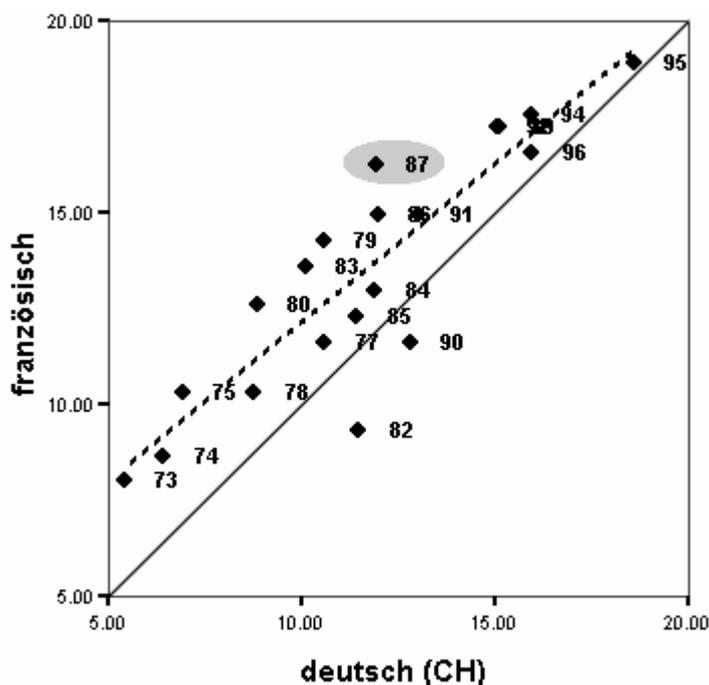


Abbildung 44: Delta-Plot für Quantitative und formale Probleme (deutsch-französisch)

In der französischen Version fällt Item 87 auf. Die Items 90 und vor allem 82 wären eher zu leicht ausgefallen, werden aber nicht ausgeglichen. Die übrigen Aufgaben liegen eng an der Regressionsgeraden.

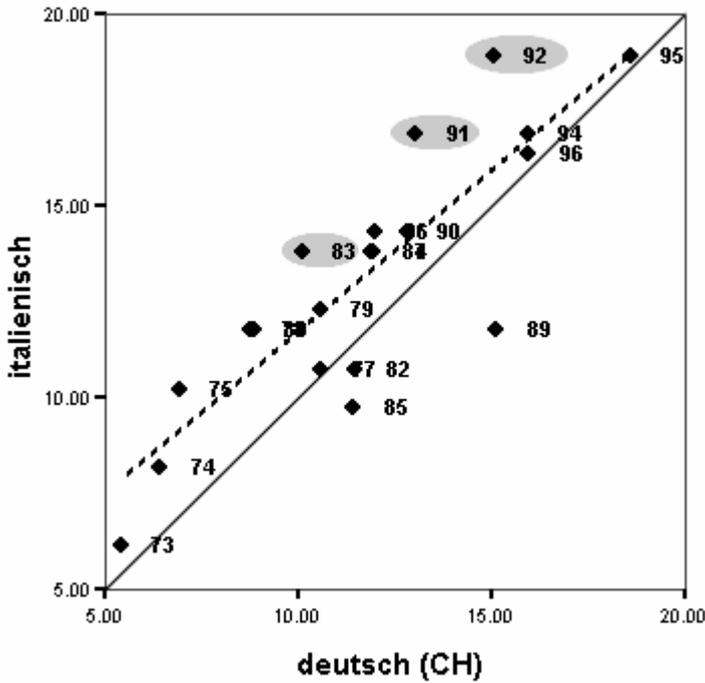


Abbildung 45: Delta-Plot für Quantitative und formale Probleme (deutsch-italienisch)

Die Verschiebung der Regressionsgeraden in der italienischen Fassung stimmt praktisch mit derjenigen der französischen Version überein (Achsen Schnittpunkte 3.9 bzw. 3.3, Steigung 0.83 bzw. 0.84). Allerdings liegen hier drei Items ausserhalb der kritischen Grenze, nämlich die Aufgaben 83, 91 und 92. Die Items 82, 85 und 89 wurden durch die Übersetzung eher „vereinfacht“.

Die ausführlichen (gerundeten) Werte sind Tabelle 57 zu entnehmen. In diesem Untertest sind maximal 0.2 (f) respektive 0.54 (i) Punkte gutzuschreiben.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
73	0.85	0.72	0.81	5.39	8.03	6.18	0.24	1.27		
74	0.80	0.69	0.71	6.42	8.69	8.22	0.39	0.37		
75	0.77	0.62	0.62	6.92	10.34	10.27	-0.57	-0.87		
77	0.60	0.55	0.60	10.59	11.67	10.78	0.75	1.10		
78	0.69	0.62	0.55	8.75	10.34	11.80	0.60	-0.86		
79	0.61	0.43	0.52	10.56	14.31	12.31	-1.31	-0.09		
80	0.69	0.51	0.55	8.84	12.66	11.80	-1.13	-0.81		
82	0.56	0.66	0.60	11.47	9.35	10.78	3.09	1.68		
83	0.63	0.46	0.45	10.09	13.65	13.84	-1.10	<b>-1.56</b>		0.18
84	0.54	0.49	0.45	11.89	12.99	13.84	0.56	-0.40		
85	0.57	0.52	0.64	11.42	12.33	9.76	0.77	2.42		
86	0.54	0.40	0.43	11.97	14.97	14.36	-0.92	-0.74		
87	0.54	0.34	0.45	11.92	16.29	13.84	<b>-1.97</b>	-0.38	0.20	
89	0.39	0.29	0.55	15.11	17.28	11.80	-0.70	3.24		
90	0.50	0.55	0.43	12.81	11.67	14.36	2.16	-0.20		
91	0.49	0.40	0.31	13.03	14.97	16.91	-0.24	<b>-2.01</b>		0.18
92	0.40	0.29	0.21	15.03	17.28	18.96	-0.75	<b>-2.29</b>		0.18
94	0.36	0.28	0.31	15.92	17.61	16.91	-0.44	-0.15		
95	0.23	0.22	0.21	18.58	18.93	18.96	0.24	0.00		
96	0.35	0.32	0.33	15.95	16.62	16.40	0.34	0.26		

Tabelle 57: DIF-Analyse Quantitative und formale Probleme

### 8.4.2.5 Textverständnis

Für den Untertest Textverständnis kann im Vergleich der deutschsprachigen mit der französischen Version (Abbildung 46) eine Parallelverschiebung der Regressionsgeraden beobachtet werden (Achsenabschnitt = 0.88, Steigung = 1.07). Für die italienischsprachige Fassung wird hier kein Sprachausgleich vorgenommen, da sich die Punktzahlen nicht signifikant unterscheiden.

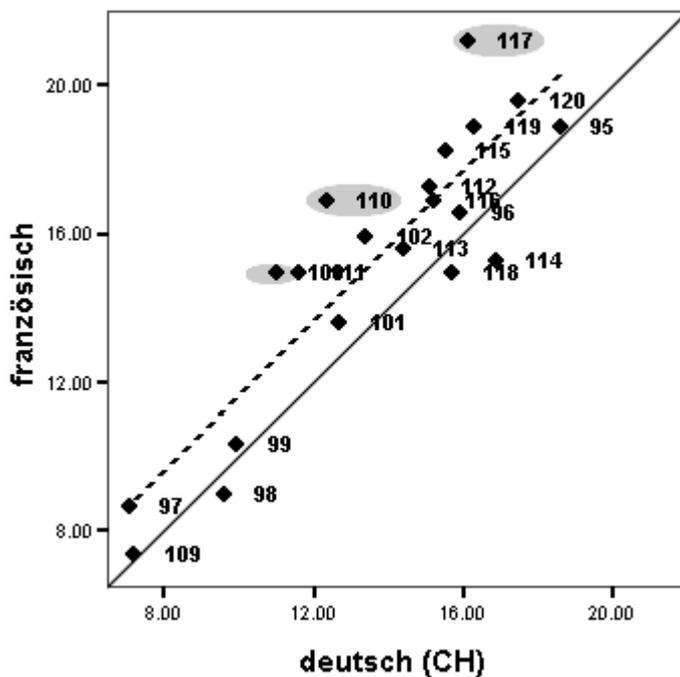


Abbildung 46: Delta-Plot für Textverständnis (deutsch-französisch)

In der französischen Version sind drei Items auszugleichen (100, 110, 117).

Verschiedene Items im unteren und mittleren Schwierigkeitsbereich deuten auf eine Vereinfachung hin (insbesondere 118, 114).

Der französischen Sprachgruppe können hier maximal 0.63 Punkte gutgeschrieben werden (Tabelle 58).

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
97	0.77	0.69	0.74	7.09	8.69	7.71	-0.16			
98	0.65	0.68	0.62	9.61	9.02	10.27	1.46			
99	0.63	0.62	0.52	9.95	10.34	12.31	0.80			
100	0.58	0.40	0.43	11.03	14.97	14.36	-1.56		0.18	
101	0.51	0.46	0.52	12.70	13.65	12.31	0.55			
102	0.47	0.35	0.36	13.39	15.96	15.89	-0.52			
109	0.76	0.75	0.81	7.23	7.37	6.18	0.85			
110	0.52	0.31	0.29	12.36	16.95	17.42	-1.94		0.21	
111	0.56	0.40	0.48	11.61	14.97	13.33	-1.14			
112	0.39	0.29	0.29	15.11	17.28	17.42	-0.16			
113	0.43	0.37	0.40	14.39	15.63	14.87	0.44			
114	0.31	0.38	0.29	16.86	15.30	17.42	2.47			
115	0.37	0.25	0.29	15.53	18.27	17.42	-0.54			
116	0.39	0.31	0.48	15.22	16.95	13.33	0.14			
117	0.35	0.11	0.31	16.11	21.24	16.91	-2.14		0.24	
118	0.36	0.40	0.36	15.72	14.97	15.89	1.86			
119	0.34	0.22	0.26	16.31	18.93	17.93	-0.42			
120	0.28	0.18	0.26	17.50	19.59	17.93	0.00			

Tabelle 58: DIF-Analyse Textverständnis

### 8.4.2.6 Fakten lernen

Bei diesem sprachabhängigen Untertest wurden weder in der französischen noch in der italienischen Fassung signifikante Differenzen zur deutschen Version festgestellt. Es wird hier kein Ausgleich durchgeführt.

### 8.4.2.7 Diagramme und Tabellen

Der Untertest Diagramme und Tabellen ist für beide Sprachgruppen zu analysieren. Für die französische Version kann auch hier eine saubere Parallelverschiebung beobachtet werden (Abbildung 47).

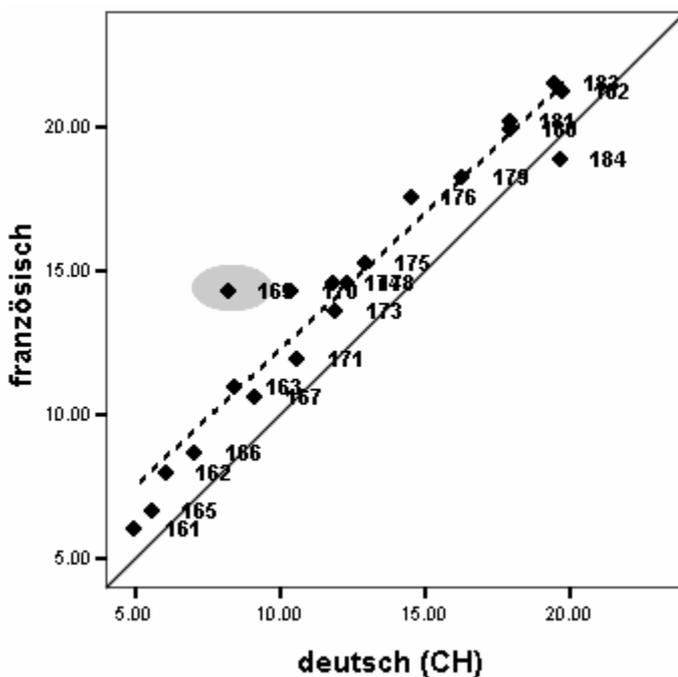


Abbildung 47: Delta-Plot für Diagramme und Tabellen (deutsch-französisch)

Die Items in der französischen Version liegen eng an der Regressionsgeraden.

Einziger ausgleichender „Ausreisser“ ist die Aufgabe 169.

Item 184 wurde durch die Adaptation vereinfacht.

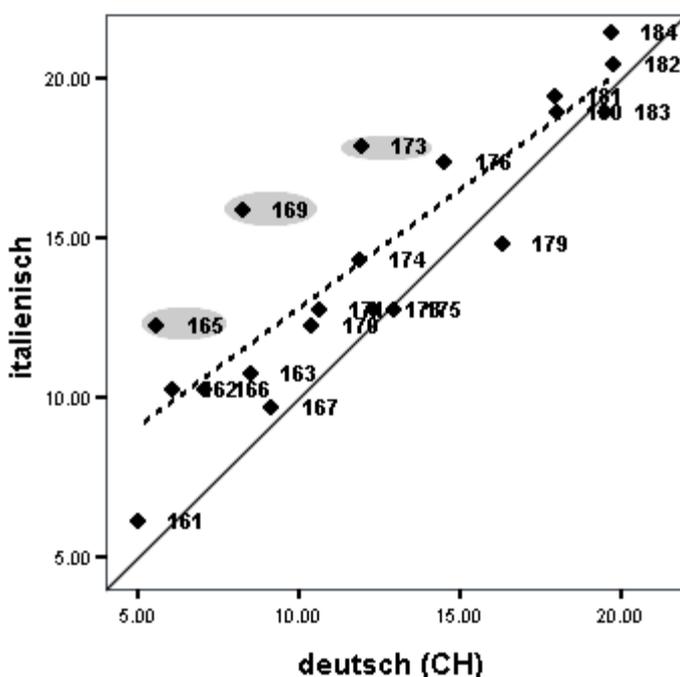


Abbildung 48: Delta-Plot für Diagramme und Tabellen (deutsch-italienisch)

Die italienischsprachige Fassung weist eine Tendenz zur Vereinfachung der schwierigen Items auf, die aber hauptsächlich auf die starke Erschwerung der drei ausgleichenden Items 165, 169 und 173 zurückzuführen ist.

Tabelle 59 zeigt die detaillierten Angaben für beide Sprachversionen. In diesem Untertest sind maximal 0.28 (f) respektive 0.95 (i) Punkte gutzuschreiben.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
161	0.87	0.82	0.81	4.98	6.05	6.18	1.05	2.34		
162	0.82	0.72	0.62	6.06	8.03	10.27	0.35	-0.30		
163	0.70	0.58	0.60	8.48	11.01	10.78	-0.14	0.73		
165	0.84	0.78	0.52	5.59	6.71	12.31	0.99	<b>-2.22</b>		0.31
166	0.77	0.69	0.62	7.06	8.69	10.27	0.56	0.30		
167	0.67	0.60	0.64	9.14	10.68	9.76	0.55	1.95		
169	0.71	0.43	0.36	8.23	14.31	15.89	<b>-2.72</b>	<b>-3.52</b>	0.28	0.36
170	0.61	0.43	0.52	10.39	14.31	12.31	-1.23	0.65		
171	0.60	0.54	0.50	10.61	12.00	12.82	0.61	0.37		
173	0.54	0.46	0.26	11.92	13.65	17.93	0.30	<b>-2.95</b>		0.28
174	0.54	0.42	0.43	11.86	14.64	14.36	-0.45	-0.12		
175	0.49	0.38	0.50	12.95	15.30	12.82	-0.19	1.76		
176	0.42	0.28	0.29	14.53	17.61	17.42	-0.78	-0.99		
178	0.52	0.42	0.50	12.34	14.64	12.82	-0.13	1.40		
179	0.34	0.25	0.40	16.31	18.27	14.87	-0.04	2.13		
180	0.26	0.17	0.21	17.97	19.92	18.96	-0.09	-0.16		
181	0.26	0.15	0.19	17.95	20.25	19.47	-0.35	-0.59		
182	0.18	0.11	0.14	19.78	21.24	20.49	0.19	-0.32		
183	0.19	0.09	0.21	19.47	21.57	18.96	-0.26	0.73		
184	0.18	0.22	0.10	19.67	18.93	21.51	1.79	-1.20		

Tabelle 59: DIF-Analyse Diagramme und Tabellen

#### 8.4.2.8 Effekte der Korrektur

Die durch die Ausgleichsprozedur resultierenden Bonuspunkte können Tabelle 60 entnommen werden. Durchschnittlich werden den französischsprachigen Teilnehmern 0.91 Punkte, den italienischsprachigen Teilnehmern 1.36 Punkte gutgeschrieben. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Der Bonus in der französischen Gruppe beträgt also 0 oder 1 Punkt, in der italienischsprachigen Gruppe 0 bis 2 Punkte.

		n	Minimum	Maximum	m	s
französisch	Med.-naturwiss. Grundv.	65	.00	.23	.16	.11
	Quant. und form. Probleme	65	.00	.20	.13	.10
	Textverständnis	65	.00	.63	.47	.18
	Diagramme und Tabellen	65	.00	.28	.16	.14
	<b>GESAMT</b>	<b>65</b>	<b>.20</b>	<b>1.34</b>	<b>.91</b>	<b>.32</b>
italienisch	Med.-naturwiss. Grundv.	42	.00	.61	.41	.25
	Quant. und form. Probleme	42	.00	.54	.36	.17
	Diagramme und Tabellen	42	.00	.95	.59	.30
	<b>GESAMT</b>	<b>42</b>	<b>.18</b>	<b>2.10</b>	<b>1.36</b>	<b>.50</b>

Tabelle 60: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen

Gegenüber früheren Testdurchführungen bedeutet diese Bonusverteilung eine Reduktion der Gutschriften. Zumindest teilweise kann dies wohl mit der zunehmenden Erfahrung der Übersetzer (insbesondere das Ziel des Erhalts der Textschwierigkeiten) und damit qualitativ besseren Übersetzungen begründet werden.

### 8.4.2.9 Evaluation: Gleichbehandlung der Sprachgruppen

Wie in den Vorjahren zeigen sich bezüglich der Zulassungsquoten zur Humanmedizin signifikante Unterschiede zwischen den Sprachgruppen. Bedingt durch die geringe Stichprobengröße und die hohe Zulassungsquote in der Veterinärmedizin lassen sich dort keine Differenzen nachweisen.

Tabelle 61 zeigt die Zulassungsquoten, differenziert nach Sprachgruppen und Disziplinen.

Disziplin	Sprache	Warteliste	zugelassen	Total
Humanmedizin	deutsch	103 15.9%	546 84.1%	649 100.0%
	französisch	11 44.0%	14 56.0%	25 100.0%
	italienisch	14 40.0%	21 60.0%	35 100.0%
	total	128 18.1%	581 81.9%	709 100.0%
Veterinärmedizin	deutsch	4 3.2%	120 96.8%	124 100.0%
	französisch	2 5.0%	38 95.0%	40 100.0%
	italienisch	-	7 100.0%	7 100.0%
	total	6 3.5%	165 96.5%	171 100.0%

Tabelle 61: Zulassung nach Sprache des Tests

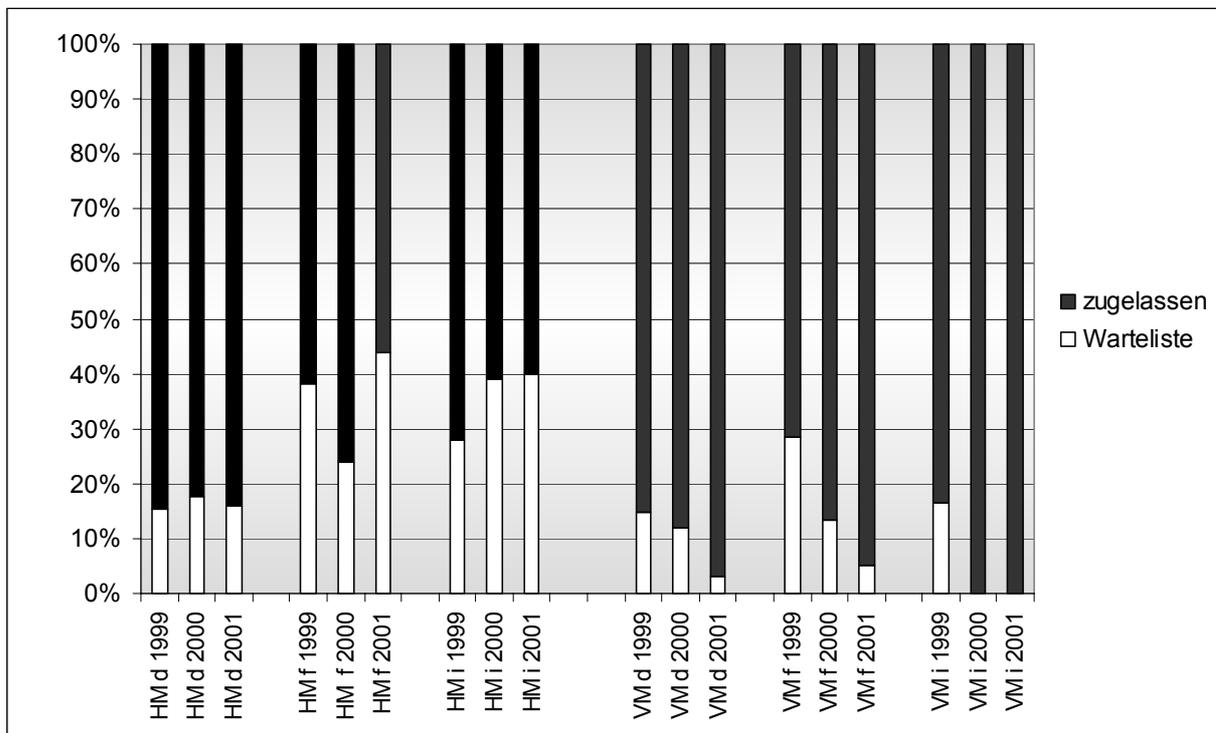
Die Signifikanzprüfung für die Zulassungsunterschiede kann Tabelle 62 entnommen werden.

Disziplin		Wert	df	Asymp. Sig. (2seitig)
Humanmedizin	Pearson Chiquadrat	24.862	2	.000
	Likelihood Ratio	20.282	2	.000
Veterinärmedizin	Pearson Chiquadrat	.547	2	.761
	Likelihood Ratio	.763	2	.683

Tabelle 62: Prüfung auf Verteilungsunterschiede für Sprachgruppen getrennt nach Disziplinen: signifikanter Unterschied für Humanmedizin, kein signifikanter Unterschied für Veterinärmedizin (aber geringe Stichprobengröße)

Abbildung 49 zeigt die grafische Aufbereitung der Zulassungsquoten. Für die französische Gruppe muss nach der Annäherung an die deutschsprachigen Bewerber von 2000 wieder eine Verringerung des Zulassungsanteiles konstatiert werden. Die Zulassungsquote ist aber

vergleichbar mit derjenigen aus dem Jahr 1999. Wir vermuten, dass dies vor allem ein Repräsentativitätsproblem der Bewerbergruppen ist. In der italienischen Gruppe bleiben die Verhältnisse gegenüber dem Vorjahr stabil.



**Abbildung 49:** Zulassungsquoten nach Disziplin Humanmedizin (HM) und Veterinärmedizin (VM) und Sprache d(Deutsch), f(Französisch), i(Italienisch) 1999 bis 2001

Es muss auch in diesem Jahr sichergestellt sein, dass die Unterschiede nicht am Test selber liegen. Dies kann durch eine Analyse des Verhältnisses zwischen sprachabhängigen und wenig sprachabhängigen Untertests recht gut belegt werden. In den **wenig sprachabhängigen Tests** wären keine oder geringere Unterschiede zwischen den Sprachgruppen vorhanden, wenn lediglich die Übersetzung des Tests für diesen Effekt verantwortlich wäre. Da dort Unterschiede aber mindestens ebenso deutlich sind wie die in den sprachabhängigen Tests, müssen diese Unterschiede eher auf „tatsächliche“ Unterschiede zwischen den Personen zurückgeführt werden. Die unterschiedliche Repräsentativität der Bewerbergruppen und kulturelle Unterschiede wurden als mögliche Ursachen genannt.

