

EMS • Eignungstest für das Medizinstudium in der Schweiz 1999

Bericht über die Durchführung und Ergebnisse

avec un résumé en français

K.-D. Hänsgen und B. Spicher

Bericht 5 (1999)



**Zentrum für Testentwicklung
und Diagnostik** am Psychologischen Institut
der Universität Freiburg - Schweiz

Hänsgen, Klaus-Dieter; Spicher, Benjamin

EMS Eignungstest für das Medizinstudium in der Schweiz 1999

Bericht über die Durchführung und Ergebnisse

Berichte des Zentrums für Testentwicklung, Nummer 5

im Auftrag der Schweizerischen Hochschulkonferenz

Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik, 1999

unter Mitarbeit von Sabine Dobler, Sébastien Simonet,
Nicolas Burger, Steve Fürst, Eva Heim, Maurizio Strazzeri

Redaktion: Valérie Favre

© Zentrum für Testentwicklung und Diagnostik 1999
am Departement für Psychologie der Universität Freiburg
Route d'Englisberg 9, CH-1763 Granges-Paccot
Email: ztd@unifr.ch
Internet: <http://www.unifr.ch/ztd>

Druck: Cric-Print, Freiburg/Schweiz

Inhaltsverzeichnis

ZUSAMMENFASSUNG	5
RÉSUMÉ	10
EINLEITUNG.....	15
NUMERUS CLAUSUS UND MEDIZINSTUDIUM	18
ANMELDUNG ZU MEDIZINSTUDIUM UND TEST 1999.....	21
Anmeldungen, Testantritte und Zulassungen.....	21
Grösse der Testlokale	24
Testorte und Wunschuniversitäten.....	24
Testorte und Wohnkantone	26
TestteilnehmerInnen nach Alter und Geschlecht.....	28
Sprachgruppen nach Alter und Geschlecht	30
Alter und Maturitätsjahr	32
BESCHREIBUNG DES VERWENDETEN EIGNUNGSTESTS.....	33
Aufbau des Tests.....	33
Beispielaufgaben für die Untertests	35
1. Untertest: Muster zuordnen	35
2. Untertest: Medizinisch-naturwissenschaftliches Grundverständnis.....	36
3. Untertest: Schlauchfiguren	37
4. Untertest: Quantitative und formale Probleme.....	37
5. Untertest: Konzentriertes und sorgfältiges Arbeiten	38
6. Untertest: Figuren lernen	38
7. Untertest: Fakten lernen	39
8. Untertest: Textverständnis.....	39
9. Untertest: Diagramme und Tabellen.....	41
Berechnung der Werte	43
Vorhersage des Studienerfolges in Deutschland.....	44
Konsequenzen für eine Testevaluation in der Schweiz	46
Geringe Trainierbarkeit – ein Vorzug des EMS.....	47
TESTANWENDUNG IN DER SCHWEIZ 1998.....	49
Organisation der Testabnahme.....	49
Verteilungsprüfung	49
Äquivalenz der Testformen 1 und 2.....	52
Vergleich der Testfassungen 1998 und 1999	53

Äquivalenz der Sprachversionen	54
Sprachvergleich für die Untertests	58
Darstellung des Korrekturverfahrens	60
Vergleichbarkeit der Testlokale	79
Vergleich der Geschlechter	84
Vergleich der Kantone innerhalb der Deutschschweiz.....	92
Vergleiche für Altersgruppen und den Zeitpunkt des Maturitätsabschlusses.....	99
Vergleiche nach Wunschuniversitäten.....	104
ERGEBNISSE ZUR TESTGÜTE	106
Zuverlässigkeit	106
Binnenstruktur	107
Item-Trennschärfen	113
Konzentriertes und sorgfältiges Arbeiten.....	116
LITERATUR.....	118
Originaltests zur Information und Vorbereitung	119
Frühere Berichte des ZTD.....	120
ANHANG ZUM BERICHT 1999.....	123

Zusammenfassung

Die in diesem Bericht vorgestellten Ergebnisse beziehen sich auf Bewerberinnen und Bewerber für ein Studium der Human- und Veterinärmedizin in der Schweiz ab Wintersemester 1999/2000, für die ein Numerus clausus zur Anwendung kam. Alle Ergebnisse sind daher nicht repräsentativ für Schlussfolgerungen hinsichtlich einer gesamtartigen Evaluation von Bildungseinrichtungen, Bildungswegen oder von Regionen.

Die Teilnahme am Eignungstest für das Medizinstudium in der Schweiz (EMS) war 1999 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.

Statistik

Bei einer Kapazität der vier am NC beteiligten Universitäten von insgesamt 696 Studienplätzen (Humanmedizin: 546, Veterinärmedizin: 150) lagen der Schweizerischen Hochschulkonferenz (SHK) am 15.2.1999 1075 Anmeldungen (Humanmedizin: 845, Veterinärmedizin: 230) für beide Disziplinen an den genannten Universitäten vor.

Davon haben sich insgesamt 909 Personen bis zum Stichtag 18.5.1999 zum Test angemeldet. Zum Test angetreten sind 857 Personen - 52 Personen haben ihre Testanmeldung zurückgezogen bzw. sind nicht zum Test erschienen. Zwei Personen haben den Test nicht vollständig bearbeitet und vorzeitig beendet. Es haben somit 855 Personen den EMS erfolgreich absolviert (1998 waren es 750 – nur für das Fach Humanmedizin). Fünf weitere Personen haben den Test bereits im Vorjahr absolviert, sich ihr Testergebnis für 1999 anrechnen lassen und auf eine erneute Testteilnahme verzichtet.

In diesem Jahr zeigt sich eine dissuasive Wirkung des Tests: (1) Die Zahl der Anmeldungen zum Medizinstudium steigt wesentlich geringer an als diejenige anderer Studienrichtungen. Ein Anstieg wäre u.a. aufgrund der steigenden Maturandenzahlen zu erwarten gewesen; (2) die Zahl der Rückzüge verändert sich. 1999 betragen die Anmeldungen für Humanmedizin 155% im Verhältnis zur Studienplatz-Kapazität – 1998 waren es noch 186%. Stellt man die 860 Personen mit einem gültigen Testergebnis 1999 den 1075 Anmeldungen bei der SHK gegenüber, so sind in diesem Jahr aber nur 20% Rückzüge bis zum Testtermin erfolgt. Im Jahre 1998 waren dies noch 26% für Humanmedizin. Dies spricht auch für ein gründlicheres Überlegen bereits bei der Abgabe der Anmeldung zum Medizinstudium.

Insgesamt hat sich der Numerus clausus geringfügig verschärft: 1999 entspricht die Studienplatzkapazität 80,5% der Personen mit einem gültigen Testwert - 1998 waren dies noch 82,7%. Im Jahre 1998 wurden in der ersten Runde nur 613 von 620 Plätzen vergeben, weil die Berücksichtigung des nächstkleineren Testwertes zu einer Überschreitung der Kapazität geführt hätte. Aufgrund der Erfahrungen mit Rückzügen wurde in diesem Jahr die Kapazität in der ersten Runde der Zulassung um ca. 10% überschritten. Alle Personen, die zunächst nicht berücksichtigt werden konnten, bildeten eine Warteliste.

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

Organisation und Ablauf

Der Test fand am 2.7.1999 gleichzeitig an sieben Testorten in drei Sprachen (Deutsch: Basel, Bern, Chur, St. Gallen, Zürich; Französisch: Freiburg; Italienisch: Bellinzona) statt. Zu beachten ist, dass die Wahl des Testortes unabhängig vom zukünftigen Studienort möglich war und auch erfolgte - am häufigsten nach der Wohnortnähe. Die Testleiter¹ wurden vorher ausführlich geschult, viele hatten bereits Erfahrungen aus 1998.

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 – etwa durch irreguläre Abläufe oder gar Pannen in einzelnen Testlokalen - traten nicht auf. Die Rate von insgesamt nur einer Verwarnung (nicht schwerwiegend) und keinem Ausschluss vom Test ist sehr gering und spricht für einen sehr 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) erreichen nahezu dieselben Werte wie im Vorjahr und liegen im langjährig in Deutschland bei der Anwendung des Eignungstests ermittelten Bereich. Die etwas geringeren Reliabilitätswerte bei "Muster zuordnen" sind dabei tolerierbar und nicht auf Qualitätsunterschiede der Testvorlagen zurückzuführen.

Die Faktorenstruktur kann identisch und mit der gleichen Varianzaufteilung wie 1998 reproduziert werden. Sie entspricht ebenfalls den aus Deutschland bekannten Ergebnissen.

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

Die insgesamt gleichbleibend hohe Testgüte rechtfertigt u.a. auch eine Übertragbarkeit der Testwerte aus dem Vorjahr.

Gleichbehandlung der Sprachgruppen

In diesem Jahr wurde ein differenziertes Konzept zur Gleichbehandlung der Sprachgruppen angewendet, welches auf den Erfahrungen aus 1998 und international üblichen Vorgehensweisen (z.B. Israel, USA) basiert.

Es kann unterschieden werden ein weniger sprachabhängiger Teil des EMS (bis auf die vergleichsweise einfache Instruktion setzen die Aufgaben keine spezielle Sprachkompetenz voraus – es müssen keine Aufgaben übersetzt werden) und ein sprachabhängiger Teil (die Aufgabenlösung setzt Sprachkompetenz voraus – für die Übersetzung ergeben sich erhöhte Anforderungen, um die Aufgaben in eine vergleichbare Schwierigkeit zu übertragen).

Unterschiede zwischen den Sprachgruppen treten in beiden Bereichen etwa gleich stark auf, die italienischsprachige Gruppe ist in den sprachabhängigen Tests etwas besser als in den

¹ Im Interesse der Kürze und 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.

weniger sprachabhängigen Tests – jeweils verglichen mit der deutschsprachigen Gruppe. Diese Unterschiede lassen sich deshalb nicht auf sprachliche Bedingtheit (und damit auch nicht auf mögliche Übersetzungsunterschiede) reduzieren. Wegen des in Faktorenanalysen ermittelten bedeutsamen Generalfaktors würde man in etwa gleiche Leistungen in beiden Test-Teilen erwarten.

Um mögliche Schwierigkeitsunterschiede der Testaufgaben in den Sprachgruppen zu minimieren, kommt ein Ausgleichsverfahren zur Anwendung. Dies beruht auf der Überlegung, dass die Unterschiede aufgrund einer anderen Zusammensetzung der Bewerbergruppen (die französisch- und italienischsprachige Gruppe ist jeweils nicht repräsentativ für die Population aller Medizinbewerber der entsprechenden Sprachgruppe) und aufgrund von kulturellen Unterschieden nicht ausgeglichen werden, sondern nur solche, die aufgrund einer unterschiedlichen Schwierigkeit von Testaufgaben zustande kommen.

Von der SHK wurde bereits 1998 beschlossen, dass allfällige Ausgleichsverfahren (1) nur in sprachabhängigen Tests und (2) bei Vorliegen eines signifikanten Unterschiedes zwischen den Gruppen zur Anwendung kommen.

Die französisch- bzw. italienischsprachige Gruppe wurde jeweils zur deutschsprachigen Gruppe ins Verhältnis gesetzt und ein vierstufiges Vorgehen gewählt: (1) Identifikation signifikanter Mittelwertunterschiede für die fünf sprachabhängigen Tests, (2) Identifikation der Aufgaben, die sich in der Schwierigkeit gegen den allgemeinen Trend der übrigen Aufgaben unterschieden, (3) Ermittlung des Korrekturwertes als Schwierigkeitsunterschied der französisch- bzw. italienischsprachigen zur deutschsprachigen Gruppe für diese Aufgaben und (4) Gutschrift dieses Korrekturwertes für die Personen der französisch- bzw. italienischsprachigen Gruppe, wenn sie diese Aufgabe nicht gelöst haben.

Die Evaluation der Testwerte für den sprachabhängigen Teil (mit der Korrektur) und den weniger sprachabhängigen Teil des Tests zeigt, dass es in keinem Fall spezielle Benachteiligungen einer Gruppe im sprachlichen Bereich des Tests gibt. Leistungen in den sprachabhängigen Teilen sind im Mittel besser als in den weniger sprachabhängigen Teilen. Die in diesem Jahr vorhandenen Testwert- und Zulassungsunterschiede zwischen den Sprachgruppen sind daher mit grosser Wahrscheinlichkeit tatsächliche Unterschiede in der Eignung.

Wir vermuten, dass die Repräsentativität der Gruppe der Testteilnehmenden für die Medizinbewerber der jeweiligen Sprachgruppe der wichtigste Faktor zur Erklärung dieser Unterschiede ist: Besonders die Zusammensetzung der französischsprachigen Gruppe hat sich gegenüber 1998 offenbar verändert. Wegen der Verwendung identischer Einstreuaufgaben 1998 und 1999 war in diesem Jahr ein direkter Vergleich beider Jahrgänge durch einen Wertevergleich für diese Aufgaben möglich. Es zeigte sich für die deutsch- und die italienischsprachige Gruppe eine Leistungsverbesserung – für die französischsprachige dagegen eine Verschlechterung, die in dieser Deutlichkeit innerhalb eines Jahres wohl kaum auf die Wirkung kultureller Unterschiede zurückgeführt werden kann. Da die Aufgaben identisch waren, scheiden auch testbedingte Ursachen weitgehend aus.

Differenzierung nach Alter

Jüngere Bewerberinnen und Bewerber (Jahrgänge 1979-1982) haben wie im Vorjahr die besten Chancen bei der Zulassung. Es lässt sich erneut bestätigen, dass die Zulassungsquote innerhalb der Gruppe der Älteren (ab Jahrgang 1978) bei denen etwa gleich ist, welche die Mittelschulbildung sofort nach der Primar- bzw. Sekundarstufe aufgenommen und die Maturitätsprüfung zum frühestmöglichen Zeitpunkt abgelegt haben. Ihre Zulassungsquote

liegt nur 5% unter derjenigen für jüngere Bewerberinnen und Bewerber. Die Gruppe der Personen, welche die Maturitätsprüfung zu einem späteren Zeitpunkt abgelegt haben, weist eine geringere Testleistung und um rund 15% geringere Zulassungsquote als die der Jüngeren auf. Diese Testergebnisse entsprechen allerdings auch den Erfahrungen, die bezüglich des Erfolgs der genannten Personengruppen im Studium vorliegen.

Für die Älteren bringt eine späte Maturitätsprüfung und damit das "Noch-Trainiertsein" durch einen kürzlich abgelegten Schulabschluss keine Vorteile gegenüber den Personen, deren Schulbesuch zeitlich schon etwas zurück liegt. Dies ist ein erneuter Beleg, dass der Test nur sehr bedingt trainiert werden kann.

Differenzierung nach Geschlecht

Es wurde mehrfach festgestellt, dass in Deutschland wie auch in der Schweiz die Notenwerte in den Propädeutikumsprüfungen des Medizinstudiums für Männer geringfügig besser sind als für Frauen. Man könnte daher erwarten, dass der Testwert, sofern er die Studieneignung vorhersagen will, ebenfalls etwas bessere Leistungen für Männer als für Frauen zeigen würde.

Wie schon 1998 werden auch in diesem Jahr für den Testwert allerdings keine signifikanten Mittelwertunterschiede zwischen den Geschlechtern nachgewiesen. Die Zulassungsquoten unterscheiden sich ebenfalls nicht signifikant.

Die vorhandenen Trends differenzieren sich für Human- und Veterinärmedizin: Sind im Fach Humanmedizin (wie schon 1998) die Testwerte der Männer im Trend besser als die der Frauen, ist es bei der Veterinärmedizin genau umgekehrt. Spätestens dieses Ergebnis macht klar, dass Unterschiede nicht als generelle Leistungsunterschiede zwischen den Geschlechtern interpretierbar sind, sondern dass auch hier Repräsentativitätsprobleme in Bezug auf die Gesamtheit der Maturandinnen und Maturanden eine Rolle spielen (wer wählt Human-, wer Veterinärmedizin?). **Die Ergebnisse sind daher grundsätzlich nicht geeignet, Aussagen zu generellen Eignungsunterschieden für die Frauen und die Männer zu treffen.**

Differenzierung nach Maturitätsquote und Kanton

Wiederum zeigt eine Zuordnung der deutschsprachigen Teilnehmer in Kantonsgruppen nach der Maturitätsquote (gleiche Gruppen wie 1998) keine signifikanten Unterschiede – weder beim Testwert noch in einzelnen Untertests.

Ein Vergleich der Testwerte von Bewerberinnen und Bewerber nach ihrem Wohnsitz gemittelt für die sieben Kantone mit den grössten Teilnehmerzahlen (mehr als 38 Personen) zeigt in diesem Jahr erstmals signifikante Unterschiede. Es deutet sich an, dass sich die bereits 1998 im Trend vorhandenen Unterschiede verstärkt haben, wobei zwei Kantone am oberen und ein Kanton am anderen Ende der Rangreihe in diesem Jahr deutlichere Abweichungen vom Mittelwert 100 aufweisen, während die vier Kantone im mittleren Bereich sich nicht unterscheiden.

Besonders hieraus sollten keine vorschnellen Schlussfolgerungen für die Maturanden der Kantone gezogen werden, weil es sich nur um Bewerbungen für das Medizinstudium handelt. Es wäre beispielsweise auch zu analysieren, ob sich das "Studienwahlverhalten" für Medizin in den Kantonen unterscheidet – bzw. ob es auch durch den Test unterschiedlich geworden ist.

Zusammenfassender Vergleich zur Testanwendung 1998

Es wird für die Schweiz nachgewiesen, dass die Kennwerte und die Faktorenstruktur des Tests über zwei Jahre nahezu identisch geblieben sind. Deshalb kann davon ausgegangen werden, dass die langjährig hohe Testgüte, welche in Deutschland nachgewiesen wurde, auch für weitere Testversionen, die gegebenenfalls in der Schweiz eingesetzt würden, gilt.

Eine Stabilität des Niveaus der Testwerte zeigt sich bei der deutschsprachigen Gruppe – eine ausreichende Grösse und Repräsentativität für die Medizinbewerber der Deutschschweiz liessen dies erwarten. Die italienisch- und v.a. die französischsprachige Gruppe zeigen dagegen Veränderungen. Neben der geringen Stichprobengrösse scheinen dafür zeitabhängige Differenzierungen des Bewerbungsverhaltens mit verantwortlich zu sein (Anmeldung für eine am NC beteiligte oder nicht beteiligte Universität), die weiter beobachtet werden müssen.

Schlussfolgerungen

Die Zulassungsbeschränkung mittels Eignungstest hat sich auch 1999 bewährt. Der Test bleibt bezüglich der Machbarkeit, der wissenschaftlichen Absicherung und der Ökonomie als Kriterium für die Studienzulassung bei einem Numerus clausus weiterhin konkurrenzlos. Vor allem die in Bundesgerichtsurteilen zum Numerus clausus wiederholt geforderte Rechtsgleichheit und Willkürfreiheit scheinen für die Gesamtheit der Bewerberinnen und Bewerber unter den konföderalen Bedingungen der Schweiz letztlich nur durch den einheitlichen Einsatz des Tests realisierbar zu sein.

Eine zukünftige Weiterentwicklung des Tests – etwa durch Anpassung an durch die Studienreformen sich verändernde Anforderungen des Medizinstudiums – ist dabei nicht ausgeschlossen. Die im Jahre 2000 beginnende Evaluation wird die Frage beantworten helfen, ob auch in der Schweiz die Studieneignung mit gleich hoher Güte wie in Deutschland durch den Test vorhergesagt werden kann. Sind solche Weiterentwicklungen notwendig, sollte damit rechtzeitig begonnen werden. Ein Faktor für die hohe Qualität des Test ist die empirische Vorerprobung aller neuen Aufgaben. Dies wäre auch für den EMS unabdingbar.

Résumé

Les résultats présentés dans ce rapport se réfèrent aux candidats aux études de médecine humaine et vétérinaire en Suisse au semestre d'hiver 1999/2000 et auxquels un numerus clausus a été appliqué. Les résultats ne sont donc pas représentatifs d'une évaluation globale des établissements de formation, des voies d'études ou des régions et ne permettent pas de tirer de conclusions allant en ce sens.

En 1999, la participation au test d'aptitudes pour les études de médecine en Suisse (AMS) était obligatoire 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 à la **médecine dentaire**.

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 1075 inscriptions au 15.2.1999 (médecine humaine: 845, médecine vétérinaire: 230) pour les deux disciplines dans les universités mentionnées.

Parmi celles-ci, 909 personnes se sont inscrites au test jusqu'à l'échéance du 18.5.1999. 857 personnes se sont présentées au test, ce qui signifie que 52 personnes ont retiré leur inscription ou ne se sont pas présentées au test. Deux personnes n'ont pas effectué le test entièrement et l'ont donc terminé prématurément. Ainsi, 855 personnes ont accompli l'AMS avec un résultat valide (en 1998 elles étaient 750, uniquement pour la médecine humaine). En plus, cinq personnes ayant subi le test l'année précédente ont décidé de faire valider leur résultat en 1999, renonçant ainsi à une nouvelle participation au test.

Cette année, le test a eu un effet dissuasif: (1) le nombre d'inscriptions aux études de médecine a beaucoup moins augmenté que celui des autres voies d'études, alors qu'on aurait pu s'attendre à une augmentation, entre autres en raison du plus grand nombre de bacheliers; (2) le nombre de retraits a diminué. En 1999, les inscriptions en médecine humaine se situaient à 155% par rapport à la capacité des places d'études, alors que cette relation était de 186% en 1998. Si l'on compare les 860 personnes ayant un résultat valide en 1999 à l'ensemble des 1075 inscriptions auprès de la CUS, le taux de retraits jusqu'à l'échéance du test est de 20% seulement. En 1998, il s'élevait à 26% pour la médecine humaine. Cela pourrait également s'expliquer par une réflexion plus approfondie, au moment déjà de la remise de l'inscription pour des études de médecine.

Globalement, le numerus clausus s'est légèrement renforcé: en 1999, la capacité des places d'études correspond à 80,5% des personnes ayant un résultat de test valide, tandis qu'en 1998 cette relation s'élevait à 82,7%. En 1998, lors de la première série d'attributions, seules 613 des 620 places disponibles furent attribuées, car la prise en considération du résultat juste inférieur à la valeur retenue aurait débouché sur un dépassement de la capacité. Cette année, grâce à l'expérience des retraits acquise l'année précédente, la capacité lors de la première série d'attributions fut dépassée d'env. 10%. Toutes les personnes qui, dans un premier temps, n'ont pas pu être prises en considération, ont été mises sur une liste d'attente.

A la suite de retraits supplémentaires (par ex. à cause du lieu d'études attribué), des places d'études ont pu être octroyées aux personnes figurant sur la liste d'attente, dans l'ordre décroissant des résultats obtenus au test. Ainsi, suite à la deuxième série d'attributions, neuf pour cent des personnes ayant un résultat de test valide seulement sont restées sans offre de place d'études.

Organisation et déroulement

Le test a eu lieu simultanément le 2.7.1999 dans sept villes et en trois langues (en allemand : Bâle, Berne, Coire, St-Gall, Zurich; en français: Fribourg; en italien: Bellinzone). Il s'agit de noter que le choix du lieu de test était indépendant du futur lieu d'études ; ainsi les participants ont le plus souvent choisi le lieu de test en fonction de la proximité de leur domicile. Les chefs de test ont été au préalable soigneusement formés, la plupart d'entre eux ayant déjà participé au déroulement du test en 1998.

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 remplies et des discriminations durant le test – en raison d'imprévus ou même de pannes dans un local particulier – ne se sont pas produites. Le petit nombre d'avertissements (un seul et de faible gravité) et l'absence d'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 chefs de test, les suppléants et les surveillants, le déroulement du test peut être considéré comme étant un 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 (consistance, fiabilité selon la méthode "split-half") sont très proches des valeurs de l'an précédent et sont comparables à celles qui prévalent de longue date en Allemagne pour l'application du test. Les valeurs de fiabilité un peu moindre obtenues au sous-test "Reconnaissance de fragments de figures" restent tolérables et ne sont pas attribuables à des différences de qualité des modèles du test.

La structure factorielle est identique à celle obtenue en 1998 et comprend la même distribution de la variance. Elle correspond également aux résultats connus en Allemagne.

Les deux formes parallèles du test utilisées ne comportent pas de différences dans les résultats et peuvent être considérées comme étant équivalentes.

La qualité du test globalement élevée justifie aussi, entre autres, la transposition des résultats du test de l'année précédente.

L'égalité de traitement des groupes linguistiques

Cette année, un concept différencié fut appliqué afin d'assurer l'égalité de traitement des groupes linguistiques, basé sur l'expérience acquise en 1998 et les procédures internationales usuelles (par ex. Israël, Etats-Unis).

Une distinction peut être faite entre une partie de l'AMS peu dépendante de la langue (à l'exception des instructions simples, les problèmes n'exigent pas de compétence linguistique particulière – les problèmes ne nécessitent pas de traduction) et une partie dépendante de la langue (la résolution des problèmes nécessite des compétences linguistiques – le niveau de difficulté doit dans toute la mesure du possible être conservé lors de la traduction).

Des différences entre les groupes linguistiques apparaissent dans les deux domaines de manière comparable. Par rapport au groupe germanophone, le groupe linguistique italien se révèle même un peu meilleur dans la partie dépendante de la langue que dans celle peu dépendante de la langue. Par conséquent, les différences ne peuvent pas être réduites à la langue (et par là, à de possibles différences de traduction). Au regard du facteur général significatif déterminé par les analyses factorielles, on s'attendrait à des performances équivalentes dans les deux parties du test.

Afin de minimiser d'éventuelles différences de difficulté entre les groupes linguistiques, une procédure de compensation a été appliquée. Elle s'appuie sur les réflexions que tant les différences résultant d'une autre composition des groupes de candidats (les groupes linguistiques français et italien ne sont pas représentatifs de la population des candidats aux études de médecine de leur groupe linguistique respectif) que les différences culturelles ne doivent pas être compensées. Seules les différences dues à un niveau de difficulté inégal des problèmes du test doivent l'être.

La CUS avait décidé en 1998 déjà que les procédures de compensation nécessaires ne seront appliquées que (1) dans les tests dépendant de la langue et (2) en cas de différences significatives entre les groupes.

Les groupes linguistiques français et italien furent respectivement mis en rapport avec le groupe germanophone dans le cadre d'une procédure comprenant quatre niveaux: (1) identification des différences significatives de la moyenne pour les cinq tests dépendants de la langue, (2) identification des problèmes qui se distinguent par leur difficulté de la tendance générale des autres problèmes, (3) calcul pour ces problèmes de la valeur de correction correspondant à la différence de difficulté entre les groupes linguistiques français ou italien et le groupe germanophone et (4) crédit de la valeur de correction aux personnes francophones et de langue italienne qui n'ont pas résolu le problème en question.

L'évaluation des valeurs du test pour la partie dépendante de la langue (avec la correction) et pour la partie peu dépendante de la langue montre qu'il n'existe en aucun cas des discriminations quelconques pour un groupe concernant l'aspect linguistique du test. En moyenne, les performances dans les parties dépendantes de la langue sont supérieures aux parties peu dépendantes de la langue. Ainsi, cette année, les différences entre les groupes linguistiques au niveau des résultats obtenus au test et des admissions sont vraisemblablement dues à de véritables différences d'aptitude.

Nous supposons, que la représentativité des groupes linguistiques respectifs pour des études de médecine est le principal facteur explicatif des différences observées: en particulier, la composition du groupe francophone s'est apparemment modifiée par rapport à 1998. Grâce aux quelques questions d'essai identiques disséminées dans le texte en 1998 et 1999, une comparaison directe entre les deux versions a pu être établie. On constate une amélioration de la performance pour les groupes linguistiques alémanique et italien, tandis que le groupe francophone montre une baisse qui, de par son ampleur en l'espace d'une année, ne peut être attribuée à un effet culturel. Etant donné qu'il s'agissait de problèmes identiques, des causes inhérentes au test n'entrent pas non plus en ligne de compte.

Différenciation selon l'âge

Les candidats les plus jeunes (années de naissance 1979-1982) ont, comme en 1998, les meilleures chances d'admission. Il a à nouveau été confirmé que le quota d'admission de toutes les personnes plus âgées (années de naissance 1978 et inférieures) équivaut à celui des plus jeunes, pour autant qu'ils aient suivi des études secondaires directement après l'école obligatoire et qu'ils aient passé leur examen de maturité au plus tôt. Leur quota

d'admission ne se situe que 5% au-dessous de celui des candidats les plus jeunes. Le groupe des personnes ayant passé leur examen de maturité ultérieurement montre une performance moindre et présente un quota d'admission inférieur de près de 15% à celui des plus jeunes. Toutefois, pour ce groupe de personnes, ces résultats de test correspondent aux connaissances relatives à leur réussite dans les études.

Parmi les personnes plus âgées, le fait d'avoir passé les examens de maturité plus tardivement et d'être ainsi "encore entraînées" par l'obtention récente de la maturité ne présente pas d'avantages par rapport aux personnes qui ont obtenu depuis un certain temps leur maturité. Cela illustre une nouvelle fois les limites des possibilités d'entraînement au test.

Différenciation selon le sexe

On a constaté à plusieurs reprises, tant en Allemagne qu'en Suisse, que les notes obtenues par les hommes lors des examens propédeutiques de médecine étaient légèrement meilleures que celles des femmes. On pourrait donc s'attendre à ce que les résultats du test, dans la mesure où celui-ci veut prédire l'aptitude aux études, montrent également des performances quelques peu meilleures pour les hommes que pour les femmes.

Toutefois, comme en 1998, les résultats obtenus au test cette année ne montrent pas de différences significatives de la moyenne entre les sexes. Les quotas d'admission ne se différencient pas non plus de manière significative.

Les tendances existantes se différencient pour la médecine humaine et vétérinaire : tandis qu'en médecine humaine (comme déjà en 1998) le résultat des hommes est en moyenne meilleur que celui des femmes, c'est exactement l'inverse qui se produit en médecine vétérinaire. Ces faits montrent clairement que des différences ne peuvent pas être interprétées comme étant des différences d'aptitude générales entre les sexes, mais, qu'ici aussi, ce sont des problèmes de représentativité en référence avec l'ensemble des bacheliers qui entrent en considération (qui choisit la médecine humaine, resp. la médecine vétérinaire?). **Manifestement, les résultats ne permettent pas de tirer conclusions concernant des différences d'aptitude générales entre les femmes et les hommes.**

Différenciation selon le taux de maturités et les cantons

Un regroupement des cantons alémaniques selon leur taux de maturités (mêmes groupes qu'en 1998) ne révèle à nouveau pas de différences significatives – tant au niveau du résultat obtenu au test qu'au niveau des différents sous-tests.

Une comparaison des moyennes des résultats au test par les candidats selon leur domicile, pour les sept cantons ayant le plus grand nombre de participants (plus de 38 personnes), montre cette année pour la première fois des différences significatives. Les tendances esquissées en 1998 se sont accentuées: les deux cantons les plus haut classés ainsi que celui le plus bas classé montrent cette année des écarts plus marqués à la valeur moyenne de 100, tandis que les quatre cantons intermédiaires ne se différencient pas.

Il s'agit de ne pas tirer de conclusions prématurées à propos des bacheliers de ces cantons, car cela concerne uniquement les candidatures pour les études de médecine. Il faudrait par exemple également analyser si la proportion des personnes qui choisissent les études de médecine se différencie d'un canton à l'autre – ou si la proportion a changé à cause du test.

Brève comparaison avec l'application du test en 1998

Il a été démontré pour la Suisse que les indices statistiques de référence et la structure factorielle du test sont restés pratiquement identiques au cours des deux années. On peut dès lors penser que la fiabilité du test qui prévaut de longue date en Allemagne serait également valable pour de futures versions du test auxquelles la Suisse pourrait avoir recours à l'avenir.

On constate une stabilité des résultats obtenus au test par le groupe germanophone – le nombre suffisant de candidats suisses alémaniques ainsi que leur représentativité permettaient de s'attendre à ce résultat. Le groupe de langue italienne et plus particulièrement le groupe francophone ont en revanche montré des changements. La faible taille de l'échantillon, associée à un comportement de candidature différencié au cours du temps (s'inscrire à une université participant ou non au NC), semblent en être responsables et doivent continuer à être observés.

Conclusions

La limitation de l'admission au moyen du test d'aptitudes a également fait ses preuves en 1999. En ce qui concerne la faisabilité, le fondement scientifique et les aspects économiques, le test reste sans concurrence en tant que critère d'admission aux études dans le cadre d'un *numerus clausus*. En particulier l'égalité des droits et l'interdiction de l'arbitraire exigées à plusieurs reprises dans les jugements du Tribunal fédéral à propos du NC semblent pouvoir être réalisables, dans les conditions fédéralistes de la Suisse, pour l'ensemble des candidats uniquement par l'application uniforme du test.

Un développement futur du test – en l'adaptant par exemple aux changements d'exigences des études de médecine à la suite de réformes – n'est pas à exclure. L'évaluation qui débutera en 2000 permettra de déterminer si le test peut prédire l'aptitude aux études en Suisse avec la même fiabilité qu'en Allemagne. Si un développement du test devait s'avérer nécessaire, il s'agira de s'y prendre à temps. Un gage de qualité du test réside dans la mise à l'épreuve empirique de tous les nouveaux problèmes. Cela serait également indispensable pour l'AMS.

