

EMS • Eignungstest für das Medizinstudium in der Schweiz • 2000

*Bericht über die Durchführung und Ergebnisse
Evaluation der Vorhersage von Prüfungserfolg
Zusammenfassende Vergleiche 1998 bis 2000*

avec un résumé en français

K.-D. Hänsgen und B. Spicher

Bericht 6 (2000)



**Zentrum für Testentwicklung
und Diagnostik** am Departement für Psychologie
der Universität Freiburg-Schweiz



Hänsgen, Klaus-Dieter; Spicher, Benjamin (2000).

EMS Eignungstest für das Medizinstudium in der Schweiz 2000

Bericht über die Durchführung und Ergebnisse

Berichte des Zentrums für Testentwicklung, Nummer 6

im Auftrag der Schweizerischen Hochschulkonferenz

Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik, 2000

unter Mitarbeit von Sabine Dobler, Sébastien Simonet,

Nicolas Burger, Steve Fürst, Maurizio Strazzeri

Redaktion: Valérie Favre

Die Evaluation des EMS 1998 erfolgte im Auftrag des Beirates „Eignungstest“ der Schweizerischen Hochschulkonferenz. Der Evaluationskommission gehören an: Prof. R. Bloch (i.A. des Beirates „Eignungstest“ der SHK), K. Wechsler (i.A. der SHK), H. Neuhaus (i.A. des BAG), K. Holenstein (i.A. des BfS), Prof. K.-D. Hänsgen (i.A. des ZTD).

Für die Unterstützung bei der Datenerhebung im BAG danken wir Herrn H. Neuhaus und Herrn D. Megert ganz herzlich.

Inhaltsverzeichnis

ZUSAMMENFASSUNG	5
RESUME	10
EINLEITUNG.....	15
NUMERUS CLAUSUS UND MEDIZINSTUDIUM	16
ANMELDUNG ZU MEDIZINSTUDIUM UND TEST 2000.....	18
Anmeldungen, Testantritte und Zulassungen	18
Grösse der Testlokale	22
Testorte und Wunschuniversitäten	22
Testorte und Wohnkantone	24
Testabsolvierung nach Alter und Geschlecht.....	26
Personen mit Übernahme des Testergebnisses aus 1999	28
Sprachgruppen.....	29
Alter und Maturitätsjahr	31
BESCHREIBUNG DES VERWENDETEN EIGNUNGSTESTS	32
Aufbau des Tests	32
Beispielaufgaben für die Untertests	34
1. Untertest: Muster zuordnen.....	34
2. Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	35
3. Untertest: Schlauchfiguren.....	36
4. Untertest: Quantitative und formale Probleme	37
5. Untertest: Konzentriertes und sorgfältiges Arbeiten.....	37
6. Untertest: Figuren lernen	38
7. Untertest: Fakten lernen	38
8. Untertest: Textverständnis	39
9. Untertest: Diagramme und Tabellen.....	40
Berechnung der Werte	42
EVALUATION DES EMS IN DER SCHWEIZ	43
Testwert und Prüfungserfolg.....	45
Testwert und Prüfungsnote.....	51
Schlussfolgerungen für den EMS	56
TESTANWENDUNG IN DER SCHWEIZ 2000.....	57
Organisation der Testabnahme	57
Verteilungsprüfung.....	57
Äquivalenz der Testformen 1 und 2.....	60

Vergleich der Testfassungen 1998 bis 2000	61
Äquivalenz der Sprachversionen	63
Sprachvergleich für die Untertests	65
Darstellung des Korrekturverfahrens.....	68
Leistungen in identischen Testaufgaben 1998 – 2000.....	83
Vergleichbarkeit der Testlokale	84
Vergleich der Geschlechter	87
Prüfungserfolg im Medizinstudium und Geschlecht	87
Ergebnisse im Jahr 2000.....	88
Vergleich der Kantone	94
Vergleich 2000 zwischen Kantonen innerhalb der Deutschschweiz.....	94
Zusammenfassende Analyse für Kantone 1998 bis 2000.....	99
Vergleiche für Altersgruppen	103
Prüfungserfolg im Medizinstudium und Alter	103
Ergebnisse im Jahr 2000.....	104
Vergleiche nach Wunschuniversitäten	108
ERGEBNISSE ZUR TESTGÜTE	109
Zuverlässigkeit	109
Binnenstruktur	110
Item-Trennschärfen	116
Konzentriertes und sorgfältiges Arbeiten	120
LITERATUR	122
Originaltest zur Information und Vorbereitung	123
Frühere Berichte des ZTD	124
ANHANG ZUM BERICHT 2000	127

Zusammenfassung

Die in diesem Bericht vorgestellten Ergebnisse evaluieren den Testeinsatz für die Zulassung zum Studium der Human- und Veterinärmedizin in der Schweiz ab Wintersemester 2000/2001. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen 1998 und 1999 herangezogen. Ausserdem werden erste Daten zur Vorhersagbarkeit von Prüfungserfolg vorgestellt.

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

Die Teilnahme am Eignungstest für das Medizinstudium in der Schweiz (EMS) war im Jahr 2000 wie im Vorjahr für die Aufnahme eines Studiums der **Humanmedizin** und der **Veterinärmedizin** an den Universitäten **Basel, Bern, Freiburg** und **Zürich** notwendig. Für **Zahnmedizin** kam kein Numerus clausus (NC) zur Anwendung. Aufgrund des Testergebnisses erfolgt die Zulassung zum Studium entsprechend der Kapazität für die Testbesten. Ausserdem wird das Testergebnis als Kriterium für die Berücksichtigung der für die Studienaufnahme gewünschten Universität mit herangezogen.

Statistik

Bei einer Anfängerkapazität der vier Universitäten von insgesamt 696 Studienplätzen (Humanmedizin: 546, Veterinärmedizin: 150) lagen der Schweizerischen Hochschulkonferenz (SHK) am 15.2.2000 Anmeldungen von 1056 Personen (Humanmedizin: 843, Veterinärmedizin 213) für beide Disziplinen vor. Die Kapazität in Humanmedizin wurde um 54%, die in Veterinärmedizin um 42% überschritten.

Insgesamt 854 Personen haben sich fristgemäss bis zum 23.5.2000 bei der SHK zum Test angemeldet, tatsächlich angetreten sind am Testtag 801 Personen. 53 Personen haben ihre Testanmeldung zurückgezogen bzw. sind nicht zum Test erschienen.

Infolge von Rückzügen (vor allem wegen Nichtzusagen notwendiger Umleitungsentscheidungen) konnte weiteren Personen der Warteliste in der Reihenfolge absteigender Testwerte ein Studienplatz angeboten werden. Dadurch bleiben nach einer zweiten Zulassungsrunde letztlich nur sieben Prozent (1999 neun Prozent) der Personen mit gültigem Testwert ohne Studienplatzangebot.

Die Situation bei den Bewerbungen am Testtag ist in der Humanmedizin mit jener von 1999 in etwa vergleichbar. In der Veterinärmedizin ging das Ausmass der Kapazitätsüberschreitung zurück, was für die Humanmedizin im Vergleich 1998 zu 1999 fast analog festgestellt wurde. Offenbar wirken dissuasive Effekte vor allem bei der Einführung des Tests für eine Disziplin stärker als später, wenn der Test zur Realität geworden ist.

Erstmals haben 38 Personen (5% der Kapazität) ihr Testergebnis aus dem Vorjahr übernommen. Diese wurden 1999 als „Rückzüge“ gewertet. Deren Testwert ist im Mittel sogar besser als der Durchschnitt der Testteilnehmer 2000. Es ist damit zu rechnen, dass auch aus den Rückzügen 2000 sich Personen erneut bewerben.

Organisation und Ablauf

Der Test fand am 7.7.2000 gleichzeitig an sieben Testorten in drei Sprachen (deutsch: Basel, Bern, Chur, St. Gallen, Zürich; französisch: Freiburg; italienisch: Bellinzona) statt. Zu beachten ist, dass der Testort unabhängig vom zukünftigen Studienort gewählt werden kann.

Die Testleiter¹ und Stellvertreter wurden vorher ausführlich geschult, viele hatten bereits Erfahrungen aus den Jahren 1998 und 1999.

Der Ablauf des EMS ist an allen Testorten gleich, die Bearbeitungszeiten und die Instruktionen wurden genau vorgegeben. Die Bedingungen der Gleichbehandlung sind eingehalten worden und Benachteiligungen bei der Testabnahme traten nicht auf, etwa durch irreguläre Abläufe oder gar Pannen in einzelnen Testlokalen. Die Rate von insgesamt nur zwei Verwarnungen und keinem Ausschluss vom Test ist sehr gering und spricht für einen disziplinierten Testablauf. Dank einer engagierten und konstruktiven Mitwirkung der Koordinatoren der Testorte, der Testleiter und Stellvertreter sowie des Aufsichtspersonals kann die Testabnahme als erfolgreich eingeschätzt werden. Alle Protokolle sind sorgfältig angefertigt worden und bilden eine rechtsfähige Grundlage für den Nachweis einer ordnungsgemässen Testabnahme.

Testgüte

Die Zuverlässigkeitswerte (Konsistenzen, Reliabilitäten nach der Testhalbierungsmethode) liegen auch in dritten Jahr des Tests im gewohnten sehr hohen Bereich und entsprechen den deutschen Kennwerten. Auch die Faktorenstruktur (faktorielle Validität) ist mit derjenigen der Vorjahre nahezu identisch. Der Test wurde in den letzten drei Jahren mit einer gleichbleibenden Qualität angeboten. Dadurch ist auch gewährleistet, dass die Übernahme von Testwerten zwischen den Jahren ohne einschränkende Bedingungen möglich ist.

Unterschiede zwischen den beiden in der Schweiz verwendeten parallelen Testformen treten nicht auf, sie können als gleichwertig angesehen werden.

Gleichbehandlung der Sprachgruppen

In diesem Jahr galt wiederum der Beschluss der Schweizerischen Hochschulkonferenz, nur für sprachabhängige Tests bei Vorliegen signifikanter Mittelwertunterschiede zwischen den Sprachgruppen mögliche testbedingte Unterschiede auszugleichen. Dieses Kriterium erfüllten für beide Sprachgruppen alle fünf entsprechenden Untertests (vier sind weniger sprachabhängig).

Es wurde das bereits 1999 angewendete Verfahren eingesetzt, welches sich dort bewährt hatte. Diejenigen Unterschiede, die auf kulturell bedingte Unterschiede sowie auf unterschiedliche Repräsentativität der Testteilnehmer für ihre jeweilige Kohorte zurückzuführen sind, werden nicht ausgeglichen, sondern nur die testbedingten Unterschiede.

Das Ausgleichsverfahren beruht auf einer Analyse der Aufgabenschwierigkeiten zwischen den Sprachgruppen und dem Ausgleich von Unterschieden für Aufgaben, die einen Grenzwert überschreiten. Die objektiven Schwierigkeitsunterschiede werden bei

¹ Im Interesse der Flüssigkeit des Textes werden in diesem Bericht alle erwähnten Personengruppen nur in der männlichen Form genannt ("Testleiter", "Testteilnehmer", "Maturanden"), obwohl es sich selbstverständlich jeweils um Frauen und Männer handeln kann.

Nichtlösung einer Aufgabe der jeweiligen Person gutgeschrieben. Als Vergleichsmaßstab diente jeweils die deutschsprachige Form.

Nach dem Ausgleich ist gewährleistet, dass die Unterschiede der französisch- bzw. italienischsprachigen zur deutschsprachigen Gruppe in den sprachabhängigen Teilen relativ gesehen sogar geringer sind als in den sprachunabhängigen Teilen (erwartet würde wegen des Faktorenmodells etwa der gleiche Unterschied). Die Korrektur schliesst deshalb das Vorhandensein weiterer testbedingter Unterschiede mit hoher Sicherheit aus.

Die nach dem Ausgleich noch vorhandenen Unterschiede sind vor allem auf eine unterschiedliche Repräsentativität der Testteilnehmenden für die Kohorte bzw. kulturelle Unterschiede zurückzuführen. Zu beachten sind auch die geringen Stichprobengrößen der französisch- und der italienischsprachigen Gruppe. Vor allem aus dem Vergleich der zwischen zwei Jahren jeweils identischen Einstreuaufgaben zeigen sich für die französisch- und die italienischsprachige Gruppe andere Trends als in der deutschen Sprachgruppe, bei welcher eine kontinuierliche Verbesserung festgestellt werden kann.

Differenzierung nach Alter

Wie in beiden Vorjahren gibt es Unterschiede im Testwert bezüglich des Alters. Die jüngste Gruppe erreicht die besten Werte, gefolgt von denjenigen, die ihre Mittelschulbildung sofort nach der Primar- bzw. Sekundarstufe aufgenommen haben (unabhängig vom Alter).

Die Zulassungswahrscheinlichkeit ist wiederum für diejenigen Personen etwas geringer, die älter sind und zudem ihre Maturitätsprüfung in späteren Jahren abgelegt haben.

Es kann durch die Evaluationsdaten erstmalig auch gezeigt werden, dass es sich hier um keine „Verzerrung“ durch den Test handelt, sondern dass sich die Kriterien des Prüfungserfolges im Studium ebenso verhalten. Für die 1. Vorprüfungen 1999 (Testteilnehmer 1998) wird nachgewiesen, dass die älteste Gruppe mit später Maturität eine geringere Bestehenswahrscheinlichkeit und schlechtere Noten als alle anderen Gruppen erreicht. Der Test bildet die Realität des Studiums also genau ab.

Differenzierung nach Geschlecht

Auch im dritten Jahr des Eignungstests treten keine signifikanten Unterschiede bezüglich des Testwertes zwischen den Geschlechtern auf. Die Testwerte unterscheiden sich nicht signifikant (5%-Niveau) für Männer und Frauen in einer Analyse nach Disziplin und Sprachgruppe. Im Trend liegen die Mittelwerte der männlichen Gruppe wie in den Vorjahren wiederum etwas höher als die der weiblichen Gruppe (ausser in Humanmedizin der italienischen Gruppe).

Zu beachten ist, dass sich die Zusammensetzung der Bewerbergruppe in den letzten drei Jahren änderte: der Männeranteil nimmt seit 3 Jahren absolut und relativ zur Zahl der Frauen ab (v.a. Männer, die etwas älter sind, bewerben sich seltener) – derjenige der Frauen nimmt von 1999 zu 2000 wieder zu, nachdem auch 1998 zu 1999 eine Abnahme zu verzeichnen war. Die unterschiedlichen Quoten können die Repräsentativität der Teilnehmer für die jeweilige Kohorte beeinflussen.

Schliesslich wird gezeigt, dass die tendenziellen Unterschiede der Testwerte in der Realität des Studiums ihre Entsprechung haben: Frauen bestehen die 1. Vorprüfung weniger häufig als Männer und erreichen geringfügig schlechtere Noten. Auch für diesen Effekt können die unterschiedlichen Anteile der Geschlechter an der Studierendenzahl verantwortlich sein. Der Test ist mit den festgestellten Trends wiederum ein Abbild der Realität.

Summa summarum gibt es wiederum keine Hinweise, dass die Chancengleichheit der Geschlechter beim Test nicht gegeben wäre.

Differenzierung nach Maturitätsquote und Kanton

Es finden sich auch zum dritten Mal keine Unterschiede beim Testwert, wenn man die Kantone nach der Maturitätsquote gruppiert. Zwischen den Kantonen (deutschsprachige Kantone, deutschsprachige Teilnehmer) treten Unterschiede auf, die in den Vorjahren bereits diskutiert worden sind. Im Trend sind die Unterschiede stabil, wobei die Abweichungen am unteren Ende der Rangreihe in diesem Jahr erfreulich abgenommen haben. Es können aufgrund der Zusammenfassung der Testergebnisse 1998 bis 2000 jetzt mehr Kantone miteinander verglichen werden. Vor Überinterpretationen sei ausdrücklich gewarnt, weil es sich nicht um repräsentative Stichproben für die Maturanden handelt.

Vergleich zur Testanwendung 1998 und 1999

Auf den ersten Blick sind die Testaufgaben im Vergleich zwischen den drei Jahren schwieriger geworden (Basis sind die Aufgabenschwierigkeiten aus Deutschland). Betrachtet man die Veränderung der Testwerte in vergleichbaren Aufgaben zwischen 1998 und 1999 bzw. 1999 und 2000, dann sind die Ergebnisse der deutschen Sprachgruppe im Trend von Jahr zu Jahr besser geworden. Diese beiden Trends gleichen sich in etwa aus und der Test behält in etwa die gleiche optimale Differenzierungsfähigkeit.

Bei der französischsprachigen Gruppe zeigt sich für 2000 gegenüber 1999 eine Verbesserung – in der italienischsprachigen eine Verschlechterung. Beide Trends waren im Vergleich der ersten beiden Testanwendungen entgegengesetzt. Dies hängt sicher ebenso mit der geringen Stichprobengrösse wie mit der Dynamik innerhalb der Sprachgruppen hinsichtlich der Repräsentativität für die Maturanden des jeweiligen Geburtsjahres zusammen.

Evaluation: Prognose der Studieneignung

Untersucht wurde, inwieweit der Testwert im Eignungstest für das Medizinstudium in der Schweiz (EMS) eine Vorhersage von Studienerfolg ermöglicht. Der Nachweis dieser Prognosegüte rechtfertigt die Anwendung des EMS als Zulassungskriterium für das Medizinstudium, indem durch die Testbesten (1) das Studium schneller beendet wird und (2) bessere Studienleistungen erzielt werden.

Zur Verfügung standen Prüfungsdaten der Personen, die 1998 ein Studium der Humanmedizin in Basel, Bern, Freiburg und Zürich begonnen haben und somit 1999 im Sommer oder Herbst erstmals die 1. Vorprüfung ablegen konnten. Diese wurden mit den Testwerten im EMS verglichen.

Es kann gezeigt werden, dass sich die Personen, welche die Prüfung im ersten Anlauf bestehen, bezüglich des Testwertes signifikant von den anderen Gruppen (Nicht angetreten, Nicht bestanden) unterscheiden. Der Unterschied fällt sehr deutlich aus. Auch die Prüfungsdurchschnitte korrelieren mit 0.50 mit den Testwerten, was die aus Deutschland bekannten Korrelationswerte sehr genau bestätigt. Personen, die sich trotz Zulassung nicht immatrikulieren, unterscheiden sich bezüglich des Testwertes signifikant von den Gruppen, welche die Prüfung im ersten Anlauf bestehen – ein Hinweis, dass die Studieneignung nicht unerheblich für dissuasive Effekte sein muss.

Es kann als bestätigt gelten, dass Personen mit besseren Testleistungen die 1. Vorprüfung eher bestehen als andere Personen und dass bessere Testleistungen auch mit besseren Prüfungsergebnissen einhergehen. Die Verwendung des EMS als Zulassungskriterium ist aufgrund der Daten für diesen ersten Studienabschnitt gerechtfertigt.

Schlussfolgerungen

Auch im dritten Jahr des Einsatzes wurde der Eignungstest erfolgreich durchgeführt, die Vergabe der Studienplätze basierte auf einem bezüglich Zuverlässigkeit mit den Vorjahren voll vergleichbaren Verfahren. Indem der EMS seine hohe Prognosefähigkeit für Studienerfolg in den Evaluationsuntersuchungen unter Beweis gestellt hat, kann nunmehr auch für die Schweiz von einem sehr gut begründetem Auswahlverfahren ausgegangen werden, welches die Anforderungen an einen Numerus clausus erfüllt: diejenigen bevorzugt studieren lassen, die in kürzerer Zeit mit besserem Erfolg das Studium abschliessen.

Noch deutlicher als in den Vorjahren zeigen sich auch Dissuasionseffekte im Verlauf der Anmeldung zum Studium. In den Evaluationsstudien finden sich Hinweise, dass diese von den Kandidatinnen und Kandidaten selbst getroffene Entscheidung, ein Studium nicht aufzunehmen, zu einem Teil offenbar auch auf einer realistischen Selbsteinschätzung der eigenen Leistungsfähigkeit basiert – ein Effekt, der durchaus gewollt ist.

Die weitere Bereitstellung eines Testverfahrens mit gleichbleibender Qualität ist ein ebensolcher Schwerpunkt der weiteren Arbeit wie die Fortführung der Evaluation für den gesamten Studienverlauf.

Résumé

Les résultats présentés dans ce rapport évaluent l'application du test pour l'admission aux études de médecine humaine et vétérinaire en Suisse au semestre d'hiver 2000/2001. Certains points sont éclaircis à l'aide des données comparables obtenues lors des tests effectués en 1998 et 1999. Ce rapport contient également les premiers chiffres sur la prévisibilité de réussite des examens.

Les résultats ne sont pas assez représentatifs pour permettre de tirer des conclusions dans le sens d'une évaluation globale des établissements de formation, des filières d'études ou des régions.

La participation au test d'aptitudes pour les études de médecine en Suisse (AMS) était obligatoire en l'an 2000, comme l'année précédente, pour l'admission aux études de **médecine humaine** et de **médecine vétérinaire** dans les universités de **Bâle, Berne, Fribourg** et **Zurich**. Le numerus clausus (NC) n'a pas été appliqué en **médecine dentaire**. L'admission aux études a lieu sur la base des résultats au test et en fonction de la capacité des places d'études. Le résultat au test est, de plus, un des critères qui détermine l'attribution du lieu d'études lorsqu'il s'agit de tenir compte du choix des universités souhaitées par le candidat.

Statistique

Alors que la capacité des quatre universités participant au NC s'élève à 696 places d'études (médecine humaine: 546, médecine vétérinaire: 150), la Conférence universitaire suisse (CUS) comptait 1056 personnes inscrites au 15.2.2000 pour les deux disciplines (médecine humaine: 843, médecine vétérinaire 213). La capacité a été dépassée de 54% en médecine humaine et de 42% en médecine vétérinaire.

854 personnes au total se sont inscrites au test jusqu'au délai requis du 23.5.2000 auprès de la CUS. Parmi celles-ci, 801 personnes se sont effectivement présentées au test. 53 personnes ont retiré leur inscription ou ne se sont pas montrées le jour du test.

A la suite de retraits (surtout à cause d'un désaccord avec les décisions d'attribution du lieu d'étude), des places d'études ont pu être attribuées aux personnes figurant sur la liste d'attente, dans l'ordre décroissant des résultats obtenus au test. Ainsi, seuls sept pour cent (neuf pour cent en 1999) des personnes avec un résultat de test valable sont restées sans offre de places d'études après la deuxième série d'attribution.

La situation des candidatures en médecine humaine le jour du test est comparable à celle de 1999. Le dépassement des capacités en médecine vétérinaire a diminué de manière analogue à ce qui avait été constaté en médecine humaine entre 1998 et 1999. Visiblement, le test a un effet dissuasif plus marqué lors de son introduction dans une discipline que plus tard, lorsqu'il est devenu réalité.

Pour la première fois, 38 personnes (5% des capacités) ont repris leur résultat au test de l'année précédente. Elles avaient été qualifiées, en 1999, de "retraits". Leur résultat moyen est même supérieur à la moyenne. Il est probable qu'en 2001 également, cette procédure sera répétée par certaines personnes qui se sont retirées en 2000.

Organisation et déroulement

Le test a eu lieu simultanément le 7.7.2000 dans sept villes et en trois langues (en français: Fribourg; en allemand: Bâle, Berne, Coire, St-Gall, Zurich; en italien: Bellinzone). Il est à noter que le lieu de test peut être choisi indépendamment du futur lieu d'études.

Les responsables de test¹ et leurs remplaçants ont été au préalable soigneusement formés, la plupart d'entre eux ayant déjà participé au déroulement du test en 1998 et 1999.

Le test s'est déroulé de manière identique dans toutes les villes, les temps de passation et les consignes étant rigoureusement réglementés. Les conditions de l'égalité de traitement ont été respectées et des discriminations durant le test – en raison d'imprévis ou même de pannes dans un local particulier – ne se sont pas produites. Deux avertissements et aucune exclusion du test parlent en faveur d'un déroulement du test très discipliné. Grâce à une collaboration intense et constructive entre les coordinateurs des lieux de test, les responsables de test, les suppléants et les surveillants, le déroulement du test peut être considéré comme succès. Tous les protocoles ont été rédigés soigneusement et constituent une base solide pour la preuve d'un déroulement régulier.

Qualité du test

Les valeurs de fiabilité (consistances, fiabilité selon la méthode "split-half") sont pour cette troisième année de test encore une fois très élevées et correspondent aux valeurs allemandes. La structure factorielle (validité factorielle) est presque identique à celle des années précédentes. Ces trois dernières années, la qualité du test est restée constante ce qui permet de reprendre sans autres les valeurs de test d'une année à l'autre.

Aucune différence n'est apparue entre les deux formes de tests utilisées parallèlement en Suisse, elles peuvent donc être considérées comme équivalentes.

Egalité de traitement des groupes linguistiques

La décision de la Conférence universitaire suisse de compenser d'éventuelles différences liées au test, uniquement pour les tests dépendants de la langue, lorsque des différences de moyennes significatives se présentent entre les groupes linguistiques, a été appliquée cette année encore. Les cinq sous-tests correspondants ont rempli ce critère pour les deux groupes linguistiques (quatre sous-tests sont peu dépendants de la langue).

La procédure appliquée a été la même que celle qui avait déjà fait ses preuves en 1999. Les différences d'ordre culturel ainsi que celles liées au fait que les participants au test ne sont pas représentatifs de leur volée respective ne sont pas compensées. Seules les différences liées au test le sont.

La procédure de compensation se base sur une analyse du niveau de difficulté des problèmes entre les groupes linguistiques et sur une compensation des différences pour les tâches qui dépassent une valeur-limite. Lorsqu'un participant au test ne résout pas une tâche, la différence de difficulté objective constatée pour cette tâche est portée à l'avoir de la personne en question. Lors des comparaisons c'est la forme allemande qui faisait foi.

¹ Pour des questions de fluidité du texte, seule la forme masculine des différents groupes de personnes est nommée ("chef de test", "participant au test", "bachelier"). Elle comprend bien entendu tant les femmes que les hommes.

La procédure de compensation garantit que les différences entre le groupe linguistique français ou italien et le groupe linguistique allemand soient moins élevées dans les parties dépendantes de la langue que dans les parties indépendantes de la langue (plus ou moins la même différence était attendue en raison du modèle de facteurs). Ainsi, la correction exclue de manière très fiable toute présence d'autres différences liées au test.

Les différences subsistant après la procédure de compensation peuvent notamment être dues à une représentativité différente des participants au test au sein de chaque volée ou à des différences culturelles. Il s'agit de relever également la faible représentation des groupes linguistiques français et italien. En comparant les questions d'essai (qui ne comptent pas dans le résultat final) identiques disséminées dans les tests des deux années, on constate que les groupes linguistiques français et italien montrent d'autres tendances que le groupe linguistique allemand (celui-ci présente une amélioration continue).

Différenciation selon l'âge

Comme les deux années précédentes, des différences ont pu être notées au niveau de l'âge. Le groupe le plus jeune atteint les meilleurs résultats, suivi des personnes qui ont commencé leur formation du secondaire II directement après l'école obligatoire (indépendamment de l'âge).

La probabilité d'admission est, comme par le passé, moins élevée pour les personnes qui sont plus âgées et qui ont de plus passé leur maturité plus tard.

Pour la première fois, on peut également montrer à travers les données de l'évaluation qu'il ne s'agit pas de "déformation" de la part du test puisque les mêmes critères de réussite ont un effet identique lors des examens pendant les études. Il a été prouvé lors des premiers examens propédeutiques de 1999 (participants au test de 1998) que le groupe le plus âgé ayant obtenu une maturité plus tard dispose d'une probabilité de réussite moindre et obtient de moins bonnes notes que tous les autres groupes. Le test reflète donc exactement la réalité.

Différenciation selon le sexe

Pour la troisième fois, aucune différence significative au niveau du résultat au test n'est à relever entre les sexes. Dans une analyse par discipline et groupe de langue, les résultats au test des hommes et des femmes ne manifestent pas de différences significatives (à un niveau de 5%). Les moyennes du groupe masculin tendent, comme par le passé, à être un peu meilleures que celles du groupe féminin, sauf en ce qui concerne le groupe linguistique italien en médecine humaine.

On remarquera que la composition des groupes de candidats a changé au cours de ces trois dernières années: la proportion des hommes est, depuis trois ans, en diminution absolue et relative par rapport au nombre de femmes (les hommes un peu plus âgés surtout se présentent plus rarement) – la proportion des femmes a à nouveau augmenté entre 1999 et 2000, après avoir affiché une baisse entre 1998 et 1999. Les différents taux peuvent influencer la représentativité des participants pour la volée en question.

Enfin, on peut montrer que les différences qui se dessinent dans les résultats au test reflètent une réalité: les femmes réussissent le premier examen propédeutique moins fréquemment que les hommes et obtiennent légèrement de moins bonnes notes. Mais cet effet peut également, être engendré par la proportion différente de chaque sexe parmi le nombre d'étudiants. Le test est donc, de par les tendances constatées, un reflet de la réalité, qu'il peut dès lors prédire.

Somme toute, il n'existe toujours pas de preuve que l'égalité des chances entre les sexes ne serait pas atteinte lors du test.

Différenciation selon le taux de maturités et les cantons

Pour la troisième fois, aucune différence n'est à noter dans le résultat au test lorsque l'on regroupe les cantons selon leur taux de maturités. Des différences apparaissent entre les cantons (cantons germanophones, participants germanophones) qui ont déjà été discutées les années précédentes. La tendance va vers une stabilité des différences. Les écarts des cantons les plus bas classés ont toutefois diminué de manière réjouissante. Davantage de cantons peuvent maintenant être comparés entre eux grâce au résumé des résultats au test de 1998 à 2000. Toute extrapolation est toutefois déconseillée puisqu'il ne s'agit pas d'échantillons représentatifs pour chaque année de maturité.

Comparaison avec l'application du test en 1998 et 1999

De premier abord, les tâches du test semblent être devenues plus compliquées par rapport aux deux autres années (la base est le degré de difficulté des tâches en Allemagne). Si l'on considère l'évolution des résultats au test atteints pour des tâches similaires entre 1998 et 1999 ou entre 1999 et 2000, il s'avère que les résultats du groupe linguistique allemand esquissent une tendance vers la hausse d'une année à l'autre. Ces deux tendances s'égalisent plus ou moins, ce qui permet au test de conserver sa capacité optimale de différenciation.

Le groupe francophone présente une amélioration par rapport à 1999, le groupe italoophone par contre une diminution. Ces deux tendances sont opposées aux deux premiers tests. Cela dépend certainement autant de la petite taille de l'échantillon que de la dynamique au sein des groupes linguistiques du point de vue de la représentativité des bacheliers de l'année en question.

Evaluation: pronostic sur l'aptitude aux études

La question de savoir dans quelle mesure le résultat du test d'aptitudes pour les études de médecine en Suisse (AMS) permet de prévoir la réussite des études a fait l'objet d'une étude. Celle-ci montre que la validité du pronostic justifie l'application de l'AMS comme critère d'admission pour les études de médecine, puisque les personnes qui ont obtenu un meilleur résultat au test (1) terminent plus rapidement leurs études et (2) atteignent de meilleurs résultats.

Ces calculs ont été effectués sur la base des données d'examen des personnes qui ont commencé des études de médecine humaine en 1998 à Bâle, Berne, Fribourg et Zurich et qui ont pu se présenter pour la première fois en été ou en automne 1999 au premier examen propédeutique. Les résultats de cet examen ont été comparés avec les résultats au test AMS.

On peut montrer que les personnes qui ont réussi l'examen au premier essai se distinguent de manière significative des autres groupes (pas présentés, échoués) au niveau de leur résultat au test. La différence est manifeste. Même les moyennes d'examen ont un coefficient de corrélation de 0,5 avec les résultats au test, ce qui confirme exactement les coefficients de corrélation connus en Allemagne. Les personnes qui, malgré leur admission, ne s'immatriculent pas, avaient des résultats sensiblement différents des groupes qui ont réussi l'examen au premier essai – un indice qui révèle que l'aptitude aux études a son importance parmi les effets dissuasifs.

Il est confirmé que les personnes qui ont obtenu de meilleurs résultats au test réussissent le premier examen propédeutique plus tôt que les autres personnes et que de meilleurs résultats au test s'accompagnent également de meilleurs résultats d'examen. Ces données justifient l'utilisation de l'AMS comme critère d'admission pour cette première partie d'études.

Conclusions

Pour la troisième année consécutive, le test d'aptitudes a été mis en œuvre avec succès. L'attribution des places d'études s'est basée sur une procédure pleinement comparable aux années précédentes pour ce qui est de la fiabilité. L'AMS a prouvé sa haute capacité à pronostiquer le succès des études lors des évaluations. Il forme une base très solide pour la procédure de sélection, en Suisse également, qui remplit les conditions posées au numerus clausus: favoriser les études aux personnes qui les terminent plus rapidement et avec davantage de succès.

Les effets dissuasifs liés à la procédure d'inscription aux études, constatés déjà les années précédentes, étaient encore plus prononcés cette année. Les études d'évaluation indiquent que la décision de ne pas commencer d'études de médecine, prise du propre gré des candidats, se base apparemment pour une part sur une auto-évaluation réaliste des propres compétences – un effet qui est entièrement souhaité.

La suite des travaux mettra l'accent aussi bien sur la constance de la qualité du test utilisé que sur la poursuite de l'évaluation du déroulement de l'ensemble des études.

Einleitung

Der Eignungstest für das Medizinstudium und damit ein Numerus Clausus (NC) ist nach nunmehr dreijährigem Einsatz eine Realität geworden, die offenbar weithin akzeptiert wird. Dafür sind verschiedene Gründe verantwortlich: die Dringlichkeit von Einsparungen (besser der Eindämmung einer Kostenexplosion) im Gesundheitsbereich; die Besonderheiten des Arztberufes, der sich besser nicht nach den Gesetzen eines freien Marktes regulieren sollte, weil Patienten diese freie Wahl nicht haben und schliesslich auch die notwendigen Sparanstrengungen im Bereich der Hochschulbildung. Obwohl letzteres für alle Fächer gilt, hat die Medizin auch hier zwei Besonderheiten: es ist das wohl teuerste Studium und die notwendigen Kapazitäten der Universitätskliniken lassen sich ohne Qualitäts- und Imageeinbussen nicht beliebig vergrössern.

Internationale Konkurrenzfähigkeit (zu erhalten oder zu erreichen) erfordert Studienreformen, wie etwa einen problemorientierten Unterricht. Letzterer scheint wegen des erhöhten Betreuungsaufwandes dann nicht durchführbar, wenn die Kapazitäten deutlich überschritten werden, wie vor Einführung des NC die Realität aussah. Auch die für den Arztberuf notwendige Ausbildung sozialer und kooperativer Fähigkeiten ist eigentlich nicht vorstellbar, wenn in den ersten Semestern ein hoher „Selektionsdruck“ herrscht und alle Beteiligten wissen, dass noch ein hoher Anteil Studierender „herausgeprüft“ werden muss.

Dieser Nachteil der inneruniversitären Selektion wird bisher z.B. in der Westschweiz in Kauf genommen, weil man sich von den Anforderungen im Studium und deren Bestehen/Nichtbestehen eine spezifischere Selektion für die Anforderungen des Medizinstudiums erhofft und einen Eignungstest dafür als nicht spezifisch genug ansieht. Die in diesem Bericht vorgestellte Studie zur Prognosefähigkeit des Tests für die 1. Vorprüfung relativiert diese Aussage allerdings: Die Ergebnisse der Evaluation zeigen, dass der Prüfungserfolg (Bestehensrate, Zeitdauer, Noten) mit hoher Güte durch den Testwert vorhergesagt werden kann. So bleiben keine gravierenden Nachteile einer voruniversitären Selektion übrig. Sie stehen den deutlichen und von vielen Seiten beklagten Einschränkungen der Studienqualität durch eine Kapazitätsüberschreitung bei inneruniversitärer Selektion gegenüber.

Bisher wurde nur die erste Prüfung evaluiert, könnte man einwenden. Das ist richtig, auch die weiteren Prüfungen sollen hinsichtlich ihrer Vorhersagefähigkeit analysiert werden. Doch Abbrecherrate und die Wiederholungsrate sind eben genau in dieser ersten Prüfung am höchsten – vielleicht auch, weil sie als „versteckter Numerus clausus“ genutzt werden musste, um überhaupt mit den Kapazitäten zurecht zu kommen. Wenn man diese Leistungen vorhersagen kann, werden die Ressourcen von Anfang an besser genutzt und nicht überlastet.

Bei alledem muss beachtet werden, dass der Numerus Clausus bisher in der Schweiz ein sehr milder ist und durch die vorhandene dissuasive Wirkung nur wenigen Personen gar kein Studienplatz angeboten werden kann. Notwendige Umleitungsentscheidungen finden beispielsweise nicht immer die Zustimmung der Bewerber und „Nachrücker“ können die freigebliebenen Plätze erhalten. Insgesamt wird sich das Profil der Bewerber pro Geburtsjahr daher nur wenig ändern. Der unmittelbare Vorteil des NC ist, dass definitiv keine Kapazitätsüberschreitungen mehr auftreten und die Studienqualität dadurch auf ausreichendem Niveau gehalten werden kann. Personen, die ohne Studienplatzangebot bleiben, haben nachweislich nur geringe Chancen, die Prüfungen zu bestehen. Insofern ist die Auswahl gut begründet.

Numerus clausus und Medizinstudium

Betrachtet man die Anmeldungen bei der Schweizerischen Hochschulkonferenz (SHK) (Abbildung 1), zeigen sich auf den ersten Blick nur geringe Veränderungen gegenüber 1999. Die Zahl der Anmeldungen hat sich in Bern erhöht und in Zürich ist das sehr hohe Niveau der Kapazitätsüberschreitung weiter vorhanden. Abnahmen finden sich in Freiburg, Neuenburg und Basel – zwischen Lausanne und Genf hat sich der Trend der Vorjahre umgekehrt. Die Kapazitäten der am NC beteiligten Universitäten werden ausser in Freiburg überall deutlich überschritten (einschliesslich eines 20%-Zuschlages, der mögliche spätere Rückzüge ausgleichen soll). Auch gesamthaft wird die Kapazität um über 50% überschritten. Deshalb kann die Kapazitätsüberschreitung nicht durch Umleitungen abgebaut werden.

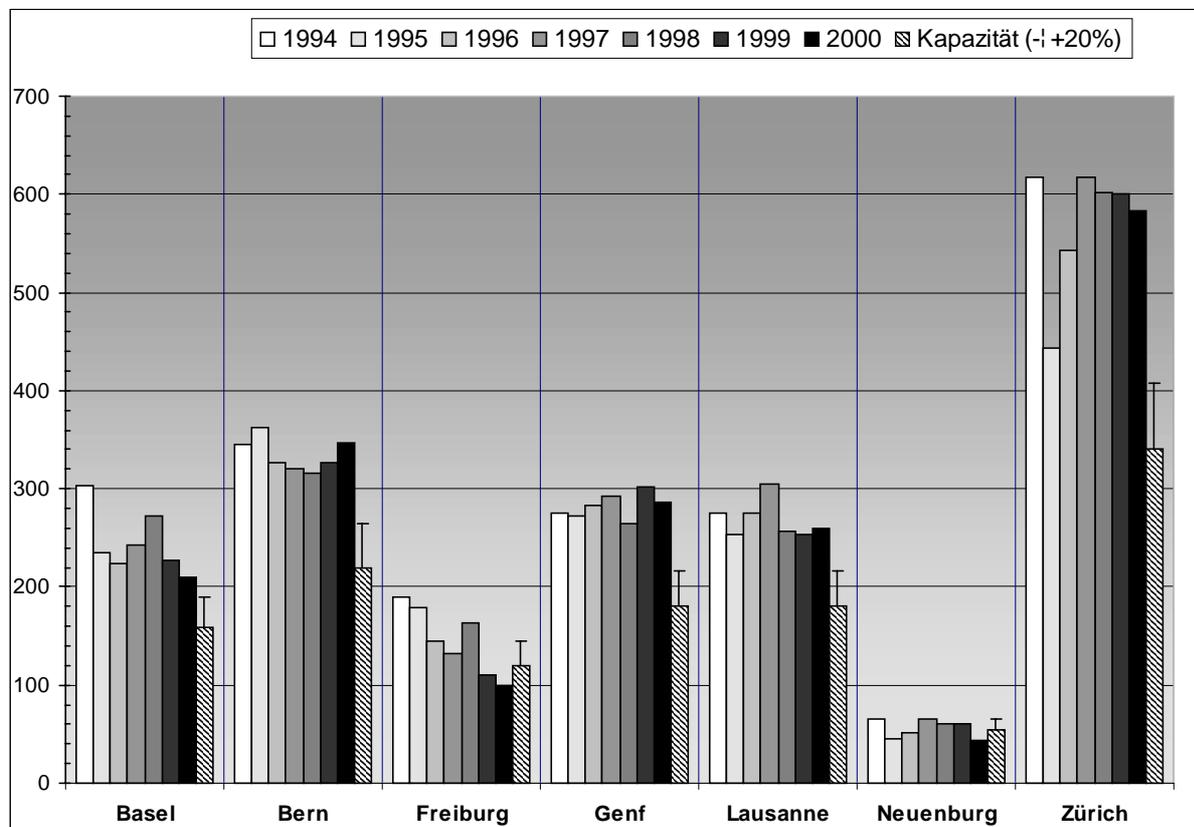


Abbildung 1. Anmeldungen zum Medizinstudium 1994 bis 2000, Anfängerprognose und Kapazitäten 2000 (letzte Säule mit Markierung des 20%-Überschreitungskriteriums für 2000 als Grenze für die Notwendigkeit einer Zulassungsbeschränkung)

Der Numerus clausus hat auch im dritten Jahr des Tests nicht zu einer Abwanderung in die Westschweiz geführt (Abbildung 2). Beide Landesteile haben im Jahr 2000 in etwa die gleiche geringfügige Abnahme bei den Anmeldezahlen zu verzeichnen. In diesem Jahr haben zwei kleinere Kantone (OW, AI) Doppeljahrgänge zu verzeichnen, die durch die Verkürzung der Maturitätszeiten von 13 auf 12 Jahre entstehen.

Zusätzliche Maturitäten 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). Wenn es keine zusätzlichen dissuasiven Effekte gibt, sollten auch die Anmeldezahlen besonders in den nächsten beiden Jahren steigen.

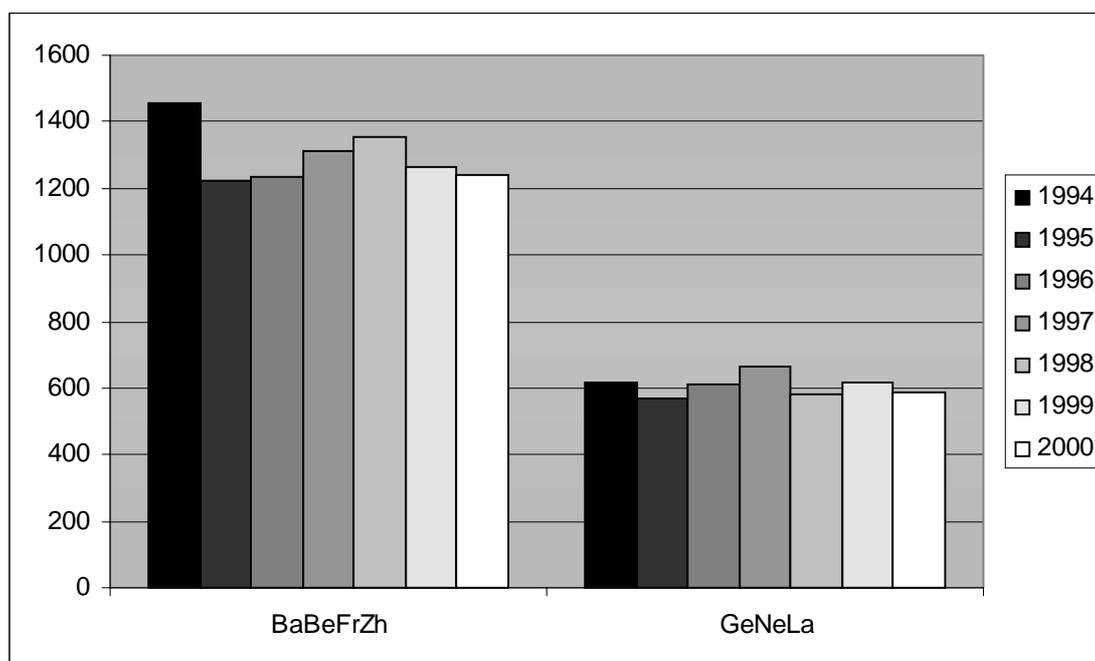


Abbildung 2. Anmeldungen zum Medizinstudium 1994 bis 2000 nach Gruppen (**B**asel, **B**ern **F**reiburg, **Z**ürich vs. **N**euenburg, **G**enf, **L**ausanne)

Der Nutzen des EMS besteht zusammengefasst in einer kapazitätsentsprechenden Zulassung zum Studium. Ausserdem wird die Berücksichtigung der Wunschuniversität vom Testergebnis mit beeinflusst, es gibt keine „Willkürentscheidungen“ bei der Umleitung. Man kann davon ausgehen, dass auch dissuasive Effekte auf den EMS zurückzuführen sind. Zum einen bleibt es eine Anforderung, der man sich „stellen“ muss mit dem Risiko, dass das Ergebnis nicht für eine Zulassung ausreicht (im Vergleich zu anderen Studiengängen, wo diese zusätzliche Leistung nicht notwendig ist). Bekanntlich ist die Zahl der Bewerbungen in Deutschland nach der Abschaffung des Tests sofort um 25% gestiegen, was für solche dissuasiven Effekte spricht,

Auch für die Kandidaten kristallisiert sich ein Nutzen heraus: das Testergebnis erlaubt eine realistischere Einschätzung der Chancen im Studium, die Prüfungen zu bestehen. Statistische Ergebnisse einer „Bestehenswahrscheinlichkeit“ beziehen sich natürlich auf die gesamte Kohorte und sie sollen weder zu Fatalismus, noch einer verminderten Anstrengung führen (weil man denkt, dass man die Prüfung leicht bestehen kann).