Eine Evaluation der Korrektur zeigt, dass auch in diesem Jahr durch die Korrektur testbedingter Unterschiede eher eine Überkompensation als nur die Vermeidung einer Benachteiligung stattgefunden hat. Aufgrund des Faktormodells, des Generalfaktors mit sehr hohem Varianzanteil, ist zu erwarten, dass die Fähigkeiten in den Bereichen sprachabhängiger und wenig sprachabhängiger Tests hoch korrelieren und allfällige Abweichungen der Sprachgruppen in beiden Bereichen etwa gleich sein müssen, wenn es sich um keine testbedingten Unterschiede handelt. Damit Vergleiche möglich sind, wurden die Mittelwerte für die französisch- und italienischsprachige Gruppe in die Verteilung der deutschsprachigen Gruppe transformiert:

$$\text{Delta } z = \frac{M_{\text{Sprachgruppe}} - M_{\text{deutsch}}}{S_{\text{deutsch}}}$$

$m$  ist der Mittelwert und  $s$  die Standardabweichung der entsprechenden Sprachgruppe. Die resultierenden Werte „Delta  $z$ “ drücken die Abweichung der jeweiligen Sprachgruppe von der deutschen Sprachgruppe in Standardabweichungen aus. 0.5 hiesse, dass die betroffene Sprachgruppe um eine halbe Standardabweichung schlechter ist als die deutschsprachige Referenzgruppe. Abbildung 50 zeigt den Testwertvergleich für die französischsprachige Gruppe.

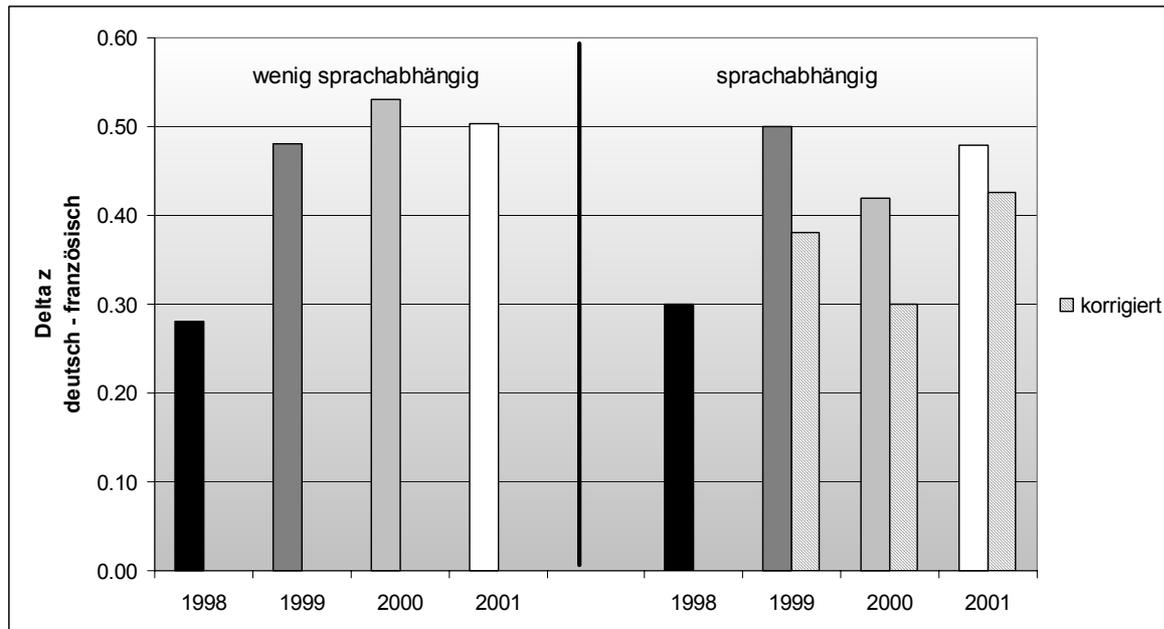


Abbildung 50: Abweichung (Delta  $z$ ) für den Testwert zwischen deutscher und französischer Sprachgruppe. Für die sprachabhängigen Untertests korrigierte und unkorrigierte Werte

Die schraffierten Balken zeigen die korrigierten Werte, die Differenz zwischen schraffiertem Balken und links anliegendem Balken entspricht dem Korrektoreffekt. 1999 schnitt die französischsprachige Gruppe in den sprachabhängigen Untertests geringfügig schlechter ab als in den wenig sprachabhängigen. Durch den Sprachausgleich wurde diese Differenz korrigiert und sogar überkompensiert. In den letzten beiden Jahren wurde in den wenig sprachabhängigen Untertests eine stärkere Differenz zwischen französischer und deutscher Gruppe beobachtet als in den sprachabhängigen Untertests vor der Korrektur. Durch die zusätzliche Kompensation darf davon ausgegangen werden, dass die französischsprachige Gruppe nicht durch die Adaptation des Test benachteiligt wird.

Abbildung 51 zeigt die entsprechende Grafik für die italienische Sprachgruppe. Hier wurde im Jahr 2000 durch die Korrektur eine grössere Differenz in den sprachabhängigen Tests vermieden. In den Jahren 1999 und 2001 liegt die Differenz dort bereits bei unkorrigierten Punktzahlen unter derjenigen in den wenig sprachabhängigen Untertests. Auch hier ist somit sichergestellt, dass keine Benachteiligung der italienischsprachigen Teilnehmer aufgrund textintensiver Aufgaben erfolgt. Für die diesjährige Durchführung muss in den wenig sprachabhängigen Untertests eine deutliche Verschlechterung der italienischsprachigen Gruppe im Vergleich zu den deutschsprachigen Bewerbern konstatiert werden.

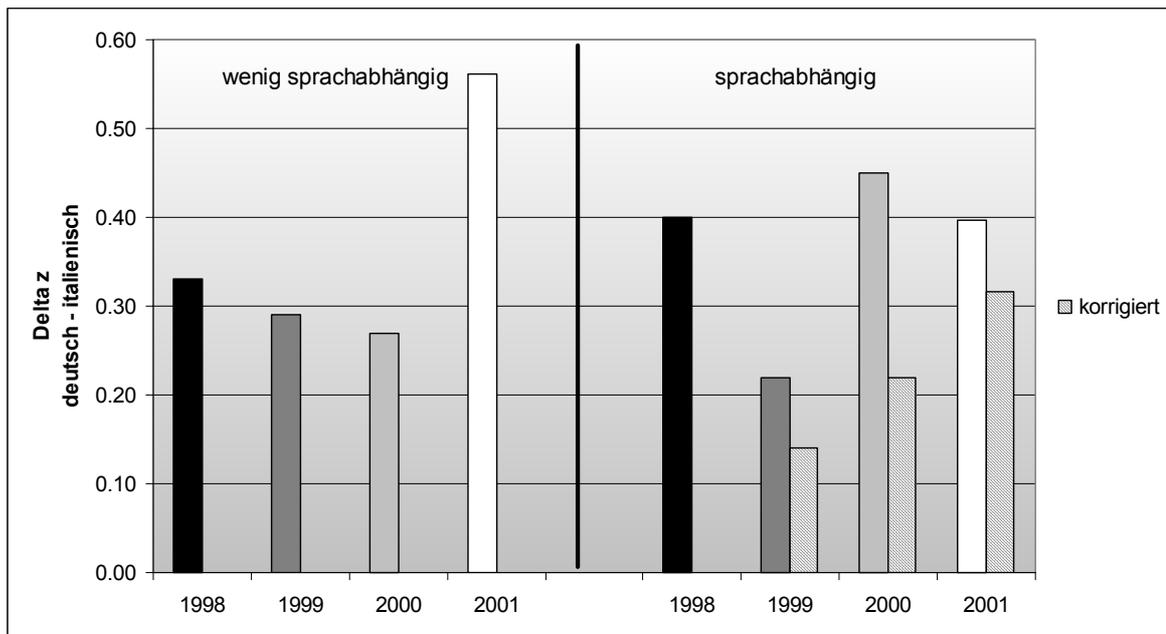


Abbildung 51: Abweichung (Delta z) für den Testwert zwischen deutscher und italienischer Sprachgruppe. Für die sprachabhängigen Untertests korrigierte und unkorrigierte Werte

1998 wurde für beide Sprachversionen kein aufgabenspezifisches Korrekturverfahren verwendet. An dessen Stelle wurde eine separate Normierung des Testwerts vorgenommen, wodurch ebenfalls Benachteiligung der Sprachminderheiten vermieden wird, allerdings ohne Differenzierung zwischen testbedingten und „realen“ Differenzen in den Testleistungen.

**8.4.3 Leistungen in identischen Testaufgaben 1998 – 2001**

Zum Vergleich der Leistungen über die Jahre hinweg werden weiterhin „Einstreuaufgaben“ verwendet. Dabei handelt es sich um Items, welche sich im Vorjahr aufgrund ihrer Kennwerte bewährt haben. Die Werte in diesen Einstreuaufgaben lassen sich demnach zwischen zwei benachbarten Jahren vergleichen. Dabei ist zu beachten, dass ein anderer Aufgabenkontext und andere Positionen der Aufgaben im Test bereits für Unterschiede verantwortlich sein können. Da diese Positionen in den Sprachgruppen aber gleich sind, können die tatsächlichen Unterschiede der Sprachgruppen besser abgeschätzt werden.

Nach einer signifikanten Verbesserung der deutschen Sprachgruppe im letzten Jahr können dieses Jahr keine signifikanten Unterschiede in den Einstreuaufgaben beobachtet werden (Tabelle 63). Die Mittelwerte liegen allerdings in allen Gruppen höher als im Vorjahr.

	Jahr	n	Mittelwert	Standardabw.	Sig.
deutsch	2000	703	14.49	3.60	.28
	2001	773	14.70	3.94	
französisch	2000	69	12.71	3.26	.76
	2001	65	12.90	3.54	
italienisch	2000	29	12.07	3.48	.77
	2001	42	12.36	4.44	

Tabelle 63: Verteilungskennwerte identischer Einstreuaufgaben 2000 und 2001

Abbildung 52 zeigt den Differenzverlauf für die Einstreuaufgaben. Die französischsprachige Gruppe zeigte 1999 eine Verschlechterung gegenüber 1998, die italienischsprachige Gruppe schnitt im Jahr 2000 schlechter ab als 1999. Ansonsten kann durchwegs eine Verbesserung der Leistungen beobachtet werden. Aufgrund der Stichprobengrößen ist allerdings nur die Verbesserung der deutschsprachigen Gruppe im Jahr 2000 gegenüber 1999 signifikant.

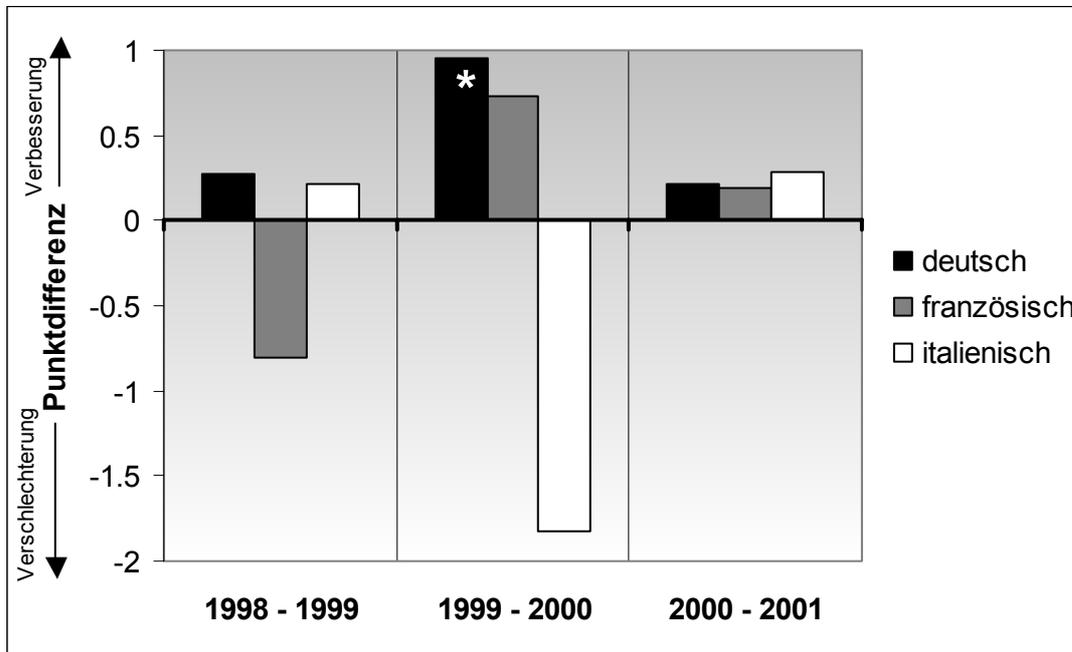


Abbildung 52: Verlauf der Leistungen in identischen Testaufgaben 1998 bis 2001 (Differenzverlauf); \*= signifikante Differenz

Die tendenzielle Verbesserung der deutschsprachigen Testteilnehmer über die Jahre ist neben den Vergleichswerten aus Deutschland ein weiteres Indiz für die weiter oben erwähnte „Erschwerung“ des Tests. Ein jahrgangsabhängiger realer Leistungsrückgang kann aufgrund dieser Daten nicht vermutet werden. Die rückläufigen Summen der Absolutpunktwerte sind auf die unterschiedliche Aufgabenmischung zurückzuführen.

## 8.5 Vergleichbarkeit der Testlokale

Die Durchführungsbedingungen in den einzelnen Testorten sind standardisiert. Dazu gehören beispielsweise die genaue Einhaltung von Zeiten, das wörtlich genaue Vorlesen von Instruktionen und einheitliches Reagieren auf mögliche Fragen und Probleme. Aus den Testprotokollen ergeben sich keine Hinweise, die als Einschränkungen dieser Chancengleichheit zu bewerten wären.<sup>1</sup> Die Einhaltung lässt sich teilweise auch anhand der Daten überprüfen. Wenn systematische Unterschiede zwischen Testlokalen bzw. Testorten auftreten würden, könnte dies auch an unterschiedlichen Durchführungs-Bedingungen liegen (beispielsweise wenn tatsächlich mehr oder auch weniger Zeit für die Bearbeitung einer Aufgabe eingeräumt worden wäre).

Die Varianzanalyse (Tabelle 64) zeigt, dass in diesem Jahr keine Unterschiede zwischen den Testlokalen vorhanden sind. Daraus ergeben sich keine Hinweise auf unterschiedliche Bedingungen.

<sup>1</sup> Die Zeiten von Beginn und Ende jedes Untertests sind im Protokoll zu notieren; Abweichungen müssen in jedem Falle ebenfalls protokolliert werden, sofern welche auftreten.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testwert	<i>Zwischen Gruppen</i>	838.483	10	83.848	.861	.570
	<i>Innerhalb Gruppen</i>	74205.498	762	97.383		
	<i>Total</i>	75043.982	772			
Muster zuordnen	<i>Zwischen Gruppen</i>	101.833	10	10.183	1.278	.239
	<i>Innerhalb Gruppen</i>	6072.785	762	7.970		
	<i>Total</i>	6174.618	772			
Med.-naturwiss. Grundverständnis	<i>Zwischen Gruppen</i>	36.730	10	3.673	.308	.979
	<i>Innerhalb Gruppen</i>	9077.818	762	11.913		
	<i>Total</i>	9114.549	772			
Schlauchfiguren	<i>Zwischen Gruppen</i>	200.488	10	20.049	1.497	.136
	<i>Innerhalb Gruppen</i>	10203.305	762	13.390		
	<i>Total</i>	10403.793	772			
Quant. und formale Probleme	<i>Zwischen Gruppen</i>	142.991	10	14.299	.959	.478
	<i>Innerhalb Gruppen</i>	11365.097	762	14.915		
	<i>Total</i>	11508.088	772			
Textverständnis	<i>Zwischen Gruppen</i>	66.089	10	6.609	.547	.857
	<i>Innerhalb Gruppen</i>	9199.057	762	12.072		
	<i>Total</i>	9265.146	772			
Figuren lernen	<i>Zwischen Gruppen</i>	57.858	10	5.786	.485	.900
	<i>Innerhalb Gruppen</i>	9086.975	762	11.925		
	<i>Total</i>	9144.833	772			
Fakten lernen	<i>Zwischen Gruppen</i>	123.077	10	12.308	.894	.538
	<i>Innerhalb Gruppen</i>	10488.910	762	13.765		
	<i>Total</i>	10611.987	772			
Diagramme und Tabellen	<i>Zwischen Gruppen</i>	71.903	10	7.190	.759	.668
	<i>Innerhalb Gruppen</i>	7215.036	762	9.469		
	<i>Total</i>	7286.939	772			
Konzentr. u. sorgf. Arbeiten	<i>Zwischen Gruppen</i>	216.612	10	21.661	1.582	.107
	<i>Innerhalb Gruppen</i>	10434.511	762	13.694		
	<i>Total</i>	10651.123	772			

**Tabelle 64:** Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den Testlokalen

## 8.6 Vergleich der Geschlechter

Die Forderung nach Gleichbehandlung der Geschlechter konnte in den bisherigen Testdurchführungen jeweils als erfüllt betrachtet werden. Chancengleichheit bedeutet, dass bei gleicher Eignung die gleichen Chancen auf eine Zulassung bestehen müssen. Beide Teilgruppen sind unterschiedlich repräsentativ in der Bewerberkohorte vertreten – der Männeranteil sinkt absolut seit 1998 immer weiter ab. In Humanmedizin sind in diesem Jahr 61.5% Frauen, in Veterinärmedizin 79.5%. Die Ursachen dieser Abnahme des Interesses der Männer für ein Medizinstudium bedürfen einer gesonderten Diskussion (siehe Seite 74).

Was heisst nun Chancengleichheit? Es kann nun über 2 Jahre nachgewiesen werden, dass sich die mittleren Testwerte für Personen, welche die erste Vorprüfung bestehen, nicht zwischen Männern und Frauen unterscheiden. Wenn der Testwert Eignung erfasst, so haben gleich Geeignete tatsächlich die gleichen Chancen.

Die Erfolgswahrscheinlichkeit, die erste Vorprüfung zu bestehen, unterscheidet sich allerdings um 7% zugunsten der Männer. Wegen der Mittelwertgleichheit der Erfolgreichen ist es sehr wahrscheinlich, dass dieser Unterschied in einer unterschiedlichen Repräsentativität der Bewerbergruppen für die beiden Geschlechter begründet liegt.

Chancengleichheit bedeutet also nicht die Nivellierung von **tatsächlich vorhandenen** Unterschieden. Der Test muss die vorhandenen Realitäten unverzerrt abbilden.

Tabelle 65 zeigt eine Zusammenfassung der Testwerte geschichtet nach Disziplin, Sprache und Geschlecht. Aufgrund der geringen Stichprobengrössen in der französisch- und italienischsprachigen Gruppe (Humanmedizin) hat dort die Angabe von Mittelwerten und Standardabweichungen nur illustrativen Charakter. Für die Veterinärmedizin sind einzig die deutsch- und französischsprachigen Gruppen der Testteilnehmerinnen für einen Vergleich genügend gross.

		Humanmedizin			Veterinärmedizin		
		deutsch	französisch	italienisch	deutsch	französisch	italienisch
männlich	Mittelwert	<b>102.74</b>	<b>92.21</b>	<b>96.25</b>	<b>100.35</b>	<b>94.03</b>	<b>97.59</b>
	Stand.-abw.	9.93	7.58	9.95	9.11	6.21	7.20
	N	250	7	16	25	8	2
weiblich	Mittelwert	<b>100.47</b>	<b>93.99</b>	<b>93.57</b>	<b>97.98</b>	<b>93.44</b>	<b>88.22</b>
	Stand.-abw.	10.87	9.47	9.15	9.12	8.60	4.31
	N	399	18	19	99	32	5

**Tabelle 65:** Statistiken für den Testwert nach Geschlechtern und Disziplin: Werte für Gruppen unter 30 Personen (kursiv) nicht ausreichend für statistische Vergleiche

Aufgrund der geringen Stichprobengrössen kann nur im Vergleich der deutschsprachigen Bewerber zur Humanmedizin ein signifikantes Ergebnis erwartet werden. Dieses wird wie aus Tabelle 66 ersichtlich in diesem Jahr erstmals erreicht. Die Bewerber zur Humanmedizin erzielen signifikant höhere Resultate als die Bewerberinnen.

Sprache	Disziplin	Levene's Test Gleichheit der Varianzen			t-Test Gleichheit des Mittelwerts		
			F	Sig.	t	df	Sig.
deutsch	Humanmedizin	Varianzen homogen	0.75	.784	2.838	647	<b>.005</b>
	Veterinärmedizin	Varianzen homogen	.003	.959	1.162	122	.248
französisch	Humanmedizin	Varianzen homogen	.111	.742	-.444	23	.661
	Veterinärmedizin	Varianzen homogen	1.185	.283	.181	38	.857
italienisch	Humanmedizin	Varianzen homogen	.024	.878	.830	33	.412

**Tabelle 66:** Prüfung Varianzhomogenität und Mittelwert (t-Test) für Testwert zwischen Geschlechtern. Italienische Sprachgruppe Veterinärmedizin wg. Fallzahl nicht berechnet

Eine zweifaktorielle Varianzanalyse nach Disziplin und Geschlecht zeigt signifikante Einflüsse für beide Faktoren, nicht jedoch deren Wechselwirkung (Disziplin:  $F = 4.79$ ,  $p < 0.03$ ; Geschlecht:  $F = 11,0$ ,  $p \leq 0.001$ ; Wechselwirkung  $p \leq 0.99$ ).

Abbildung 53 zeigt den sprachunabhängigen Vergleich zwischen männlichen und weiblichen Bewerbern, aufgeteilt nach Disziplin.

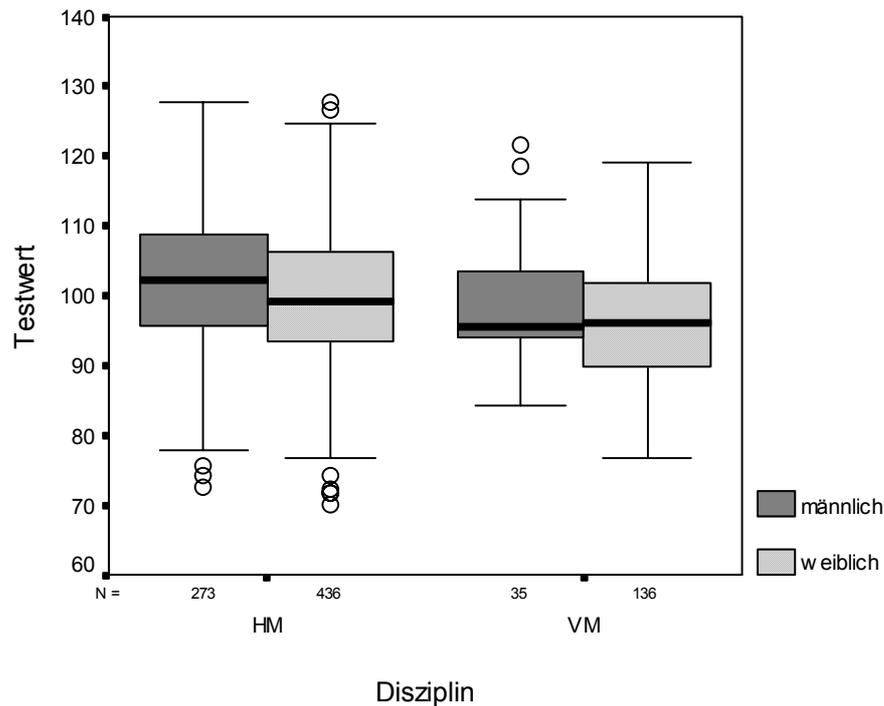


Abbildung 53: Boxplot für den Testwert nach Geschlecht und Disziplinen

Der Median der männlichen Bewerber für Veterinärmedizin liegt leicht unter demjenigen der Bewerberinnen. Durch die schiefe Verteilung ist dennoch der Mittelwert etwas höher. Abbildung 54 zeigt den Boxplot für die Testwerte nach Sprache und Geschlecht geschichtet. Die schiefen Verteilungen der französischen und italienischen Sprachgruppen kommen durch die geringen Teilnehmerzahlen zustande.