Einleitung

Zum zweiten Mal kam 1999 in der Schweiz der Eignungstest für das Medizinstudium (EMS) zum Einsatz. Die medizinischen Fächer mit Numerus clausus (NC) wurden um die Veterinärmedizin erweitert.

Der Test zeigt eine dissuasive Wirkung: Die Anmeldezahl hat gegenüber der Erwartung bei einer Entwicklung ohne Numerus clausus um schätzungsweise 25 bis 30% abgenommen. Dies kommt allerdings nicht unerwartet. In Deutschland hat es z.B. nach der Umstellung der Hochschulzulassung in der Medizin auf das allgemeine NC-Auswahlverfahren und die damit verbundene Abschaffung des Tests eine Zunahme der Bewerbungen um 25% gegeben. Es ist wahrscheinlich, dass wegen des Tests ungefähr dieser Anteil von einer Bewerbung für ein Medizinstudium in den Vorjahren Abstand genommen hat.

Man muss nun genau beobachten, ob sich die „richtigen“ Bewerber dem Test stellen. Die Ergebnisse sprechen – im Vergleich zu 1998 – auch für Änderungen des Studienwahlverhaltens. Diese Veränderungen differieren allerdings zwischen den Sprachgruppen und innerhalb der Deutschschweiz. Die Risikobereitschaft, sich dem Test zu unterziehen, die Anstrengungsbereitschaft, sich auf den Test optimal vorzubereiten, eine gesunde Portion Selbstsicherheit und die Fähigkeit, sich dabei selbst zu motivieren, sind Werte, die auch recht gut zum Berufsbild der Medizin passen. Grundsätzlich scheint es also nicht schlecht zu sein, wenn solche Anforderungen indirekt beim EMS mit vorausgesetzt werden. Der Test misst die Studieneignung – direkt bestimmte kognitive Fähigkeiten und indirekt auch die beschriebenen Persönlichkeitsqualitäten. Die Dauer des Tests sollte auch an die Belastbarkeit der Kandidaten Anforderungen stellen – eine Eigenschaft, die im späteren Berufsbild einen ausserordentlich hohen Stellenwert hat.

Motivation kann allerdings auch „von aussen“ beeinflusst werden – man kann ermutigen oder Resignation fördern. Es sollte daher geprüft werden, ob unterschiedliche Einflüsse und Diskussionen auch zu regional unterschiedlicher Dissuasion beigetragen haben könnten. Ein Vorzug des EMS ist in diesem Zusammenhang, dass regionale Unterschiede zwischen den schulischen Anforderungen sehr wenig Einfluss haben, weil Schulwissen nicht vorausgesetzt wird und auch das Training durch die Schule eine geringere Rolle spielt als angenommen (siehe die Befunde zur Altersspezifität). Die grosse Chance des EMS ist es, eine wirklich vergleichbare und von der tatsächlichen Studieneignung geleitete Zulassung zu erreichen, die nicht oder wenig von den regionalen Variationen der Bildungswege oder Bildungseinrichtungen abhängt.

Wiederholt wurde auf die fehlende Prüfung sozial-kommunikativer Fähigkeiten bei der Eignungsabklärung zum Medizinstudium hingewiesen. Die für solche Überprüfungen verwendbaren Methoden und im europäischen Ausland tatsächlich verwendeten Methoden sind allerdings in der Regel nicht mit gleicher Testgüte durchführbar wie der jetzt verwendete Test und praktisch nicht bezahlbar, wenn sie für alle Kandidatinnen und Kandidaten angewendet werden sollten. So erfordern Interviews bzw. Eignungsgespräche ein standardisiertes Beurteilungssystem, ein Training der Beurteiler und mindestens drei Beurteiler pro Kandidat. Eine Dauer von 30 Minuten ist dabei ein absolutes Minimum. Die Kosten für 1000 Kandidaten wären entsprechend sehr hoch.

Die Notwendigkeit des NC für die Medizin findet in den öffentlich geführten Diskussionen vermehrt Anerkennung. Zudem ist die Rechtslage bezüglich des NC und der zu

verwendenden Kriterien eindeutig: Das Bundesgericht hat in seinen jüngsten Urteilen¹ den NC grundsätzlich gutgeheissen – aber gleichzeitig zwei Forderungen an das Zulassungsverfahren gestellt, die ausserordentlich bedeutsam scheinen: Es müssen Rechtsgleichheit und Willkürfreiheit gewährleistet sein². Diese beiden Bedingungen sind somit die wohl wichtigsten Evaluationskriterien für alle Zulassungsverfahren. Willkürfreiheit muss zum wissenschaftlichen Begriff der Objektivität in Beziehung gesetzt werden. Durchführung, Auswertung und die Bewertung der Ergebnisse müssen frei von Subjektivität sein. Man muss sich fragen, ob beispielsweise Maturitätsnoten dieses Kriterium erfüllen würden, weil die Notengebung stark variiert und die Ergebnisse sachlich und numerisch folglich kaum vergleichbar sind. Zudem würden die bereits erwähnten Eignungsgespräche hohe Anforderungen an die Definition von übereinstimmenden Beurteilungskriterien stellen und ein Training der Beurteiler auch aus diesen Gründen notwendig machen.

Die Anwendung des Tests ist im Jahre 1999 ohne grosses Medienecho erfolgt. Waren im ersten Jahr Berichterstattungen mit Reportern vor Ort die Regel, erwähnte 1999 nur ein Teil der Medien den Test überhaupt noch in einer Kurzmeldung. Die detaillierte Information über alle Testauswertungen in den Berichten des ZTD wurde gut aufgenommen. In den Medien dominiert zunehmend eine sachliche und der Wirklichkeit angemessene Berichterstattung, die auf eine Wiederholung der pauschal vorgebrachten und zumeist nicht zutreffenden Kritikpunkte verzichtet, die sich auf Benachteiligungen von Gruppen bezogen. Wir möchten besonders diese Tendenz durch die vorliegende Darstellung der Ergebnisse der zweiten EMS in der Schweiz verstärken.

Wir sind uns der Tatsache bewusst, dass die Ergebnisse des Tests auch dazu verleiten können, generalisierte Urteile über die Maturanden bestimmter Bildungseinrichtungen, Bildungswege, Kantone oder Sprachgruppen zu bilden. Die wohl wichtigste Einschränkung ist, dass es sich „nur“ um Bewerbungen für das Fach Medizin handelt. Diese sind nicht repräsentativ für alle Maturanden – es können regional durchaus unterschiedliche Kriterien für die Studienfachwahl gelten, die sich auch in den Resultaten der Studieneignung für die Medizin niederschlagen. Dies gilt besonders für die Sprachgruppen. Auch Dissuasionseffekte können unterschiedlich wirken. Alle Ergebnisse – besonders die zu den Vergleichen der Sprachgruppen – legen nahe, dass es sich um sehr komplexe Sachverhalte handelt. Sie dürfen daher auch nicht auf einzelne „erwünschte“ Faktoren reduziert werden.

Der Bedarf an vergleichender empirischer Bildungsforschung und an Untersuchungen über die Rolle der Maturität innerhalb der Schweiz kann jedoch durch die Befunde des EMS nicht abgedeckt werden. Wir haben in diesem Jahr deshalb im Bericht wiederholt vermerkt, warum welche Schlüsse aus den Ergebnissen nicht gezogen werden können und Begründungen dafür angeben.

Die Aussagekraft des EMS ist spezifisch. Die darin zu prüfenden Funktionen wurden bekanntlich nicht willkürlich ausgewählt, sondern sind Ergebnis einer vergleichsweise aufwendig betriebenen Anforderungsanalyse für das Medizinstudium. Die in anderen Ländern eingesetzten Tests führten zu ähnlichen Ergebnissen. Verfahren zur Feststellung der Eignung für eine Tätigkeit werden ihre Methoden immer von den Anforderungen für diese Tätigkeit her definieren müssen, für die ein erfolgreiches Abschneiden vorhergesagt

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

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

werden soll. Deshalb kann sich der Eignungsbegriff beim EMS nur über die (prognostizierte) Studierfähigkeit definieren.

Im Übrigen bleibt entsprechendes Wissen weiterhin notwendig und wird in der Schweiz für die Studienzulassung durch die abgelegte Maturitätsprüfung vorausgesetzt (in anderen Ländern werden zusätzliche Wissenstests durchgeführt).

Dieser Bericht wendet sich an unterschiedliche Leserkreise, weshalb wir uns bemüht haben, die Darstellung auf die verschiedenen Interessentengruppen auszurichten. Dennoch lässt es sich auch in diesem Jahr nicht vermeiden, statistische Prüfgrößen und -verfahren mit anzuführen. Die Abschnitte beginnen jeweils mit Darstellungen, die für einen breiteren Leserkreis gedacht sind. Die Zahlenbefunde und statistischen Prüfungen folgen anschliessend im entsprechenden Kapitel. Der Abschnitt zur Testbeschreibung wurde gegenüber dem Bericht 4 nur wenig verändert. Wir erachten es aber für notwendig, dass man auch in diesem Bericht leicht nachschlagen kann, welche Aufgabentypen sich hinter den jeweiligen Untertests verbergen.

Im Internet (<http://www.unifr.ch/ztd>) sind ergänzende Informationen und alle weiteren Berichte des ZTD abrufbar. Um eine hohe Transparenz zu erreichen, bleiben diese Informationen dauerhaft verfügbar – was natürlich auch zu einem beträchtlichen Anstieg der verfügbaren Informationsmenge führt. Deshalb sollte beachtet werden, dass das Studium dieser Berichte zur wissenschaftlichen Begleitung des EMS keine Voraussetzung für eine erfolgreiche Teilnahme am Test ist. Materialien, die der Testvorbereitung dienen, wurden deshalb besonders gekennzeichnet.

Es verbleibt, allen Personen und Institutionen wiederum ganz herzlich zu danken, die zum Gelingen des Eignungstests im Jahre 1999 beigetragen haben.

Numerus clausus und Medizinstudium

Die Zahl der Anmeldungen bei der Schweizerischen Hochschulkonferenz zum Medizinstudium hat sich auf den ersten Blick nur unwesentlich verändert (Abbildung 1). Vor allem in Zürich, wo die Veränderungen seit drei Jahren marginal sind, bleibt die Überschreitung der Kapazität deutlich vorhanden. Basel ist zum Niveau zurückgekehrt, welches in den letzten drei Jahren vor 1998 vorhanden war, in Bern hat sich das Niveau der letzten vier Jahre gehalten. Hier ist allerdings zu beachten, dass für das Fach Veterinärmedizin 1999 in den am NC beteiligten Universitäten eine Aufnahme des Studiums bereits ab dem ersten Studienjahr in Bern oder Zürich erforderlich ist.

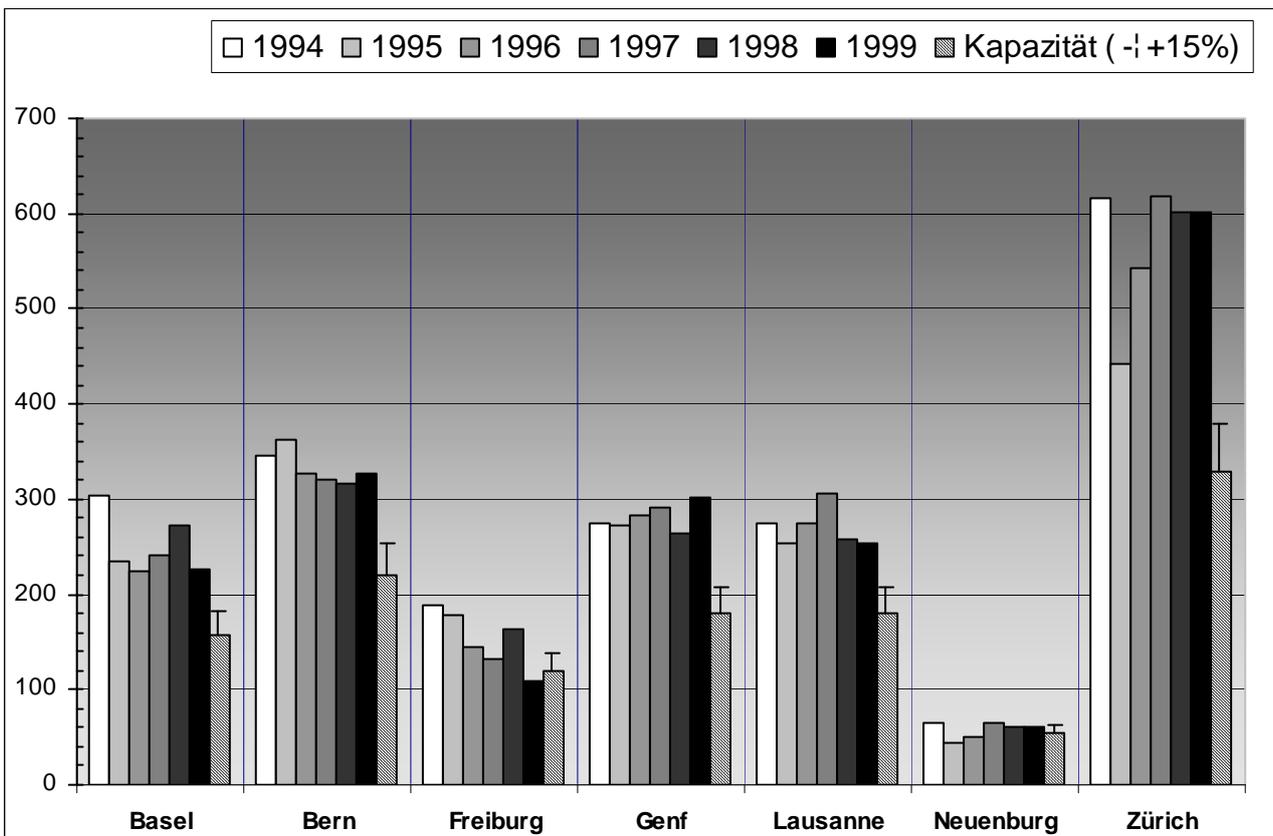
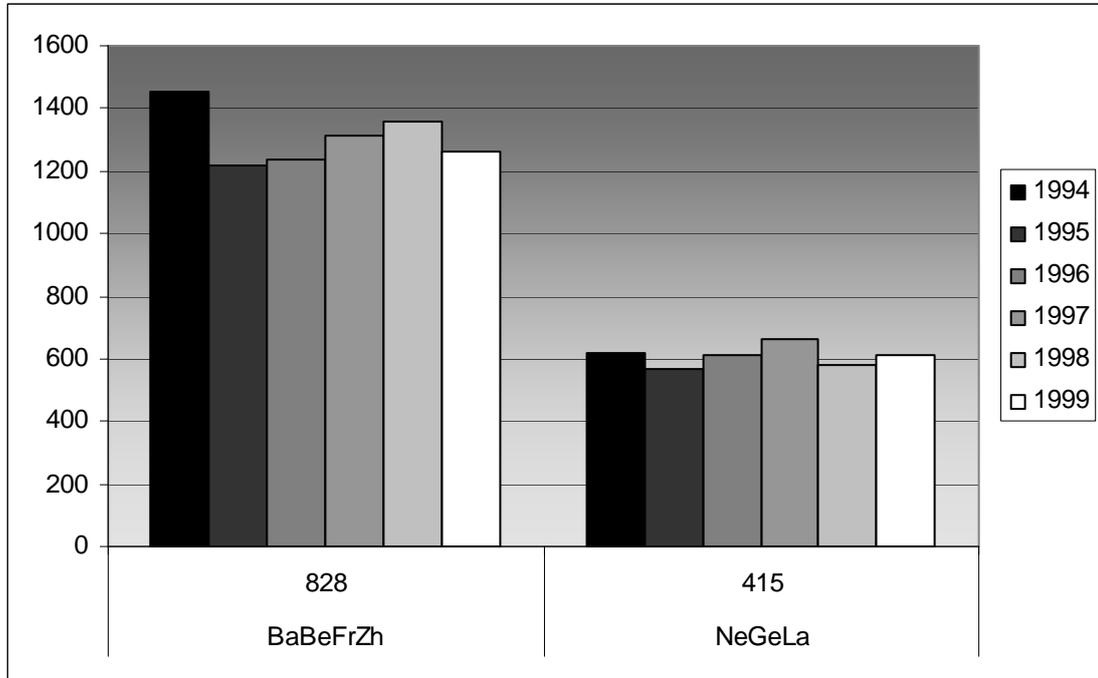


Abbildung 1. Anmeldungen zum Medizinstudium 1994 bis 1999, Anfängerprognose und Kapazitäten 1999 (letzte Säule mit Markierung des 15%-Überschreitungskriteriums als Grenze für die Notwendigkeit einer Zulassungsbeschränkung).

In Freiburg hat die Nachfrage in diesem Jahr wieder abgenommen – der Trend der Jahre 1995 bis 1997 setzt sich fort und die „Umkehr“ des Trends von 1998 konnte nicht stabilisiert werden. Die Nachfrage in Genf nahm wieder auf das Niveau von 1997 zu - Lausanne hat das Vorjahresniveau gehalten. In Neuenburg bleibt die Nachfrage auf niedrigerem Niveau konstant. Bis auf Freiburg und Neuenburg haben alle Universitäten das Kriterium „Nachfrage grösser als 115% der Kapazität“ deutlich überschritten – dies gilt auch für die Gesamtzahl aller Kandidaten der am NC beteiligten Universitäten.

Die „Zweiteilung“ der Universitäten in der Schweiz bezüglich der Anwendung eines NC hat nicht zu einer deutlichen Verschiebung der Bewerbungen zu Universitäten ohne Anwendung eines NC geführt (Abbildung 2). Es wäre denkbar gewesen, dass die deutschsprachigen Kandidaten auf die Universitäten „ohne NC“ ausweichen. Vor drei

Jahren – ohne Anwendung des NC – hat es aber bereits einmal das gleiche Verhältnis wie 1999 gegeben. Offenbar wird im Wechsel „über die Sprachgrenze“ keine Alternative gesehen, um dadurch die Testteilnahme zu umgehen. Bei der Bewertung des Rückganges in Freiburg könnten solche Verschiebungen allerdings als Mitursache ins Auge gefasst werden. Zur Bewertung der Situation in Freiburg vergleiche man z.B. den Beitrag in der



NZZ von Celio (29.9.1999).

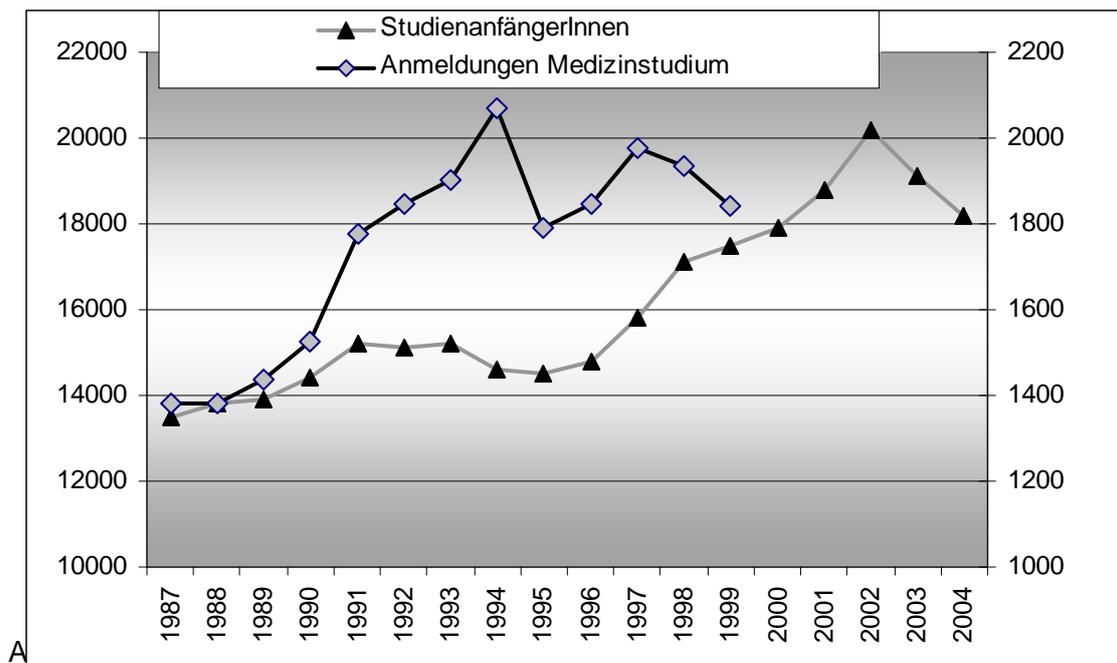


Abbildung 3. Linke Y-Achse: Prognose Studienanfänger 1987 bis 2004 (aus: "Wie viele Studierende an Schweizerischen Hochschulen?" Bildung und Wissenschaft, Heft 15). Rechte Y-Achse: Anmeldungen zum Medizinstudium bei der Schweizerischen Hochschulkonferenz.

Eine spannende Frage bleibt die nach der Prognose für die nächsten Jahre. Seit zwei Jahren – seit der Anwendung eines Tests – nimmt die Zahl der Anmeldungen für ein Medizinstudium wieder etwas ab – gegen den Trend der steigenden Zahl von Studienanfängern insgesamt. Das tiefere Niveau von 1995 ist noch nicht wieder erreicht, wo durch die Ankündigung eines Tests und diverse Aufklärung über den Medizinerberuf ein starker Dissuasionseffekt wirksam war.

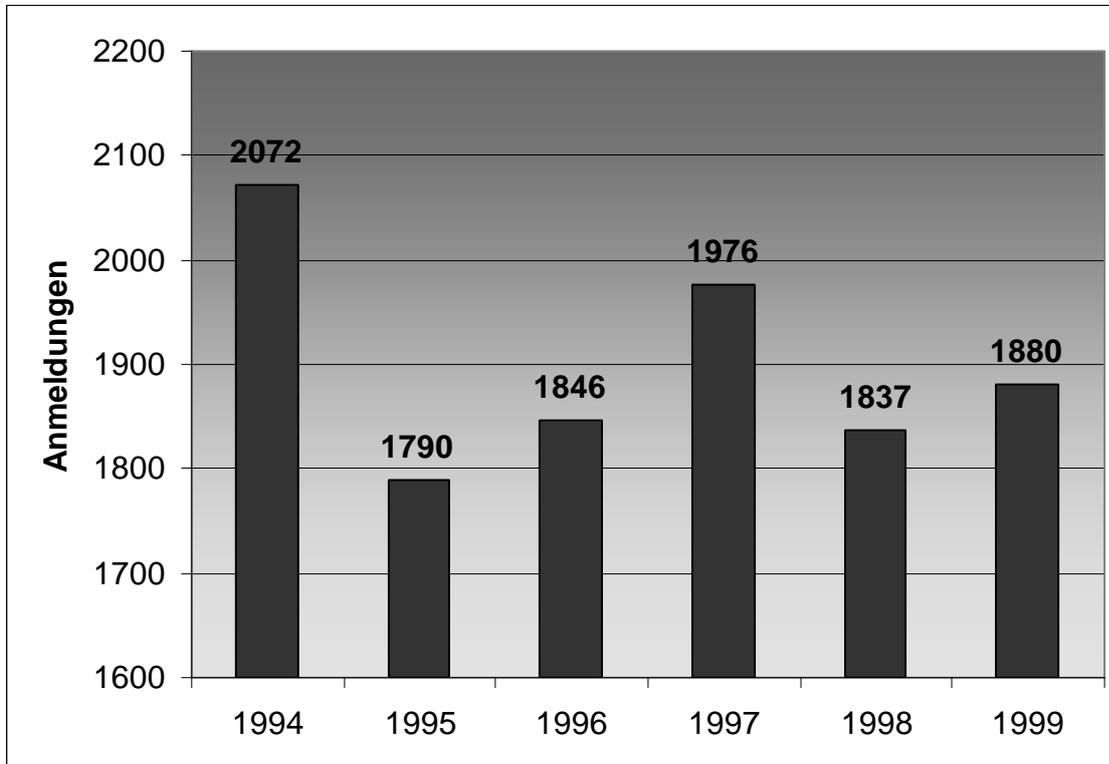


Abbildung 4. Bereinigte Anmeldezahlen zum Medizinstudium bei der SHK (1998: 100 zusätzliche Anmeldungen wegen Doppeljahrgängen bei Maturität abgerechnet).

Dieser Rückgang kommt nicht unerwartet – bekanntlich ist die Zahl der Bewerbungen in Deutschland im ersten Jahr ohne Einsatz des TMS-Eignungstests um 25% angestiegen.

Bei den Zahlenvergleichen ist in Rechnung zu stellen, dass infolge der Verkürzung der gymnasialen Ausbildung von 13 auf 12 Jahre, die im Zeitraum 1998–2003 kantonsweise wirksam wird, „Doppeljahrgänge“ entstehen. Nach neuesten Dokumenten (Dokument 71299 der Schweizerischen Hochschulkonferenz) sind bereits 1998 780 zusätzliche Maturitäten (Schwyz und St. Gallen) abgelegt worden. 1990 bis 1995 wurden jährlich durchschnittlich 12590 Maturitätszeugnisse erworben – die Bewerberzahl für Medizin mit 1800 weist auf eine Rate von 14% hin. Dies wären rein rechnerisch rund 100 zusätzliche Bewerber für Medizin in 1998 gewesen. Die „bereinigten“ Zahlen wären dann die in der Abbildung 4 dargestellten: Der „Dissuasionseffekt“ durch den Test 1998 wäre dann wie erwartet deutlich (und wurde durch die zusätzlichen Maturanden verdeckt). In 1999 – ohne Doppeljahrgänge – würde sich der erwartete Anstieg aufgrund der allgemeinen Zunahme der Maturandenquote zeigen.

Mehrere Doppeljahrgänge folgen noch - 2000 werden es 60 sein und im Jahr 2001 1810, im Jahr 2002 3550 und im Jahr 2003 670 sein – bei der Bewertung der erwarteten Schwankungen in den Folgejahren ist dies zu berücksichtigen.

Anmeldung zu Medizinstudium und Test 1999

Anmeldungen, Testantritte und Zulassungen

Im Jahr 2 der Testanwendung in der Schweiz zeigen sich Veränderungen; bereits die Anmeldung zum Medizinstudium wird genauer überlegt. Nach der Anmeldung bei der SHK im Februar 1999 überschreitet die Bewerberzahl die Studienplatzkapazität gegenüber 1998 um 30% weniger – dies sogar trotz einer Reduktion der Kapazitäten. Im weiteren Verlauf treten aber gegenüber 1998 7% weniger Rückzüge auf. Die Angemeldeten sind motivierter, ein Studium der Medizin auch beginnen zu wollen. Im Endeffekt bewerben sich mehr Kandidaten (mit gültigem Testwert) um einen Studienplatz als 1998, was durch den „NC-Bewerbungsindex“ (Tabelle 1) ausgedrückt wird. Dieser Index soll in den Folgejahren zu Vergleichszwecken herangezogen werden. Es wird auch deutlich, dass es sich in der Schweiz im internationalen Massstab um einen milden NC handelt.

	Human- medizin	Veterinär- medizin	Total
Anfängerkapazitäten	546	150	696
Anmeldung bei SHK am 15.2.99 (% im Verhältnis zur Kapazität)	845 (155%) (1998 = 1020 bzw. 186%)	230 (153%)	1075 (154%)
Anmeldung zum Test am 18.5.99	717 + 5*	194	911 + 5*
Erfolgreiche Absolvierung Test	673 + 5*	182	855 + 5*
Rückzugsquote zwischen Anmeldung bei SHK und Testabsolvierung	19,8% (1998: 26,5%)	20,8%	20,0%
Prozent der Personen mit gültigem Testwert, denen ein Studienplatz zugewiesen werden kann	80,5% (1998: 82,7%)	82,4%	80,9%
„NC-Bewerbungsindex“ (Bewerber mit gültigem Testwert pro Studienplatz)	1,24 : 1 (1998: 1,21 : 1)	1,21 : 1	1,23 : 1

Tabelle 1. Disziplinspezifische Statistiken für die am NC beteiligten Universitäten Basel, Bern, Freiburg und Zürich (*: 5 Personen mit gültigem Testwert aus 1998). Drei Personen, die am Test teilgenommen haben, benötigen aktuell keinen Studienplatz an einer der am NC beteiligten Universitäten.

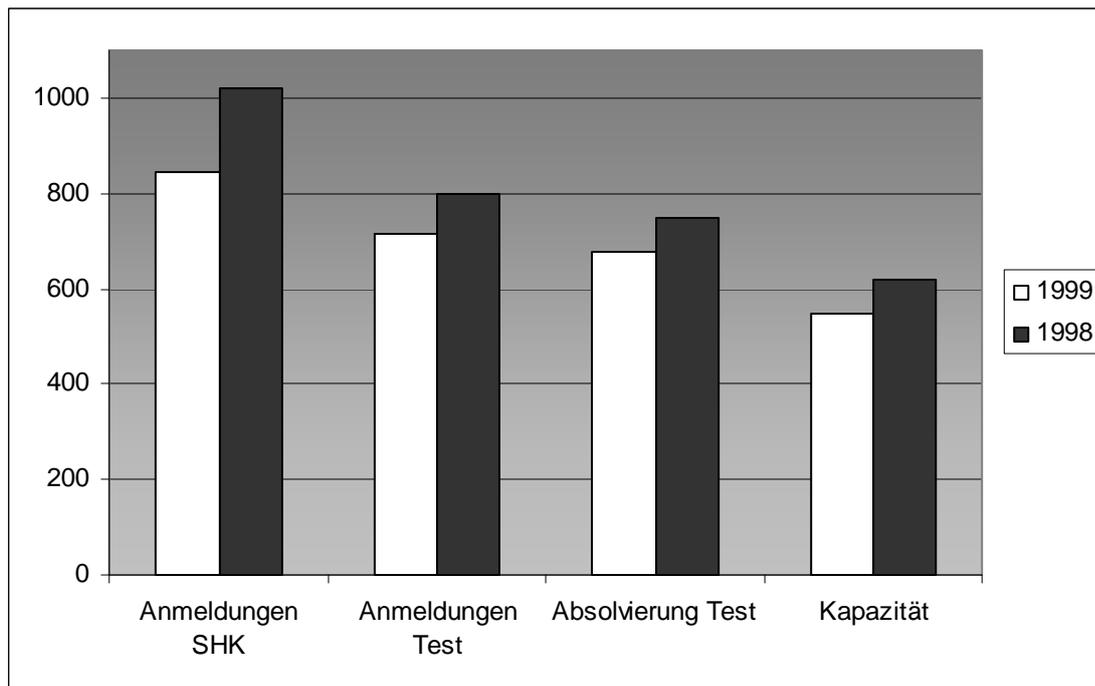


Abbildung 5. Anmeldeverlauf für Humanmedizin 1998 und 1999 im Vergleich.

Die Abbildung 5 verdeutlicht die Unterschiede zwischen 1998 und 1999: Diese werden im Verlauf des Bewerbungsverfahrens (erste drei Säulenpaare) immer geringer – die Differenz der Testabsolvierungen ist sogar etwas geringer als die Differenz der Kapazitäten. Bei allen Planungen des Rückzugsverhaltens müssen diese neuen Verhältnisse berücksichtigt werden.

Disziplin	Testteilnehmer	In 2 Zuteilungsrunden konnten Studienplätze angeboten werden		Keine Studienplätze konnten angeboten werden für	
		Absolut	In Prozent	Absolut	In Prozent
HM	673	617	92 %	56	8 %
VM	182	160	88 %	22	12 %
Total	855	777	91 %	78	9 %

Tabelle 2. Ergebnis der zwei Runden des Zulassungsverfahrens. Humanmedizin (HM), Veterinärmedizin (VM).

Die Studienplätze wurden 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 angebotenen 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. Letztendlich mussten so nur 9% der Personen effektiv abgewiesen werden.

Die Abbildung 6 zeigt, dass das Anmeldeverhalten für Human- und Veterinärmedizin vergleichbar ist. „Testergebnis nicht ausreichend“ bedeutet hier den Personenkreis, der vor der ersten Runde der Verteilung der Studienplätze die Kapazität überschreitet. In der Praxis wurde bereits in der ersten Runde ein möglicher Rückzug eingerechnet und es erhielten 10% mehr Personen, als die Kapazität betrug, einen Studienplatz angeboten. 1998 wurden relativ zur Anmeldezahl bei der SHK weniger Personen zugelassen – dennoch war der NC „milder“ (relativ mehr Rückzüge – relativ weniger Personen mit Testwert auf der Warteliste).