Anmeldung zu Medizinstudium und Test 2000

Anmeldungen, Testantritte und Zulassungen

Die folgende Tabelle 1 fasst den jährlichen Anmeldeverlauf bis zur Testabsolvierung zusammen. Bei Humanmedizin findet sich ein mit 1999 vergleichbarer Verlauf, wobei erstmals eine grössere Zahl von Personen bereits im Vorjahr am Test teilnahm und sich erneut anmeldete. In Veterinärmedizin findet sich möglicherweise der gleiche Rückgang der Bewerberzahlen, der zwischen 1998 und 1999 in Humanmedizin zu verzeichnen war (wenn im Jahr 2 mit einem Test gerechnet werden muss). Allerdings gab es 1998 auch 780 zusätzliche Maturitäten wegen der Doppeljahrgänge

	Humanmedizin			Veterinärmedizin		Total	
	1998	1999	2000	1999	2000	1999	2000
Anfängerkapazitäten	620	546	546	150	150	696	696
Anmeldungen bei SHK Mitte Februar d.J.	1020	845	844	231	213	1076	1057
... in Prozent der Kapazität	186%	155%	154%	153%	142%	154%	152%
Anmeldungen zum Test im Mai d.J.	797	717	678	194	176	911	854
Antrag Übernahme Testergebnis aus Vorjahr	0	5	33	0	5	5	38
Total Bewerbungen Mai	797	722	711	194	181	916	892
... in Prozent der Kapazität	129%	132%	130%	129%	121%	132%	128%
Absolvierung Test im Juli	750	673	637	182	164	855	801
Total Bewerbungen Juli (mit Test oder Ergebnisübernahme)	750	678	670	182	169	860	839
Rückzugsquote zwischen Anmeldung SHK und Testabsolvierung	26,5%	19,8%	20,6%	21,2%	20,7%	20,1%	20,6%
Anteil Personen mit gültigem Testwert, denen ein Studien- platz zugewiesen werden kann	82,7%	80,5%	81,4%	82,4%	88,7%	80,9%	82,9%
„NC-Bewerbungsindex“ (Bewerber mit gültigem Testwert pro Studienplatz)	1,21	1,24	1,23	1,21	1,13	1,23	1,21

Tabelle 1. Disziplinspezifische Statistiken für die am NC beteiligten Universitäten Basel, Bern, Freiburg und Zürich; NC 1998 nur für Humanmedizin – 1999 und 2000 für Human- und Veterinärmedizin

Der Bewerbungsindex für Humanmedizin liegt zwischen demjenigen der beiden Vorjahre, es ist keine Entspannung der Situation zu verzeichnen. Im Fach Veterinärmedizin ist er abgesunken. Die beiden folgenden Abbildungen veranschaulichen die Anmeldesituation grafisch.

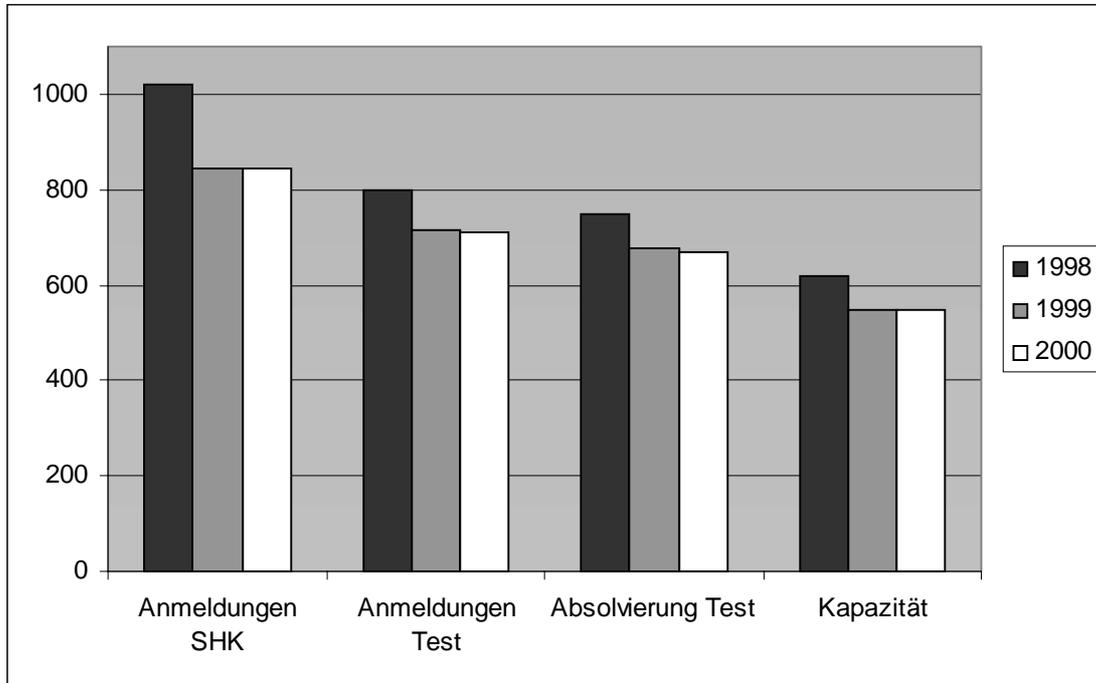


Abbildung 3. Anmeldeverlauf für Humanmedizin 1998 bis 2000 im Vergleich (mit NC)

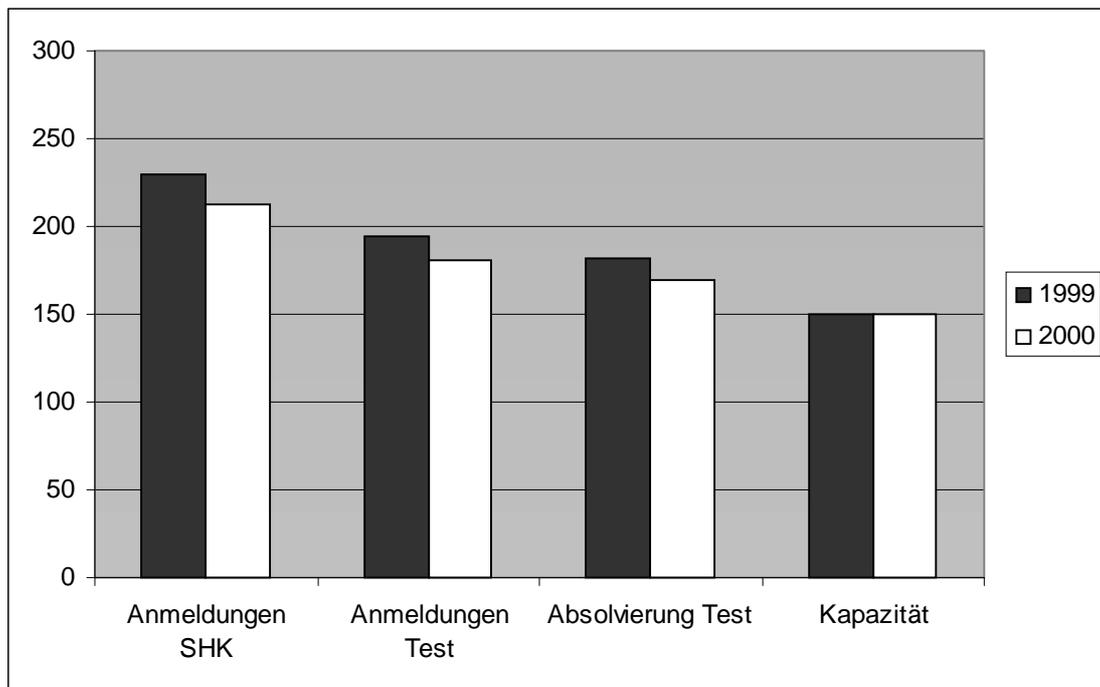


Abbildung 4. Anmeldeverlauf Veterinärmedizin 1999 und 2000 im Vergleich (mit NC)

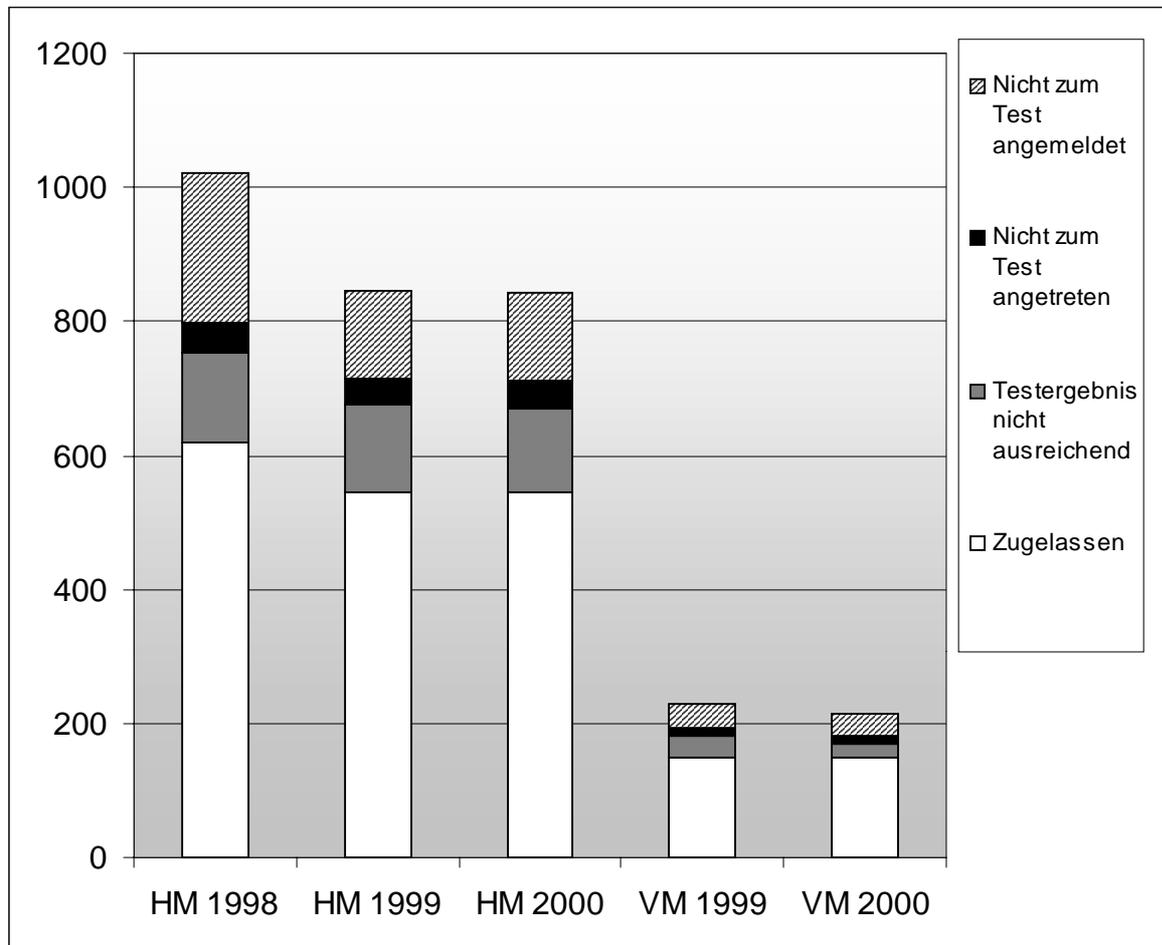


Abbildung 5. Kapazität, Anmeldungen und Testteilnahme für Human- (HM) und Veterinärmedizin (VM). Aufteilung der Voranmeldungen vom Februar des jeweiligen Jahres

Bezogen auf die einzelnen Universitäten und Disziplinen ergibt sich folgendes Bild:

Universität	WS 1999/2000					WS 2000/2001				
	Anmeldungen am 15.2.	Personen mit gültigem Testwert			Rückzüge in %	Anmeldungen am 15.2.	Personen mit gültigem Testwert			Rückzüge in %
		1998	1999	Total			1999	2000	Total	
Basel	189	2	149	151	20.1	174	9	124	133	23.6
Bern	164	2	130	132	19.5	177	5	139	144	18.6
Fribourg	98	0	79	79	19.4	88	2	64	66	25.0
Zürich	394	1	315	316	19.8	405	17	310	327	19.3
Total	845	5	673	678	19.8	844	33	637	670	20.6

Tabelle 2. Anmeldeverlauf für Universitäten: Humanmedizin

Bei im Mittel vergleichbaren Rückzugsraten ist die Zunahme der Rückzüge in Freiburg hervorzuheben. Auch in Veterinärmedizin ist die Rückzugsrate zwischen den Jahren vergleichbar, wobei es zwischen Bern und Zürich offenbar zu einer Umkehr der Trends gekommen ist. Hier würde erst eine längere Zeitspanne der Testanwendung Schlussfolgerungen hinsichtlich der Trends zulassen.

Universität	WS 1999/2000					WS 2000/2001				
	Anmeldungen am 15.2.	Personen mit gültigem Testwert			Rückzüge in %	Anmeldungen am 15.2.	Personen mit gültigem Testwert			Rückzüge in %
		1998	1999	Total			1999	2000	Total	
Bern	112	0	85	85	24.1	117	0	99	99	15.4
Zürich	119	0	97	97	18.5	96	5	65	70	27.1
Total	231	0	182	182	21.2	213	5	164	169	20.7

Tabelle 3. Anmeldeungsverlauf für Universitäten; Veterinärmedizin

Der Verlauf für 2000 wird in der nachfolgenden Tabelle grafisch veranschaulicht. Die Kapazitätsüberschreitung in Zürich bleibt am auffälligsten.

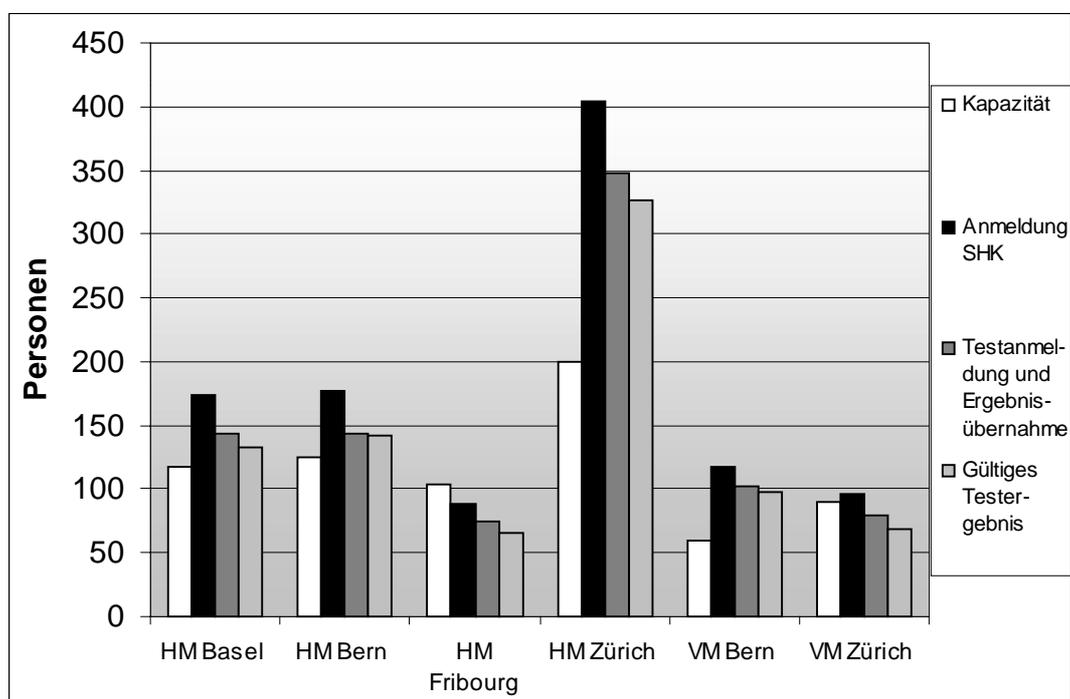


Abbildung 6. Kapazität, Anmeldungen und Testteilnahme für Fächer und Universitäten

Die Studienplätze wurden wie 1999 in zwei „Runden“ vergeben. Allen Personen mit einem bestimmten Testwert und darüber, der für jede Disziplin nach Kapazität festgelegt wurde, konnte in der ersten Runde ein Studienplatz angeboten werden. In Härtefällen und bei Identität von Wohnkanton und Kanton der Wunschuniversität wurde der gewünschte Ort berücksichtigt. Allen anderen Kandidaten wurden nach der Rangreihe des Testwertes die Universitäten so angeboten, dass möglichst wenige Umleitungen notwendig werden und die Übereinstimmung zwischen Rangreihe der Wunschuniversitäten und angebotenem Studienort am grössten ist. Dennoch waren, vor allem für Personen mit geringeren Testwerten, Umleitungen notwendig. Nicht alle Kandidaten bestätigten die Aufnahme des Studiums – die freiwerdenden Plätze wurden in einer zweiten Runde den Personen der Warteliste (wiederum nach der Höhe des Testwertes) angeboten.

Für die Berechnungen zur Zulassungsquote in diesem Bericht wurden wie in den Vorjahren Zulassungsraten festgelegt, die den Kapazitäten genau entsprechen.

Grösse der Testlokale

Wiederum wurde der Test in 14 Testlokalen in 7 Testorten durchgeführt. Es ergibt sich in etwa die gleiche Aufteilung der Kandidaten auf die Testlokale wie 1999. In Freiburg wurde der Test auf französisch, in Bellinzona auf italienisch angeboten. In allen anderen Testlokalen wurde der Test auf deutsch durchgeführt. Die unterschiedliche Grösse der Lokale hatte keinen Einfluss auf die Betreuungskapazität: die Zahl der Assistenten wurde in Abhängigkeit davon gewählt. Im übrigen wurde darauf geachtet, dass die Abstände zwischen den Arbeitsplätzen auch in den grösseren Testlokalen ausreichend waren.

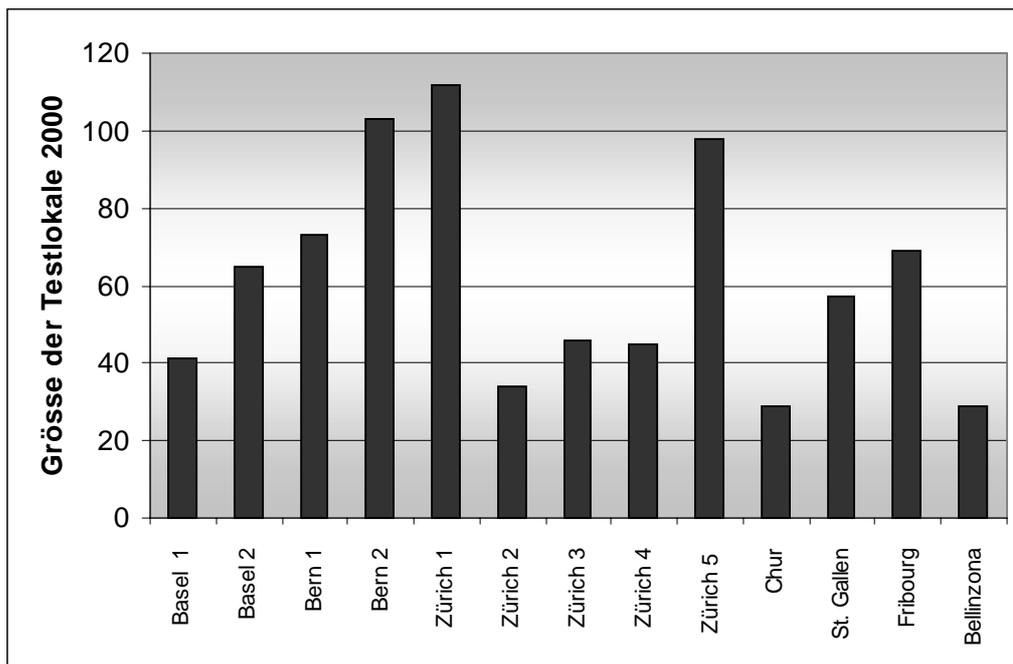


Abbildung 7. Grösse (Zahl der Personen) und Aufteilung der Testlokale 2000

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.

Testort	Studienort				Total
	Basel	Bern	Freiburg	Zürich	
BASEL	89 84.0%	11 10.4%	2 1.9%	4 3.8%	106
BELLINZONA	9 31.0%		3 10.3%	17 58.6%	29
BERN	6 3.4%	150 85.2%	14 8.0%	6 3.4%	176
CHUR	9 31.0%	7 24.1%		13 44.8%	29
FREIBURG	1 1.4%	42 60.9%	22 31.9%	4 5.8%	69
ST. GALLEN	2 3.5%	12 21.1%	8 14.0%	35 61.4%	57
ZÜRICH	8 2.4%	16 4.8%	15 4.5%	296 88.4%	335
Total	124	238	64	375	801

Tabelle 4. Testorte und Wunschuniversität (erste Wahl)

Abbildung 8 macht durch die "Streckung" aller Universitäten auf 100% das Verhältnis der "lokalen" Bewerber zu den übrigen vergleichbar.

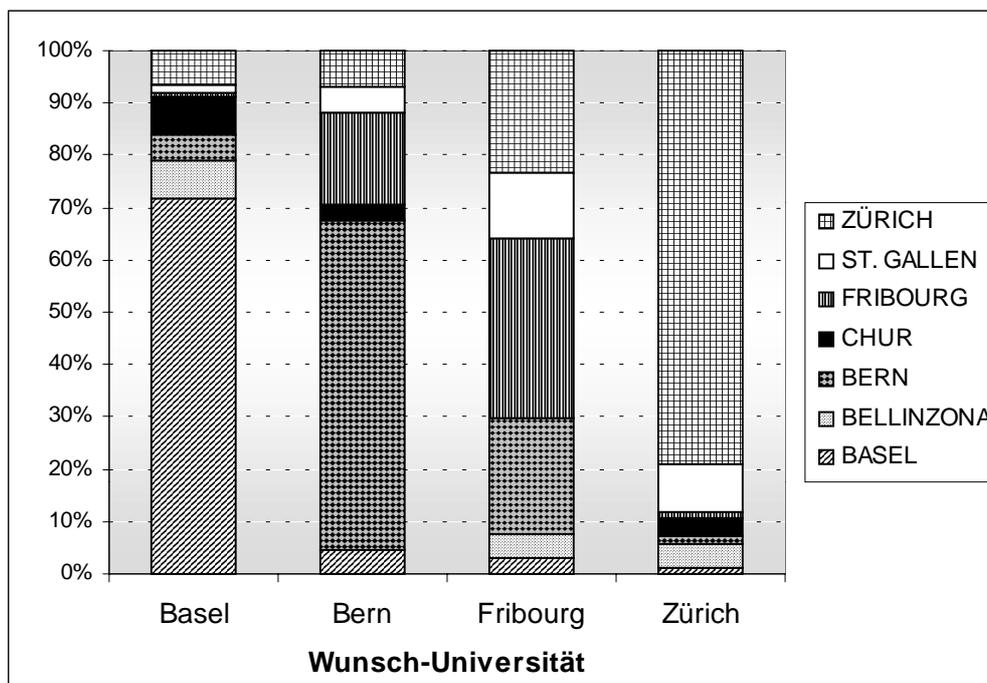


Abbildung 8. Aufteilung der Kandidaten nach Wunschuniversität auf die Testorte

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.

Wegen der geringen Absolutzahl ist in Kantonen mit Doppeljahrgängen allenfalls ein Trend zur Zunahme zu erkennen. Auch in den anderen Kantonen finden sich starke Fluktuationen. Die Zahlen 1998 für St. Gallen und Schwyz entsprechen aber keinesfalls einer Verdoppelung der Bewerberzahlen.

Wohnkanton	BASEL	BELLIN-ZONA	BERN	CHUR	FRI-BOURG	ST. GALLEN	ZÜRICH	2000	1999	1998 (HM)
AG	17		3				51	71	84	66
AI						4		<u>4</u>	0	2
AR						4		4	6	4
BE	5		107				2	114	113	100
BL	39						1	40	59	50
BS	28							28	39	40
FR			15		29			44	47	34
GE			1		10			11	13	1
GL							5	5	4	4
GR		1		18				19	29	19
JU					1			1	2	0
LU	2		14	1			24	41	44	43
NE					6			6	5	2
NW			1				1	2	3	6
OW	2		3				4	<u>9</u>	2	3
SG			1	7		35	12	55	61	<u>64</u>
SH							8	8	4	7
SO	8		12					20	28	26
SZ	2		2				14	18	18	<u>21</u>
TG						10	13	23	20	22
TI		27						27	29	30
UR							2	2	1	3
VD			1		16			17	11	3
VS			12		6			18	15	22
ZG			1		1		7	9	10	6
ZH		1	1	1		2	189	194	191	164
AUSLAND	3		2	2		2	2	11	17	8
Total	106	29	176	29	69	57	335	801	855	750

Tabelle 5. Wohnkanton und Testorte, Vergleich zu 1999 und 1998 (dort nur Humanmedizin). Unterstrichen: Doppeljahrgänge der Maturanden

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 1999 und 2000 nur in Bern und Zürich angeboten worden ist.

Wohnkanton	Studienort				Total 2000	1999
	Basel	Bern	Freiburg	Zürich		
AG	17	7	2	45	71	84
AI		2		2	4	0
AR			2	2	4	6
BE	8	102	1	3	114	113
BL	32	5	2	1	40	59
BS	24	2		2	28	39
FR	2	12	29	1	44	47
GE		9		2	11	13
GL		1	1	3	5	4
GR	5	4		10	19	29
JU		1			1	2
LU	5	16	5	15	41	44
NE		6			6	5
NW	1	1			2	3
OW	1	4		4	9	2
SG	3	11	6	35	55	61
SH			1	7	8	4
SO	7	11	1	1	20	28
SZ	2	1	1	14	18	18
TG		2	1	20	23	20
TI	9		3	15	27	29
UR			1	1	2	1
VD		15	1	1	17	11
VS		15	1	2	18	15
ZG		4		5	9	10
ZH	3	4	6	181	194	191
Ausland	5	3		3	11	17
Total	124	238	64	375	801	855

Tabelle 6. gewünschte Studienorte nach Wohnkantonen, Vergleich zu 1999

Testabsolvierung nach Alter und Geschlecht

Die Zahl der Frauen überwiegt auch im Jahr 2000 in den Fächern Human- und Veterinärmedizin die Zahl der Männer. Der relative Frauenanteil nimmt in diesem Jahr in beiden Disziplinen deutlich zu. In der Abbildung 9 wird deutlich, dass die Zahl der Frauen absolut gesehen nur geringfügig zugenommen und vor allem die Zahl der Männer abgenommen hat. Noch weiter differenziert wird dieser Befund in der Abbildung 10. Es sind vor allem die älteren Bewerbergruppen (Geburtsjahre 1958 bis 1977 im Jahre 2000 bzw. in den Vorjahren jeweils um 1 Jahr älter), die sich in Humanmedizin weniger häufig bewerben.

Im Fach Veterinärmedizin ist vor allem die Zahl der Männer gesunken, die sich unmittelbar nach der Mittelschule für dieses Fach bewerben. Für Männer und Frauen gilt hier, dass sich mehr „Ältere“ bewerben als in den Vorjahren.

	Geburtsjahre			Total
	1958 – 1976	1977 – 1979	1980 - 1982	
Humanmedizin				
männlich	23 37,7%	91 46,9%	141 36,9%	255 40,0% von Gesamt 1999 45,5% 1998 44,8%
weiblich	38 62,3%	103 53,1%	241 63,1%	382 60,0% von Gesamt 1999 54,5% 1998 55,2%
Gesamt	61 9,6% v. Gesamt 1999 10,4% 1998 10,4%	194 30,5% v. Gesamt 1999 32,5% 1998 29,5%	382 59,9% v. Gesamt 1999 57,1% 1998 60,1%	637
Veterinärmedizin				
männlich	2 10,5%	20 30,3%	13 16,5%	35 21,3% von Gesamt 1999 24,2%
weiblich	17 89,5%	46 69,7%	66 83,5%	138 78,7% von Gesamt 1999 75,8%
Gesamt	19 11,6% v. Gesamt 1999 9,9%	66 40,2% v. Gesamt 1999 26,9%	79 48,2% v. Gesamt 1999 63,2%	164

Tabelle 7. Testabsolventen nach Geburtsjahr und Geschlecht. (Vergleiche der Geburtsjahre zu den Analysen aus 1998 und 1999 um je ein Jahr verschoben)

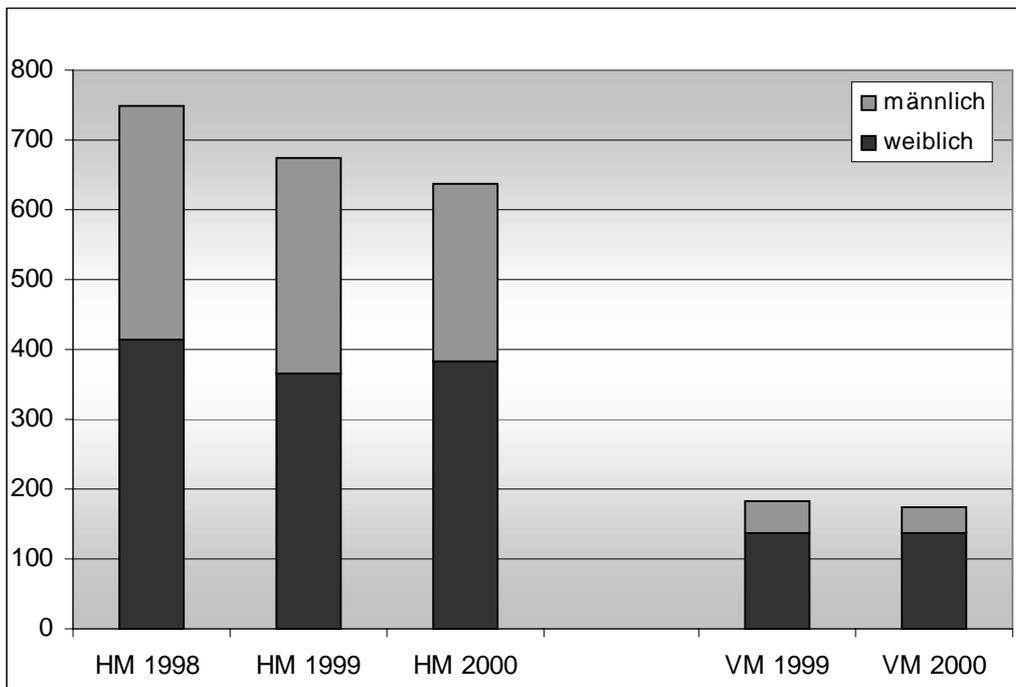


Abbildung 9. Anteil der Geschlechter bei der Bewerbung für Humanmedizin (HM) 1998 bis 2000 und Veterinärmedizin (VM) 1999 und 2000

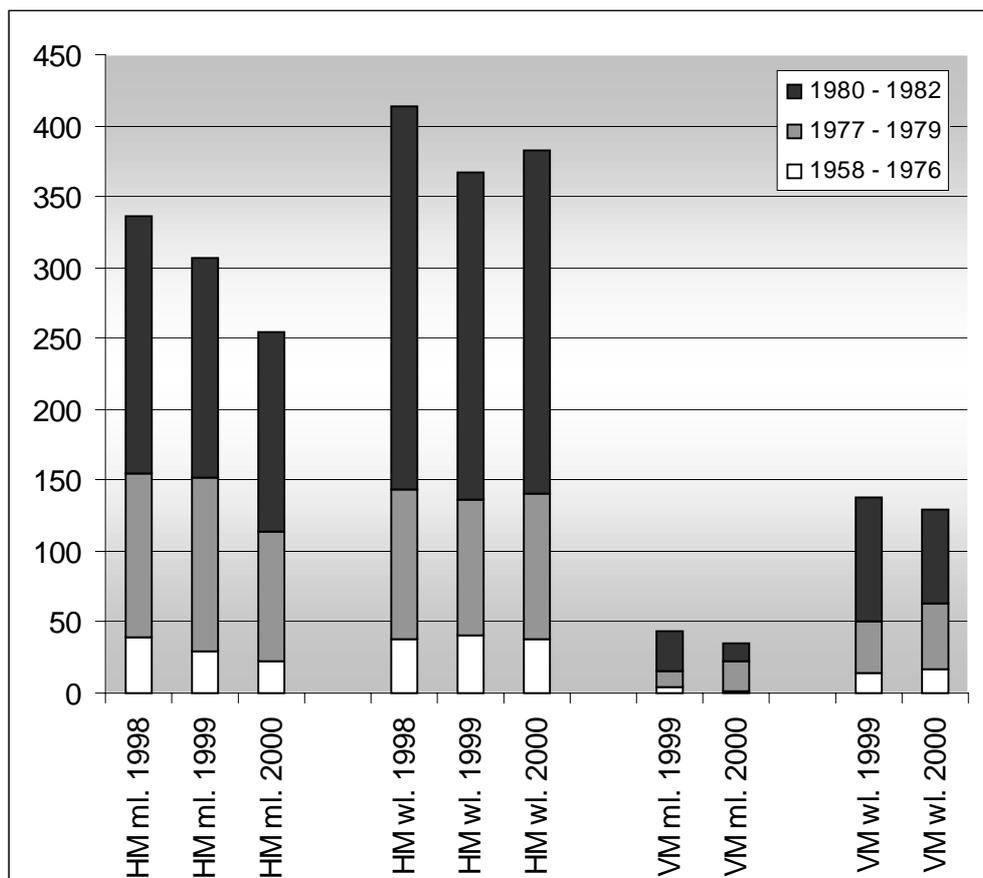


Abbildung 10. Zusammensetzung der Bewerbergruppen nach Geschlecht ml: männlich wl: weiblich und Alter (Geburtsjahre)

Personen mit Übernahme des Testergebnisses aus 1999

Erstmals übernehmen 38 Personen ihr Testergebnis aus dem Vorjahr. Es kann gezeigt werden, dass dieser Personenkreis im Mittel über überdurchschnittliche Testwerte verfügt. Nur 5,3% dieser Personen liegen unter dem kapazitätsentsprechenden Zulassungskriterium für 2000. Rückzüge nach dem Testtermin bedeuten also nicht die Aufgabe des Studienwunsches – eventuell wird sich dieser Trend weiter verstärken und ein Jahr „Auszeit“ genommen.

Disziplin	N	Minimum	Maximum	Mittelwert	Standardabw.
Humanmedizin	33	85	119	102.85	7.25
Veterinärmedizin	5	97	107	102.00	3.81

Tabelle 8. Testwert für Personen mit Übernahme des Testergebnisses aus 1999

16 Personen sind männlichen, 22 Personen weiblichen Geschlechts. Die Altersspanne umfasst die Geburtsjahre 1967 bis 1981 (Der Median ist 1979).

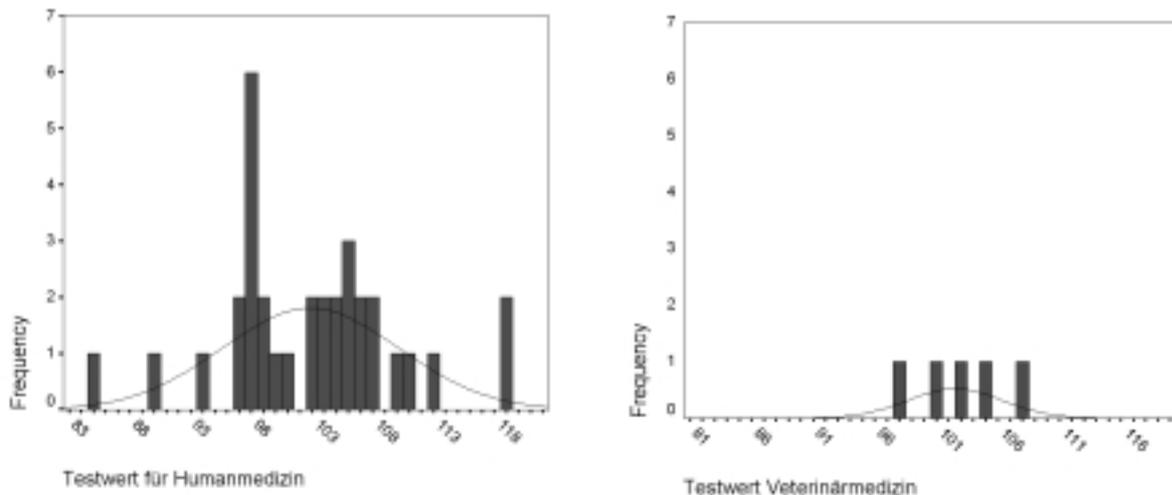


Abbildung 11. Werteverteilung der Personen mit Testübernahme aus 1999

	Test 2000	Übernahme Testergebnis 1999
Nicht zugelassen	17.9%	5.3%
Zugelassen	82.1%	94.7%

Tabelle 9. Vergleich der Zulassungsquoten für Testteilnehmer 2000 und Personen, die ihr Ergebnis aus 1999 übernommen haben

Sprachgruppen

In Bellinzona konnte der Test in italienischer Sprache und in Fribourg in französischer Sprache absolviert werden. Leider sind beide Sprachgruppen sehr klein und auch die Repräsentativität für die Bewerberinnen und Bewerber der Sprachregion lässt sich nicht genau einschätzen. Dies wird bei der Äquivalenzprüfung der Sprachversionen noch eine Rolle spielen.

Schlüsselt man die Alters- und Geschlechtsverteilung nach den Sprachgruppen auf (Tabelle 10), wird deutlich, dass in der französisch- und der italienischsprachigen Gruppe die Anteile der Jüngeren wesentlich grösser sind. Es gibt dort kaum Personen beim Test, die ihr Studium nicht zum frühestmöglichen Zeitpunkt nach dem Ablegen der Maturitätsprüfung beginnen wollen.

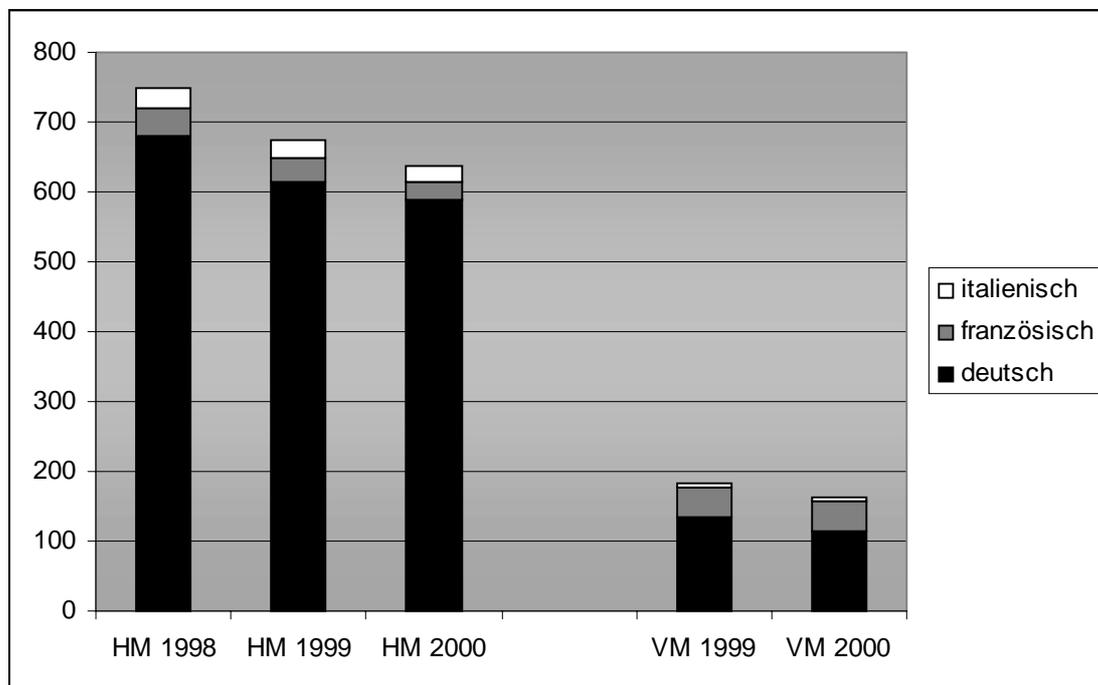


Abbildung 12. Vergleich der Sprachgruppen 1998 bis 2000 - Testteilnehmer

Sprache	Geschlecht	Geburtsjahr			Total
		1958-1976	1977-1979	1980-1982	
Humanmedizin					
deutsch	männlich	21 9.1%	86 37.4%	123 53.5%	230
	weiblich	37 10.3%	100 27.8%	223 61.9%	360
französisch	männlich		2 15.4%	11 84.6%	13
	weiblich	1 9.1%	1 9.1%	9 81.8%	11
italienisch	männlich	2 16.7%	3 25.0%	7 58.3%	12
	weiblich		2 18.2%	9 81.8%	11
Veterinärmedizin					
deutsch	männlich	2 9.5%	15 71.4%	4 19.0%	21
	weiblich	14 15.2%	40 43.5%	38 41.3%	92
französisch	männlich		4 36.4%	7 63.6%	11
	weiblich	3 8.8%	5 14.7%	26 76.5%	34
italienisch	männlich		1 33.3%	2 66.7%	3
	weiblich		1 33.3%	2 66.7%	3

Tabelle 10. Sprachgruppen nach Alter und Geschlecht aufgeschlüsselt

Alter und Maturitätsjahr

Die nachfolgende Tabelle verdeutlicht, dass die ältere Gruppe sich wiederum aus zwei Teilgruppen zusammensetzt: Personen, die ihre Maturitätsprüfung unmittelbar vor der Bewerbung absolviert haben (ausserhalb der Diagonalen der Tabelle) und solche, die ihre Maturitätsprüfung zum frühestmöglichen Zeitpunkt abgelegt haben und dann noch anderen Tätigkeiten nachgegangen sind (in der Diagonalen der Tabelle). Bei der Gruppe, die ihre Maturitätsprüfung später abgelegt hat, handelt es sich offenbar v.a. um Personen, welche die Maturität auf dem sogenannten "zweiten Bildungsweg" erworben haben.

Jhrg.	Maturitätsjahr															Total
	78	85	88	89	90	91	92	93	94	95	96	97	98	99	00	
1958	1															1
1962															1	1
1965															1	1
1966		1							1						1	3
1967															2	2
1968			2												1	3
1969			1	1	1		1							1		5
1970					1									3	2	6
1971					2						1		1		1	5
1972						2	2	1							4	9
1973							2	1						1	1	5
1974						1	1	2	3		1			1	5	14
1975								1	3	2	1	2			6	15
1976										1	2			1	6	10
1977											1	4	12	8	12	37
1978												2	19	18	27	66
1979												3	6	66	82	157
1980													1	57	289	347
1981														8	98	106
1982															8	8
	1	1	3	1	4	3	6	5	7	3	6	11	39	164	547	801

Tabelle 11. Geburtsjahr und Maturitätsjahr, dicke Linien zeigen die Aufteilung der Gruppen

Beschreibung des verwendeten Eignungstests

Aufbau des Tests

Der EMS besteht aus neun verschiedenen Untertests, die zu einem Gesamtwert verrechnet werden. Die Tabelle 12 gibt einen Überblick über den Testaufbau, die Aufgabenzahl und die Dauer der einzelnen Untertests:

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
MITTAGSPAUSE 60 Minuten			
Lernphase zu den Gedächtnistests: Figuren lernen Fakten lernen			4 6
Textverständnis	Verständnis und Interpretation medizinischer und naturwissenschaftlicher 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
Gesamttest		204 (178)*	5 Std. 7 Min.

Tabelle 12. Struktur und Ablauf des EMS. * Aufgabenzahl und in Klammern die Zahl der gewerteten Aufgaben pro Untertest und für Gesamt, 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 sollten Struktur und Testlänge mit denen des deutschen TMS identisch sein, um dort langjährig erprobte Abläufe und damit auch die Güte-Kennwerte nicht zu verändern.

Wie wurden die Aufgaben entwickelt? Ausgangspunkt war eine differenzierte Anforderungsanalyse des Medizinstudiums, an der zahlreiche Lehrbeauftragte und Experten mitarbeiteten. Gemäss diesen Anforderungen wurden einzelne Aufgabengruppen (Untertests) konstruiert, von denen neun in die entsprechend endgültige Testform Aufnahme fanden. 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 12) erfolgte in Deutschland im Rahmen sogenannter „**Einstreuaufgaben**“. Nur bei ausreichender Bewährung wurden solche Aufgaben dann 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.

Damit Testlänge und -schwierigkeit des EMS in der Schweiz mit derjenigen der deutschen Version vergleichbar sind, wurden auch in der Schweizer Version pro entsprechendem Untertest vier Einstreuaufgaben verwendet. Diese wurden wie in Deutschland aus Forschungsgesichtspunkten ausgewählt und bei der Berechnung des Testwertes nicht berücksichtigt. Die beiden in der Schweiz verwendeten Testformen unterscheiden sich hinsichtlich der Einstreuaufgaben, sodass insgesamt acht zusätzliche Aufgaben pro Untertest „eingestreut“ werden konnten. Im Jahre 2000 wurden die Einstreuaufgaben erneut aus Aufgaben gebildet, die 1999 bereits verwendet worden sind. Dies schien notwendig, um Veränderungen vor allem zwischen den Sprachgruppen besser abschätzen zu können. So ist ein direkter Leistungsvergleich beider Geburtsjahre möglich. Insgesamt wurden 204 Aufgaben bearbeitet, von denen 178 gewertet worden sind.