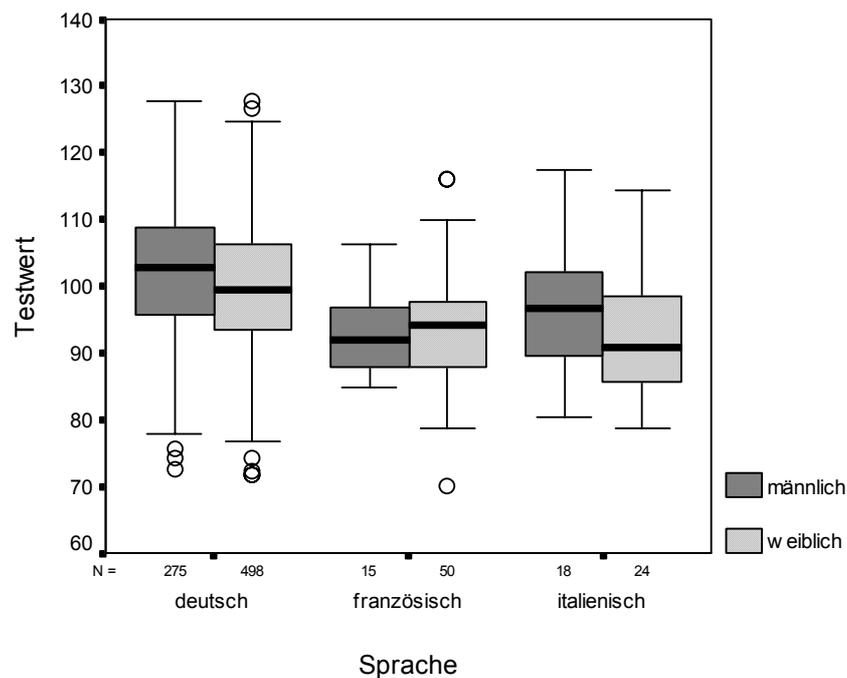


Abbildung 54: Boxplots für die Testwerte nach Geschlechtern getrennt für die drei Sprachgruppen

Wiederum zeigen sich Differenzen für die einzelnen Untertests zwischen den Geschlechtern. Bis auf den Untertest Muster zuordnen sind alle Unterschiede signifikant. In vier (davon drei signifikant) Untertests zeigen die Frauen bessere Leistungen, in fünf Untertests die Männer. Dabei bleibt die Tendenz der einzelnen Untertests über die Jahre stabil. Frauen erzielen tendenziell bessere Werte in den Untertests Muster zuordnen, Figuren lernen, Fakten lernen, Konzentriertes und sorgfältiges Arbeiten. Die männlichen Testteilnehmer erreichen in den übrigen Untertests höhere Punktzahlen.

Untertest	Geschlecht	m	s	Levene's Test Varianz			t-test Mittelwert		
					F	Sig.	t	df	Sig.
Muster zuordnen	männlich	10.05	2.89	Varianzhomogenität	.553	.457	-.670	878	.503
	weiblich	10.18	2.77						
Med.-naturwiss. Grundverst.	männlich	10.86	3.48	Varianzhomogenität	.277	.599	5.052	878	.000
	weiblich	9.64	3.39						
Schlauchfiguren	männlich	14.00	3.51	Keine Varianzhomogenität	6.132	.013	4.366	675.153	.000
	weiblich	12.88	3.81						
Quantitative und formale Probleme	männlich	12.54	3.57	Varianzhomogenität	1.731	.189	9.512	878	.000
	weiblich	10.04	3.81						
Textverständnis	männlich	9.33	3.65	Keine Varianzhomogenität	4.731	.030	5.022	573.194	.000
	weiblich	8.08	3.28						
Figuren lernen	männlich	10.11	3.33	Varianzhomogenität	1.835	.176	-3.116	878	.002
	weiblich	10.87	3.52						
Fakten lernen	männlich	10.20	3.59	Varianzhomogenität	.426	.514	-3.468	878	.001
	weiblich	11.09	3.66						
Diagramme und Tabellen	männlich	11.15	3.16	Varianzhomogenität	.592	.442	6.186	878	.000
	weiblich	9.80	3.02						
Konzent. u. sorgfältiges Arbeiten	männlich	12.73	3.67	Varianzhomogenität	.531	.466	-1.983	878	.048
	weiblich	13.25	3.71						

**Tabelle 67:** Mittelwerte (m) und Standardabweichungen (s) geschlechtsspezifisch und Ergebnisse der Prüfungen auf Varianzhomogenität und Mittelwertunterschiede

Abbildung 55 zeigt eine grafische Aufbereitung des Geschlechtervergleichs über die letzten drei Jahre. Die tendenzielle Richtung der Differenzen bleibt in allen Untertests konstant. Bei Muster zuordnen scheint es sich trotz der stabilen Differenzrichtung um den am wenigsten geschlechtsabhängigen Untertest zu handeln. Hier wurden in keiner Testdurchführung signifikante Unterschiede festgestellt.

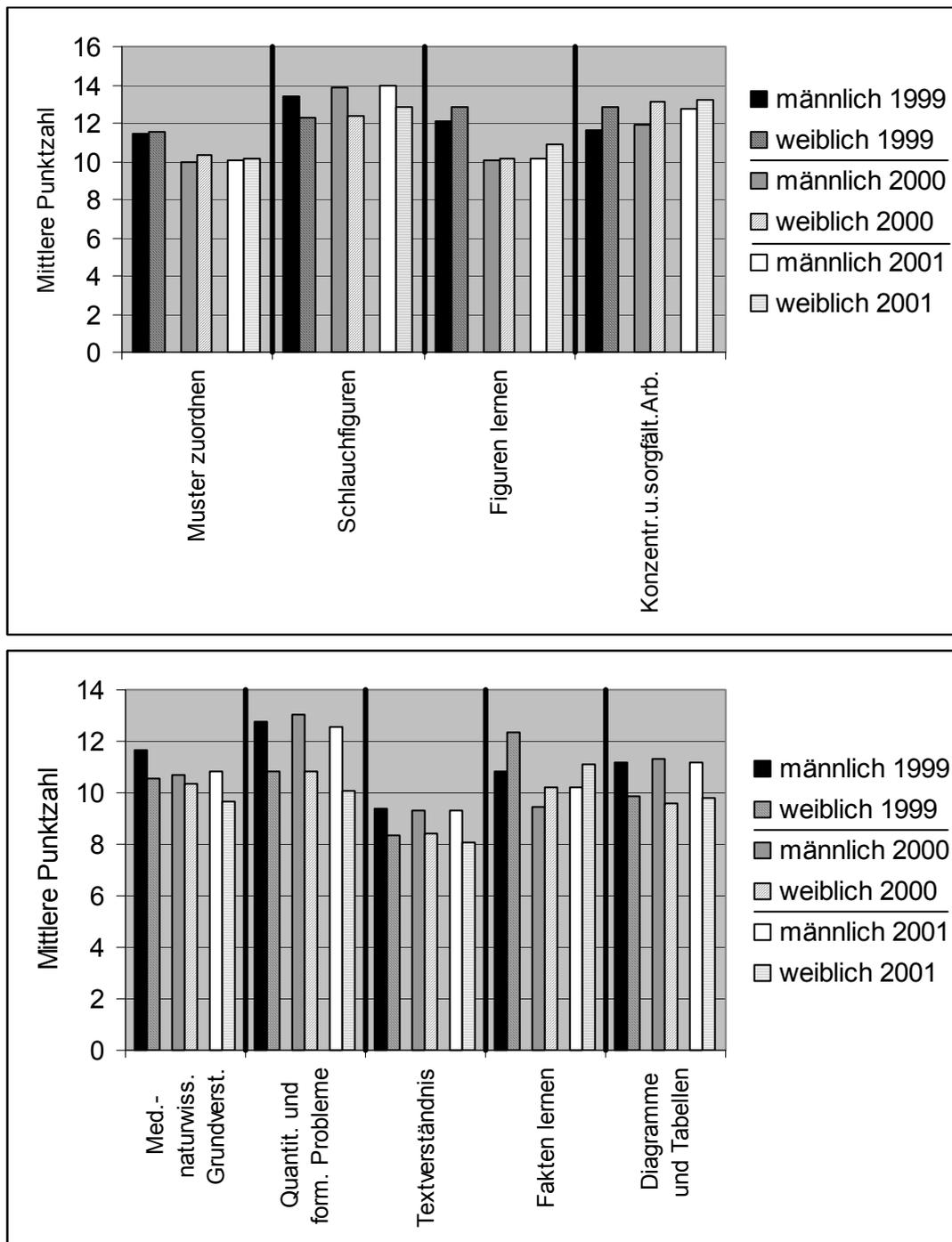


Abbildung 55: Mittelprofile für Punktwerte der Untertests geschlechtsspezifisch (getrennt nach sprachabhängigen und wenig sprachabhängigen Untertests)

Bleibt zu untersuchen, ob sich die Differenz im Testwert auch in einer unterschiedlichen Zulassungsquote niederschlägt. Die Angaben sind in Tabelle 68 dargestellt. Die Zulassungsquoten in Humanmedizin für beide Geschlechter unterscheiden sich nicht signifikant (Tabelle 69). Für die Veterinärmedizin ist die geringe Zahl männlicher Bewerber zu beachten, aber auch hier kann kein Unterschied in der Zulassungsquote festgestellt werden. Zusammenfassend kann auch in diesem Jahr davon ausgegangen werden, dass es trotz den diesjährigen (geringen) Unterschieden in den Testwerten zu keiner testbedingten Benachteiligung von Bewerberinnen gekommen ist.

Disziplin	Geschlecht	Warteliste	zugelassen	Total
Humanmedizin	männlich	41 15.0%	232 85.0%	273 100.0%
	weiblich	87 20.0%	349 80.0%	436 100.0%
	Total	128 18.1%	581 81.9%	709 100.0%
Veterinärmedizin	männlich	0	35 100.0%	35 100.0%
	weiblich	6 4.4%	130 95.6%	136 100.0%
	Total	6 3.5%	165 96.5%	171 100.0%

Tabelle 68: Zulassungsquoten geschlechtsspezifisch

Disziplin		Wert	df	Asymp. Sig. (2-seitig)
Humanmedizin	Pearson Chiquadrat	2.765	1	.096
	Likelihood Ratio	2.819	1	.093
Veterinärmedizin	Pearson Chiquadrat	1.600	1	.206
	Likelihood Ratio	2.804	1	.094

Tabelle 69: Prüfung Unterschiede der Zulassung – geschlechtsspezifisch: keine signifikanten Unterschiede bei Humanmedizin und Veterinärmedizin

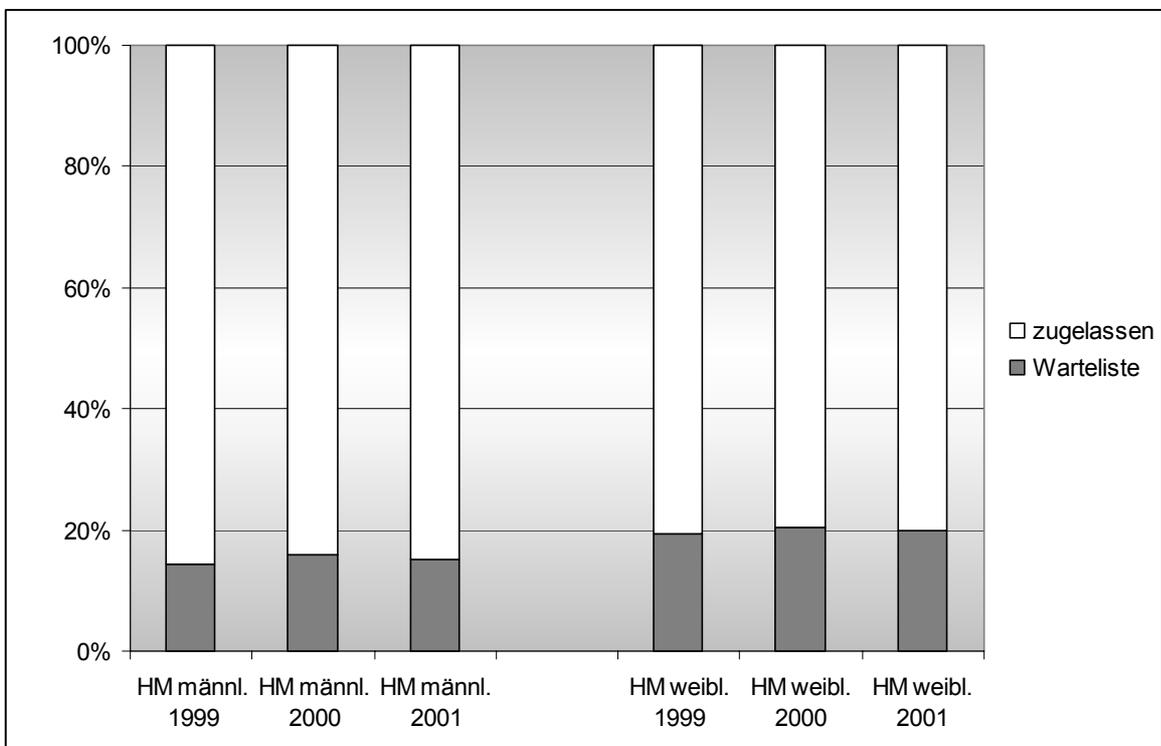
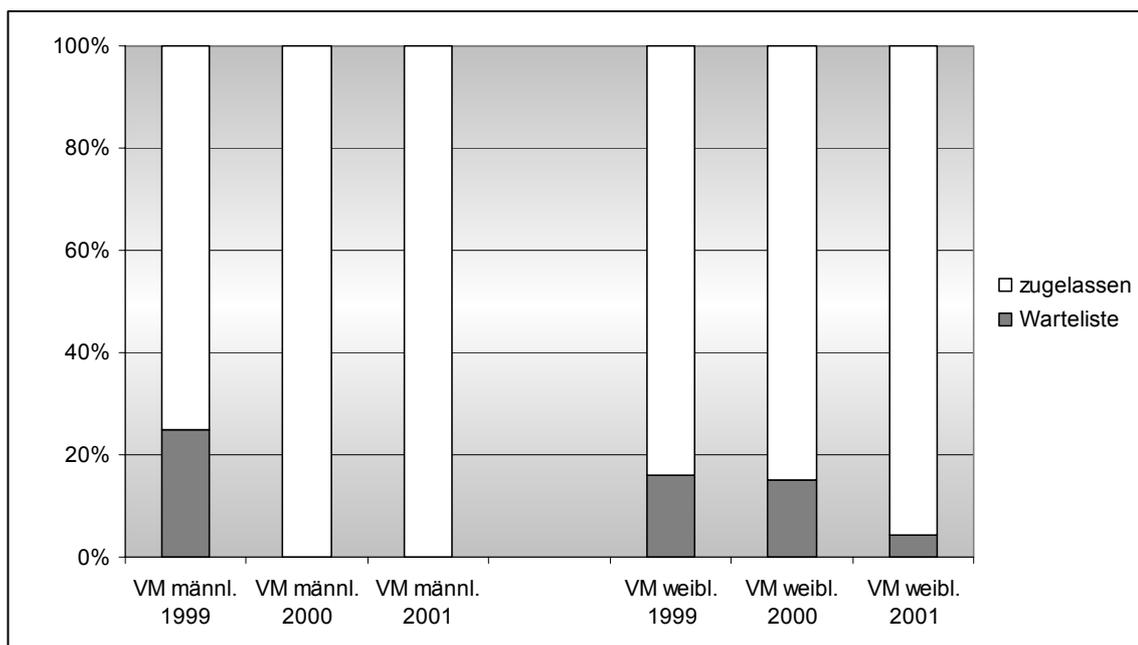


Abbildung 56: Zulassung Humanmedizin (HM) für die Geschlechter; alle Testteilnehmer auf je 100% bezogen

Die Zulassungsquoten in der Humanmedizin bleiben über die Jahre konstant (siehe Abbildung 56). Den diesjährigen Doppel-Maturitätsabgängen im Kanton Bern wurde mit einer Erhöhung der Kapazitäten von Seiten der Universitäten begegnet. Es darf festgestellt werden, dass das Ausmass der Kapazitätserhöhung angemessen war.

In der Veterinärmedizin stand eine leichte Erhöhung der Testteilnehmerzahl einer grösseren Steigerung der Kapazitäten gegenüber, was zu einer massiven Erhöhung der Zulassungsquote geführt hat. Es ist zu beachten, dass die Anzahl der Testteilnehmerinnen jeweils weit über derjenigen der männlichen Bewerber liegt. Hier wurde der Effekt der Doppelabgänge überschätzt, eine geringere Kapazitätserhöhung für die Veterinärmedizin wäre im Nachhinein gesehen vertretbar gewesen.



**Abbildung 57:** Zulassung Veterinärmedizin (VM) für die Geschlechter; alle Testteilnehmer auf je 100% bezogen

## 8.7 Vergleich der Kantone

Bereits in den letzten Jahren bestand ein grosses Interesse an Ergebnissen, die sich auf die Vergleiche zwischen den Kantonen beziehen. Die Ursache scheint auch ein gewisser Mangel an Vergleichsstudien im Bildungsbereich zu sein. Wiederum ist vor unzulässigen Generalisierungen vorab zu warnen: es handelt sich nicht um eine Repräsentativerhebung für „die Altersgruppe“ oder „die Maturanden“. Medizinbewerbungen werden nach verschiedenen Gesichtspunkten erfolgen und sich zwischen Kantonen auch möglicherweise unterscheiden – dies würde aber Gegenstand einer eigenen Untersuchung sein müssen.

Dennoch können die Ergebnisse Hinweise darauf liefern, ob sich bestimmte systematische Unterschiede zeigen. Die Interpretation muss mit der gebotenen Vorsicht für den jeweiligen Einzelfall vorgenommen werden.

### 8.7.1 Vergleich 2001 innerhalb der Deutschschweiz

Eine der wichtigsten Fragen ist jeweils, ob sich die Leistungen für die Kantone nach unterschiedlichen Maturitätsquoten unterscheiden. Diesbezüglich konnten in den letzten Jahren keine Unterschiede verzeichnet werden. Basierend auf den Quoten für 1999<sup>1</sup> wurden die Kantone erneut in die drei Gruppen geteilt, die hohe, mittlere und niedrige Maturitätsquoten repräsentieren (Tabelle 70).

hoch			mittel			niedrig		
Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl
BS	21.2	37	SH	18.4	18	SG	12.8	48
BL	20.4	45	NW	16.5	3	SZ	12.7	15
ZH	18.8	147	AG	15.9	49	AI	12.3	1
			GL	15.3	3	GR	12.3	36
			ZG	15	23	LU	11.5	47
			SO	13.9	25	TG	10.4	38
			AR	13.8	10	UR	10.2	5
			BE	13.1	185*	OW	9.9	3
Summe		229	Summe		316	Summe		193

\*= Doppeljahrgang

Tabelle 70: Gruppierung der Kantone nach der Maturitätsquote (Basis 1998)

Auch im Jahr 2001 können keine Unterschiede bezüglich des Testwerts nach der Maturandenquote nachgewiesen werden (Tabelle 71). Einzig im Untertest Quantitative und formale Probleme tritt ein signifikanter Wert auf.

Die Ergebnisse belegen zumindest, dass es gegenwärtig keinen Grund dafür gibt, in der unterschiedlichen Maturitätsquote eine generelle Ursache zu sehen, aus der sich Schlussfolgerungen für einzelne Kantone bezüglich einer Auswahl für die Mittelschule ergeben.

<sup>1</sup> Zahlen aus: Bundesamt für Statistik (1999). Maturitäten 1998: Statistik der Schweiz Bd. 15., Tabelle 4 Seite 13.

	Maturitäts- quote	Mittel- wert	Stand.- abw.		Quadrat- summe	df	MQ	F	Sig.
Testwert	hoch	101.59	10.033	Zwischen Gruppen	157.403	2	78.702	.805	.448
	mittel	100.57	9.606						
	niedrig	100.61	10.167						
Muster zuordnen	hoch	10.62	2.777	Zwischen Gruppen	43.715	2	21.858	2.794	.062
	mittel	10.06	2.670						
	niedrig	10.17	3.016						
Medizin.-natur- wiss. Grundver- ständnis	hoch	10.36	3.451	Zwischen Gruppen	.598	2	.299	.025	.975
	mittel	10.31	3.429						
	niedrig	10.28	3.427						
Schlauchfiguren	hoch	13.97	3.528	Zwischen Gruppen	35.971	2	17.986	1.332	.265
	mittel	13.46	3.802						
	niedrig	13.59	3.633						
Quantitat. u.formale Probleme	hoch	10.72	4.025	Zwischen Gruppen	96.703	2	48.352	3.296	<b>.038</b>
	mittel	11.09	3.736						
	niedrig	11.67	3.743						
Textverständnis	hoch	8.80	3.549	Zwischen Gruppen	22.180	2	11.090	.922	.398
	mittel	8.70	3.321						
	niedrig	8.36	3.607						
Figuren lernen	hoch	10.89	3.395	Zwischen Gruppen	15.703	2	7.851	.667	.514
	mittel	10.88	3.331						
	niedrig	10.55	3.630						
Fakten lernen	hoch	11.18	3.922	Zwischen Gruppen	40.857	2	20.428	1.505	.223
	mittel	10.66	3.489						
	niedrig	10.70	3.704						
Diagramme und Tabellen	hoch	10.58	3.105	Zwischen Gruppen	1.728	2	.864	.090	.914
	mittel	10.49	3.036						
	niedrig	10.46	3.162						
Konzentr.u. sorgf. Arb.	hoch	13.68	3.929	Zwischen Gruppen	40.187	2	20.093	1.432	.240
	mittel	13.21	3.758						
	niedrig	13.13	3.497						

**Tabelle 71:** Mittelwerte und Standardabweichungen für drei Gruppen der Kantone nach der Maturitätsquote, keine signifikanten Unterschiede für Testwert und die Untertests

Für Vergleiche zwischen den Kantonen werden wiederum wie in den Vorjahren die gleichen, zahlenmässig stärksten Kantone berücksichtigt (Tabelle 72). Die Abbildung 58 zeigt den Verlauf der Testwerte für die Kantone über die vier bisherigen Testdurchführungen.

Die Rangreihe der Kantone ist über die einzelnen Jahre relativ stabil. Zürich und St. Gallen liegen an der Spitze. Basel-Land erreichte dieses Jahr das zweithöchste Ergebnis. Basel-Stadt fällt nach einem stärkeren Jahrgang wieder auf das Ergebnis von 1999 zurück. Möglicherweise fehlt hier die organisierte Vorbereitung (Probelauf), wie sie im vergangenen Jahr stattfand.