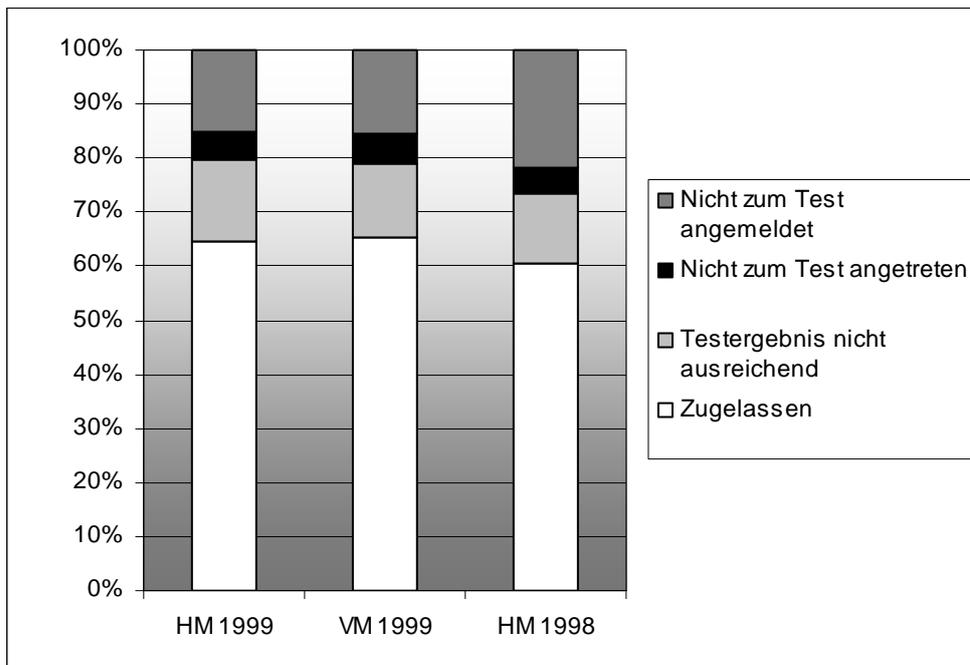


Abbildung 6. Kapazität, Anmeldungen und Testteilnahme für Human- (HM) und Veterinärmedizin (VM). Aufteilung der Voranmeldungen vom 2.1998 und 1999.

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

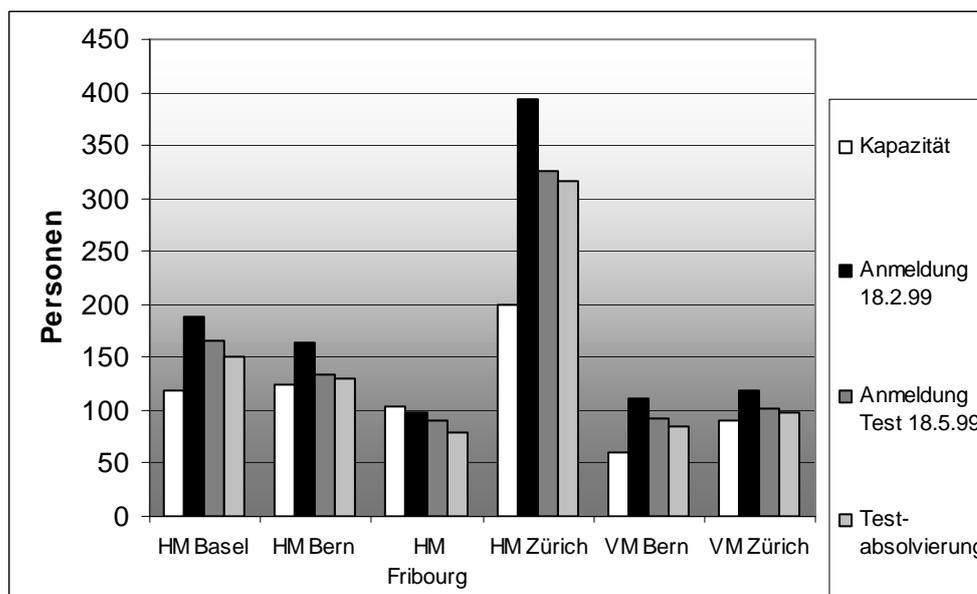


Abbildung 7. Kapazität, Anmeldungen und Testteilnahme für Fächer und Universitäten.

Grösse der Testlokale

In diesem Jahr wurde der Test in den gleichen Orten wie 1998 angeboten – in Freiburg in französischer Sprache, in Bellinzona in italienischer Sprache und in Basel, Bern, Chur, St. Gallen und Zürich in deutscher Sprache. Es wurden fast immer die gleichen Testlokalitäten wie 1998 verwendet. In Basel reduzierte sich die Zahl der Testlokale von vier auf zwei. Entsprechend der Beobachtungen des Vorjahres kann der Test auch in grösseren Lokalitäten problemlos abgenommen werden. Dies minimiert Kosten und Organisationsaufwand.

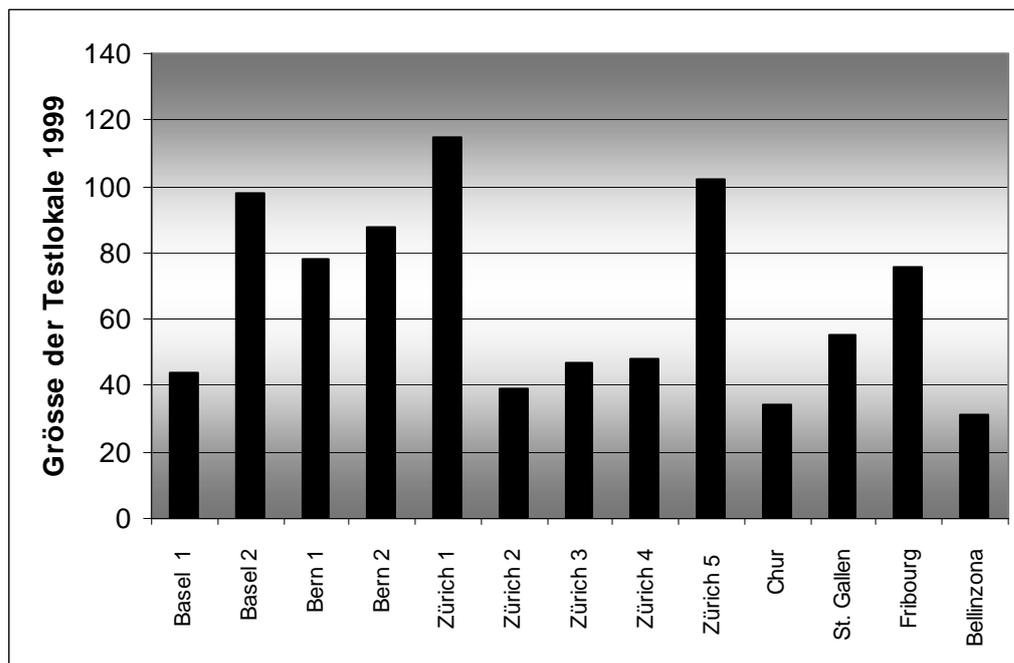


Abbildung 8. Grösse (Zahl der Personen) und Aufteilung der Testlokale 1999.

Testorte und Wunschuniversitäten

Die Wahl des Testortes hatte auch in diesem Jahr 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 im Vorjahr 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. Es zeigt sich in diesen Häufungen auch, dass die von den Kandidaten selbst gewünschte Mobilität gar nicht so hoch ist.

Testort	Studienort				Total
	Basel	Bern	Fribourg	Zürich	
Basel	119 83.8%	10 7.0%	5 3.5%	8 5.6%	142
Bern	6 3.6%	140 84.3%	16 9.6%	4 2.4%	166
Zürich	4 1.1%	14 4.0%	15 4.3%	318 90.6%	351
Chur	6 17.6%	5 14.7%		23 67.6%	34
St.Gallen	5 9.1%	4 7.3%	10 18.2%	36 65.5%	55
Freiburg	3 3.9%	38 50.0%	30 39.5%	5 6.6%	76
Bellinzona	6 19.4%	4 12.9%	3 9.7%	18 58.1%	31
Total	149 17.4%	215 25.1%	79 9.2%	412 48.2%	855

Tabelle 3. Testorte und Wunschuniversität (erste Wahl)

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

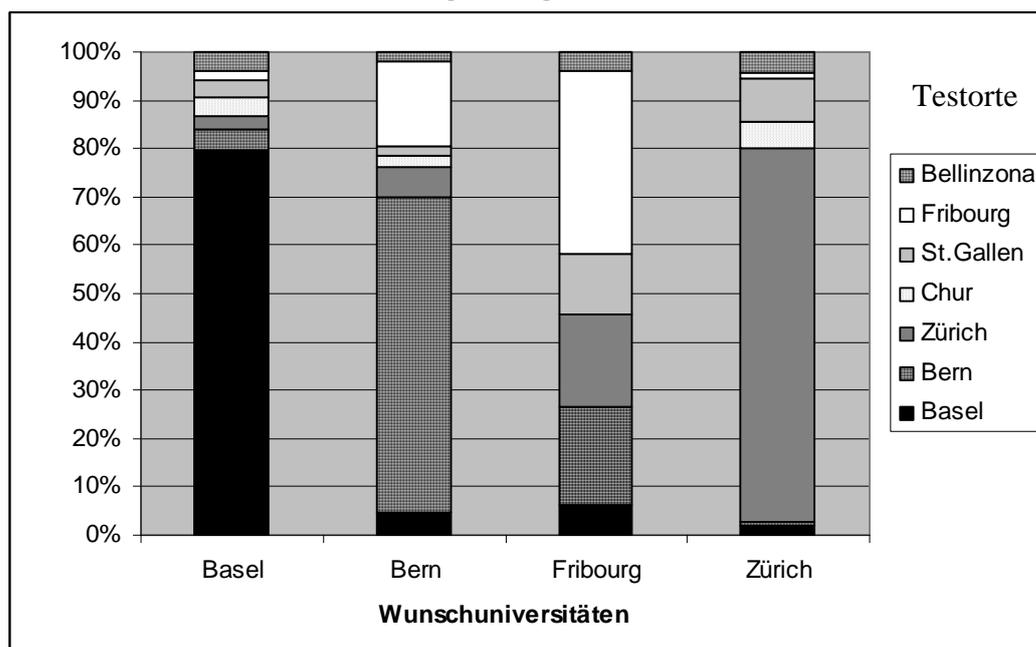


Abbildung 9. Aufteilung der Kandidaten nach Wunschuniversität auf die Testorte.

Testorte und Wohnkantone

Für differenziertere Analysen steht die Tabelle 4 zur Verfügung. Hier sind die Testorte im Verhältnis zum Wohnkanton tabelliert. Angegeben sind auch die Vergleichswerte für Humanmedizin 1998.

Wohnkanton	Testorte							HM	VM	Gesamt	HM 1998
	Basel	Bern	Zürich	Chur	St. Gallen	Fribourg	Bellinzona				
AG	25	2	57					62	22	84	66
AI										0	2
AR					6			6		6	4
BE		109	2			2		87	26	113	100
BL	56				1	2		52	7	59	50
BS	38	1						37	2	39	40
FR		14				33		39	8	47	34
GE						13			13	13	1
GL			4					4		4	4
GR			2	24	1		2	22	7	29	19
JU						2		1	1	2	0
LU	4	14	26					34	10	44	43
NE						5		1	4	5	2
NW		1	2					3		3	6
OW		1	1					2		2	3
SG	1	1	10	8	41			51	10	61	64
SH			4					4		4	7
SO	11	14	3					22	6	28	26
SZ			18					16	2	18	21
TG			15		5			16	4	20	22
TI							29	23	6	29	30
UR	1							1		1	3
VD						11			11	11	3
VS	3	5	1			6		12	3	15	22
ZG			10					4	6	10	6
ZH		1	188		1	1		159	32	191	164
Ausland	3	3	8	2		1		15	2	17	8
	142	166	351	34	55	76	31				

Tabelle 4: Wohnkanton und Testorte, Humanmedizin (HM) Veterinärmedizin (VM) und Vergleich für Humanmedizin 1998 (HM 1998).

Die Wahl des Studienortes abhängig vom Wohnkanton wird in der Tabelle 5 dargestellt. Vor allem **in Freiburg** treten Veränderungen auf, die für die Interpretation der Ergebnisse

bedeutsam scheinen: In diesem Jahr entfallen 36 der 79 (46%) gewünschten Studienplätze in Freiburg auf im Kanton Freiburg Ansässige. Im Jahr 1998 waren dies 29 von 101 (29%). Im Jahr 1999 haben 30 der 79 (38%) Personen den Test auf Französisch abgelegt – im Vorjahr waren dies 31 von 101 (31%). Über 50% der Bewerber stammen aus anderen deutschsprachigen Kantonen – nur zu einem geringen Teil aus französischsprachigen Kantonen. Die Zweisprachigkeit von Freiburg ist als attraktiver Faktor offenbar wirksam, wobei die erhöhte Nachfrage aus dem Kanton einen Rückgang der Nachfrage aus anderen deutschsprachigen Kantonen nicht ganz kompensieren kann. Bewerberinnen und Bewerber aus dem Tessin, die am Test teilnahmen, strebten 1998 und 1999 vornehmlich ein Studium in Zürich ab.

Wohnkanton	Studienort				Total
	Basel	Bern	Fribourg	Zürich	
AG	21	5	1	57	84
AR			1	5	6
BE	3	101	5	4	113
BL	46	8	3	2	59
BS	34	2	1	2	39
FR		11	36		47
GE		11		2	13
GL			1	3	4
GR	5	3	1	20	29
JU	1	1			2
LU	4	16	5	19	44
NE		4		1	5
NW		2	1		3
OW		1	1		2
SG	7	7	9	38	61
SH	1	2	1		4
SO	12	10	1	5	28
SZ		1	1	16	18
TG		1		19	20
TI	5	4	3	17	29
UR	1				1
VD		10		1	11
VS	5	5	3	2	15
ZG			1	9	10
ZH	1	4	4	182	191
Ausland	3	6		8	17
	149	215	79	412	855

Tabelle 5. gewünschte Studienorte nach Wohnkantonen

Testteilnehmer nach Alter und Geschlecht

	Jahrgänge			Total
	1957 – 1975	1976 – 1978	1979 - 1982	
Humanmedizin				
männlich	29 41,4%	123 56,2%	154 40,1%	306 45,5% von Gesamt (1998: 44,8%)
weiblich	41 58,6%	96 43,8%	230 59,9%	367 54,5% von Gesamt (1998: 55,2%)
Gesamt	70 10,4% v. Gesamt 1998: 10,4%	219 32,5% v. Gesamt 1998: 29,5%	384 57,1% v. Gesamt 1998: 60,1%	673
Veterinärmedizin				
männlich	4 22,2%	12 24,5%	28 24,3%	44 24,2%
weiblich	14 77,8%	37 75,5%	87 75,7%	138 75,8%
Gesamt	18 9,9% v. Gesamt	49 26,9% v. Gesamt	115 63,2% v. Gesamt	182

Tabelle 6. Testabsolventen nach Jahrgang und Geschlecht. (Vergleiche der Jahrgänge zu 1998 um je ein Jahr verschoben).

In beiden Fächern überwiegt die Zahl der Frauen – bei Veterinärmedizin beträgt der Frauenanteil sogar $\frac{3}{4}$. In Humanmedizin bleibt das Verhältnis von 55% Frauen zu 45% Männer gegenüber 1998 in etwa konstant. Der Anteil älterer Bewerber hat sich in Humanmedizin nicht verändert und entspricht auch in etwa dem Verhältnis in der Veterinärmedizin. Die nachgewiesene etwas schlechtere Studieneignung aus 1998 hat keine spezifische dissuasive Wirkung auf diese Gruppe ausgeübt.

Die mittlere Altersgruppe hat in Humanmedizin in ihrem Anteil etwas zugenommen. In den Abbildungen 10 und 11 werden die Anteile noch einmal grafisch gegenübergestellt.

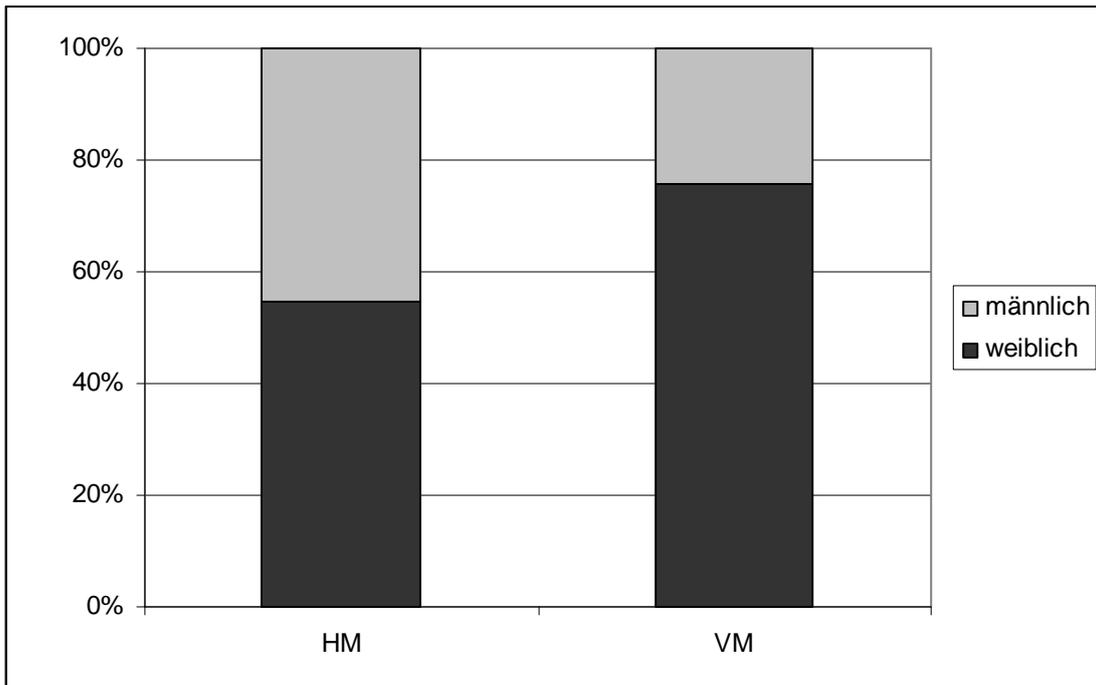


Abbildung 11. Anteil der Geschlechter bei der Bewerbung für Humanmedizin (HM) und Veterinärmedizin (VM).

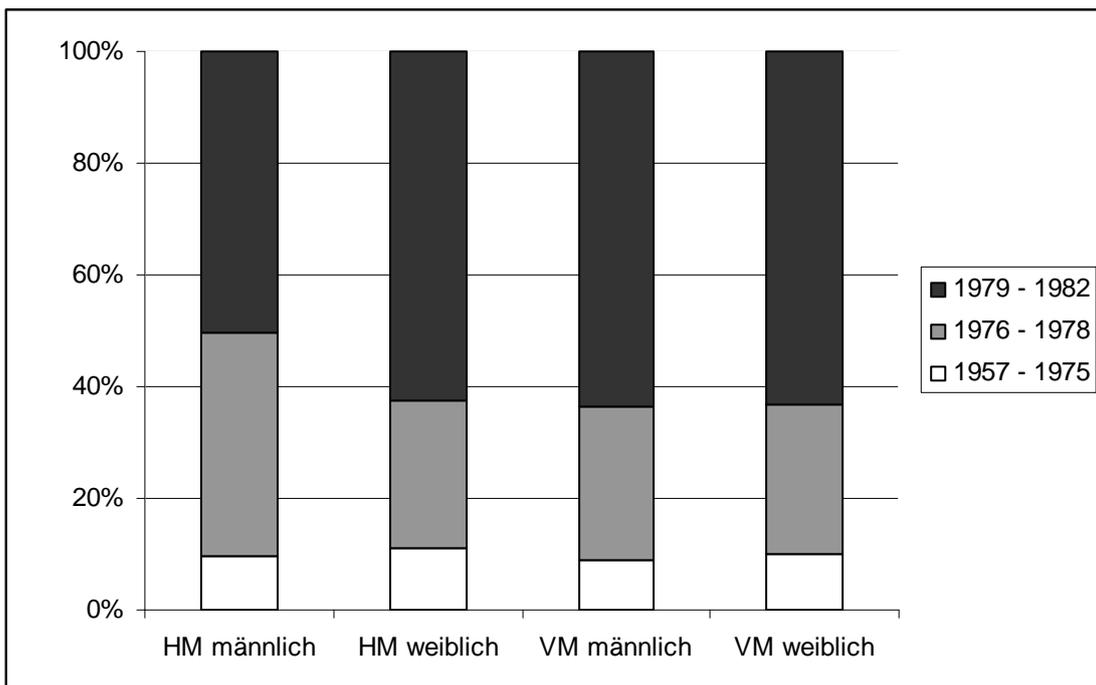


Abbildung 10. Zusammensetzung der Bewerbergruppen nach Geschlecht und Alter (Jahrgänge).

Sprachgruppen nach Alter und Geschlecht

Die Aufschlüsselung der Sprachgruppen nach Alter und Geschlecht zeigt, dass in der französisch- und der italienischsprachigen Gruppe der Anteil der Jüngeren wesentlich höher ist als in der deutschsprachigen. Dies gilt gleichermaßen für Human- und Veterinärmedizin.

Bei allen statistischen Berechnungen wirkt sich die geringe Gruppengrösse der italienisch- und bedingt auch die der französischsprachigen Gruppe ungünstig aus. Die jetzt möglichen Vergleiche mit dem Vorjahr können dies nur bedingt kompensieren.

Sprache	Geschlecht	Jahrgänge			Total
		1957 - 1975	1976 - 1978	1979 - 1982	
deutsch	männlich	29 10,4%	117 42,1%	132 47,5%	278
	weiblich	37 11,0%	92 27,4%	207 61,6%	336
	Total	66 10,7%	209 34,0%	339 55,2%	614
französisch	männlich		4 26,7%	11 73,3%	15
	weiblich	3 15,8%	2 10,5%	14 73,7%	19
	Total	3 8,8%	6 17,6%	25 73,5%	34
italienisch	männlich		2 15,4%	11 84,6%	13
	weiblich	1 8,3%	2 16,7%	9 75,0%	12
	Total	1 4,0%	4 16,0%	20 80,0%	25

Tabelle 7. Humanmedizin: Sprachgruppen nach Alter und Geschlecht aufgeschlüsselt

Sprache	Geschlecht	Jahrgänge			Total
		1957 - 1975	1976 – 1978	1979 - 1982	
deutsch	männlich	4	10	13	27
		14,8%	37,0%	48,1%	100,0%
	weiblich	13	35	59	107
		12,1%	32,7%	55,1%	100,0%
	Total	17	45	72	134
		12,7%	33,6%	53,7%	100,0%
französisch	männlich		2	13	15
			13,3%	86,7%	100,0%
	weiblich	1	2	24	27
		3,7%	7,4%	88,9%	100,0%
	Total	1	4	37	42
		2,4%	9,5%	88,1%	100,0%
italienisch	männlich			2	2
				100,0%	100,0%
	weiblich			4	4
				100,0%	100,0%
	Total			6	6
				100,0%	100,0%

Tabelle 8. Veterinärmedizin: Sprachgruppen nach Alter und Geschlecht aufgeschlüsselt

Analysen, die Sprache und Disziplin gleichzeitig berücksichtigen, sind leider in ihrer Aussagefähigkeit stark eingeschränkt – die Zahl der Bewerber für Veterinärmedizin in der italienischsprachigen Gruppe ist dafür zu gering.

Alter und Maturitätsjahr

Die nachfolgende Tabelle zeigt, 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 haben, handelt es sich offenbar v.a. um Personen, welche die Maturität auf dem sogenannten "zweiten Bildungsweg" erworben haben.

Jhrg.	Maturitätsjahr 19..																Total
	75	80	85	86	87	88	90	91	92	93	94	95	96	97	98	99	
1957	1																1
1958																1	1
1959															1	1	
1961		1												1		2	
1962			1													1	
1963															2	2	
1965															1	1	
1967				1				1							1	3	
1968					1	2								1	2	6	
1969					1								1	1	1	2	6
1970						1	2							2	2	7	
1971													1		4	5	
1972								1	1					1	3	6	
1973									2	1		2		1	4	10	
1974										2	4	1			7	14	
1975										1	1	1	1	4	1	8	17
1976												1	4	5	8	6	24
1977													1	16	21	21	59
1978														6	82	82	170
1979															59	277	336
1980															5	60	65
1981																9	9
1982																1	1
	1	1	1	1	2	3	2	2	3	4	5	5	8	32	183	494	747

Tabelle 9. Jahrgang und Maturitätsjahr

Beschreibung des verwendeten Eignungstests

Aufbau des Tests

Der EMS besteht aus neun verschiedenen Untertests, die zu einem Gesamtwert verrechnet werden. Die Tabelle 10 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	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: 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 10. 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 die langjährig erprobten 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 10) 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 1999 wurden die Einstreuaufgaben ausnahmsweise aus Aufgaben gebildet, die 1998 bereits verwendet worden sind. So ist ein direkter Leistungsvergleich beider Jahrgänge 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).

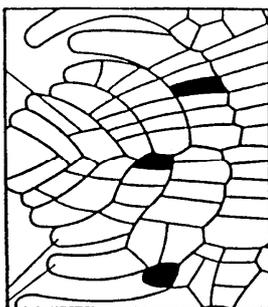
Bei der Beurteilung der Aufgaben wird die **Nähe zu Studienanforderungen** deutlich: an der Entwicklung haben neben Gymnasiallehrern, Medizinerinnen und Psychologen auch Lehrbeauftragte des Grundstudiums Medizin mitgearbeitet. Die Struktur der Untertests ist auf detaillierte Anforderungsanalysen eines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

1. Untertest: Muster zuordnen

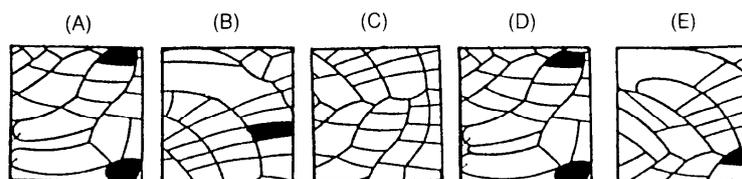
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wiederzuerkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

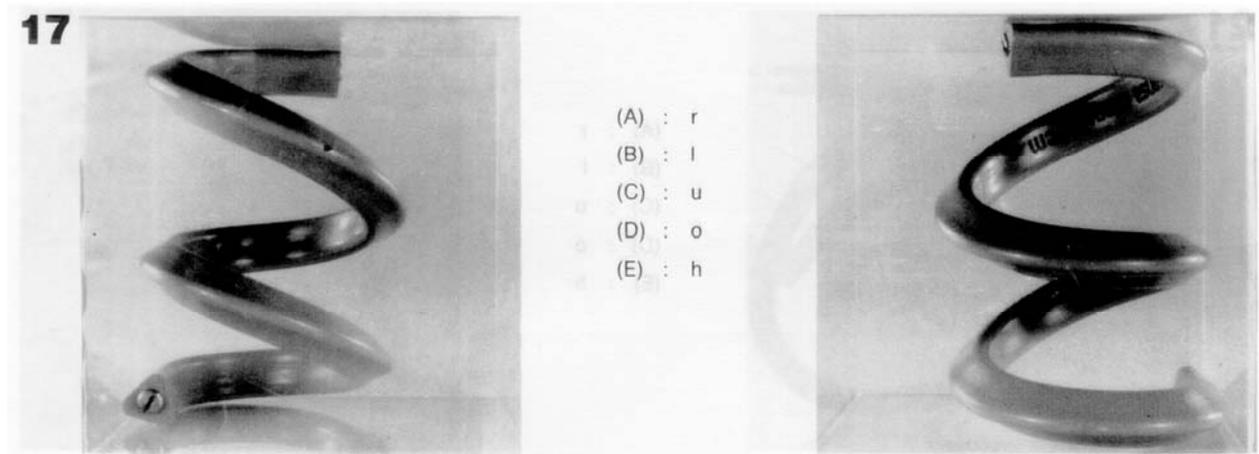
- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind 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 Querstreichen versehen sind, die entweder beide unten, beide oben oder je einer unten und oben angebracht sind, markiert werden. Die Lösungsmenge ist ebenso wichtig wie die Fehlerfreiheit der Bearbeitung. Dieser Test ist trainierbar – im Test-Info wird darauf hingewiesen, diesen Untertest vor der Testabnahme mehrfach zu üben.

b b b

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

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

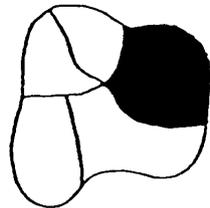
6. Untertest: Figuren lernen

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.

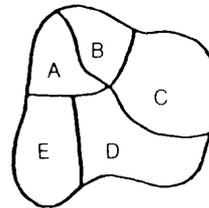
Der Untertest „Figuren lernen“ prüft, wie gut man sich Einzelheiten von Gegenständen einprägen und merken kann.

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 Jodatomb 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 T₃ 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 Bewerber Eignungsgespräche durchzuführen).
- Die Trainierbarkeit der Eignungsabklärung muss gering bleiben – da andernfalls der Erfolg vom Trainingsaufwand und damit auch von finanziellen Möglichkeiten 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 12). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Jahrgänge 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.

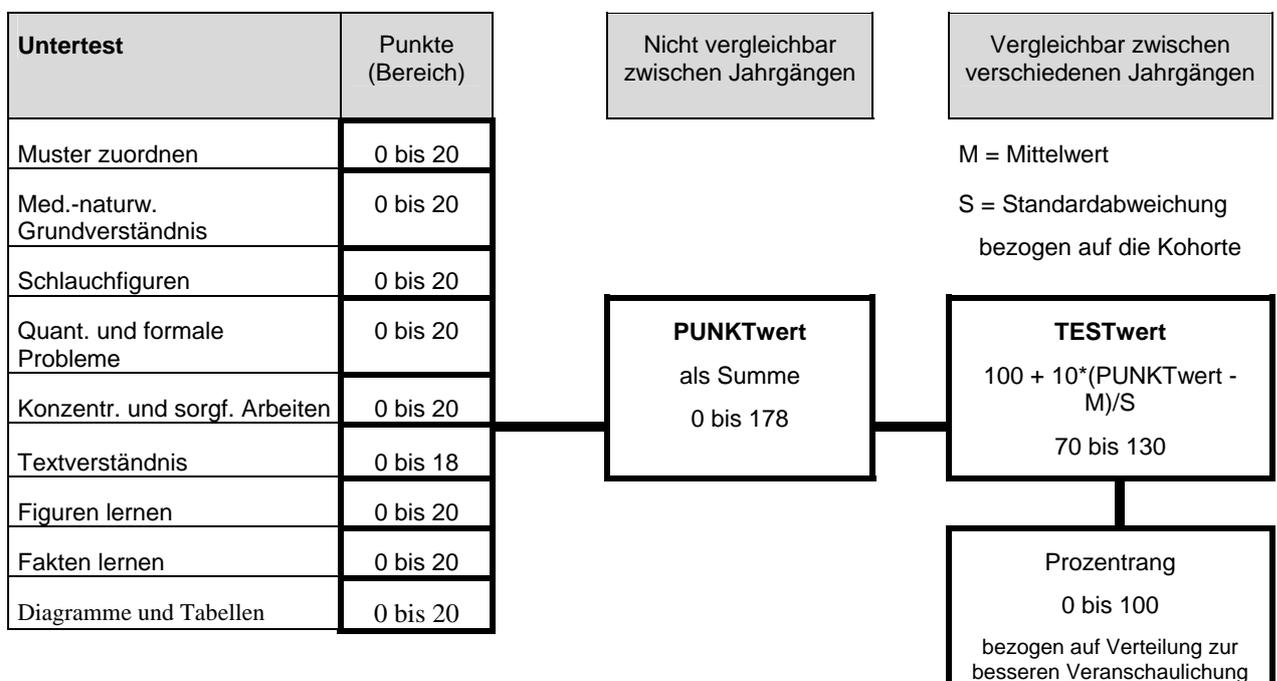


Abbildung 12. Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert.

Vorhersage des Studienerfolges in Deutschland

Der Test wurde in Deutschland über zehn Jahre angewendet. In dieser Zeit fanden mehrere Evaluationen statt, um seine Gütekriterien zu überprüfen (vgl. Bericht von Trost u.a., 1998). Die Vorhersagefähigkeit von Studienerfolg ist dabei das zentrale Kriterium. Mit Hilfe des Tests sollen diejenigen bevorzugt eine Chance erhalten, die mit grösserer Wahrscheinlichkeit das Studium auch erfolgreich beenden. Die folgenden Ergebnisse sind dem 18. Arbeitsbericht des Instituts für Test- und Begabungsforschung (Trost, 1994) entnommen. Zu beachten ist, dass in Deutschland fünf Zugangsmöglichkeiten zum Studium bestanden. Bei der Platzvergabe nach einer kombinierten Abitur-Test-Quote (45% aller Plätze) und einer Test-Quote (10%) geht das Testergebnis als Kriterium ein. Zwanzig Prozent der Plätze werden aufgrund einer Wartezeit, 15% aufgrund von Auswahlgesprächen und 10% als "Vorab-Quote" (z.B. bereits früher zugesagte Plätze, Zweitstudienplätze, Härtefälle) vergeben. Die Berücksichtigung der Quoten erfolgt sequentiell: Die Auswahl nach der Abitur-Test-Quote erfolgt zuerst, dann die nach der Testquote. Personen mit guten Abitur- und Testleistungen werden also bereits in der ersten Quote berücksichtigt. Die Testquote beinhaltet dann Personen, die bei Kombination Abitur und Test nicht zugelassen werden können, aber eine gute Testleistung erreicht haben. Die Auswahlgesprächs-Quote trifft dann nur für Personen zu, die nicht mittels einer der beiden erstgenannten zugelassen worden sind.