Die in der Schweiz verwendete Version des Eignungstests wurde in Deutschland bereits einmal im Zeitraum der letzten zehn Jahre angewendet. Dabei können einzelne Aufgaben aus verschiedenen deutschen Versionen entnommen worden sein, um die Zusammenstellung hinsichtlich ihrer Kennwerte zu optimieren. Der grosse Vorteil der in der Schweiz verwendeten Aufgaben besteht darin, dass sie bereits mindestens zweimal von einer grösseren Stichprobe bearbeitet worden sind: mindestens einmal als Einstreuaufgabe in einem früheren Test (bei notwendiger Revision ggf. auch wiederholt) und einmal als Aufgabe, die für den Testwert gezählt wurde. Zweimal wurden die Gütekriterien von diesen Aufgaben also bereits erfüllt.

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 beträchtlich. Auch die Vielfalt der Aufgaben ist grösser als hier vermittelt werden kann. Einen besseren Überblick geben die veröffentlichten Originalversionen. Idealerweise ist das gesamte Schwierigkeitsspektrum repräsentiert, damit der Testwert ausreichend gut zwischen Personen differenziert. Komplette Versionen von Tests sind im Buchhandel für die Vorbereitung auf den Test erhältlich (siehe Literaturverzeichnis).

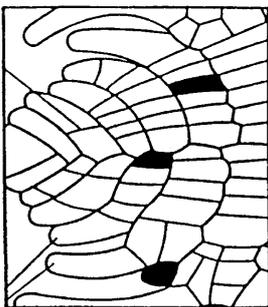
Bei der Beurteilung der Aufgaben wird die **Nähe zu Studienanforderungen** deutlich: an der Entwicklung haben neben Gymnasiallehrern, Medizinern und Psychologen auch Lehrbeauftragte des Grundstudiums Medizin mitgearbeitet. Die Struktur der Untertests ist auf detaillierte Anforderungsanalysen eines Medizinstudiums zurückzuführen, die im Rahmen der Testentwicklung durchgeführt worden sind (Trost, 1997). Die manchmal verwendete Bezeichnung des Tests als „Probestudium“ ist vor allem darauf zurückzuführen.

1. Untertest: Muster zuordnen

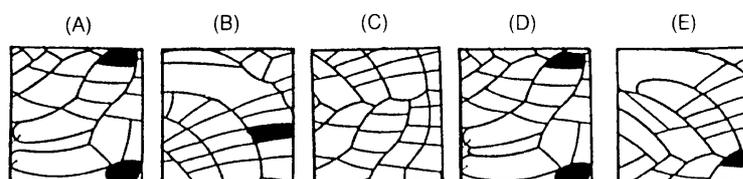
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wiederzuerkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

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:

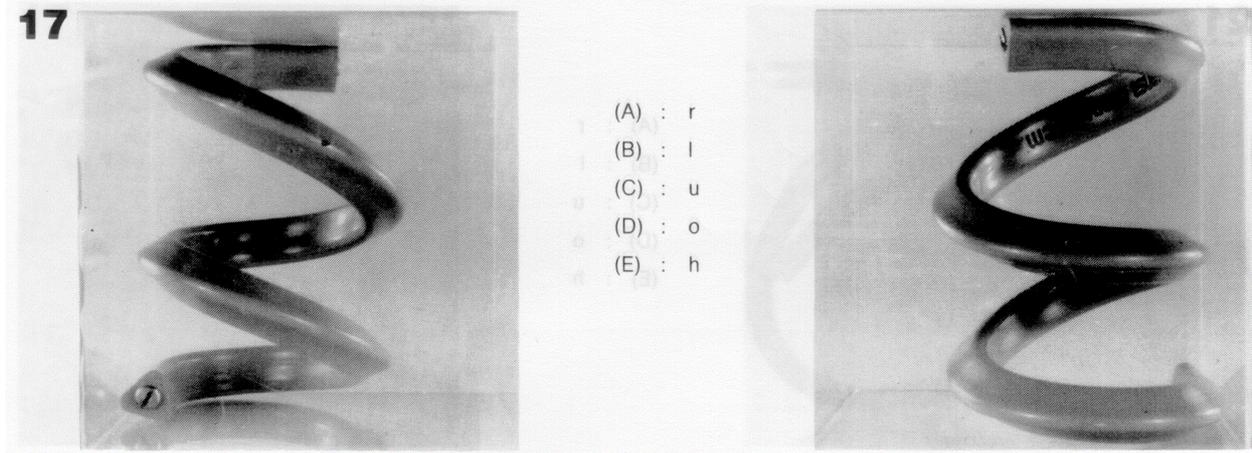
- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind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

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, wo 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!)

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.

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 $\overline{\overline{b}}$ $\overline{\underline{b}}$

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.

$\overline{\overline{b}}$ $\overline{\overline{q}}$ $\underline{\underline{b}}$ $\overline{\underline{b}}$ $\underline{\overline{b}}$ $\overline{\overline{b}}$ $\overline{\overline{b}}$ $\underline{\underline{q}}$ $\underline{\underline{q}}$ $\underline{\underline{b}}$ $\underline{\underline{b}}$ $\overline{\overline{b}}$ $\overline{\overline{q}}$

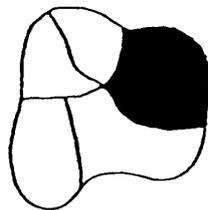
6. Untertest: Figuren lernen

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

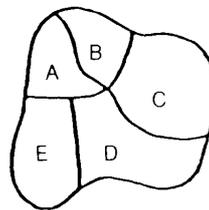
Der Untertest „Figuren lernen“ prüft, wie gut man sich Einzelheiten von Gegenständen einprägen und merken kann.

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.

7. Untertest: Fakten lernen

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

Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

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

Der Patient mit dem Schädelbruch ist von Beruf ...

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

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. 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 Jodatombau eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schliesslich T_4 , während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T_3 hervorgeht. T_3 und T_4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt.

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

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 nicht alle heute erfassten Anforderungen weniger wichtig für Studienerfolg sein – 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.

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 13). 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.

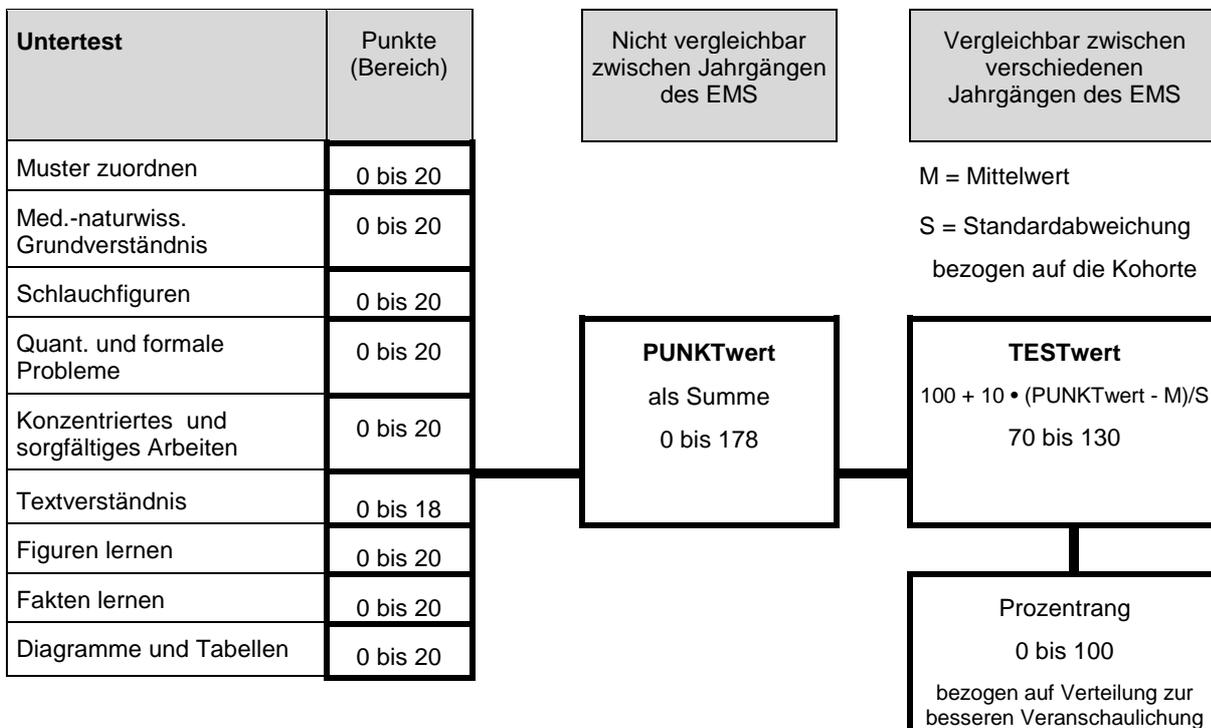


Abbildung 13. Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert

Evaluation des EMS in der Schweiz

Seit 1998 muss die Zulassung für das Medizinstudium an den Universitäten Basel, Bern, Freiburg und Zürich durch einen Numerus clausus beschränkt werden, weil die Anmeldungen die Kapazitäten deutlich überschreiten. Während 1998 ein Numerus clausus nur für Humanmedizin galt, ist seit 1999 auch Veterinärmedizin betroffen.

Der Numerus clausus basiert auf dem Eignungstest für das Medizinstudium in der Schweiz (EMS), welcher aus Deutschland übernommen worden ist und dort bereits sehr gute Evaluationsergebnisse erzielt hat.¹

Neben den juristisch geforderten Kriterien Rechtsgleichheit und Willkürfreiheit bei der Zulassung muss gelten, dass die Zugelassenen das Studium in angemessen kurzer Zeit beenden, um die Ressource der vorhandenen Studienplätze optimal zu nutzen. Die Zulassung ist dann fair, wenn gute Leistungen im EMS mit guten Studienleistungen in Beziehung stehen – wenn die „testbesten“ Zugelassenen das Studium mit besseren Leistungen in kürzerer Zeit beenden als die Personen mit weniger guten Testleistungen.

Bisher wurde anhand der Testdaten der Jahre 1998 und 1999 auch in der Schweiz nachgewiesen, dass der EMS wichtige Evaluationskriterien qualitativ hochstehend erfüllt:²

- Durch eine strenge Standardisierung der Durchführung, die Schulung der Testleiter und vergleichbare Bedingungen an allen Testorten ist die Rechtsgleichheit gegeben.
- Dies gilt insbesondere für die Sprachgruppen, wo durch ein vergleichsweise aufwändiges Analyseverfahren mögliche testbedingte Unterschiede ausgeglichen werden, die tatsächlich vorhandenen Unterschiede jedoch nicht nivelliert werden.
- Ein objektives Auswerteverfahren mit mehreren Kontrollschritten bei der Ermittlung des Testwertes und ein transparentes, gut definiertes Zulassungsverfahren garantieren auch die geforderte Willkürfreiheit.
- Der Test ist fair bezüglich der Geschlechter, der Altersgruppen, der Regionen mit unterschiedlicher Maturitätsquote innerhalb der Deutschschweiz und auch der einzelnen Kantone: der Testwert zeigt keine signifikanten Unterschiede (z.B. Geschlechter, Maturitätsquoten) bzw. bildet vorhandene Realitäten objektiv ab (z.B. Alter).
- Die Testgütekriterien wie Zuverlässigkeit und faktorielle Struktur entsprechen in der Schweiz bei beiden bisherigen Testdurchführungen dem von Deutschland her bekannten hohen Niveau.

Bei der Bewertung der Auswirkungen des Numerus clausus in der Schweiz muss man allerdings beachten, dass es sich um einen vergleichsweise „milden“ NC handelt:

¹ Trost, G. (Hrsg.) (1996). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (20. Arbeitsbericht). ITB: Bonn. – Auszug auch im Bericht 4 (1998) und Bericht 5 (1999) des ZTD und im Internet unter http://www.unifr.ch/ztd/ems/berichte/b5/vorhersage_des_studienerfolges.htm. Erwähnenswert sind die hohen Zuverlässigkeitswerte, die mehrfach bestätigte Faktorenstruktur und vor allem die Prognosegüte für schriftliche und auch mündliche Prüfungsleistungen.

² Vgl. Berichte 4 (1998) und 5 (1999) des Zentrums für Testentwicklung.

- Nur etwa 1/5 der Personen mit einem gültigen Testergebnis kann kein Studienplatz in der ersten Zuteilungsrunde angeboten werden. Da notwendige Umleitungen an andere Universitäten manchmal nicht die Zustimmung der betroffenen Personen finden, können freigewordene Studienplätze weiteren Personen angeboten werden. Berücksichtigt man dies, erhielten 1998 insgesamt nur 12% der Testabsolventen gar kein Studienplatzangebot. Es handelt sich also um keine extreme Auslese bei der Zulassung nach der Testleistung.
- Entgegen dem allgemeinen Trend verstärkter Nachfrage nach Hochschulausbildung zeigt sich für das Medizinstudium auch eine dissuasive Wirkung des Tests: die Anmeldezahlen bleiben stabil bzw. sinken sogar leicht ab. Diese „Selbstregulation“ ist beabsichtigt; sie wird wahrscheinlich nicht nur nach der Studieneignung erfolgen.
- Schliesslich muss bei der Bewertung von Veränderungen an den Universitäten beachtet werden, dass zunächst noch eine hohe Zahl von Repetenten die Kapazitäten in Anspruch nimmt, bevor sich die Situation entspannt. Die jahrelange Überschreitung der Kapazitäten lässt sich nicht sofort abbauen, die Entlastung wird sich erst allmählich einstellen.

Zahl der Personen, die immatrikuliert worden sind	597		100%
Noch nie angetreten (inkl. Studienabbrecher)	54		9,0 %
Im Sommer 1999 erster Versuch bestanden	297		49,7 %
Im Herbst 1999 erster Versuch bestanden	64		10,7 %
Im Sommer 1999 erster Versuch nicht bestanden	138		23,1 %
	<i>Davon im Herbst bestanden</i>	32	23 %
	<i>Davon im Herbst erneut nicht bestanden</i>	6	4 %
	<i>Davon nicht erneut angetreten</i>	100	73 %
Im Herbst 1999 erster Versuch nicht bestanden	44		7,4 %
Zahl der Personen, die nicht immatrikuliert wurden	153		
	<i>Nicht zugelassen (aufgrund niedrigem Testwert)</i>	88	
	<i>Zugelassen, aber nicht immatrikuliert</i>	65	

Tabelle 13. Personengruppen nach der Teilnahme an der 1. Vorprüfung (Humanmedizin, Studienbeginn 1998) und Personen, die nicht zugelassen worden sind oder sich trotz Zulassung nicht immatrikuliert haben.

Eine wichtige Frage der Evaluation ist natürlich, inwieweit sich durch den Testwert Studienerfolg tatsächlich vorhersagen lässt. Bisher haben lediglich die 1998 für Humanmedizin Zugelassenen zwei Prüfungsmöglichkeiten gehabt (Sommer 1999 und Herbst 1999). Aufgrund dieser Daten ist eine erste Evaluation erfolgt und soll hier vorgestellt werden.¹

Untersucht wurden alle Studienanfänger in Humanmedizin, die 1998 aufgrund des EMS an den Universitäten Basel, Bern, Freiburg und Zürich zugelassen wurden. Die Prüfungsnoten

¹ Die Evaluation erfolgt im Auftrag des Beirates „Eignungstest“ der Schweizerischen Hochschulkonferenz. Der Evaluationskommission gehören an: Prof. R. Bloch (i.A. des Beirates „Eignungstest“ der SHK), K. Wechsler (i.A. der SHK), H. Neuhaus (i.A. des BAG), K. Holenstein (i.A. des BfS), Prof. K.-D. Hänsgen (i.A. des ZTD).

wurden aus dem Prüfungsregister des Bundesamts für Gesundheit (BAG)¹ ermittelt und von den Prüfungsprotokollen individuell neu erfasst. Die Information über das Bestehen der Prüfung wurde direkt aus der Datenbank des BAG extrahiert. Berücksichtigt wurden die Prüfungssessionen Sommer 1999 und Herbst 1999.

Neun Prozent der 1998 in der Humanmedizin Immatrikulierten haben noch keinen Prüfungsversuch unternommen. Darunter sind auch Personen, die das Studium bereits abgebrochen haben. Eine Bestimmung dieses Anteiles würde Studienverlaufsdaten des Bundesamtes für Statistik erfordern. Wegen der geringen Gruppengröße wurden diese Daten in diesem Jahr noch nicht beigezogen.

Rund 60% der Immatrikulierten bestehen die Prüfungen beim ersten Antritt. Von den rund 25% Personen, welche die Prüfung im Sommer nicht bestanden haben, bestand nur ¼ im Herbst den zweiten Prüfungsversuch. Summa summarum wird es bei 1/3 der ursprünglich Immatrikulierten kein Bestehen der 1. Vorprüfung nach zwei Semestern geben, was eine wahrscheinliche Studienverlängerung und damit zusätzliche Blockierung von Ressourcen bedeutet.

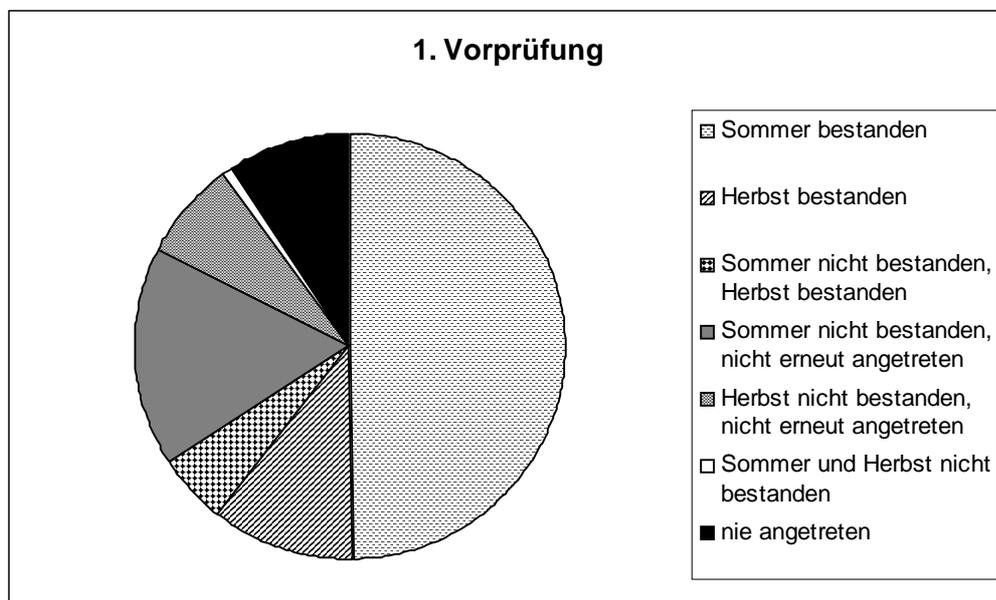


Abbildung 14. Anteile der einzelnen Gruppen nach dem Abschneiden in den Prüfungen

153 Personen haben am Test teilgenommen, sind aber nicht immatrikuliert: 88 Personen wurden nicht zugelassen (Testwert unter dem NC-Kriterium). 65 Personen wurden zugelassen, haben aber den Studienplatz nicht in Anspruch genommen (z.B. wegen einer Umleitung, der Aufnahme eines anderen Studiums, wegen persönlicher Gründe etc.).

Testwert und Prüfungserfolg

Besteht ein Zusammenhang zwischen Testleistung im Eignungstest und Prüfungserfolg? In der Tabelle 14 sind die Mittelwerte und Standardabweichungen für den Testwert der in Tabelle 13 genannten Gruppen dargestellt und in Abbildung 15 die Verteilungen als Boxplots gezeichnet.

¹ Besonders sei Herrn D. Megert vom BAG für die Aufbereitung der Daten für diese Studie gedankt.

	N	Mittelwert	Standard- abweichung	95% Vertrauensintervall	
				Untergrenze	Obergrenze
nicht zugelassen	88	82.8	4.4	81.9	83.8
Sommer - Herbst -	6	96.2	4.6	91.3	101.0
Herbst -	44	96.7	5.1	95.2	98.3
nicht immatrikuliert (aber zugelassen)	65	96.8	7.8	94.9	98.8
Sommer -	100	98.5	6.1	97.3	99.8
nie angetreten	54	99.4	6.8	97.6	101.3
Sommer - Herbst +	32	100.9	5.2	99.1	102.8
Herbst +	64	105.3	7.5	103.5	107.2
Sommer +	297	105.7	7.8	104.8	106.6
Total	750	100.0	10.0	99.3	100.7

Tabelle 14. Kennwerte des Testwertes für die Gruppen (+ bestanden, - nicht bestanden)

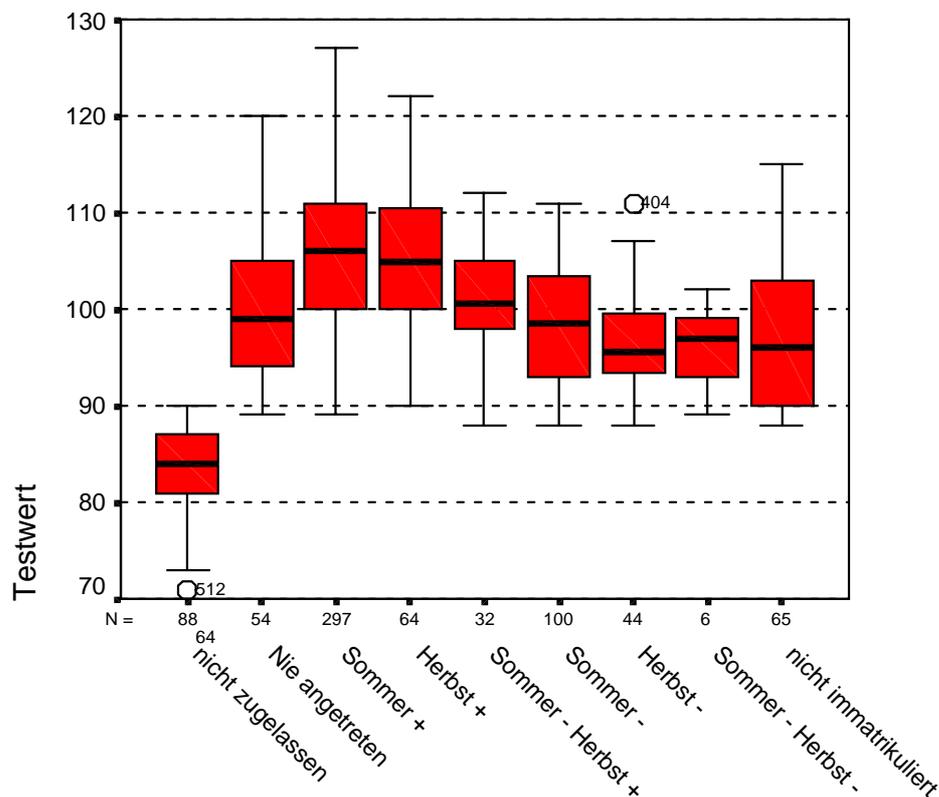


Abbildung 15. Boxplots des Testwertes für die Prüfungsgruppen. Dicke schwarze Linie: Median. Dargestellt sind die Quartile: 2. und 3. Quartil als Balken, 1. und 4. Quartil als Linie. Ausreisser sind markiert (z.B. bei Herbst Erstantritt nicht bestanden; eine Person mit Testwert 110 als nicht typisch für diese Gruppe); + bestanden, - nicht bestanden

Der Gesamt-Mittelwert des Testwertes für alle Immatrikulierten liegt mit 102,9 ca. $\frac{1}{3}$ Standardabweichung über dem Mittelwert aller Testteilnehmer von 100 (die Standardabweichung beträgt 10). Die nicht zugelassenen Personen unterscheiden sich erwartungsgemäss deutlich – erfolgte die Zulassung doch nach dem Testwert.

Interessant ist allerdings, dass die Personen, die sich trotz Zulassung nicht immatrikuliert haben, ebenfalls 1/3 Standardabweichung unter dem Mittelwert von 100 liegen. Wie sich beim multiplen Mittelwertvergleich (Tabelle 15) zeigt, unterscheidet sich diese Gruppe signifikant von den Erfolgreichen, welche die Prüfung im ersten Anlauf bestehen. Die „Dissuasion“ *nach* dem Zeitpunkt der Testteilnahme scheint also nicht völlig unabhängig von der Studieneignung zu sein, obwohl sehr verschiedene Gründe für die Nichtaufnahme des Medizinstudiums denkbar sind.

		Subset for alpha = .05		
		1	2	3
<i>nicht zugelassen</i>	88	82.85		
Sommer - Herbst -	6		96.17	
Herbst -	44		96.75	
<i>nicht immatrikuliert (aber zugelassen)</i>	65		96.85	
Sommer -	100		98.56	
nie angetreten	54		99.44	
Sommer - Herbst +	32		100.94	100.94
Herbst +	64			105.38
Sommer +	297			105.71
Sig.		1.000	.151	.151

Tabelle 15. Multipler Mittelwertvergleich (Tukey) für Testwert: Drei homogene Gruppen nach Nicht zugelassen, Bestehen oder Nichtbestehen/Nicht-Angetretensein getrennt. + bestanden, - nicht bestanden.

Im multiplen Mittelwertvergleich werden drei homogene Gruppen unterschieden: Personen, welche die Prüfung bestehen, bilden die eine Gruppe. Klar getrennt davon werden Personen, welche die Prüfungen nicht bestanden haben oder noch nicht zur Prüfung angetreten sind. Die dritte Gruppe sind erwartungsgemäss die Nicht-Zugelassenen, was ja aufgrund des Testwertes erfolgte.

Personen, welche zweimal die Prüfung nicht bestanden haben, unterscheiden sich dabei um fast 10 Punkte (eine Standardabweichung) von denjenigen, welche die Prüfung im ersten Anlauf bestehen. Personen, die erstmals im Herbst antreten und dort nicht bestehen, haben fast genauso niedrige Testwerte wie die Personen, die bereits zweimal nicht bestanden haben.

Personen, die noch nie zur Prüfung angetreten sind, unterscheiden sich bezüglich dem Testwert auch deutlich von denjenigen, welche die Prüfung im ersten Anlauf bestehen. Die Personengruppe, welche die Prüfung im zweiten Anlauf besteht, lässt sich beiden Mittelwertgruppen zuordnen.

Es zeigt sich insgesamt, dass das Bestehen der Prüfung sehr eng mit den Testwerten zusammenhängt. Fasst man die Prüfungsteilnehmer nach „Bestanden“, und „Nicht bestanden“ zusammen, wird dies am deutlichsten. Die Erfolgreichen in der ersten Vorprüfung unterscheiden sich signifikant von den Personen, welche noch nicht angetreten

oder nicht bestanden haben. Im Boxplot (vgl. Abbildung 16) zeigt sich die sehr geringe Überlappung der Quartile 2 und 3 für „Bestanden“ und „Nicht bestanden“.

	N	Mittelwert	Standardabweichung
nicht bestanden	150	97.9	5.8
nicht angetreten	54	99.4	6.8
bestanden	393	105.2	7.6
Total	597	102.9	7.9

Tabelle 16. Zusammenfassung für Testwert nach Bestehen und Nichtbestehen der 1. Vorprüfung

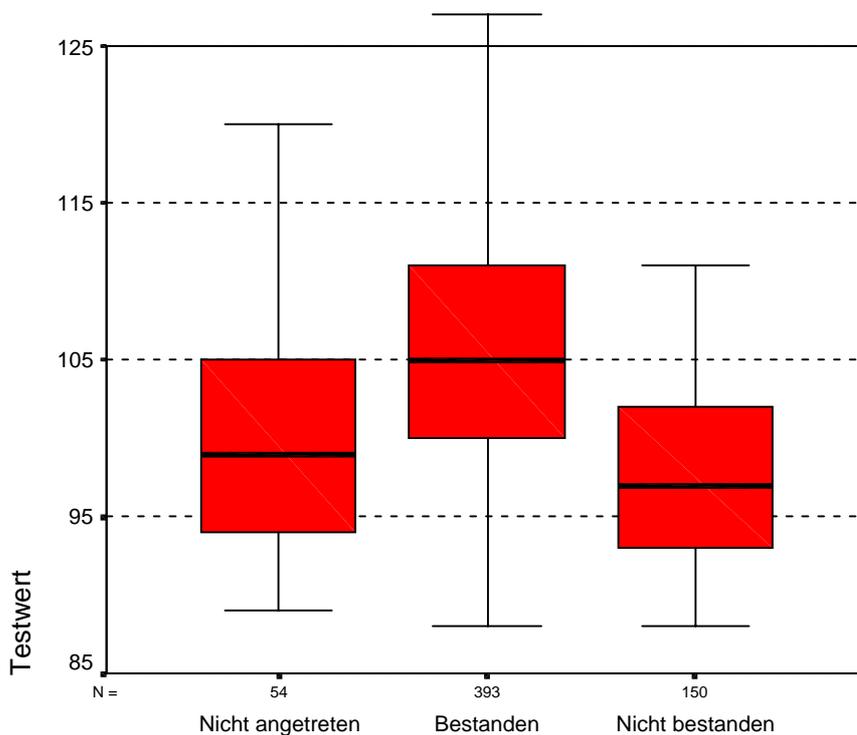


Abbildung 16. Boxplots für Zusammenfassung nach "bestanden", "nicht bestanden" und "nicht angetreten"

	N	Subset for alpha = .05	
		1	2
nicht bestanden	150	97.9	
nicht angetreten	54	99.4	
bestanden	393		105.3
Sig.		.270	1.000

Tabelle 17. Multipler Mittelwertvergleich: Prüfung bestanden wird von "nicht bestanden" und "nicht angetreten" signifikant unterschieden

Die Abbildung 17 zeigt das Prozentverhältnis von "Bestanden" und "Nicht bestanden" (ohne immatrikulierte Personen, die nicht zur Prüfung angetreten sind) pro Testwert.¹ Alle Personen mit einem Testwert > 112 haben die Prüfung bestanden. Bei niedrigeren Testwerten zeigt sich eine sehr deutliche monotone Zunahme der Bestehenswahrscheinlichkeit in Abhängigkeit vom Testwert.

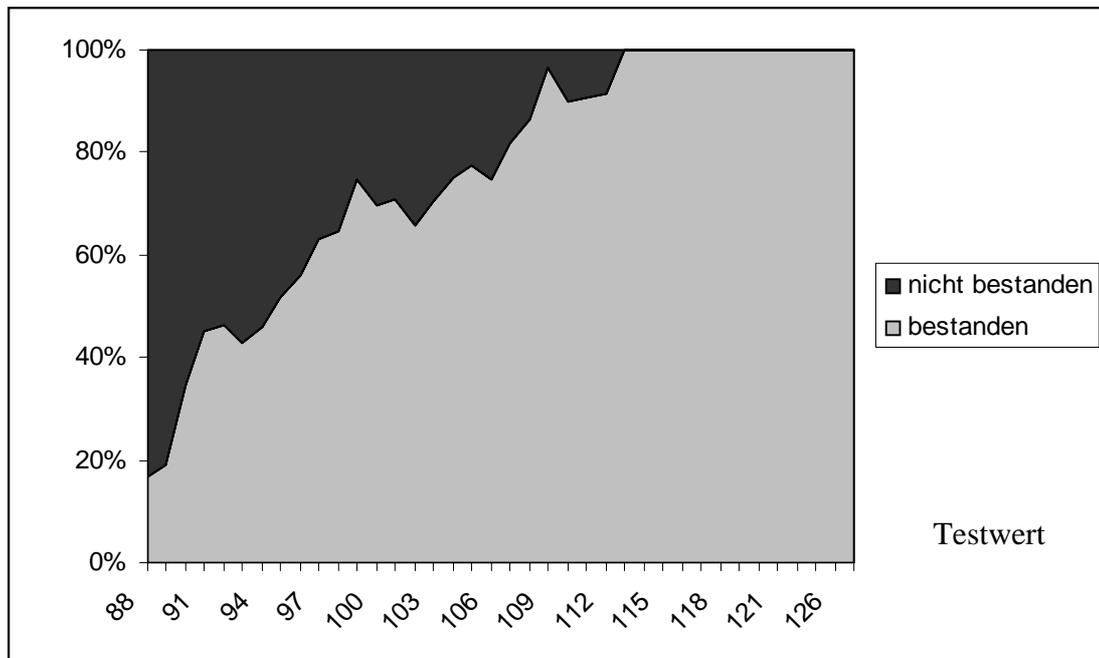


Abbildung 17. Prozentverhältnis "bestanden" und "nicht bestanden" pro individuellem Testwert (x-Achse)

Berechnet man die Bestehenswahrscheinlichkeit der Prüfung für jeden einzelnen Testwert lässt sich für den „linearen“ Teil der Kurve (Wahrscheinlichkeit kleiner als 1, d.h. Testwerte sind kleiner als 112) ein signifikanter linearer Zusammenhang zwischen Testwert und der Bestehenswahrscheinlichkeit nachweisen. Bei einem Bestimmtheitsmass r^2 von 0,76 ($F = 81,8$; $p < 0,001$) fällt dieser ausserordentlich hoch aus (Abbildung 18).

In der Abbildung 19 wurden die Personen einbezogen, die noch zu keiner Prüfung angetreten sind. Auch dieser Zusammenhang ist monoton: für Nichtantreten sind ebenfalls niedrigere Testwerte charakteristischer. Er ist allerdings wie erwartet nicht so deutlich wie bei den Personen, welche die Prüfung nicht bestehen. Sehr verschiedene Möglichkeiten sind denkbar, warum noch keine Prüfung absolviert worden ist, nicht alle müssen dabei von der Leistung abhängen.

¹ Eine Glättung der Kurve gleicht Einzelschwankungen aufgrund der Schwankung der Personenzahl pro Testwert aus. Dazu wurden für jeden Testwert die Häufigkeiten der beiden Nachbarwerte zu seiner Häufigkeit addiert.

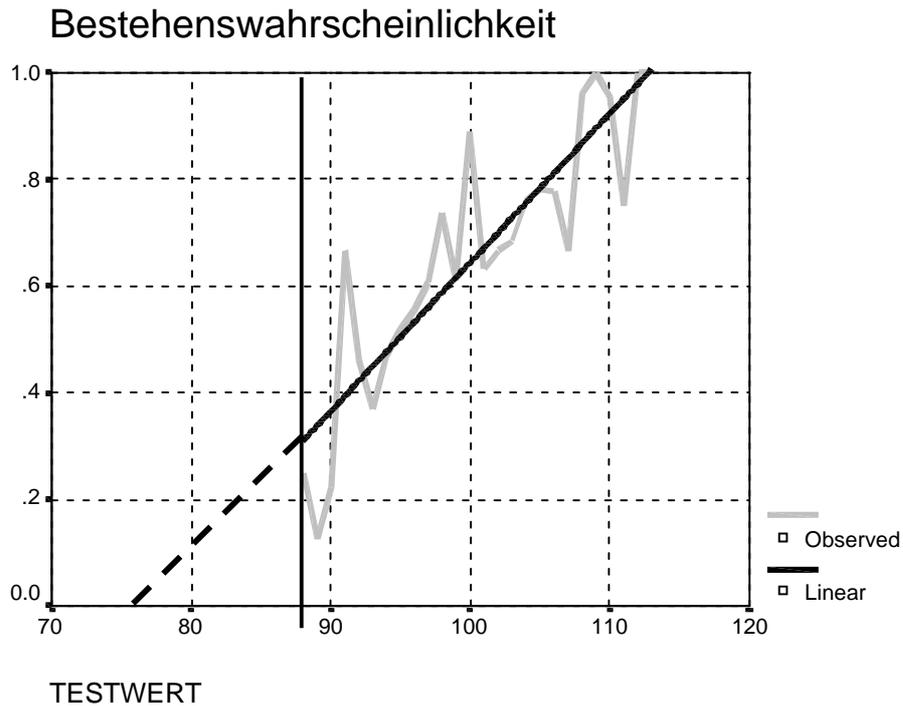


Abbildung 18. Lineare Regression der Bestehenswahrscheinlichkeit für jeden Testwert. Grau: beobachtete Werte, schwarz: Regressionsgerade. Gestrichelt: Schätzung für die Testwerte von Personen, die nicht an Prüfungen teilnahmen (nicht zugelassen worden sind)

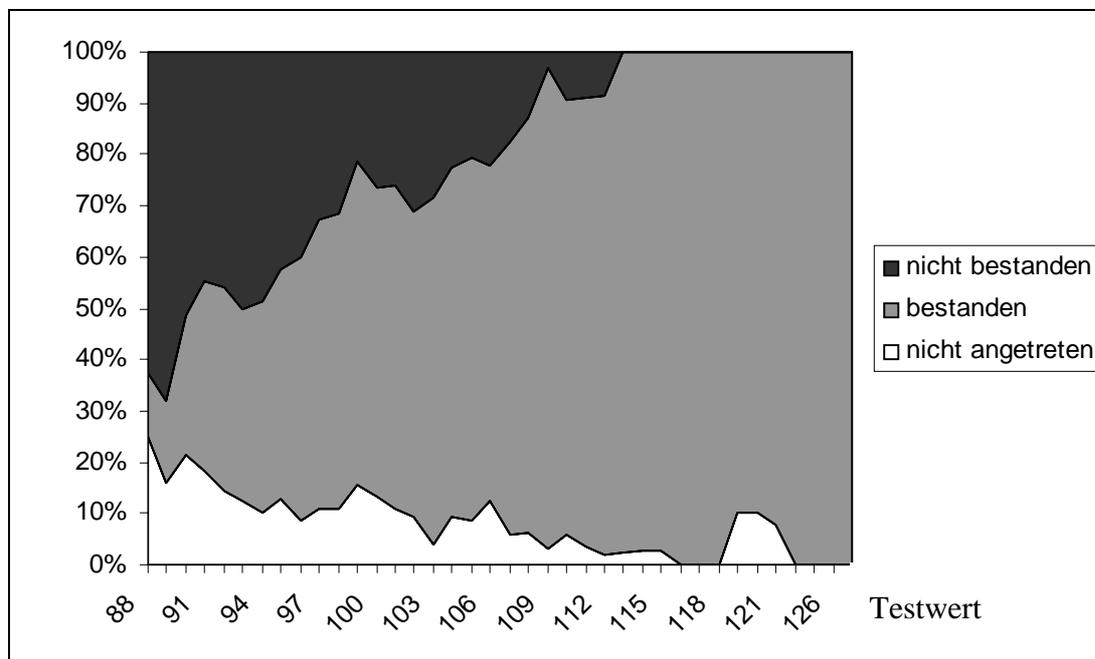


Abbildung 19. Prozentverhältnis "bestanden", "nicht bestanden" und "nicht angetreten" pro individuellem Testwert (x-Achse)

Die recht deutliche Abhängigkeit des Prüfungserfolges vom Testwert und damit der Studieneignung in der ersten Vorprüfung belegt, dass der Test tatsächlich so etwas wie ein „Probestudium“ ist, dass die Komplexität und die Art der Aufgaben richtig gewählt sind. Wie im Studium erfasst der Test auch die Fähigkeit, sich in einer neuen Situation zurechtzufinden und in kurzer Zeit effektive Lösungsstrategien zu entwickeln. Die zum Teil extreme Höhe dieses Zusammenhanges zeigt, dass sich zumindest in der Anfangsphase des Studiums die Test- und Studienanforderungen sehr ähneln.

In dieser ersten Studienphase ist auch die Abbrecherrate bekanntlich am höchsten. Deshalb ist es gerechtfertigt, die Studienzulassung vor allem auf den Erfolg in dieser kritischen Phase auszurichten und die begrenzte Ressource der Studienplätze fair an Personen mit den besten Erfolgsaussichten zu verteilen.

Es bleibt abzuwarten, ob die nachfolgenden Studienabschnitte ebenso genau vorhergesagt werden können. Dies wäre in dieser Höhe wie bei der ersten Vorprüfung sogar etwas überraschend.

Testwert und Prüfungsnote

Erfasst wurden die Prüfungsnoten, um innerhalb der Gruppen weiter differenzieren zu können. Es muss zunächst festgehalten werden, dass zwei Typen von Examen angewendet worden sind.¹ Die Mehrzahl der Personen hat ein Examen nach Typ 1 abgelegt – in Bern kam neu ein Examen vom Typ 2 zur Anwendung. Gleiche Benennung der Examen und gleiche Noten bedeuten allerdings nicht, dass die Anforderungen und die Noten-Massstäbe der Universitäten vergleichbar sein müssen (siehe Abbildung 7). Einige Prüfungen wurden schriftlich, einige mündlich abgenommen. Bei einigen schriftlichen Prüfungen kam ein Multiple-Choice-Verfahren zur Anwendung, einige wurden frei beantwortet. Dies ist in der Tabelle 18 entsprechend gekennzeichnet.

Die Analysen werden vor allem durchgeführt, um die aus Deutschland bekannten Ergebnisse zu überprüfen. In Deutschland wurde die Beziehung zur 1. Vorprüfung berechnet, die frühestens nach 2 Jahren stattfindet. Der Prognosezeitraum ist also mindestens um das Doppelte länger.

Die Benotungen verschiedener Universitäten wurden zusammengefasst, obwohl sie sich jeweils systematisch unterscheiden können. Erhoben wurde pro Person die jeweils letzte Note pro Teilprüfung – bei Wiederholungsprüfungen also die Note der Wiederholung.

Für die Examen vom Typ 1 können Korrelationsanalysen durchgeführt werden, weil die Fallzahl ausreichend hoch ist und auch die Notengebung zwischen den Stufen ausreichend variiert. Für Examen vom Typ 2 muss wegen der geringen Variation der Noten eine grössere Fallzahl abgewartet werden, wobei sich allerdings bereits jetzt deutliche Unterschiede zum Examenstyp 1 zeigen (s.u.).

¹ Die Medizinalprüfungen für Ärzte an allen medizinischen Fakultäten sind durch eidg. Prüfungsverordnungen geregelt. Während für den ersten Examenstyp die Verordnung über die Prüfung von Ärzten vom 19.11.80 gilt, regelt beim zweiten Examenstyp z.Zt. eine eidg. Verordnung vom 1.11.99 das besondere Prüfungsverfahren im Rahmen des im Herbst 96 eingeführten Berner Reformcurriculums (dies löst die erste Derogationsverordnung vom 24.11.96 ab).

Note	Examen Typ 1 BE: 64, BS: 132, FR: 75, ZH: 215				Examen Typ 2 BE: 49			
	Physik und Physiologie	Chemie und Biochemie	Allgemeine u. Humanbiologie I	Allgemeine u. Humanbiologie II	Examen 1 alle Lehrinhalte	Examen 2 alle Lehrinhalte	Examen 3 alle Lehrinhalte	Examen 4 mündlich
1	3	2	5	2				
2	35	40	23	15	2	2	3	
3	105	89	60	61	5	4	3	4
4	180	187	119	109	23	14	12	16
5	134	147	170	185	18	29	26	21
6	29	20	108	113	1		5	8
N	486	485	485	485	49	49	49	49

Tabelle 18. Notenverteilung für die Examen (Quelle: Prüfungsprotokolle im BAG), jeweils letzte Note pro Fach für jede Person. Note 6 ist die beste Note, ab Note 4 gilt eine Teilprüfung als bestanden

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Durchschnittsnote	Vergleich Deutschland ¹
Testwert	.46	.44	.48	.50	.54	.50
Muster zuordnen	.15	.18	.21	.20	.21	.17
Med.-naturw. Grundverständnis	.35	.32	.34	.38	.40	.41
Schlauchfiguren	.23	.21	.23	.26	.28	.28
Quantitative und formale Probleme	.44	.42	.37	.37	.46	.45
Textverständnis	.38	.32	.37	.40	.42	.41
Figuren lernen	.14	.16	.18	.20	.20	.22
Fakten lernen	.18	.22	.26	.26	.26	.17
Diagramme und Tabellen	.34	.31	.29	.32	.36	.41
Konzentr. und sorgf. Arbeiten	.14	.19	.24	.22	.22	.24
N	487	487	487	487	487	

Tabelle 19. Korrelation der Prüfungsnoten (alle vorliegenden Fälle mit Daten) mit dem EMS; Rangkorrelation nach Spearman – der Masskorrelationskoeffizient nach Pearson weicht um maximal 0.02 ab

Die Korrelationen liegen für Prognose-Koeffizienten im Niveau sehr hoch (Werte zwischen 0,3 und 0,4 werden üblicherweise erwartet). Mit 0,54 liegt die prognostische Validität des

¹ Trost, G. (Hrsg.) (1994). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht). ITB: Bonn, Seite 144; Noten für schriftliche Vorprüfung nach 2 Jahren

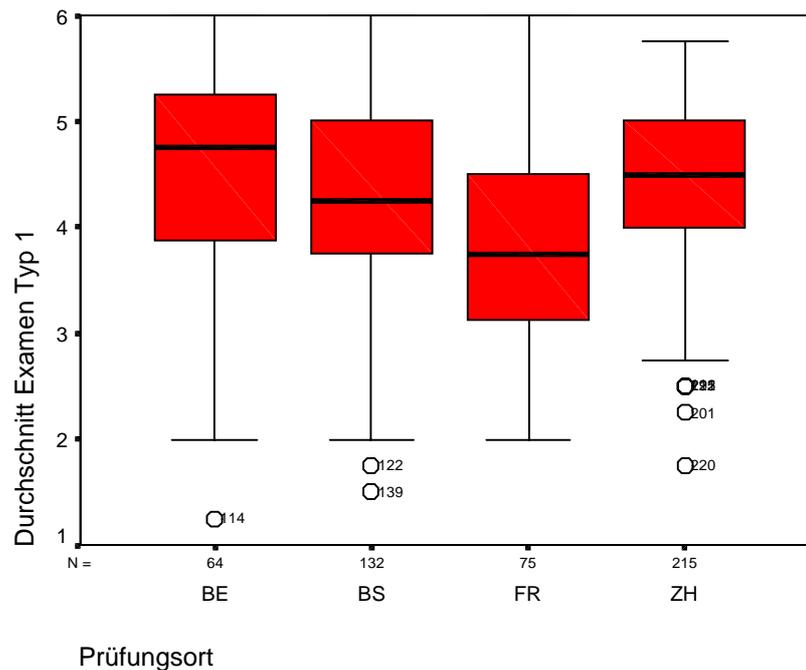


Abbildung 20. Boxplot der Durchschnittsnoten für Examen Typ 1 nach Prüfungsorten

Testwertes für den Notendurchschnitt ausserordentlich hoch und bestätigt die aus Deutschland bekannten Ergebnisse (dort Korrelation mit der Ärztlichen Vorprüfung, die in der Regel nach 4 Semestern abgelegt wird).