Kanton		Testwert	Musterzuordnen	Medizin-naturw. Grundv.	Schlauchfiguren	Quantit. u. form. Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arb.
AG	M	<b>101.0</b>	9.7	10.3	13.7	11.8	8.7	11.2	10.2	10.8	13.3
	S	<b>9.1</b>	2.7	2.9	3.8	3.8	3.1	3.9	3.4	3.0	3.6
BE	M	<b>99.5</b>	10.0	10.3	13.1	10.9	8.6	10.7	10.5	10.2	12.5
	S	<b>9.8</b>	2.7	3.5	3.9	3.9	3.2	3.3	3.5	3.0	3.6
BL	M	<b>101.8</b>	10.9	10.0	14.1	10.1	8.6	11.8	11.1	10.8	13.7
	S	<b>8.9</b>	2.5	3.2	3.3	3.4	2.8	3.2	3.7	2.7	3.7
BS	M	<b>95.3</b>	9.8	8.9	12.7	9.3	7.3	9.4	9.6	9.4	12.1
	S	<b>9.6</b>	3.2	3.1	3.4	3.7	3.3	3.7	3.2	2.9	3.4
LU	M	<b>100.4</b>	9.1	10.7	13.1	12.0	8.9	10.6	10.7	11.0	12.3
	S	<b>10.4</b>	2.6	3.2	3.6	3.9	3.5	3.7	3.7	3.1	3.5
SG	M	<b>101.3</b>	10.8	10.3	13.9	11.7	8.7	10.6	10.8	10.4	13.1
	S	<b>11.6</b>	3.1	3.3	4.2	3.7	4.0	3.8	3.9	3.5	3.1
ZH	M	<b>103.1</b>	10.7	10.8	14.3	11.3	9.3	11.0	11.6	10.8	14.1
	S	<b>101.0</b>	9.7	10.3	13.7	11.8	8.7	11.2	10.2	10.8	13.3

Tabelle 72: Vergleich der Kantone für Testwert und die einzelnen Untertests

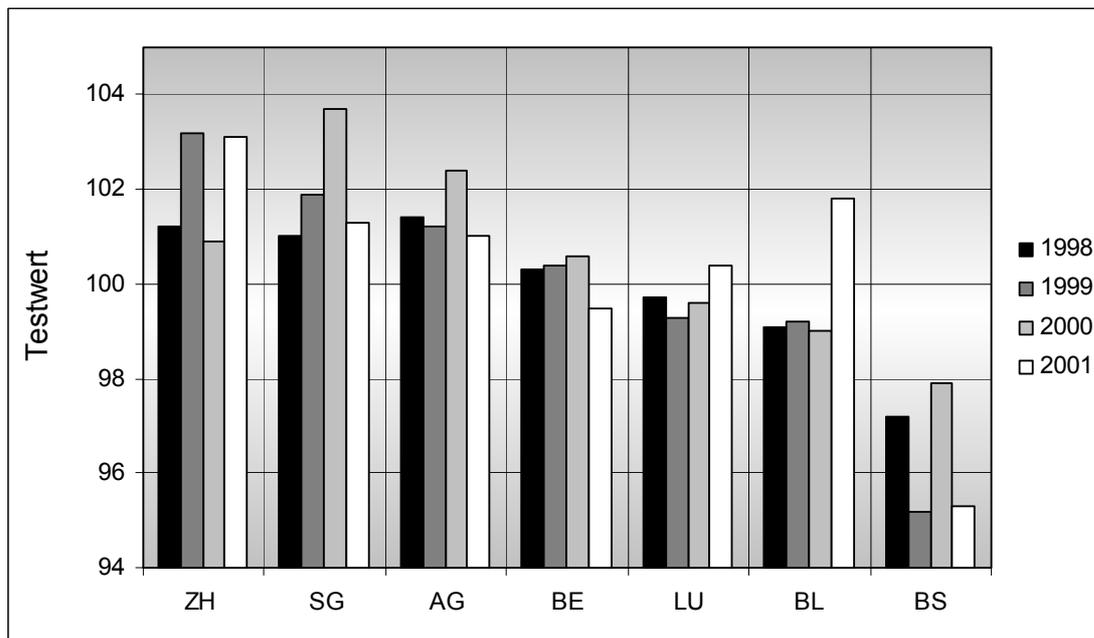


Abbildung 58: Testwert für Kantone im Vergleich

Die nachfolgende Tabelle 73 zeigt für den Testwert und fünf Untertests die Ergebnisse der Homogenitätsprüfung der Mittelwerte. Es werden je zwei homogene Mittelwertgruppen identifiziert. Markiert sind die Werte mit signifikanten Differenzen.

Testwert	n	Subset for alpha = .05	
		1	2
BS	37	95.27	
BE	185	99.50	99.50
LU	47	100.36	100.36
AG	49		101.04
SG	48		101.27
BL	45		101.78
ZH	147		103.12
Sig.		.092	.453

Muster zuordnen	n	Subset for alpha = .05	
		1	2
LU	47	9.09	
AG	49	9.73	9.73
BS	37	9.76	9.76
BE	185	9.97	9.97
ZH	147		10.74
SG	48		10.77
BL	45		10.91
Sig.		.607	.260

Med.-naturwiss. Grundverständnis	n	Subset for alpha = .05	
		1	2
BS	37	8.86	
BL	45	10.04	10.04
BE	185	10.28	10.28
SG	48	10.29	10.29
AG	49	10.33	10.33
LU	47	10.70	10.70
ZH	147		10.83
Sig.		.058	.878

Quantitative und formale Probleme	n	Subset for alpha = .05	
		1	2
BS	37	9.27	
BL	45	10.09	10.09
BE	185	10.85	10.85
ZH	147	11.27	11.27
SG	48		11.73
AG	49		11.80
LU	47		11.96
Sig.		.094	.146

Textverständnis	n	Subset for alpha = .05	
		1	2
BS	37	7.27	
BL	45	8.58	8.58
BE	185	8.63	8.63
SG	48	8.67	8.67
AG	49	8.69	8.69
LU	47	8.85	8.85
ZH	147		9.26
Sig.		.173	.939

Figuren lernen	n	Subset for alpha = .05	
		1	2
BS	37	9.43	
SG	48	10.56	10.56
LU	47	10.60	10.60
BE	185	10.72	10.72
ZH	147	10.98	10.98
AG	49	11.22	11.22
BL	45		11.80
Sig.		.086	.477

**Tabelle 73:** Multiple Mittelwertvergleiche für Kantone (Untertests mit signifikanten Unterschieden)

Beim Testwert lassen sich 2 Mittelwertsgruppen unterscheiden. Basel-Stadt unterscheidet sich hier signifikant von den Ergebnissen in Aargau, St. Gallen, Basel-Land und Zürich.

Im Untertest Muster zuordnen ist der Überlappungsbereich etwas grösser, hier unterscheidet sich nur Luzern von Zürich, St. Gallen und Basel-Land.

Nur ein signifikanter Unterschied kann in den Untertests Medizinisch-naturwissenschaftliches Grundverständnis und Textverständnis beobachtet werden. Es unterscheidet sich einzig Basel-Stadt von Zürich.

Der Untertest Quantitative und formale Probleme weist wieder einen geringeren Überlappungsbereich auf. Signifikante Differenzen findet man hier zwischen Basel-Stadt und St. Gallen, Aargau, Luzern.

Die höchsten Punktzahlen im Untertest Figuren lernen wurden im Kanton Basel-Land erzielt. Die einzige signifikante Differenz ist jene zum Ergebnis in Basel-Stadt.

### 8.7.2 Zusammenfassende Analyse für Kantone 1998 bis 2001

Seit 1998 haben 3286 Personen am Test teilgenommen. In der Tabelle 74 sind alle Teilnehmer am EMS 1998 bis 2001 nach Kantonen aufgeführt. Zu beachten ist, dass 1998 nur Humanmedizin, 1999 bis 2001 Human- und Veterinärmedizin einem NC unterlag.

Wohnkanton/ Wohnort	Disziplin		Geschlecht		Sprache			Total
	Human- medizin	Veterinär- medizin	männlich	weiblich	deutsch	franzö- sisch	italie- nisch	
AG	232	38	107	163	270			270
AI	6	1	2	5	7			7
AR	22	2	10	14	24			24
BE	423	91	191	323	503	11		514
BL	181	14	85	110	193	2		195
BS	135	11	51	95	145		1	146
FR	130	22	62	90	48	104		152
GE	2	33	9	26	1	34		35
GL	14	2	4	12	16			16
GR	87	16	47	56	98		5	103
JU	1	3	2	2		4		4
LU	148	27	65	110	175			175
NE	3	13		16		16		16
NW	14		4	10	14			14
OW	17		6	11	17			17
SG	207	21	83	145	228			228
SH	31	6	14	23	37			37
SO	82	17	46	53	99			99
SZ	68	4	27	45	72			72
TG	88	15	36	67	103			103
TI	105	19	63	61	1		123	124
UR	10	1	7	4	11			11
VD	7	47	20	34	2	52		54
VS	57	17	28	46	51	23		74
ZG	35	13	20	28	47	1		48
ZH	622	76	274	424	693	1	4	698
FL	14	2	7	9	16			16
Ausland	28	6	14	20	33	1		34
Total	2769		1284	2002	2904	249	133	3286

Tabelle 74: Gesamtzahl der Teilnehmer am EMS 1998 bis 2001; das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt

Die Testwerte sind durch die Standardisierung auf den jeweiligen Mittelwert und die Streuung des Geburtsjahres zwischen den Jahren direkt vergleichbar. Deshalb kann ein Vergleich über alle drei Jahre gezogen werden. Dies lässt, wie mehrfach betont, keine Bewertung von Schulbildung zu, weil der Test eher bildungsunabhängig ist und die Bewerber nicht repräsentativ für die Maturanden sind.

Wohnkanton/ Wohnort	n	Mittelwert	Std. Abweichung	Minimum	Maximum
AG	270	101.54	8.850	74	127
AI	7	102.29	7.566	95	116
AR	24	102.25	9.483	89	127
BE	514	100.12	9.953	72	125
BL	195	99.75	9.419	74	125
BS	146	96.50	9.915	71	122
FR	152	94.31	9.067	70	121
GE	35	94.00	8.530	76	117
GL	16	104.69	10.448	89	130
GR	103	99.47	9.169	77	118
LU	175	99.79	9.311	72	124
NE	16	92.12	9.408	79	116
NW	14	102.43	12.132	78	121
OW	17	100.59	7.890	92	126
SG	228	101.98	10.281	70	128
SH	37	100.59	9.394	82	120
SO	99	101.95	9.239	79	127
SZ	72	99.58	9.159	76	127
TG	103	101.99	9.530	77	122
TI	124	96.85	9.147	75	120
UR	11	101.18	9.590	87	120
VD	54	96.06	8.682	80	119
VS	74	98.05	10.407	70	120
ZG	48	99.96	9.843	78	120
ZH	698	101.99	10.384	70	130
FL	16	96.63	9.818	82	120
Ausland	34	94.74	9.955	78	113
Total	3286	100.006	9.9793	70	130

**Tabelle 75:** Kennwerte für den Testwert von Kantonen; das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt

Nachfolgend werden getrennt für Human- und Veterinärmedizin die Testwertverteilungen und die theoretischen (!) Zulassungsquoten verglichen. Letztere sind nach der genauen Kapazität festgelegt und berücksichtigen nicht die Rückzüge und erneute Vergabe dieser Plätze an Personen mit niedrigeren Testwerten als dem theoretischen Zulassungskriterium. Die mittleren Testwerte für beide Disziplinen unterscheiden sich, die Zusammensetzung der Bewerber ist nicht vergleichbar (grösserer Bewerberanteil für Veterinärmedizin aus der Westschweiz) und auch die Zulassungsquoten sind unterschiedlich. Bei der Interpretation sind die Stichprobengrössen der Tabelle 74 unbedingt zu beachten.

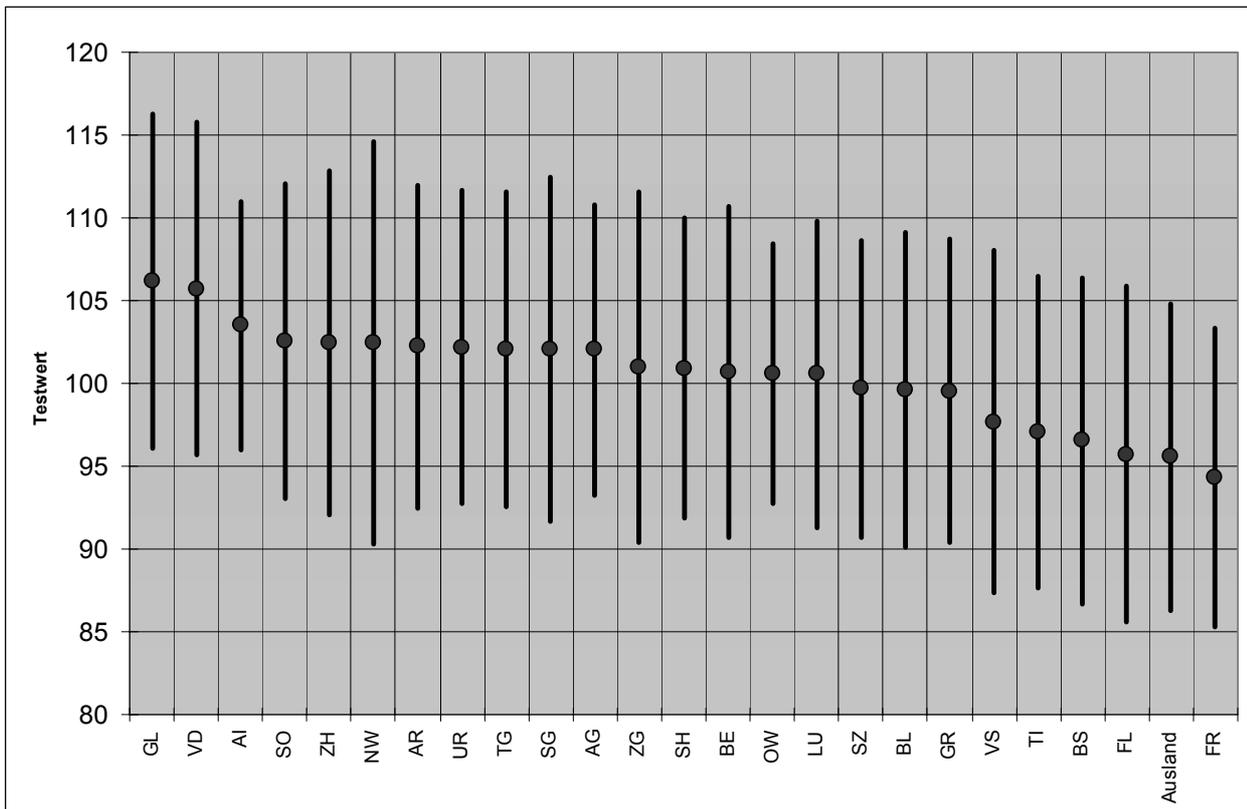


Abbildung 59: Mittelwerte und Streubereiche ( $m \pm s$ ) der Testwerte für Kantone mit mehr als 5 Bewerbern 1998 bis 2001 zusammengefasst; **Humanmedizin**

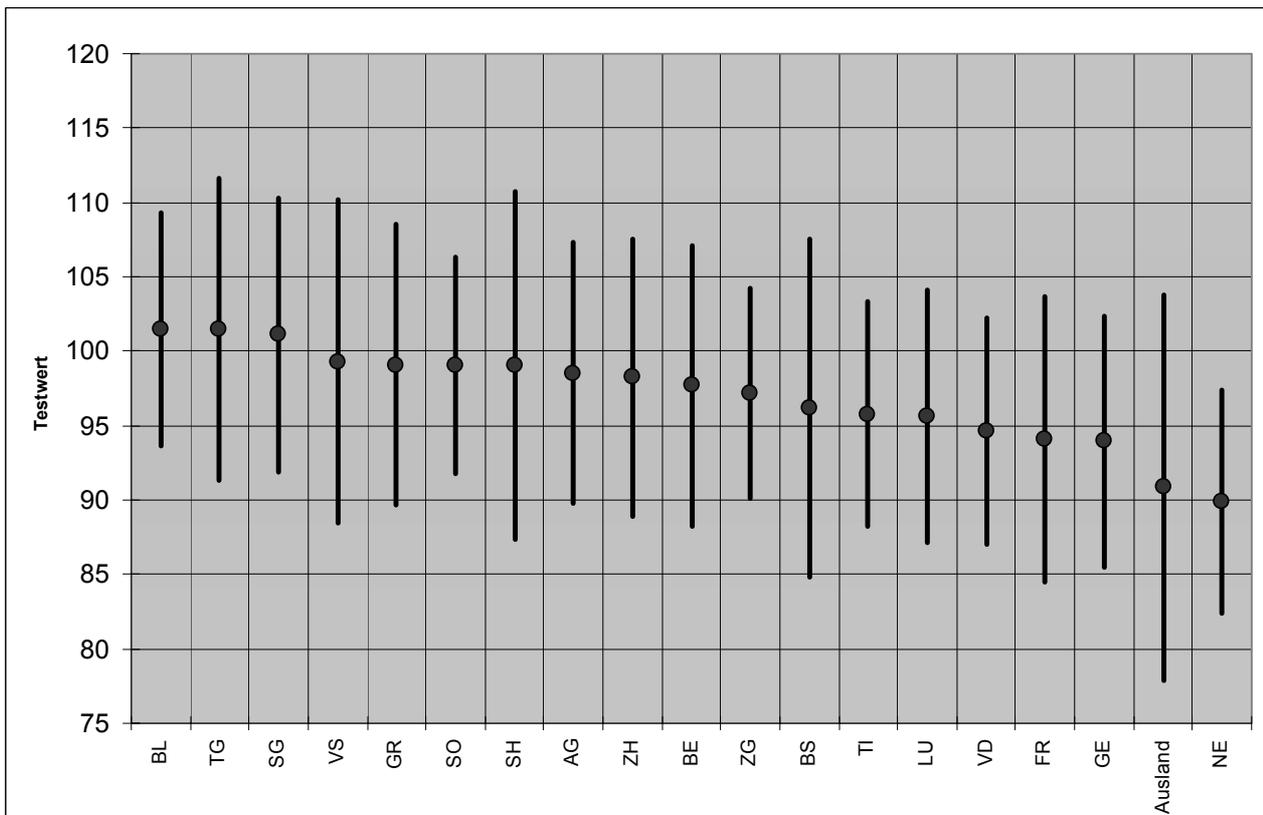


Abbildung 60: Mittelwerte und Streubereiche ( $m \pm s$ ) der Testwerte für Kantone mit mehr als 5 Bewerbern 1998 bis 2001 zusammengefasst; **Veterinärmedizin**

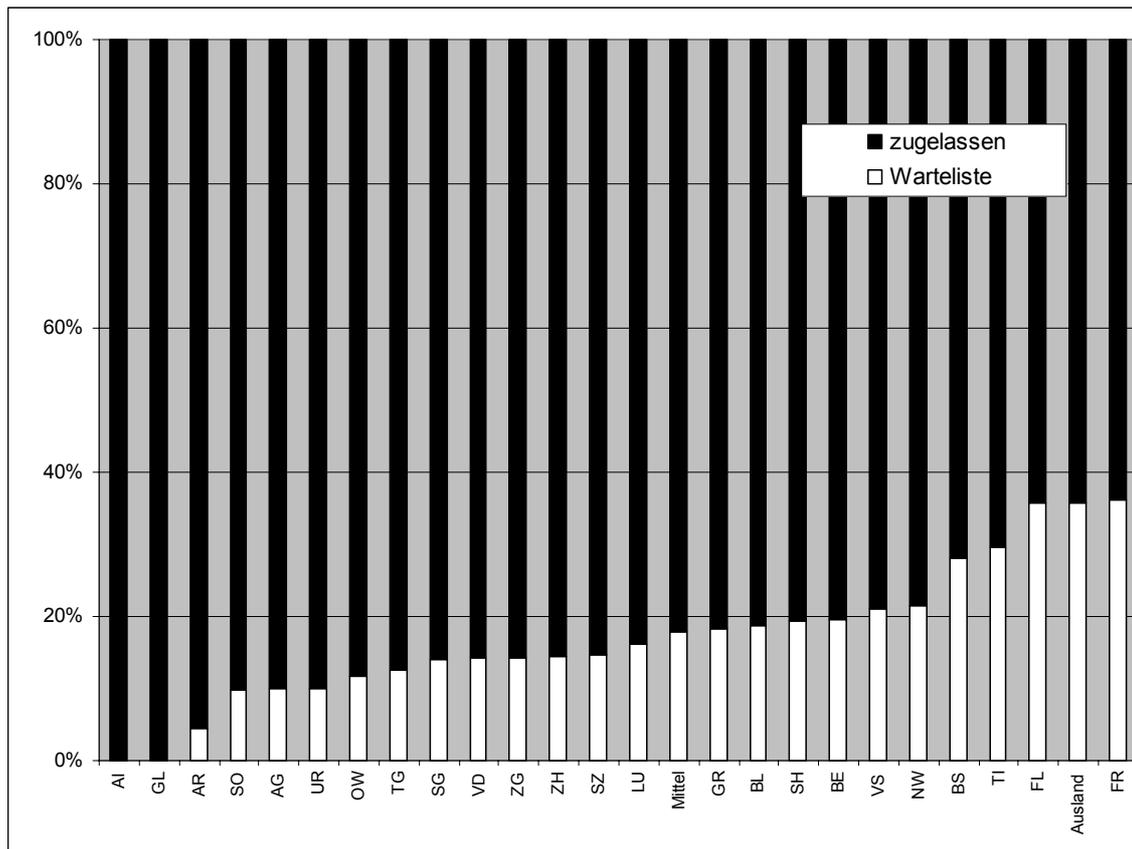


Abbildung 61: Mehrjährige theoretische Zulassungsquoten auf der Basis der Studienplatzvergabe nach dem Testtermin (ohne Nachrücker) 1998 bis 2001; **Humanmedizin**

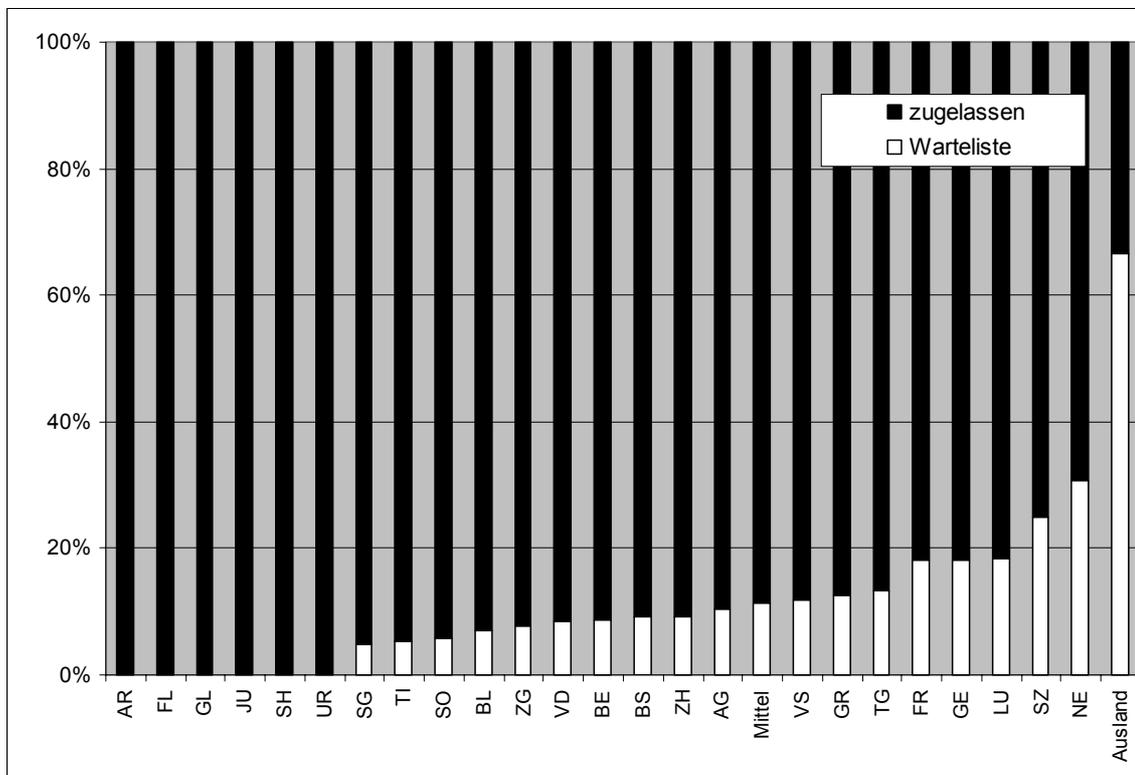


Abbildung 62: Mehrjährige theoretische Zulassungsquoten auf der Basis der Studienplatzvergabe nach dem Testtermin (ohne Nachrücker) 1999 bis 2001; **Veterinärmedizin**

## 8.8 Vergleiche für Altersgruppen

Bisher wurde jedes Jahr nachgewiesen, dass Personen, die älter sind, etwas geringere Zulassungschancen haben. Dies war besonders der Fall, wenn die Personen die Maturitätsprüfung relativ spät abgelegt haben, also der Schulbesuch nicht zeitlich zusammenhängend erfolgte. Diese Personengruppe wird aus verschiedensten Gründen nicht frühestmöglich die Mittelschule besucht haben.