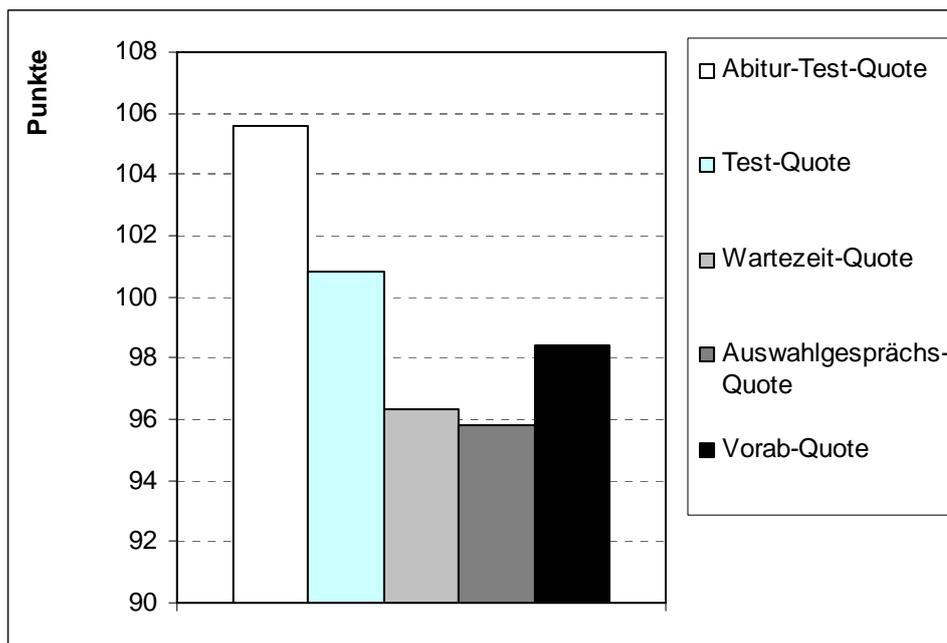


Abbildung 13. Durchschnittliche Punkte für schriftliche Prüfungsleistungen während des Studiums nach der Zulassung (aus Trost, 1994, S. 144).

In der schriftlichen Prüfung (Abbildung 13) werden von den Personen mit Zulassung nach der Abitur/Test- und der Testquote die besten Leistungen erzielt. Bemerkenswert ist, dass die Personen der Auswahlgesprächs-Quote hier die schlechtesten Leistungen erreichen. Auch in der mündlichen Prüfung (Abbildung 11; die Note 1 ist in Deutschland die beste Leistung!) zeigt sich dieser Trend. Die Korrelationen (Tabelle 11) verdeutlichen, dass vor allem die schriftliche Prüfungsnote durch den Test sehr gut vorhergesagt werden kann, der Test seiner Bezeichnung als „Probeklausur“ gerecht wird. Die Korrelation von 0.45 (TMS

Gesamtwert mit Prüfungs-Gesamtnote) ist ein sehr hoher Wert im Vergleich zu anderen bekannten Prognosekorrelationen.

Die in Deutschland ermittelte sehr gute Prognosegüte für Studienerfolg lässt hoffen, dass sich bei einer Zulassung auf der Basis der Testwerte die Zahl der Studienabbrecher in der Schweiz deutlich verringert und die Zugelassenen gute Chancen für einen erfolgreichen Studienabschluss in der jeweils üblichen Studienzeit haben. Auch die Studiendauer bis zum erstmaligen Antritt zur Prüfung ist für die Personen der Abitur-Test-Quote und der Testquote deutlich kürzer als für die anderen drei Gruppen. Dies gilt für Human-, Zahn- und Tiermedizin gleichermaßen (Troost u.a., 1998).

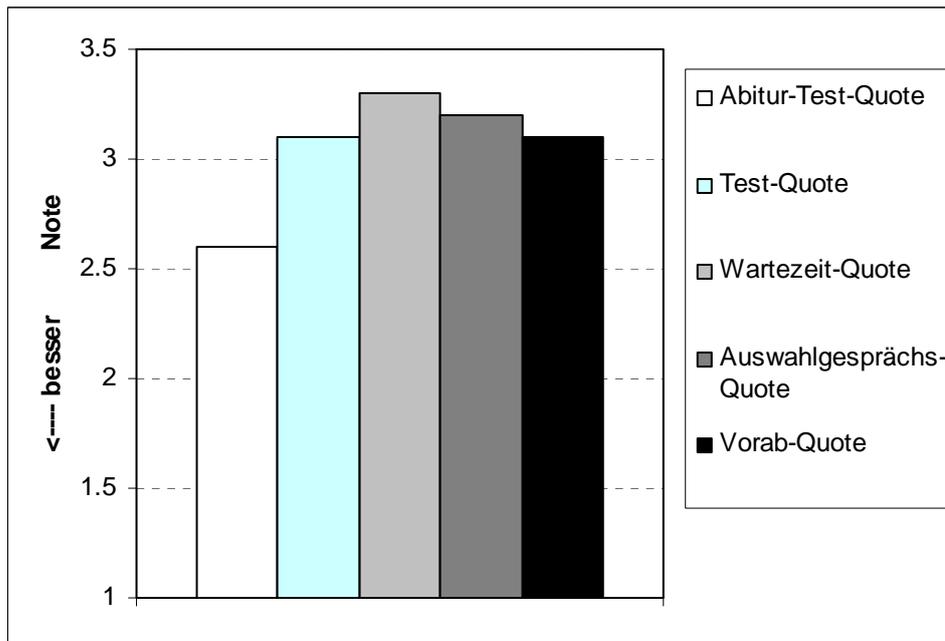


Abbildung 14. Durchschnittliche Note für mündliche Prüfungsleistungen während des Studiums nach der Zulassung. Aus Trost 1994, S. 144).

Prädiktorvariable	Korrelationen mit den Ergebnissen der Ärztlichen Vorprüfungen		
	schriftlicher Teil	mündlicher Teil	Gesamtnote
TMS-Gesamtwert	<u>0.50</u>	0.27	<u>0.45</u>
Muster zuordnen	0.17	0.10	0.14
Med.-naturw. Grundverständnis	0.41	0.21	0.36
Schlauchfiguren	0.28	0.14	0.25
Quantitative und formale Probleme	0.45	0.21	0.39
Konzentriertes und sorgf. Arbeiten	0.24	0.16	0.22
Textverständnis	0.41	0.21	0.35
Figuren lernen	0.22	0.14	0.21
Fakten lernen	0.17	0.11	0.16
Diagramme und Tabellen	0.41	0.21	0.37

Tabelle 11. Korrelationen zwischen den Prädiktorvariablen (TMS-Ergebnisse) und den Ergebnissen der ärztlichen Vorprüfungen; aus Trost (1994).

Konsequenzen für eine Testevaluation in der Schweiz

Es ist notwendig, die Testgüte auch für die Schweiz zu evaluieren – dabei aber sind die zu Deutschland unterschiedlichen Bedingungen zu beachten, weil

- hier nur ein Zulassungskriterium aufgrund des Testwertes verwendet wird,
- die Zulassungsquote wesentlich grösser ist (80% bezogen auf die Testabsolventen – durch die nachfolgenden Rückzüge bis zu 90%) und
- Unterschiede im Studienablauf zu Deutschland vorhanden sind.

Daraus ergeben sich die Bewertungskriterien des Zulassungsverfahrens und des EMS, die einmal die gesamthafte Verbesserung der Situation einbeziehen und zum anderen auf die spezifische Leistung des Testes zugeschnitten sind. Spürbare Verbesserungen werden im Übrigen auch davon abhängig sein, dass der „Stau“ bei Repetitionen der Prüfungen abgebaut ist. Die folgende Tabelle stellt die Bewertungsebenen zusammen:

Kriterium	Evaluation des Zulassungsverfahrens	Evaluation des Tests
	Auswirkung gesamthaft durch die verbesserte Studiensituation und geeignetere Studierende – wird durch eine geringe „Selektionsrate“ des Tests 1998 eher wenig Auswirkung zeigen.	Abschätzung des Anteils, den der Eignungstest daran hat
Wahrscheinlichkeit, dass das Studium mit Erfolg abgeschlossen wird (Verringerung der Abbrecherquote)	Gesamthafte Verringerung gegenüber den Vorjahren (aber: z. Zt. geringe Selektionsrate des Tests!)	Studienabbruch korreliert negativ mit der Testleistung im EMS
Verkürzung der Studienzeiten – zunächst 1. und 2. Propädeutikum	Gesamthafte Verkürzung der Studienzeiten gegenüber den Vorjahren (aber: z. Zt. geringe Selektionsrate des Tests!)	Studiendauer korreliert negativ mit der Testleistung im EMS; Bestehensrate der Prüfungen im ersten Versuch korreliert positiv mit der Testleistung
Verbesserung der Studienleistungen	Gesamthafte Verbesserung der Studienleistungen (aber: z. Zt. geringe Selektionsrate des Tests!)	Studienleistungen korrelieren positiv mit der Testleistung im EMS

Tabelle 12. Evaluation des EMS in der Schweiz

Auswirkungen in der Kohorte werden sich eher nur langfristig und in Universitäten mit besonders hoher Überlastung zeigen, weil durch den recht milden Numerus clausus keine starke Selektion nach der Leistung erfolgte. Die Rückzüge (die sicher nicht unbeeinflusst durch den Test erfolgten) sind hinsichtlich ihrer Ursachen nicht abschätzbar.

Geringe Trainierbarkeit – ein Vorzug des EMS

Bei der Diskussion über Zulassungskriterien spielt deren soziale Verträglichkeit eine grosse Rolle. Der Vorbereitungsaufwand und die dafür notwendigen Kosten sollen möglichst gering sein. Da der Test in Deutschland über Jahre an sehr grossen Personengruppen durchgeführt worden ist, hatte sich dort ein „Trainingsmarkt“ etabliert, welcher den Kandidatinnen und Kandidaten Testvorbereitungskurse – und manchmal noch peripherere Kurse und Materialien – zu kommerziell recht unterschiedlichen Bedingungen angeboten hat. Zumindest in der Deutschschweiz wird nun versucht, diese Konzepte ebenfalls zu vermarkten. Die Argumentation dieser Kurs-Anbieter beruht vor allem darauf, dass man den Test gut trainieren und sich (mit Hilfe der Kurse) den Erfolg sichern kann. Die verständlicherweise vorhandene Besorgnis der Kandidatinnen und Kandidaten findet hier einen „Anker“ – bei nur 8% effektiver Ablehnungsquote in der Schweiz ist es nicht schwer, einen Erfolg vorherzusagen.

Die Frage der Trainierbarkeit hat auch in Deutschland aus den genannten Gründen eine wichtige Rolle bei der Testevaluation gespielt. Zumindest die wissenschaftlichen Fakten sprechen gegen eine Notwendigkeit von speziellem Training. Mittlerweile konnte zweifach wissenschaftlich abgesichert werden, dass durch eine selbständige gezielte Vorbereitung auf den EMS anhand der kostenlos abgegebenen Test Info und einer veröffentlichten Originalversion ebensolche Leistungen erreicht werden können, wie sie durch Nutzung zusätzlicher Trainingskurse erzielt werden (zu Details vgl. Beitrag von Hofer u.a. im Bericht 3 des ZTD, 1997).

Ein zumeist gezieltes Missverständnis in der Argumentation von Trainingsanbietern besteht in einer Vermischung von Vorbereitung und Training. Es wurde nie bestritten, dass man sich auf den Test vorbereiten kann und muss. Genau diesem Zweck dienen die Test-Info und die veröffentlichten Originalversionen – einschliesslich der dort gegebenen Hinweise, wie dies optimal geschehen kann. Diese empfohlene Vorbereitung ist aber grundsätzlich ausreichend, um den Teil möglicher Leistungsverbesserung beim EMS auszuschöpfen, der überhaupt durch „Training“ – besser Vorbereitung – erreicht werden kann. Bekanntlich sind alle Testaufgaben so konstruiert, dass kein spezielles Vorwissen notwendig ist. Auch beim Untertest "Konzentriertes und sorgfältiges Arbeiten" ist nach einigen wenigen Übungsdurchgängen ein Niveau erreicht, welches sich für den Einzelnen durch weiteres Üben kaum noch steigern lässt. Die Übungs-Vorlage dazu finden sich in der Test-Info.

Der Besuch zusätzlicher Trainingskurse von privaten Anbietern ist daher weder notwendig noch nachgewiesenermassen effizienter als diese empfohlene Vorbereitung. Eine Gefahr der Benutzung von Materialien und Kursen, die von privaten Anbietern jetzt auf den Markt gebracht werden, besteht darin, dass diese den Originalaufgaben nur nachempfunden sind und sogar auf Irrwege bei der Lösung der tatsächlich verwendeten EMS-Aufgaben führen können. Ihnen fehlt - nach unserem Wissen ausschliesslich - die aufwendige Überprüfung und Revision, denen die tatsächlich verwendeten Testaufgaben unterliegen. Bekanntlich werden alle neuen Aufgaben an vorhandenen Aufgaben geeicht, um über die Jahre eine vergleichbare Testschwierigkeit und -qualität zu erhalten. Offenbar wird auch der Versuch unternommen, unter Bezugnahme auf den EMS Materialien und Kursinhalte zu verkaufen, die nur wenig oder nichts mit dem EMS zu tun haben.

Es kann natürlich für die Aufnahme eines Studiums von Vorteil sein, in entsprechenden Kursen allgemeine Lern- und Studienstrategien genauer kennenzulernen und sich mit studententypischen Problemstellungen auseinanderzusetzen. Solche Kurse – angeboten zu

fairen Bedingungen – sind sicherlich nicht schädlich für den weiteren Lebensweg. Sie stehen aber in keinem nachgewiesenen Zusammenhang mit dem Erfolg beim EMS, wie meistens von den Anbietern behauptet wird. Die Ausnutzung von auf den Test bezogenen Ängsten als Verkaufsargument für diese Dinge bewegt sich am Rande der moralischen Tolerierbarkeit.

Ein organisiertes Bearbeiten der veröffentlichten Originalversion unter "Ernstfallbedingungen", die in einem Kurs simuliert werden, kann dagegen sinnvoll sein – allein fällt es erfahrungsgemäss schwerer, gleichzeitig auf die Einhaltung der Zeiten zu achten und die Aufgaben zu bearbeiten. Diese Kurse werden in einigen Kantonen zu finanziell angemessenen Bedingungen angeboten und es wird seitens des ZTD angestrebt, diese Anbieter ab 2000 zu schulen.

Testanwendung in der Schweiz 1998

Organisation der Testabnahme

Die Testabnahme fand an allen Testorten einheitlich am 2.7.1999 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 beteiligt waren und so über Erfahrungen verfügten.

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.

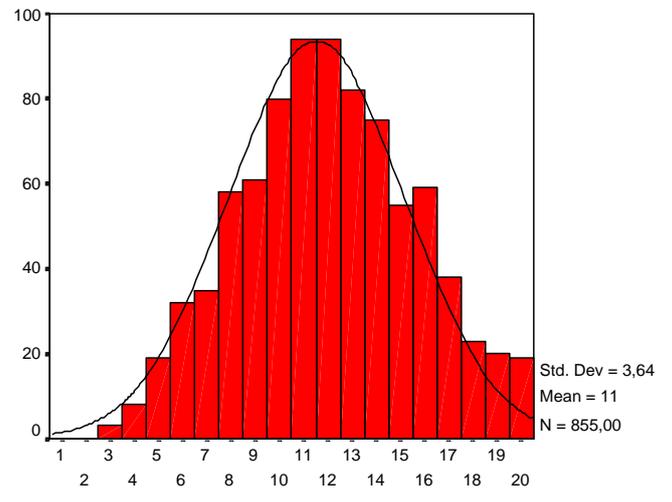
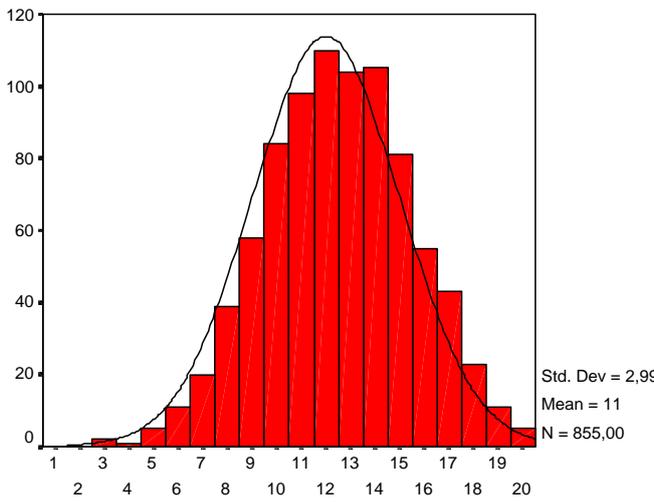
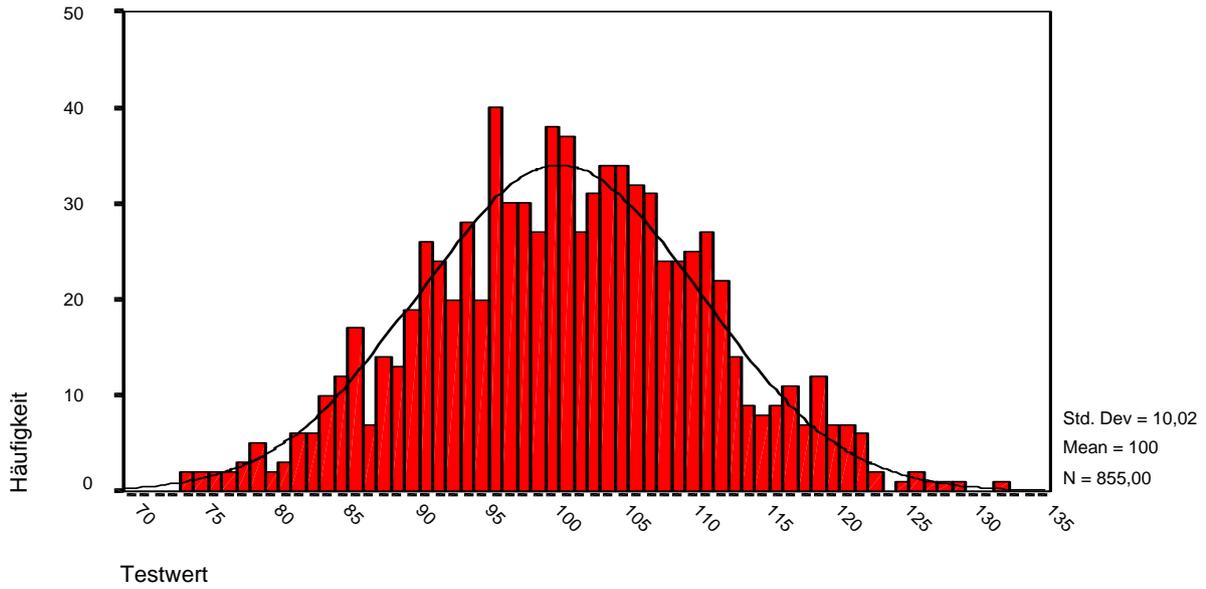
Die formalen Bedingungen an allen Testorten waren vergleichbar und es traten keine Benachteiligungen auf.

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.

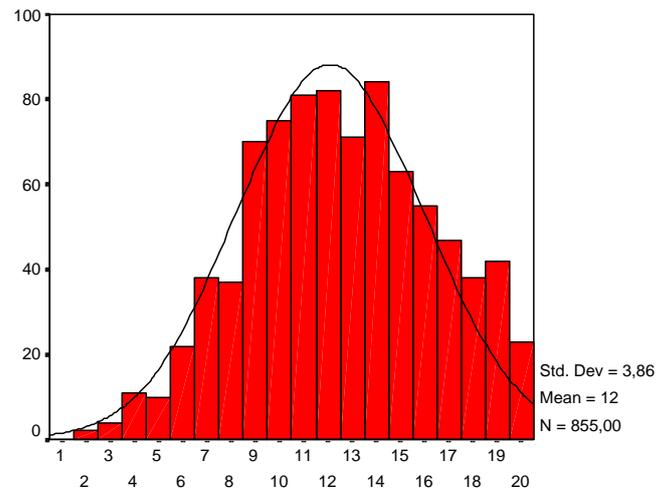
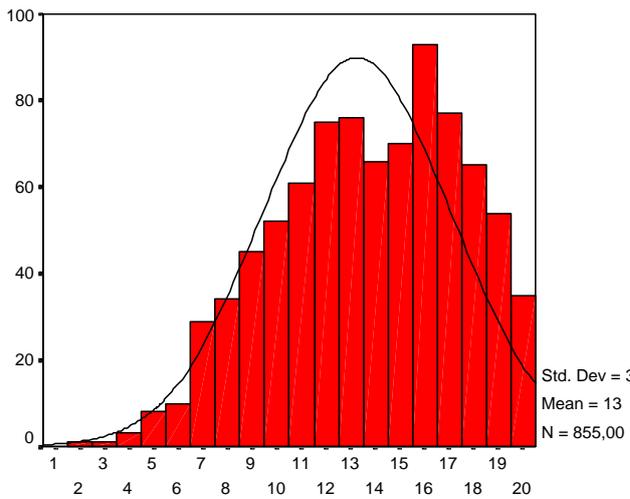
	Testwert	Muster zuordnen	Schlauch figuren	Figuren lernen	Konzent. u. sorgfält. Arb.	Med.-nat.wiss. Grundverst.	Quantit. u. form. Probleme	Textverständnis	Fakten lernen	Diagramme und Tabellen
Mittelwert	100.0	11.4	12.7	12.5	12.3	11.0	11.5	8.7	11.7	10.3
Stand.-abweich.	10.0	2.9	3.7	3.6	3.3	3.6	3.8	3.3	3.7	3.1
extremste Differenz	.029	.070	.103	.069	.088	.064	.065	.093	.058	.059
Kolmog.-Smirnov Z	.835	2.048	3.009	2.017	2.574	1.873	1.892	2.723	1.685	1.737
Asymp. Sig.	.488	.000	.000	.001	.000	.002	.002	.000	.007	.005

Tabelle 13. Verteilungsprüfung auf Normalverteilung (One-Sample Kolmogorov-Smirnov Test).



Muster zuordnen

Med.-naturwiss. Grundverst.



Schlauchfiguren

Quantit. und form. Probleme

Abbildung 15. Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 1-4.

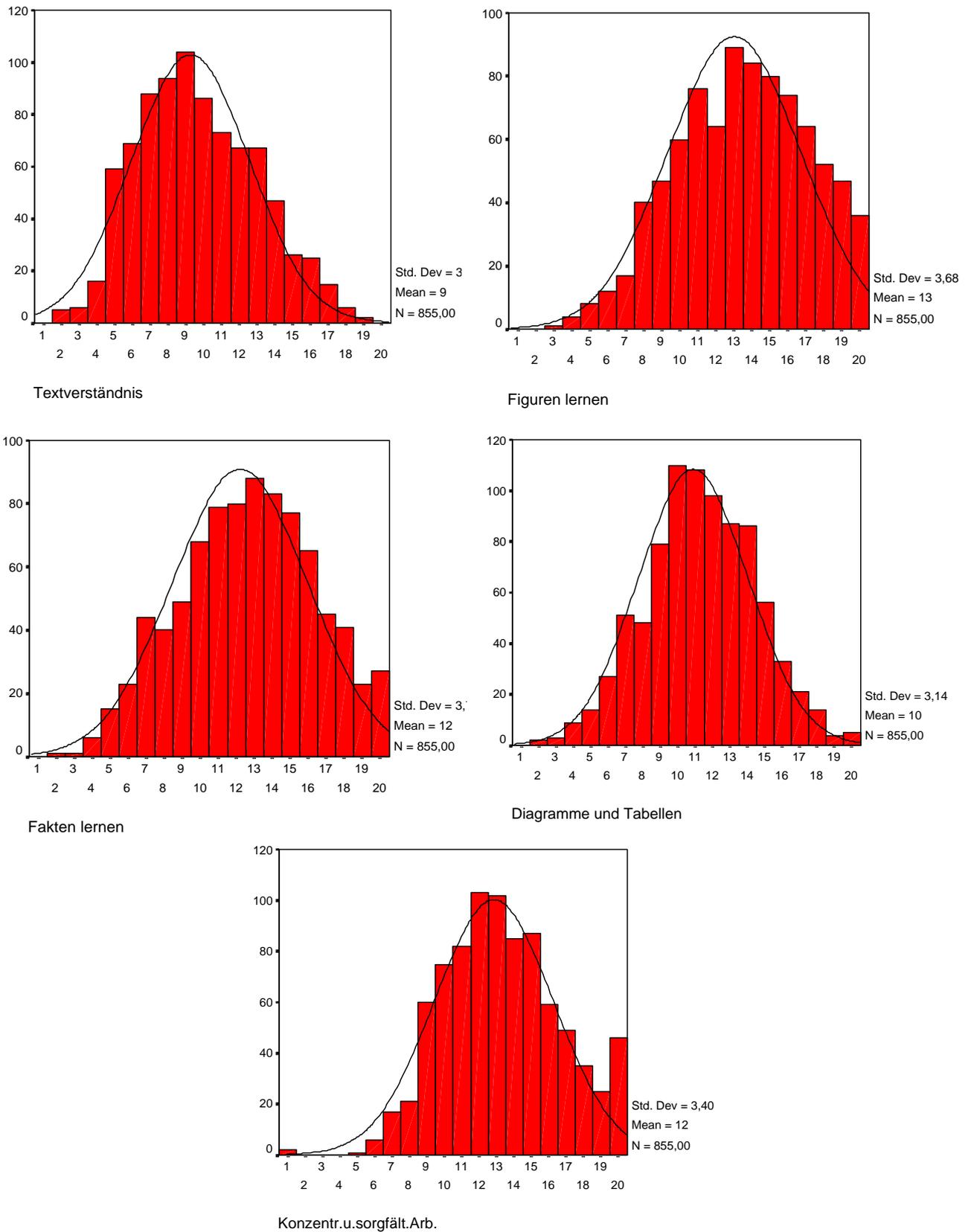


Abbildung 16. Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 5-9 (Wert 0 wird nicht dargestellt, dieser kommt insgesamt nur einmal vor – siehe auch Anhang).

Äquivalenz der Testformen 1 und 2

Wie im Vorjahr wurden zwei pseudoparallele Testformen verwendet. Die nachfolgende Tabelle stellt Mittelwerte und Standardabweichungen beider Formen gegenüber. Die statistische Prüfung (Anhang Tabelle A 4) 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	Mittelwert	Standardabweichung
Testwert	1	100.3	10.2
	2	99.7	9.7
Muster zuordnen	1	11.5	3.1
	2	11.4	2.8
Schlauchfiguren	1	12.6	3.6
	2	12.8	3.8
Figuren lernen	1	12.4	3.6
	2	12.5	3.7
Konzentr. u. sorgfält.Arbeiten	1	12.4	3.4
	2	12.2	3.3
Med.-naturwiss. Grundverständnis	1	11.1	3.5
	2	10.9	3.7
Quantit. und form. Probleme	1	11.7	3.9
	2	11.4	3.7
Textverständnis	1	8.9	3.2
	2	8.6	3.3
Fakten lernen	1	11.7	3.8
	2	11.6	3.6
Diagramme und Tabellen	1	10.5	3.1
	2	10.2	3.1

Tabelle 14. Mittelwerte und Standardabweichungen für die Formen 1 und 2, alle Mittelwertunterschiede sind statistisch nicht signifikant (siehe Anhang Tabelle A 4).

Vergleich der Testfassungen 1998 und 1999

Durch die Umwandlung der Punktwerte in Testwerte ist gewährleistet, dass die Ergebnisse aus verschiedenen Jahren miteinander verglichen werden können. Die "absolute" Schwierigkeit als Punktwert unterliegt dabei über die Jahre geringfügigen Schwankungen.

- Ein Vergleich der aus Deutschland mitgeteilten Summe der Itemschwierigkeiten zeigt, dass der Test 1999 um 3.6 Punkte "schwerer" ist als 1998 – jeweils bezogen auf die deutsche Ausgangsform und ohne "Konzentriertes und sorgfältiges Arbeiten". Diese 3.6 Punkte sind verglichen mit der Standardabweichung des Punktwertes von 20.2 allerdings vernachlässigbar gering.
- In der Schweiz werden 1999 im Fach Humanmedizin 4.9 Punkte weniger erreicht als 1998 – die grössere Schwierigkeit wirkt sich entsprechend aus.
- Im letzten Jahr war die Deutschschweizer Gruppe um 7.3 Punkte besser als die deutschen Vergleichswerte – in diesem Jahr sind es 6.0 Punkte. Bei der Bewertung dieser Zahlen ist allerdings die grössere Zahl von Testteilnahmen in Deutschland – bezogen auf die Zahl der Maturanden – zu beachten.

	D 1999	CH 1999 Gesamt	CH 1999 Humanmed.	D 1998	CH 1998 Humanmed.
Muster zuordnen	10.99	11.63	11.72	12.2	13.2
Med.-naturwiss. Grundverständnis	11.26	11.26	11.34	11.6	12.5
Schlauchfiguren	11.94	12.90	13.03	11.8	13.3
Quantitative und formale Probleme	9.75	11.62	11.78	9.9	11.9
Textverständnis	8.87	8.88	8.95	10.2	10.3
Figuren lernen	11.81	12.67	12.60	11.6	12.6
Fakten lernen	10.97	11.81	11.82	11.2	11.6
Diagramme und Tabellen	10,28	10,55	10.63	11,0	11,4

Tabelle 15. Mittelwerte (Punkte) der Untertests 1998 und 1999 deutsche Sprachgruppe Schweiz und Schätzung aufgrund der Item-Schwierigkeitswerte aus Deutschland.

Bezogen auf die einzelnen Untertests sind vor allem "Muster zuordnen" und "Textverständnis" schwerer 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.

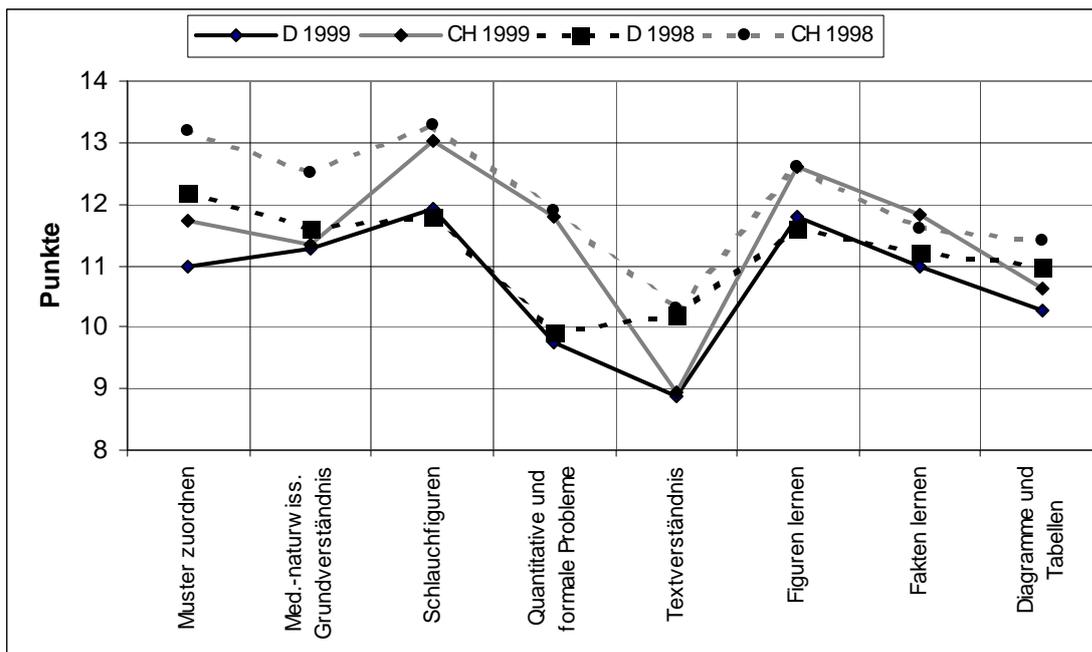


Abbildung 17. Mittelwerte der Untertests für Schweiz (deutschsprachig) und Vergleichswerte Deutschland 1999 und 1998 – jeweils Humanmedizin.

Äquivalenz der Sprachversionen

Das Kriterium der Rechtsgleichheit muss für die drei Sprachgruppen eingehalten sein. Der Test wurde für den Einsatz im Jahr 1999 nach den international üblichen Kriterien aus der deutschen Fassung in die französisch- und italienischsprachige Fassung übertragen, wofür die Bezeichnung „Adaptation“ üblich ist. Im Bericht 4 (1998) des ZTD wurde dieses Verfahren bereits ausführlich erläutert. Dieser Prozess geht über die einfache Übersetzung eines Textes hinaus, mit dem Ziel, in den unterschiedlichen Sprachgruppen eine vergleichbare Schwierigkeit, Zuverlässigkeit und Gültigkeit eines Items zu erzielen. Dabei müssen sprach- und kulturspezifische Eigenheiten berücksichtigt werden.