In Tabelle 20 werden die Korrelationen der Benotungen für die einzelnen Universitäten gegenübergestellt. Gemessen an der Erwartung für Koeffizienten prognostischer Validität sind alle im Bereich einer ausreichenden Höhe. Die „niedrigste“ Korrelation in Basel entspricht von Niveau her immer noch einer guten prognostischen Validität und lässt die Annahme eines gesicherten Zusammenhangs zwischen Noten und Testwert uneingeschränkt zu.

Die Tabelle enthält auch Angaben über die Art der Prüfung. Zwei Prüfungen werden mündlich, der Rest wird schriftlich abgenommen. Bei den schriftlichen Prüfungen gibt es Multiple-Choice-Aufgaben ebenso wie „Essay“-Aufgaben (Fragen sind durch kurze Texte zu beantworten), in einigen schriftlichen Prüfungen wird dies kombiniert.

Es deutet sich keine Regelmäßigkeit an, dass die Korrelationen auf die methodische Variante „Multiple-Choice“ zurückgeführt werden können (die bekanntlich auch im EMS eingesetzt wird). Vor allem in Zürich besteht die höchste Korrelation zu einer Prüfung vom „Essay“-Aufgabentyp, auch die kombinierten Prüfungen korrelieren nicht geringer mit dem Testwert als diejenigen, die nur Multiple-Choice-Aufgaben verwenden. In Basel schliesslich werden die drei „geringsten“ Korrelationen des Testwertes auch mit Multiple-Choice-Prüfungen gefunden. Der Frage wird in späteren Evaluationen natürlich weiter nachzugehen sein – es ergeben sich aber bisher keine Hinweise, dass es sich bei den Korrelationen von Studieneignung (Testwert) mit Wissen (Prüfung) um „artifizielle“ methodenbedingte Korrelationen handelt.

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Prüfungsdurchschnitt
Bern (Ex. 1) (n=64)	SCH-EM	SCH-M	SCH-M	SCH-M	
	.64	.51	.48	.58	.59
Basel (n=132)	SCH-M	SCH-EM	SCH-M	SCH-M	
	.33	.35	.30	.35	.38
Freiburg (n=75)	MÜ	SCH-M	MÜ	SCH-M	
	.42	.52	.44	.46	.56
Zürich (n=215)	SCH-EM	SCH-EM	SCH-E	SCH-M	
	.51	.47	.57	.51	.60

Tabelle 20. Korrelationen des Testwertes mit den Prüfungsnoten (Examen Typ 1) für die Universitäten (n.s.: nicht signifikant). SCH: schriftlich, MÜ: mündlich, E: Essay, M: Multiple Choice, EM: Kombination E und M

Höhere Korrelationen zwischen Testwert und Prüfungsnoten deuten auf einen grösseren Einfluss der individuellen Studieneignung auf den Prüfungserfolg hin – sehr hohe bedeuten aber auch, dass es weniger an Kompensationsmöglichkeiten gibt, eventuelle Defizite auf dem Gebiet der Fähigkeiten auszugleichen – etwa durch erhöhte Anstrengungen, Aufwendung von mehr Zeit (was beim Test bekanntlich nicht möglich ist) oder die Nutzung von Hilfe beim Wissenserwerb. Bei den Examen handelt es sich im Unterschied zum Test in der Regel um Wissensprüfungen, der Wissenserwerb sollte idealerweise so erfolgen, dass durch solche Kompensationsmechanismen trotz unterschiedlicher Ausgangsniveaus Wissen auf vergleichbaren Niveaustufen erworben wird. Insofern sollte eine „zu hohe“ Korrelation auch Denkanstoss sein, über optimale und individuell abgestimmte Fördermöglichkeiten nachzudenken.

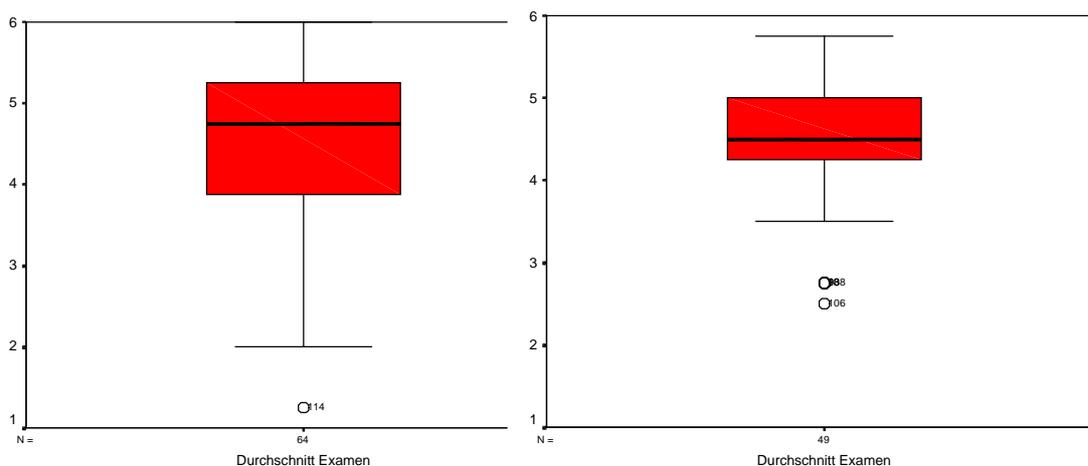


Abbildung 21. Boxplot für Examensdurchschnitte Typ 1 (links) und Typ 2 (rechts) für Universität Bern

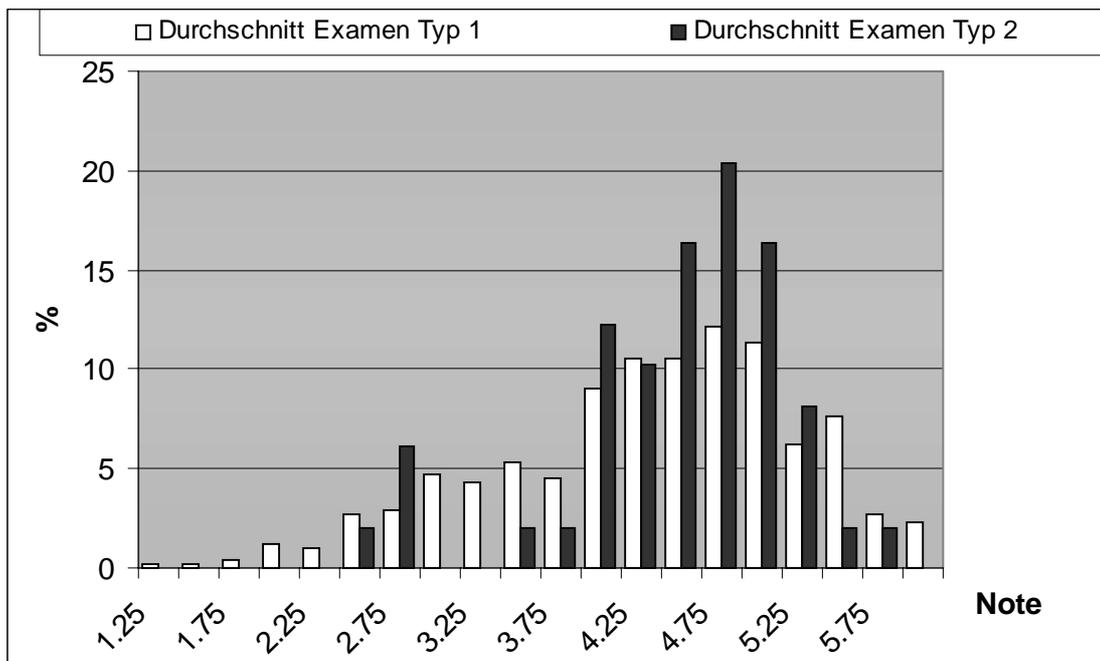


Abbildung 22. Verteilung der Prüfungsdurchschnittsnoten für die Examen Typ 1 (alle) und Typ 2 (Bern)

Zum Examen Typ 2 sei eine Anmerkung gestattet, weil sich hier ein Problem anzudeuten scheint. In Abbildung 21 (nur Bern) und 22 (Gesamt Typ 1 und Typ 2) werden die Verteilungen der Prüfungsdurchschnitte beider Examenstypen verglichen. Die Prüfungsdurchschnittsnoten des Examen Typ 2 treten gehäuft innerhalb von 1,5 Notenstufen auf und differenzieren deutlich weniger als die vom Examen Typ 1. Der EMS-Testwert der beider Gruppen in Bern unterscheidet sich dabei nicht (Examen 1 Mittelwert=103,3; Examen 2 Mittelwert=103,7), man muss also von vergleichbaren Fähigkeiten der Geprüften ausgehen.

Im Mittelwert unterscheiden sich Notendurchschnitte des Examenstyps 1 (4,3) nicht vom Examenstyp 2 (4,5). Dennoch ist die Bestehensquote für Examen Typ 2 mit 87,8% deutlich höher als Examen Typ 1 mit 71,8%. Hierfür können natürlich bessere Leistungen ebenso verantwortlich sein wie Probleme der Leistungsbewertung. Letztendlich wäre es eine erwartete Auswirkung der Zulassung nach der Studieneignung, dass alle Personen gute Leistungen erreichen und die Prüfung bestehen.

Da die Extremmeidung aber beide Extreme betrifft, scheint es sich um ein Problem der Notengebung zu handeln: Die Benotung im Examen Typ 2 könnte u.a. auf einen „Extremscheue-Effekt“ zurückgeführt werden, der eine Meidung schlechterer wie besserer Noten und eine bevorzugte Vergabe mittlerer Noten beinhaltet (siehe auch Tabelle 6). Bei neu eingeführten Notensystemen kann dies in der Anfangsphase auftreten, solange nur wenige Vergleichsmöglichkeiten existieren und/oder Misserfolge (für das Individuum wie das Ausbildungssystem) vermieden werden sollen. Wenn dies so bliebe, würde sich mangels einer Variation der Leistungsbewertung durch die Prüfung die Korrelationsanalysen mit dem EMS-Testwert erübrigen. Die Evaluation des EMS liesse sich unter diesen Umständen auf die Analyse des Bestehens der Prüfung beschränken.

Schlussfolgerungen für den EMS

Für die Evaluation des EMS bleibt festzuhalten:

1. Der Testwert **prognostiziert den Prüfungserfolg** mit sehr hoher Güte. Personen mit hohen Testwerten treten früher zur 1. Vorprüfung an und bestehen diese auch mit höherer Wahrscheinlichkeit im ersten Anlauf. Die Verwendung des EMS als Zulassungskriterium für den NC bietet die Gewähr, dass Personen bevorzugt einen Studienplatz erhalten, die das Studium wahrscheinlich schneller beenden (zunächst nur für die 1. Vorprüfung nachgewiesen). Damit können die vorhandenen Studienplatz-Ressourcen besser genutzt werden und letztendlich mehr Personen studieren.
2. Personen mit hohen Testleistungen erreichen mit hoher Wahrscheinlichkeit auch **bessere Prüfungs-Einzelnoten und Notendurchschnitte**. Geht man davon aus, dass die Prüfungsleistungen nebst weiteren Fähigkeiten und Fertigkeiten künftiger Ärzte für die Qualität der Berufsausbildung nicht unerheblich sind, so werden diese Leistungsniveaus angehoben.
3. Es finden sich Hinweise, dass Personen, die sich trotz zugeteiltem Studienplatz **nicht immatrikulierten, geringere Testwerte** erzielt haben als die erfolgreichen Personen (Prüfung im ersten Anlauf bestanden). Unter Berücksichtigung der Punkte 1 und 2 kann man dies – mit der gebotenen Vorsicht – als Hinweis interpretieren, dass die Studienmotivation und somit Dissuasionseffekte zumindest teilweise von der Studieneignung in gewünschter Weise beeinflusst sind und nicht unbedingt die Geeigneten „abgeschreckt“ werden.
4. Die insgesamt milde Wirkung des Numerus clausus innerhalb der Schweiz verhindert, dass bei der Zulassung eine Überbewertung des Eignungsaspektes erfolgt. Für den tatsächlichen Studienverlauf ist er jedoch in einer recht starken Art und Weise wirksam.

Die Eignung des EMS als Selektionsinstrument für die Studienzulassung kann aufgrund dieser Studie bestätigt werden.

Die Analyse der nachfolgenden Prüfungen und Geburtsjahre ist geplant und wird diese Ergebnisse weiter differenzierbar machen.

Technische Anmerkungen:

Der Datenschutz bei dieser Studie erfolgte unter Beachtung der gesetzlichen Vorschriften: In den Auswertedateien wurden alle Personen-Identifikationsmerkmale gelöscht. Ergebnisse werden so dargestellt, dass ein konkreter Personenbezug nicht ersichtlich ist.

Die Erfassung der Prüfungsnoten und der Prüfungsergebnisse im BAG erfolgte ohne Kenntnis der Testleistungen im EMS. Es wurden durch die Schweizerische Hochschulkonferenz (SHK) lediglich die Immatrikulationsnummern der Studienanfänger mitgeteilt und daraufhin im BAG die zugehörigen Prüfungsdaten ermittelt. Die Datei mit den Prüfungsergebnissen wurde der SHK übergeben und erst dort mit den Testergebnissen im EMS verbunden. Diese Datei wurde anschliessend anonymisiert und sie liegt der SHK als Masterdatei auch zu Kontrollzwecken der Ergebnisse vor, um eine ausreichende Objektivität des Evaluationsverfahrens zu gewährleisten. Alle Berechnungen im ZTD erfolgten dann mit einer Kopie dieser Masterdatei.

Testanwendung in der Schweiz 2000

Organisation der Testabnahme

Die Testabnahme fand an allen Testorten einheitlich statt. Pro Testlokal war ein Testleiter für den Ablauf verantwortlich. Ein Stellvertreter, ein Assistent und pro 50 Teilnehmer ein weiterer Assistent standen diesen für die Betreuung der Teilnehmer zur Seite. Testleiter und Stellvertreter wurden vorher auf einer eintägigen Veranstaltung geschult und insbesondere auf das genaue Einhalten des Ablaufes und einheitliche Reagieren auf Besonderheiten vorbereitet. In der Regel wurden wieder die Personen eingesetzt, die bereits 1998 und 1999 beteiligt waren und bereits über Erfahrungen verfügten. Im Juni fand eine eintägige Schulung aller Testleiter und Stellvertreter in Zürich statt.

Zur Gewährleistung der Rechtsfähigkeit wurde pro Testlokal ein ausführliches Protokoll geführt. Die Zeiten für Beginn, Dauer der Untertests und Dauer der Pause waren genau vorgegeben und wurden laut Protokoll in jedem Fall genau eingehalten. Die Identität der Teilnehmerinnen und Teilnehmer wurde anhand eines gültigen Personaldokumentes überprüft. Eine telefonische Hotline zum ZTD und zur Hochschulkonferenz war für die Testleiter eingerichtet.

Aus den Protokollen ergeben sich keine Hinweise, welche die Gültigkeit der Testabnahme einschränken. Eine Verwarnung betraf einen leichten Verstoss gegen Regeln zum Testablauf. In diesem Jahr brachen drei Personen die Testbearbeitung vor dem Ende ab.

Die formalen Bedingungen an allen Testorten waren vergleichbar und es traten keine Benachteiligungen auf.

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	Muster zuordnen	Med.-naturw. Grundverst.	Schlauchfiguren	Quantit. und formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konz. und sorgf. Arbeiten
Mittelwert	100	10.18	10.45	12.92	11.60	8.74	10.09	9.95	10.20	12.70
Stand.-abweich.	10,0	2.81	3.18	3.59	3.88	3.75	3.40	3.42	3.23	3.56
extremste Differenz	.038	.074	.074	.068	.070	.094	.073	.093	.075	.075
Kolmog.-Smirnov Z	1.070	2.094	2.108	1.911	1.991	2.666	2.073	2.638	2.109	2.121
Asymp. Sig.	.203	.000	.000	.001	.001	.000	.000	.000	.000	.000

Tabelle 21. Verteilungsprüfung auf Normalverteilung (Kolmogorov-Smirnov Test)

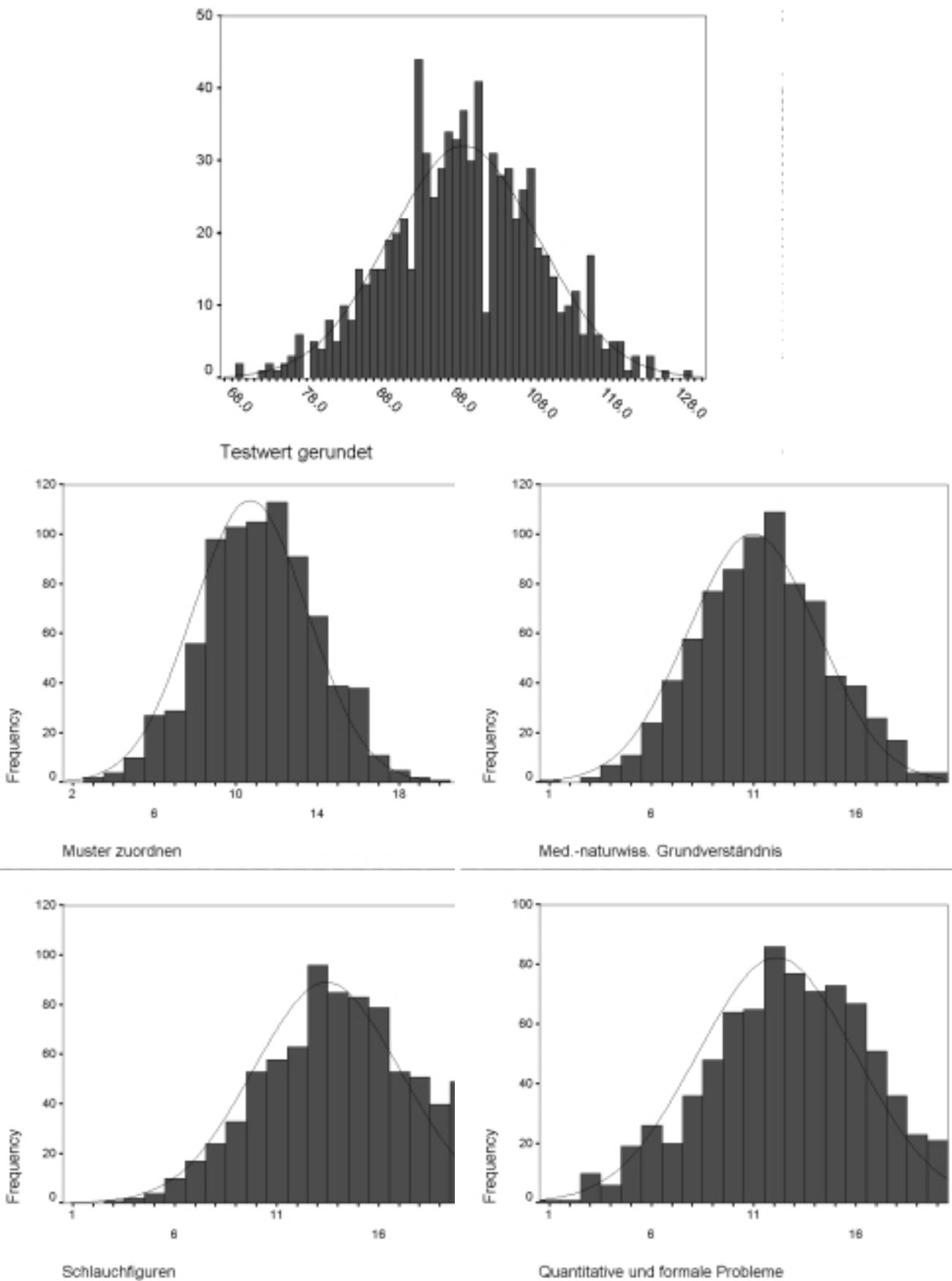


Abbildung 23. Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 1-4

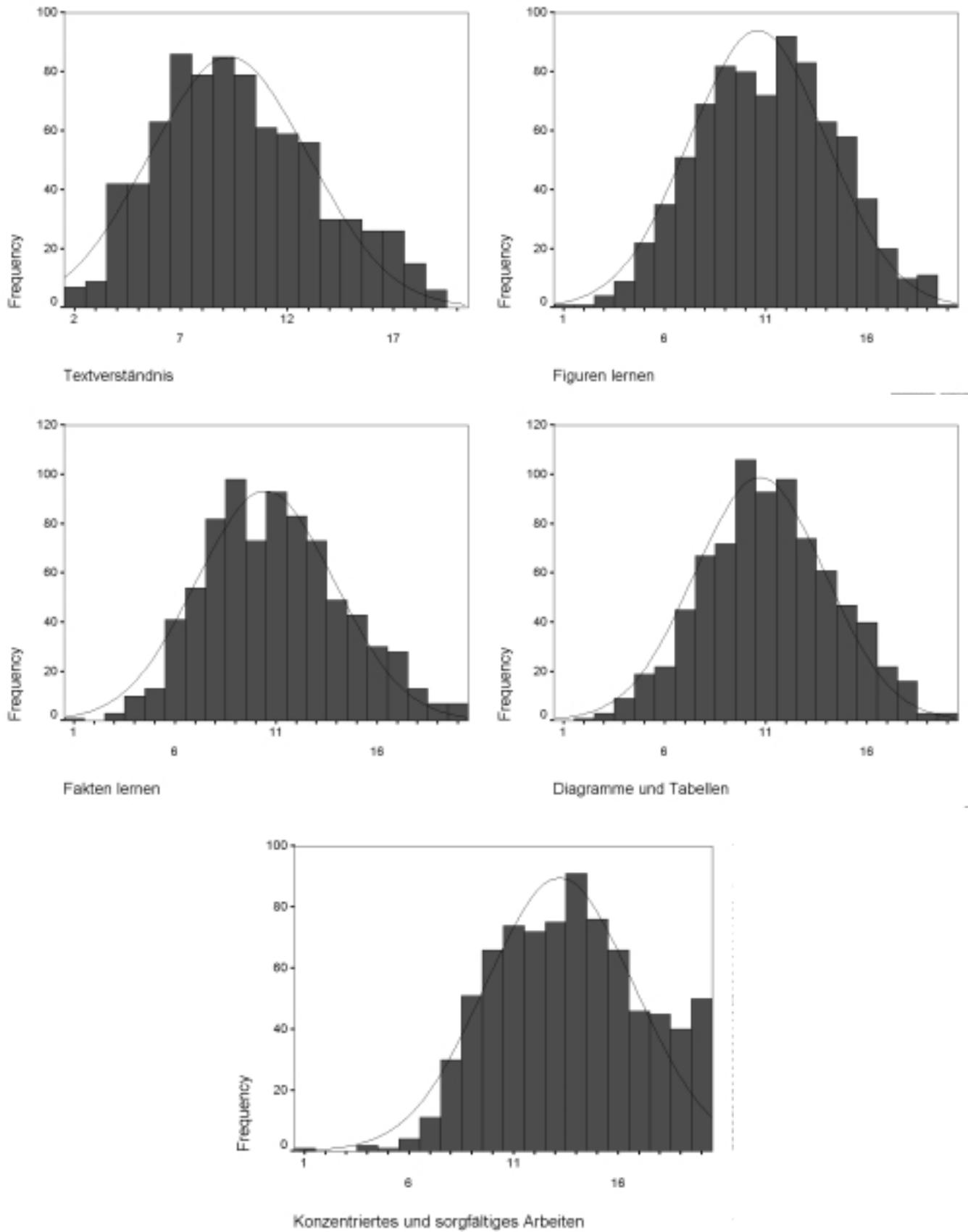


Abbildung 24. Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 5-9

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 so ausreichend, dass keine Testveränderung (z.B. durch Zeitverkürzung) notwendig wird.

Äquivalenz der Testformen 1 und 2

Wie in den Vorjahren wurden zwei pseudoparallele Testformen verwendet. Die nachfolgende Tabelle stellt Mittelwerte und Standardabweichungen beider Formen gegenüber. Die statistische Prüfung zeigt für den Testwert und alle Punktwerte der Untertests sowohl Varianzhomogenität als auch keine signifikanten Mittelwertunterschiede. Es hat keine Benachteiligungen bei einer der Testformen gegeben. Deshalb können beide Testformen in allen nachfolgenden Analysen zusammengefasst werden.

	FORM	N	M	S	Levene's Test for Equality of Variance		t-Test für unabhängige Stichproben		
					F	Sig.	t	df	Sig. (2seitig)
Testwert	1	405	100.14	9.90	.006	.937	.444	799	.657
	2	396	99.83	10.02					
Muster zuordnen	1	405	10.24	2.74	.162	.687	.634	799	.526
	2	396	10.12	2.88					
Med.-naturwiss. Grundverständnis	1	405	10.53	3.25	.971	.325	.698	799	.485
	2	396	10.37	3.12					
Schlauchfiguren	1	405	13.02	3.64	1.204	.273	.765	799	.444
	2	396	12.83	3.53					
Quantitative und formale Probleme	1	405	11.61	3.85	.132	.716	.051	799	.959
	2	396	11.59	3.91					
Textverständnis	1	405	8.80	3.86	.605	.437	.389	799	.698
	2	396	8.69	3.65					
Figuren lernen	1	405	10.11	3.40	.064	.801	.157	799	.875
	2	396	10.08	3.40					
Fakten lernen	1	405	9.95	3.44	.001	.969	-.047	799	.962
	2	396	9.96	3.40					
Diagramme und Tabellen	1	405	10.26	3.24	.109	.741	.494	799	.621
	2	396	10.14	3.22					
Konzentr. Und sorgf. Arbeiten	1	405	12.64	3.52	1.250	.264	-.409	799	.683
	2	396	12.75	3.61					

Tabelle 22. Mittelwerte und Standardabweichungen für die Formen 1 und 2, alle Mittelwert- und Varianzunterschiede sind statistisch nicht signifikant (F bzw. t-Test)

Vergleich der Testfassungen 1998 bis 2000

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. Indem die Gütekriterien in den Jahren identisch sind, 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 dabei über die Jahre geringfügigen Schwankungen. Diese wird in der deutschen Version als Addition der Einzelschwierigkeiten der einzelnen Aufgaben berechnet.

	D 2000	CH 2000 Gesamt	CH 2000 HM	D 1999	CH 1999 Gesamt	CH 1999 HM	D 1998	CH 1998 HM
Muster zuordnen	9.3	10.3	10.4	11.0	11.6	11.7	12.2	13.2
Med.-naturwiss. Grundverständnis	10.8	10.7	10.8	11.2	11.3	11.3	11.6	12.5
Schlauchfiguren	12.3	13.2	13.4	11.9	12.9	13.0	11.8	13.3
Quantitative und formale Probleme	10.2	11.8	12.0	9.7	11.6	11.8	9.9	11.9
Textverständnis	9.3	9.0	9.2	8.9	8.9	8.9	10.2	10.3
Figuren lernen	9.3	10.2	10.2	11.8	12.7	12.6	11.6	12.6
Fakten lernen	8.9	10.1	10.2	11.0	11.8	11.8	11.2	11.6
Diagramme und Tabellen	9.7	10.4	10.5	10.3	10,5	10.6	11.0	11,4

Tabelle 23. Mittelwerte (Punkte) der Untertests 1998 bis 2000 deutsche Sprachgruppe Schweiz und Schätzung aufgrund der Item-Schwierigkeitswerte aus Deutschland (HM: Humanmedizin)

Bezogen auf die einzelnen Untertests sind vor allem "Muster zuordnen" und "Textverständnis" schwieriger geworden. Auch die Unterschiede Deutschland-Schweiz bleiben nicht gleich: die in 1998 vorhandene bessere Leistung der Schweizer Gruppe in medizinisch-naturwissenschaftlichem Grundverständnis lässt sich 1999 nicht nachweisen. Bei "Schlauchfiguren" und "Quantitativen und formalen Problemen" bleibt sie dagegen nahezu identisch erhalten.

Bis auf „Medizinisch-naturwissenschaftliches Grundverständnis“ und „Textverständnis“ liegen die Schweizer Ergebnisse im Punktwert höher als in der entsprechenden deutschen Stichprobe. Dieser Unterschied wäre dann begründbar, wenn man die viel höhere Maturitätsquote und die höhere Bewerbungsquote im Fach Medizin in Deutschland in

Rechnung stellt. Dennoch geben die Ergebnisse zu denken, warum es genau diese beiden Untertests sind, die keine Unterschiede aufweisen. Es handelt sich um Aufgaben, die vor allem Textanforderungen enthalten. Oswald (1999) stellt fest, dass die Fähigkeiten, Textaufgaben zu lösen – d.h. in Texten elementare logische Zusammenhänge zu erkennen – relativ schlecht entwickelt seien. Wenn diese Aufgaben relativ zu anderen, wo es um vergleichsweise elementarere (nicht einfachere!) Zusammenhänge geht, schlechter gelöst werden, dann könnte dies ein Beleg für die Hypothese von Oswald sein.

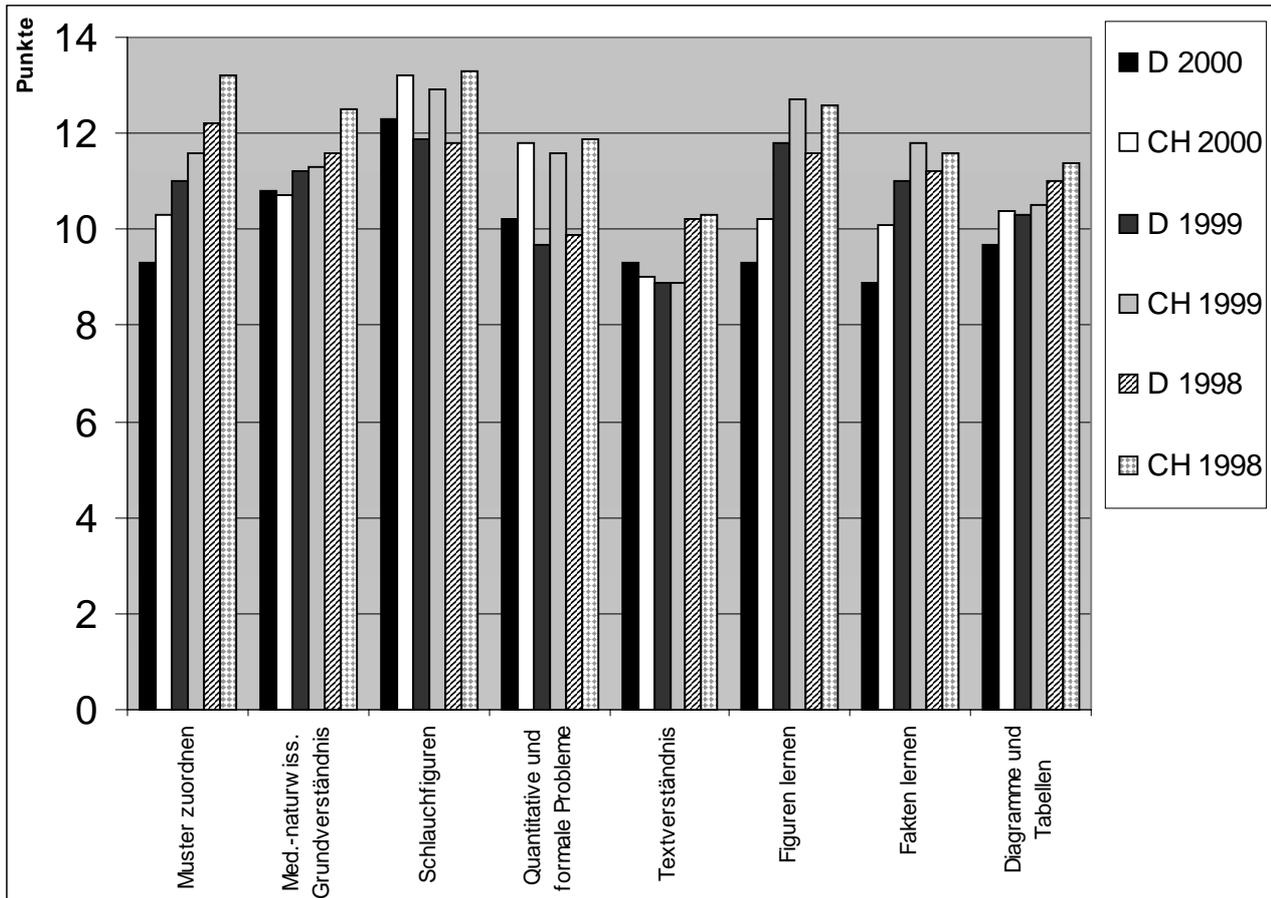


Abbildung 25. Mittelwerte für Punkte der Untertests für Schweiz (deutschsprachig) und Vergleichswerte Deutschland 1998 bis 2000 – jeweils Humanmedizin

Äquivalenz der Sprachversionen

Der Test wurde für den Einsatz im Jahr 2000 nach den international üblichen Kriterien aus der deutschen Fassung in die französisch- und die italienischsprachige Fassung adaptiert. In den Berichten 4 (1998) und 5 (1999) des ZTD wurde dieses Verfahren bereits ausführlich erläutert.

Die Erstübersetzungen erfolgten durch einen Fachübersetzer. Danach fanden je zwei Kontrollen pro Sprachform durch Zweisprachler und Bearbeitungen durch je einen zweisprachigen Mitarbeiter des ZTD statt. Für die Kontrollen wurden Personen ausgewählt, die mit den Studienanforderungen und dem Niveau der Maturität vertraut sind. Die Beteiligten verfügten bereits über Erfahrungen aus 1999.

Als Qualitätssicherungsinstrument wurden die Schwierigkeitsdiagramme aus 1998 und 1999 verwendet, die unterschiedliche Aufgabenschwierigkeiten sehr genau aufdecken. Als Zielgröße wurde eine Schwierigkeitsübereinstimmung vorgegeben. Der Satzbau sollte unbedingt der Zielsprache angepasst werden, wobei die Aufgabenschwierigkeit bestmöglich erhalten bleiben sollte.

Besonders bei den komplizierteren sprachabhängigen Aufgaben sind in der deutschsprachigen Fassung sehr komplexe Sätze verwendet worden, bei deren Übertragung in die Zielsprachen nicht dem deutschen Satzbau gefolgt werden kann.

Ein wichtiger Grundsatz bleibt, dass der Test in den drei Sprachgruppen gleich schwer sein muss – dass die Lösungswahrscheinlichkeit einer Aufgabe bei gleicher Fähigkeit tatsächlich gleich sein muss. 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 folgende Abbildung fasst zusammen, welche Ursachen für Unterschiede in den drei Sprachgruppen verantwortlich sein können:

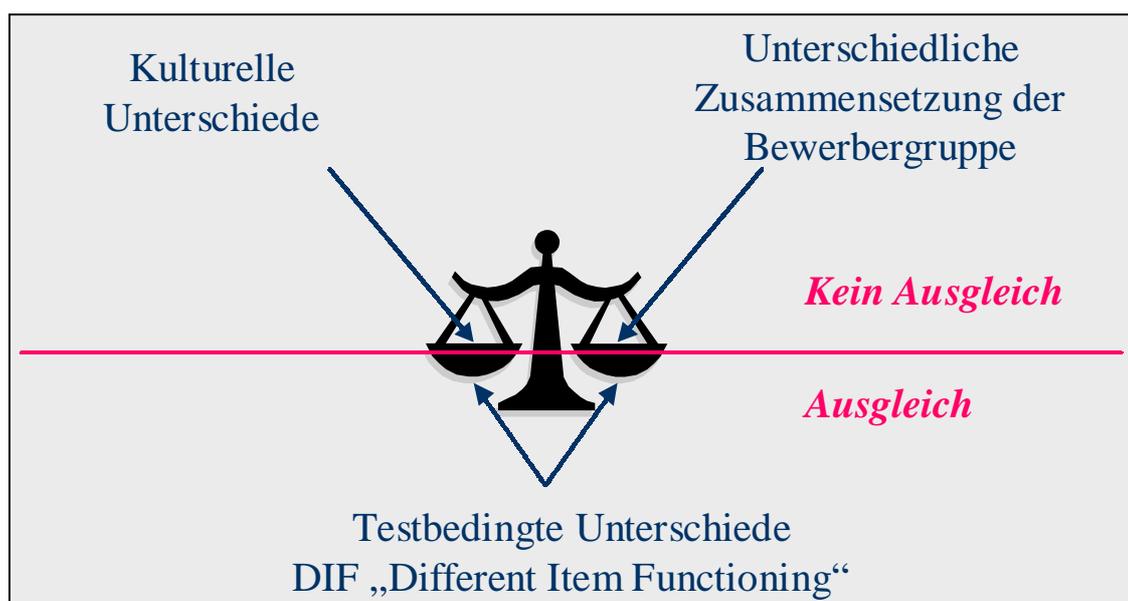


Abbildung 26. Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen

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 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.

Die ersten beiden Unterschiede sollten logischerweise nicht ausgeglichen werden – der dritte muss möglichst vollständig ausgeglichen werden. Gesucht ist eine Schätzmethode, die das Auffinden solcher testbedingten Unterschiede ermöglicht. Bei der Anwendung statistischer Verfahren machen sich die unterschiedlichen Stichprobengrößen und besonders die geringe Personenzahl in der italienischsprachigen Gruppe stark einschränkend bemerkbar. Die Anwendungen von Lineargleichungen zur Testwertschätzung wie in Israel ist beispielsweise nicht möglich.

Für das Jahr 2000 ist der folgende Beschluss der Hochschulkonferenz weiter bindend. Er beinhaltet,

1. einen Ausgleich nur bei sprachabhängigen Tests vorzunehmen. Dies sind fünf von neun Untertests. Der Sprachausgleich beschränkt sich demnach prinzipiell 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.

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.

Eine Adaptation von Tests führt beinahe zwangsläufig zu einem Sprach-Bias der einzelnen Items, also im Endeffekt zu einer möglichen Benachteiligung der einen oder anderen Sprachgruppe. In den USA und Israel liegen langjährig entsprechende Erfahrungen vor. Das Auftreten eines Bias auf Itemebene wird als "Differential Item Functioning" (DIF) bezeichnet. Solche Effekte sind vor allem in den eher sprachabhängigen Untertests zu erwarten, während weitgehend sprachfreie Untertests ("Figuren lernen", "Schlauchfiguren", "Konzentriertes und sorgfältiges Arbeiten") davon kaum betroffen sind. Spricht man von DIF, so muss dieser klar von systematischen Mittelwertsdifferenzen, welche den gesamten Untertest betreffen, abgegrenzt werden. Die Ergebnisse des letzten Jahres haben gezeigt, dass sich die Sprachgruppen in manchen Untertestscores signifikant unterscheiden. Diese Differenzen kommen nicht nur durch einzelne Items zustande, sondern auch durch eine systematische Verschiebung der Lösungsschwierigkeit über alle Items eines Untertests. Für diese systematischen Unterschiede ist es unwahrscheinlich, dass diese testbedingt sind. Sie sind vielmehr den Faktoren „Kulturunterschiede“ und „unterschiedliche Repräsentativität“ zuzurechnen. Deshalb sollen diese nicht ausgeglichen werden.

Als DIF wird nun verstanden, wenn die Abweichung eines Items in signifikanter Weise über diese systematische Verschiebung hinausgeht. Hier ist die Testbedingtheit des Unterschiedes sehr wahrscheinlich, weil nur dieses eine Item vom allgemeinen Trend abweicht.

Werden DIF-Statistiken untersucht, so finden sich im Normalfall auch Items, welche die Vergleichsgruppe gegenüber der Referenzgruppe bevorteilen. Wie solche Items behandelt werden, richtet sich nach dem Verständnis des Ausdrucks "Bias". Wird darunter ein Effekt verstanden, welcher die Vergleichsgruppe benachteiligt, so ist es sinnvoll, nur jene DIF-Items als "kritisch" zu betrachten, welche sich **nachteilig auf diese Gruppe** auswirken (Zieky, 1993). Da in unserem Fall die deutsche Testversion als "Referenz" gilt und eine durch die Übersetzung verursachte Benachteiligung der französisch- und italienischsprachigen Testteilnehmer ausgeschlossen werden sollte, wurde von einer "symmetrischen" Behandlung der DIF-Items abgesehen, und nur jene Items ausgeglichen, welche die nicht-deutschsprachigen Personen benachteiligten.

Während die DIF-Erkennung zur Identifikation der kritischen Items dient, sind für den effektiven Ausgleichsbetrag die Itemschwierigkeiten der Vergleichs- und Referenzgruppe massgebend. Weist ein Item DIF auf, so wird der betroffenen Sprachgruppe die Differenz der ermittelten Itemschwierigkeit zu derjenigen der Referenzgruppe (deutschsprachige Version) gutgeschrieben. Dies betrifft allerdings nur jene Testteilnehmer, welche das Item nicht gelöst haben.

Sprachvergleich für die Untertests

Wie 1998 und 1999 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Untertests auf. 1998 und 1999 erzielt die italienischsprachige Gruppe den besten Wert bei „Quantitative und formale Probleme“, in diesem Jahr nicht mehr – ansonsten sind die Werte der deutschen Sprachgruppe am besten.

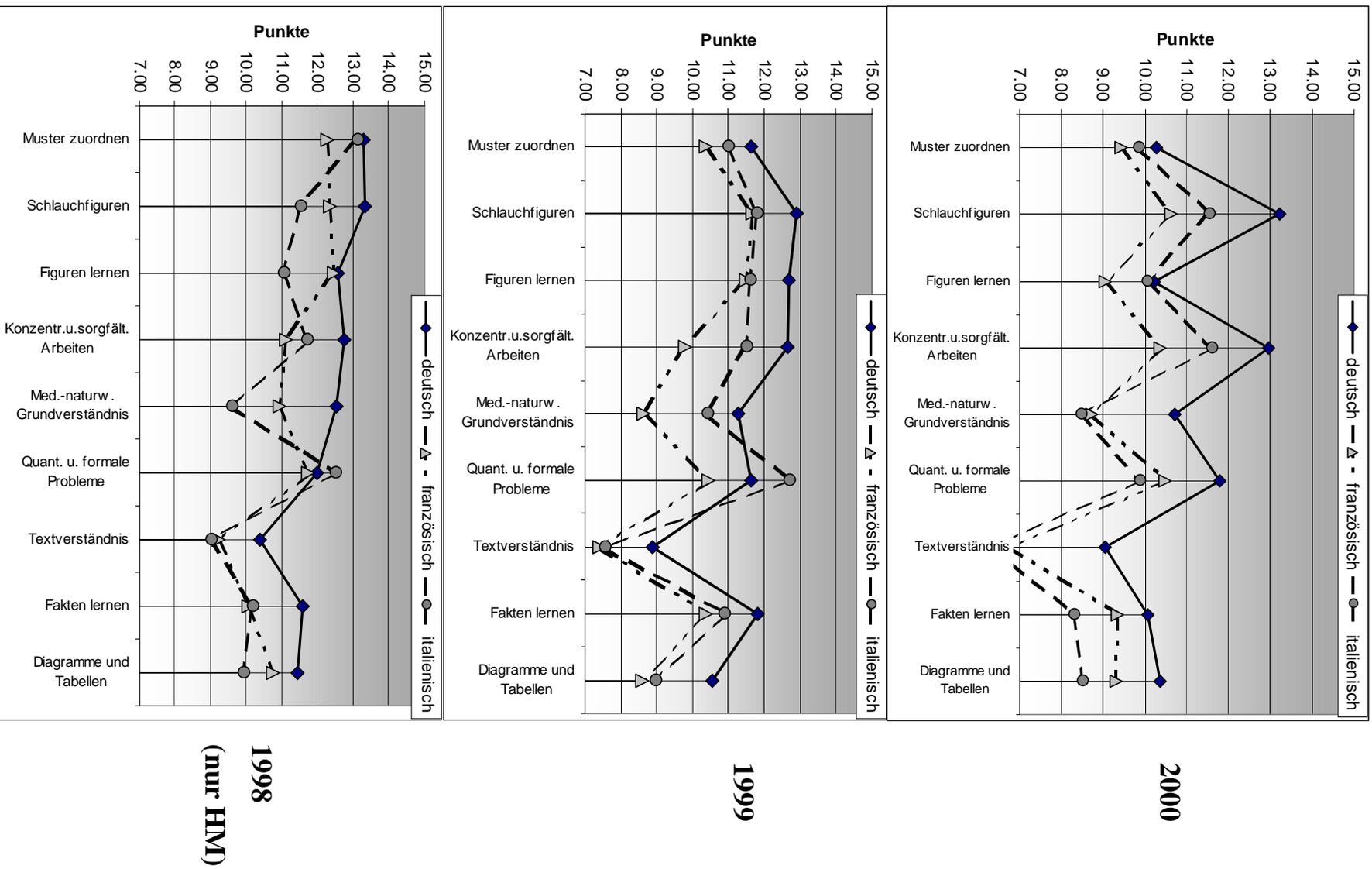


Abbildung 27. Mittelwerte der Untertests für die Sprachgruppen 1998-2000 (unkorrigiert)

		N	Mittelwert	Standardabw.	Minimum	Maximum
Testwert ** <i>D > (F,I)</i>	deutsch	703	100.75	9.93	70	130
	französisch	69	94.38	8.26	73	117
	italienisch	29	94.90	8.60	75	117
Muster zuordnen ** <i>homogen</i>	deutsch	703	10.27	2.82	2	19
	französisch	69	9.43	2.60	3	14
	italienisch	29	9.83	2.62	4	15
Med.-naturwiss. Grundverständnis ** <i>D > (F,I)</i>	deutsch	703	10.70	3.17	0	20
	französisch	69	8.72	2.64	3	15
	italienisch	29	8.45	2.67	3	13
Schlauchfiguren ** <i>D > (F,I)</i>	deutsch	703	13.21	3.53	3	20
	französisch	69	10.62	3.48	2	18
	italienisch	29	11.52	2.91	5	17
Quantitative und formale Probleme ** <i>D > I</i>	deutsch	703	11.78	3.85	0	20
	französisch	69	10.46	3.92	2	18
	italienisch	29	9.86	3.74	3	16
Textverständnis ** <i>D > (F,I)</i>	deutsch	703	9.04	3.75	1	18
	französisch	69	6.77	2.88	1	16
	italienisch	29	6.31	3.32	1	15
Figuren lernen ** <i>homogen</i>	deutsch	703	10.20	3.36	1	20
	französisch	69	9.04	3.71	0	18
	italienisch	29	10.03	3.35	4	18
Fakten lernen ** <i>D > I</i>	deutsch	703	10.08	3.48	0	20
	französisch	69	9.33	2.74	4	16
	italienisch	29	8.28	2.86	3	14
Diagramme und Tabellen ** <i>D > I</i>	deutsch	703	10.36	3.22	1	19
	französisch	69	9.32	3.16	2	15
	italienisch	29	8.48	2.89	3	13
Konzentriertes und sorgfältiges Arbeiten ** <i>D > F</i>	deutsch	703	12.97	3.52	3	20
	französisch	69	10.35	3.00	0	19
	italienisch	29	11.59	3.61	7	20

Tabelle 24. Punktwert und Untertests für die Sprachgruppen (deskriptive Statistiken)
 **: signifikanter Einfluss des Faktors "Sprache" bei varianzanalytischer Prüfung; daneben
 sind die signifikanten Unterschiede bei multiplem Mittelwertvergleich angegeben. In
 französischer und italienischer Sprachgruppe bereits korrigierte Punktwerte

Darstellung des Korrekturverfahrens

Es wurde ein differenziertes Korrekturverfahren erarbeitet, welches auf den ersten Blick etwas kompliziert scheint und statistische Vorkenntnisse voraussetzt. Im anschliessenden Kapitel „Evaluation“ wird nachgewiesen, dass es der Komplexität angemessen ist.