Für die Überprüfung der Alterseffekte im Jahr 2001 wurden die Gruppen auf der Basis der Tabelle 76 gebildet. Die Altersgruppen wurden um 1 Jahr gegenüber 2000 verschoben. Die älteste Gruppe wurde nahe dem Median des Maturitätsalters geteilt. Zusätzlich ist der Doppeljahrgang zu beachten, wodurch jüngere Bewerber (bis Jahrgang 1985) berücksichtigt werden müssen.

Maturitätsjahr	Geburtsjahr			Total
	1954-1977	1978-1980	1981-1985	
1973	1			1
1981	1			1
1986	1			1
1987	1			1
1990	1			1
1992	3			3
1994	2			2
1995	1			1
1996	1	1		2
1997	7	1		8
1998	7	17		24
1999	2	28	2	32
2000	7	68	58	133
2001	29	126	510	665
	64	241	570	875

Tabelle 76: Gruppenbildung für Jahr der Maturitätsprüfung bezogen auf die Geburtsjahre

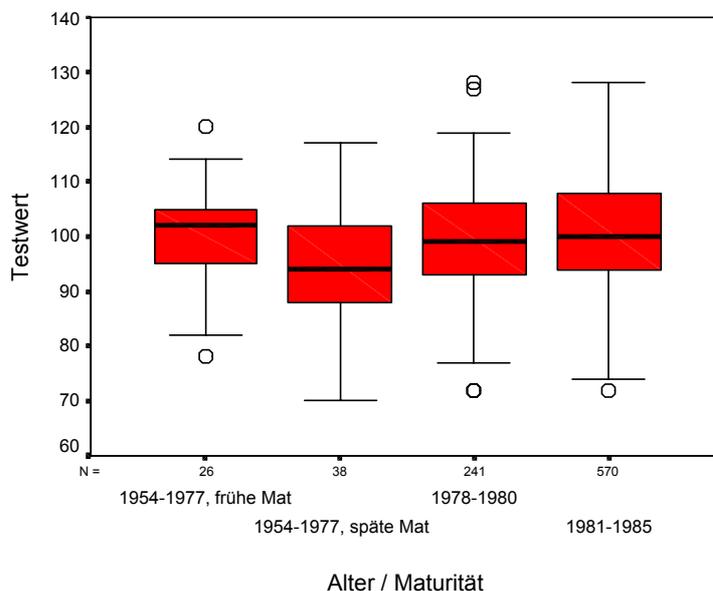


Abbildung 63: Boxplots für Testwert der Altersgruppen (Geburtsjahre) und Maturität

Die älteste Bewerbergruppe mit frühem Maturitätsabschluss erreicht im Mittel höhere durchschnittliche Testwerte als die zweitjüngste Gruppe (Tabelle 78), aber gleichzeitig durch die schiefe Verteilung eine geringere Zulassungsquote. Der Einfluss des Maturitätsalters kann bei den älteren Bewerbern hinsichtlich des Testwerts wieder beobachtet werden – die älteren Personen mit später Maturität haben die geringsten Mittelwerte. Aufgrund der geringeren Bewerberzahl älterer Personen schlug sich dieser Effekt diesmal allerdings kaum in der Zulassungsquote (Tabelle 77) nieder. Hier ist nur ein schwacher Unterschied beobachtbar.

Geburtsjahr und Maturität	Warteliste	zugelassen	n
1954-77, frühe Maturität	6 23.1%	20 76.9%	26
1954-77, späte Maturität	9 23.7%	29 76.3%	38
1978 - 80	45 18.5%	198 81.5%	243
1981 - 85	74 13.0%	496 87.0%	570
	134 15.3%	743 84.7%	877

Tabelle 77: Zulassungsquoten für die drei Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung

Bezüglich Mittelwertsunterschieden zeigen sich im Testwert und vier Untertests signifikante Differenzen zwischen den Gruppen (Tabelle 79). Im Testwert sowie den Untertests Quantitative und formale Probleme, Textverständnis, Diagramme und Tabellen stellt die älteste Gruppe mit später Maturität jeweils die schwächste Gruppe dar. Mit Ausnahme des Untertests Diagramme und Tabellen sind aber nur die Vergleiche mit der jüngsten Gruppe signifikant. Im Untertest Muster zuordnen erreichte die älteste Gruppe mit früher Maturität signifikant geringere Punktzahlen als die jüngste Gruppe.

Der zeitliche Verlauf der Wartelisten-Quoten für die Alters- und Maturitätskategorien ist in Abbildung 64 dargestellt. Trotz leicht höherem Mittelwert liegt die Wartelistenquote der älteren Gruppe mit frühem Maturitätsabschluss diesmal über derjenigen der zweitjüngsten Gruppe. Die besten Leistungen bezüglich Testwert und Zulassungsquote zeigt die jüngste Gruppe.

	Geburtsjahr und Maturität	Mittelwert	Standardabw.	Quadratsumme	df	MQ	F	Sig.
Testwert	54-77, frühe Matur.	99.38	10.288	1854.142	3	618.047	6.320	<b>.000</b>
	54-77, späte Matur.	94.13	11.309					
	1978-80	99.12	9.774					
	1981-85	100.78	9.821					
Muster zuordnen	54-77, frühe Matur.	8.88	2.732	90.426	3	30.142	3.843	<b>.009</b>
	54-77, späte Matur.	9.29	2.680					
	1978-80	10.00	2.812					
	1981-85	10.31	2.807					
Med.-naturwiss. Grundverständnis	54-77, frühe Matur.	9.69	4.018	27.439	3	9.146	.766	.513
	54-77, späte Matur.	9.32	3.146					
	1978-80	10.02	3.489					
	1981-85	10.13	3.434					
Schlauchfiguren	54-77, frühe Matur.	13.73	3.683	51.537	3	17.179	1.225	.300
	54-77, späte Matur.	12.24	4.129					
	1978-80	13.42	3.743					
	1981-85	13.25	3.722					
Quantitative und formale Probleme	54-77, frühe Matur.	10.42	4.456	255.172	3	85.057	5.663	<b>.001</b>
	54-77, späte Matur.	8.97	4.220					
	1978-80	10.47	3.709					
	1981-85	11.24	3.894					
Textverständnis	54-77, frühe Matur.	8.31	3.988	173.077	3	57.692	4.896	<b>.002</b>
	54-77, späte Matur.	6.89	3.220					
	1978-80	8.14	3.466					
	1981-85	8.77	3.405					
Figuren lernen	54-77, frühe Matur.	10.31	3.172	128.657	3	42.886	3.587	<b>.013</b>
	54-77, späte Matur.	9.11	3.652					
	1978-80	10.37	3.656					
	1981-85	10.82	3.369					
Fakten lernen	54-77, frühe Matur.	11.12	3.702	105.513	3	35.171	2.649	<b>.048</b>
	54-77, späte Matur.	9.82	3.439					
	1978-80	10.38	3.655					
	1981-85	11.01	3.650					
Diagramme und Tabellen	54-77, frühe Matur.	10.42	3.049	163.255	3	54.418	5.646	<b>.001</b>
	54-77, späte Matur.	8.55	2.787					
	1978-80	9.97	3.014					
	1981-85	10.49	3.164					
Konzentriertes und sorgfältiges Arbeiten	54-77, frühe Matur.	13.46	3.744	51.208	3	17.069	1.246	.292
	54-77, späte Matur.	12.00	4.268					
	1978-80	13.19	3.776					
	1981-85	13.06	3.627					

**Tabelle 78.** Testwert und Punktwerte für die drei Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und späterer Maturitätsprüfung; varianz-analytische Prüfung des Unterschiedes

Testwert	n	Subset for alpha = .05	
		1	2
1954-1977, späte Matur	36	94.92	
1978-1980	230	99.40	99.40
1954-1977, frühe Matur	23	100.09	100.09
1981-1985	481		102.03
Sig.		.064	.597

Muster zuordnen	n	Subset for alpha = .05	
		1	2
1954-1977, frühe Matur	23	8.87	
1954-1977, späte Matur	36	9.50	9.50
1978-1980	230	10.01	10.01
1981-1985	481		10.47
Sig.		.236	.387

Quantitative und formale Probleme	n	Subset for alpha = .05	
		1	2
1954-1977, späte Matur	36	9.25	
1978-1980	230	10.50	10.50
1954-1977, frühe Matur	23	10.87	10.87
1981-1985	481		11.53
Sig.		.201	.605

Textverständnis	n	Subset for alpha = .05	
		1	2
1954-1977, späte Matur	36	7.08	
1978-1980	230	8.20	8.20
1954-1977, frühe Matur	23	8.35	8.35
1981-1985	481		9.01
Sig.		.321	.689

Diagramme und Tabellen	n	Subset for alpha = .05	
		1	2
1954-1977, späte Matur	36	8.64	
1978-1980	230	10.09	10.09
1954-1977, frühe Matur	23		10.74
1981-1985	481		10.83
Sig.		.116	.671

Tabelle 79: Multiple Mittelwertvergleiche für Altersgruppen (nur deutschsprachige Teilnehmer, Untertests mit signifikanten Mittelwertsunterschieden)

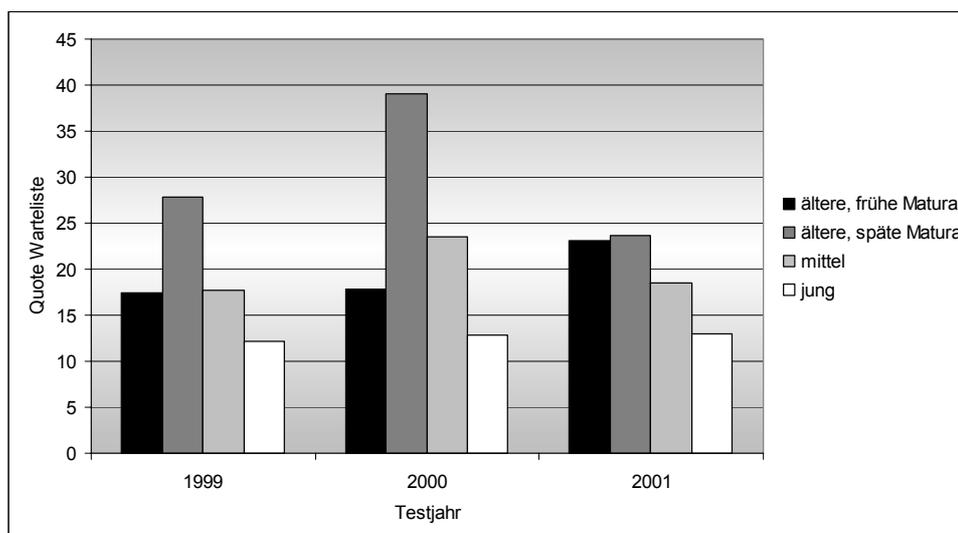


Abbildung 64: Wartelisten-Quote nach Altersgruppen – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung

Wegen der Doppelmaturitäten bewarben sich in diesem Jahr erstmals Personen, welche die Schule 12 Jahre bis zur Maturitätsprüfung besuchten. Für einen exakten Vergleich mit den Personen, die dies in 13 Jahren absolvierten, fehlen Daten zur genauen Unterscheidung. Näherungsweise wurde deshalb die jüngste Gruppe noch einmal geteilt unter der Annahme, dass sich die Anteile beider „Maturitätstypen“ unterschiedlich verteilen. Hinsichtlich des Testwertes konnte zwischen diesen Gruppen keine Differenz gefunden werden. Einzig in den Untertests Diagramme und Tabellen sowie konzentriertes und sorgfältiges Arbeiten schnitt die jüngste Kategorie (1983-85) schlechter ab als der Jahrgang 1981 (Tabelle 80).

Diagramme und Tabellen	n	Subset for alpha = .05		Konz. u. sorgf. Arbeiten	n	Subset for alpha = .05	
		1	2			1	2
1983-85	30	9.53		1983-85	30	11.73	
1982	147		10.79	1982	147	12.98	12.98
1981	304		10.98	1981	304		13.87
Sig.		1.000	.925	Sig.		.101	.312

Tabelle 80: Multiple Mittelwertvergleiche für junge (deutschsprachige) Testteilnehmer (Untertests mit signifikanten Unterschieden)

## 8.9 Vergleiche nach Wunschuniversitäten

Die Zulassung zu den Universitäten erfolgt aufgrund des Testwertes. Wie auch in den vorangegangenen Jahren unterscheiden sich die Wunschuniversitäten nicht signifikant hinsichtlich ihrer Zulassungsquoten.

Disziplinen	Universität	Zulassung		Total
		Warteliste	zugelassen	
Humanmedizin	Bern	37 16.3%	190 83.7%	227 100.0%
	Basel	25 19.4%	104 80.6%	129 100.0%
	Freiburg	19 28.4%	48 71.6%	67 100.0%
	Zürich	47 16.4%	239 83.6%	286 100.0%
	Total	128 18.1%	581 81.9%	709 100.0%
Veterinärmedizin	Bern	4 4.0%	96 96.0%	100 100.0%
	Zürich	2 2.8%	69 97.2%	71 100.0%
	Total	6 3.5%	165 96.5%	171 100.0%

Tabelle 81: Zulassungsquoten für Disziplinen und Wunschuniversität (erste Wahl)

## 9 Ergebnisse zur Testgüte

### 9.1 Zuverlässigkeit

Die Reliabilität (deutsche Ergebnisse vgl. Trost et al., 1995) kann anhand zweier Koeffizienten verglichen werden. Zunächst wurde die Zuverlässigkeit der Untertests abgeschätzt. Beim Untertest Konzentriertes und sorgfältigen Arbeiten ist diese Berechnung aufgrund seiner anderen Struktur nicht möglich. Seine Zuverlässigkeit wird indirekt dadurch nachgewiesen, dass das Weglassen dieses Untertests die Güte des Punktwertes verringern würde.

Die internen Konsistenzen (Cronbachs Alpha) schätzen die Messgenauigkeit anhand der Korrelationen jeder Aufgabe mit allen anderen des entsprechenden Untertests. Wie sich zeigt, liegen einige der in der Schweiz ermittelten Werte unterhalb der deutschen Vergleichsangaben. Die Konsistenz des Untertests Quantitative und formale Probleme liegt höher als diejenige in Deutschland.

Die Reliabilitäten nach der Testhalbierungsmethode (Teilung nach gerad- und ungeradzahligen Aufgaben) wurden ebenfalls verglichen. Wegen der Stichprobengrößen werden nur die Ergebnisse des deutschsprachigen Tests der Schweiz verwendet. Die Koeffizienten liegen hier eher im gleichen Wertebereich wie in Deutschland.

Untertest	Innere Konsistenz (Cronbachs Alpha)					Reliabilität nach Testhalbierungsmethode				
	Deutsch- land	Schweiz				Deutsch- land	Schweiz			
		1998	1999	2000	2001		1998	1999	2000	2001
Muster zuordnen	.64 - .69	.69	.60	.53	.56	.67 - .73	.72	.61	.62	.59
Med.-nat. Grundv.	.64 - .73	.70	.70	.61	.66	.65 - .72	.72	.71	.60	.69
Schlauchfiguren	.76 - .80	.71	.74	.71	.73	.75 - .82	.75	.74	.74	.72
Quant. u. form. Prob.	.69 - .71	.76	.75	.76	.74	.70 - .75	.78	.75	.74	.74
Textverständnis	.68 - .71	.73	.75	.74	.70	.69 - .74	.75	.70	.75	.75
Figuren lernen	.69 - .74	.72	.70	.62	.64	.72 - .75	.72	.70	.62	.62
Fakten lernen	.62 - .70	.68	.72	.64	.69	.68 - .73	.70	.70	.61	.72
Diagr. und Tab.	.67 - .71	.67	.75	.63	.61	.68 - .72	.66	.62	.61	.63
Gesamter Test						.91-.93	.92	.91	.90	.91

Tabelle 82: Zuverlässigkeit 1998 bis 2001; Vergleich der Skalenkonsistenzen für die Untertests, Ergebnisse für Deutschland aus Trost et al. (1997), S. 26 ff.

Die **Reliabilität des Punktwertes** insgesamt – geschätzt nach der Testhalbierungsmethode (gerade vs. ungerade Itemnummer) auf der Basis aller neun Untertests liegt mit 0.91 in einem sehr hohen Bereich. Hier liegt der Vergleichswert für Deutschland zwischen 0.91 und 0.93 (gemessen für rund 3000 Teilnehmer).

Die nachfolgende Tabelle 83 zeigt, dass das Weglassen von Untertests zu einer Verminderung der Gesamtkonsistenz führen würde. Die Konsistenz des Profils aufgrund der Untertests liegt bei 0.80.

Untertest	Konsistenz ohne den Untertest 1998	Konsistenz ohne den Untertest 1999	Konsistenz ohne den Untertest 2000	Konsistenz ohne den Untertest 2001
Muster zuordnen	.82	.80	.80	.78
Med.-nat. Grundv.	.79	.78	.78	.77
Schlauchfiguren	.80	.79	.79	.78
Quant. u. form. Probl.	.80	.79	.78	.78
Textverständnis	.80	.79	.78	.77
Figuren lernen	.81	.80	.80	.78
Fakten lernen	.82	.81	.80	.79
Diagramme und Tabellen	.80	.78	.78	.77
Konz. u. sorgf. Arbeiten	.82	.81	.80	.79
<b>TESTPROFIL</b>	<b>.83</b>	<b>.81</b>	<b>.81</b>	<b>.80</b>

Tabelle 83: Innere Konsistenzen, wenn der jeweilige Untertest weggelassen worden wäre und Konsistenz des Testprofils („standardized alpha“)

## 9.2 Binnenstruktur

Aufgrund der Korrelationen zwischen den Untertests kann mittels Strukturanalyse geprüft werden, ob sich die einzelnen Untertests bestimmten Dimensionen zuordnen. Die so gewonnene Struktur bietet Vergleichsmöglichkeiten mit theoretischen Vorstellungen zum Fähigkeitsbereich. Die Gruppierung der Untertestleistungen kann zu den bekannten Faktormodellen der Intelligenz in Beziehung gesetzt werden.

Ähnlich wie in früheren Jahren sind diesmal alle Korrelationen sogar auf dem 1%-Niveau signifikant. Die Korrelationen der Untertests mit dem Punktwert liegen ausnahmslos über 0.50 (Tabelle 84).

	Muster zuordnen	Med.-naturwiss. Grundv.	Schlauchfiguren	Quantitat. und formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Punktwert
Muster zuordnen	1	.265	.437	.182	.229	.316	.261	.227	.429	.577
Med.-naturw. Grundverst.	.265	1	.287	.544	.658	.222	.235	.541	.142	.694
Schlauchfiguren	.437	.287	1	.281	.259	.337	.269	.278	.404	.639
Quant. formale Probl	.182	.544	.281	1	.532	.191	.139	.555	.056	.630
Textverständnis	.229	.658	.259	.532	1	.240	.230	.546	.125	.681
Figuren lernen	.316	.222	.337	.191	.240	1	.399	.244	.358	.591
Fakten lernen	.261	.235	.269	.139	.230	.399	1	.244	.352	.568
Diagramme und Tabellen	.227	.541	.278	.555	.546	.244	.244	1	.179	.672
Konzent. u. sorgfält. Arbeiten	.429	.142	.404	.056	.125	.358	.352	.179	1	.548
Punktwert	.577	.694	.639	.630	.681	.591	.568	.672	.548	1

Tabelle 84: Korrelationen zwischen Punktwerten und Untertests

Die bereits bekannte Drei-Faktoren Struktur bleibt auch in diesem Jahr stabil. Die Ladungsmuster stimmen weiterhin gut überein. Tabelle 85 zeigt die Vergleichbarkeit der unrotierten Lösungen über die letzten drei Jahre. Erneut laden alle Untertests signifikant im ersten unrotierten Faktor (Tabelle 87).

Unrotierte Lösung									
Faktor	Eigenwert			% Varianz			Kumuliert %		
	2001	2000	1999	2001	2000	1999	2001	2000	1999
1	3.53	3.70	3.67	39.21	41.12	40.78	39.21	41.12	40.78
2	1.63	1.46	1.36	18.14	16.18	15.16	57.35	57.30	55.94
3	.85	.84	.91	9.39	9.31	10.16	66.74	66.61	66.10
4	.62	.66	.64	6.93	7.34	7.10	73.67	73.95	73.20
5	.57	.58	.59	6.35	6.40	6.50	80.03	80.35	79.70
6	.55	.53	.54	6.06	5.93	5.95	86.09	86.28	85.65
7	.50	.47	.53	5.53	5.23	5.86	91.61	91.51	91.50
8	.42	.39	.42	4.63	4.31	4.64	96.24	95.82	96.15
9	.34	.38	.35	3.76	4.18	3.85	100.00	100.00	100.00

Tabelle 85: Verlauf der unrotierten Lösung

In Tabelle 86 sind die rotierten Zwei- und Dreifaktorenlösungen dargestellt. Die Ladungen der einzelnen Untertests können Tabelle 88, beziehungsweise Tabelle 89 entnommen werden.

Varimax-rotierte Lösungen								
Eigenwert			% Varianz			Kumuliert %		
2001	2000	1999	2001	2000	1999	2001	2000	1999
Zwei-Faktorenlösung								
2.717	2.744	2.723	30.191	30.491	30.254	30.191	30.491	30.254
2.444	2.412	2.312	27.155	26.804	25.687	57.346	57.295	55.940
Drei-Faktorenlösung								
2.706	2.711	2.683	30.064	30.123	29.806	30.064	30.123	29.806
1.818	1.916	1.811	20.205	21.289	20.122	50.269	51.411	49.928
1.482	1.368	1.455	16.470	15.197	16.172	66.739	66.608	66.100

Tabelle 86: Varianzanteile der einzelnen Faktorenlösungen (rotierte Lösungen)

Im „Scree-Plot“ (Abbildung 65) wird der Eigenwerteverlauf der Faktoren dargestellt. Der Verlauf für den EMS 2001 befindet sich links, rechts davon zum Vergleich der Verlauf aus dem Vorjahr. Die Ein-Faktorenlösung ist erneut deutlich hervorgehoben, die Zwei-Faktorenlösung und bedingt auch die Drei-Faktorenlösung unterscheiden sich weniger von den jeweils nachfolgenden. Ab der Drei-Faktorenlösung sind die Eigenwerte kleiner als 1 – der Varianzanteil des Faktors ist kleiner als der eines einzelnen Untertests.