Für diese Adaptation konnte auf Erfahrungen aus 1998 zurückgegriffen werden: Neben der Erstübersetzung durch einen Fachübersetzer erfolgten je zwei Kontrollen pro Sprachform durch Zweisprachler und Bearbeitungen durch je einen zweisprachigen Mitarbeiter des ZTD. Für die Kontrollen wurden Personen ausgewählt, die mit den Studienanforderungen und dem Niveau der Maturität vertraut sind. Die Beteiligten wurden in diesem Jahr intensiv geschult – auch unter Nutzung der Ergebnisse aus 1998. So wurde auf eine Rückübersetzung aus der Sprache der Adaptation verzichtet und auch die Übersetzer wurden ausdrücklich darauf hingewiesen, dass keine Rückübersetzungen stattfinden. Bei Bekanntheit der Tatsache einer Rückübersetzung orientieren sich die Übersetzer sehr stark auch in der Zielsprache am Satzbau der Herkunftssprache – was dann die Rückübersetzung erleichtert, nicht aber das Verständnis der Aufgaben in der Zielsprache. Als

Qualitätssicherungsinstrument wurden die Schwierigkeitsdiagramme aus 1998 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.

Im Jahre 1998 wurden die Testwerte der drei Sprachgruppen getrennt berechnet, indem die jeweiligen Mittelwerte und Standardabweichungen als Standardisierungsgrundlage verwendet worden sind. Der Vorteil war die gleiche Zulassungswahrscheinlichkeit für jede Sprachgruppe. Der „Preis“ war, dass alle Unterschiede ausgeglichen worden sind – also auch nicht testbedingte, deren Ausgleich eigentlich nicht gerechtfertigt ist. Die geringen Stichprobengrössen der französisch- und italienischsprachigen Gruppe und fehlende statistische Schätzverfahren (die zu diesem Zeitpunkt nur an grossen Stichproben überprüft waren) lieferten eine Begründung für dieses Vorgehen, die noch heute Bestand hat.

In diesem Jahr wurde ein differenziertes Vorgehen angewendet, welches nicht zuletzt auf neuen Ergebnissen – v.a. bezüglich der statistischen Prüfverfahren für kleine Stichproben – beruht, welche im Frühjahr 1999 anlässlich eines Workshops der Internationalen Testkommission zu Testadaptation bekanntgegeben wurden.

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 Abbildung 18 fasst zusammen, welche Ursachen für Unterschiede in den drei Sprachgruppen verantwortlich sein können:

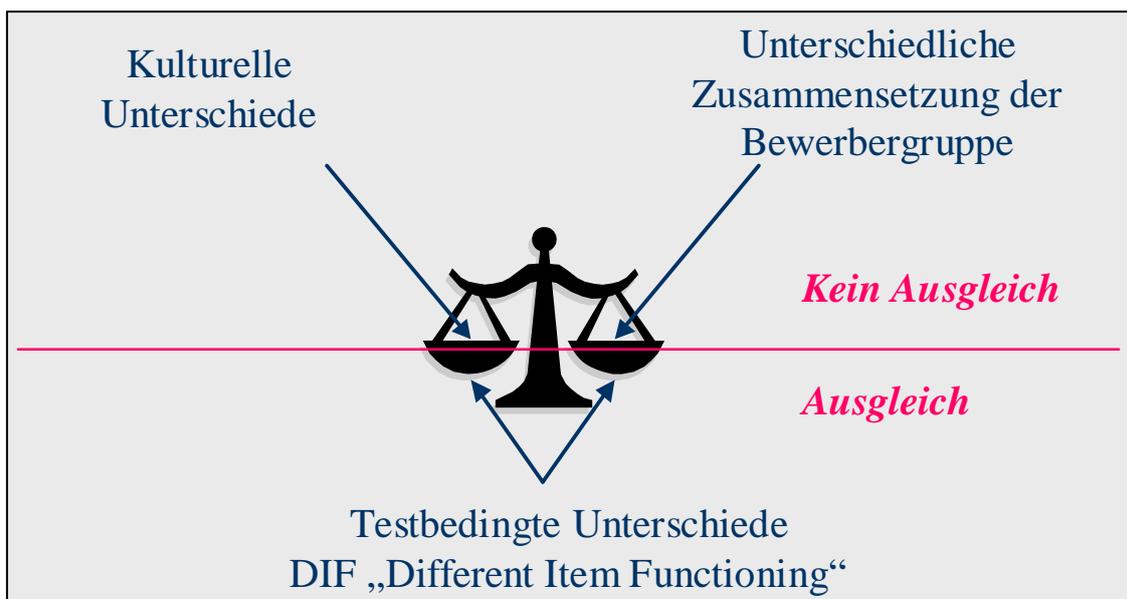


Abbildung 18. Faktoren, die sich auf die Unterschiede zwischen den Sprachversionen auswirken.

Kulturelle Unterschiede zwischen den Sprachgruppen

- Mögliche generelle Fähigkeitsunterschiede oder, präziser ausgedrückt, 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.

Aufgrund der Analyse der Ergebnisse von 1998 hat die Schweizerische Hochschulkonferenz den Beschluss gefasst,

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.

Dieser Beschluss legt die Bedingungen des Ausgleiches fest – gesucht war ein begründetes und faires Verfahren, diesen Ausgleich vorzunehmen. In diesem Jahr liegen auch neue Evaluationsverfahren vor, welche die Richtigkeit des Ausgleiches überprüfbar machen.

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 "Different 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 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Gruppen auf. 1998 und 1999 erzielt die italienischsprachige Gruppe den besten Wert bei „Quantitative und formale Probleme“ – ansonsten sind die Werte der deutschen Sprachgruppe am besten. In der Tabelle 16 sind die signifikanten Mittelwertunterschiede und die Ergebnisse der varianzanalytischen Prüfung dargestellt.

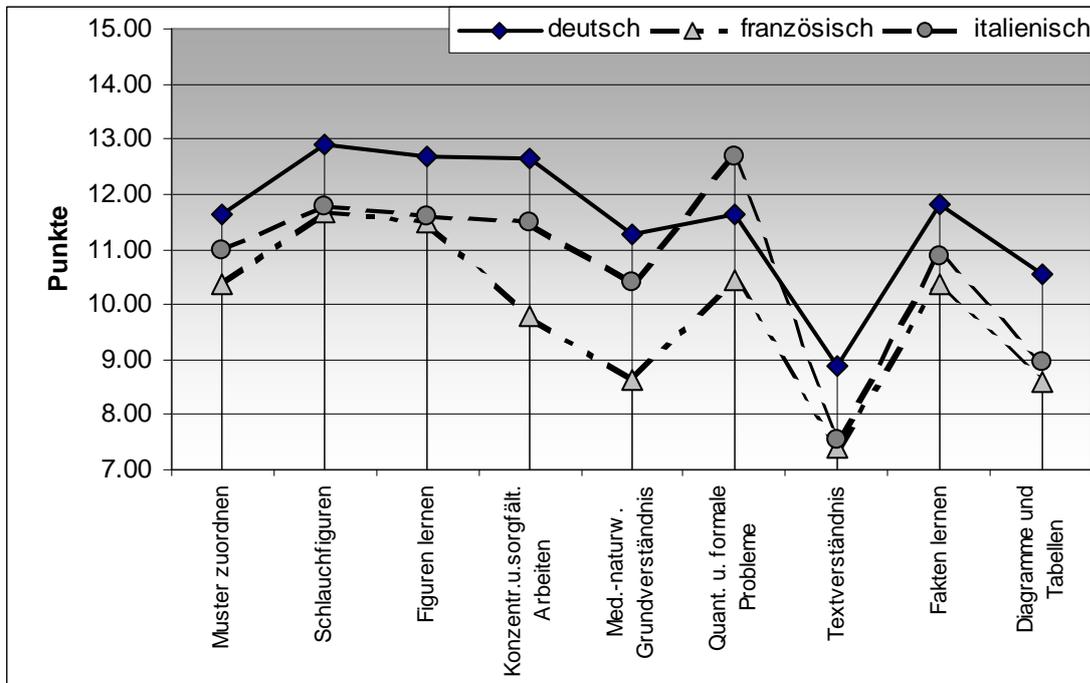


Abbildung 19. Mittelwerte der Untertests für die Sprachgruppen 1999 (unkorrigiert).

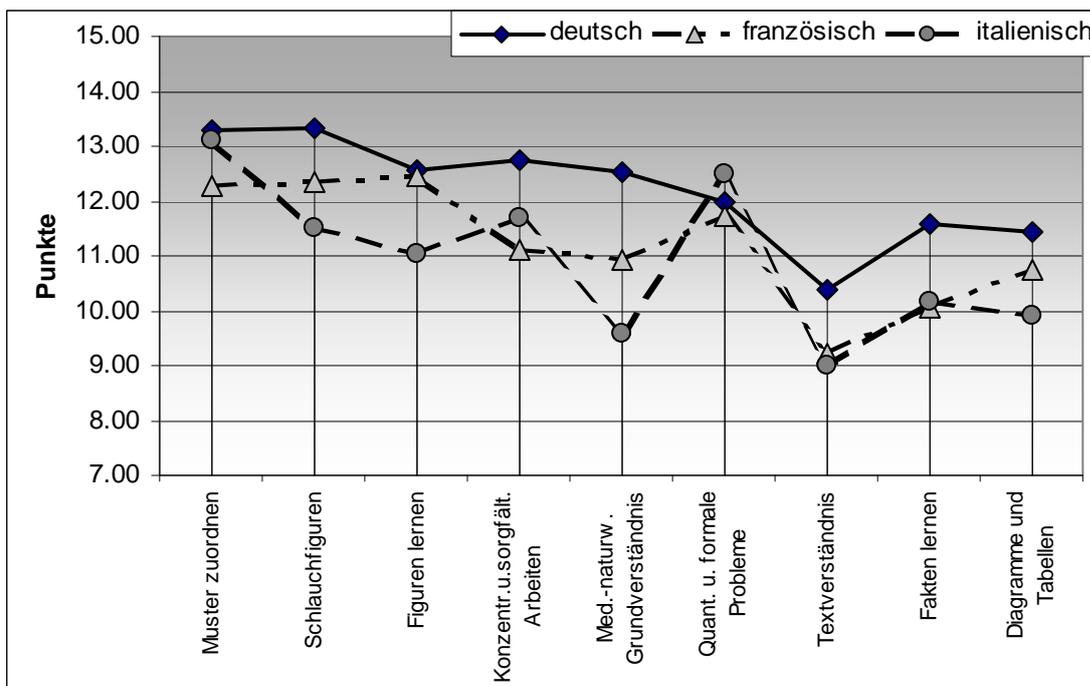


Abbildung 20. Mittelwerte der Untertests für die Sprachgruppen – Vergleichswerte aus 1998 (unkorrigiert).

		Mittelwert	Stand.- abweichung	Minimum	Maximum
Testwert** d > f	deutsch	100.7	9.9	73	131
	französisch	94.1	9.1	73	120
	italienisch	97.5	8.4	81	111
Muster zuordnen** d > f	deutsch	11.6	2.9	2	20
	französisch	10.3	3.1	3	17
	italienisch	10.9	2.7	6	15
Schlauchfiguren**	deutsch	12.8	3.7	1	20
	französisch	11.6	4.1	2	20
	italienisch	11.7	3.5	4	18
Figuren lernen**	deutsch	12.6	3.6	3	20
	französisch	11.4	3.9	2	20
	italienisch	11.5	4.1	3	19
Konzentr. u. sorgfält. Arb.** d, i > f	deutsch	12.6	3.3	4	20
	französisch	9.7	3.4	0	20
	italienisch	11.4	2.6	6	16
Med.-naturwiss. Grundverst.** d > f	deutsch	11.2	3.6	2	20
	französisch	8.9	3.3	3.58	18.30
	italienisch	10.3	3.3	4	19
Quantit. und form. Probleme i > f	deutsch	11.6	3.8	1	20
	französisch	10.8	3.7	4.66	19
	italienisch	12.6	3.3	5	20
Textverständnis**	deutsch	8.8	3.3	1	18
	französisch	7.7	2.4	3.51	12.51
	italienisch	8.1	2.6	1.95	13.77
Fakten lernen	deutsch	11.8	3.7	1	20
	französisch	11.0	3.4	3.68	18.46
	italienisch	10.8	3.3	5	19
Diagramme und Tabellen** d > f	deutsch	10.5	3.1	1	19
	französisch	8.9	3.2	1.55	18
	italienisch	9.6	2.4	5.83	17.59

Tabelle 16. Punktwert und Untertests für die Sprachgruppen (deskriptive Statistiken)
 **: signifikanter Einfluss des Faktors "Sprache" bei varianzanalytischer Prüfung; darunter
 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 (Different 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 nachfolgenden Tabellen zeigen, dass in den französischen Sprachgruppe alle 5 möglichen Tests ausgleichend werden müssen – in der italienischsprachigen nur drei.

Untertest	Sprache	N	M	s	Sig.	Ausgleich
Französischsprachige Form						
Med.-naturwiss. Grundverst.	d	748	11.26	3.61	.000	Ja
	f	76	8.62	3.45		
Quantit. und form. Probleme	d	748	11.62	3.88	.011	Ja
	f	76	10.42	4.01		
Textverständnis	d	748	8.88	3.39	.000	Ja
	f	76	7.38	2.52		
Fakten lernen	d	748	11.81	3.79	.001	Ja
	f	76	10.36	3.57		
Diagramme und Tabellen	d	748	10.55	3.12	.000	Ja
	f	76	8.59	3.34		

Tabelle 17. Signifikanzprüfung deutsche vs. italienische Sprachform.

Untertest	Sprache	N	M	s	Sig.	Ausgleich
Italienischsprachige Form						
Med.-naturwiss. Grundverst.	d	748	11.26	3.61	.185	Nein
	i	31	10.39	3.32		
Quantit. und form. Probleme	d	748	11.62	3.88	.135	Nein
	i	31	12.68	3.37		
Textverständnis	d	748	8.88	3.39	.031	Ja
	i	31	7.55	2.78		
Fakten lernen	d	748	11.81	3.79	.176	Nein
	i	31	10.87	3.34		
Diagramme und Tabellen	d	748	10.55	3.12	.000	Ja
	i	31	8.06	2.53		

Tabelle 18. Signifikanzprüfung deutsche vs. französische 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 21). 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.

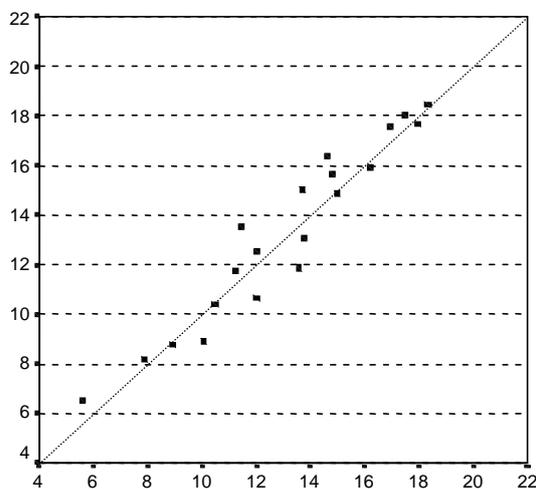


Abbildung 21. Beispiel für einen Delta-Plot.

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

Tabelle 19 zeigt die DIF-Analyse für die französischsprachige Version des Untertests "Medizinisch-naturwissenschaftliches Grundverständnis". Für jedes Item ist die Itemschwierigkeit in den verglichenen Sprachgruppen (d und f), die Δ -Werte und die Distanz zur Regressionsgeraden (D_i) angegeben. Die letzte Spalte listet für die relevanten Items den "Ausgleichsbetrag" auf. Ein Wert von 0.3 bedeutet beispielsweise, dass allen Personen der benachteiligten Gruppe (hier die französischsprachigen), welche das betreffende Item nicht korrekt beantwortet haben, 0.3 Punkte als Sprachausgleich gutgeschrieben werden.

Im Vergleich zwischen französischsprachiger und deutscher Version weisen drei Items (Item 27, 30 und 41) DIF-Werte über 1.5 in der relevanten "Richtung" auf. Item 39 weist einen DIF-Wert von 2.99 auf, allerdings in Richtung der Diagonalen (positiver Wert). Aufgrund der Entscheidung, nicht symmetrisch auszugleichen, wird dieses Item nicht berücksichtigt.

Item	Schwierigkeit		Δ -Wert		D_i	Diff. d-f
	d	f	d	f		
25	0.81	0.71	5.72	8.04	1.41	
27	0.79	0.49	6.22	13.24	-2.31	0.30
28	0.74	0.66	7.28	9.26	1.45	
29	0.70	0.59	8.36	10.79	0.95	
30	0.71	0.43	7.96	14.46	-2.15	0.28
32	0.69	0.54	8.46	12.01	0.06	
33	0.65	0.50	9.48	12.93	-0.01	
34	0.60	0.50	10.54	12.93	0.66	
35	0.72	0.54	7.87	12.01	-0.31	
36	0.55	0.43	11.84	14.46	0.30	
38	0.58	0.46	11.00	13.85	0.24	
39	0.56	0.59	11.59	10.79	2.99	
40	0.53	0.36	12.28	16.29	-0.85	
41	0.55	0.32	11.72	17.21	-1.91	0.23
42	0.45	0.33	14.20	16.91	-0.11	
44	0.33	0.17	16.81	20.57	-1.30	
45	0.29	0.24	17.78	19.05	0.49	
46	0.36	0.28	16.16	18.13	0.18	
47	0.34	0.26	16.57	18.43	0.20	
48	0.31	0.22	17.44	19.35	0.04	
Regressionsgerade: A = 0.81 B = 5.2						

Tabelle 19. DIF-Analyse "Med.-nat. Grundverständnis" (französisch)

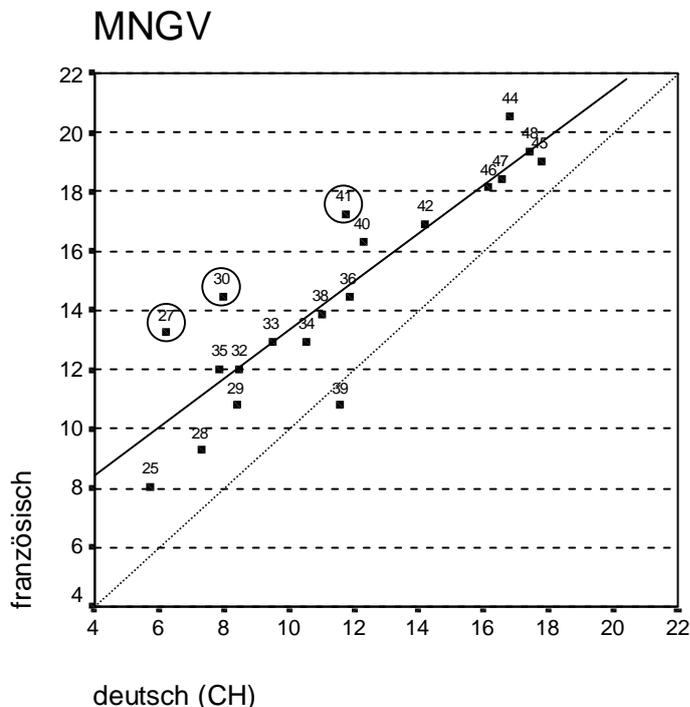


Abbildung 22 zeigt den Delta-Plot für den Untertest "MNGV". Die auszugleichenden Items sind eingekreist. Die Differenz der Itemschwierigkeiten für den Sprachvergleich deutsch – französisch betragen 0.3, 0.28 und 0.23. Je nachdem ob und wie viele dieser drei Aufgaben von einer (französischsprachigen) Person gelöst wurden, werden ihr bei diesem Untertest also maximal 0.81 Punkte gutgeschrieben.

Abbildung 22. Delta-Plot für "MNGV" (deutsch-französisch).

Quantitative und formale Probleme

Der Vergleich zwischen deutscher und französischsprachiger Version des Untertests "Quantitative und formale Probleme" zeigt vier Items mit kritischen DIF-Werten. Fünf weitere Items weisen ebenfalls Werte über 1.5 auf, allerdings in Richtung der Diagonalen.

Item	Schwierigkeit		Δ -Wert		D_i	Diff. d-f
	d	f	D	f		
73	0.79	0.68	7.57	9.92	0.08	
74	0.82	0.57	6.67	12.66	-2.64	0.25
76	0.85	0.70	6.18	9.62	-0.54	
77	0.70	0.71	9.54	9.31	1.77	
78	0.66	0.43	10.44	15.70	-2.73	0.23
79	0.70	0.70	9.45	9.62	1.47	
81	0.62	0.46	11.40	15.09	-1.66	0.16
82	0.68	0.59	9.92	12.05	-0.17	
83	0.68	0.71	10.01	9.31	2.06	
84	0.73	0.79	8.80	7.49	2.75	
85	0.59	0.47	12.17	14.79	-0.95	
87	0.65	0.47	10.75	14.79	-1.82	0.18
88	0.43	0.37	15.82	17.22	-0.64	
89	0.45	0.51	15.26	13.88	1.66	
90	0.58	0.67	12.26	10.23	2.71	
91	0.33	0.28	18.04	19.35	-0.96	
93	0.41	0.34	16.19	17.83	-0.89	
94	0.33	0.36	18.17	17.53	0.56	
95	0.34	0.36	17.76	17.53	0.31	
96	0.26	0.25	19.74	19.96	-0.40	

Regressionsgerade:
A = 0.78 B = 4.15

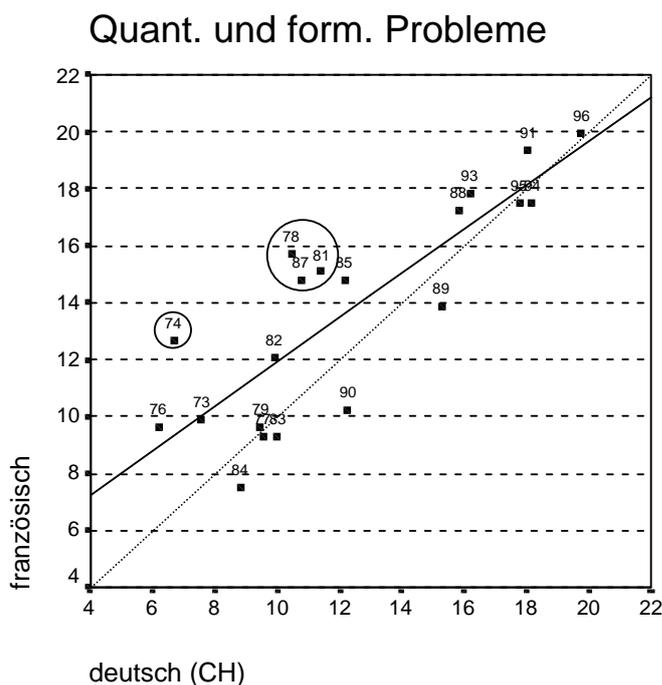


Tabelle 20. DIF-Analyse "Quantitative und formale Probleme" (französisch)

Die auszugleichenden Items sind in Abbildung 23 eingekreist dargestellt. In diesem Untertest können maximal 0.82 Punkte gutgeschrieben werden.

Abbildung 23. Delta-Plot für "Quantitative und formale Probleme" (deutsch-französisch).

Textverständnis

Der Untertest "Textverständnis" musste sowohl in der französisch- wie auch in der italienischsprachigen Version ausgeglichen werden. Tabelle 21 zeigt den Vergleich zwischen deutsch- und französischsprachiger Ausführung. Die Items 111 und 115 weisen hier hohe Abweichungen auf. Diese beiden Items sind auch in Abbildung 24 leicht zu identifizieren.

Item	Schwierigkeit		Δ-Wert		D _i	Diff. d-f
	d	f	d	f		
97	0.66	0.53	8.10	11.35	-0.57	
98	0.71	0.66	6.76	8.13	0.91	
99	0.53	0.50	11.27	11.99	1.08	
100	0.72	0.70	6.73	7.17	1.60	
101	0.53	0.41	11.24	14.24	-0.61	
102	0.35	0.32	15.65	16.49	0.69	
109	0.70	0.70	7.22	7.17	1.93	
110	0.52	0.38	11.43	14.88	-0.95	
111	0.57	0.36	10.16	15.52	-2.28	0.21
112	0.54	0.41	11.01	14.24	-0.76	
113	0.30	0.22	16.92	18.74	-0.11	
114	0.28	0.21	17.28	19.06	-0.11	
115	0.52	0.22	11.43	18.74	-3.80	0.30
116	0.65	0.58	8.23	10.06	0.47	
117	0.42	0.37	13.98	15.20	0.53	
118	0.49	0.42	12.12	13.92	0.22	
119	0.31	0.22	16.72	18.74	-0.25	
120	0.23	0.28	18.65	17.45	2.00	
Regressionsgerade: A = 0.91 B = 3.21						

Tabelle 21. DIF-Analyse "Textverständnis" (französisch)

Die Itemschwierigkeiten der betroffenen Items unterscheiden sich um 0.21, beziehungsweise 0.3 Punkte, sodass in diesem Untertest in der französischsprachigen Version maximal 0.51 Punkte gutgeschrieben werden.

Tabelle 22 (und Abbildung 25) zeigt die entsprechende Auswertung für den Vergleich zwischen deutscher und italienischsprachiger Version. Hier handelt es sich um vier "kritische" Items mit einer maximalen Ausgleichssumme von 0.95 Punkten.

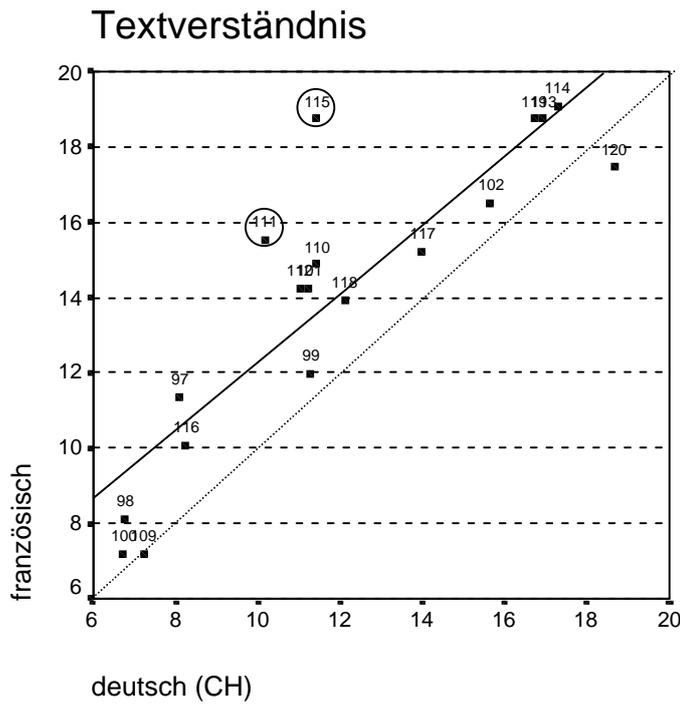


Abbildung 24. Delta-Plot für "Textverständnis" (deutsch-französisch)

Item	Schwierigkeit		Δ-Wert		D _i	Diff. d-i
	d	i	d	i		
97	0.66	0.32	8.74	16.53	-4.61	0.34
98	0.71	0.48	7.47	12.80	-2.75	0.23
99	0.53	0.35	11.74	15.78	-2.01	0.18
100	0.72	0.74	7.44	6.82	1.60	
101	0.53	0.48	11.71	12.80	0.15	
102	0.35	0.39	15.89	15.04	1.36	
109	0.70	0.65	7.90	9.06	0.28	
110	0.52	0.58	11.90	10.56	1.91	
111	0.57	0.52	10.69	12.05	0.00	
112	0.54	0.55	11.49	11.30	1.10	
113	0.30	0.10	17.09	21.75	-2.71	0.20
114	0.28	0.16	17.43	20.26	-1.39	
115	0.52	0.61	11.90	9.81	2.46	
116	0.65	0.74	8.86	6.82	2.57	
117	0.42	0.39	14.31	15.04	0.29	
118	0.49	0.52	12.55	12.05	1.27	
119	0.31	0.23	16.91	18.77	-0.66	
120	0.23	0.26	18.73	18.02	1.13	
Regressionsgerade: A = 0.93 B = 2.06						

Tabelle 22. DIF-Analyse "Textverständnis" (italienisch)

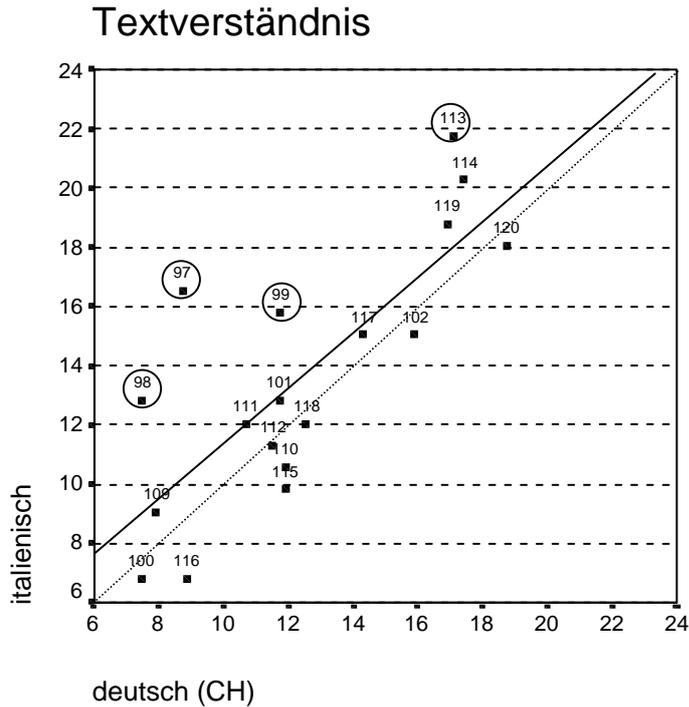
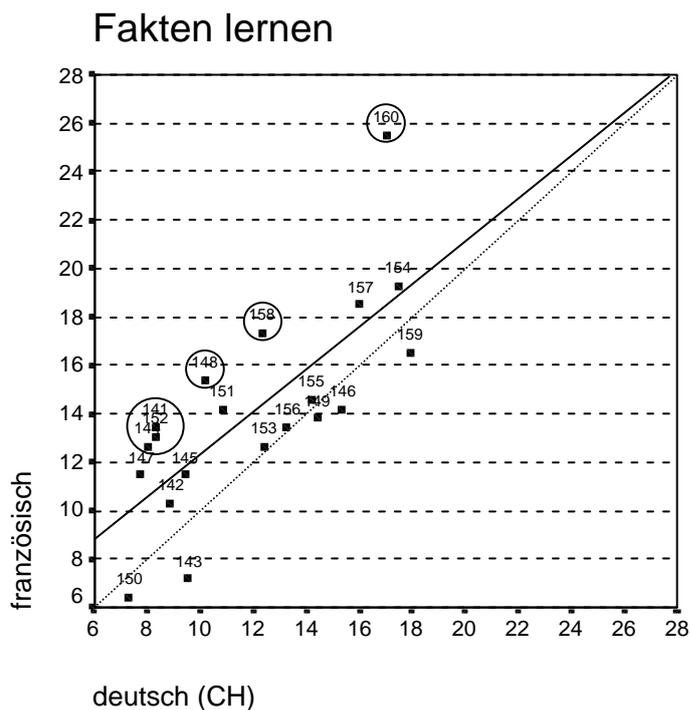


Abbildung 25. Delta-Plot für "Textverständnis" (deutsch-italienisch).

Fakten lernen

Bezüglich des Untertests "Fakten lernen" zeigte sich nur im Vergleich deutsch - französischsprachig ein signifikanter Mittelwertsunterschied. Aus Tabelle 23 ist ersichtlich, dass sechs Items dieses Untertests ausgeglichen werden müssen.

Abbildung 26 stellt den Sachverhalt grafisch dar.



In diesem Untertest können den französischsprachigen Teilnehmern bis maximal 1.12 Punkte gutgeschrieben werden.

Abbildung 26. Delta-Plot für "Fakten lernen" (deutsch-französisch).