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 das sprachabhängige Verständnis der Instruktion scheint so einfach, dass Verständnisunterschiede unwahrscheinlich sind.
- 2) Damit ein Sprachausgleich in einem bestimmten sprachabhängigen Untertest erfolgt, müssen sich die entsprechenden Punktzahlen signifikant unterscheiden (Vergleich deutsch - französisch, beziehungsweise deutsch – italienisch mittels t-Test). Dies entspricht einem Beschluss der Schweizerischen Hochschulkonferenz. Wenn keine Mittelwert- und Varianzunterschiede vorhanden sind, trägt der Untertest auch nicht zu 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 Standardisierung der Untertests auf die jeweiligen Mittelwerte und Standardabweichungen und die Addition der Standardwerte zu einem Punktwert bzw. Testwert hätte den Nachteil, dass sich die Gewichte der Untertests beim Testwert ändern. Diese Abweichung vom aus Deutschland bewährten Prinzip würde möglicherweise die Prognosegüte für Studienerfolg beeinflussen – deshalb wurde davon abgesehen.

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. Es kann nicht mehr als einen Punkt pro gelöster Aufgabe geben. Auch dies würde sich auf die Gewichtung der Untertests negativ auswirken.

Welche Untertests ausgleichen?

Die folgenden Tabellen zeigen die Sprachvergleiche für die einzelnen Untertests. Referenz ist jeweils die deutschsprachige Testversion. Es zeigt sich, dass in allen sprachabhängigen Untertests signifikante Differenzen auftreten. Das heisst, das sowohl für die französische wie auch für die italienische Sprachgruppe alle fünf Untertests ausgeglichen werden müssen.

Untertest	Sprache	N	M	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	703	10.7027	3.1742	.000	Ja
	f	69	8.7246	2.6395		
Quantitative und formale Probleme	d	703	11.7838	3.8466	.007	Ja
	f	69	10.4638	3.9167		
Textverständnis	d	703	9.0384	3.7546	.000	Ja
	f	69	6.7681	2.8756		
Fakten lernen	d	703	10.0839	3.4772	.037	Ja
	f	69	9.3333	2.7422		
Diagramme und Tabellen	d	703	10.3585	3.2176	.011	Ja
	f	69	9.3188	3.1646		

Tabelle 25. Signifikanzprüfung deutsche vs. französische Sprachform

Untertest	Sprache	N	M	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	703	10.7027	3.1742	.000	Ja
	i	29	8.4483	2.6671		
Quantitative und formale Probleme	d	703	11.7838	3.8466	.008	Ja
	i	29	9.8621	3.7390		
Textverständnis	d	703	9.0384	3.7546	.000	Ja
	i	29	6.3103	3.3177		
Fakten lernen	d	703	10.0839	3.4772	.006	Ja
	i	29	8.2759	2.8648		
Diagramme und Tabellen	d	703	10.3585	3.2176	.002	Ja
	i	29	8.4828	2.8862		

Tabelle 26. Signifikanzprüfung deutsche vs. italienische Sprachform

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 " Δ -Werte" transformiert. Die Transformation erfolgt über die Formel

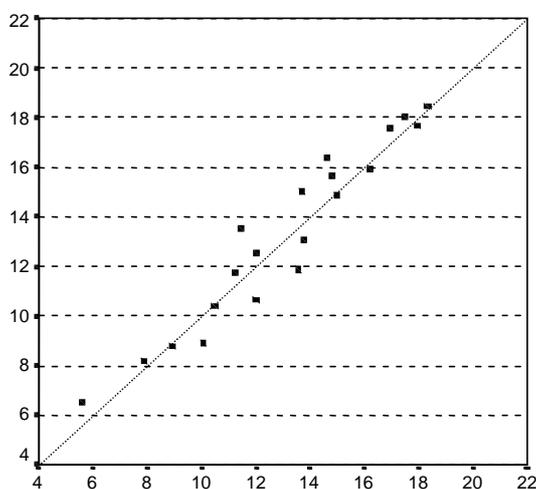
$$\Delta = 13 - 4z$$

Dies bedeutet, dass Δ 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 28). Dies würde bedeuten, dass sowohl die Itemschwierigkeiten wie auch deren Reihenfolge in beiden Sprachversionen vergleichbar sind.

Die Diagonale wird in der Folge durch eine gestrichelte 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) zu vermuten sind.

Auffällige Abweichungen einzelner Punkte von dieser Regressionsgeraden hingegen weisen



auf Items hin, die zusätzlich zu einer eventuellen systematischen Verschiebung spezielle Eigenschaften aufweisen. Diese Items stimmen in der untersuchten Sprachgruppe bezüglich der "Schwierigkeitshierarchie" nicht mit der Referenzgruppe überein. 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.

Abbildung 28. Beispiel für einen Delta-Plot

Regressionsgeraden werden in der Folge als durchgezogene Linien dargestellt. Massgeblich für die Entscheidung, ob bei einem Item DIF vorliegt, ist die Distanz des entsprechenden Punktes von dieser Geraden. Die Distanz 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 Untertest 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 (als Absolutbetrag) 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.

Medizinisch-naturwissenschaftliches Grundverständnis

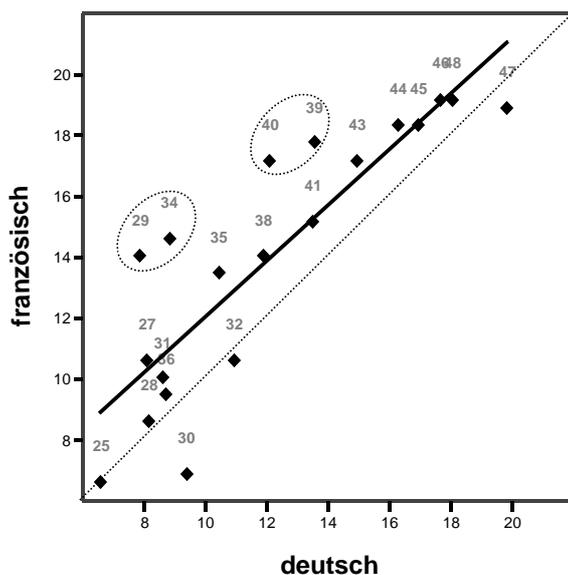


Abbildung 29. Delta-Plot für "MNGV" (deutsch-französisch)

Abbildung 29 zeigt den Delta-Plot für den Untertest „MNGV“ für den Vergleich zwischen deutsch- und französischsprachiger Version, die Abbildung 30 denjenigen zwischen deutsch- und italienischsprachiger Version.

Insgesamt zeigen beide Abbildungen ein recht klares Muster in Form einer Parallelverschiebung der Regressionsgeraden von der Diagonalen.

In der französischen Version weisen die Items 29, 34, 39 und 40 auffällige DIF-Werte auf. Für die italienische Version sind es die Items 25 und 29. Die Items 30, respektive 35 weisen starke Abweichungen in der entgegengesetzten Richtung auf („Erleichterung“ für die übersetzten Versionen). Da keine symmetrische DIF-Behandlung verwendet wird, werden diese Differenzen nicht ausgeglichen.

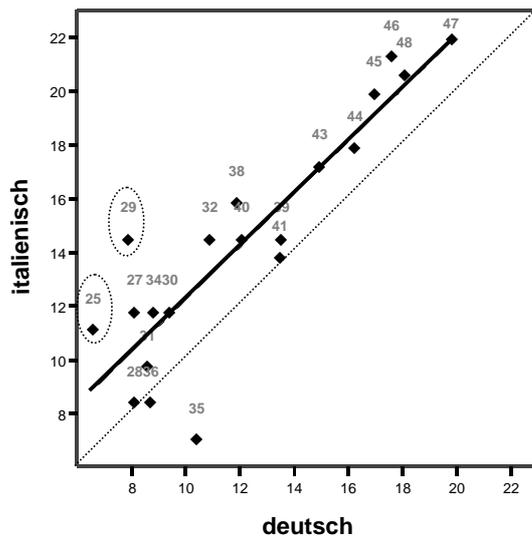


Abbildung 30. Delta-Plot für "MNGV" (deutsch-italienisch)

Die genauen Werte für beide Sprachvergleiche sind aus Tabelle 27 ersichtlich. Maximal können hier den französisch- beziehungsweise italienischsprachigen Teilnehmern 1.10 respektive 0.57 Punkte gutgeschrieben werden.

Item	Schwierigkeit			Δ -Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
25	0.82	0.81	0.59	6.58	6.64	11.09	1.74	-1.57		0.23
27	0.74	0.61	0.55	8.09	10.65	11.77	-0.18	-0.99		
28	0.74	0.71	0.72	8.12	8.64	8.37	1.31	1.45		
29	0.75	0.43	0.41	7.84	14.07	14.49	-2.87	-3.11	0.32	0.34
30	0.67	0.80	0.55	9.38	6.93	11.77	3.43	-0.09		
31	0.71	0.64	0.66	8.59	10.07	9.73	0.58	0.82		
32	0.60	0.61	0.41	10.90	10.65	14.49	1.72	-0.97		
34	0.70	0.41	0.55	8.82	14.65	11.77	-2.63	-0.48	0.30	
35	0.62	0.46	0.79	10.42	13.50	7.01	-0.71	4.04		
36	0.71	0.67	0.72	8.71	9.50	8.37	1.08	1.87		
38	0.55	0.43	0.34	11.85	14.07	15.85	-0.16	-1.28		
39	0.46	0.25	0.41	13.50	17.79	14.49	-1.77	0.85	0.22	
40	0.54	0.28	0.41	12.07	17.22	14.49	-2.32	-0.15	0.26	
41	0.47	0.38	0.45	13.45	15.22	13.81	0.08	1.30		
43	0.39	0.28	0.28	14.93	17.22	17.21	-0.38	-0.09		
44	0.32	0.22	0.24	16.25	18.36	17.89	-0.33	0.35		
45	0.29	0.22	0.14	16.93	18.36	19.93	0.13	-0.64		
46	0.25	0.17	0.07	17.63	19.22	21.29	-0.03	-1.12		
47	0.14	0.19	0.03	19.82	18.93	21.97	1.66	-0.07		
48	0.23	0.17	0.10	18.05	19.22	20.61	0.25	-0.34		

Tabelle 27. DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“

Quantitative und formale Probleme

Die folgenden Abbildungen zeigen die Delta-Plots für den Untertest „Quantitative und formale Probleme“. Auch hier kann eine klare Parallelverschiebung der Regressionsgeraden beobachtet werden (Achsenabschnitte: 1.24 / 1.79; Steigung: jeweils 1.01)

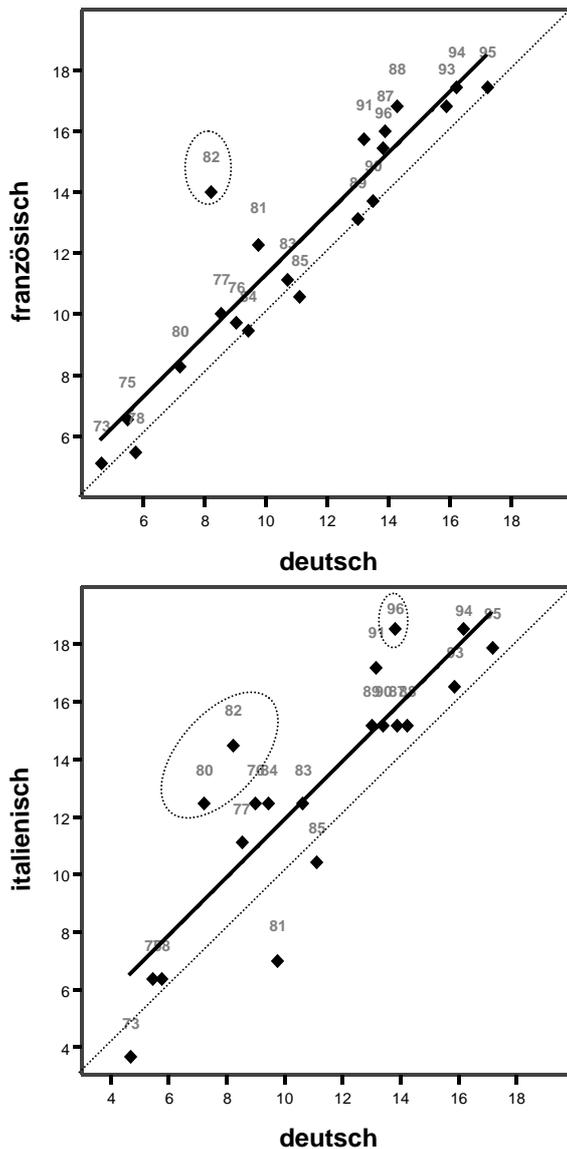


Abbildung 31. Delta-Plot für "Quantitative und formale Probleme" (deutsch-französisch)

In der französischen Version weicht einzig das Item 82 auffällig von der Regressionsgeraden ab. Die übrigen Items liegen eng an der Regressionsgeraden.

Abbildung 32. Delta-Plot für "Quantitative und formale Probleme" (deutsch-italienisch)

Die Items der italienischen Version streuen stärker um die Regressionsgerade (Abbildung 32). Die Items 80, 82 und 96 weisen Differenzen über 1.5 auf und werden korrigiert. Das Item 81 wurde möglicherweise durch die Übersetzung „vereinfacht“.

Insgesamt können hier den Kandidaten maximal 0.3 beziehungsweise 0.83 Punkte gutgeschrieben werden.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
73	0.91	0.88	0.97	4.67	5.21	3.61	0.52	2.04		
75	0.87	0.81	0.83	5.45	6.64	6.33	0.07	0.68		
76	0.69	0.65	0.52	9.02	9.79	12.45	0.39	-1.09		
77	0.72	0.64	0.59	8.54	10.07	11.09	-0.15	-0.47		
78	0.86	0.87	0.83	5.73	5.50	6.33	1.08	0.88		
80	0.78	0.72	0.52	7.22	8.36	12.45	0.12	-2.37		0.27
81	0.65	0.52	0.79	9.75	12.36	7.01	-0.90	3.25		
82	0.73	0.43	0.41	8.20	14.07	14.49	-3.20	-3.11	0.30	0.32
83	0.61	0.58	0.52	10.67	11.22	12.45	0.56	0.08		
84	0.67	0.67	0.52	9.44	9.50	12.45	0.89	-0.79		
85	0.59	0.61	0.62	11.09	10.65	10.41	1.26	1.82		
87	0.45	0.33	0.38	13.87	16.08	15.17	-0.58	0.44		
88	0.43	0.29	0.38	14.23	16.93	15.17	-0.93	0.70		
89	0.49	0.48	0.38	13.00	13.22	15.17	0.81	-0.18		
90	0.47	0.45	0.38	13.45	13.79	15.17	0.73	0.14		
91	0.48	0.35	0.28	13.20	15.79	17.21	-0.86	-1.47		
93	0.34	0.29	0.31	15.89	16.93	16.53	0.25	0.92		
94	0.33	0.26	0.21	16.20	17.50	18.57	0.07	-0.30		
95	0.28	0.26	0.24	17.18	17.50	17.89	0.76	0.88		
96	0.45	0.36	0.21	13.81	15.50	18.57	-0.22	-1.99		0.24

Tabelle 28. DIF-Analyse „Quantitative und formale Probleme“

Textverständnis

Für den Untertest „Textverständnis“ kann im Vergleich der deutschsprachigen mit der französischen Version (Abbildung 33) erneut eine Parallelverschiebung der Regressionsgeraden beobachtet werden (Achsen Schnittpunkt=3.17, Steigung = 0.96). Die italienischsprachige Version zeigt allerdings deutlichere Unterschiede (Abbildung 34) in Richtung einer stärkeren „Erschwerung“ der leichteren Aufgaben.

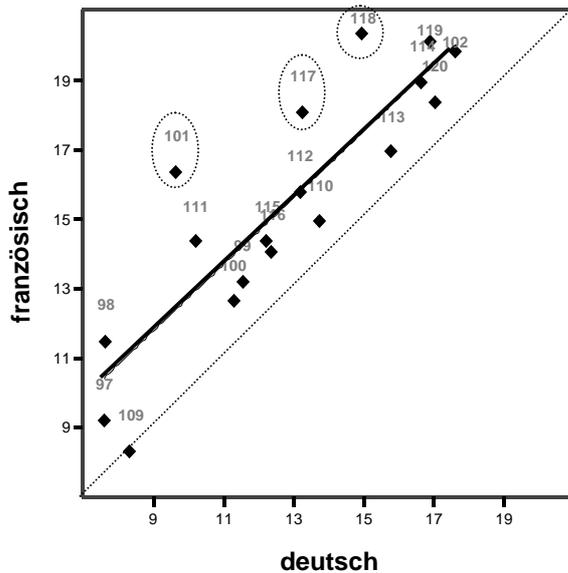


Abbildung 33. Delta-Plot für "Textverständnis" (deutsch-französisch)

In der französischen Version sind drei Items auszugleichen (101, 117, 118). Die übrigen Items liegen recht nahe an der Regressionsgeraden. Item 109 liegt praktisch auf der Diagonalen, weist also in beiden Versionen fast identische Schwierigkeiten auf. Aufgrund der Lage der Regressionsgeraden ist zu vermuten, dass es für die französische Version durch die Übersetzung eher „erleichtert“ wurde.

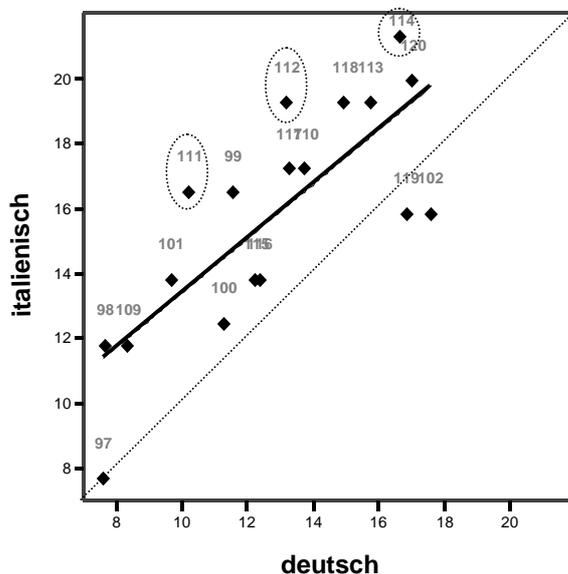


Abbildung 34. Delta-Plot für "Textverständnis" (deutsch-italienisch)

In der italienischsprachigen Version sind in beiden Richtungen einige klare Abweichungen zu beobachten. Durch das nicht-symmetrische Ausgleichsverfahren werden jedoch nur die Items 111, 112 und 114 berücksichtigt. Aus Tabelle 30 sind die genauen Werte für beide Sprachversionen ersichtlich. Insgesamt werden für diesen Untertest maximal 0.85 beziehungsweise 0.86 Punkte gutgeschrieben.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
97	0.76	0.68	0.76	7.59	9.22	7.69	0.89	2.87		
98	0.76	0.57	0.55	7.67	11.50	11.77	-0.70	-0.20		
99	0.56	0.48	0.31	11.57	13.22	16.53	0.76	-1.34		
100	0.58	0.51	0.52	11.29	12.65	12.45	0.98	1.60		
101	0.66	0.32	0.45	9.66	16.36	13.81	-2.83	-0.49	0.34	
102	0.26	0.14	0.34	17.60	19.79	15.85	0.20	3.06		
109	0.73	0.72	0.55	8.34	8.36	11.77	2.03	0.23		
110	0.45	0.39	0.28	13.76	14.93	17.21	1.04	-0.45		
111	0.63	0.42	0.31	10.22	14.36	16.53	-0.99	-2.21		0.32
112	0.48	0.35	0.17	13.20	15.79	19.25	0.03	-2.38		0.31
113	0.35	0.29	0.17	15.78	16.93	19.25	1.00	-0.72		
114	0.30	0.19	0.07	16.67	18.93	21.29	0.18	-1.70		0.23
115	0.53	0.42	0.45	12.24	14.36	13.81	0.40	1.17		
116	0.52	0.43	0.45	12.38	14.07	13.81	0.71	1.27		
117	0.48	0.23	0.28	13.25	18.08	17.21	-1.58	-0.78	0.24	
118	0.39	0.12	0.17	14.96	20.36	19.25	-2.04	-1.24	0.27	
119	0.29	0.13	0.34	16.87	20.08	15.85	-0.51	2.59		
120	0.28	0.22	0.14	17.04	18.36	19.93	0.84	-0.43		

Tabelle 29. DIF-Analyse „Textverständnis“

Fakten lernen

Die Regressionsgeraden für den Untertest „Fakten lernen“ weisen für beide Sprachversionen auf eine Vereinfachung der „schwierigen“ Items hin. Die Steigungen der Geraden betragen 0.63 (französisch), beziehungsweise 0.72 (italienisch).

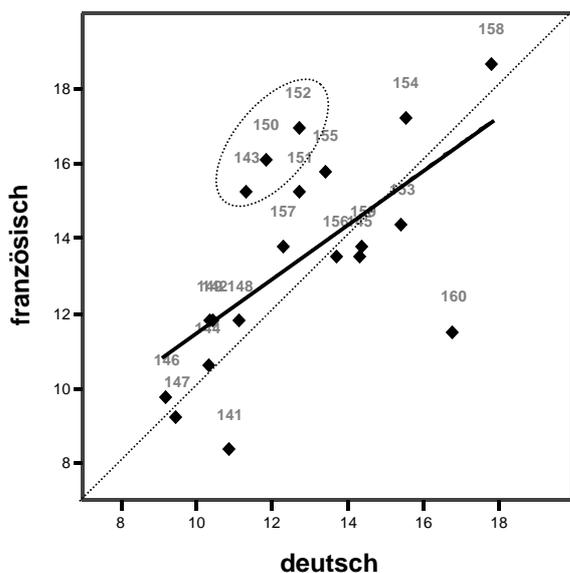


Abbildung 35. Delta-Plot für "Fakten lernen" (deutsch-französisch)

In der französischen Version müssen die Items 143, 150 und 152 ausgeglichen werden. Die Items 141 und 160 werden aufgrund der Abweichungsrichtung nicht berücksichtigt.

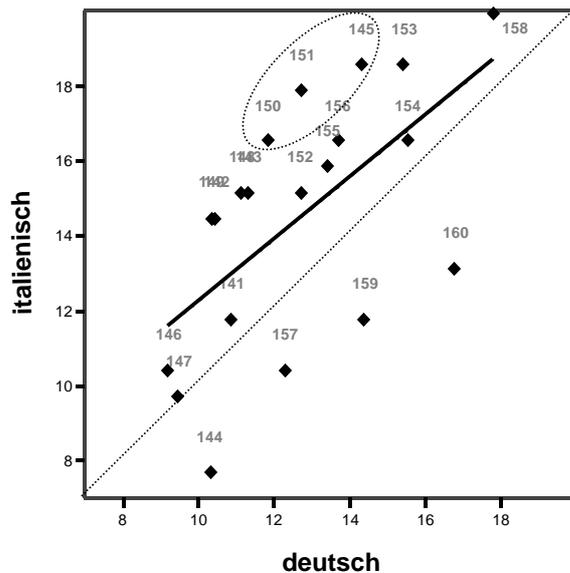


Abbildung 36. Delta-Plot für "Fakten lernen" (deutsch-italienisch)

Auch in der italienischen Version müssen drei Items ausgeglichen werden. Betroffen sind hier die Nummern 145, 150 und 151. Größere Abweichungen in der entgegengesetzten Richtung weisen ausserdem die Items 144, 157, 159 und 160 auf.

In diesem Untertest werden maximal 0.63 (französisch) beziehungsweise 0.72 Punkte (italienisch) als Bonus gutgeschrieben (Tabelle 30).

Item	Schwierigkeit			Δ -Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
141	0.60	0.72	0.55	10.87	8.359	11.77	3.05	0.96		
142	0.62	0.55	0.41	10.47	11.79	14.49	0.05	-1.38		
143	0.57	0.38	0.38	11.32	15.22	15.17	-2.23	-1.37	0.20	
144	0.63	0.61	0.76	10.31	10.65	7.689	0.87	3.74		
145	0.42	0.46	0.21	14.32	13.5	18.57	0.93	-2.07		0.22
146	0.68	0.65	0.62	9.212	9.788	10.41	0.92	0.95		
147	0.67	0.68	0.66	9.465	9.216	9.729	1.53	1.64		
148	0.58	0.55	0.38	11.15	11.79	15.17	0.44	-1.47		
149	0.62	0.55	0.41	10.36	11.79	14.49	-0.02	-1.45		
150	0.55	0.33	0.31	11.85	16.08	16.53	-2.61	-2.07	0.21	0.24
151	0.50	0.38	0.24	12.72	15.22	17.89	-1.40	-2.56		0.26
152	0.50	0.29	0.38	12.69	16.93	15.17	-2.80	-0.49	0.22	
153	0.37	0.42	0.21	15.44	14.36	18.57	0.90	-1.35		
154	0.36	0.28	0.31	15.55	17.22	16.53	-1.35	0.29		
155	0.47	0.35	0.34	13.42	15.79	15.85	-1.45	-0.55		
156	0.45	0.46	0.31	13.7	13.5	16.53	0.56	-0.89		
157	0.52	0.45	0.62	12.3	13.79	10.41	-0.49	2.92		
158	0.25	0.20	0.14	17.8	18.65	19.93	-1.18	-0.89		
159	0.42	0.45	0.55	14.4	13.79	11.77	0.75	3.22		
160	0.30	0.57	0.48	16.79	11.5	13.13	4.00	3.70		

Tabelle 30. DIF-Analyse „Fakten lernen“

Diagramme und Tabellen

Für die französische Version des Untertests "Diagramme und Tabellen" ist eine leichte Parallelverschiebung der Regressionsgeraden zu beobachten (Achsenabschnitt: 0.77; Steigung: 1.02). Das Diagramm für die italienische Version (Abbildung 38) spricht eher für eine Vereinfachung der schwierigen Items.

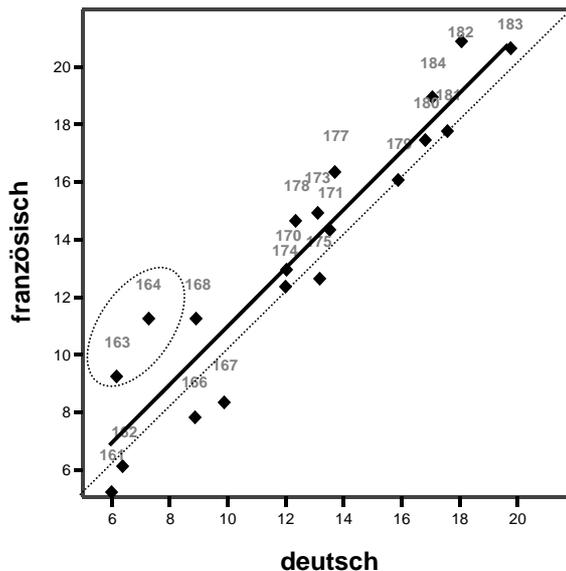


Abbildung 37. Delta-Plot für "Diagramme und Tabellen" (deutsch-französisch)

Im Vergleich deutsch-französisch liegen die Punkte eng an der Regressionsgeraden. Einzig die beiden Items 163 und 164 müssen ausgeglichen werden.

In der entgegengesetzten Richtung würde nur Item 167 eine auffällige Differenz aufweisen.

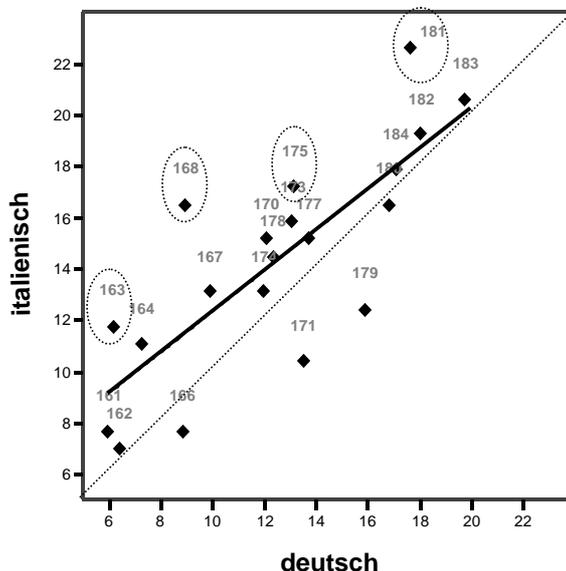


Abbildung 38. Delta-Plot für "Diagramme und Tabellen" (deutsch-italienisch)

Wie bei den vorhergehenden Untertests streut auch hier die italienische Version stärker um die Regressionsgerade. Folgende 4 Items müssen ausgeglichen werden: 163, 168, 175, 181.

Items, welche in der italienischsprachigen Version eher „vereinfacht“ wurden, sind 166, 171, 179.

Für die französischsprachigen Teilnehmer werden maximal 0.36 Punkte ausgeglichen, für die italienischsprachigen sind es maximal 1.12.

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
161	0.84	0.88	0.76	5.99	5.21	7.69	1.16	1.14		
162	0.82	0.84	0.79	6.41	6.07	7.01	0.86	1.93		
163	0.84	0.68	0.55	6.15	9.22	11.77	-1.52	-1.94	0.16	0.28
164	0.78	0.58	0.59	7.25	11.22	11.09	-2.14	-0.73	0.20	
166	0.7	0.75	0.76	8.85	7.79	7.69	1.41	2.93		
167	0.65	0.72	0.48	9.89	8.36	13.13	1.75	-0.67		
168	0.7	0.58	0.31	8.93	11.22	16.53	-0.94	-3.92		0.39
170	0.54	0.49	0.38	12.07	12.93	15.17	0.11	-0.90		
171	0.46	0.42	0.62	13.53	14.36	10.41	0.15	3.73		
173	0.48	0.39	0.34	13.11	14.93	15.85	-0.55	-0.78		
174	0.54	0.52	0.48	11.99	12.36	13.13	0.45	0.64		
175	0.48	0.51	0.28	13.17	12.65	17.21	1.09	-1.81		0.20
177	0.45	0.32	0.38	13.73	16.36	15.17	-1.11	0.14		
178	0.52	0.41	0.41	12.35	14.65	14.49	-0.89	-0.19		
179	0.34	0.33	0.52	15.92	16.08	12.45	0.65	3.63		
180	0.29	0.26	0.31	16.84	17.50	16.53	0.31	1.02		
181	0.25	0.25	0	17.63	17.79	22.65	0.67	-3.27		0.25
182	0.23	0.09	0.17	18.05	20.93	19.25	-1.23	-0.35		
183	0.15	0.1	0.1	19.76	20.65	20.61	0.19	-0.34		
184	0.28	0.19	0.24	17.09	18.93	17.89	-0.51	0.11		

Tabelle 31. DIF-Analyse „Diagramme und Tabellen“

Effekte der Korrektur

Aus Tabelle 32 sind die resultierenden Korrekturwerte für die Sprachgruppen ersichtlich. Für die Bestimmung des Testwertes wurde auf ganze Punkte gerundet.

		N	Minimum	Maximum	M	s
französisch	Med.-naturwiss. Grundv.	69	.00	1.10	.71	.30
	Quant. und form. Probleme	69	.00	.30	.17	.15
	Textverständnis	69	.24	.85	.65	.22
	Fakten lernen	69	.00	.63	.42	.19
	Diagramme und Tabellen	69	.00	.36	.14	.13
	GESAMT	69	.67	3.24	2.09	.56
italienisch	Med.-naturwiss. Grundv.	29	.00	.57	.29	.22
	Quant. und form. Probleme	29	.00	.83	.51	.29
	Textverständnis	29	.00	.86	.69	.23
	Fakten lernen	29	.22	.72	.54	.18
	Diagramme und Tabellen	29	.25	1.12	.79	.25
	GESAMT	29	1.30	3.71	2.82	.67

Tabelle 32. Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen

Die Korrekturwerte wurden auf ganze Punkte gerundet. Somit wurden allen französischsprachigen Kandidatinnen und Kandidaten zwischen 1 und 3 Punkten gutgeschrieben – den italienischsprachigen zwischen 1 und 4.

Evaluation: Gleichbehandlung der Sprachgruppen

Wie 1999 zeigen sich signifikante Unterschiede bei der Zulassung für Humanmedizin für die Sprachgruppen. Im Fach Veterinärmedizin lassen sich keine Unterschiede nachweisen – dabei ist allerdings die geringe Stichprobengröße zu beachten.

		Warteliste	zugelassen	Total
Humanmedizin	deutsch	110 17.7%	512 82.3%	622 100.0%
	französisch	6 24.0%	19 76.0%	25 100.0%
	italienisch	9 39.1%	14 60.9%	23 100.0%
	total	125 18.7%	545 81.3%	670 100.0%
Veterinärmedizin	deutsch	14 11.9%	104 88.1%	118 100.0%
	französisch	6 13.3%	39 86.7%	45 100.0%
	italienisch		6 100.0%	6 100.0%
	total	20 11.8%	149 88.2%	169 100.0%

Tabelle 33. Zulassung nach Sprache des Tests

Disziplin		Value	df	Asymp. Sig. (2seitig)
Humanmedizin	Pearson Chi-Square	7.210	2	.027
	Likelihood Ratio	6.047	2	.049
	Linear-by-Linear Association	6.950	1	.008
Veterinärmedizin	Pearson Chi-Square	.902	2	.637
	Likelihood Ratio	1.605	2	.448
	Linear-by-Linear Association	.106	1	.744

Tabelle 34. Prüfung auf Verteilungsunterschiede für Sprachgruppen getrennt nach Disziplinen: signifikanter Unterschied für Humanmedizin, kein signifikanter Unterschied für Veterinärmedizin (aber geringe Stichprobengröße)

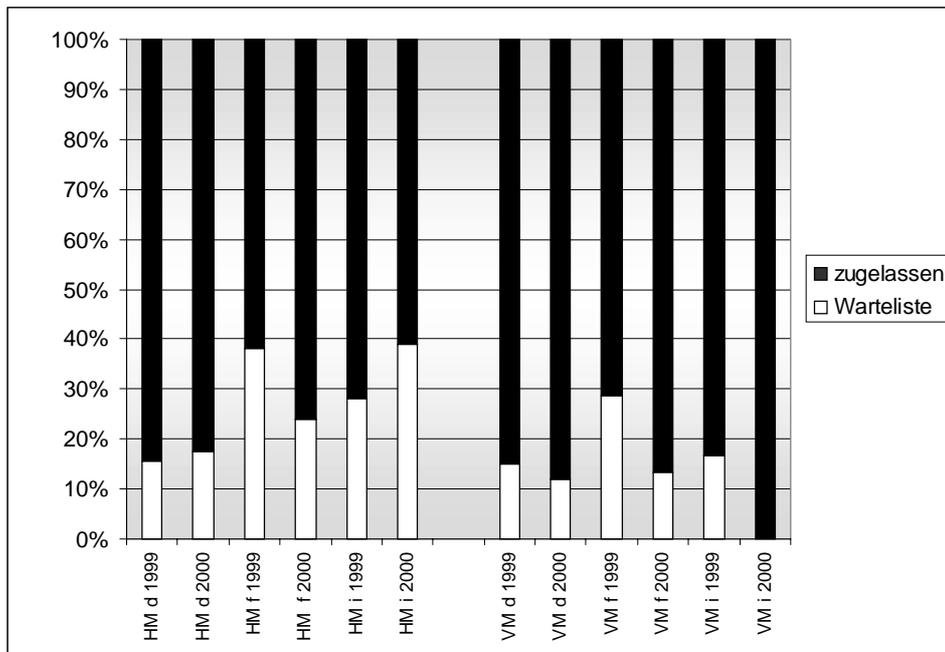


Abbildung 39. Zulassungsquoten nach Disziplin HM und VM und Sprache d f i 1999 und 2000 jeweils dunkler: zugelassen, heller: Warteliste

Dass diese Unterschiede nicht am Test liegen, kann recht gut belegt werden. Eine bereits im Vorjahr diskutierte Frage war die nach dem Verhältnis der Unterschiede in sprachabhängigen und wenig sprachabhängigen Tests. In den sprachunabhängigen Tests wären keine Unterschiede zwischen den Sprachgruppen vorhanden, wenn die Übersetzung des Tests für diesen Effekt verantwortlich wäre. Da diese Unterschiede aber mindestens ebenso deutlich sind wie die in den sprachabhängigen Tests, sind diese Unterschiede auf „tatsächliche“ Unterschiede zwischen den Personen zurückzuführen. Wir hatten die unterschiedliche Repräsentativität der Bewerbergruppen und kulturelle Unterschiede als Ursachen genannt.

Man kann den Effekt der Korrektur evaluieren und kommt zu dem Ergebnis, dass die Korrektur mögliche testbedingte Unterschiede eher etwas überkorrigiert hat: Aufgrund des Faktormodells, des Generalfaktors mit sehr hohem Varianzanteil, ist zu erwarten, dass die Fähigkeiten in den Bereichen sprachabhängiger und sprachunabhängiger Tests hoch korreliert sind 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 deutschen bzw. der jeweiligen Sprachgruppe. Die resultierenden Werte „Delta z“ drücken die Abweichung der jeweiligen Sprachgruppe von der deutschen Sprachgruppe in Standardabweichungen aus. 0.5 hiesse, dass die jeweilige Sprachgruppe um $\frac{1}{2}$ Standardabweichung schlechter ist als die deutschsprachige Referenzgruppe. In den beiden nachfolgenden Abbildungen sind die sprachunabhängigen Tests wiederum zuerst dargestellt. Beim sprachunabhängigen Konzentrationstest treten in der französischen Sprachgruppe die grössten Unterschiede auf. In der italienischen Sprachgruppe ist auffällig, dass „Quantitative und formale Probleme“ erstmals höhere Abweichungen in Richtung geringere Punktzahl hat – bisher erreichte diese Gruppe immer die besten Werte. Dieser Untertest ist wegen der in allen Sprachgruppen

identischen mathematischen Formeln sogar noch recht wenig sprachabhängig. Offenbar wirkt sich hier die geringe Stichprobengröße störend aus.

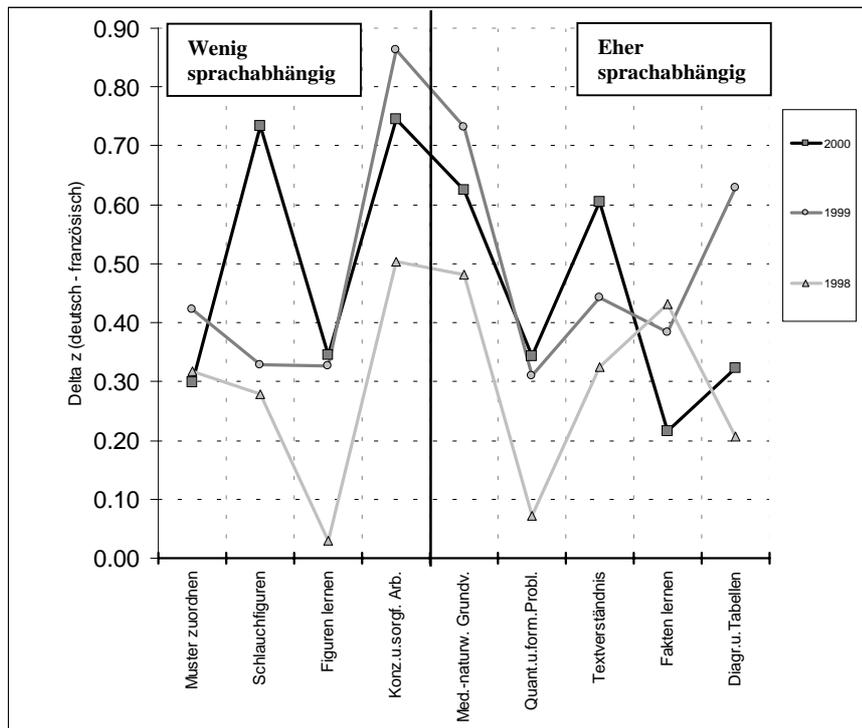


Abbildung 40. Abweichung (Delta z) der deutschen und französischen Sprachgruppe für die Untertests (unkorrigiert)

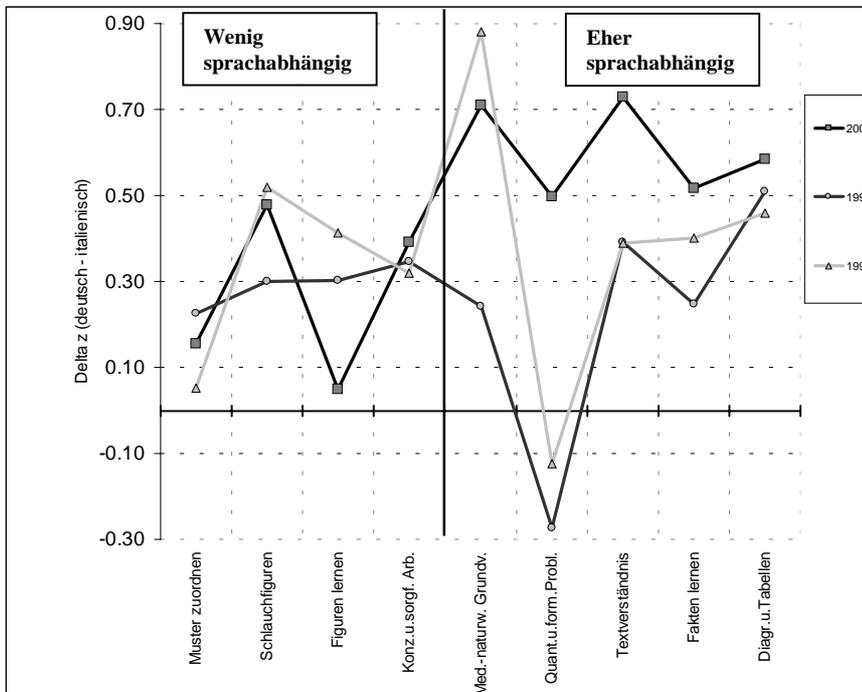


Abbildung 41. Abweichung (Delta z) der deutschen und italienischen Sprachgruppe für die Untertests (unkorrigiert)

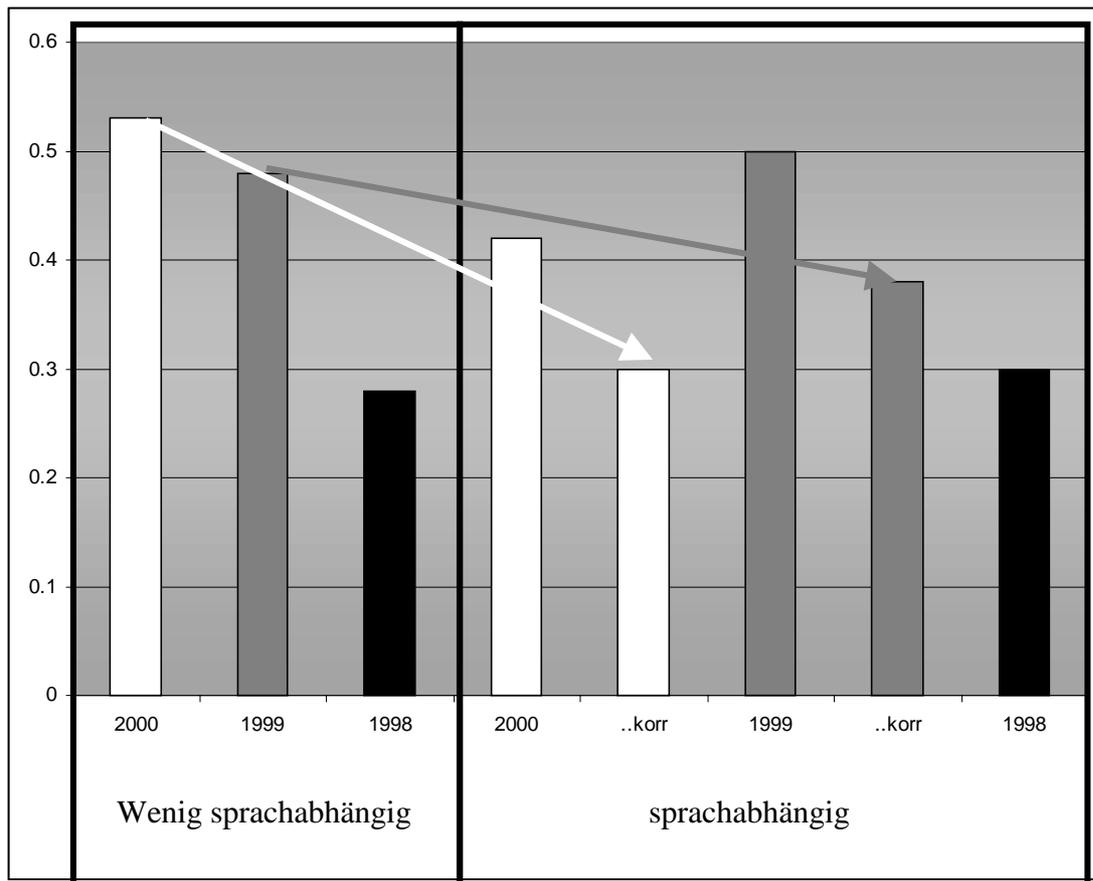


Abbildung 42. Abweichung (Delta z) für den Testwert zwischen deutscher und französischer Sprachgruppe. korr: korrigierter Wert des vorstehenden Jahres

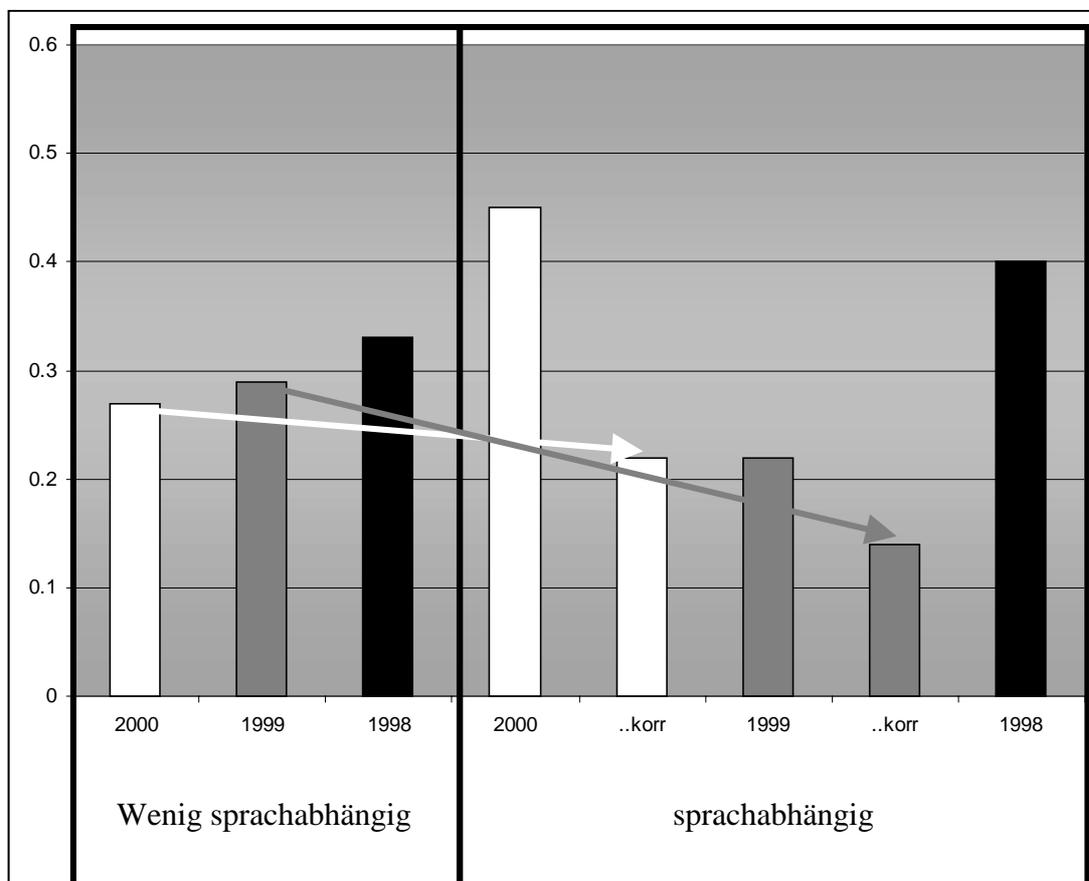


Abbildung 43. Abweichung (Delta z) für den Testwert zwischen deutscher und italienischer Sprachgruppe. korr: korrigierter Wert des vorstehenden Jahres

In Abbildung 42 und 43 werden die mittleren Abweichungen für sprachunabhängige und sprachabhängige Tests der letzten 3 Jahre gegenübergestellt – für 1999 und 2000 unkorrigiert und korrigiert. Für die französische und die italienische Sprachgruppe sowie beide Jahre mit Korrektur gilt, dass nach der Korrektur der mittlere Unterschied in den sprachabhängigen Tests zur deutschen Sprachgruppe geringer ist als derjenige in den sprachunabhängigen Tests. Dies belegt, dass noch vorhandene Unterschiede mit grosser Wahrscheinlichkeit nicht testbedingt sind.