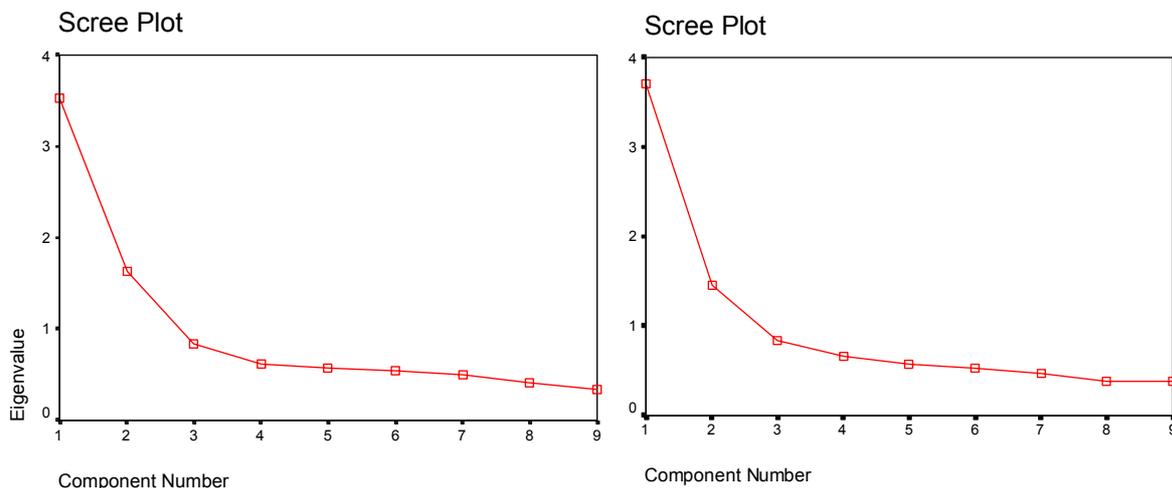


Abbildung 65: Scree-Plot (Eigenwerteverlauf) der Faktoren 1 bis 9 für die Untertests des EMS 2001 (links) und 2000 (rechts)

Die Abbildung 66 zur Struktur der Untertests fasst die Ergebnisse zusammen. Die Ladungsmatrix des ersten Faktors der unrotierten Lösung zeigt substantielle Ladungen für alle Untertests. Ein „Generalfaktormodell“ ist indiziert. Der allgemeine Fähigkeitsfaktor ist sehr varianzstark. Die Varianzaufklärung bleibt über die vergangenen vier Testdurchführungen konstant.

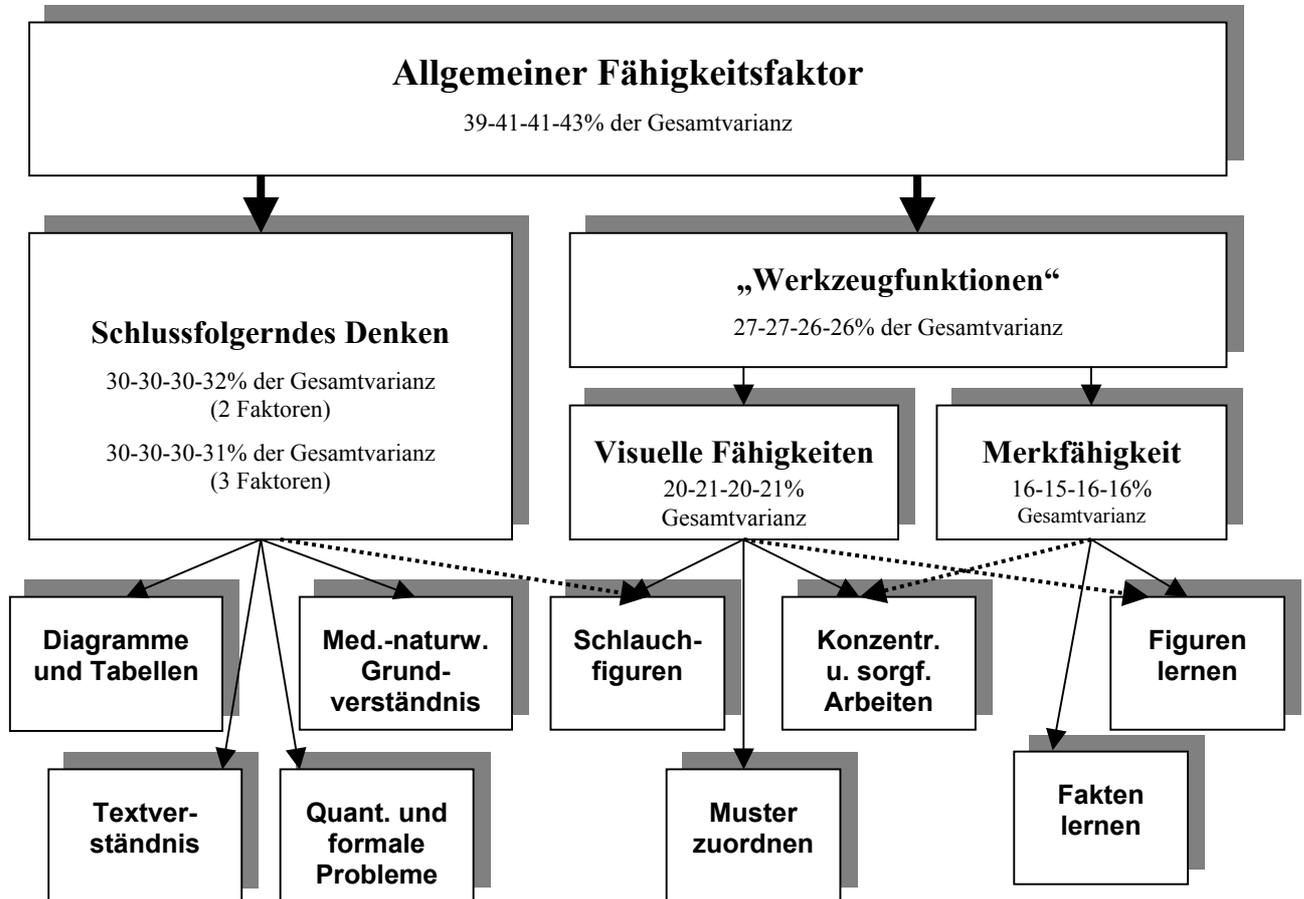


Abbildung 66: Struktur der Untertests des EMS, hierarchische Faktorenstruktur (2001 erster Wert, 2000 zweiter, 1999 dritter und 1998 vierter Wert der Varianzaufklärung durch den Faktor)

	Ladungen 1998	Ladungen 1999	Ladungen 2000	Ladungen 2001
Muster zuordnen	.56	.53	.57	.57
Med.-naturwiss. Grundverst.	.76	.76	.74	.74
Schlauchfiguren	.69	.64	.65	.62
Quant. und formale Probleme	.70	.70	.72	.66
Textverständnis	.72	.70	.71	.72
Figuren lernen	.62	.59	.53	.56
Fakten lernen	.53	.52	.55	.52
Diagramme und Tabellen	.72	.75	.71	.72
Konzentr. u. sorgf. Arbeiten	.47	.51	.55	.49

Tabelle 87: Ladungen im ersten Faktor der unrotierten Lösung 1998 bis 2001 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor)

	Faktor 1				Faktor 2				Kommunalitäten (h <sup>2</sup> )			
	1998	1999	2000	2001	1998	1999	2000	2001	1998	1999	2000	2001
Diagramme und Tabellen	<b>.83</b>	<b>.76</b>	<b>.79</b>	<b>.77</b>				.21	.71	.65	.66	.63
Med.-naturwiss. Grundverständnis	<b>.79</b>	<b>.81</b>	<b>.78</b>	<b>.82</b>	.22	.21	.22	(.18)	.68	.71	.66	.70
Textverständnis	<b>.77</b>	<b>.80</b>	<b>.80</b>	<b>.82</b>				(.16)	.64	.66	.67	.70
Quant. und formale Probleme	<b>.77</b>	<b>.78</b>	<b>.80</b>	<b>.81</b>					.62	.63	.68	.66
Konzentr. u. sorgf. Arbeiten					<b>.78</b>	<b>.79</b>	<b>.81</b>	<b>.79</b>	.62	.62	.66	.62
Figuren lernen	.23	(.16)	(.15)	(.16)	<b>.70</b>	<b>.72</b>	<b>.64</b>	<b>.66</b>	.55	.55	.44	.46
Muster zuordnen				(.15)	<b>.66</b>	<b>.64</b>	<b>.73</b>	<b>.69</b>	.48	.43	.54	.50
Fakten lernen	.20	.25	.27	.15	<b>.59</b>	<b>.51</b>	<b>.51</b>	<b>.62</b>	.40	.38	.34	.41
Schlauchfiguren	.45	.33	.30	.25	<b>.54</b>	<b>.59</b>	<b>.65</b>	<b>.66</b>	.50	.46	.51	.50

**Tabelle 88:** Ladungen und Kommunalitäten der Zwei-Faktorenlösung, varimaxrotiert

In der Zwei-Faktorenlösung zeigen sich „Schlussfolgerndes Denken“ und „Werkzeugfunktionen“ als Faktoren. Wiederum laden alle Untertests auch in einem Faktor. Die Ladung des Untertests Schlauchfiguren im ersten Faktor wurde gegenüber den letztjährigen Ergebnissen gemindert. Auch die Drei-Faktorenlösung bleibt stabil: der Faktor „Werkzeugfunktionen“ teilt sich in die Komponenten „Wahrnehmung“ und „Gedächtnis“ auf – der Faktor „Schlussfolgerndes Denken“ bleibt bezüglich der Ladungen und des Varianzanteiles gegenüber der Zwei-Faktorenlösung stabil.

Faktor	Schweiz Gesamt								Deutschland		
	1		2		3		h <sup>2</sup>		1	2	3
	2000	2001	2000	2001	2000	2001	2000	2001			
Diagramme und Tabellen	<b>.80</b>	<b>.77</b>		.13		.17	.67	.64	<b>.82</b>		
Medizin.-naturwiss. Grundverständnis	<b>.78</b>	<b>.82</b>		.15	.19	.12	.66	.70	<b>.81</b>		.20
Quantitat. u. formale Probleme	<b>.81</b>	<b>.81</b>	.17	.11			.69	.66	<b>.80</b>	.18	
Textverständnis	<b>.80</b>	<b>.82</b>			.18	.15	.67	.70	<b>.79</b>		.20
Muster zuordnen		.16	<b>.79</b>	<b>.80</b>		.10	.66	.68		<b>.81</b>	
Konzentr. u. sorgfält. Arbeiten			<b>.75</b>	<b>.68</b>	.33	.40	.67	.63		<b>.70</b>	.40
Schlauchfiguren	.32	.25	<b>.72</b>	<b>.75</b>		.11	.62	.64	.35	<b>.71</b>	
Fakten lernen	.22	.14		.11	<b>.85</b>	<b>.85</b>	.78	.76	.21		<b>.87</b>
Figuren lernen	.12	.15	.34	.28	<b>.67</b>	<b>.71</b>	.58	.61	.13	.47	<b>.64</b>

**Tabelle 89:** Faktorenanalyse: Varimaxrotierte Drei-Faktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42)

### 9.3 Item-Trennschärfen

Die Item-Trennschärfen (korrigierte Item-Total-Korrelation) folgen recht gut den deutschen Kennwerten. Fehlende Itemnummern bezeichnen Einstreuaufgaben.

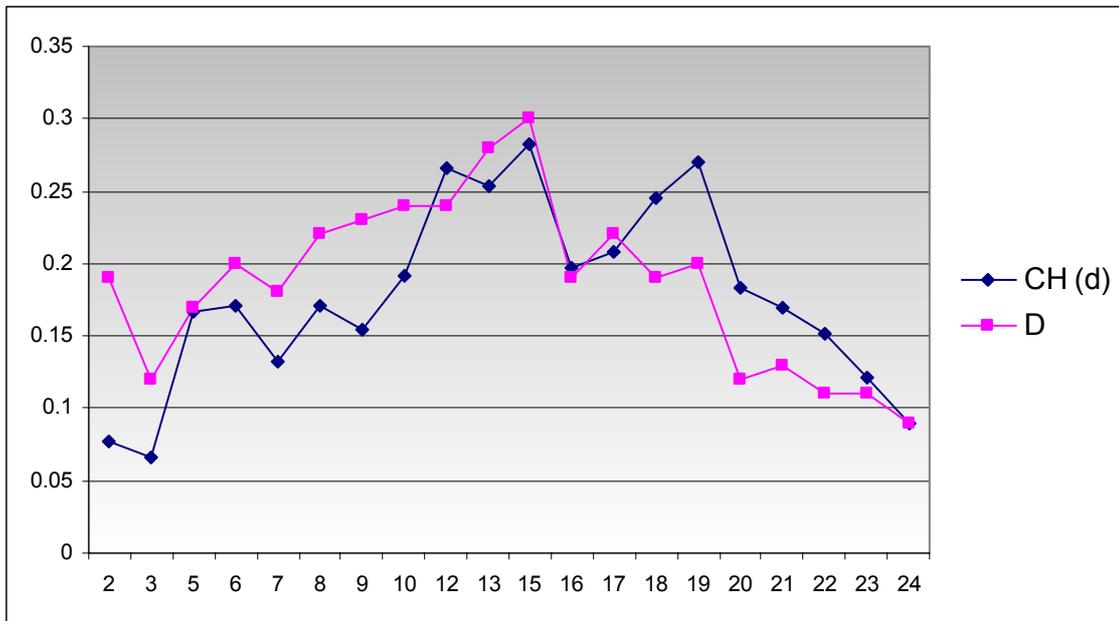


Abbildung 67: Trennschärfen für den Untertest Muster zuordnen

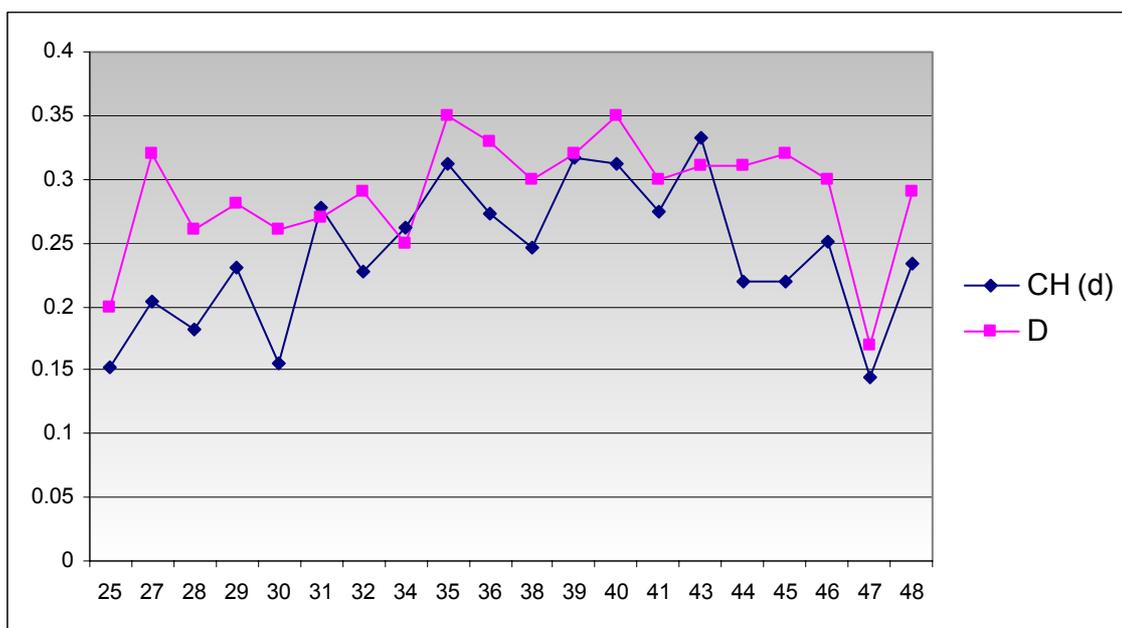


Abbildung 68: Trennschärfen für den Untertest Medizinisch-naturwissenschaftliches Grundverständnis