Item	Schwierigkeit		Δ -Wert		D_i	Diff.
	d	f	d	f		
141	0.71	0.54	8.28	13.42	-1.92	0.17
142	0.69	0.64	8.84	10.30	0.79	
143	0.67	0.75	9.51	7.17	3.59	
144	0.72	0.57	8.01	12.64	-1.51	0.15
145	0.68	0.61	9.40	11.47	0.28	
146	0.47	0.51	15.35	14.20	2.15	
147	0.73	0.61	7.73	11.47	-0.82	
148	0.65	0.47	10.15	15.37	-2.16	0.18
149	0.51	0.53	14.40	13.81	1.82	
150	0.75	0.78	7.29	6.39	2.72	
151	0.63	0.51	10.82	14.20	-0.83	
152	0.71	0.55	8.28	13.03	-1.62	0.16
153	0.57	0.57	12.41	12.64	1.39	
154	0.40	0.34	17.45	19.28	-0.29	
155	0.51	0.50	14.20	14.59	1.10	
156	0.55	0.54	13.24	13.42	1.35	
157	0.45	0.37	16.02	18.50	-0.64	
158	0.58	0.41	12.33	17.33	-2.19	0.17
159	0.39	0.43	17.93	16.55	2.08	
160	0.42	0.13	17.01	25.53	-5.27	0.29
Regressionsgerade: A = 0.88 B = 3.61						

Tabelle 23. DIF-Analyse "Fakten lernen" (französisch)

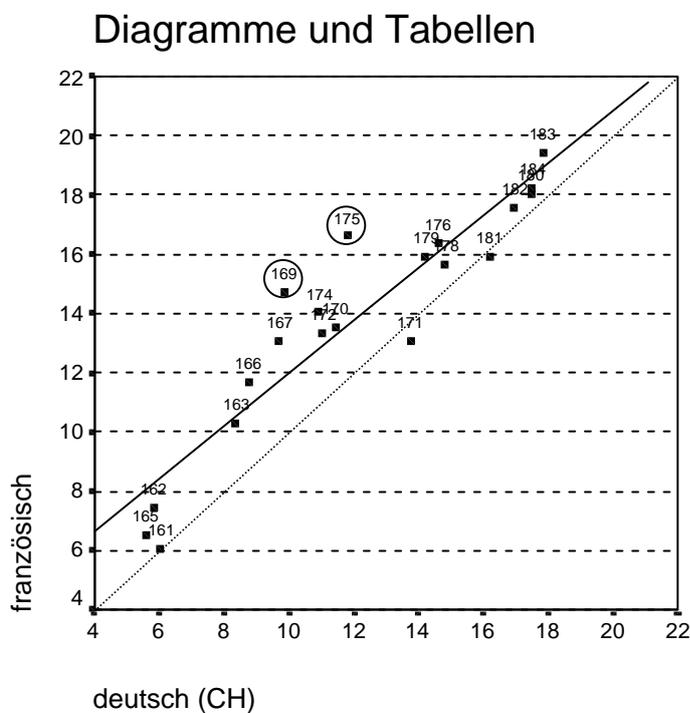
Diagramme und Tabellen

Dieser Untertest musste wiederum für beide Sprachversionen ausgeglichen werden. Tabelle 24 (und Abbildung 27) zeigt die Daten für die französischsprachige Gruppe. Die Items 169 und 175 weisen hier Distanzen über 1.5 DIF von der Regressionsgeraden (in der relevanten Richtung) auf.

Item	Schwierigkeit		Δ -Wert		D_i	Diff.
	d	f	d	f		
161	0.87	0.87	6.03	6.04	1.82	
162	0.88	0.79	5.81	7.45	0.63	
163	0.74	0.63	8.35	10.27	0.19	
165	0.89	0.84	5.58	6.51	1.17	
166	0.72	0.55	8.75	11.68	-0.59	
167	0.66	0.47	9.68	13.09	-1.03	
169	0.66	0.38	9.85	14.73	-2.15	0.28
170	0.57	0.45	11.43	13.56	-0.23	
171	0.44	0.47	13.74	13.09	1.66	
172	0.59	0.46	11.02	13.32	-0.32	
174	0.60	0.42	10.90	14.03	-0.93	
175	0.55	0.28	11.78	16.61	-2.27	0.27
176	0.39	0.29	14.60	16.38	-0.23	
178	0.38	0.33	14.82	15.67	0.44	
179	0.41	0.32	14.22	15.91	-0.13	
180	0.23	0.20	17.47	18.02	0.44	
181	0.30	0.32	16.18	15.91	1.17	
182	0.26	0.22	16.94	17.55	0.45	
183	0.21	0.12	17.82	19.43	-0.38	
184	0.23	0.18	17.51	18.26	0.29	

Regressionsgerade:
A = 0.89 B = 3.13

Tabelle 24. DIF-Analyse Diagramme und Tabellen (französisch)



Je nach Beantwortung der Items werden in diesem Untertest also 0, 0.27, 0.28 oder maximal 0.55 Punkte gutgeschrieben.

Die italienischsprachige Version (Tabelle 25, Abbildung 28) weist vier "kritische" Items auf. Hier beträgt der maximale Ausgleich 1.07 Punkte.

Abbildung 27. Delta-Plot für "Diagramme und Tabellen" (deutsch-französisch).

Item	Schwierigkeit	Δ -Wert	D_i	Diff.
------	---------------	----------------	-------	-------

	d	i	d	i		d-i
162	0.88	0.94	5.46	4.45	2.64	
163	0.74	0.65	8.09	9.83	0.35	
165	0.89	0.84	5.22	6.25	1.13	
166	0.72	0.48	8.51	12.82	-1.61	0.24
167	0.66	0.42	9.47	14.01	-1.86	0.24
169	0.66	0.39	9.65	14.61	-2.19	0.27
170	0.57	0.55	11.28	11.62	1.13	
171	0.44	0.42	13.68	14.01	0.93	
172	0.59	0.68	10.86	9.23	2.64	
174	0.60	0.52	10.73	12.22	0.32	
175	0.55	0.23	11.65	17.59	-3.09	0.32
176	0.39	0.39	14.57	14.61	1.07	
178	0.38	0.32	14.79	15.80	0.33	
179	0.41	0.29	14.18	16.40	-0.53	
180	0.23	0.13	17.54	19.38	-0.53	
181	0.30	0.35	16.20	15.20	1.72	
182	0.26	0.23	17.00	17.59	0.45	
183	0.21	0.23	17.91	17.59	1.06	
184	0.23	0.29	17.59	16.40	1.74	

Regressionsgerade:
A = 0.84 B = 3.34

Tabelle 25. DIF-Analyse Diagramme und Tabellen (italienisch)

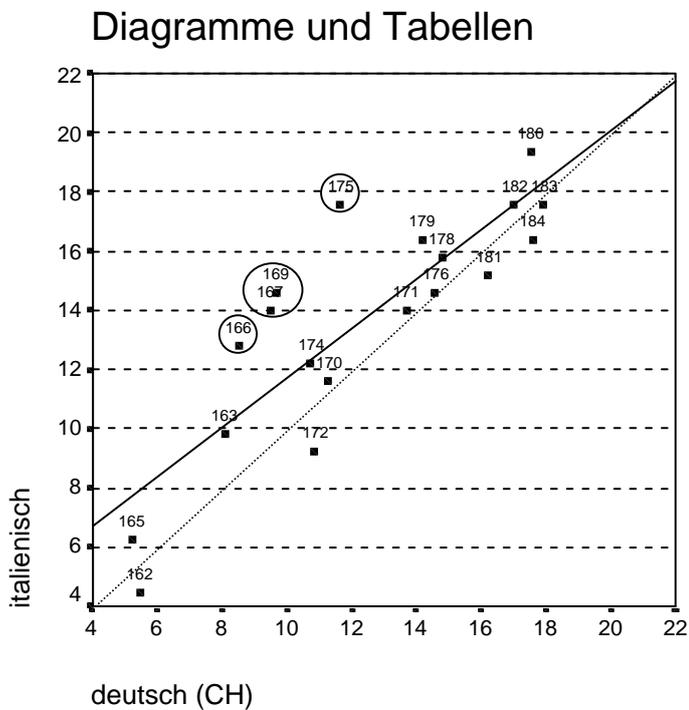


Abbildung 28. Delta-Plot für "Diagramme und Tabellen" (deutsch-italienisch).

Effekte der Korrektur

Sprache	Untertest	N	Minimum	Maximum	M	S
französisch	Med.-naturwiss, Grundv.	76	.00	.58	.31	.21
	Quantit. und form. Probleme	76	.00	.82	.41	.27
	Textverständnis	76	.00	.51	.36	.17
	Fakten lernen	76	.17	1.12	.66	.23
	Diagramme und Tabellen	76	.00	.55	.36	.20
	GESAMT		76	.82	3.23	2.13
italienisch	Textverständnis	31	.20	.95	.64	.24
	Diagramme und Tabellen	31	.00	1.07	.67	.26
	GESAMT	31	.20	2.02	1.32	.42

Tabelle 26. Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppe.

Die Korrekturwerte wurden auf ganze Punkte gerundet. Somit wurden allen französischsprachigen Kandidatinnen und Kandidaten zwischen 1 und 3 Punkten gutgeschrieben – den italienischsprachigen zwischen 0 und 2.

Evaluation: Gleichbehandlung der Sprachgruppen

Im Unterschied zu 1998 zeigen sich in diesem Jahr signifikante Unterschiede bei der Zulassung für Humanmedizin für die Sprachgruppen (Tabelle 27). Bei Veterinärmedizin sind die Unterschiede nicht signifikant – die geringe Stichprobengrösse wirkt sich hier allerdings aus. Vor allem die französische Sprachgruppe schneidet deutlich schlechter ab als die anderen beiden Gruppen.

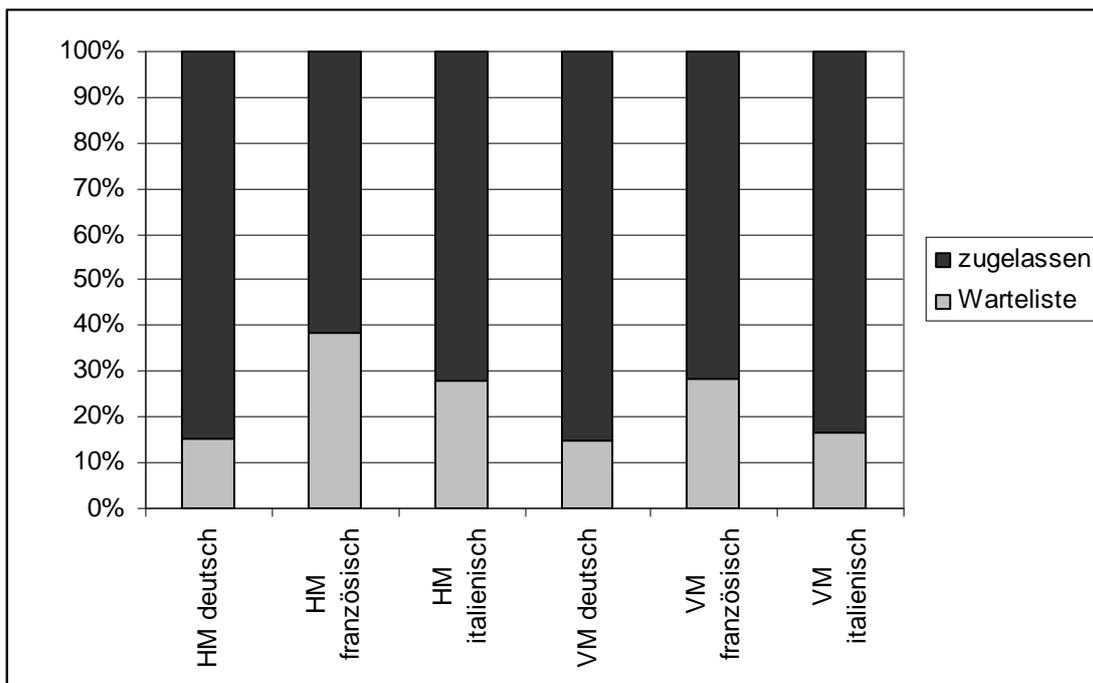
Bevor vorschnell der Schluss gezogen wird, dass es sich hier um keine Gleichbehandlung handelt, muss man die Gruppen genauer analysieren. Es handelt sich mit ziemlicher Sicherheit nicht um testbedingte Unterschiede.

- Die französische Sprachgruppe schneidet 1999 deutlich schlechter ab als 1998 (siehe Punktwerte für gleiche Aufgaben für 1998 und 1999), während sich die deutsch- und die italienischsprachige Gruppe verbessern.
- Die Unterschiede bei sprachabhängigen Tests sind nach der Korrektur tendenziell sogar geringer als bei den nicht sprachabhängigen Tests.

Da es sehr unwahrscheinlich ist, dass sich innerhalb eines Jahres die Leistungen einer ganzen Sprachgruppe derart verändern und dass die Unterschiede auf den Test zurückzuführen sind (weil auch alle sprachunabhängigen Untertests betroffen sind), liegt als Ursache die **Änderung der Zusammensetzung der Bewerbergruppe in Freiburg** an erster Stelle. Bereits bei der Diskussion der demografischen Daten deutete sich das an.

Humanmedizin			
	Zulassung		Total
	Warteliste	zugelassen	
deutsch	95 15,5%	519 84,5%	614 100,0%
französisch	13 38,2%	21 61,8%	34 100,0%
italienisch	7 28,0%	18 72,0%	25 100,0%
Total	115 17,1%	558 82,9%	673 100,0%
Veterinärmedizin			
deutsch	20 14,9%	114 85,1%	134 100,0%
französisch	12 28,6%	30 71,4%	42 100,0%
italienisch	1 16,7%	5 83,3%	6 100,0%
Total	33 18,1%	149 81,9%	182 100,0%

Tabelle 27. Prüfung Homogenität Zulassung nach Sprache des Tests



Die Abbildung 29 zeigt, dass die Zulassungsquote für Humanmedizin der französischen Sprachgruppe am niedrigsten ist – gefolgt von Humanmedizin in der italienischen Sprachgruppe.

Disziplin		Value	df	Asymp. Sig. (2-sided)
Humanmedizin	Pearson Chi-Square	13.965	2	.001
	Likelihood Ratio	11.568	2	.003
	Linear-by-Linear Association	9.126	1	.003
	N of Valid Cases	673		
Veterinärmedizin	Pearson Chi-Square	4.020	2	.134
	Likelihood Ratio	3.713	2	.156
	Linear-by-Linear Association	2.375	1	.123

Abbildung 29. Zulassungsquoten nach Fächern und Sprache

	N of Valid Cases	182		
--	------------------	-----	--	--

Tabelle 28. Prüfung auf Verteilungsunterschiede für Sprachgruppen getrennt für Disziplinen: signifikanter Unterschied für Humanmedizin, kein signifikanter Unterschied für Veterinärmedizin (aber geringe Stichprobengröße).

Eine bereits im Vorjahr diskutierte Frage war die nach dem Verhältnis der Unterschiede in sprachabhängigen und wenig sprachabhängigen Tests. Aufgrund des Faktormodells, des Generalfaktors mit sehr hohem Varianzanteil, ist zu erwarten, dass die Fähigkeiten in beiden Bereichen 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 ½ Standardabweichung schlechter ist als die deutschsprachige Referenzgruppe.

In den beiden nachfolgenden Abbildungen sind die sprachunabhängigen Tests wiederum zuerst dargestellt. Die Abbildung 30 zeigt, dass 1999 in der französischen Sprachgruppe bis auf "Fakten lernen" in jedem Untertest grössere Unterschiede als 1998 vorhanden sind. Beim sprachunabhängigen Konzentrationstest treten die grössten Unterschiede auf.

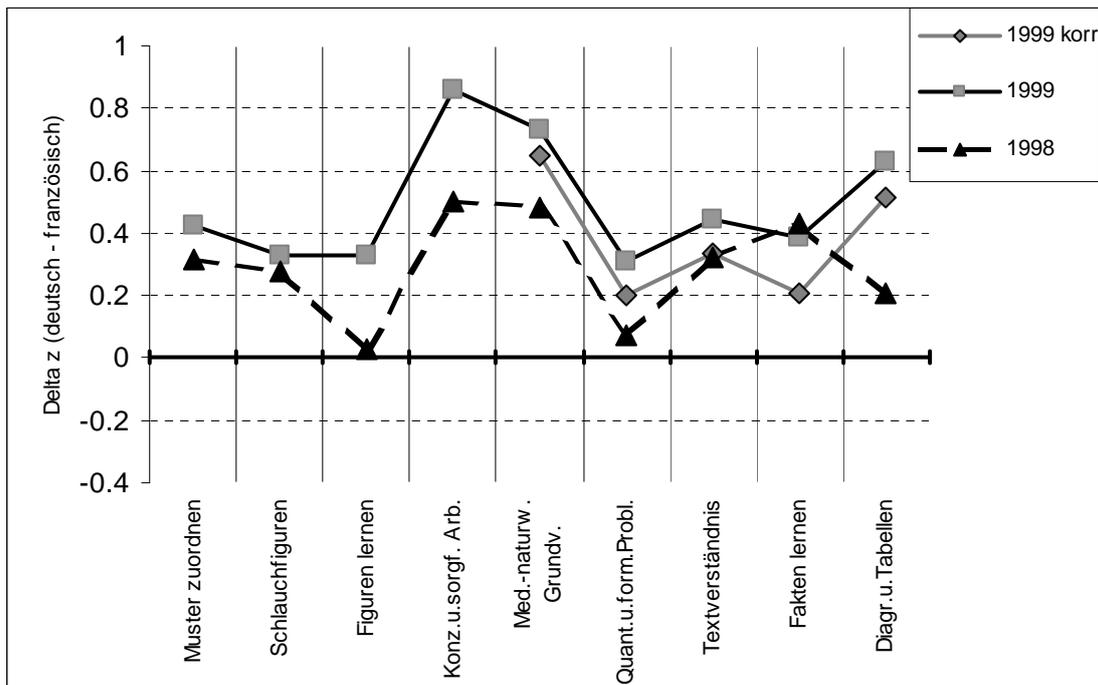


Abbildung 30. Abweichung (Delta z) der deutschen und französischen Sprachgruppe für die Untertests.

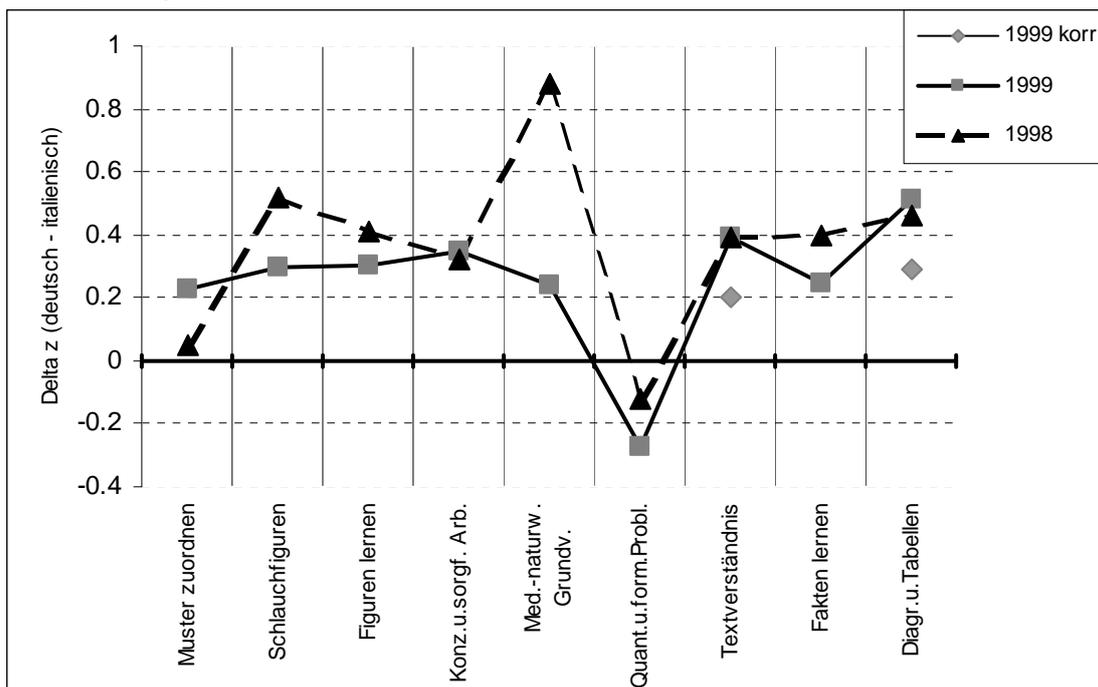


Abbildung 31. Abweichung (Delta z) der deutschen und italienischen Sprachgruppe für die Untertests.

In der italienischen Sprachgruppe zeigt sich 1999 ein ausgeglichenes Bild – die Abweichung ist in allen Untertests etwa gleich. 1998 war beim Untertest "Medizinisch-naturwissenschaftliches Grundverständnis" ein sehr grosser Unterschied vorhanden, der sich nicht wiederholt. In beiden Jahren wurden die besten Leistungen bei "Quantitativen und formalen Problemen" erzielt. Dies scheint ein stabiler Effekt zu sein – der Test selbst ist im Übrigen wegen der Verwendung von Formeln und zumeist kurzen Sätzen auch nicht so anfällig für Sprachunterschiede.

Die Delta-z-Werte können für die Untertests auch gemittelt werden. Wir tun dies in den Abbildungen 30 und 31 getrennt für sprachabhängige und wenig sprachabhängige Tests. Dies liefert die eigentliche Bewertung des Korrekturverfahrens. Wiederum ist vorauszusetzen, dass die Unterschiede in beiden Bereichen etwa gleich sein sollten, weil ein bedeutender allgemeiner Leistungsfaktor nachgewiesen wurde.

- 1999 wichen in der französischen Sprachgruppe die Leistungen in sprachabhängigen Tests mehr von der deutschen Sprachgruppe ab als die weniger sprachabhängigen. Nach der Korrektur sind die sprachabhängigen Tests weniger abweichend als die weniger sprachabhängigen. Das Korrekturverfahren hat eine mögliche Benachteiligung in den sprachabhängigen Tests mit ziemlicher Sicherheit ausgeschlossen.
- 1999 zeigen die Unterschiede in der italienischen Sprachgruppe bereits vor der Korrektur keine deutlichen Differenzen für sprachabhängige Tests (hier gehen die besseren Leistungen bei "Quantitativen und formalen Problemen" mit ein). Das Korrekturverfahren erhöht die Sicherheit ebenfalls, dass keine Benachteiligung durch testbedingte Unterschiede auftreten.
- Es wird auch deutlich, dass die Unterschiede in der französischen Sprachgruppe 1998 in beiden Bereichen wesentlich geringer waren. In der italienischen Gruppe trat vor allem in den sprachabhängigen Tests ein grösserer Unterschied als bei den weniger sprachabhängigen auf.

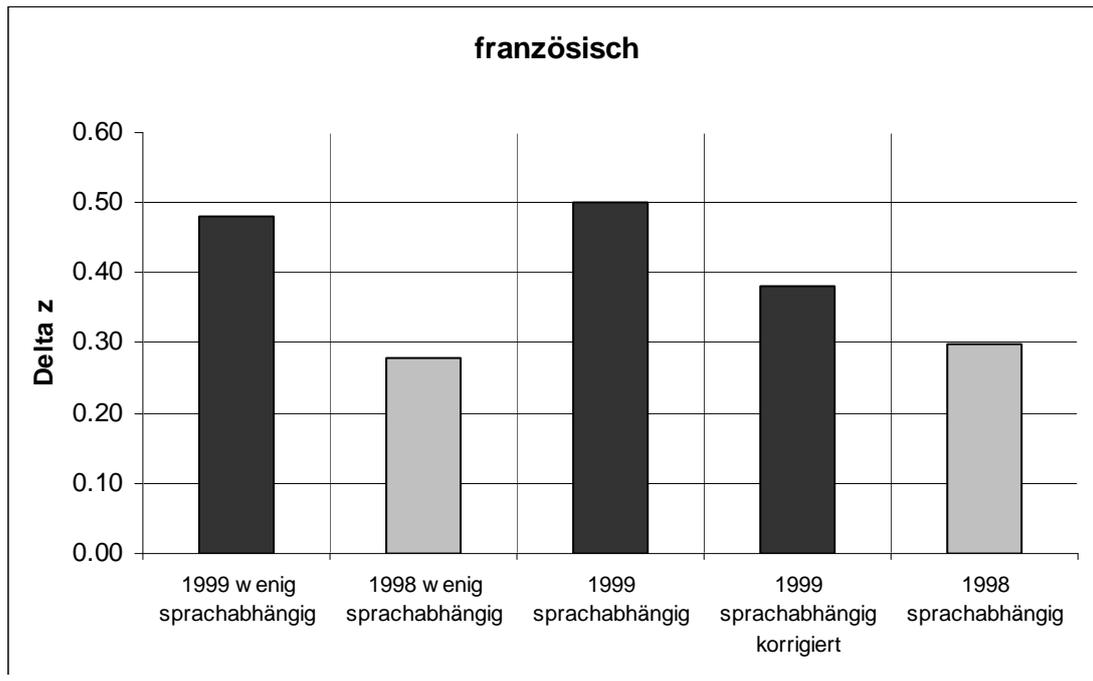


Abbildung 32. Abweichung (Delta z) für den Testwert zwischen deutscher und französischer Sprachgruppe.

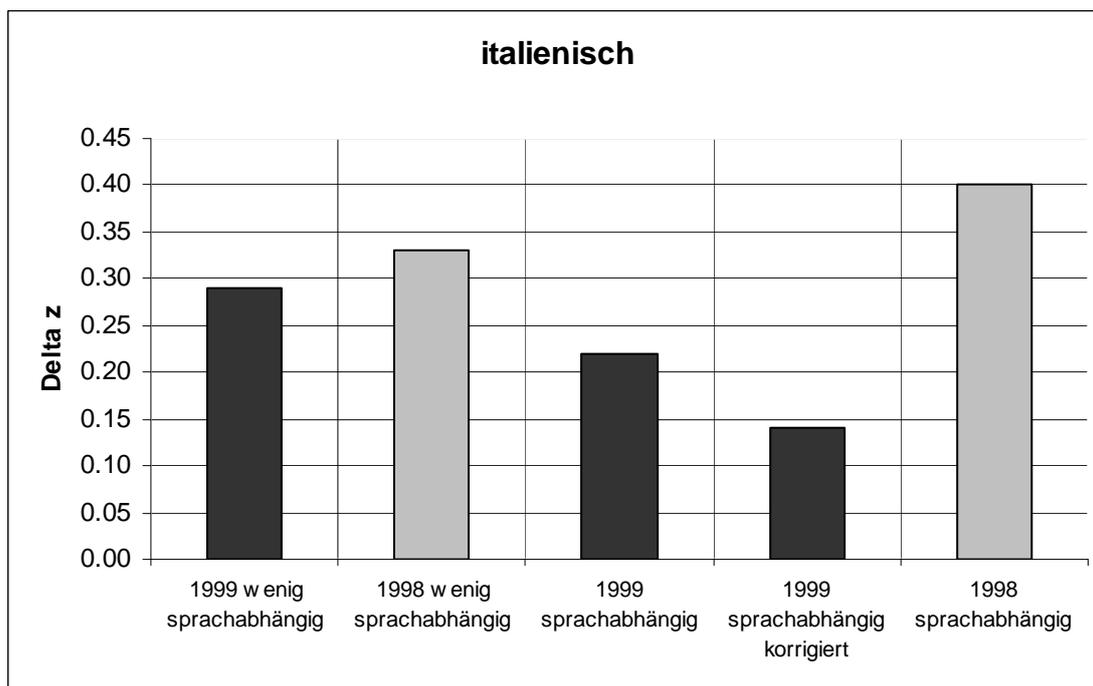


Abbildung 33. Abweichung (Delta z) für den Testwert zwischen deutscher und italienischer Sprachgruppe.

Ein weiteres Argument, dass die schlechteren Leistungen der französischsprachigen Gruppe nicht testbedingt sind, liefert die nachfolgende Analyse:

Aus technischen Gründen wurden im Test 1999 in jeder der beiden Testformen 22 Aufgaben des Tests von 1998 als Einstreuaufgaben erneut verwendet. Dies betrifft die Untertests "Muster zuordnen", "Medizinisch-naturwissenschaftliches Grundverständnis",

"Quantitative und formale Probleme", "Textverständnis" und "Diagramme und Tabellen". Dies geschah auch mit dem Ziel, die Stabilität der Leistungen zu untersuchen. Die Befunde unterstreichen, dass besonders die französischsprachige Gruppe mit der des Vorjahres nicht vergleichbar ist: deren Leistungen nehmen im Gegensatz zu den anderen beiden Gruppen deutlich ab. Erwartet hätten wir auch hier den Leistungsanstieg der beiden anderen Sprachgruppen, der vermutlich v.a. auf eine bessere Vorbereitung auf den Test mit den veröffentlichten Materialien zurückzuführen ist – die Erfahrungen aus 1998 haben den Test und die Notwendigkeit der Vorbereitung deutlicher gemacht als es bei der erstmaligen Durchführung war.

	1998		1999			
	Humanmedizin		Humanmedizin		Veterinärmedizin	
	<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>
deutsch	13.14	3.52	13.41	3.71	13.18	3.80
französisch	12.09	3.22	11.29	3.97	11.40	4.40
italienisch	12.03	3.25	12.24	3.02	15.00	2.83

Tabelle 29. Mittelwerte und Standardabweichungen (in Klammern) der gleichen Einstreuaufgaben 1998 und 1999.

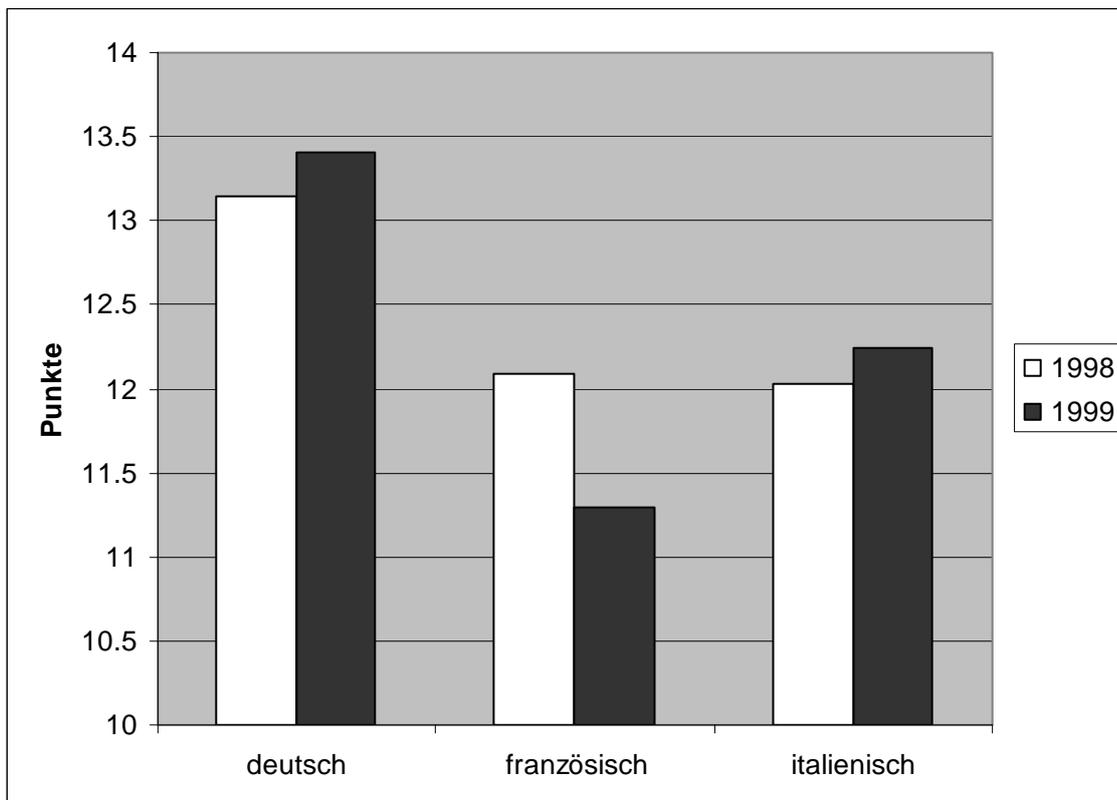


Abbildung 34. Punkte in den 1998 und 1999 gleichen Aufgaben für die Sprachgruppen.

Vergleichbarkeit der Testlokale

Gleiche Durchführungsbedingungen in den einzelnen Testorten sind eine wichtige Voraussetzung für Chancengleichheit. 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 30 stellt die Testwerte aller Testlokale gegenüber. Die Varianzanalyse zeigt, dass Verteilungsunterschiede beim Testwert auf 5%-Niveau vorhanden sind – der multiple Mittelwertvergleich (Tabelle 31) zeigt allerdings keine signifikanten Unterschiede – alle Mittelwerte des Testwertes für die Testlokale sind homogen. Auffällig ist, dass in beiden Testlokalen in Basel die niedrigsten Werte auftreten. Hier zeigt sich offenbar eine Konfundierung mit regionalen Unterschieden, die weiter unten diskutiert werden.

	N	Mittelwert	Stand.-abweichg	Minimum	Maximum
Basel 1	44	98.2	10.4	76	127
Basel 2	98	97.8	10.0	73	120
Bern 1	78	100.0	10.6	75	120
Bern 2	88	100.2	9.3	82	125
Zürich 1	115	101.1	9.6	77	128
Zürich 2	39	102.6	8.7	80	121
Zürich 3	47	102.6	11.0	79	126
Zürich 4	48	101.7	10.7	82	131
Zürich 5	102	101.7	9.7	74	122
Chur	34	101.2	7.6	89	115
St.Gallen	55	102.5	10.0	81	122

Tabelle 30. Statistiken für den Testwert nach Testlokalen (nur deutsche Sprachgruppe).

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.