Leistungen in identischen Testaufgaben 1998 – 2000

Betrachtet man die bisherigen Vergleiche über die Jahre, fehlt wegen der in jedem Jahr unterschiedlichen Aufgaben ein Vergleichsmassstab. Deshalb verwenden wir 1999 und 2000 als Einstreuaufgaben solche, die sich im jeweiligen Vorjahr bewährt haben. Die Werte in diesen Einstreuaufgaben lassen sich zwischen 2 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. Für die deutsche Sprachgruppe zeigt sich ein signifikanter Punktzuwachs zwischen 1999 und 2000, der im Trend 1998 zu 1999 schon vorhanden war. Wegen der geringen Stichprobengrössen sind die Unterschiede in der französisch- und italienischsprachigen Gruppe nicht signifikant, es kam allerdings zu einer „Trendumkehr“ - die französische Sprachgruppe erhält jetzt mehr, die italienische weniger Punkte.

	Jahr	N	Mittelwert	Standardabw.	Sig.
deutsch	1999	748	14.71	3.68	.00
	2000	703	15.66	4.26	
französisch	1999	76	12.84	3.69	.25
	2000	69	13.57	3.92	
italienisch	1999	31	13.65	3.57	.93
	2000	29	11.83	4.63	

Tabelle 35. Mittelwerte und Standardabweichungen der gleichen Einstreuaufgaben 1999 und 2000

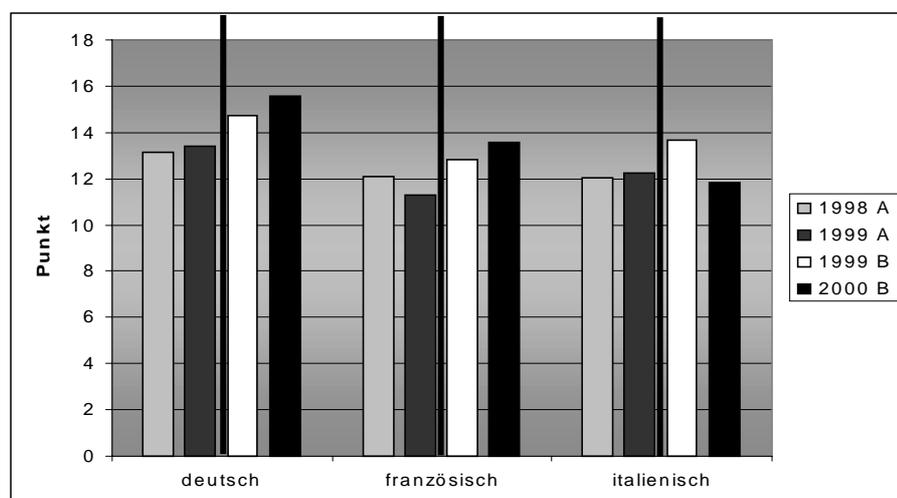


Abbildung 44. Vergleiche der identischen Einstreuaufgaben Set A (1998 zu 1999) und Set B (1999 zu 2000). Zulässig sind nur Vergleiche innerhalb eines Sets

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.¹ 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 Wahl des Testlokals erfolgt allerdings in der Regel nach der Wohnortnähe. Deshalb werden sich mögliche regionale Unterschiede auch in diesen Vergleichen auswirken. In Orten mit mehreren Testlokalen (Basel, Bern Zürich) würde ein Unterschied dann eher auf Durchführungsbedingungen zurückzuführen sein, wenn sich die Testlokale innerhalb eines Testortes deutlich unterscheiden.

Tabelle 36 stellt die Testwerte aller Testlokale mit Unterschieden gegenüber (10%-Niveau). Die Varianzanalyse (Tabelle 37) zeigt, dass in diesem Jahr keine Unterschiede beim Testwert auf 5%-Niveau vorhanden sind. Im Verhältnis zu den Vorjahren sind die Unterschiede geringer (damals waren die Testwerte signifikant auf 5%-Niveau). Daraus ergeben sich keine Hinweise auf unterschiedliche Bedingungen zwischen den Testlokalen.

		N	Mittelwert	Standardabw.
Testwert	Basel 1	41	98.17	9.18
	Basel 2	65	99.03	9.29
	Bern 1	73	102.04	9.34
	Bern 2	103	98.77	10.21
	Zürich 1	112	101.77	11.13
	Zürich 2	34	102.03	8.70
	Zürich 3	46	100.37	8.64
	Zürich 4	45	100.02	8.56
	Zürich 5	98	100.92	10.15
	Chur	29	103.38	9.63
	St. Gallen	57	102.95	10.60

Fortsetzung nächste Seite

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

		N	Mittelwert	Standardabw.
Muster zuordnen	Basel 1	41	9.07	2.11
	Basel 2	65	10.58	3.09
	Bern 1	73	10.12	2.90
	Bern 2	103	10.17	2.77
	Zürich 1	112	10.38	2.84
	Zürich 2	34	9.94	2.59
	Zürich 3	46	10.48	2.88
	Zürich 4	45	9.76	2.66
	Zürich 5	98	10.12	2.83
	Chur	29	11.28	2.43
	St. Gallen	57	11.11	3.06
Schlauchfiguren	Basel 1	41	13.07	3.70
	Basel 2	65	13.22	3.20
	Bern 1	73	13.52	3.71
	Bern 2	103	12.20	3.26
	Zürich 1	112	13.93	3.58
	Zürich 2	34	13.74	3.55
	Zürich 3	46	13.39	3.59
	Zürich 4	45	12.62	3.38
	Zürich 5	98	12.88	3.48
	Chur	29	13.62	3.52
	St. Gallen	57	13.65	3.81
Textverständnis	Basel 1	41	8.54	3.67
	Basel 2	65	7.91	3.25
	Bern 1	73	10.36	3.66
	Bern 2	103	8.52	3.88
	Zürich 1	112	9.19	4.13
	Zürich 2	34	9.09	4.00
	Zürich 3	46	9.07	3.54
	Zürich 4	45	9.44	3.43
	Zürich 5	98	9.15	3.60
	Chur	29	9.21	3.89
	St. Gallen	57	8.98	3.64
Konzentriertes und sorgfältiges Arbeiten	Basel 1	41	11.63	2.95
	Basel 2	65	13.35	3.17
	Bern 1	73	12.51	3.51
	Bern 2	103	12.11	3.61
	Zürich 1	112	12.91	3.76
	Zürich 2	34	13.56	2.99
	Zürich 3	46	13.46	3.47
	Zürich 4	45	13.07	3.37
	Zürich 5	98	13.24	3.62
	Chur	29	12.79	3.14
	St. Gallen	57	14.58	3.53

Tabelle 36. Statistiken für den Testwert nach Testlokalen (nur deutsche Sprachgruppe und Untertests mit signifikanten Unterschieden – vgl. Tabelle 37)

		Quadrat- summe	df	Mittl. quadr. Abweich- ung	F	Sig.
Testwert gerundet	<i>Zwischen Gruppen</i>	1672.256	10	167.226	1.714	.074
	<i>Innerhalb Gruppen</i>	67524.674	692	97.579		
	<i>Total</i>	69196.930	702			
Muster zuordnen	<i>Zwischen Gruppen</i>	157.725	10	15.772	2.005	<u>.030</u>
	<i>Innerhalb Gruppen</i>	5442.924	692	7.865		
	<i>Total</i>	5600.649	702			
Med.-naturwiss. Grundverständnis	<i>Zwischen Gruppen</i>	87.258	10	8.726	.864	.567
	<i>Innerhalb Gruppen</i>	6985.606	692	10.095		
	<i>Total</i>	7072.865	702			
Schlauchfiguren	<i>Zwischen Gruppen</i>	223.055	10	22.305	1.810	.055
	<i>Innerhalb Gruppen</i>	8526.624	692	12.322		
	<i>Total</i>	8749.679	702			
Quantitative und formale Probleme	<i>Zwischen Gruppen</i>	171.344	10	17.134	1.161	.314
	<i>Innerhalb Gruppen</i>	10215.791	692	14.763		
	<i>Total</i>	10387.135	702			
Textverständnis	<i>Zwischen Gruppen</i>	259.734	10	25.973	1.865	<u>.047</u>
	<i>Innerhalb Gruppen</i>	9636.229	692	13.925		
	<i>Total</i>	9895.963	702			
Figuren lernen	<i>Zwischen Gruppen</i>	193.674	10	19.367	1.738	.069
	<i>Innerhalb Gruppen</i>	7711.046	692	11.143		
	<i>Total</i>	7904.720	702			
Fakten lernen	<i>Zwischen Gruppen</i>	175.146	10	17.515	1.458	.151
	<i>Innerhalb Gruppen</i>	8312.902	692	12.013		
	<i>Total</i>	8488.048	702			
Diagramme und Tabellen	<i>Zwischen Gruppen</i>	147.877	10	14.788	1.437	.159
	<i>Innerhalb Gruppen</i>	7119.790	692	10.289		
	<i>Total</i>	7267.667	702			
Konzentriertes und sorgfältiges Arbeiten	<i>Zwischen Gruppen</i>	354.507	10	35.451	2.936	<u>.001</u>
	<i>Innerhalb Gruppen</i>	8356.924	692	12.076		
	<i>Total</i>	8711.431	702			

Tabelle 37. Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den Testlokalen

Vergleich der Geschlechter

Im Vorfeld des Tests wurde die Forderung nach Gleichbehandlung der Geschlechter mehrfach thematisiert. Dies betraf vor allem Fragen der Testfairness: Es darf keine Gruppe beim Test benachteiligt sein. Chancengleichheit bedeutet allerdings nicht die Nivellierung von tatsächlich vorhandenen Unterschieden. Woraus könnten Unterschiede begründet sein? Zunächst sind die unterschiedlichen Anteile der Geschlechter an der Bewerberzahl zu beachten, indem die Testteilnehmenden unterschiedlich repräsentativ für die jeweilige Kohorte sein können (analog zum Problem der Repräsentativität der Sprachgruppen). Eine andere Ursache könnte sein, dass in der Realität tatsächlich Unterschiede bezüglich des Prüfungserfolges zwischen den Geschlechtern vorhanden sind, welche der Test vorhersagen und daher abbilden muss. Die Klärung der Ursachen für diese Unterschiede ist dann eine zweite Frage, die allerdings nicht den Test betrifft.

Prüfungserfolg im Medizinstudium und Geschlecht

Die weiter vorn vorgestellten Evaluationsergebnisse der Testteilnehmer 1998 erlauben erstmals auch eine geschlechtsspezifische Auswertung zum Prüfungserfolg. In der Tabelle 38 werden Bestehen, Nichtbestehen und Nichtantreten zur 1. Vorprüfung geschlechtsspezifisch tabelliert. Die statistischen Prüfungen (Tabelle 39) belegen, dass der Unterschied auf 5%-Niveau signifikant ist.

Es findet sich eine Bestätigungen für die ebenfalls aus Deutschland bekannten Fakten, dass die Bestehenswahrscheinlichkeit der 1. Vorprüfung für Männer höher (hier um 10%) liegt als für Frauen. Die Rate für „Nichtantreten“ zur Prüfung unterscheidet sich prozentual nicht.

	1. Ärztliche Vorprüfung			Total
	Nicht angetreten	Bestanden	Nicht bestanden	
Männlich	23 8.4%	196 71.5%	55 20.1%	274 100.0%
Weiblich	31 9.6%	197 61.0%	95 29.4%	323 100.0%
Total	54 9.0%	393 65.8%	150 25.1%	597 100.0%

Tabelle 38. Prüfungsergebnis 1. Vorprüfung 1999 nach Geschlecht

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.886	2	.019
Likelihood Ratio	7.963	2	.019

Tabelle 39. Signifikanzprüfung für Prüfungsergebnis 1. Vorprüfung nach Geschlecht

Verglichen werden auch die Prüfungsnoten (Tabelle 40). Die Prüfungsnoten der Männer sind signifikant besser. Für diesen Unterschied bei Bestehenswahrscheinlichkeit und Prüfungsnoten können wohlgermerkt die genannten Repräsentativitätsunterschiede (mehr Frauen als Männer) genauso verantwortlich sein wie „tatsächliche“ Unterschiede. Dies lässt sich aufgrund dieser Daten nicht weiter differenzieren.

Durchschnitt	Geschlecht	N	Mittelwert	Standardabw.	F-Test		t-Test		
					F	Sig.	T	df	Sig. (2s)
Examen Typ 1	Männlich	228	4.44	.93	2.179	.141	3.029	484	<u>.003</u>
	Weiblich	258	4.17	.96					

Tabelle 40. Prüfungsergebnis für 1. Vorprüfung (Durchschnitt) nach Geschlecht

Ergebnisse im Jahr 2000

In der Tabelle 41 werden die Mittelwerte und Standardabweichungen für den Testwert nach Sprache und Geschlecht dargestellt. Aufgrund der geringen Stichprobengrößen in der französisch- und italienischsprachigen Gruppe hat dort die Angabe von Mittelwerten und Standardabweichungen nur illustrativen Charakter – auch die Zahl der männlichen Bewerber im Fach Veterinärmedizin ist für die deutsche Sprachgruppe recht gering für statistische Vergleiche.

		Humanmedizin			Veterinärmedizin		
		deutsch	französisch	italienisch	deutsch	französisch	italienisch
männlich	Mittelwert	102.24	100.31	93.83	101.05	94.45	100.00
	Stand.-abw.	9.34	6.49	7.59	7.28	7.37	4.36
	N	230	13	12	21	11	3
weiblich	Mittelwert	100.80	94.27	94.18	96.74	92.12	96.67
	Stand.-abw.	10.18	8.06	11.30	9.91	8.36	2.52
	N	360	11	11	92	34	3

Tabelle 41. Statistiken für den Testwert nach Geschlechtern und Disziplin: Werte für Gruppen unter 30 Personen (kursiv) nicht ausreichend für statistische Vergleiche

Sprache	Disziplin	Levene's Test Gleichheit der Varianzen			t-Test Gleichheit des Mittelwerts		
			F	Sig.	t	df	Sig.
deutsch	Humanmedizin	Varianzen homogen	1.285	.258	1.737	588	.083
	Veterinärmedizin	Varianzen homogen	2.964	.088	1.878	111	.063
französisch	Humanmedizin	Varianzen homogen	.732	.402	2.033	22	.054
	Veterinärmedizin	Varianzen homogen	.122	.728	.827	43	.413
italienisch	Humanmedizin	Varianzen homogen	1.365	.256	-.088	21	.931

Tabelle 42. Prüfung Varianzhomogenität und Mittelwert (t-Test) für Testwert zwischen Geschlechtern. Italienische Sprachgruppe Veterinärmedizin wg. Fallzahl nicht berechnet

Nach den strengen statistischen Gesichtspunkten ist kein Unterschied auf 5%-Niveau signifikant. Trendunterschiede (10%-Niveau) finden sich allerdings wie in den Vorjahren auch in diesem Jahr. Stellt man die nachgewiesenen signifikanten Unterschiede bei Prüfungserfolg und -noten in Rechnung, entspräche der Unterschied im Trend beim Testwert dieser Realität. Er bleibt aber bezüglich seiner Nachweisbarkeit hinter dem Effekt für die Prüfung zurück.

Eine zweifache Varianzanalyse nach Disziplin und Geschlecht zeigt, dass hier die Unterschiede für den Faktor Disziplin (Human- vs. Veterinärmedizin) bezüglich des Testwertes signifikant sind – nicht aber für das Geschlecht oder eine Wechselwirkung.

	Quadratsumme	df	Mittlere quadrat. Abweichung	F	Sig.
Disziplin	8010697.783	2	4005348.892	11979.954	.002
Geschlecht	485.825	1	485.825	4.865	.727
Disziplin x Geschlecht	96.726	1	96.726	1.018	.313

Tabelle 43. Zweifache Varianzanalyse für den Einfluss von Geschlecht und Disziplin auf den Testwert

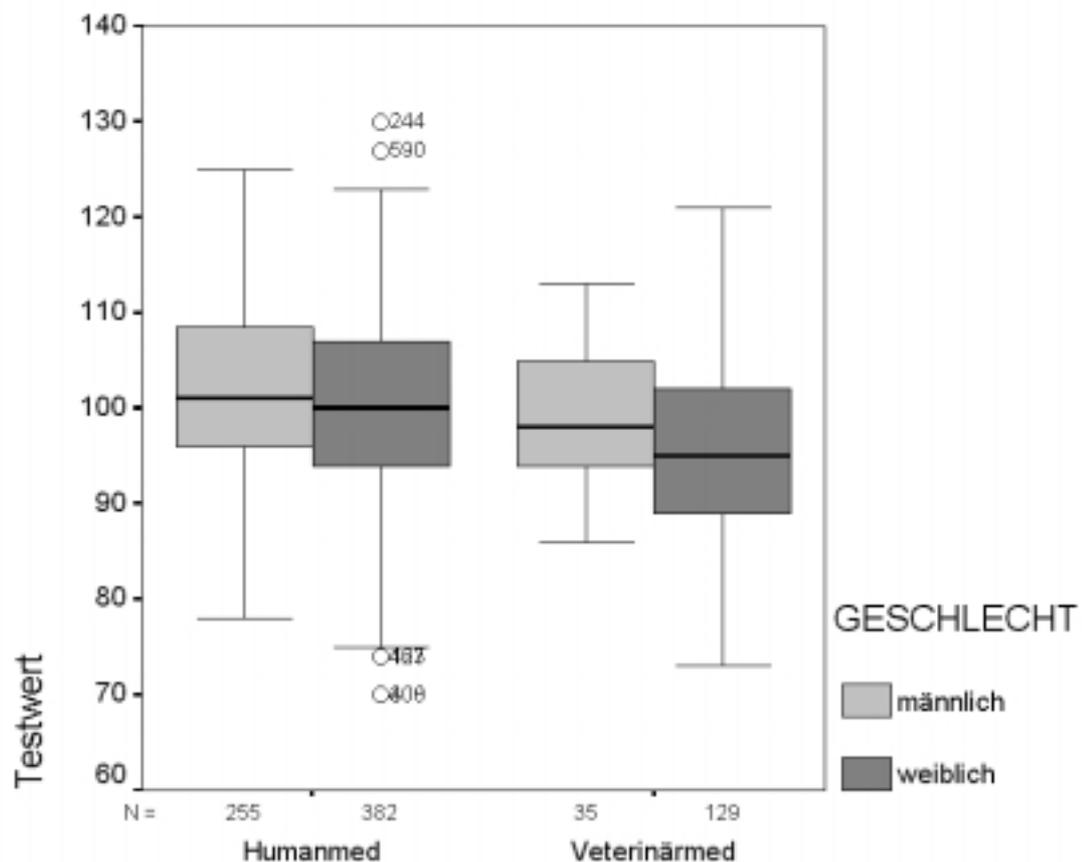


Abbildung 45. Boxplot für den Testwert nach Geschlecht und Disziplinen

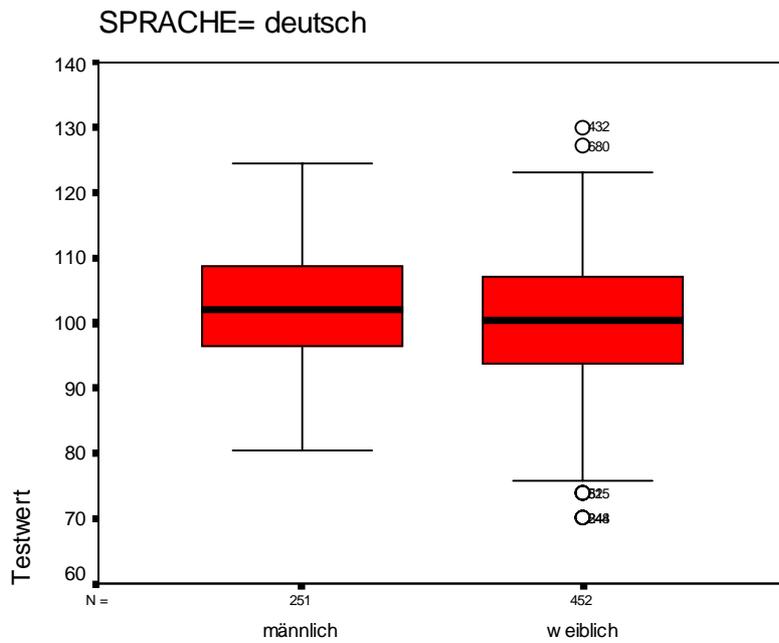
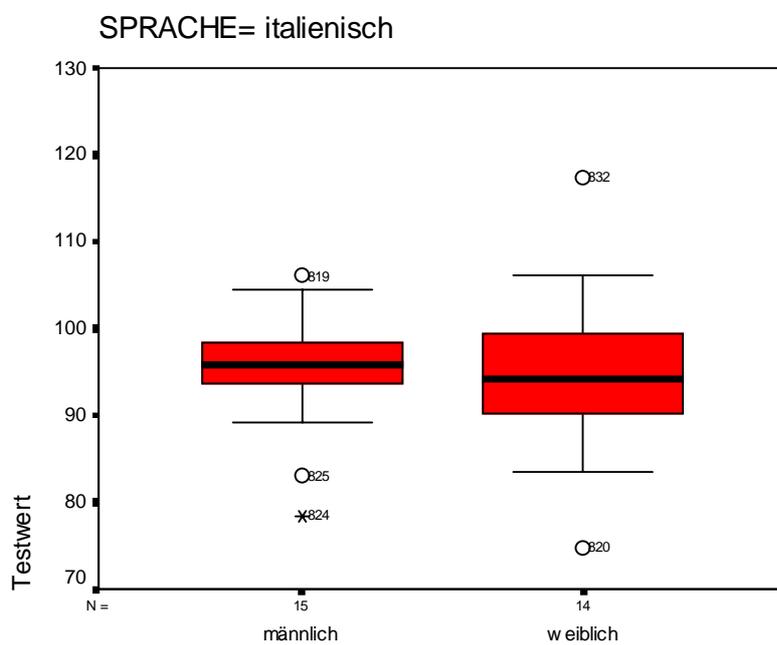
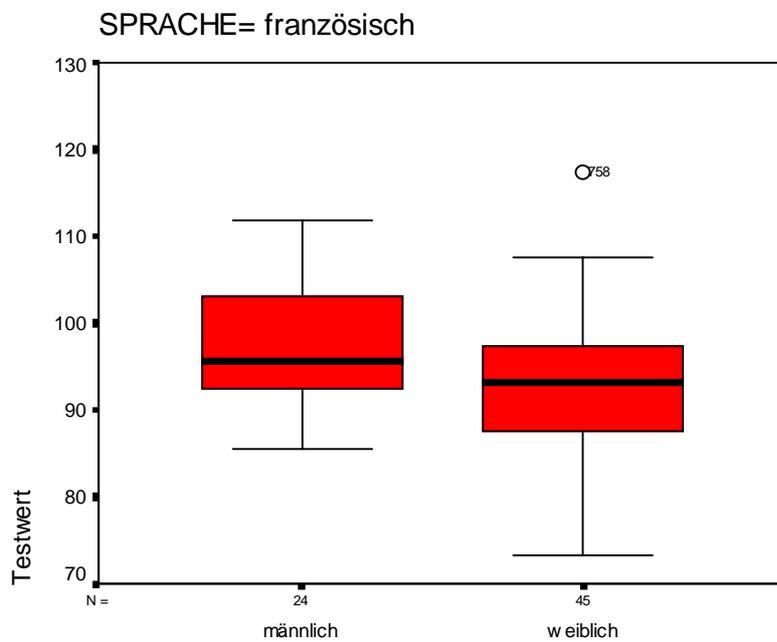


Abbildung 46. Boxplots für die Testwerte nach Geschlechtern getrennt für die drei Sprachgruppen (Werte siehe Tab. 41)



Wiederum zeigen sich signifikante Unterschiede für die einzelnen Untertests zwischen den Geschlechtern. Bis auf drei Untertests sind alle Unterschiede signifikant. In vier Untertests zeigen die Frauen bessere Leistungen, in fünf Untertests die Männer – dies deckt sich mit den Befunden aus den Vorjahren. Ein wichtiger Vorzug des Tests ist es, dass beide Geschlechter durch die Kombination der Untertests vergleichbare Leistungen erzielen. Dies dürfte der Erfahrung im Leben entsprechen, dass gleich gute Leistungen auf verschiedenen Wegen erzielt werden können.

Untertest	Geschlecht	M	S	Levene's Test Varianzgleichh.			t-test Gleichheit des Mittelwerts		
					F	Sig.	t	df	Sig. (2- seitig)
Muster zuordnen	männlich	9.97	2.75	<i>Varianz- homogenität</i>	.350	.554	-1.622	799	.105
	weiblich	10.30	2.83						
Med.-naturwiss. Grundverständnis	männlich	10.66	3.04	<i>Varianz- homogenität</i>	1.088	.297	1.393	799	.164
	weiblich	10.33	3.26						
Schlauchfiguren	männlich	13.88	3.35	<i>Varianz- homogenität</i>	2.076	.150	5.817	799	.000
	weiblich	12.38	3.61						
Quantitative und formale Probleme	männlich	13.02	3.37	<i>Keine Varianz- homogenität</i>	6.139	.013	8.441	676.4	.000
	weiblich	10.80	3.92						
Textverständnis	männlich	9.30	3.74	<i>Varianz- homogenität</i>	.059	.807	3.196	799	.001
	weiblich	8.43	3.73						
Figuren lernen	männlich	10.03	3.25	<i>Varianz- homogenität</i>	.802	.371	-4.22	799	.673
	weiblich	10.13	3.48						
Fakten lernen	männlich	9.47	3.26	<i>Varianz- homogenität</i>	.612	.434	-3.061	799	.002
	weiblich	10.23	3.48						
Diagramme und Tabellen	männlich	11.30	2.93	<i>Varianz- homogenität</i>	1.515	.219	7.530	799	.000
	weiblich	9.58	3.22						
Konzentriertes und sorgfältiges Arbeiten	männlich	11.92	3.43	<i>Varianz- homogenität</i>	1.354	.245	-4.674	799	.000
	weiblich	13.13	3.57						

Tabelle 44. Mittelwerte und Standardabweichungen geschlechtsspezifisch und Ergebnisse der Prüfungen auf Varianzhomogenität und Mittelwertunterschiede

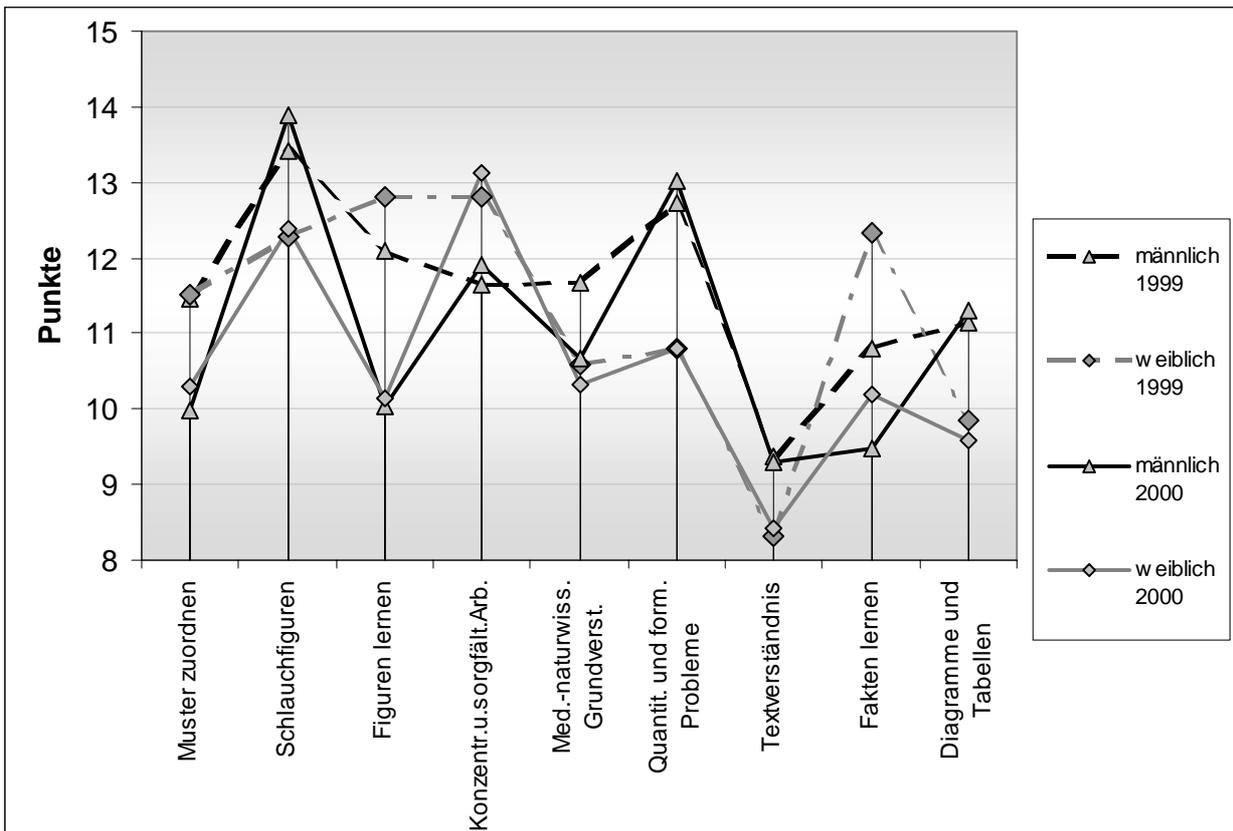


Abbildung 47. Mittelprofile für Punktwerte der Untertests geschlechtsspezifisch

Die Zulassungsquoten in Humanmedizin für beide Geschlechter (Tabelle 45) unterscheiden sich ebenfalls nicht signifikant (Tabelle 46). In Veterinärmedizin zeigt sich ein Unterschied. Hier werden alle männlichen Bewerber zugelassen. Zu beachten ist, dass sich in diesem Jahr weniger Männer beworben haben. Summa summarum kann auch in diesem Jahr davon ausgegangen werden, dass es zu keiner Benachteiligung von Frauen gekommen ist. Die Ergebnisse bleiben seit 3 Jahren stabil.

Disziplin		Warteliste	zugelassen	Total
Humanmedizin	männlich	43 15.9%	228 84.1%	271 100.0%
	weiblich	82 20.6%	317 79.4%	399 100.0%
	Total	125 18.7%	545 81.3%	670 100.0%
Veterinärmedizin	männlich	0	35 100.0%	35 100.0%
	weiblich	20 14.9%	114 85.1%	134 100.0%
	Total	20 11.8%	149 88.2%	169 100.0%

Tabelle 45. Zulassungsquoten geschlechtsspezifisch

Disziplin		Value	df	Asymp. Sig. (2-seitig)
Humanmedizin	Pearson Chi-Square	2.199	1	.138
	Continuity Correction	1.906	1	.167
	Likelihood Ratio	2.232	1	.135
Veterinärmedizin	Pearson Chi-Square	6.180	1	<u>.013</u>
	Continuity Correction	4.817	1	.028
	Likelihood Ratio	10.332	1	.001

Tabelle 46. Prüfung Unterschiede der Zulassung – geschlechtsspezifisch: keine signifikanten Unterschiede bei Humanmedizin, signifikant bei Veterinärmedizin

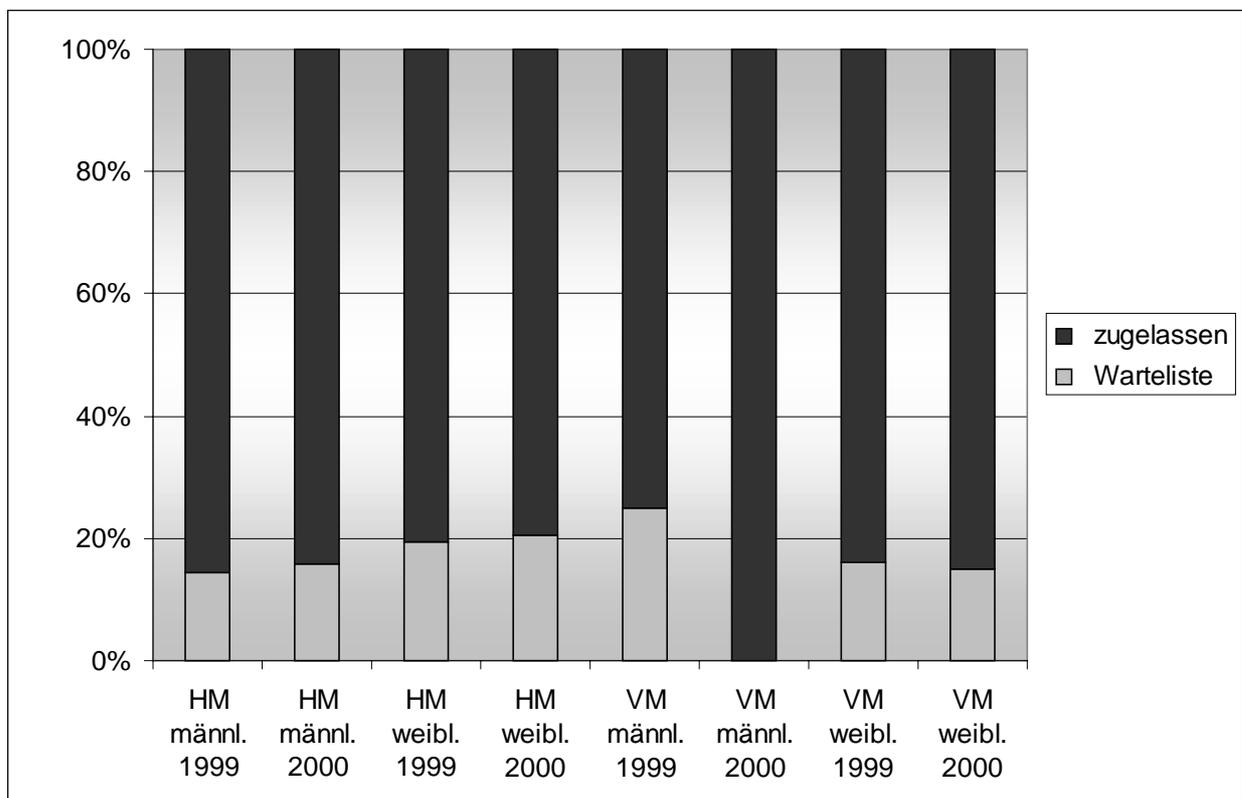


Abbildung 48. Zulassung nach Disziplinen Humanmedizin (HM) und Veterinärmedizin (VM) für die Geschlechter alle Testteilnehmer auf je 100% bezogen

Vergleich der Kantone

Bereits in den letzten beiden 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.

Vergleich 2000 zwischen Kantonen innerhalb der Deutschschweiz

Eine der wichtigsten Fragen war die, ob sich die Leistungen für die Kantone nach unterschiedlichen Maturitätsquoten unterscheiden. Basierend auf den Quoten für 1998¹ wurden die Kantone in drei Gruppen geteilt, die hohe, mittlere und niedrige Maturitätsquoten repräsentieren.

hoch			mittel			niedrig		
Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl
BS	21.2	28	SH	18.4	8	SG	12.8	55
BL	20.4	40	NW	16.5	2	SZ	12.7	18
ZH	18.8	194	AG	15.9	71	AI	12.3	4
			GL	15.3	5	GR	12.3	19
			ZG	15	9	LU	11.5	41
			SO	13.9	20	TG	10.4	23
			AR	13.8	4	UR	10.2	2
			BE	13.1	114	OW	9.9	9
Summe		262	Summe		233	Summe		171

Tabelle 47. Gruppierung der Kantone nach der Maturitätsquote (Basis 1998)

Im Jahre 1998 und 1999 wurden keine Testwertunterschiede für die Kantonsgruppen gefunden, wenn man sie nach der Maturitätsquote gruppiert. Diese Aussage kann auch im Jahr 2000 bestätigt werden. Es wurden die Kantone ähnlich wie 1998 und 1999 gruppiert, wobei allerdings die Maturitätsquoten insgesamt angestiegen sind. Einige Verschiebungen gegenüber 1998 und 1999 sind vorhanden.

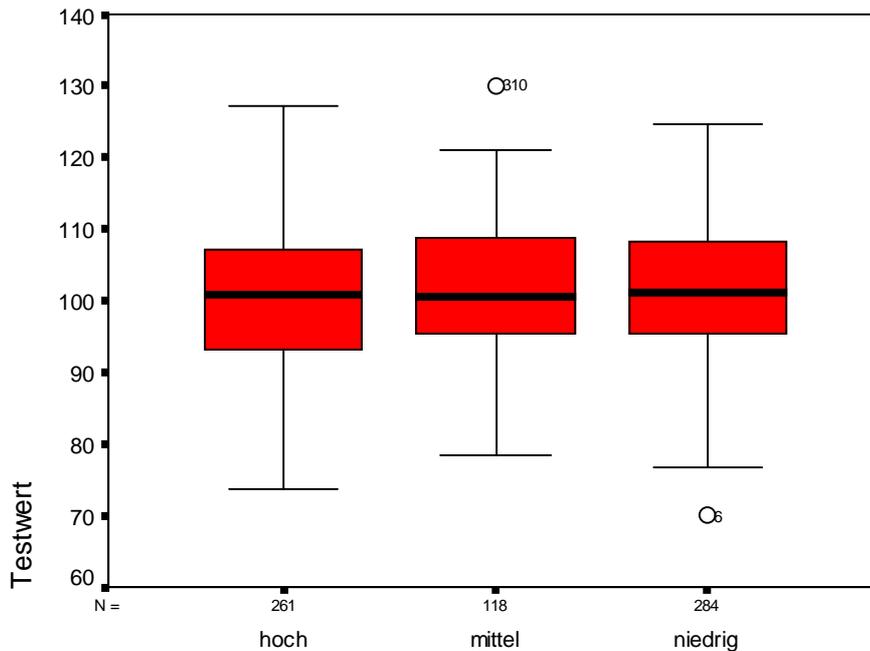
¹ Zahlen aus: Bundesamt für Statistik (1999). Maturitäten 1998: Statistik der Schweiz Bd. 15., Tabelle 4 Seite 13

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.

	Maturitäts- quote	Mittel- wert	Stand.- abw.		Quadrat- summe	df	MQ	F	Sig.
Testwert	hoch	100.307	10.302	Zwischen Gruppen	182.568	2	91.284	.931	.395
	mittel	101.449	9.536						
	niedrig	101.349	9.670						
Muster zuordnen	hoch	10.088	2.868	Zwischen Gruppen	25.363	2	12.682	1.591	.204
	mittel	10.559	2.681						
	niedrig	10.447	2.839						
Medizin.- naturwiss. Grundver- ständnis	hoch	10.617	3.285	Zwischen Gruppen	14.511	2	7.256	.732	.481
	mittel	11.025	3.104						
	niedrig	10.820	3.037						
Schlauch- figuren	hoch	13.387	3.584	Zwischen Gruppen	27.553	2	13.777	1.111	.330
	mittel	13.483	3.371						
	niedrig	13.011	3.525						
Quantitat. u.formale Probleme	hoch	11.586	3.779	Zwischen Gruppen	36.484	2	18.242	1.262	.284
	mittel	12.254	3.652						
	niedrig	11.831	3.883						
Text- verständnis	hoch	8.992	3.810	Zwischen Gruppen	2.090	2	1.045	.073	.929
	mittel	9.051	3.948						
	niedrig	9.116	3.659						
Figuren lernen	hoch	9.862	3.397	Zwischen Gruppen	77.780	2	38.890	3.550	.092
	mittel	10.627	3.143						
	niedrig	10.532	3.295						
Fakten lernen	hoch	9.851	3.534	Zwischen Gruppen	68.926	2	34.463	2.868	.075
	mittel	9.822	3.332						
	niedrig	10.493	3.457						
Diagramme und Tabellen	hoch	10.330	3.225	Zwischen Gruppen	3.402	2	1.701	.163	.850
	mittel	10.322	3.255						
	niedrig	10.472	3.233						
Konzentr.u. sorgfält.Arb.	hoch	13.042	3.664	Zwischen Gruppen	1.612	2	.806	.066	.936
	mittel	12.949	3.170						
	niedrig	13.088	3.476						

Tabelle 48. Mittelwerte und Standardabweichungen für drei Gruppen der Kantone nach der Maturitätsquote, keine signifikanten Unterschiede für Testwert und die Untertests

Die Abbildung 49 verdeutlicht die Unterschiede zwischen den drei Gruppen als Boxplots.