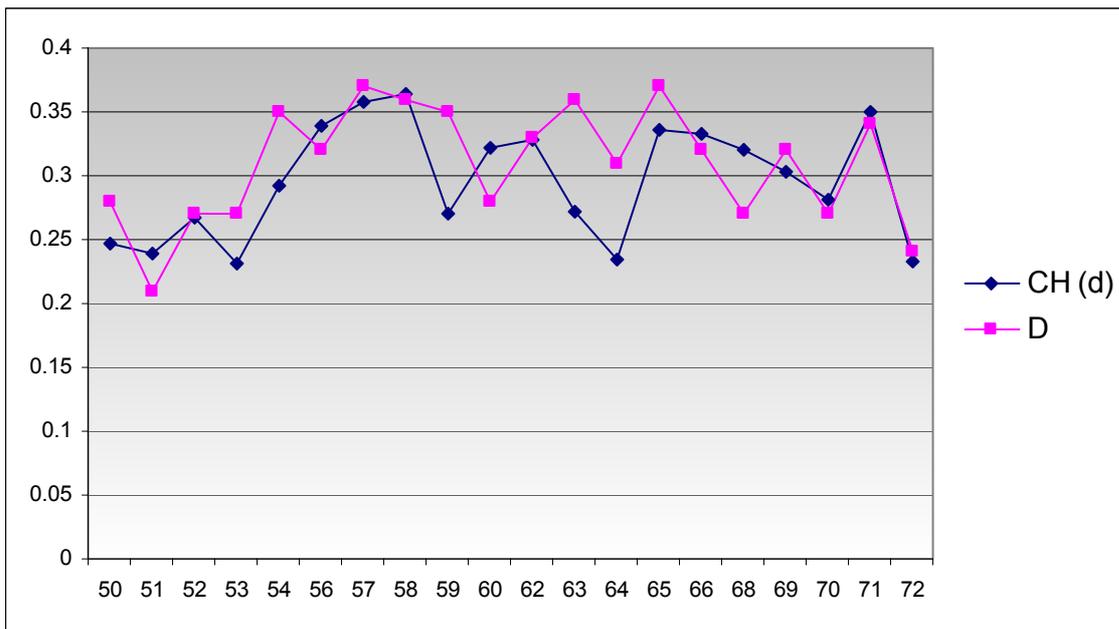


Abbildung 69: Trennschärfen für den Untertest Schlauchfiguren

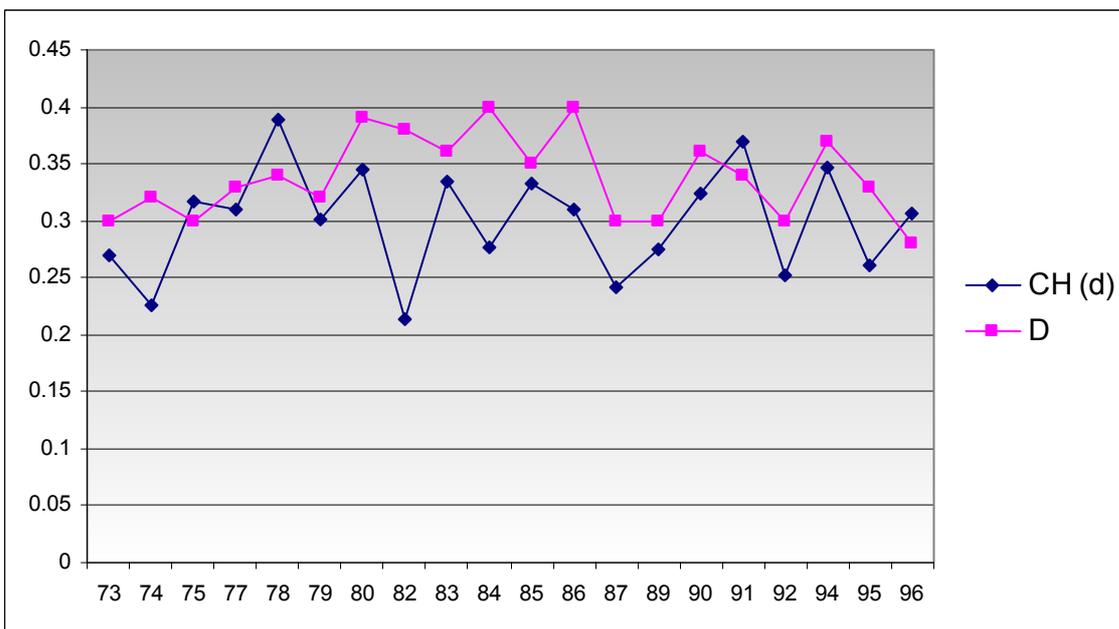


Abbildung 70: Trennschärfen für den Untertest Quantitative und formale Probleme

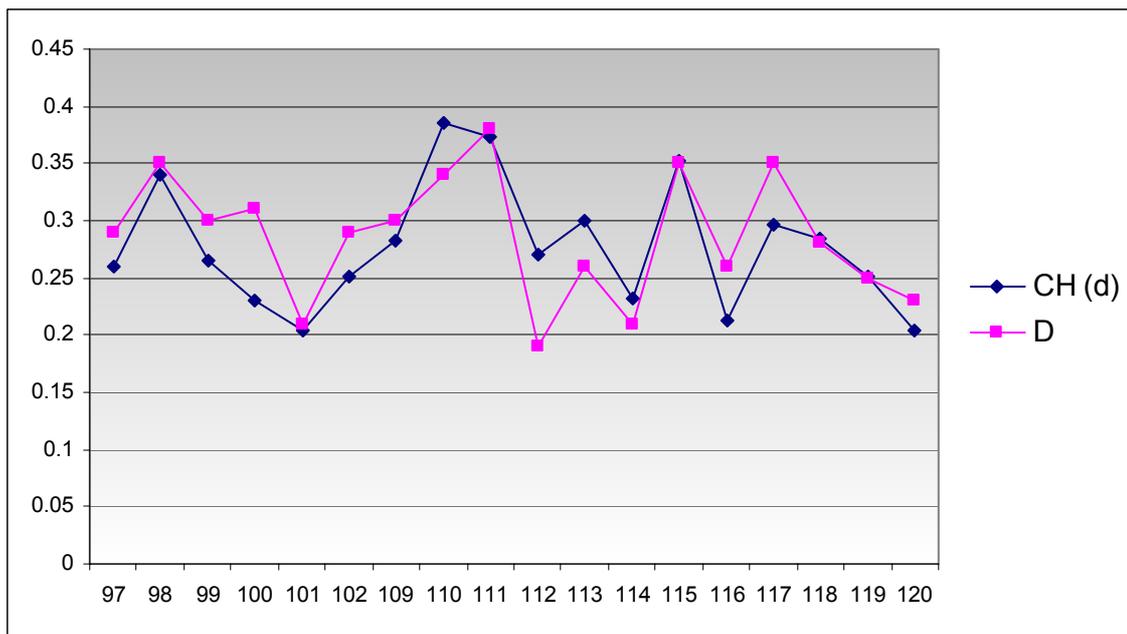


Abbildung 71: Trennschärfen für den Untertest Textverständnis

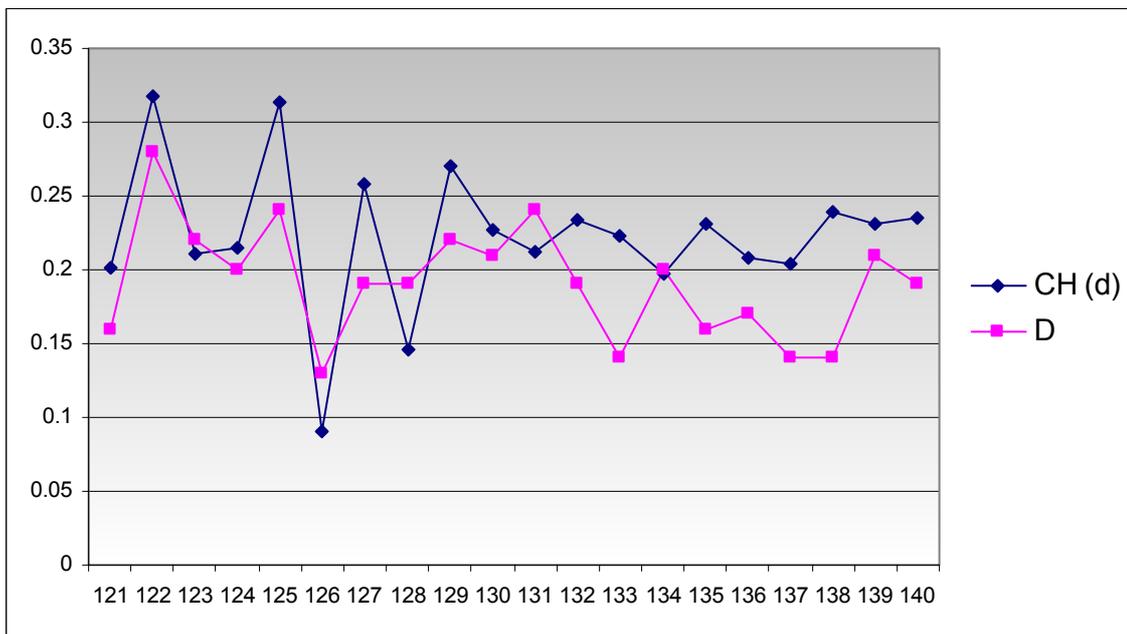
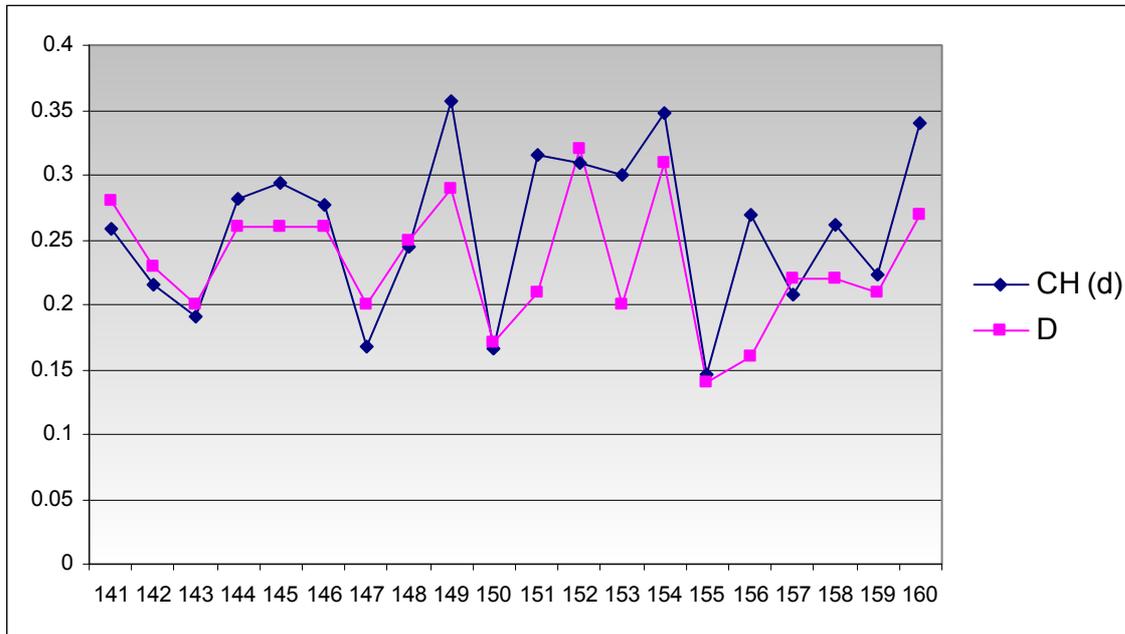
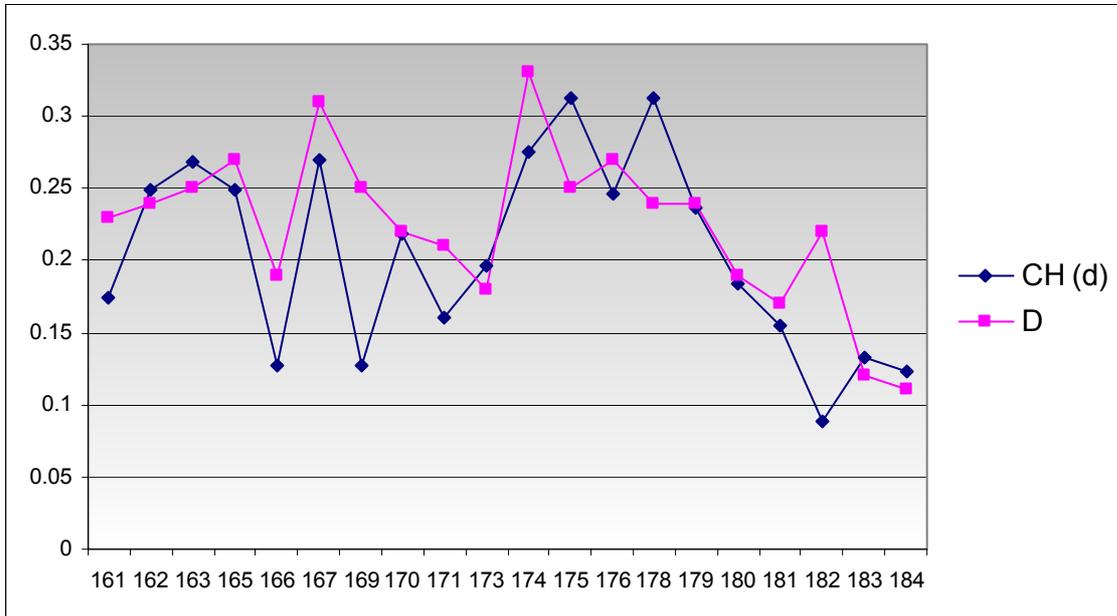


Abbildung 72: Trennschärfen für den Untertest Figuren lernen



**Abbildung 73:** Trennschärpen für den Untertest Fakten lernen



**Abbildung 74:** Trennschärpen für den Untertest Diagramme und Tabellen

## 9.4 Konzentriertes und sorgfältiges Arbeiten

Für diesen Untertest ist zunächst anzumerken, dass für den EMS 2001 erstmals mit einer neuen Einlesetechnologie gearbeitet wurde. Das Einlesen einer solchen Menge von Zeichen auf kleinem Raum stellt hohe Anforderungen an die Einlese- und Verarbeitungstechnologie. ZTD-interne Tests zeigten dabei gegenüber den Vorjahren eine weitere Steigerung der Zuverlässigkeit seitens der verwendeten Hardware und Software. Neben leistungsbedingten Schwankungen der Testteilnehmer können Unterschiede zu den Vorjahren also möglicherweise auch auf Unterschiede in der Technologie zurückzuführen sein.

In diesem Untertest kann die Zuverlässigkeit nicht nach dem Itemkonzept eingeschätzt werden. Der Untertest besteht aus 1200 Zeichen, die innerhalb vorgegebener Zeit bearbeitet werden müssen. Jedes Zeichen kann dabei als ein Item aufgefasst werden. Der Punktwert wird aufgrund der Menge der korrekt bearbeiteten Zeichen (total bearbeitete minus falsch bearbeitete Zeichen) berechnet. Dabei können zwei Fehlertypen auftreten: übersehene oder fälschlicherweise markierte Zeichen (Tabelle 90).

2001	Deutsch		Französisch		Italienisch	
	m	s	m	s	m	s
Richtig	420.37	109.99	338.72	83.54	350.98	87.43
Falsch	20.05	20.81	13.18	10.45	15.95	17.49
... übersehen	12	17.72	8.69	9.28	10.24	15.13
... fälschlich markiert.	8.05	9.01	4.49	4.49	5.71	5.77
<b>2000</b>						
Richtig	404.70	105.37	322.26	84.44	355.59	103.73
Falsch	16.41	20.61	12.87	36.50	9.03	14.92
... übersehen	12.95	19.27	10.65	36.43	6.59	12.88
... fälschlich markiert.	3.46	4.14	2.22	2.56	2.45	2.67
<b>1999</b>						
Richtig	391.24	98.11	308.12	93.35	352.16	75.76
Falsch	12.72	13.03	20.99	63.18	9.23	7.99
<b>1998</b>						
Richtig	393.38	95.55	345.74	97.57	359.48	73.05
Falsch	11.77	16.92	13.82	17.64	8.55	7.86

**Tabelle 90:** Kennwerte für den Untertest Konzentriertes und sorgfältiges Arbeiten

Anders als in den Vorjahren ist der Zusammenhang zwischen richtigen und falschen Markierungen diesmal signifikant (5%), wenn auch auf einem äusserst geringen Niveau (Tabelle 91). Die folgenden Abbildungen zeigen den Verlauf richtiger und falsch bearbeiteter Zeichen über die vergangenen Jahre, nach Sprachgruppen aufgeteilt. Die diesjährigen Ergebnisse folgen trotz der neuen Technologie dem Trend der Vorjahre. Dies spricht für im Endeffekt (früher durch zusätzliche Handauswertungen erreichte) vergleichbare Zuverlässigkeiten zwischen den beiden Einlese- und Auswertungskonzepten.

	Falsch	...übersehen	...fälschlich durchgestrichen
Richtig	.077* .022	-.021 .530	.221** .000
Falsch		.904** .000	.537** .000
...übersehen			.126** .000

Tabelle 91: Korrelationen der Parameter im Konzentrationstest; kursiv: Signifikanzniveau

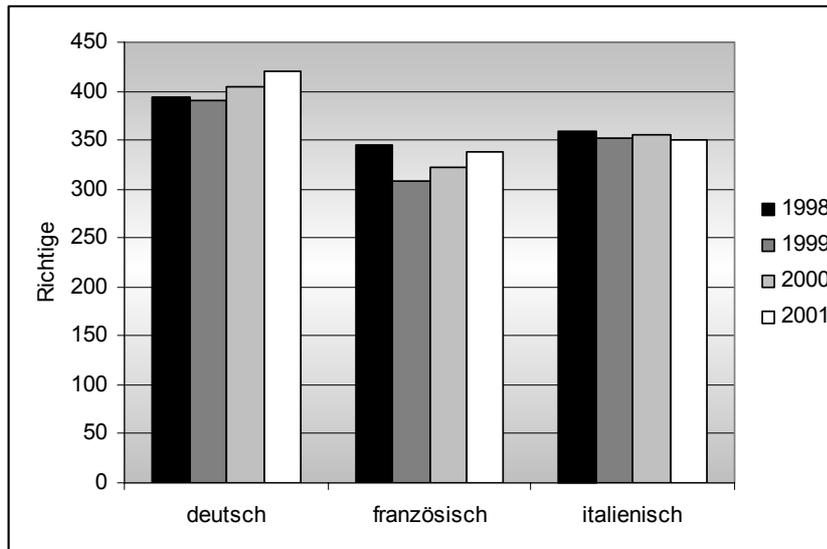


Abbildung 75: Richtige 1998 bis 2001 nach Sprachgruppen

Insbesondere bei der Darstellung der Fehler sind bei den Sprachminderheiten die geringen Bewerberzahlen zu beachten. Erfahrungsgemäss kann bereits eine einzige Person etwa durch falsch verstandene Instruktionen hier grosse Schwankungen in der durchschnittlichen Fehlerzahl verursachen.

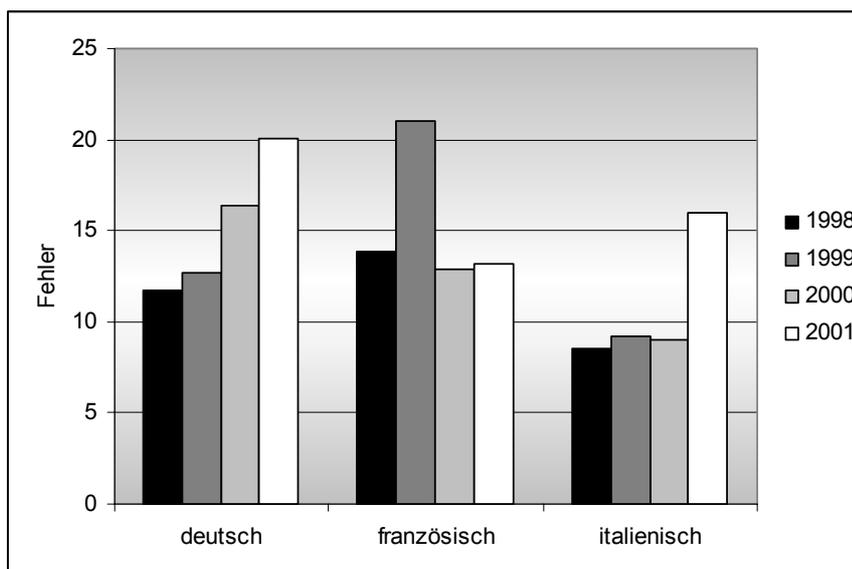


Abbildung 76: Fehler im Untertest Konzentriertes und sorgfältiges Arbeiten 1998 bis 2001 nach Sprachgruppen

## 10 Literatur

- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In Holland, P.W., Wainer, H. (Eds.), Differential Item Functioning. Hillsdale, NJ: Erlbaum.
- Beller, M., Gafni, N. (1995). Translated Versions of Israel's inter-university Psychometric Entrance Test (PET). In T. Oakland & R.K. Hambleton (Eds.), International Perspectives of Academic Assessment, S.207-218. Boston: Kluwer.
- Beller, M. (1996). Translating, equating and validating Scholastic Aptitude Tests: The Israeli Case. In Hänsgen, K.-D., Hofer, R. & Ruefli, D. (Hrsg). Eignungsdiagnostik und Medizinstudium, (S. 14-29), Bericht 2. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Celio, M.R. (1999) Ausbau des Medizinstudiums in Freiburg? Reformdruck als Chance der zweisprachigen Universität. Neue Zürcher Zeitung. 29.9. 1999 [http://www.nzz.ch/online/01\\_nzz\\_aktuell/sonderbeilagen/studium98/studium98celio.htm](http://www.nzz.ch/online/01_nzz_aktuell/sonderbeilagen/studium98/studium98celio.htm)
- Cook, L.L. (1998). Can Scores Obtained on Test Given in Different Languages to Examinees of Different Cultures be Equally Valid? ICAP San Francisco.
- Deidesheimer Kreis (1997). Hochschulzulassung und Studieneignungstests: studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus-Clausus und andere Studiengänge. Göttingen, Zürich: Vandenhoeck und Ruprecht.
- Ebach, J., Trost, G. (1997). Admission to Medical Schools in Europe. Lengerich: Pabst.
- Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus Clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 1: Erfahrungen mit dem EMS als Zulassungskriterium. Schweizerische Ärztezeitung Heft 12 S. 666 – 672.
- Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus Clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 2: EMS und Chancengleichheit. Schweizerische Ärztezeitung Heft 13 S. 723-730.
- Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995a). Un test d'aptitudes aux études de médecine est-il faisable en Suisse? Bulletin des médecins suisses, 7, S. 267 - 274.
- Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995b). Der Eignungstest für das Medizinstudium in der Schweiz. Schweizerische Ärztezeitung, 37, S. 1476 - 1496.
- Oswald, U. (1999). Der Eignungstest 1998 für das Medizinstudium. Schweizerische Ärztezeitung 80, S. 1313 – 1317.
- 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.

ZVS Press: Pressedienst der Zentralstelle für die Vergabe von Studienplätzen (1998). Deutlicher Anstieg der Bewerbungen für das Medizin-Studium. Dortmund: ZVS.

## 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, Université de Fribourg (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 Frühere Berichte des ZTD

Alle Berichte auch im Internet: <http://www.unifr.ch/ztd/ems/welcome.htm>

### Bericht 6 (2000)

#### **EMS Eignungstest für das Medizinstudium in der Schweiz 2000**

*Bericht über die Durchführung und Ergebnisse, Evaluation der Vorhersage von Prüfungserfolg – Zusammenfassender Vergleich 1998 bis 2000 - avec un résumé en français*

K.-D. Hänsgen und B. Spicher

<b>Bericht 5 (1999)</b>	
<b>EMS Eignungstest für das Medizinstudium in der Schweiz 1999</b> <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher	
<b>Bericht 4 (1998)</b>	
<b>EMS Eignungstest für das Medizinstudium in der Schweiz 1998</b> <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher	
<b>Bericht 3 (1997) - Rapport 3 (1997)</b>	
<b>Eignungsprüfung für das Medizinstudium - Kriterien und Testverfahren</b> <i>Bericht über das Internationale Symposium in Bern am 8. November 1996</i> Hrsg. von K.-D. Hänsgen und N. Ischi	<b>L'examen d'aptitude aux études de médecine en Suisse - Critères et procédés d'application du test</b> <i>Rapport sur le Symposium international à Berne</i> Editeurs: K.-D. Hänsgen; N. Ischi
<p><b>Günter Trost</b> Eignungskriterien bei der Zulassung zum Medizinstudium in Europa: Ergebnisse einer Erhebung</p> <p><b>Ingemar Wedman &amp; Widar Henriksson</b> The Swedish Scholastic Aptitude Test. Research and main findings</p> <p><b>John L. Hackett</b> The Medical College Admission Test (MCAT) - its use in U.S. and Canada and some results of validation</p> <p><b>Piet J. Janssen</b> Admission to the study in medicine in Belgium: two 'different' solutions to the 'same' problem; reflections of a Flemish school psychologist</p> <p><b>Klaus-Dieter Hänsgen</b> Der Eignungstest für das Medizinstudium in der Schweiz als Instrument für die Studienzulassung</p> <p><b>Rainer Hofer &amp; Klaus-Dieter Hänsgen</b> Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium</p> <p><b>Urs Schallberger</b> Anforderungen an das Zulassungsverfahren für das Medizinstudium in der Schweiz: Leitlinien</p>	

<b>Bericht 2 (1996) - Rapport 2 (1996)</b>	
<p><b>Eignungsdiagnostik und Medizinstudium</b> <i>Symposiumsbericht</i> Hrsg. von K.-D. Hänsgen, R. Hofer und D. Ruefli</p>	<p><b>Diagnostic d'aptitudes et études de médecine -</b> <i>Rapport d'un symposium,</i> Edité par K.-D. Hänsgen, R. Hofer et D. Ruefli</p>
<p><b>Klaus-D. Hänsgen</b> Vorwort: Eignungstests und Medizinstudium</p> <p><b>Christina Stage</b> Experiences with the Swedish Scholastic Aptitude Test</p> <p><b>Michal Beller</b> Translating, equating and validating Scholastic Aptitude Tests: The Israeli Case</p> <p><b>Eckhard Klieme</b> Erfolgsprognose in medizinischen Studiengängen - Zur Validität des Tests für medizinische Studiengänge und anderer Auswahlinstrumente</p> <p><b>Günter Trost</b> Testergebnisse versus Schulnoten als Auswahlkriterien: Paternoster-Effekt, Filter-Effekt, Kosten-Nutzen-Effekte und Auswirkungen auf die Fairness der Zulassung</p> <p><b>Urs Schallberger</b> Nutzen, Fairness, Validität und Akzeptanz von Selektionsverfahren</p> <p><b>Rainer Hofer, Daniel Ruefli &amp; Klaus-D. Hänsgen</b> Der "Test des Tests" - Ergebnisse eines Probelaufs des Eignungstests in der Schweiz in deutscher und französischer Sprache</p>	

<b>Bericht 1 (1995) - Rapport 1 (1995)</b>	
<p><b>Der Eignungstest für das Medizinstudium in der Schweiz - ein Probelauf</b> Rainer Hofer, Daniel Ruefli &amp; Klaus-D. Hänsgen</p>	<p><b>Le test d'aptitudes pour les études de médecine en Suisse - Epreuve d'essai</b> Rainer Hofer, Daniel Ruefli &amp; Klaus-D. Hänsgen</p>

## **11 Anhang zum Bericht 2001**

## 11.1 Beispielaufgaben für die Untertests

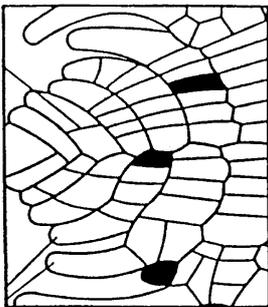
Nachfolgend wird pro Untertest eine Beispielaufgabe dargestellt. So können lediglich die Prinzipien der Aufgabenstruktur verdeutlicht werden – die Aufgaben unterscheiden sich innerhalb jedes Untertests bezüglich des Schwierigkeitsgrades und der Anforderung.

### 11.1.1 Untertest: Muster zuordnen

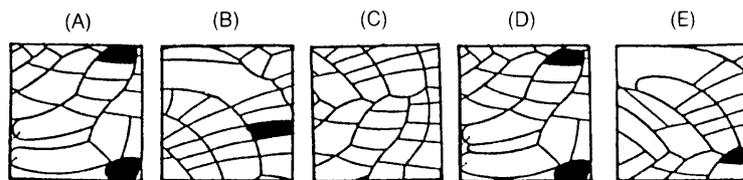
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wiederzuerkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

### 11.1.2 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 sogenannten 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:

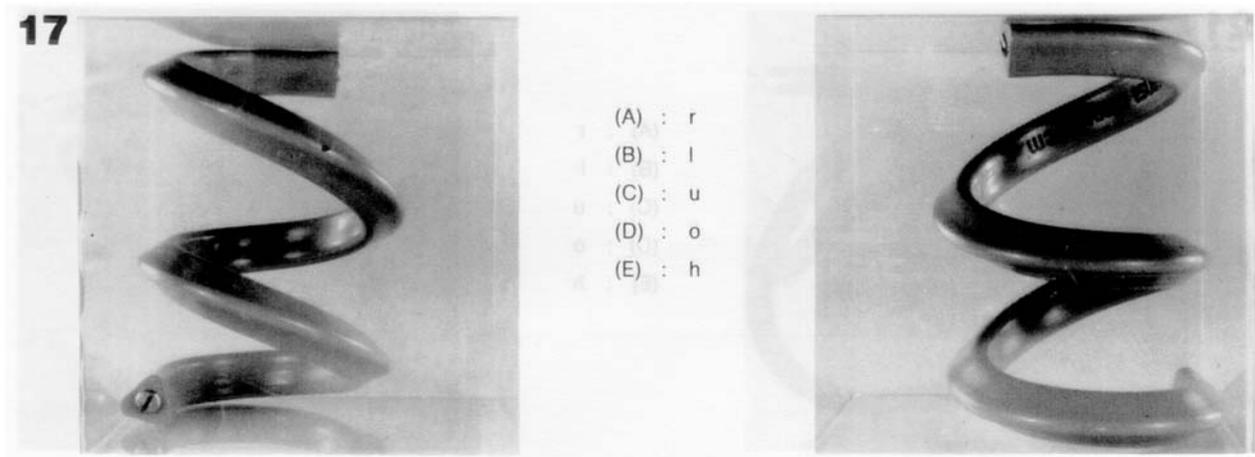
- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind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

### 11.1.3 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!)

### 11.1.4 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.

### 11.1.5 Untertest: Konzentriertes und sorgfältiges Arbeiten

Bei diesem Untertest soll die Fähigkeit, rasch, sorgfältig und konzentriert zu arbeiten, gemessen werden. Dabei sollen möglichst alle b, die mit zwei Querstrichen versehen sind, die entweder beide unten, beide oben oder je einer unten und oben angebracht sind, markiert werden. Die Lösungsmenge ist ebenso wichtig wie die Fehlerfreiheit der Bearbeitung. Dieser Test ist trainierbar – im Test-Info wird darauf hingewiesen, diesen Untertest vor der Testabnahme mehrfach zu üben.

b   b   b

Diese Buchstaben b mit zwei Querstrichen sind eingestreut unter b mit einem, drei oder vier Querstrichen sowie unter q mit einem oder mehreren Querstrichen. Im folgenden Beispiel wären also das 1., 4., 6., 8., 9. und 13. Zeichen zu markieren.

b   q   b   b   b   b   b   b   b   b   q   b   b   b   b   q

### 11.1.6 Untertest: Figuren lernen

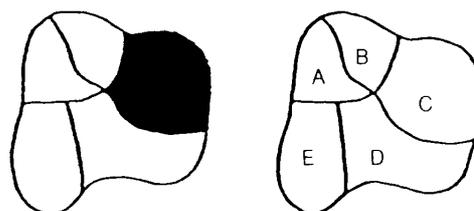
Für beide nachfolgende Gedächtnistests wird nach der Mittagspause das Material zum Einprägen ausgeteilt. Vor der Abfrage des Gelernten wird der Untertest „Textverständnis“ bearbeitet, damit liegt die Zeit des Behaltens der gelernten Inhalte über einer Stunde. Gedächtnisleistungen sind wichtige Voraussetzungen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

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

Ein Beispiel:

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.