		Quadrat- summe	df	Mittl. quadr. Abweichun- g	F	Sig.
Testwert	Zwischen Gruppen	1861.629	10	186.163	1.896	.043
	Innerhalb Gruppen	72351.414	737	98.170		
	Total	74213.043	747			

Tabelle 31. Varianzanalytische Prüfung des Einflusses des Faktors „Testlokal“ auf den Testwert: Signifikanz auf 4%-Niveau.

Testlokal	N	Untergruppen für alpha = .05
Basel 2	98	97.8
Basel 1	44	98.2
Bern 1	78	100.0
Bern 2	88	100.2
Zürich 1	115	101.1
Chur	34	101.2
Zürich 5	102	101.7
Zürich 4	48	101.7
St.Gallen	55	102.5
Zürich 3	47	102.6
Zürich 2	39	102.6
Sig.		.225

Tabelle 32. Multipler Mittelwertsvergleich zwischen den Testlokalen – alle Mittelwerte sind homogen.

Bei den beiden Untertests mit varianzanalytisch geführtem Nachweis eines Einflusses des Faktors "Testlokal" ("Quantitative und formale Probleme", "Fakten lernen"; vgl. Tabelle 33) bleiben die Mittelwerte homogen. Dies spricht gegen eine systematische "Verschiebung" der Verteilung – etwa, wenn Arbeitszeiten nicht genau eingehalten worden wären. Zwei homogene Mittelwertgruppen lassen sich dagegen bei "Medizinisch-naturwissenschaftlichem Grundverständnis" nachweisen. Aufgrund der Konstellationen scheinen allerdings auch hierfür vor allem die regionalen Unterschiede verantwortlich zu sein.

		Quadrat- summe	df	Mittl. quadr. Abweich- ung	F	Sig.
Muster zuordnen	Zwischen Gruppen	121.574	10	12.157	1.393	.179
	Innerhalb Gruppen	6434.323	737	8.730		
	Total	6555.897	747			
Schlauchfiguren	Zwischen Gruppen	140.234	10	14.023	1.002	.440
	Innerhalb Gruppen	10309.632	737	13.989		
	Total	10449.866	747			
Figuren lernen	Zwischen Gruppen	175.990	10	17.599	1.356	.196
	Innerhalb Gruppen	9562.454	737	12.975		
	Total	9738.444	747			
Konzentr. u. sorgfält. Arb.	Zwischen Gruppen	136.191	10	13.619	1.246	.258
	Innerhalb Gruppen	8053.706	737	10.928		
	Total	8189.897	747			
Med.-naturwiss. Grundverst.	Zwischen Gruppen	183.187	10	18.319	1.411	.171
	Innerhalb Gruppen	9567.455	737	12.982		
	Total	9750.642	747			
Quantit. und form. Probleme	Zwischen Gruppen	271.193	10	27.119	1.823	.053
	Innerhalb Gruppen	10960.978	737	14.872		
	Total	11232.171	747			
Textverständnis	Zwischen Gruppen	130.906	10	13.091	1.141	.328
	Innerhalb Gruppen	8452.266	737	11.468		
	Total	8583.171	747			
Fakten lernen	Zwischen Gruppen	265.575	10	26.557	1.875	.045
	Innerhalb Gruppen	10436.703	737	14.161		
	Total	10702.278	747			
Diagramme und Tabellen	Zwischen Gruppen	79.394	10	7.939	.815	.614
	Innerhalb Gruppen	7181.359	737	9.744		
	Total	7260.753	747			

Tabelle 33. Varianzanalytische Prüfung des Einflusses des Faktors „Testlokal“ auf die Punktwerte der Untertests.

Testlokal	N	Untergruppen für alpha = .05 1	2
Basel 1	44	10.2	
Chur	34	10.5	10.5
Basel 2	98	10.7	10.7
Bern 1	78	11.0	11.0
Bern 2	88	11.2	11.2
St.Gallen	55	11.3	11.3
Zürich 1	115	11.3	11.3
Zürich 4	48	11.4	11.4
Zürich 5	102	11.6	11.6
Zürich 2	39	11.8	11.8
Zürich 3	47		12.4
Sig.		.378	.153

Tabelle 34. Multipler Mittelwertsvergleich für "Med.-naturwiss. Grundverständnis": zwei homogene Mittelwertgruppen.

	N	Untergruppen für alpha = .05 1
Basel 2	98	10.2
Basel 1	44	10.8
Freiburg	76	10.8
Bern 1	78	11.5
Chur	34	11.6
Zürich 2	39	11.7
Bern 2	88	11.8
Zürich 1	115	11.8
St.Gallen	55	11.9
Zürich 5	102	12.0
Zürich 4	48	12.1
Zürich 3	47	12.1
Bellinzona	31	12.6
Sig.		.057

Tabelle 35. Multipler Mittelwertsvergleich für "Quantitative und form. Probleme": eine homogene Mittelwertgruppe.

Testlokal	N	Untergruppen für alpha = .05 1
Basel 2	98	10.8
Bellinzona	31	10.8
Freiburg	76	11.0
Bern 1	78	11.2
Zürich 3	47	11.3
Bern 2	88	11.4
Basel 1	44	11.6
Zürich 4	48	11.7
St.Gallen	55	12.1
Zürich 1	115	12.3
Zürich 5	102	12.3
Zürich 2	39	12.5
Chur	34	12.8
Sig.		.202

Tabelle 36. Multipler Mittelwertsvergleich für "Fakten lernen": eine homogene Mittelwertgruppe.

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. Die Definition, was Testfairness ist, spielt dabei allerdings eine wichtige Rolle. Eine Vergabe der Plätze entsprechend der Bewerberquoten würde dann „unfair“ sein, wenn unterschiedliche Bewerberquoten Einfluss auf die Verteilung der Testleistung hätten (vgl. die Diskussion zu den Sprachgruppen). Auch die Erwartung, dass Männer und Frauen numerisch gleiche Testwerte im Mittel erreichen müssen, wäre kein Fairnesskriterium: Es wurde mehrfach empirisch nachgewiesen, dass Frauen im statistischen Notendurchschnitt die Medizinprüfungen etwas schlechter abschliessen als Männer (vgl. Trost, 1995, für die Schweiz Hänsgen, Hofer & Ruefli, 1996). In Deutschland erreichten Frauen im Test langjährig etwa um 2 Punkte tiefere Werte als Männer. Relativ gesehen (bezogen auf die Streuungen) war der Geschlechterunterschied bei den Noten dort übrigens grösser als im Test. Wenn der Test die Realität abbilden und ein „Probestudium“ sein soll, welches die Studienleistungen gut vorhersagt, dann müsste sich dieser Unterschied im Test wiederfinden.

Bei der Bewertung der vorliegenden Ergebnisse muss auch berücksichtigt werden, dass sich mehr Frauen zum Studium beworben haben als Männer. Damit könnten sich unterschiedliche Quoteneffekte auf die Testleistungen auswirken. Besonders bei „Veterinärmedizin“ unterscheiden sich diese Quoten beträchtlich. 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
	Mittelwerte					
männlich	101.89	95.67	97.54	98.81	95.53	(96.50)
weiblich	100.39	92.21	96.92	99.21	93.96	(100.00)
	Standardabweichungen					
männlich	9.78	12.17	7.57	9.34	8.36	(19.09)
weiblich	10.39	8.13	9.41	8.98	8.41	(4.69)
	Stichprobengrößen					
männlich	278	15	13	27	15	2 (!)
weiblich	336	19	12	107	27	4 (!)

Tabelle 37. 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. (2-seitig)
deutsch	Humanmedizin	Varianzen homogen	.265	.607	1.831	612	.068
	Veterinärmedizin	Varianzen homogen	.000	.999	-.200	132	.841
französisch	Humanmedizin	Varianzen homogen	2.795	.104	.991	32	.329
	Veterinärmedizin	Varianzen homogen	.092	.763	.581	40	.564
italienisch	Humanmedizin	Varianzen homogen	1.792	.194	.183	23	.857

Tabelle 38. Prüfung Varianzhomogenität und Mittelwertvergleich (t-Test) für den Testwert zwischen den Geschlechtern, keine Signifikanz auf 5%-Niveau. Italienische Sprachgruppe Veterinärmedizin wegen geringer Fallzahl nicht berechnet.

Anhand der konkreten Ergebnisse der deutschen Sprachgruppe (Tabelle 37) kann festgestellt werden, dass sich die Testwerte für Humanmedizin bei einer Streuung von 10 nur um 1.5 unterscheiden. Im Vorjahr betrug diese Differenz des Testwertes 0.9 Dieser Unterschied ist statistisch allerdings wie im Jahre 1998 nicht signifikant, er liegt bei Annahme von 5% Irrtumswahrscheinlichkeit im Bereich des Zufalls. Die offenbar vorhandene tendenzielle Zunahme von Unterschieden entspricht allerdings der oben diskutierten Erwartung.

Im Vergleich zu Deutschland scheint der Leistungsunterschied zwischen den Geschlechtern in der Schweiz geringer zu sein. Dies wird auch durch die Ergebnisse des Probelaufes des EMS im Jahre 1995 gestützt, wo Frauen sogar bessere Werte erreichten als Männer.

	Typ III Quadratsumme	df	Mittlere quadrat. Abweichung	F	Sig.
Modell	8555107.928	4	2138776.982	21564.883	.000
Geschlecht	24.388	1	24.388	.246	.620
Disziplin	827.362	1	827.362	8.342	.004
Geschlecht x Disziplin	129.787	1	129.787	1.309	.253
Fehler	84401.072	851	99.179		
Total	8639509.000	855			

Tabelle 39. Zweifache Varianzanalyse für den Einfluss von Geschlecht und Disziplin auf den Testwert.

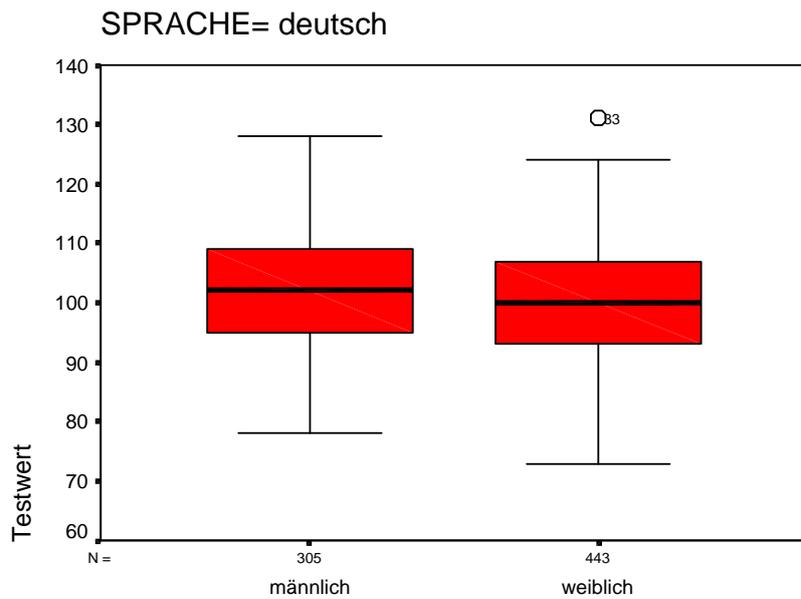
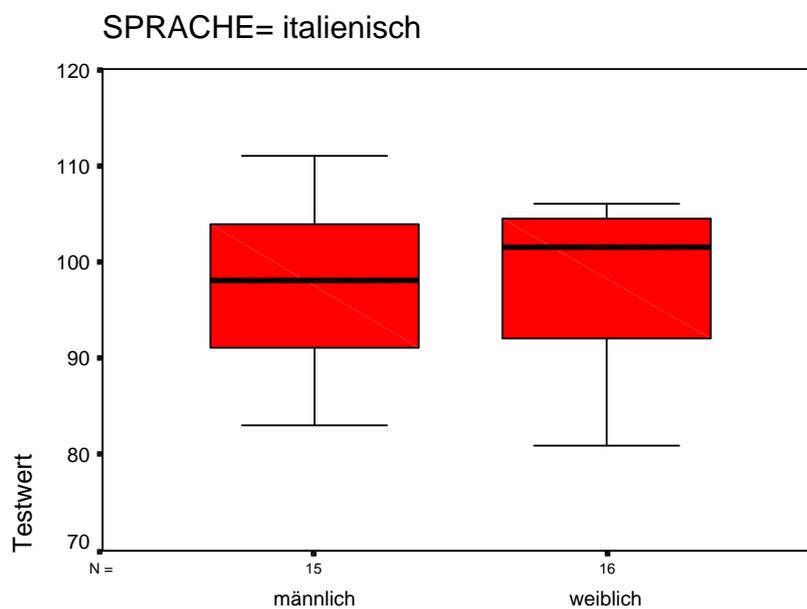
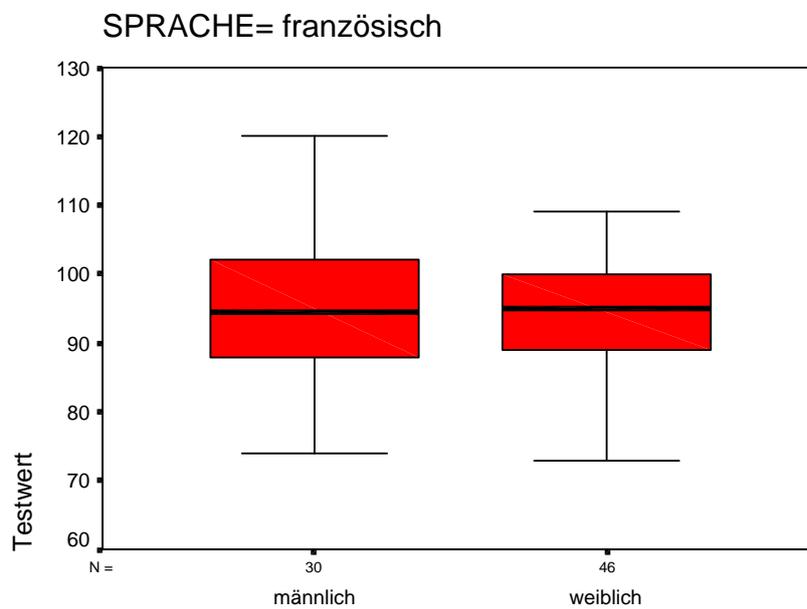
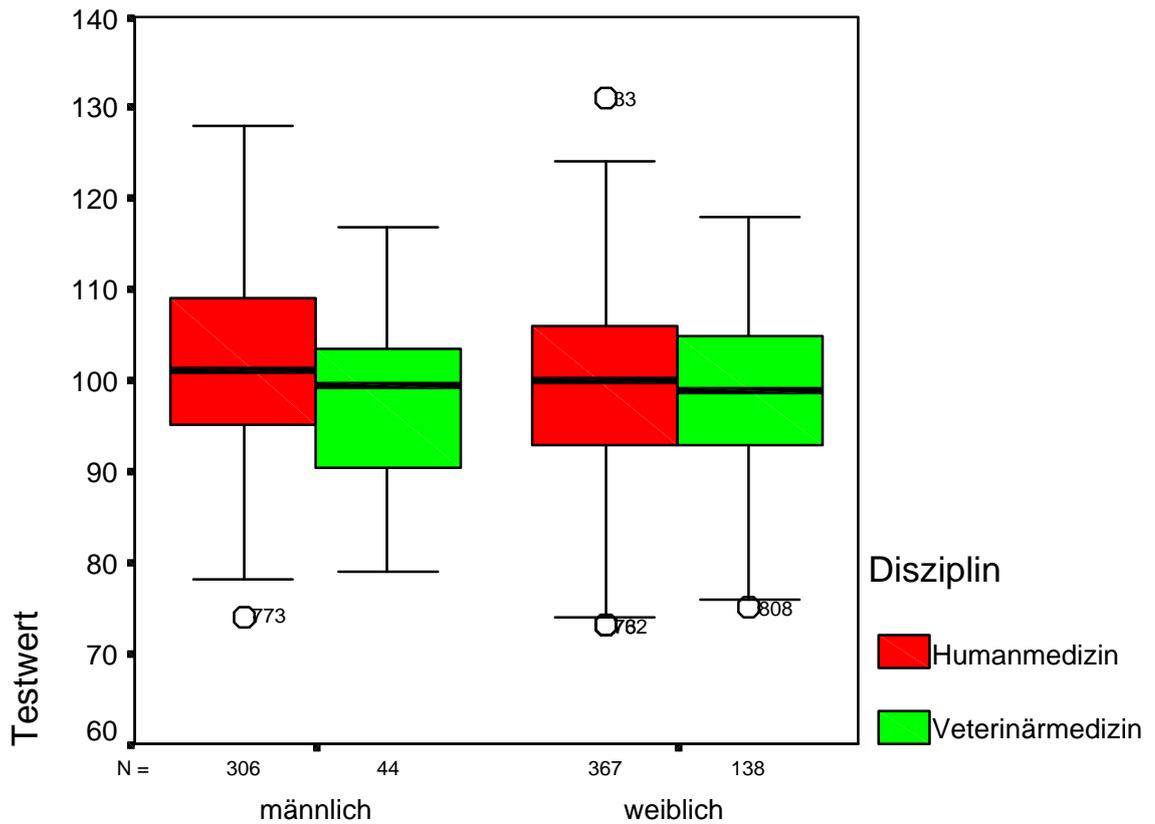


Abbildung 35. Boxplots für die Testwerte nach Geschlechtern getrennt für die drei Sprachgruppen (Werte siehe Tab. 38).



GESCHLECHT



GESCHLECHT

Abbildung 36. Boxplot für den Testwert nach Geschlecht und Disziplinen.

	deutsch		französisch		Italienisch	
	m	S	m	S	m	S
männlich	101.62	9.77	95.60	10.26	97.40	8.68
weiblich	100.10	10.07	93.24	8.25	97.69	8.44
Gesamt	100.72	9.97	94.17	9.10	97.55	8.41

Tabelle 40. Statistiken für den Testwert nach Geschlechtern und Sprachgruppen für alle Testteilnehmenden.

Bekanntlich zeigen beide Geschlechtergruppen auch ein unterschiedliches Wahlverhalten für die Disziplinen Human- und Veterinärmedizin. Eine zweifache Varianzanalyse des Testwertes mit den Faktoren „Geschlecht“ und „Disziplin“ zeigt jedoch nur eine signifikante Hauptwirkung (Tabelle 38): Die Unterschiede bestehen zwischen Human- und Veterinärmedizin – nicht zwischen den Geschlechtern. Bedeutsam ist, dass auch die Wechselwirkung „Geschlecht x Disziplin“ keine Signifikanz zeigt. Der Eindruck, dass Männer tendenziell in Human- und Frauen in Veterinärmedizin bessere Leistungen haben, bleibt im Verhältnis zu den generellen Unterschieden zwischen den Leistungen beider

Disziplinen unbedeutend. Die geringe Stichprobengröße bei männlichen Bewerbern für Veterinärmedizin muss hier erneut als Einschränkung erwähnt werden.

Wiederum zeigen sich signifikante Unterschiede für die einzelnen Untertests zwischen den Geschlechtern. Bis auf den Untertest "Muster zuordnen" sind alle Unterschiede signifikant. In vier Untertests zeigen die Frauen bessere Leistungen, in fünf Untertests die Männer. Dies deckt sich genau mit den Befunden von 1998. Der 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.

		Mittelwert	Standardabw	F	Sig.
Muster zuordnen	männlich	11.5	2.9	.274	.601
	weiblich	11.6	2.9		
Schlauchfiguren	männlich	13.5	3.5	18.161	.000
	weiblich	12.4	3.8		
Figuren lernen	männlich	12.2	3.7	5.958	.015
	weiblich	12.9	3.5		
Konzentr. u. sorgfält. Arb.	männlich	11.9	3.0	19.653	.000
	weiblich	13.0	3.4		
Med.-naturwiss. Grundverst.	männlich	11.9	3.6	17.350	.000
	weiblich	10.8	3.5		
Quantit. und form. Probleme	männlich	12.6	3.5	41.450	.000
	weiblich	10.8	3.9		
Textverständnis	männlich	9.5	3.4	17.918	.000
	weiblich	8.4	3.2		
Fakten lernen	männlich	10.8	3.8	31.131	.000
	weiblich	12.4	3.6		
Diagramme und Tabellen	männlich	11.2	2.9	30.168	.000
	weiblich	10.0	3.1		

Tabelle 41. Mittelwerte und Standardabweichungen geschlechtsspezifisch und Ergebnisse der varianzanalytischen Prüfung F und Irrtumswahrscheinlichkeit (Sig.): bis auf "Muster zuordnen" alle Untertests signifikant unterschiedlich. Dieses Muster der Unterschiede ist mit dem von 1998 identisch.

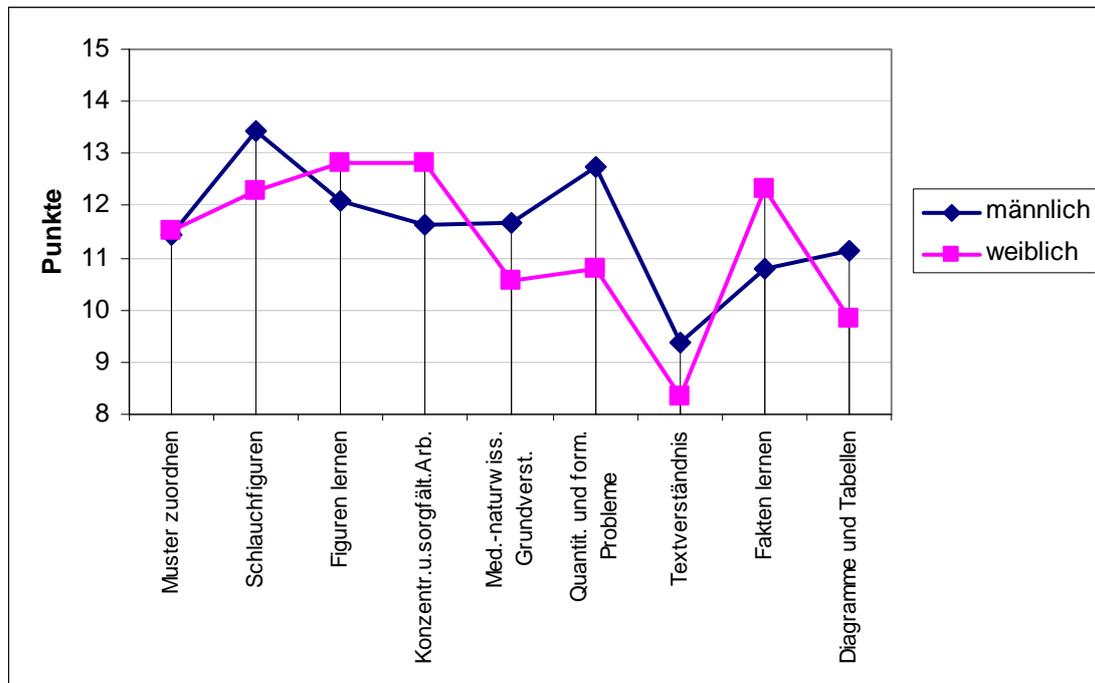


Abbildung 37. Mittelprofile für Punktwerte der Untertests geschlechtsspezifisch.

Die Zulassungsquoten (Gesamt) für beide Geschlechter (Tabelle 41) unterscheiden sich ebenfalls nicht signifikant (Tabelle 42).

	Zulassung		Total
	Warteliste	Zugelassen	
männlich	55 15.7%	295 84.3%	350 100.0%
weiblich	93 18.4%	412 81.6%	505 100.0%
Gesamt	148 17.3%	707 82.7%	855 100.0%

Tabelle 42. Zulassungsquoten Gesamt geschlechtsspezifisch.

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	1.054	1	.305	
Continuity Correction	.874	1	.350	
Likelihood Ratio	1.063	1	.303	
Fisher's Exact Test				.314

Tabelle 43. Prüfung Unterschiede der Zulassung – geschlechtsspezifisch: keine signifikanten Unterschiede.

		Zulassung	
		Warteliste	zugelassen
Humanmedizin	männlich	44 14.4%	262 85.6%
	weiblich	71 19.3%	296 80.7%
Veterinärmedizin	männlich	11 25.0%	33 75.0%
	weiblich	22 15.9%	116 84.1%

Tabelle 44. Zulassungsquoten für Disziplinen - geschlechtsspezifisch.

Differenziert man die Zulassung noch nach der Disziplin, haben Männer in Veterinärmedizin um ca. 10% geringere Zulassungschancen als in Humanmedizin. Bei Frauen beträgt dieser Unterschied nur 3%. Auch dieser Befund spricht dafür, dass die Unterschiede vor allem das Studienwahl-Verhalten betreffen und nicht als "generelle" Unterschiede der Leistungen zwischen Männern und Frauen interpretiert werden dürfen. In der Abbildung 38 werden die prozentualen Verhältnisse gegenübergestellt.

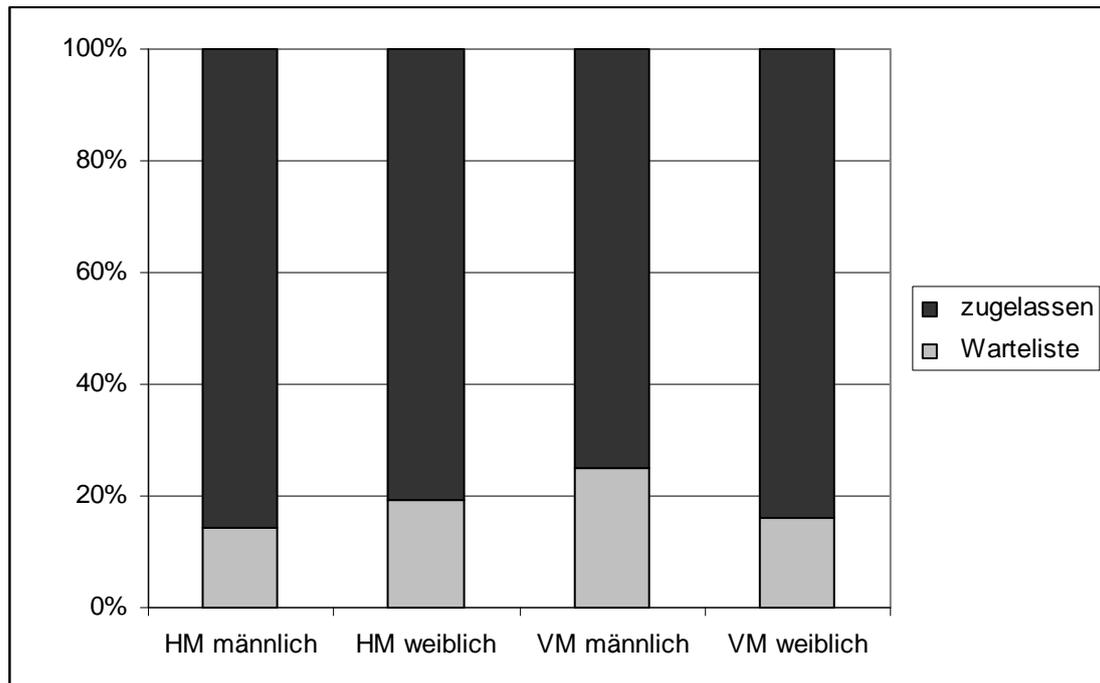


Abbildung 38. Zulassung nach Disziplinen Humanmedizin (HM) und Veterinärmedizin (VM) für die Geschlechter – alle Testteilnehmer auf je 100% bezogen.

Vergleich der Kantone innerhalb der Deutschschweiz

Dass alle Vergleiche sich nur auf die Medizinbewerbungen beziehen, wurde bereits mehrfach betont. Bei allen regionalen Vergleichen muss vor einem weiteren Missverständnis gewarnt werden, welches Oswald (1999) aufzeigt: Er sieht einen "Gegensatz, in welchem sich der EMS zur gegenwärtigen Bildungslandschaft befindet" (S. 1314) und begründet diesen u.a. mit der Gegenüberstellung einer "Schülerzentriertheit" der Bildungspolitik und einer "Gegenstandsorientiertheit" des Tests. Wenn die im EMS geprüften Funktionen bzw. Fähigkeiten in "herrschenden pädagogisch-didaktischen Dogmen" (ebenda) tatsächlich geringgeschätzt werden, so ist vor der "Konstruktion" eines Widerspruches zu klären, ob diese Gegenüberstellung überhaupt gerechtfertigt ist.

Die im EMS zu prüfenden Funktionen sind Ergebnis einer aufwendig betriebenen Anforderungsanalyse des Medizinstudiums, für welches ein erfolgreiches Abschneiden vorhergesagt werden soll. So definiert sich auch die **Fairness der Zulassung**: weil Studienplätze ein "beschränkt zur Verfügung stehendes Gut" sind, sollen diejenigen vor allem eine Chance erhalten, die das Studium in kürzerer Zeit und mit besseren Ergebnissen abschneiden.

Ein Verfahren wie der EMS, das die **Eignung für eine Tätigkeit** feststellen will, wird seine Methoden **immer von den Anforderungen für diese Tätigkeit** her definieren müssen, für welche eine Erfolgs-Vorhersage notwendig ist. Deshalb kann sich der Eignungsbegriff beim EMS nur über die Studierfähigkeit definieren und muss primär wenig Rücksicht darauf nehmen, wie sich die schulische Bildungspolitik gestaltet. Dies ist an sich keine neue Wahrheit, schon den Lateinern war sie offenbar bekannt ("Non scholae, sed vitae discimus"¹).

Es besteht allerdings auch für die Bildungspolitik kein Grund, von einer Identität der bildungspolitischen Ziele und der eignungsrelevanten Fähigkeiten auszugehen. Keinesfalls sollten sich die Schulen, wie vereinzelt erwogen, nun auf das "Training" der im EMS geforderten Fähigkeiten beschränken. Sehr wahrscheinlich lässt sich eine **Förderung** dieser Fähigkeiten nur auf sehr komplexen Wegen erreichen – muss man bildungspolitisch sehr viel breiter vorgehen und beispielsweise auch soziale Komponenten mit einbeziehen. Die bedingte unmittelbare Trainierbarkeit der im EMS geprüften Fähigkeiten wäre ein solcher Beleg, dass Leistungsverbesserungen nicht auf kurzen Wegen erreichbar sind.

Die Bildungspolitik muss sich allerdings gefallen lassen, dass die Bewertung ihres Erfolges bzw. des Erfolges ihrer Absolventinnen und Absolventen von den Anforderungen des späteren Berufslebens (hier des Studienerfolges) her möglich sein muss. Die Evaluationskriterien dürfen praktisch gar nicht mit den Bildungszielen identisch sein (sonst handelt es sich um reine Zirkelschlüsse) und sie müssen "von aussen" angelegt werden. Dieses "Aussen" muss sich vor allem an beruflichen Erfolgskriterien orientieren, wenn die Schweiz bildungspolitisch im internationalen Wettbewerb erfolgreich bleiben will. Eine kontinuierliche Evaluation der Schulausbildung an diesen externen Kriterien hätte langfristig im Übrigen vielleicht auch eine bildungspolitisch regulierende Funktion.

Vor diesem Hintergrund müssen alle nachfolgenden Vergleiche gesehen werden: Unterschiede können (1) auf regional unterschiedlichem Studienwahlverhalten und unterschiedlicher Wirkung der Dissuasion beruhen (2) spezifisch für die Medizinbewerbungen sein (ohne für alle Maturandinnen und Maturanden zu gelten).

¹ Nicht für die Schule, fürs Leben lernen wir.

hoch (> 18%)			mittel (13 – 18%)			niedrig (< 13%)		
Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl
BL	23.3	57	SH	16.6	4	BE	12.9	111
BS	20.7	39	ZG	16.4	10	SG	12.8	61
ZH	18.7	190	AG	16.2	84	SZ	12.7	18
			AR	16.0	6	LU	12.2	44
			GL	14.1	4	OW	12.2	2
			AI	13.7	0	NW	10.2	3
			SO	13.6	28	TG	9.8	20
			GR	13.2	27			
			UR	13.0	1			
Summe		286	Summe		164	Summe		259

Tabelle 45. Gruppierung der Kantone nach der Maturitätsquote (1998)

Im Jahre 1998 wurden keine Testwertunterschiede für die Kantonsgruppen gefunden, wenn man sie nach der Maturitätsquote gruppiert. Diese Aussage kann 1999 erneut bestätigt werden:

Es wurden drei Gruppen der Kantone nach der Maturitätsquote wie 1998 gebildet (Tabelle 44). Dabei wurden nur die Personen dieser Kantone berücksichtigt, die den Test in deutscher Sprache abgelegt haben. Die Aussagefähigkeit wird natürlich von vornherein dadurch eingeschränkt, dass sich die einzelnen Kantone in vielen weiteren Merkmalen unterscheiden, die nicht ausbalanciert werden können (Lage, Bevölkerungsanteil, Stadt-Land-Verhältnis usw.). Dennoch bleibt es ein interessantes Ergebnis, dass sich keine signifikanten Unterschiede zwischen den drei Gruppen auf 5%-Niveau ergeben – weder beim Testwert, noch in den Untertests.