Kantone nach Maturitätsquote

Abbildung 49. Boxplot für die drei Gruppen der Kantone nach der Maturitätsquote

Für Vergleiche zwischen den Kantonen wurden die gleichen Kantone wie 1998 und 1999 berücksichtigt. Die Unterschiede haben sich gegenüber 1998 und 1999 vor allem an den Enden etwas verändert: Basel-Stadt unterscheidet sich nicht mehr so deutlich wie in den Vorjahren. Es ist möglicherweise auch darauf mit zurückzuführen, dass in Basel-Stadt für interessierte Personen eine organisierte Testvorbereitung erfolgte.

Kanton		Testwert	Muster zuordnen	Medizin.-naturw. Grundv.	Schlauchfiguren	Quantit. u. form. Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arb.
AG	M	102.4	10.1	11.5	13.4	12.9	10.0	10.5	9.8	10.8	12.8
	S	8.7	2.5	3.3	3.1	3.4	3.9	2.8	2.9	3.1	2.9
BE	M	100.6	10.2	11.0	12.7	11.9	9.5	10.1	10.2	10.7	12.2
	S	9.8	2.7	3.3	3.3	3.9	3.9	3.3	3.4	3.2	3.3
BL	M	99.0	9.7	10.1	13.4	10.9	8.9	10.4	9.3	9.9	12.5
	S	9.9	3.0	3.1	3.6	4.3	3.4	3.4	3.5	3.5	3.4
BS	M	97.9	10.4	9.7	13.2	9.4	7.1	10.4	10.1	9.5	13.4
	S	8.6	3.0	3.1	3.7	3.7	3.2	3.5	3.1	2.4	3.3
LU	M	99.6	9.9	10.5	11.9	11.7	8.7	10.6	10.3	9.9	13.0
	S	8.8	2.6	3.0	3.4	4.2	3.5	2.6	3.3	3.3	3.3
SG	M	103.7	10.9	11.4	14.2	11.8	9.5	10.9	10.8	10.4	14.3
	S	10.1	3.1	2.8	3.4	3.8	3.6	3.8	3.5	3.1	3.4
ZH	M	100.9	10.1	10.8	13.4	12.1	9.3	9.7	9.9	10.5	13.1
	S	10.6	2.8	3.3	3.6	3.5	3.9	3.4	3.6	3.3	3.8

Tabelle 49. Vergleich der Kantone mit mehr als 38 Teilnehmer für Testwert und die einzelnen Untertests

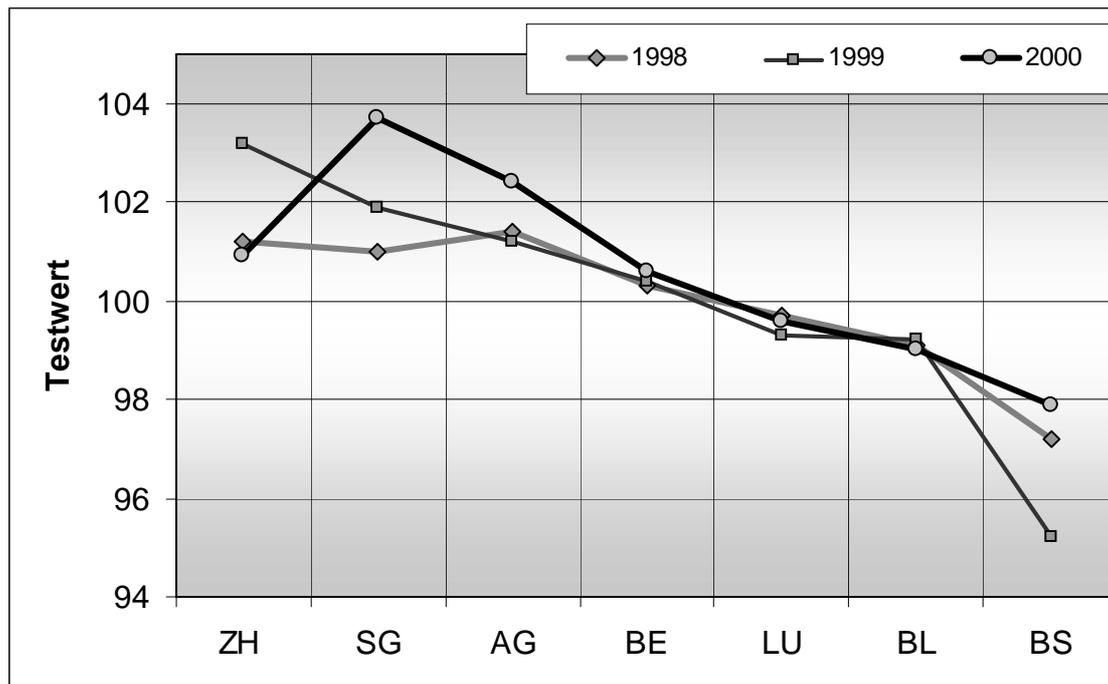


Abbildung 50. Testwert für Kantone im Vergleich

Die nachfolgende Tabelle zeigt für den Testwert und vier Untertests die Ergebnisse der Homogenitätsprüfung der Mittelwerte. Es werden je zwei homogene Mittelwertsgruppen identifiziert. Markiert sind die nur in einer Gruppe befindlichen Mittelwerte und der Überlappungsbereich.

Beim Testwert lassen sich 2 Mittelwertsgruppen unterscheiden, die sich nach höchstem und niedrigstem mittleren Testwert pro Kanton unterscheiden (nur der höchste und der niedrigste Wert unterscheiden sich praktisch signifikant). Im vergangenen Jahr war der Überlappungsbereich kleiner – der niedrigste Testwert in BS unterschied sich noch vom Mittelwert in drei Kantonen signifikant. Nur noch 3 statt 6 Untertests weisen signifikante multiple Mittelwert-Unterschiede auf. Diese sind in der Tabelle 50 aufgeführt.

War BS im Vorjahr noch in allen 6 Untertests mit signifikanten Mittelwertunterschieden auf der Position mit dem geringsten Testwert, so ist dies jetzt nur noch in 2 von 4 Untertests der Fall. Damit kann auch gezeigt werden, dass bereits eine Vorbereitung auf den Test mit dem empfohlenen Material erfolgreich ist, wenn dies durch die Schulen unterstützt wird. Es muss kein (überteuertes und bezüglich der Wirksamkeit ungeprüftes) kommerzielles Training sein!

Testwert	N	Subset for alpha = .05	
		1	2
BS	28	97.929	
BL	40	99.025	99.025
LU	41	99.634	99.634
BE	114	100.588	100.588
ZH	193	100.917	100.917
AG	71	102.366	102.366
SG	55		103.673
Sig.		.232	.184
Schlauchfiguren	N	Subset for alpha = .05	
		1	2
LU	41	11.878	
BE	114	12.667	12.667
BS	28	13.214	13.214
BL	40	13.350	13.350
AG	71	13.380	13.380
ZH	193	13.420	13.420
SG	55		14.200
Sig.		.238	.244
Quantitat. u. formale Probleme	N	Subset for alpha = .05	
		1	2
BS	28	9.429	
BL	40	10.850	10.850
LU	41		11.659
SG	55		11.764
BE	114		11.868
ZH	193		12.052
AG	71		12.944
Sig.		.448	.062
Textverständnis	N	Subset for alpha = .05	
		1	2
BS	28	7.143	
LU	41	8.707	8.707
BL	40	8.875	8.875
ZH	193	9.285	9.285
SG	55		9.491
BE	114		9.491
AG	71		9.958
Sig.		.054	.614
Konzentr.u.sorgfält.Arbeiten	N	Subset for alpha = .05	
		1	2
BE	114	12.175	
BL	40	12.525	12.525
AG	71	12.831	12.831
LU	41	13.000	13.000
ZH	193	13.098	13.098
BS	28	13.393	13.393
SG	55		14.345
Sig.		.537	.094

Tabelle 50. Multiple Mittelwertvergleiche für Kantone (Untertests mit signifikanten Unterschieden)

Zusammenfassende Analyse für Kantone 1998 bis 2000

Seit 1998 haben 2406 Personen am Test teilgenommen. Die Tabelle 51 listet alle Teilnehmer am EMS 1998 bis 2000 nach Kantonen auf. Zu beachten ist, dass 1998 nur Humanmedizin, 1999 und 2000 Human- und Veterinärmedizin einem NC und damit dem Test unterlag.

	Sprache			Disziplin		Total
	deutsch	französisch	italienisch	Humanmedizin	Veterinärmedizin	
AG	221			190	31	221
AI	6			5	1	6
AR	14			14		14
BE	321	6		275	52	327
BL	147	2		139	10	149
BS	107			103	4	107
FR	40	85		106	19	125
GE	1	24		1	24	25
GL	13			11	2	13
GR	64		3	57	10	67
JU		3		1	2	3
LU	128			106	22	128
NE		13		3	10	13
NW	11			11		11
OW	14			14		14
SG	180			166	14	180
SH	19			16	3	19
SO	74			61	13	74
SZ	57			53	4	57
TG	65			58	7	65
TI			86	74	12	86
UR	6			6		6
VD	1	30		5	26	31
VS	37	18		45	10	55
ZG	24	1		16	9	25
ZH	546	1	2	492	57	549
AUSLAND	35	1		32	4	36
Total	2131	184	91	2060	346	2406

Tabelle 51: Gesamtzahl der Teilnehmer am EMS 1998 bis 2000

Beide folgenden Abbildungen stellen die jeweiligen Anteile der Medizinbewerbungen 1998 bis 2000 an den Maturanden und an der 19jährigen Wohnbevölkerung gegenüber. Die Daten dafür befinden sich im Anhang (Tabelle A 7). Bei Humanmedizin zeigt sich, dass beide Basel einen hohen Anteil bezogen auf beide Grössen aufweisen. In beiden Kantonen liegt nicht nur eine hohe Maturandenquote vor, sondern es entscheiden sich auch relativ viele Maturanden für ein Medizinstudium. In Obwalden stösst das Medizinstudium auch auf grosses Interesse, die Maturandenquote ist aber deutlich geringer.

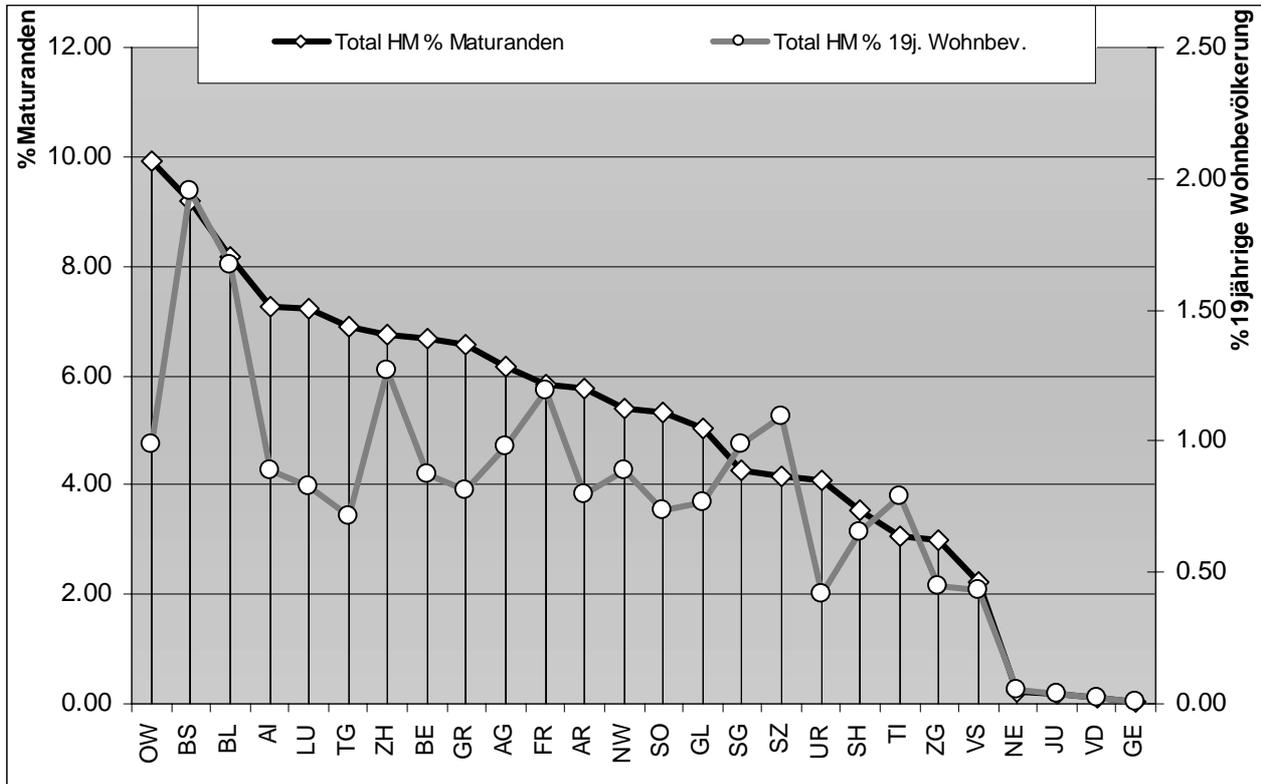


Abbildung 51. Anteile EMS-Teilnehmer für Humanmedizin (mit NC) an Maturanden und an 19jähriger Wohnbevölkerung 1998 bis 2000

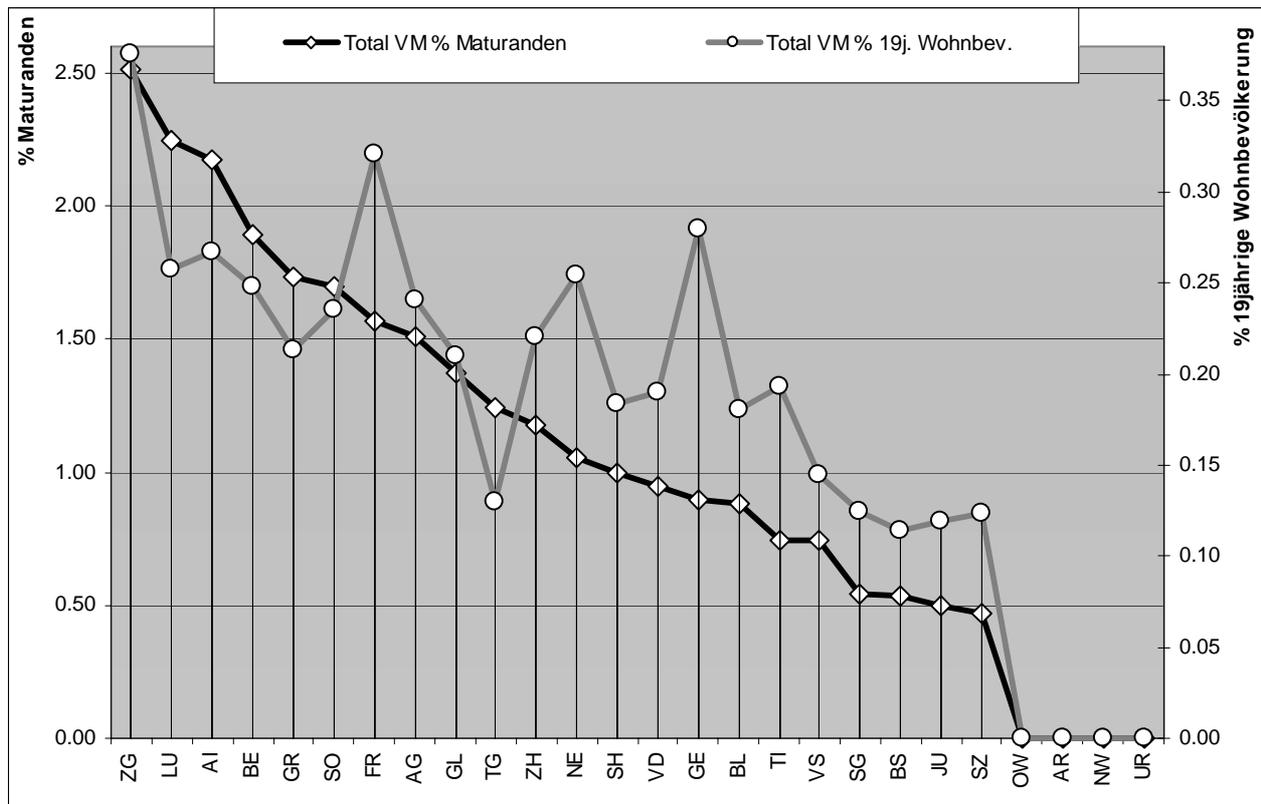


Abbildung 52. Anteile EMS-Teilnehmer für Veterinärmedizin (mit NC) an Maturanden und an 19jähriger Wohnbevölkerung 1999 und 2000

Die Testwerte sind durch die Standardisierung auf den jeweiligen Mittelwert und die Streuung des Geburtsjahres zwischen den Jahren direkt vergleichbar. Deshalb kann für die deutschsprachigen Kantone und die deutschsprachige Testform 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 auch hier Repräsentativitätsfragen auftreten.

	N	Mittelwert	Standardabw.	Minimum	Maximum
BS	107	96.729	10.001	71	122
ZG	24	98.583	9.050	78	118
SH	19	99.000	10.593	82	120
BL	147	99.129	9.601	74	125
LU	128	99.578	8.910	77	118
GR	64	99.953	8.924	77	118
SZ	57	100.421	8.948	80	127
BE	321	100.483	10.098	73	125
OW	14	100.500	8.689	92	126
AR	14	101.214	8.631	89	119
SO	74	101.311	9.349	79	122
TG	65	101.323	9.589	77	121
AG	221	101.652	8.812	74	127
ZH	546	101.731	10.500	70	130
SG	180	102.172	9.923	70	125
NW	11	104.091	11.202	88	121
GL	13	104.923	11.521	89	130
Total	2131	100.504	9.958	70	130

Tabelle 52. Kennwerte für den Testwert von Kantonen mit mehr als 10 Teilnehmern (deutschsprachige Kantone – Test in deutscher Sprache)

	N	Subset for alpha = .05		
		1	2	3
BS	107	96.729		
ZG	24	98.583	98.583	
SH	19	99.000	99.000	
BL	147	99.129	99.129	
LU	128	99.578	99.578	99.578
GR	64	99.953	99.953	99.953
SZ	57	100.421	100.421	100.421
BE	321	100.483	100.483	100.483
OW	14	100.500	100.500	100.500
AR	14	101.214	101.214	101.214
SO	74	101.311	101.311	101.311
TG	65	101.323	101.323	101.323
AG	221	101.652	101.652	101.652
ZH	546	101.731	101.731	101.731
SG	180	102.172	102.172	102.172
NW	11		104.091	104.091
GL	13			104.923
Sig.		.066	.062	.068

Tabelle 53. Multipler Mittelwertvergleich (Werte der Tabelle 52)

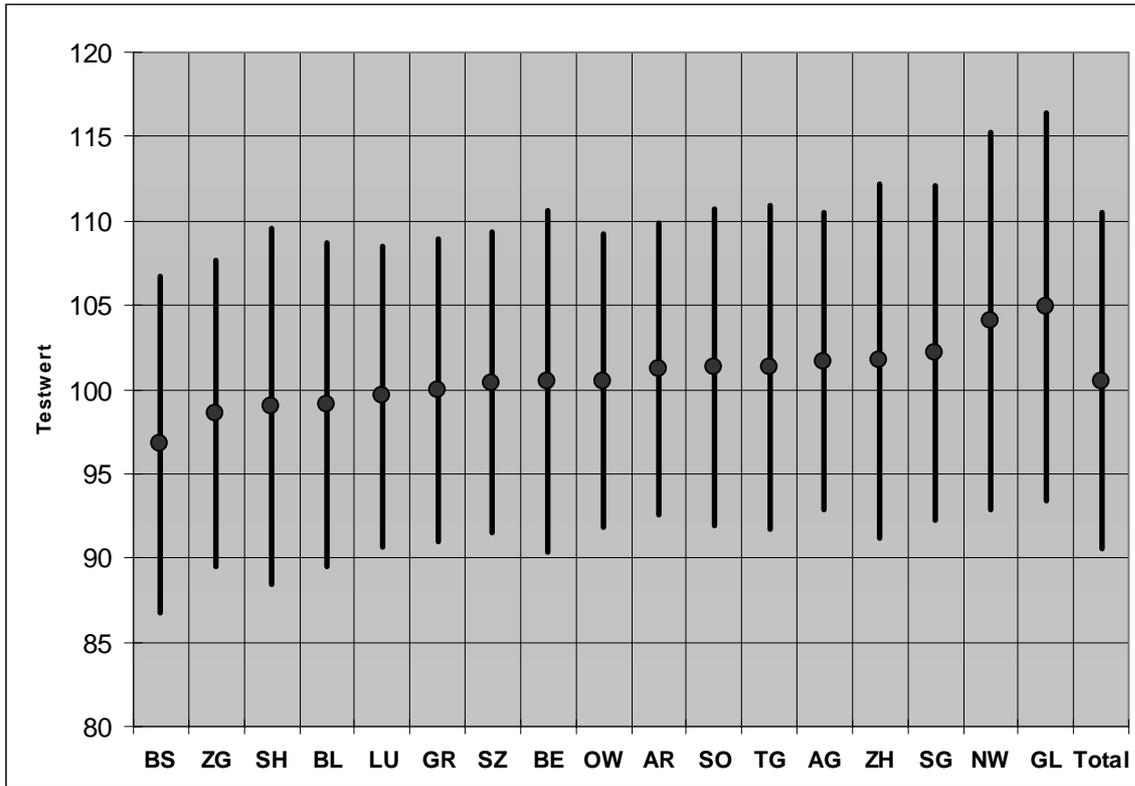


Abbildung 53. Mittelwerte und Streubereiche ($x \pm s$) der Testwerte für Kantone (dt. Sprache)

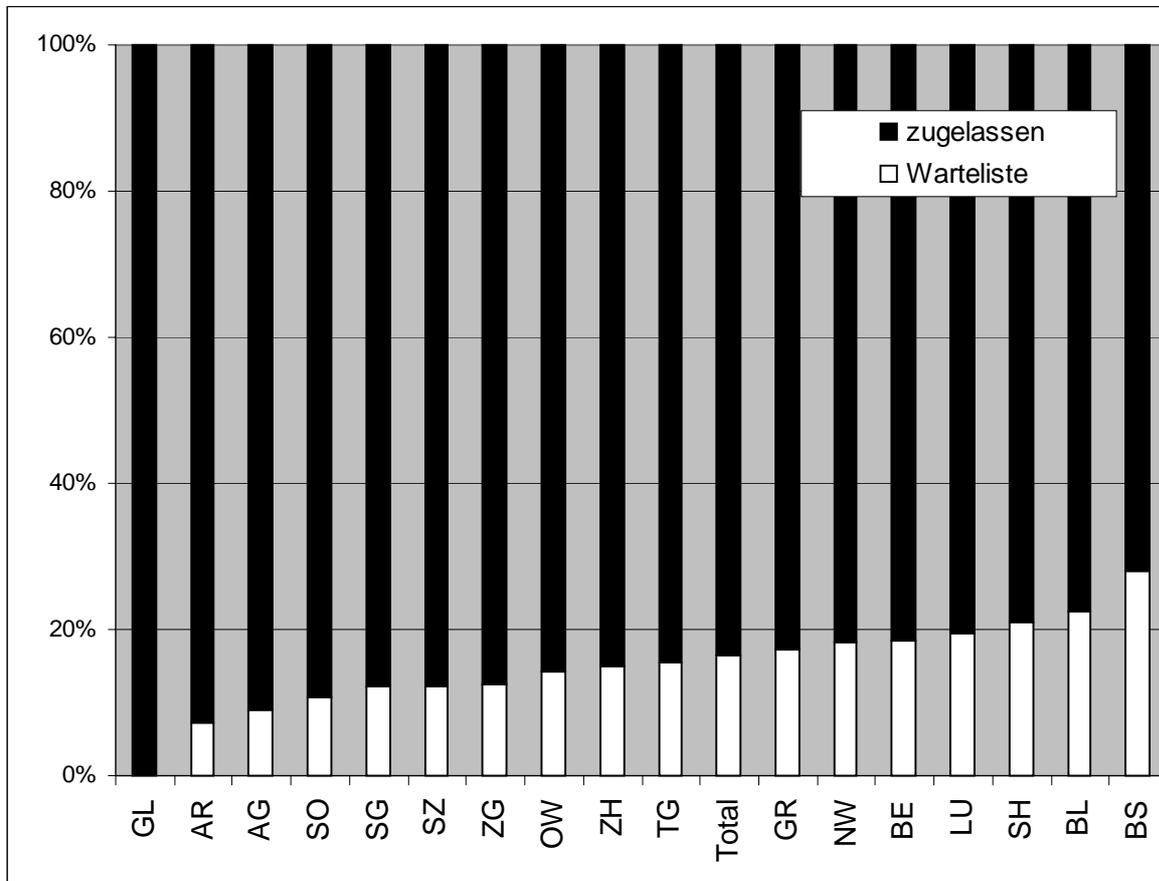


Abbildung 54. Mehrjährige Zulassungsquoten auf der Basis der Studienplatzvergabe nach dem Testtermin (ohne Nachrücker) 1998 bis 2000

Vergleiche für Altersgruppen

1998 und 1999 wurde nachgewiesen, dass Personen, die älter sind, etwas geringere Zulassungschancen haben. Diese waren besonders dann geringer, 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.

Prüfungserfolg im Medizinstudium und Alter

Der Test soll den Prüfungserfolg vorhersagen. Bezüglich der Altersabhängigkeit von Prüfungserfolg lässt die Evaluation erstmals genaue Aussagen zu. Verwendet wurde die Einteilung nach Geburtsjahren beim EMS 1998 (vgl. Bericht 4). Die Aufteilung in frühe und späte Maturität erfolgte danach, ob die Maturität bis zum 22. Lebensjahr abgelegt wurde oder später.

Die Bestehenswahrscheinlichkeit der Prüfung sinkt für ältere Personen ab Geburtsjahr 1977 um etwa 10% ab, allerdings unabhängig davon, wie alt sie sind – wenn sie die Maturität zum frühestmöglichen Zeitpunkt abgelegt haben. Nur Personen, die älter sind und die Maturität relativ spät abgelegt haben (älter als 22) haben eine geringere Bestehensquote.

Geburtsjahr und Maturität	BESTEHEN			Total
	Nicht angetreten	Bestanden	Nicht bestanden	
1952-74, späte Maturität	6 26.1%	12 52.2%	5 21.7%	23 100.0%
1952-74 frühe Maturität	5 20.8%	14 58.3%	5 20.8%	24 100.0%
1975-77	15 9.3%	96 59.3%	51 31.5%	162 100.0%
1978-80	28 7.2%	271 69.8%	89 22.9%	388 100.0%
Total	54 9.0%	393 65.8%	150 25.1%	597 100.0%

Tabelle 54. 1. Vorprüfung 2000 nach Altersgruppen

Bezüglich der Prüfungsnoten (Tabelle 55) gibt es die gleichen Effekte, die Unterschiede sind jedoch nicht signifikant (geringe Stichprobengrösse der älteren Gruppe ist zu beachten). Die Befunde der letzten beiden Jahre sind also identisch sowohl im Eignungstest als auch bei den Prüfungen zu finden. Der Test ist ein Abbild der Realität und sagt auch die altersabhängigen Effekte beim Prüfungserfolg recht gut vorher.

Geburtsjahr und Maturität	N	Subset for alpha = .05
1952-74 späte Maturität	16	3.9375
1975-77	128	4.1348
1952-74, frühe Maturität	15	4.1833
1978-80	327	4.3907
Sig.		.273

Tabelle 55. Notendurchschnitt 1. Vorprüfung nach Altersgruppen und Ergebnis multipler Mittelwertvergleich

Ergebnisse im Jahr 2000

Für die Überprüfung der Alterseffekte im Jahr 2000 wurden die Gruppen auf der Basis der Tabelle 56 gebildet. Die Altersgruppen wurden um 1 Jahr gegenüber 1999 verschoben. Die älteste Gruppe wurde nahe dem Median des Maturitätsalters geteilt. Die Befunde aus 1998 und 1999 können auch im Jahr 2000 nahezu identisch bestätigt werden. Es gibt einen generellen Alterseffekt (jüngere haben etwas höhere Testwerte und Zulassungschancen als ältere) und die älteste Gruppe differenziert sich nach dem Maturitätsalter.

Maturitätsjahr	Geburtsjahr			Total
	1958-1976	1977-1979	1980-1982	
1978	1			1
1985	1			1
1988	3			3
1989	1			1
1990	4			4
1991	3			3
1992	6			6
1993	5			5
1994	7			7
1995	3			3
1996	5	1		6
1997	2	9		11
1998	1	37	1	39
1999	7	92	65	164
2000	31	121	395	547
	80	260	461	801

Tabelle 56. Gruppenbildung für Jahr des Ablegens der Maturitätsprüfung bezogen auf die Geburtsjahre

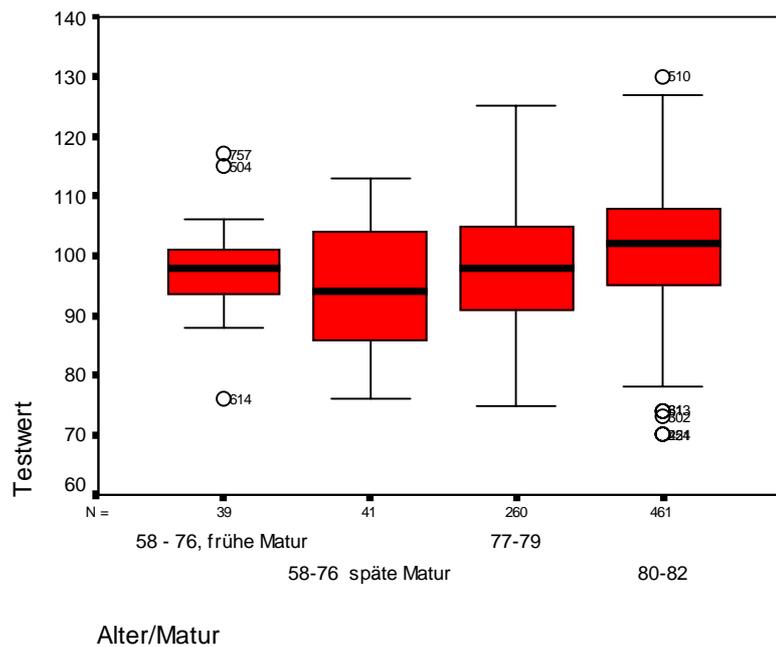


Abbildung 55. Boxplots für Testwert der Altergruppen (Geburtsjahre) und Maturität

Geburtsjahr und Maturität	Warteliste	zugelassen	
58 - 76, frühe Maturität	7	32	39
	17.9%	82.1%	100.0%
58 -76 späte Maturität	16	25	41
	39.0%	61.0%	100.0%
77-79	61	199	260
	23.5%	76.5%	100.0%
80-82	59	402	461
	12.8%	87.2%	100.0%
	143	658	801
	17.9%	82.1%	100.0%

Tabelle 57. Zulassungsquoten für die drei Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung

	Geburtsjahr und Maturität	Mittelwert	Standardabw.	Sum of Squares	df	Mean Square	F	Sig.
Testwert	58-76, frühe Matur.	97.3	7.3	3689.2	3	1229.7	12.9	.000
	58-76, späte Matur.	94.2	10.1					
	77-79	98.3	9.9					
	80-82	101.6	9.7					
Muster zuordnen	58-76, frühe Matur.	8.9	1.9	155.9	3	51.9	6.7	.000
	58-76, späte Matur.	8.9	3.4					
	77-79	10.1	2.7					
	80-82	10.4	2.8					
Med.-naturwiss. Grundverständnis	58-76, frühe Matur.	10.6	2.9	177.2	3	59.0	5.9	.001
	58-76, späte Matur.	8.8	2.8					
	77-79	10.1	3.2					
	80-82	10.7	3.1					
Schlauchfiguren	58-76, frühe Matur.	12.4	2.8	128.0	3	42.6	3.3	.019
	58-76, späte Matur.	11.3	3.8					
	77-79	12.9	3.6					
	80-82	13.1	3.5					
Quantitative und formale Probleme	58-76, frühe Matur.	11.4	3.3	392.0	3	130.6	8.9	.000
	58-76, späte Matur.	9.8	4.3					
	77-79	10.9	3.8					
	80-82	12.1	3.7					
Textverständnis	58-76, frühe Matur.	8.3	4.0	194.8	3	64.9	4.6	.003
	58-76, späte Matur.	7.7	3.6					
	77-79	8.2	3.7					
	80-82	9.1	3.6					
Figuren lernen	58-76, frühe Matur.	9.0	3.3	206.3	3	68.7	6.0	.000
	58-76, späte Matur.	8.7	3.0					
	77-79	9.7	3.3					
	80-82	10.4	3.4					
Fakten lernen	58-76, frühe Matur.	9.2	3.2	266.8	3	88.9	7.8	.000
	58-76, späte Matur.	9.2	3.4					
	77-79	9.3	3.1					
	80-82	10.4	3.5					
Diagramme und Tabellen	58-76, frühe Matur.	9.9	2.8	158.2	3	52.7	5.1	.002
	58-76, späte Matur.	9.1	3.3					
	77-79	9.7	3.2					
	80-82	10.5	3.1					
Konzentriertes und sorgfältiges Arbeiten	58-76, frühe Matur.	11.8	2.9	93.2	3	31.0	2.4	.061
	58-76, späte Matur.	11.9	3.9					
	77-79	12.5	3.7					
	80-82	12.9	3.4					

Tabelle 58. Testwert und Punktwerte für die drei Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und späterer Maturitätsprüfung; varianzanalytische Prüfung des Unterschiedes

Die nachfolgende Abbildung zeigt, dass sich der Effekt gegenüber 1999 bezüglich der ältesten Gruppe mit später Maturität sogar noch etwas verstärkt hat.

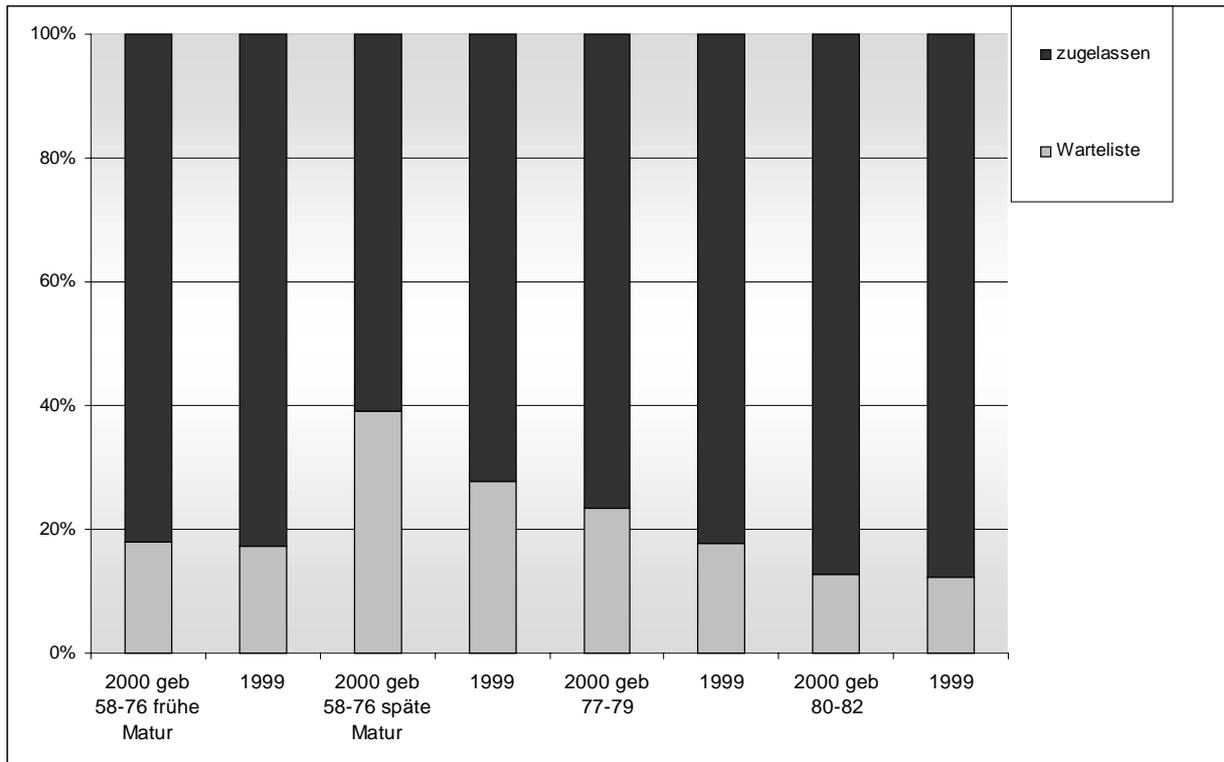


Abbildung 56. Zulassungsquoten für Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung

Vergleiche nach Wunschuniversitäten

Die Zulassung erfolgt aufgrund des Testwertes, der die Sprachunterschiede ausgleicht. Ein Vergleich der Zulassungsquoten bezogen auf die jeweiligen Wunschuniversitäten zeigt, dass es keine signifikanten Unterschiede bezüglich dieser Quoten gibt: die Bewerber an allen Universitäten unterscheiden sich bezüglich der Studieneignung nicht.

Disziplinen	Universität	Zulassung		Total
		Warteliste	zugelassen	
Humanmedizin	Basel	32	101	133
		24.1%	75.9%	100.0%
	Bern	23	121	144
		16.0%	84.0%	100.0%
	Fribourg	15	51	66
	22.7%	77.3%	100.0%	
Zürich	55	272	327	
	16.8%	83.2%	100.0%	
Total	125	545	670	
	18.7%	81.3%	100.0%	
Veterinärmedizin	Bern	15	84	99
		15.2%	84.8%	100.0%
	Zürich	5	65	70
	7.1%	92.9%	100.0%	
Total	20	149	169	
	11.8%	88.2%	100.0%	

Tabelle 59. Zulassungsquoten für Disziplinen und Wunschuniversität (erste Wahl)

Disziplin		Value	df	Asymp. Sig. (2-seitig)
Humanmedizin	Pearson Chi-Square	4.690	3	.196
	Likelihood Ratio	4.536	3	.209
Veterinärmedizin	Pearson Chi-Square	2.521	1	.112
	Likelihood Ratio	2.661	1	.103

Tabelle 60. Prüfung Homogenität Zulassung nach Wunschuniversität: keine signifikanten Unterschiede

Ergebnisse zur Testgüte

Zuverlässigkeit

Die Zuverlässigkeit (Messgenauigkeit – deutsche Ergebnisse vgl. Trost et al., 1995) kann anhand zweier Koeffizienten verglichen werden. Zunächst wurde die Zuverlässigkeit der Untertests abgeschätzt. Beim „Konzentrierten 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. Wie sich zeigt, liegen die meisten der in der Schweiz ermittelten Werte in der Bandbreite der für je 16 parallelen Testformen pro Jahr in Deutschland in den Jahren 1995 und 1996 ermittelten Konsistenzen.

Die Reliabilitäten nach der Testhalbierungsmethode (Teilung nach gerade- und ungeradzahligen Aufgaben) wurden ebenfalls verglichen. Wegen der Stichprobengrößen werden nur die Ergebnisse des deutschsprachigen Tests der Schweiz verwendet. Die Koeffizienten liegen mindestens im gleichen Wertebereich wie in Deutschland, bei zwei Untertests leicht darüber.

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

Tabelle 61. Zuverlässigkeit 1998 bis 2000; 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 0.90 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 62 zeigt, dass das Weglassen von Untertests zu einer Verminderung der Konsistenz führen würde (auch bei „Konzentriertem und sorgfältigem Arbeiten“). Die Konsistenz des Profils aufgrund der Untertests liegt bei 0.81.

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

Tabelle 62. Innere Konsistenzen, wenn der jeweilige Untertest weggelassen worden wäre und Konsistenz des Testprofils („standardized alpha“)

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. Gruppierung der Untertestleistungen können zu den bekannten Faktormodellen der Intelligenz in Beziehung gesetzt werden.

Alle Korrelationen sind mindestens auf dem 5%-Niveau signifikant. Die Korrelationen der Untertests mit dem Punktwert liegen ausnahmslos über 0.50 – ein erster Hinweis für eine starke gemeinsame Komponente.

In Deutschland wurde über viele Jahre eine Dreifaktoren-Struktur repliziert, die folgende Faktoren enthält (Trost et al. 1998, S. 44). Die Faktoren decken sich mit den Befunden von Thurstone, auf den die faktorenanalytisch begründeten Theorien letztlich zurückgehen (siehe z.B. Amelang und Bartussek, 1990).

- Die Fähigkeit zum schlussfolgernden Denken im Kontext von Medizin und Naturwissenschaften (als „reasoning“ einer der bedeutendsten Faktoren bei Thurstone),

- die Fähigkeit zur Verarbeitung visueller Information (als „perceptional speed“ bei Thurstone identifiziert),
- die Merkfähigkeit (als „memory“ bei Thurstone) - wobei verbales und figürliches Gedächtnis als die beiden „Gegensätze“ gleichermaßen berücksichtigt sind.

	Muster zuordnen	Med.-naturwiss. Grundv.	Schlauchfiguren	Quantitat. und formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Diagramme und Tabellen
Muster zuordnen	1.000	.247	.419	.243	.226	.327	.243	.251	.475	.718
Med.-naturw. Grundverst.	.247	1.000	.346	.539	.615	.243	.308	.536	.237	.315
Schlauchfiguren	.419	.346	1.000	.364	.293	.317	.266	.335	.449	.393
Quant. formale Probl	.243	.539	.364	1.000	.539	.229	.253	.611	.164	.324
Textverständnis	.226	.615	.293	.539	1.000	.212	.298	.525	.175	.231
Figuren lernen	.327	.243	.317	.229	.212	1.000	.346	.241	.355	.581
Fakten lernen	.243	.308	.266	.253	.298	.346	1.000	.249	.328	.590
Diagramme und Tabellen	.251	.536	.335	.611	.525	.241	.249	1.000	.174	.502
Konzent. u. sorgfält. Arbeiten	.475	.237	.449	.164	.175	.355	.328	.174	1.000	.265
Punktwert	.575	.709	.662	.701	.687	.565	.573	.684	.581	1.000

Tabelle 63. Korrelationen zwischen Punktwerten und Untertests

Die Skalenstruktur der Untertests des EMS wurde faktorenanalytisch geprüft. Verwendet wurde die Hauptkomponentenmethode mit anschließender Varimaxrotation. Die Dreifaktorenstruktur des deutschen TMS kann repliziert werden, die Ladungsmuster stimmen gut überein. Es finden sich darüber hinaus Belege, dass die Faktorenstruktur hierarchisch ist und der von 1998 und 1999 sehr genau entspricht:

- Alle Untertests laden signifikant im ersten unrotierten Faktor (Tabelle 65). Dies ist als Hinweis auf einen Generalfaktor der Leistungsfähigkeit zu interpretieren; es ist gemeinsame Varianz zwischen allen Untertests vorhanden. Mit einem Anteil von 41,1% an der aufgeklärten Varianz ist er genauso bedeutsam wie 1998 und 1999.
- Aufgrund mehrerer Kriterien (Eigenwerte > 1 , 50% Varianzaufklärung, Scree-Test) ist auch die Zwei-Faktorenlösung naheliegend. Hier findet sich der Faktor „Fähigkeit zum schlussfolgernden Denken“. Der zweite Faktor, der „Fähigkeit zur Verarbeitung visueller Information“ und „Merkfähigkeit“ zusammenfasst, weist auf gemeinsame Varianz der zugrundeliegenden Untertests hin. Er wäre im Sinne von „Werkzeug-

funktionen“ bzw. Leistungsvoraussetzungen (Aufmerksamkeit, Wahrnehmung, Gedächtnis) zu interpretieren. Bei dieser Lösung wird wiederum 15% Varianz mehr aufgeklärt als bei der Ein-Faktorenlösung.

- Die Drei-Faktorenlösung entspricht dem aus Deutschland bekannten Ergebnis, wobei genau wie 1998 und 1999 rund 10% mehr Varianz aufgeklärt werden als bei der Zwei-Faktorenlösung. Weil der Faktor „Merkfähigkeit“ nur durch zwei Untertests repräsentiert ist, bleibt dieser relativ geringe Zugewinn erklärlich.
- Nach dem Kriterium 5% Eigenwerteanteil wären 1999 ebenfalls sieben Faktoren noch in Betracht zu ziehen. Es würden 91% der Varianz aufgeklärt. Da dies aber fast der Zahl der Untertests entspricht, wird diese Lösung nicht weiter betrachtet.