### 11.1.7 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

### 11.1.8 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 sogenannter 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 sogenannte Tyrosin-Reste (des Thyreoglobulins) wird zunächst ein Jodatome 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 Rückkopplungsmechanismus: 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.

### 11.1.9 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 neun 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.

Wenn der Test die Studieneignung erfassen will, ist seine Struktur natürlich abhängig von der Entwicklung dieser Studienanforderungen. In der Schweiz werden solche Veränderungen gegenwärtig in Reformkonzepten diskutiert und bereits umgesetzt. Sozial-kommunikative Eignung, problemorientiertes Lernen oder mehr Verschränkung von Grundlagen und Anwendung werden hier genannt. Es werden auch zukünftig nicht alle der heute erfassten Anforderungen weniger wichtig für Studienerfolg werden – ohne intellektuelle Eignung ist die erfolgreiche Absolvierung eines Studiums kaum denkbar. Dennoch ist zu prüfen, ob sich andere Aspekte der Eignung gleichfalls erfassen lassen, um den Test ggf. zu aktualisieren.

Bei allen Aktualisierungen werden zwei Zielstellungen zu beachten sein:

- Der Aufwand für die Erfassung muss tragbar sein (z.B. wäre es unrealistisch, mit allen Bewerbern Eignungsgespräche durchzuführen).
- Die Trainierbarkeit der Eignungsabklärung muss gering bleiben – da andernfalls der Erfolg vom Trainingsaufwand und damit auch von finanziellen Möglichkeiten abhängig wird. Leider beruhen alle ernstzunehmenden Methoden der Eignungsfeststellung für den sozial-kommunikativen Bereich auf Anforderungen, die zumeist recht gut trainiert werden können.

## 11.2 Weitere Detailergebnisse

Gesamt	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen			1	2	10	28	52	57	109	105	124	121	94	77	45	24	19	8	3	1	
Med.-nat. Grundv.		5	6	14	19	37	51	74	86	91	111	84	87	86	43	31	17	18	13	7	
Schlauchfiguren				4	8	11	15	21	42	56	62	65	68	78	82	93	82	72	64	33	24
Qua.form Probl.		4	8	12	19	33	48	56	60	88	65	101	84	76	60	43	40	45	21	10	7
Textverständnis	2	5	10	37	47	82	92	85	96	101	83	70	52	39	25	27	11	12	4		
Figuren lernen		2	6	1	21	30	48	65	68	96	107	102	68	71	74	43	40	18	9	10	1
Fakten lernen			3	6	21	25	48	49	99	91	101	94	80	63	55	43	32	29	16	17	8
Diagr und Tabellen		1	1	5	17	23	60	74	89	90	107	101	95	76	52	46	28	9	3	2	1
Kon.sorgf .Arbeiten					2	5	10	33	42	72	72	88	100	80	68	69	59	42	44	51	43

Tabelle A 1: Häufigkeitsverteilung der Punkte Gesamt

Deutsch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen			1	1	8	23	46	48	94	87	106	111	87	68	40	22	19	8	3	1	
Med.-nat. Grundv.		4	5	10	15	26	38	59	72	81	100	76	83	81	40	28	17	18	13	7	
Schlauchfiguren				4	5	9	10	13	34	40	51	56	55	71	75	87	75	68	63	33	24
Qua.form Probl.		3	4	11	16	26	40	44	49	81	52	91	77	70	57	38	37	44	18	8	7
Textverständnis	1	3	8	32	37	71	76	77	81	89	74	66	46	35	25	27	11	11	3		
Figuren lernen		2	5	1	15	21	39	54	62	85	93	88	59	64	72	41	38	16	8	9	1
Fakten lernen			2	6	19	21	43	43	85	81	82	84	71	53	48	40	30	26	14	17	8
Diagr und Tabellen				3	12	20	44	63	71	73	98	93	92	67	50	46	28	9	1	2	1
Kon.sorgf .Arbeiten					2	4	8	24	35	53	57	75	89	73	61	58	57	41	42	51	43

Tabelle A 2: Häufigkeitsverteilung der Punkte deutsche Sprachgruppe

Französisch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen					2	4	4	4	10	9	12	5	4	5	4	2					
Med.-nat. Grundv.		1	1	2	2	7	7	10	10	5	7	3	3	3	2	2					
Schlauchfiguren					1	1	2	6	6	10	5	4	8	5	4	3	5	4	1		
Qua.form Probl.		1	2	1	3	4	5	8	5	3	11	4	4	4	1	3	3		2	1	
Textverständnis	1	2	1	3	6	6	10	5	8	9	5	2	4	3							
Figuren lernen			1		5	3	4	8	2	7	10	9	6	5		1	1	2		1	
Fakten lernen			1		2	3	3	1	10	6	13	6	5	7	4	1		1	2		
Diagr und Tabellen			1	1	3	1	11	7	9	9	7	6	2	7	1						
Kon.sorgf .Arbeiten						1	1	6	4	14	7	10	4	4	4	7	1	1	1		

Tabelle A 3: Häufigkeitsverteilung der Punkte französische Sprachgruppe

Italienisch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen				1		1	2	5	5	9	6	5	3	4	1						
Med.-nat. Grundv.				2	2	4	6	5	4	5	4	5	1	2	1	1					
Schlauchfiguren					2	1	3	2	2	6	6	5	5	2	3	3	2				
Qua.form Probl.			2			3	3	4	6	4	2	6	3	2	2	2		1	1	1	
Textverständnis			1	2	4	5	6	3	7	3	4	2	2	1				1	1		
Figuren lernen					1	6	5	3	4	4	4	5	3	2	2	1	1		1		
Fakten lernen						1	2	5	4	4	6	4	4	3	3	2	2	2			
Diagr und Tabellen		1		1	2	2	5	4	9	8	2	2	1	2	1				2		
Kon.sorgf .Arbeiten							1	3	3	5	8	3	7	3	3	4	1		1		

Tabelle A 4: Häufigkeitsverteilung der Punkte italienische Sprachgruppe

Human- medizin	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen			1	2	8	25	38	46	86	85	94	99	79	64	38	19	17	6	1	1	
Med.-nat. Grundv.		5	4	10	12	26	39	56	64	74	88	76	71	71	36	25	17	18	11	6	
Schlauch- figuren				2	6	8	11	16	29	43	51	50	45	64	63	83	67	66	57	29	19
Qua.form Probl.		2	7	7	14	22	39	41	43	72	52	82	71	63	56	33	32	40	17	9	7
Textver- ständnis	2	2	5	28	34	68	65	68	73	82	72	59	45	32	22	25	11	12	4		
Figuren lernen		2	6	1	16	24	36	51	54	76	84	77	58	59	66	37	34	14	7	6	1
Fakten lernen			2	5	16	22	38	39	79	76	72	75	60	52	47	35	29	27	15	12	8
Diagr und Tabellen		1	1	4	15	16	47	54	74	67	93	76	76	57	48	40	26	9	3	1	1
Kon.sorgf .Arbeiten					2	4	9	21	33	44	59	68	81	68	58	53	51	36	40	44	38

Tabelle A 5: Häufigkeitsverteilung der Punkte Humanmedizin

Veterin.- medizin	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen					2	3	14	11	23	20	30	22	15	13	7	5	2	2	2		
Med.-nat. Grundv.			2	4	7	11	12	18	22	17	23	8	16	15	7	6			2	1	
Schlauch- figuren				2	2	3	4	5	13	13	11	15	23	14	19	10	15	6	7	4	5
Qua.form Probl.		2	1	5	5	11	9	15	17	16	13	19	13	13	4	10	8	5	4	1	
Textver- ständnis		3	5	9	13	14	27	17	23	19	11	11	7	7	3	2					
Figuren lernen					5	6	12	14	14	20	23	25	10	12	8	6	6	4	2	4	
Fakten lernen			1	1	5	3	10	10	20	15	29	19	20	11	8	8	3	2	1	5	
Diagr und Tabellen				1	2	7	13	20	15	23	14	25	19	19	4	6	2			1	
Kon.sorgf .Arbeiten						1	1	12	9	28	13	20	19	12	10	16	8	6	4	7	5

Tabelle A 6: Häufigkeitsverteilung der Punkte Veterinärmedizin

	Humanmedizin		Veterinärmedizin			Humanmedizin		Veterinärmedizin	
	Punktwert	Testwert	Punktwert	Testwert		Punktwert	Testwert	Punktwert	Testwert
34	1				101	16	21	2	9
42	2				102	15	32	2	3
43	1				103	13	27	1	7
44	1				104	15	31	5	5
47	2				105	17	34	2	2
50	1				106	14	23	3	4
52	2		2		107	16	17	1	1
54	3				108	17	22	1	4
55	1		2		109	9	23	2	4
56	1				110	15	20	2	3
57	2		1		111	8	16		2
58			1		112	8	14	1	1
59	1				113	16	14	1	2
61			1		114	6	14	3	3
62	5		1		115	9	7	1	
63	5				116	14	9	4	1
64	5				117	10	11	1	2
65	2		1		118	10	9	1	1
66	5		2		119	9	7	1	1
67	2		6		120	7	6	1	
68	3		2		121	8	1		
69	5		2		122	6	1	1	1
70	6	1	3		123	5	1	2	
71	4		3		124	9	1		
72	3	3	1		125	6	2	2	
73	7	1			126	8		1	
74	4	2	1		127	5	2		
75	11		1		128	2	2		
76	6	1	2		129	3		1	
77	7	2	6	2	130	6			
78	4	3	1	1	131	4		2	
79	9	2	2	1	132	7			
80	6	3		2	133	6			
81	12		4		134	3		1	
82	8	10	3	1	135	3		1	
83	12	4	3	1	136	4			
84	8	8	3	6	137	3			
85	20	9	3	7	138	3			
86	11	10	7	5	139	1			
87	12	9	7	2	140	1		1	
88	15	14	9	2	143	1			
89	9	14	6	4	144	1			
90	10	12	4	5	146	2			
91	11	20	4	4	150	2			
92	14	19	4	7	152	2			
93	10	9	4	2					
94	19	32		11					
95	13	26	2	14					
96	21	16	6	11					
97	12	27	5	10					
98	9	30		4					
99	9	35	7	5					
100	12	20	2	8					

Tabelle A 7: Punkt- und Testwertverteilung nach Disziplinen

	Punkt- wert	Test- wert	Mus- ter- zuor- nen	Med.- naturw. Grundv.	Schlauch- figuren	Quantit. formale Proble- me	Text- verständ- nis	Figur- lernen	Fak- ten- ler- nen	Diagram- me und Tabellen	Konz. und sorgf. A.
Total											
Mittelwert	97,64	100,03	10,14	10,07	13,27	10,91	8,52	10,60	10,78	10,27	13,07
N	880	880	880	880	880	880	880	880	880	880	880
Standardabw.	19,762	9,985	2,812	3,466	3,743	3,912	3,462	3,469	3,654	3,134	3,702
Kurtosis	-,125	-,127	-,187	-,141	-,527	-,479	-,292	-,354	-,303	-,408	-,792
Skewness	-,047	-,051	,124	,080	-,332	-,017	,325	,065	,287	,069	,145
Median	97,00	100,00	10,00	10,00	14,00	11,00	8,00	10,00	10,00	10,00	13,00
Deutsch											
Mittelwert	99,47	100,89	10,23	10,34	13,61	11,11	8,68	10,77	10,85	10,52	13,36
N	773	773	773	773	773	773	773	773	773	773	773
Standardabw.	19,392	9,859	2,828	3,436	3,671	3,861	3,464	3,442	3,708	3,072	3,714
Kurtosis	-,033	-,021	-,194	-,068	-,369	-,470	-,377	-,337	-,350	-,482	-,813
Skewness	-,105	-,120	,125	,041	-,421	-,047	,306	,032	,285	,042	,065
Median	100,00	101,00	10,00	10,00	14,00	11,00	9,00	11,00	11,00	11,00	13,00
Französisch											
Mittelwert	84,11	93,58	9,46	8,08	11,12	9,37	7,17	9,48	10,08	8,54	10,82
N	65	65	65	65	65	65	65	65	65	65	65
Standardabw.	16,728	8,374	2,745	3,114	3,444	4,106	3,085	3,469	3,232	2,750	2,883
Kurtosis	1,217	,918	-,519	-,158	-,861	-,296	-,414	,191	,374	-,478	-,445
Skewness	,217	,262	,064	,263	,181	,284	-,072	,272	,098	-,001	,417
Median	84,00	94,00	9,00	8,00	11,00	10,00	7,00	10,00	10,00	8,00	10,00
Italienisch											
Mittelwert	84,81	94,17	9,36	8,19	10,36	9,74	7,69	9,21	10,69	8,43	11,14
N	42	42	42	42	42	42	42	42	42	42	42
Standardabw.	18,267	9,095	2,377	2,973	3,161	3,902	3,468	3,418	3,189	3,408	2,755
Kurtosis	-,046	-,187	,104	-,502	-,519	,041	1,401	-,309	-,750	1,797	-,346
Skewness	,459	,420	-,231	,291	-,136	,339	,996	,520	,296	,840	,314
Median	83,00	94,00	9,00	8,00	10,00	9,00	7,50	9,00	10,00	8,00	11,00
Humanmedizin											
Mittelwert	99,09	100,74	10,16	10,28	13,51	11,14	8,77	10,66	10,88	10,39	13,31
N	709	709	709	709	709	709	709	709	709	709	709
Standardabw.	19,894	10,059	2,821	3,469	3,702	3,871	3,502	3,475	3,722	3,186	3,704
Kurtosis	-,035	-,045	-,233	-,094	-,513	-,440	-,346	-,369	-,414	-,430	-,778
Skewness	-,135	-,137	,063	,042	-,394	-,038	,313	-,023	,267	,041	,069
Median	100,00	101,00	10,00	10,00	14,00	11,00	9,00	11,00	11,00	10,00	13,00
Veterinärmedizin											
Mittelwert	91,60	97,06	10,02	9,20	12,29	9,98	7,46	10,36	10,37	9,81	12,09
N	171	171	171	171	171	171	171	171	171	171	171
Standardabw.	18,046	9,116	2,782	3,325	3,768	3,956	3,084	3,445	3,334	2,870	3,542
Kurtosis	-,105	-,105	,100	-,190	-,363	-,558	-,389	-,128	,294	-,319	-,563
Skewness	,252	,232	,390	,216	-,091	,103	,207	,440	,314	,121	,469
Median	89,00	96,00	10,00	9,00	12,00	10,00	7,00	10,00	10,00	10,00	12,00

Tabelle A 8: Verteilungskennwerte für Punktwert, Testwert und Untertests nach Sprachgruppen und Disziplinen

		<b>Quadrat- summe</b>	<b>df</b>	<b>MQ</b>	<b>F</b>	<b>Sig.</b>
Testwert	Zwischen	4710,799	2	2355,400	24,911	,000
	Innerhalb	82923,600	877	94,554		
	Total	87634,399	879			
Muster zuordnen	Zwischen	62,493	2	31,246	3,978	,019
	Innerhalb	6888,415	877	7,855		
	Total	6950,908	879			
Med.-naturwiss. Grundverständnis	Zwischen	461,404	2	230,702	20,037	,000
	Innerhalb	10097,640	877	11,514		
	Total	10559,044	879			
Schlauchfiguren	Zwischen	745,639	2	372,819	28,254	,000
	Innerhalb	11572,451	877	13,195		
	Total	12318,090	879			
Quant. und formale Probleme	Zwischen	242,091	2	121,045	8,035	,000
	Innerhalb	13211,345	877	15,064		
	Total	13453,436	879			
Textverständnis	Zwischen	166,448	2	83,224	7,040	,001
	Innerhalb	10367,261	877	11,821		
	Total	10533,709	879			
Figuren lernen	Zwischen	186,262	2	93,131	7,858	,000
	Innerhalb	10394,120	877	11,852		
	Total	10580,382	879			
Fakten lernen	Zwischen	35,966	2	17,983	1,348	,260
	Innerhalb	11697,579	877	13,338		
	Total	11733,544	879			
Diagramme und Tabellen	Zwischen	385,620	2	192,810	20,503	,000
	Innerhalb	8247,379	877	9,404		
	Total	8632,999	879			
Konzentr. u. sorgf. Arbeiten	Zwischen	553,581	2	276,791	21,119	,000
	Innerhalb	11494,050	877	13,106		
	Total	12047,632	879			

Tabelle A 9: Varianzanalytische Prüfung des Einflusses der Sprache auf Testwert und die Punktwerte

Variable	(I) Sprache	(J) Sprache	Mittlere Differenz (I-J)	Standardfehler	Sig.	95% Konfidenzintervall	
						Lower	Upper
Testwert	deutsch	französisch	7,30	1,256	,000	4,35	10,25
		italienisch	6,72	1,541	,000	3,10	10,34
	französisch	deutsch	-7,30	1,256	,000	-10,25	-4,35
		italienisch	-,58	1,925	,951	-5,10	3,94
	italienisch	deutsch	-6,72	1,541	,000	-10,34	-3,10
		französisch	,58	1,925	,951	-3,94	5,10
Muster zuordnen	deutsch	französisch	,77	,362	,084	-,08	1,62
		italienisch	,88	,444	,119	-,17	1,92
	französisch	deutsch	-,77	,362	,084	-1,62	,08
		italienisch	,10	,555	,981	-1,20	1,41
	italienisch	deutsch	-,88	,444	,119	-1,92	,17
		französisch	-,10	,555	,981	-1,41	1,20
Med.-naturwiss. Grundverständnis	deutsch	französisch	2,26	,438	,000	1,23	3,29
		italienisch	2,15	,538	,000	,88	3,41
	französisch	deutsch	-2,26	,438	,000	-3,29	-1,23
		italienisch	-,11	,672	,984	-1,69	1,46
	italienisch	deutsch	-2,15	,538	,000	-3,41	-,88
		französisch	,11	,672	,984	-1,46	1,69
Schlauchfiguren	deutsch	französisch	2,49	,469	,000	1,39	3,59
		italienisch	3,25	,576	,000	1,90	4,60
	französisch	deutsch	-2,49	,469	,000	-3,59	-1,39
		italienisch	,77	,719	,536	-,92	2,45
	italienisch	deutsch	-3,25	,576	,000	-4,60	-1,90
		französisch	-,77	,719	,536	-2,45	,92
Quant. und formale Probleme	deutsch	französisch	1,74	,501	,002	,56	2,91
		italienisch	1,37	,615	,067	-,07	2,81
	französisch	deutsch	-1,74	,501	,002	-2,91	-,56
		italienisch	-,37	,768	,881	-2,17	1,44
	italienisch	deutsch	-1,37	,615	,067	-2,81	,07
		französisch	,37	,768	,881	-1,44	2,17
Textverständnis	deutsch	französisch	1,51	,444	,002	,46	2,55
		italienisch	,99	,545	,167	-,29	2,27
	französisch	deutsch	-1,51	,444	,002	-2,55	-,46
		italienisch	-,52	,681	,724	-2,12	1,08
	italienisch	deutsch	-,99	,545	,167	-2,27	,29
		französisch	,52	,681	,724	-1,08	2,12
Figuren lernen	deutsch	französisch	1,30	,445	,010	,25	2,34
		italienisch	1,56	,545	,012	,28	2,84
	französisch	deutsch	-1,30	,445	,010	-2,34	-,25
		italienisch	,26	,682	,921	-1,34	1,86
	italienisch	deutsch	-1,56	,545	,012	-2,84	-,28
		französisch	-,26	,682	,921	-1,86	1,34
Fakten lernen	deutsch	französisch	,77	,472	,232	-,34	1,88
		italienisch	,16	,579	,960	-1,20	1,52
	französisch	deutsch	-,77	,472	,232	-1,88	,34
		italienisch	-,61	,723	,673	-2,31	1,08
	italienisch	deutsch	-,16	,579	,960	-1,52	1,20
		französisch	,61	,723	,673	-1,08	2,31
Diagramme und Tabellen	deutsch	französisch	1,98	,396	,000	1,05	2,91
		italienisch	2,09	,486	,000	,95	3,23
	französisch	deutsch	-1,98	,396	,000	-2,91	-1,05
		italienisch	,11	,607	,982	-1,32	1,54
	italienisch	deutsch	-2,09	,486	,000	-3,23	-,95
		französisch	-,11	,607	,982	-1,54	1,32
Konzentr. u. sorgf. Arbeiten	deutsch	französisch	2,55	,468	,000	1,45	3,65
		italienisch	2,22	,574	,000	,88	3,57
	französisch	deutsch	-2,55	,468	,000	-3,65	-1,45
		italienisch	-,33	,717	,891	-2,01	1,36
	italienisch	deutsch	-2,22	,574	,000	-3,57	-,88
		französisch	,33	,717	,891	-1,36	2,01

Tabelle A 10: Multiple Mittelwertsprüfung für Sprachunterschiede

## Das Zentrum für Testentwicklung und Diagnostik im Internet:

[www.unifr.ch/ztd](http://www.unifr.ch/ztd)

Untitled Document - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

← Zurück → Suchen Favoriten Medien » Google » Links

Adresse <http://www.unifr.ch/ztd/ems/welcome.htm> Wechseln zu

Eignungstest für das Medizinstudium in der Schweiz  
Test d'aptitudes pour les études de médecine en Suisse  
Test attitudinale per lo studio della medicina in Svizzera

E M S  
A M S

**Kandidatinnen & Kandidaten**

- Termine & Ablauf
- Wie vorbereiten?
- Fragen & Antworten
- CRUS (mehr Info)

**Information des ZTD für Teilnehmende am EMS**

**EIGNUNGSTEST 2001 am 6.7. durchgeführt**

Am 6. Juli 2001 wurde für die Studienbewerberinnen und Studienbewerber für Human- und Veterinärmedizin der Universitäten Basel, Bern, Freiburg und Zürich ein Eignungstest in deutscher, französischer und italienischer Sprache durchgeführt.

In der Zahnmedizin steht weiterhin eine ausreichende Zahl von Studienplätzen zur Verfügung, so dass kein Numerus clausus (NC) dort notwendig ist.

Diese Situation entspricht der von 1999 und 2000, im Jahre 1998 wurde der NC nur für Humanmedizin notwendig.

Der folgende Bericht fasst den aktuellen Stand zusammen:

- [deutsch](#)
- [français](#)

Die Kandidatinnen und Kandidaten erhalten in Kürze einen Testbescheid und eine Information über die Zulassung. Bitte vergessen Sie nicht, die Zulassung ggf. rechtzeitig zu bestätigen!

Für die Organisation des Eignungstests ist ab 2001 die [Rektorenkonferenz der Schweizer Universitäten \(CRUS\)](#) verantwortlich. Die Bereitstellung des Tests, die fachliche Betreuung der Durchführung und die Auswertung sind die Aufgaben des [ZTD](#).

ZTD 13.07.2001

**Weitere Informationen der CRUS**

**CRUS**  
Rektorenkonferenz der Schweizer Universitäten  
Conférence des Recteurs des Universités Suisses  
Conferenza dei Rettori delle Università Svizzere

**E-MAIL : [ZTD@unifr.ch](mailto:ZTD@unifr.ch)**

**Weitere Informationen des ZTD**

Internet