Faktor	Unrotierte Lösung						Rotierte Lösungen					
	Eigenwert		% Varianz		Kumuliert %		Eigenwert		% Varianz		Kumuliert %	
	2000	1999	2000	1999	2000	1999	2000	1999	2000	1999	2000	1999
1	3.700	3.670	41.116	40.782	41.116	40.782	Zweifaktorenlösung					
2	1.456	1.364	16.179	15.158	57.295	55.940	2.744	2.723	30.491	30.254	30.491	30.254
3	.838	.914	9.313	10.160	66.608	66.100	2.412	2.312	26.804	25.687	57.295	55.940
4	.660	.639	7.337	7.099	73.945	73.199	Dreifaktorenlösung					
5	.576	.585	6.401	6.503	80.346	79.703	2.711	2.683	30.123	29.806	30.123	29.806
6	.534	.535	5.933	5.946	86.279	85.649	1.916	1.811	21.289	20.122	51.411	49.928
7	.471	.527	5.230	5.856	91.509	91.505	1.368	1.455	15.197	16.172	66.608	66.100
8	.388	.418	4.309	4.644	95.818	96.150						
9	.376	.347	4.182	3.850	100.000	100.000						

Tabelle 64. Varianzanteile der einzelnen Faktorenlösungen

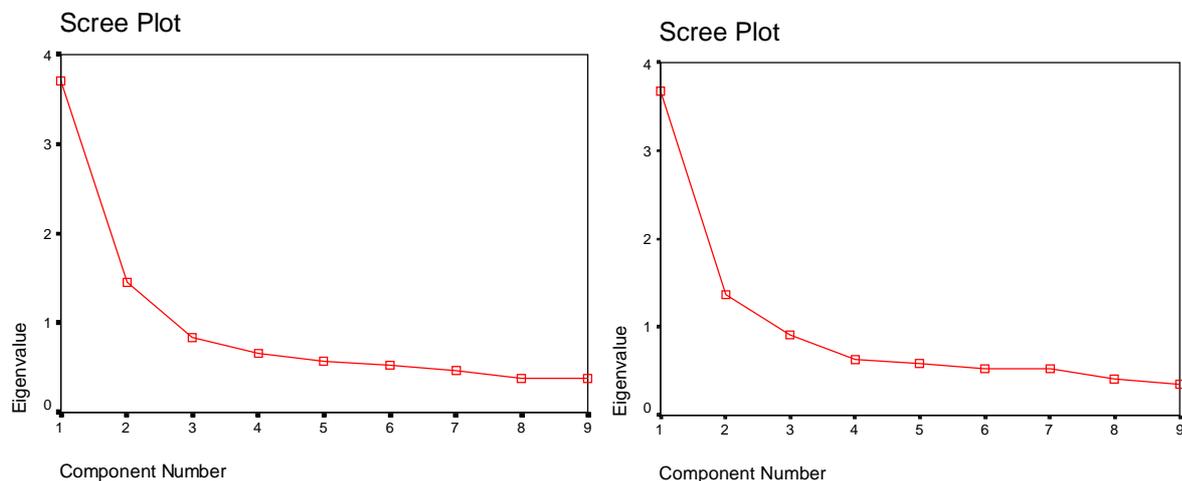


Abbildung 57. Scree-Plot (Eigenwerteverlauf) der Faktoren 1 bis 9 für die Untertests des EMS 2000 (links) und 1999 (rechts)

Im „Scree-Plot“ (Abbildung 57) wird der Eigenwerteverlauf der Faktoren dargestellt. Die Ein-Faktorenlösung ist 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. Die Abbildung 58 zur Struktur der Untertests fasst die Ergebnisse der drei nachfolgenden Analysen 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.

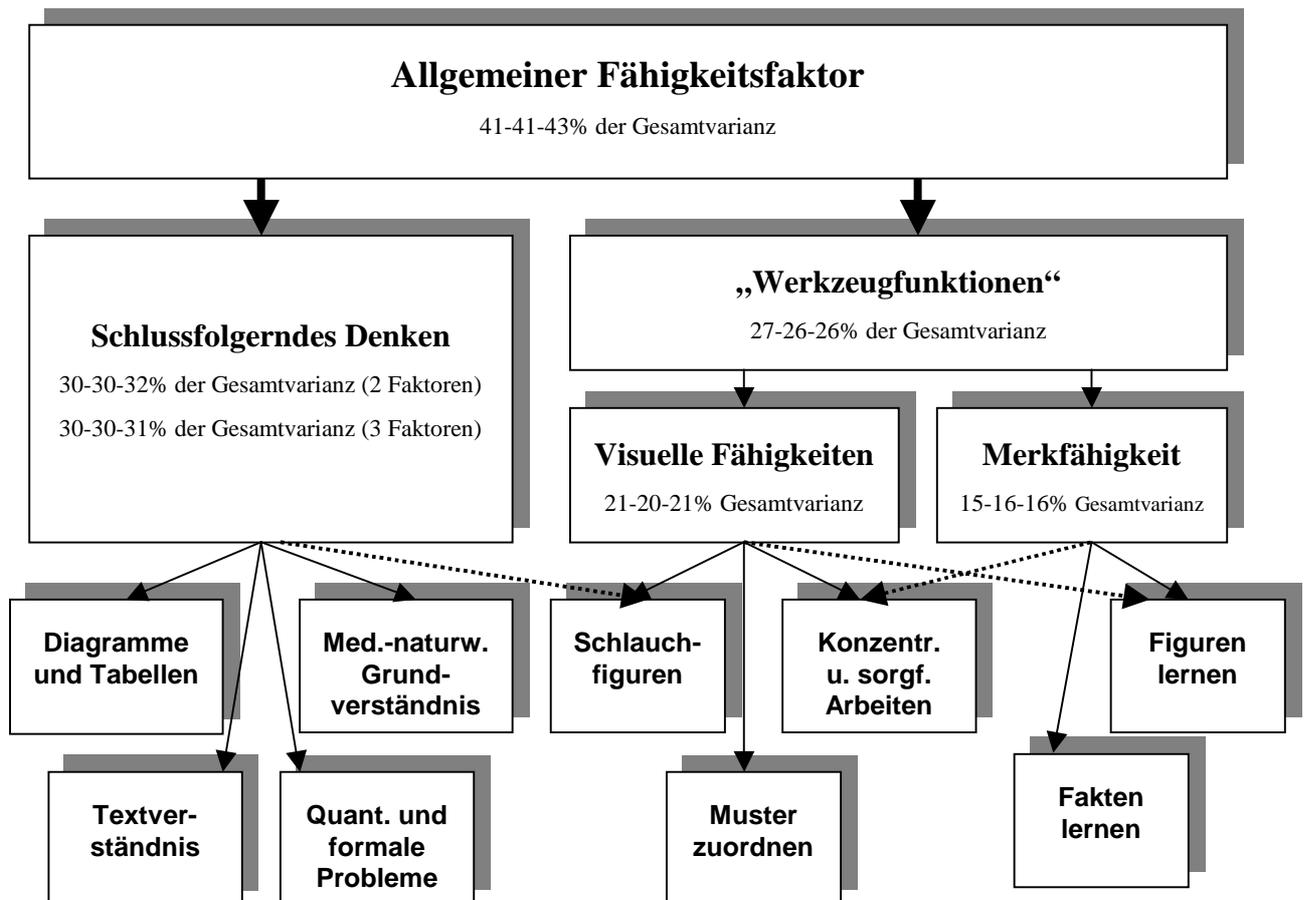


Abbildung 58. Struktur der Untertests des EMS, hierarchische Faktorenstruktur (2000 erster Wert, 1999 zweiter und 1998 dritter Wert der Varianzaufklärung durch den Faktor)

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

Tabelle 65. Ladungen im ersten Faktor der unrotierten Lösung 1998 bis 2000 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor)

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

Tabelle 66. 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. Der Untertest „Schlauchfiguren“ weist ebenfalls eine signifikante Ladung im Faktor „Schlussfolgerndes Denken“ auf (auch in der nachfolgend dargestellten Drei-Faktorenlösung). Er ist kein reiner

Wahrnehmungstest. In der Drei-Faktorenlösung teilt sich der Faktor „Werkzeugfunktionen“ in die Komponenten „Wahrnehmung“ und „Gedächtnis“ auf – der Faktor „Schlussfolgerndes Denken“ bleibt bezüglich der Ladungen und des Varianzanteiles stabil.

Faktor	Schweiz Gesamt								Deutschland		
	1		2		3		h ²		1	2	3
	2000	1999	2000	1999	2000	1999	2000	1999			
Diagramme und Tabellen	.80	.76		.29			.67	.67	.82		
Medizin.-naturwiss. Grundverständnis	.78	.81			.19	.22	.66	.71	.81		.20
Quantitat. u. formale Probleme	.81	.78	.17	(.17)			.69	.64	.80	.18	
Textverständnis	.80	.80			.18	.17	.67	.67	.79		.20
Muster zuordnen			.79	.77			.66	.62		.81	
Konzentr. u. sorgfält. Arbeiten			.75	.65	.33	.44	.67	.62		.70	.40
Schlauchfiguren	.32	.33	.72	.70			.62	.60	.35	.71	
Fakten lernen	.22				.85	.86	.78	.79	.21		.87
Figuren lernen	.12	(.14)	.34	.42	.67	.66	.58	.63	.13	.47	.64

Tabelle 67. Faktorenanalyse: Varimaxrotierte Dreifaktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42)

Item-Trennschärfen

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

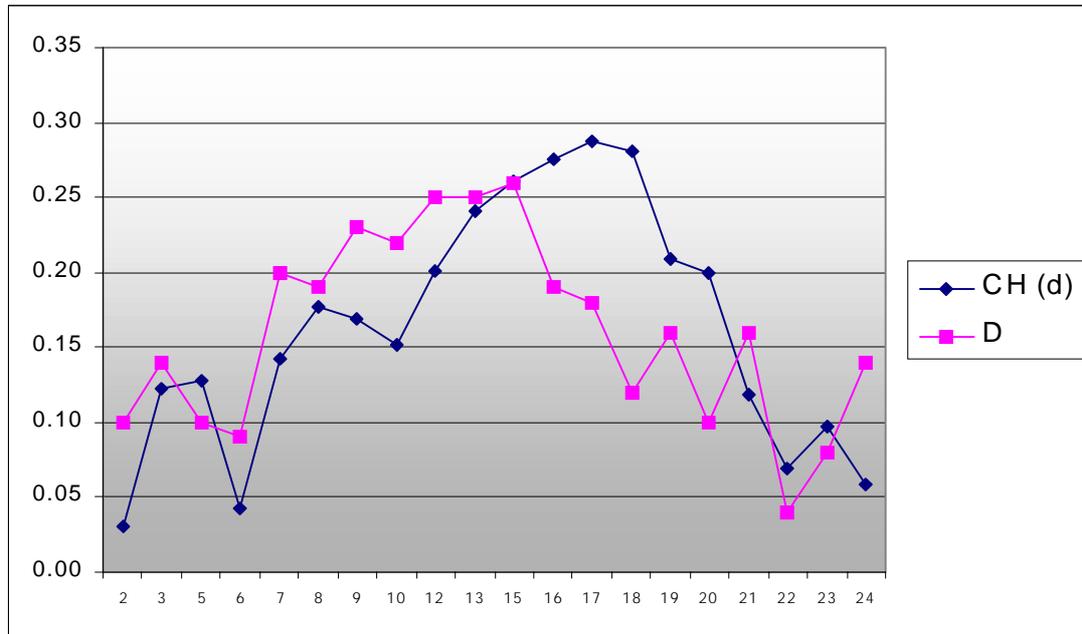


Abbildung 59. Trennschärfen für den Untertest "Muster zuordnen"

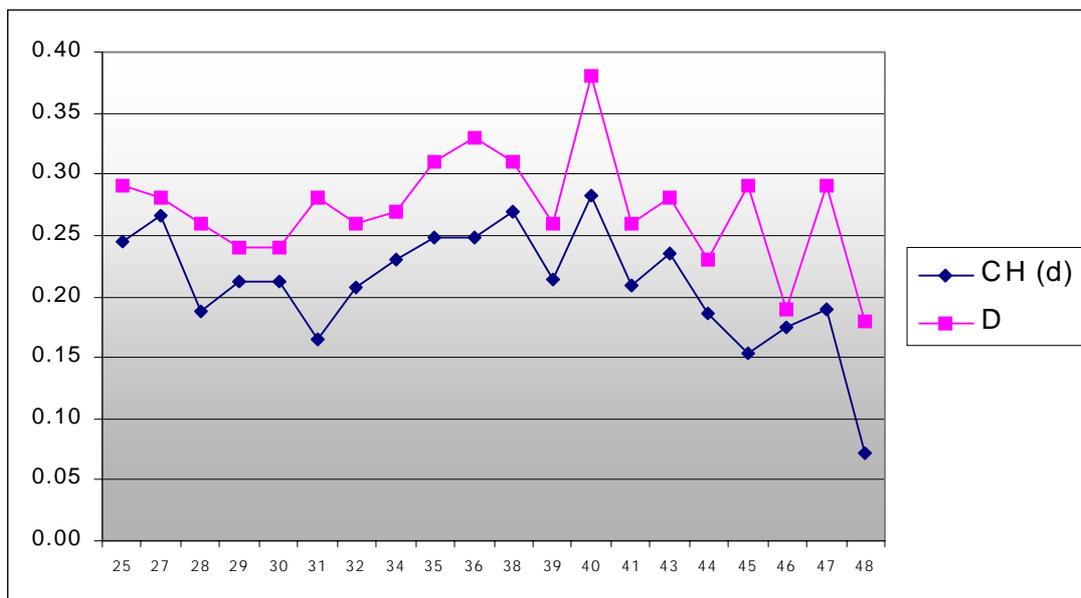


Abbildung 60. Trennschärfen für den Untertest "Medizinisch-naturwissenschaftliches Grundverständnis"

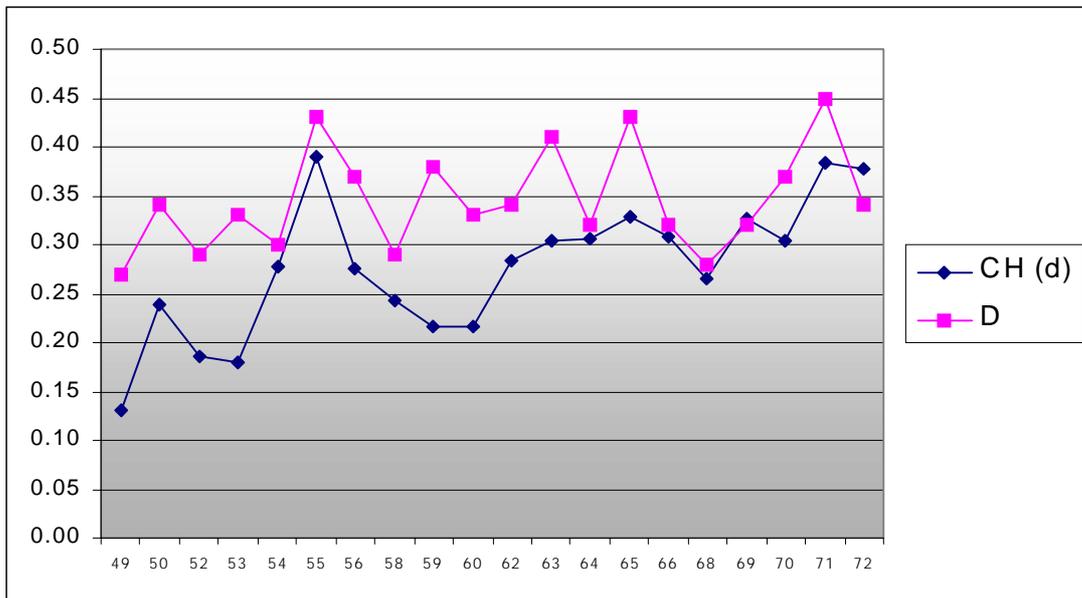


Abbildung 61. Trennschärfe für den Untertest "Schlauchfiguren"

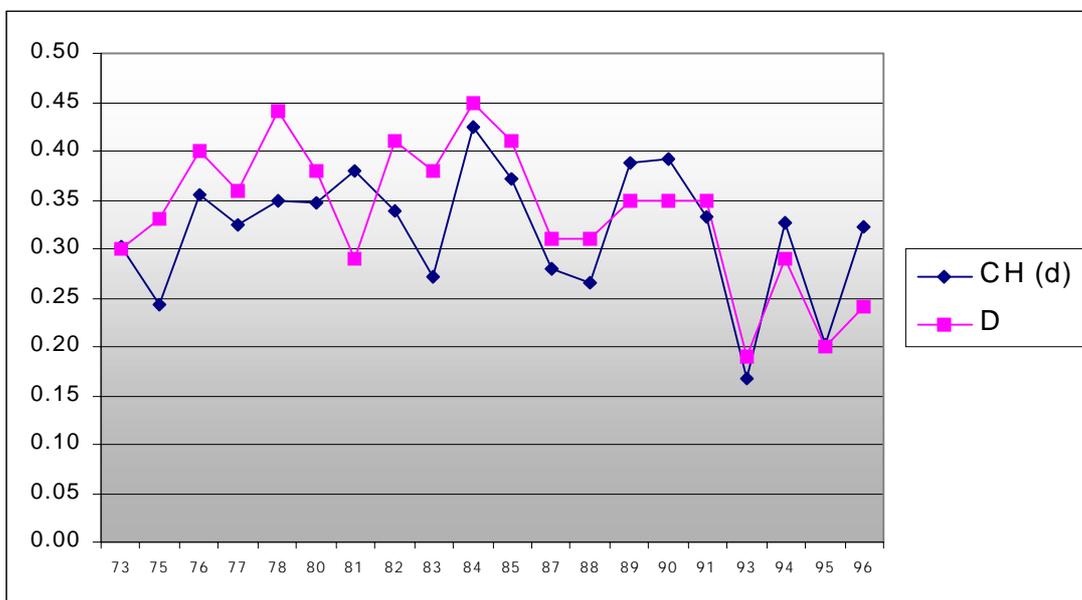


Abbildung 62. Trennschärfe für den Untertest "Quantitative und formale Probleme"

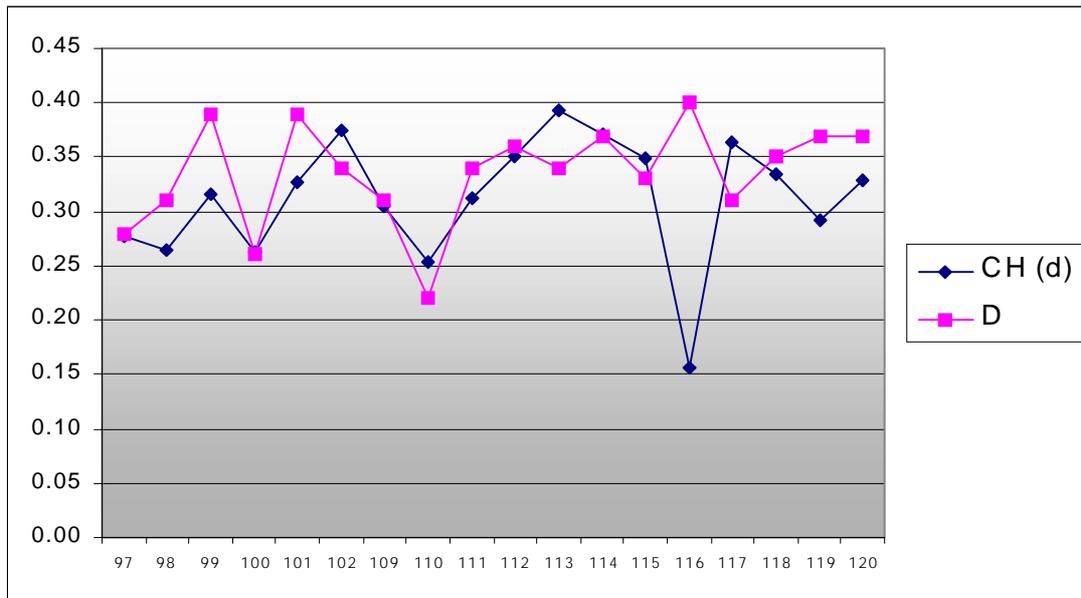


Abbildung 63. Trennschärpen für den Untertest "Textverständnis"

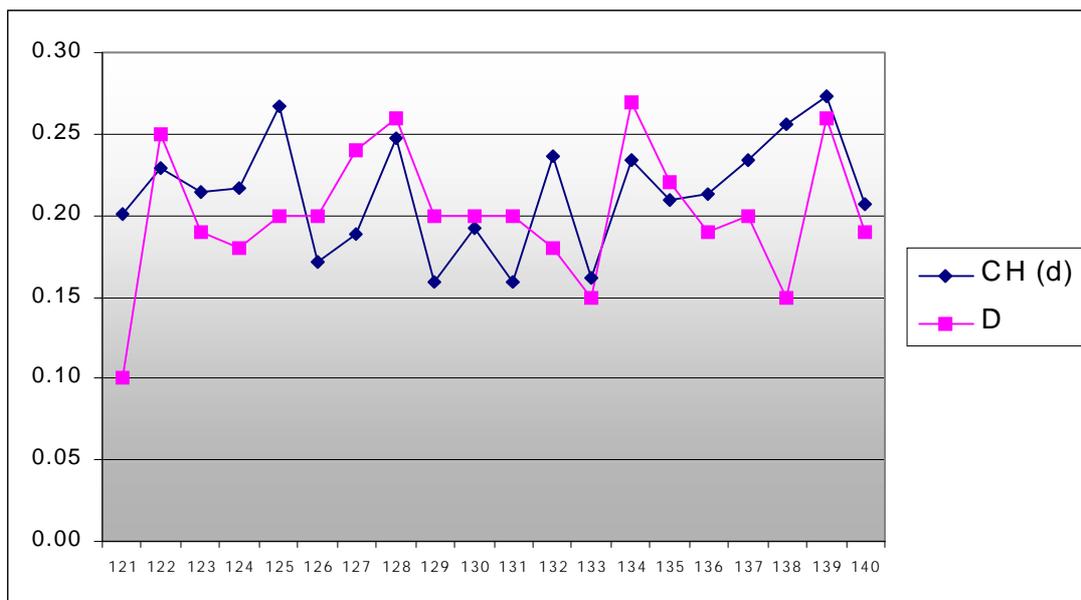


Abbildung 64. Trennschärpen für den Untertest "Figuren lernen"

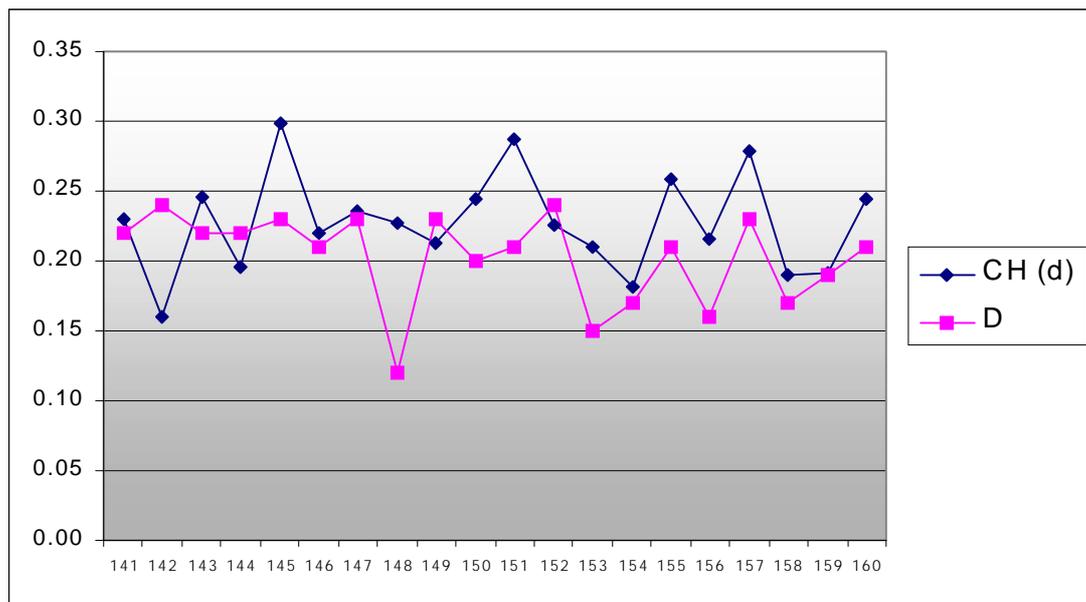


Abbildung 65. Trennschärfen für den Untertest "Fakten lernen"

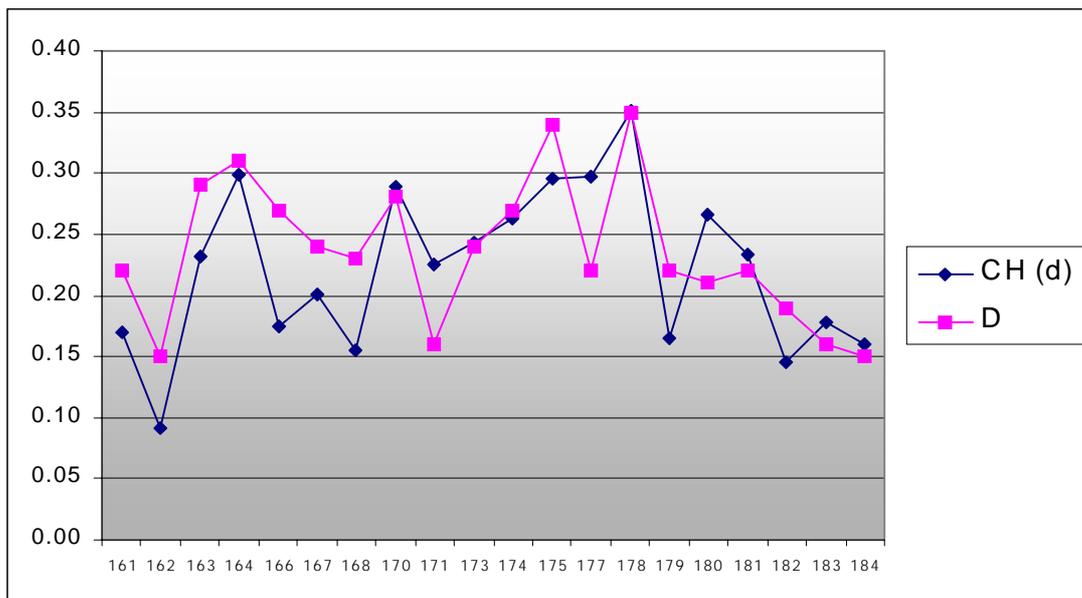


Abbildung 66. Trennschärfen für den Untertest "Diagramme und Tabellen"

Konzentriertes und sorgfältiges Arbeiten

Bei diesem Untertest kann die Zuverlässigkeit nicht nach dem Itemkonzept eingeschätzt werden. Es wird die Menge der Richtigen d.h. die Differenz bearbeiteter Zeichen minus Fehler berechnet. Zwei Fehlertypen können auftreten: übersehene Zeichen und solche, die fälschlich durchgestrichen worden sind. Die Korrelation zwischen Menge und Fehlern ist nicht signifikant: schnelles Beantworten ist nicht zwingend mit einer niedrigeren Sorgfalt verbunden. Dies entspricht den Ergebnissen aus dem Jahr 1998 und 1999.

Insgesamt gibt es aufgrund der Ergebnisse keinen Grund, die Testdauer zu verändern. In Deutschland haben sich die Mittelwerte nach mehrmaligem Einsatz des Tests zu besseren Leistungen hin verändert und der Test differenzierte nicht mehr optimal. Dies wurde vor allem auf eine bessere Testvorbereitung der Kandidatinnen und Kandidaten zurückgeführt.

2000	Deutsch		Französisch		Italienisch	
	m	s	m	s	m	s
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 durchgestr.	3.46	4.14	2.22	2.56	2.45	2.67
1999						
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
1998						
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 68. Kennwerte für den Untertest „Konzentriertes und sorgfältiges Arbeiten“

	Falsch	...übersehen	...fälschlich durchgestrichen
Richtig	.069 .051	.037 .290	.187 .000
Falsch		.984 .000	.371 .000
...übersehen	.984 .000		.203 .000

Tabelle 69: Korrelationen der Parameter im Konzentrationstest; kursiv: Signifikanzniveau

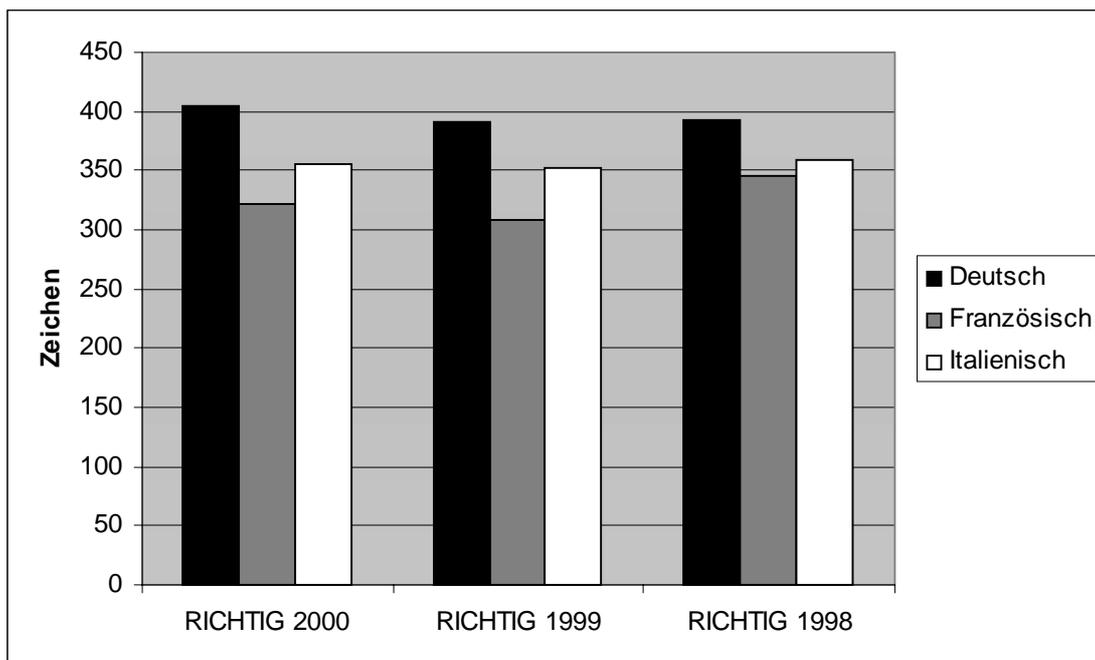


Abbildung 68: Richtige im Untertest Konzentriertes und sorgfältiges Arbeiten 1998 bis 2000 nach Sprachgruppen

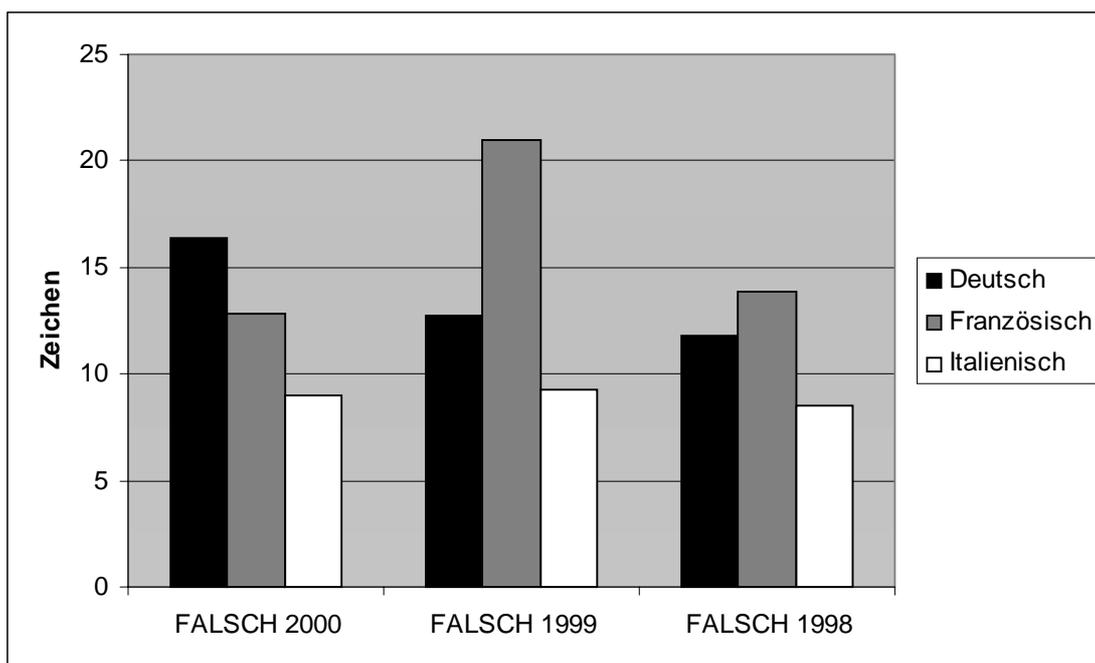


Abbildung 67: Fehler im Untertest Konzentriertes und sorgfältiges Arbeiten 1998 bis 2000 nach Sprachgruppen

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 and 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 Züricher Zeitung. 29.9. 1999 http://www.nzz.ch/online/01_nzz_aktuell/sonderbeilagen/studium98/studium98celio.htm
- Cook, LL (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 2: EMS und Chancengleichheit. Schweizerische Ärztezeitung Heft 13 S. 723-730.
- 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., Ischi, N. (Hrsg.) (1997). Eignungsprüfung für das Medizinstudium. Tagungsband. Berichte des ZTD Band 3. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Hänsgen, K.-D., Hofer, R. & Ruefli, D. (Hrsg.) (1996). Eignungsdiagnostik und Medizinstudium. Tagungsband. Berichte des ZTD Band 2. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- 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.
- Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995). Der Eignungstest für das Medizinstudium in der Schweiz. Ein Probelauf. Berichte des ZTD Band 1. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- 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.

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 Freiburg (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 Freiburg (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.

Frühere Berichte des ZTD

Bericht 5 (1999)	
<p>EMS Eignungstest für das Medizinstudium in der Schweiz 1999 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i></p> <p>K.-D. Hänsgen und B. Spicher</p> <p>http://www.unifr.ch/ztd/test/bericht5/bericht5start.html</p>	
Bericht 4 (1998)	
<p>EMS Eignungstest für das Medizinstudium in der Schweiz 1998 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i></p> <p>K.-D. Hänsgen und B. Spicher</p> <p>http://www.unifr.ch/ztd/test/bericht4/bericht4start.html</p>	
Bericht 3 (1997) - Rapport 3 (1997)	
<p>Eignungsprüfung für das Medizinstudium - Kriterien und Testverfahren <i>Bericht über das Internationale Symposium in Bern am 8. November 1996</i></p> <p>hrsg. von K.-D. Hänsgen und N. Ischi</p>	<p>L'examen d'aptitude aux études de médecine en Suisse - Critères et procédés d'application du test <i>Rapport sur le Symposium international à Berne</i></p> <p>Editeurs: K.-D. Hänsgen; N. Ischi</p>
<p>Günter Trost Eignungskriterien bei der Zulassung zum Medizinstudium in Europa: Ergebnisse einer Erhebung in allen europäischen Ländern</p> <p>Ingemar Wedman & Widar Henriksson The Swedish Scholastic Aptitude Test. Research and main findings</p> <p>John L. Hackett The Medical College Admission Test (MCAT) - its use in U.S. and Canada and some results of validation studies</p> <p>Piet J. Janssen Admission to the study in medicine in Belgium: two 'different' solutions to the 'same' problem; reflections of a Flemish school psychologist</p> <p>Klaus-Dieter Hänsgen Der Eignungstest für das Medizinstudium in der Schweiz als Instrument für die Beschränkung der Studienzulassung</p> <p>Rainer Hofer & Klaus-Dieter Hänsgen Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium in der Schweiz</p> <p>Urs Schallberger Anforderungen an das Zulassungsverfahren für das Medizinstudium in der Schweiz: Leitlinien für die Entwicklung eines eignungsdiagnostischen Verfahrens</p>	

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

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

Anhang zum Bericht 2000

Verteilung der Punktwerte

Gesamt	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	10	27	29	56	98	103	105	113	91	67	39	38	11	5	2	1	
Med.-nat. Grundv.	1		2	7	11	24	41	58	77	86	99	109	80	73	43	39	26	17	4	2	2
Schlauchfiguren			1	2	4	10	17	24	33	53	58	63	96	85	83	79	53	51	40	25	24
Qua.form Probl.	1	1	10	6	19	26	20	36	48	64	65	86	77	71	73	67	51	36	23	17	4
Textverständnis		7	9	42	42	63	86	79	85	79	61	59	56	30	30	26	26	15	6		
Figuren lernen	1	1	4	9	22	35	51	69	82	80	72	92	83	63	58	37	20	10	11		1
Fakten lernen	1		3	10	13	41	54	82	98	73	93	83	73	49	43	30	28	13	7	2	5
Diagr und Tabellen		1	3	9	19	22	45	67	72	106	93	98	74	61	47	40	22	16	3	3	
Kon.sorgf .Arbeiten	1			2	1	4	11	30	51	66	74	72	75	91	76	66	46	45	40	29	21

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			2	3	8	22	25	47	86	88	93	101	77	60	35	37	11	5	2	1	
Med.-nat. Grundv.	1		2	5	9	14	33	46	66	70	86	97	76	68	42	37	26	17	4	2	2
Schlauchfiguren				2	3	6	12	16	30	46	42	55	82	74	79	71	49	48	39	25	24
Qua.form Probl.	1	1	9	5	13	19	18	30	36	55	61	76	66	64	67	61	46	32	22	17	4
Textverständnis		5	6	31	34	51	70	65	78	71	55	54	53	29	30	25	25	15	6		
Figuren lernen		1	4	8	14	28	45	61	70	72	64	84	71	54	54	35	19	10	8		1
Fakten lernen	1		3	9	11	36	44	71	78	61	85	71	68	46	38	28	26	13	7	2	5
Diagr und Tabellen		1	2	6	15	19	39	50	62	95	82	88	66	57	42	35	22	16	3	3	
Kon.sorgf .Arbeiten				2	1	3	8	25	38	51	59	60	68	84	67	63	44	43	40	28	19

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				1	1	4	3	8	7	10	9	9	9	5	3						
Med.-nat. Grundv.				1	2	7	6	5	9	11	11	10	2	2	1	2					
Schlauchfiguren			1		1	3	4	7	3	5	10	5	9	6	4	6	3	1	1		
Qua.form Probl.			1		4	5	2	3	8	7	3	7	8	5	4	3	4	4	1		
Textverständnis		1	1	6	7	8	12	14	3	4	4	5	2	1			1				
Figuren lernen	1			1	6	6	5	6	6	7	7	3	11	3	2	2	1		2		
Fakten lernen				1	2	6	8	17	6	6	10	4	2	3	2	2					
Diagr und Tabellen			1	1	3	1	4	13	7	8	7	7	5	2	5	5					
Kon.sorgf .Arbeiten	1					1	3	3	9	12	10	8	6	4	7	3	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	1	1	5	5	3	3	5	2	1	1					
Med.-nat. Grundv.				1		3	2	7	2	5	2	2	2	3							
Schlauchfiguren						1	1	1		2	6	3	5	5		2	1	2			
Qua.form Probl.				1	2	2		3	4	2	1	3	3	2	2	3	1				
Textverständnis		1	2	5	1	4	4		4	4	2		1			1					
Figuren lernen					2	1	1	2	6	1	1	5	1	6	2				1		
Fakten lernen				1	1	3	4	3	3	6	2	2	1	1	2						
Diagr und Tabellen				2	1	2	2	4	3	3	4	3	3	2							
Kon.sorgf .Arbeiten								2	4	3	5	4	1	3	2		1	2			2

Tabelle A 4. Häufigkeitsverteilung der Punkte italienische Sprachgruppe

	Punktwert	Testwert	Muster zuordnen	Med.- naturw. Grundv.	Schlauch -figuren	Quantitat . formale Probleme	Text- verständ- nis	Figuren lernen	Fakte n lernen	Diagram -me und Tabellen	Konzent . und sorgf. Arb.
Deutsch											
Mittelwert	98.62	100.75	10.27	10.70	13.21	11.78	9.04	10.20	10.08	10.36	12.97
N	703	703	703	703	703	703	703	703	703	703	703
Standardabw.	19.45	9.93	2.82	3.17	3.53	3.85	3.75	3.36	3.48	3.22	3.52
Kurtosis	.126	-.067	-.131	-.087	-.400	-.231	-.562	-.455	-.214	-.263	-.628
Skewness	-.141	-.103	.020	.012	-.165	-.309	.300	.023	.267	.063	.008
Median	99.00	101.00	10.00	11.00	13.00	12.00	9.00	10.00	10.00	10.00	13.00
Französisch											
Mittelwert	86.17	94.38	9.43	8.72	10.62	10.46	6.77	9.04	9.33	9.32	10.35
N	69	69	69	69	69	69	69	69	69	69	69
Standardabw.	16.08	8.26	2.60	2.64	3.48	3.92	2.88	3.71	2.74	3.16	3.00
Kurtosis	.398	.411	-.521	-.203	-.529	-.779	.561	-.208	-.119	-.512	1.541
Skewness	.159	.152	-.289	.081	-.138	-.039	.675	.181	.591	.077	-.107
Median	86.00	94.00	10.00	9.00	11.00	11.00	6.00	9.00	9.00	9.00	10.00
Italienisch											
Mittelwert	87.34	94.90	9.83	8.45	11.52	9.86	6.31	10.03	8.28	8.48	11.59
N	29	29	29	29	29	29	29	29	29	29	29
Standardabw.	16.87	8.60	2.62	2.67	2.91	3.74	3.32	3.35	2.86	2.89	3.61
Kurtosis	1.226	1.199	-.181	-.629	.219	-1.017	.116	-.259	-.369	-.777	.273
Skewness	-.068	-.067	-.195	.159	-.108	-.131	.552	.059	.328	-.273	.963
Median	89.00	96.00	10.00	8.00	12.00	10.00	6.00	11.00	8.00	9.00	11.00
Total											
Mittelwert	97.14	99.99	10.18	10.45	12.92	11.60	8.74	10.09	9.95	10.20	12.70
N	801	801	801	801	801	801	801	801	801	801	801
Standardabw.	19.48	9.95	2.81	3.18	3.59	3.88	3.75	3.40	3.42	3.23	3.56
Kurtosis	.063	-.094	-.134	-.136	-.381	-.338	-.503	-.428	-.182	-.278	-.519
Skewness	-.074	-.042	.006	.061	-.154	-.275	.353	.022	.311	.063	.056
Median	97.00	100.00	10.00	10.00	13.00	12.00	8.00	10.00	10.00	10.00	13.00

Tabelle A 5. Verteilungskennwerte für Punktwert, Testwert und Untertests nach Sprachgruppen

	Humanmedizin		Veterinärmedizin	
	Punktwert	Testwert	Punktwert	Testwert
24	1			
37	1			
45			1	
46	2			
48	1			
50	1			
51			1	
52	2		1	
54	1		3	
55	2			
59	3		2	
60	2			
61	1		1	
62			3	
63	3		2	
64	3		2	
65	5		2	
66	1		2	
67	1		2	
68	4		1	
69	6		2	
70	7	2		
71	4		2	
72	5		2	
73	3		5	1
74	5	2	2	
75	4	1	1	
76	6	1	4	1
77	6	2	2	1
78	9	3	2	3
79	6		2	
80	12	3		2
81	9	3	2	1
82	7	3	4	5
83	3	3		2
84	8	6	4	4
85	15	5	8	3
86	15	13	6	2
87	13	9	2	4
88	11	8	5	7
89	12	10	1	5
90	8	15	4	4
91	9	18	2	2
92	15	16	3	6
93	11	11	8	4
94	14	30	1	14
95	10	24	2	7
96	19	20	2	5
97	13	24	7	5
98	14	25	3	9
99	14	29	4	4

Fortsetzung nächste Seite

	Humanmedizin		Veterinärmedizin	
	Punktwert	Testwert	Punktwert	Testwert
100	10	27	2	10
101	13	24	4	6
102	18	31	6	10
103	8	8	1	1
104	13	30	1	1
105	17	20		8
106	9	26	5	3
107	11	20	3	2
108	14	17	1	9
109	12	25	2	4
110	7	18		
111	13	16	2	1
112	9	12	4	2
113	8	7	5	2
114	12	10	3	
115	13	11	1	1
116	8	6		
117	10	15		2
118	9	6	1	
119	7	4		
120	6	5		
121	6	4	2	1
122	1	1		
123	6	3	2	
124	5			
125	5	3		
126	6		1	
127	5	1		
128	4			
129	2			
130	7	1	1	
131	8		1	
132	2			
133	4			
134	4			
136	4			
137	1			
138	3			
139	1		1	
140	1			
141	2			
142	1			
145	3			
150	1			
161	1			

Tabelle A 6. Punkt- und Testwertverteilung nach Disziplinen

Statistik der Testteilnehmer

	Human- medizin	Veterinär- medizin	Maturanden	19jährige Wohnbev.	HM % Matu- randen	HM % 19j. Wohnbev.	VM % Matu- randen	VM % 19j. Wohnbev.
OW	14	0	47	474	9.93	0.98	0	0
BS	103	4	374	1761	9.18	1.95	0.53	0.11
BL	139	10	566	2774	8.19	1.67	0.88	0.18
AI	5	1	23	187	7.25	0.89	2.17	0.27
LU	106	22	490	4267	7.21	0.83	2.24	0.26
TG	58	7	281	2706	6.88	0.71	1.25	0.13
ZH	492	57	2427	12926	6.76	1.27	1.17	0.22
BE	275	52	1376	10500	6.66	0.87	1.89	0.25
GR	57	10	289	2347	6.57	0.81	1.73	0.21
AG	190	31	1028	6449	6.16	0.98	1.51	0.24
FR	106	19	606	2964	5.83	1.19	1.57	0.32
AR	14	0	81	586	5.76	0.80	0	0
NW	11	0	68	413	5.39	0.89	0	0
SO	61	13	383	2763	5.31	0.74	1.70	0.24
GL	11	2	73	477	5.02	0.77	1.37	0.21
SG	166	14	1296	5620	4.27	0.98	0.54	0.12
SZ	53	4	426	1616	4.15	1.09	0.47	0.12
UR	6	0	49	481	4.08	0.42	0	0
SH	16	3	151	819	3.53	0.65	0.99	0.18
TI	74	12	805	3110	3.06	0.79	0.75	0.19
ZG	16	9	179	1197	2.98	0.45	2.51	0.38
VS	45	10	673	3464	2.23	0.43	0.74	0.14
NE	3	10	474	1968	0.21	0.05	1.05	0.25
JU	1	2	201	837	0.17	0.04	0.50	0.12
VD	5	26	1377	6850	0.12	0.02	0.94	0.19
GE	1	24	1340	4291	0.02	0.01	0.90	0.28
AUSLAND	32	4	122		8.74		1.64	

Tabelle A 7: Testteilnehmer 1998 bis 2000 und Prozentverhältnis zu Maturanden und 19jähriger Wohnbevölkerung. Basis für Maturandenzahl und 19jährige Wohnbevölkerung ist 1998 (St. Gallen und Schwyz Maturandenzahl 1997, weil 1998 Doppeljahrgänge der Maturanden vorhanden waren).