Die Ergebnisse belegen zumindest, dass es gegenwärtig keinen Grund dafür gibt, in der unterschiedlichen Maturitätsquote eine generelle Ursache für Vorteile oder Nachteile für einzelne Kantone bei der Zulassung zum Medizinstudium zu sehen und Schlussfolgerungen bezüglich einer Auswahl für die Mittelschule zu ziehen. Das Gegenteil scheint der Fall: Gleichgültig, wie hoch die Maturitätsquote ist, bewerben sich für Medizin immer gleichgute Kandidatinnen und Kandidaten. Die Variation der Maturitätsquoten innerhalb der Deutschschweiz ist dabei natürlich im Verhältnis zur Gesamtschweiz geringer (Tessin mit 29.8 und Genf mit 29.6 unterscheiden sich vom höchsten Wert der einbezogenen Kantone – BL mit 23.3 - noch einmal beträchtlich).

Bei der Gruppe mit hoher Maturitätsquote ist zu bemerken, dass 34% der Personen aus Basel Stadt und Basel Land stammen und mit dem Kanton Zürich zusammengefasst werden. Beide liegen im nachfolgenden Vergleich der "grossen" Kantone sogar an unterschiedlichen Enden der Testwerte-Verteilung. Besonders wichtig ist, dass die Zusammenfassungen der Kantone nach mittlerer und niedriger Maturitätsquote sich bezüglich des Testwertes und nicht unterscheiden. Hier gehen mindestens sieben Kantone ein.

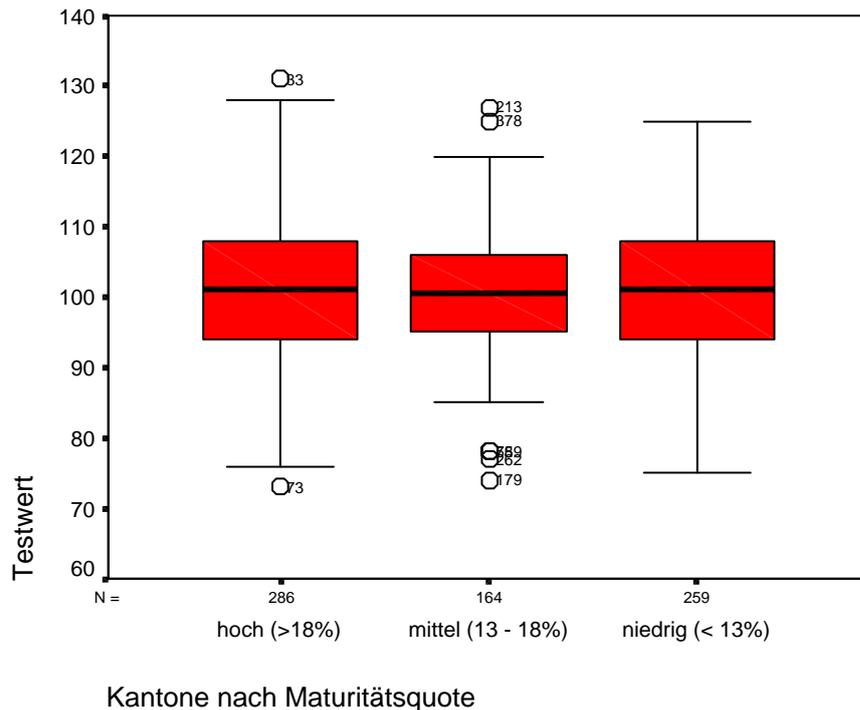


Abbildung 39. Boxplot für die drei Gruppen der Kantone nach der Maturitätsquote.

Für Vergleiche zwischen den Kantonen wurden alle Kantone berücksichtigt, die mehr als 38 Teilnehmer am Test aufweisen (im Vorjahr betrug die Grenze 40 – damit beide Basel wiederum berücksichtigt werden können, war die „Absenkung“ notwendig).

In diesem Jahr ergeben sich deutlichere Unterschiede, die beim Testwert und sechs Untertests zu signifikant unterscheidbaren Mittelwertsgruppen führen (Tabelle 47). Im Vergleich zum Vorjahr bleibt festzuhalten, dass Unterschiede vor allem an den Enden der Leistungsangereihe der Kantone nach dem Testwert auftreten: Die Leistungen in zwei Kantonen mit 1998 bereits hohen Testwerten sind besser – die Leistung in dem Kanton mit den relativ geringsten Testwerten ist schlechter geworden (Abbildung 40). Die Tabelle 48 verdeutlicht, dass diese Unterschiede in mehreren Untertests durchgängig auftreten und die Rangreihen der Mittelwerte durchaus korrespondieren. BS mit der höchsten Maturitätsquote innerhalb der Deutschschweiz (1998: 23,3%) liegt dabei nie am Beginn einer Rangreihe – es ist BL mit einer um 2,6% geringeren Maturitätsquote. Man muss allerdings festzuhalten, dass BS und BL die höchsten Maturitätsquoten innerhalb der Deutschschweiz haben (2% Differenz zu ZH, 4% Differenz zu SH) – dass beide beim Testwert und in vier Untertests die jeweils niedrigsten Werte aufweisen. Im speziellen Fall Basel sollte man deshalb die – im deutschschweizerischen Massstab geltende - Unabhängigkeit des Erfolges von der Maturitätsquote noch nicht als gänzlich unbedeutend verwerfen. Um vor vorschnellen Schlüssen zu warnen: Unterschiede bezüglich BS und BL können beeinflusst sein durch (1) die Auswahl zur Mittelschule, (2) die Mittelschule selber, (3) die Studienwahl zum Medizinstudium. Keine dieser Ebenen kann aufgrund der Ergebnisse als unbedeutend ausgeschlossen werden.

		N	Mittelwert	Stand.- abweichg	Minimum	Maximum
Testwert	hoch (>18%)	286	101.2063	10.5456	73.00	131.00
	mittel (13 - 18%)	164	101.0671	9.1485	74.00	127.00
	niedrig (< 13%)	259	100.7568	9.6789	75.00	125.00
Muster zuordnen	hoch (>18%)	286	11.6503	3.0152	4.00	19.00
	mittel (13 - 18%)	164	11.8963	2.9299	2.00	19.00
	niedrig (< 13%)	259	11.6062	2.8732	2.00	20.00
Schlauchfiguren	hoch (>18%)	286	12.9406	3.8150	1.00	20.00
	mittel (13 - 18%)	164	12.9573	3.7619	5.00	20.00
	niedrig (< 13%)	259	13.0000	3.5753	4.00	20.00
Figuren lernen	hoch (>18%)	286	12.6853	3.6657	3.00	20.00
	mittel (13 - 18%)	164	12.6098	3.5002	4.00	20.00
	niedrig (< 13%)	259	12.7799	3.6591	4.00	20.00
Konzentr. u. sorgfält. Arb.	hoch (>18%)	286	12.9965	3.3961	5.00	20.00
	mittel (13 - 18%)	164	12.6098	3.2018	7.00	20.00
	niedrig (< 13%)	259	12.4208	3.3452	4.00	20.00
Med.-naturwiss. Grundverst.	hoch (>18%)	286	11.4545	3.7318	3.00	20.00
	mittel (13 - 18%)	164	11.3171	3.7837	2.00	20.00
	niedrig (< 13%)	259	11.2008	3.2917	2.00	20.00
Quantit. und form. Probleme	hoch (>18%)	286	11.5210	3.7740	1.00	20.00
	mittel (13 - 18%)	164	11.8659	3.8156	2.00	20.00
	niedrig (< 13%)	259	11.7683	4.0530	1.00	20.00
Textverständnis	hoch (>18%)	286	9.0385	3.4031	1.00	18.00
	mittel (13 - 18%)	164	8.9451	3.4254	1.00	17.00
	niedrig (< 13%)	259	8.8378	3.3290	1.00	17.00
Fakten lernen	hoch (>18%)	286	12.0140	3.8852	1.00	20.00
	mittel (13 - 18%)	164	11.7744	3.6627	3.00	20.00
	niedrig (< 13%)	259	11.7375	3.8510	3.00	20.00
Diagramme und Tabellen	hoch (>18%)	286	10.5909	3.0660	2.00	19.00
	mittel (13 - 18%)	164	10.6463	2.9572	4.00	19.00
	niedrig (< 13%)	259	10.7181	3.1650	1.00	19.00

Tabelle 46. Mittelwerte und Standardabweichungen für drei Gruppen der Kantone nach der Maturitätsquote, keine signifikanten Unterschiede für Testwert und die Untertests (vgl. Anhang Tabelle A 7).

		Testwert	Musterzuordnen	Schlauchfiguren	Figuren lernen	Konzent. u. sorgfält. Arb.	Med.-naturwiss. Grundv.	Quantit. u. form. Proble.	Textverständnis	Fakten lernen	Diagramme und Tabellen
AG	M	101.21	11.55	12.62	12.46	12.25	11.82	12.08	9.20	12.26	10.63
	S	9.64	2.99	3.85	3.61	3.03	3.69	3.98	3.70	3.37	3.11
BE	M	100.43	11.57	13.12	12.64	11.90	11.22	11.81	9.14	11.47	10.61
	S	9.53	2.72	3.71	3.52	3.19	3.34	4.18	3.33	3.84	3.18
BL	M	99.21	11.12	12.86	12.40	12.65	11.30	10.32	8.51	11.12	10.61
	S	9.73	3.01	3.56	3.21	3.20	3.88	3.79	3.33	4.01	3.00
BS	M	95.28	10.64	11.95	11.00	13.03	9.59	9.44	7.54	10.38	9.44
	S	10.47	2.97	3.71	4.02	3.59	3.90	3.52	2.95	3.68	3.11
LU	M	99.32	11.45	12.55	12.66	12.50	10.61	11.48	7.89	11.36	10.59
	S	9.43	3.08	3.82	3.58	3.37	2.98	4.43	3.05	3.21	3.01
SG	M	101.90	12.25	13.20	12.84	13.08	11.21	11.89	8.85	12.16	10.90
	S	10.75	2.66	3.23	3.60	3.45	3.46	4.12	3.69	4.12	3.40
ZH	M	103.02	12.02	13.17	13.12	13.09	11.88	12.31	9.51	12.62	10.82
	S	10.29	2.97	3.90	3.63	3.42	3.55	3.57	3.41	3.76	3.04

Tabelle 47. Vergleich der Kantone mit mehr als 38 Teilnehmer für Testwert und die einzelnen Untertests.

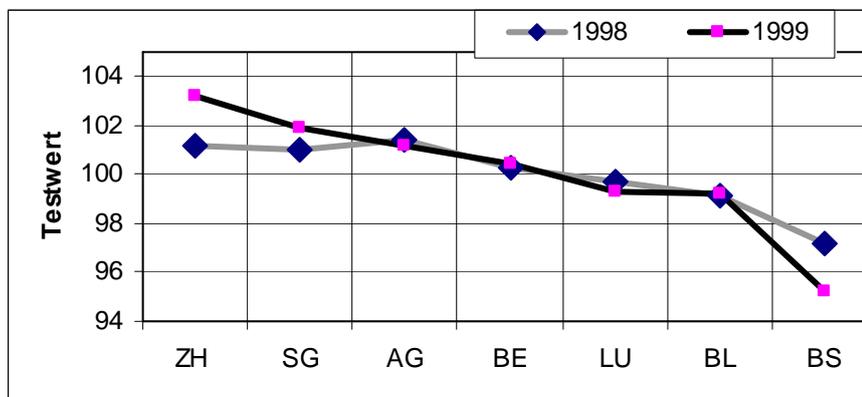


Abbildung 40. Testwert für Kantone mit mehr als 38 Teilnehmer 1998 und 1999 im Vergleich.

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

Testwert			
	N	Untergruppen für alpha = .05	
BS	39	95.2	
BL	57	99.2	99.2
LU	44	99.3	99.3
BE	111	100.4	100.4
AG	84		101.2
SG	61		101.9
ZH	190		103.0
Sig.		.054	.315
Muster zuordnen			
	N	Untergruppen für alpha = .05	
BS	39	10.6	
BL	57	11.1	11.1
LU	44	11.4	11.4
AG	84	11.5	11.5
BE	111	11.5	11.5
ZH	190	12.0	12.0
SG	61		12.2
Sig.		.102	.299
Figuren lernen			
	N	Untergruppen für alpha = .05	
BS	39	11.0	
BL	57	12.4	12.4
AG	84	12.4	12.4
BE	111	12.6	12.6
LU	44	12.6	12.6
SG	61	12.8	12.8
ZH	190		13.1
Sig.		.056	.920
Med.-naturwiss. Grundverst.			
	N	Untergruppen für alpha = .05	
BS	39	9.5	
LU	44	10.6	10.6
SG	61	11.2	11.2
BE	111	11.2	11.2
BL	57	11.2	11.2
AG	84		11.8
ZH	190		11.8
Sig.		.088	.390

... Fortsetzung nächste Seite

Quantit. und form. Probleme			
	N	Untergruppen für alpha = .05	
BS	39	9.4	
BL	57	10.3	10.3
LU	44		11.4
BE	111		11.8
SG	61		11.8
AG	84		12.0
ZH	190		12.3
Sig.		.860	.056
Textverständnis			
	N	Untergruppen für alpha = .05	
BS	39	7.5	
LU	44	7.8	7.8
BL	57	8.5	8.5
SG	61	8.8	8.8
BE	111	9.1	9.1
AG	84	9.2	9.2
ZH	190		9.5
Sig.		.081	.099
Fakten lernen			
	N	Untergruppen für alpha = .05	
BS	39	10.3	
BL	57	11.1	11.1
LU	44	11.3	11.3
BE	111	11.4	11.4
SG	61	12.1	12.1
AG	84	12.2	12.2
ZH	190		12.6
Sig.		.066	.261

Tabelle 48. Multiple Mittelwertvergleiche für Kantone mit mehr als 38 Teilnehmer

Vergleiche für Altersgruppen und den Zeitpunkt des Maturitätsabschlusses

Anhand der vorliegenden psychologischen Erkenntnisse sind zwei gegensätzliche Tendenzen beim Zusammenhang von Leistungsanforderungen und Alter abzuleiten: Jüngere sind besser, wenn es um solche Leistungsvoraussetzungen wie Aufmerksamkeit, Belastbarkeit oder Gedächtnis geht. In fast allen Normtabellen von Lern- oder Aufmerksamkeitstests findet sich dieser Trend. Älteren gelingt es dafür aufgrund ihrer Erfahrungen besser, Wesentliches von Unwesentlichem zu unterscheiden und ihre bisherigen Erfahrungen als Routine beim Lösen neuartiger Probleme einzubringen. Beide Tendenzen sollten sich so ergänzen, dass die effektive Leistungsfähigkeit mit steigendem Alter zumindest gleich bleiben kann. Für beide Tendenzen finden sich im EMS entsprechende Untertests. Wir würden aufgrund der Anforderungsanalysen zum Medizinstudium davon ausgehen, dass beide Tendenzen auf hohem Niveau benötigt werden, um gute Chancen für einen Studienerfolg zu haben.

Bezüglich der Untersuchung des Alterseinflusses bieten sich aufgrund der Daten drei Gruppen an, deren Grenzen zu Vergleichszwecken gegenüber 1998 um je ein Jahr verschoben sind: die Jahrgänge 1979-1982, die sich im 19. Lebensjahr unmittelbar nach der Maturitätsprüfung bewerben, die Jahre 1976-1978, die dies eins bis zwei Jahre später tun und die Jahre 1957-1975, die dies mehrere Jahre später tun (Tabelle 49).

	Häufigkeit	Prozent	Kumulative Prozent
1957	1	.1	.1
1958	1	.1	.3
1959	1	.1	.4
1961	2	.3	.7
1962	1	.1	.8
1963	2	.3	1.1
1965	1	.1	1.2
1967	3	.4	1.6
1968	6	.8	2.4
1969	6	.8	3.2
1970	7	.9	4.1
1971	5	.7	4.8
1972	6	.8	5.6
1973	10	1.3	7.0
1974	14	1.9	8.8
1975	17	2.3	11.1
1976	24	3.2	14.3
1977	59	7.9	22.2
1978	171	22.9	45.1
1979	336	44.9	90.0
1980	65	8.7	98.7
1981	9	1.2	99.9
1982	1	.1	100.0
Total	748	100.0	

Tabelle 49. Gruppenbildung für Jahrgänge

Die **älteste Gruppe** lässt sich nach dem Alter beim Erlangen der Maturität wiederum differenzieren: solche Personen, die ihre Maturitätsprüfung zum frühestmöglichen Zeitpunkt abgelegt haben und dann einer anderen Tätigkeit nachgegangen sind und solche Personen, die ihre Maturitätsprüfung später bestanden – also in der Regel unmittelbar vor der jetzigen Bewerbung zum Medizinstudium (siehe Tabelle 50). Hier könnte man annehmen, dass Personen, die sich unmittelbar vor dem Ablegen des Tests noch in einer Schulausbildung befunden haben, beim Test besser abschneiden. Die Technologien der Wissensaneignung wurden unmittelbar vorher gefordert und damit trainiert. Der Trennpunkt beider Gruppen wird zwischen Maturitätsjahr 1995 und 1996 gesetzt.

Maturitätsjahr	Jahrgänge		
	1957 bis 1975	1976-1978	1979-1982
1975	1		
1980	1		
1985	1		
1986	1		
1987	2		
1988	3		
1990	2		
1991	2		
1992	3		
1993	4		
1994	5		
1995	4	1	
1996	3	5	
1997	5	27	
1998	8	111	64
1999	38	109	347
	83	253	411

Tabelle 50. Gruppenbildung für Jahr des Ablegens der Maturitätsprüfung bezogen auf die Jahrgänge.

Die Leistungen der drei Altersgruppen sind in Abbildung 41 bzw. Tabelle 51 dargestellt. Bis auf einen Untertests sind die Unterschiede signifikant – die Mittelwerte der jüngeren Gruppe sind beim Testwert am besten. Es lässt sich also ein Vorteil der jüngeren Bewerberinnen und Bewerber nachweisen, die sich sofort nach einer frühestmöglich bestandenen Maturität zum Studium bewerben. Dies dürfte auch der Studienrealität entsprechen, wonach Jüngere insgesamt bessere Leistungen erreichen.

Differenziert man die älteste Gruppe nach dem Erlangen der Maturität, zeigt sich entgegen der geäußerten Erwartung auch hier ein besserer Punkt- und Testwert für Personen, welche die Maturität **früh abgelegt** haben. Sie unterscheiden sich in allen Leistungen nicht von der mittleren Altersgruppe. Tabelle A 9 des Anhanges zeigt die signifikanten Mittelwertunterschiede für beide Teilgruppen. Der Unterschied des Testwertes erreicht Signifikanz – bei den Untertests sind Unterschiede in den „Werkzeugfunktionen“ ("Figuren lernen", "Konzentriertes und sorgfältiges Arbeiten") und wie im Vorjahr beim logischen Denken ("Quantitative und formale Probleme") signifikant. Die Leistungsbesten haben sich offenbar auch früh für eine Maturität entschieden – diese Unterschiede gleichen sich später nicht aus.

	Jahrgang	N	M	S		df	MQ	F	Sig.
Testwert	vor 75, Fr. M	54	95.5	7.5	Zwischen	3	801.861	8.308	.000
	vor 75, Sp. M	29	99.7	9.4	Innerhalb	744	96.515		
	76 - 78	254	99.7	9.9	Total	747			
	79 - 82	411	102.0	10.0					
Muster zuordnen	vor 75, Fr. M	54	10.4	2.3	Zwischen	3	64.546	7.548	.000
	vor 75, Sp. M	29	10.0	2.5	Innerhalb	744	8.551		
	76 - 78	254	11.5	2.7	Total	747			
	79 - 82	411	11.9	3.0					
Schlauchfiguren	vor 75, Fr. M	54	11.6	4.2	Zwischen	3	38.939	2.804	.039
	vor 75, Sp. M	29	12.5	3.1	Innerhalb	744	13.889		
	76 - 78	254	12.7	3.6	Total	747			
	79 - 82	411	13.1	3.7					
Figuren lernen	vor 75, Fr. M	54	11.2	3.2	Zwischen	3	51.089	3.966	.008
	vor 75, Sp. M	29	13.0	4.1	Innerhalb	744	12.883		
	76 - 78	254	12.5	3.5	Total	747			
	79 - 82	411	12.9	3.6					
Konzentr. u. sorgfält. Arb.	vor 75, Fr. M	54	11.2	3.0	Zwischen	3	40.512	3.736	.011
	vor 75, Sp. M	29	12.5	3.5	Innerhalb	744	10.845		
	76 - 78	254	12.5	3.2	Total	747			
	79 - 82	411	12.8	3.3					
Med.-naturwiss. Grundverst.	vor 75, Fr. M	54	9.5	3.2	Zwischen	3	71.309	5.563	.001
	vor 75, Sp. M	29	10.8	3.6	Innerhalb	744	12.818		
	76 - 78	254	11.1	3.5	Total	747			
	79 - 82	411	11.6	3.6					
Quantit. und form. Probleme	vor 75, Fr. M	54	10.7	3.5	Zwischen	3	57.672	3.880	.009
	vor 75, Sp. M	29	12.7	3.7	Innerhalb	744	14.864		
	76 - 78	254	11.1	3.8	Total	747			
	79 - 82	411	11.9	3.8					
Textverständnis	vor 75, Fr. M	54	7.6	2.9	Zwischen	3	47.803	4.214	.006
	vor 75, Sp. M	29	8.0	3.4	Innerhalb	744	11.344		
	76 - 78	254	8.7	3.3	Total	747			
	79 - 82	411	9.1	3.4					
Fakten lernen	vor 75, Fr. M	54	11.5	3.5	Zwischen	3	14.385	1.004	.390
	vor 75, Sp. M	29	11.8	4.5	Innerhalb	744	14.327		
	76 - 78	254	11.5	3.8	Total	747			
	79 - 82	411	12.0	3.7					
Diagramme und Tabellen	vor 75, Fr. M	54	9.3	3.1	Zwischen	3	53.184	5.572	.001
	vor 75, Sp. M	29	10.3	2.8	Innerhalb	744	9.545		
	76 - 78	254	10.2	3.0	Total	747			
	79 - 82	411	10.9	3.1					

Tabelle 51. Testwert und Punktwerte für die drei Altersgruppen (nach Jahrgängen) – für die älteste Gruppe aufgeteilt nach früher (Fr. M) und späterer (Sp. M.) Maturitätsprüfung; varianzanalytische Prüfung des Unterschiedes.

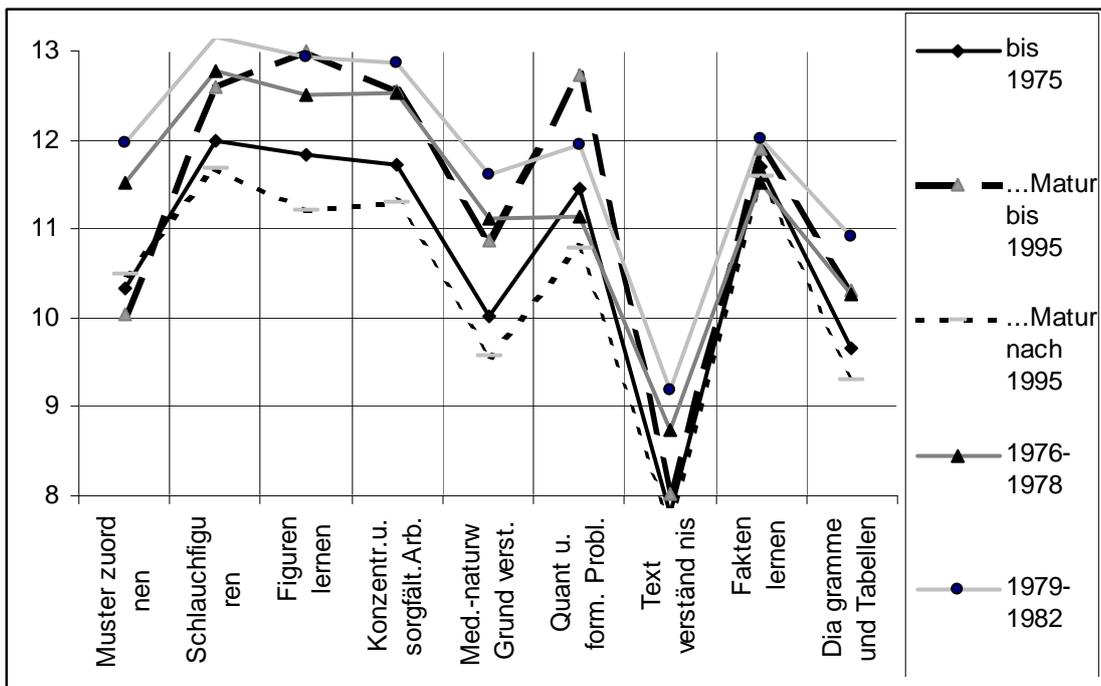


Abbildung 41. Mittelwerte der Punktwerte für die Untertests nach Altersgruppen.

Die Zulassungsquoten für die drei Altersgruppen unterscheiden sich ebenfalls signifikant. Die Bewerberinnen und Bewerber, die das Medizinstudium auf „direktem“ Wege beginnen, haben die besten Chancen. In den beiden Gruppen der Älteren sind die Zulassungsquoten jeweils geringer.

Dass trotz der Unterschiede die Gruppe der Älteren nicht chancenlos ist, wird durch den Vergleich der Zulassungsquoten gezeigt. Die Personen, die ihre Maturitätsprüfung früh abgelegt haben, erreichen die Zulassungsquote der mittleren Altersgruppe. Eine schlechtere Prognose haben ältere Personen, die ihre Maturitätsprüfung spät abgelegt haben. Die „Übung“, in welcher sie durch das kürzliche Ablegen der Maturitätsprüfungen noch sind, kann die vorhandenen Leistungsunterschiede nicht ausgleichen. Dies ist ein erneuter Hinweis dafür, dass der Test nur bedingt trainierbar ist – auch nicht indirekt durch eine Ausbildung. Wiederum sei einschränkend darauf hingewiesen, dass es sich um eine sehr geringe Stichprobengrösse und ausschliesslich um Medizinberwerber handelt. Hier sind diese Effekte allerdings stabil und entsprechen denen von 1998.

Jahrgang	Zulassung		Total
	Warteliste	zugelassen	
bis 1975	20 24.1%	63 75.9%	83 100.0%
... davon frühe Matur	5 17.2%	24 82.8%	29 100.0%
... davon späte Matur	15 27.8%	39 72.2%	54 100.0%
1976 -1978	45 17.7%	209 82.3%	254 100.0%
1979 -1982	50 12.2%	361 87.8%	411 100.0%

Tabelle 52. Zulassungsquoten für die drei Altersgruppen (nach Jahrgängen) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung.

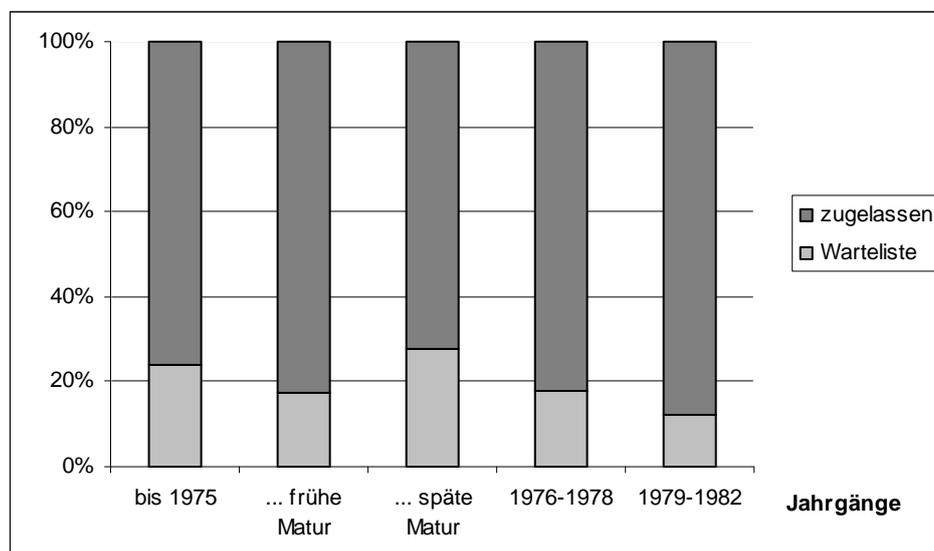


Abbildung 42. Zulassungsquoten für Altersgruppen (nach Jahrgängen) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung.

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.

Disziplin		Zulassung		Total
		Warteliste	zugelassen	
Humanmedizin	Basel	33	116	149
		22.1%	77.9%	100.0%
	Bern	23	107	130
		17.7%	82.3%	100.0%
	Freiburg	24	55	79
	30.4%	69.6%	100.0%	
Zürich	35	280	315	
	11.1%	88.9%	100.0%	
	Total	115	558	673
	17.1%	82.9%	100.0%	
Veterinärmedizin	Bern	16	69	85
		18.8%	81.2%	100.0%
	Zürich	17	80	97
	17.5%	82.5%	100.0%	
Total	33	149	182	
	18.1%	81.9%	100.0%	

Tabelle 53. Zulassungsquoten für Disziplinen und Wunschuniversität (erste Wahl)

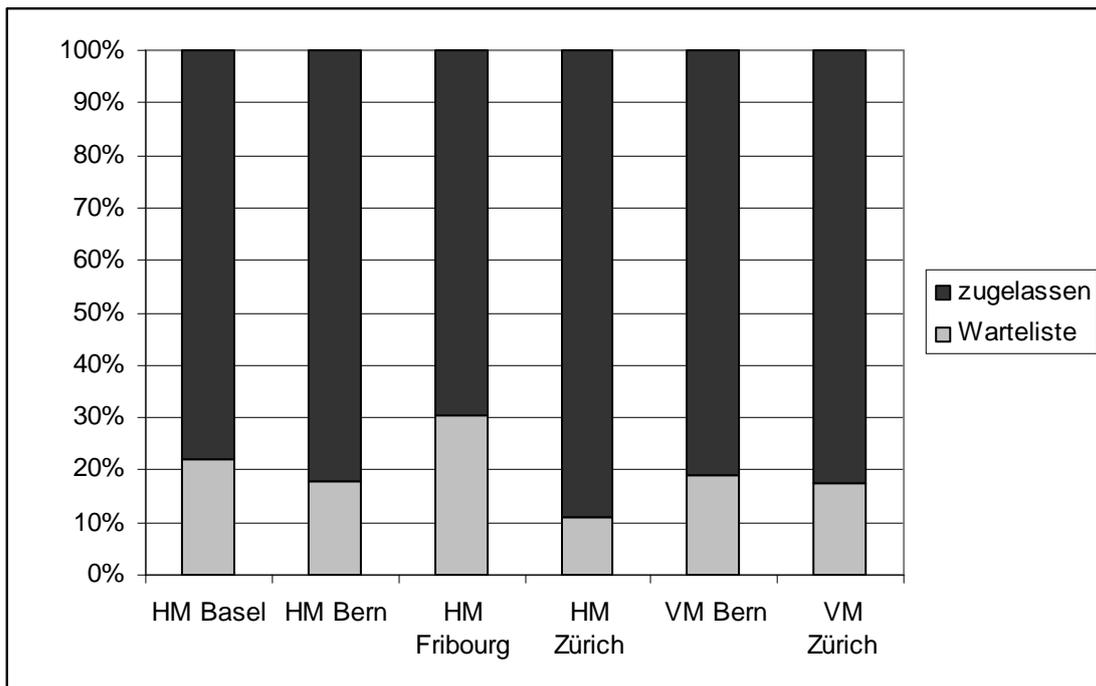


Abbildung 43: Zulassungsquoten für Disziplinen und Wunschuniversitäten.

Disziplin		Value	df	Asymp. Sig. (2-sided)
Humanmedizin	Pearson Chi-Square	20.520	3	.000
	Likelihood Ratio	19.792	3	.000
	N of Valid Cases	673		
Veterinärmedizin	Pearson Chi-Square	.051	1	.821
	Likelihood Ratio	.051	1	.821
	N of Valid Cases	182		

Tabelle 54. Prüfung Homogenität Zulassung nach Wunschuniversität: keine signifikanten Unterschiede für Veterinärmedizin, Unterschiede für Humanmedizin sind signifikant.

Ergebnisse zur Testgüte

Zuverlässigkeit

Die Zuverlässigkeit (Messgenauigkeit – deutsche Ergebnisse vgl. Trost et al., 1995) kann anhand zweier Koeffizienten verglichen werden, die üblicherweise bei Niveautests herangezogen werden können. 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 alle in der Schweiz ermittelten Werte in der Bandbreite der für 16 Formen in Deutschland in den Jahren 1995 und 1996 ermittelten Konsistenzen.

Die Reliabilitäten nach der Testhalbierungsmethode (Teilung nach gerade- und ungeradzahligem Aufgaben) wurden ebenfalls verglichen. Wegen der Stichprobengrößen werden nur die Ergebnisse der deutschen Form 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		1998	1999
Muster zuordnen	.64 - .69	.69	.60	.67 - .73	.72	.61
Med.-nat. Grundverständnis	.64 - .73	.70	.70	.65 - .72	.72	.71
Schlauchfiguren	.76 - .80	.71	.74	.75 - .82	.75	.74
Quant. u. form. Probleme	.69 - .78	.76	.75	.70 - .75	.78	.75
Textverständnis	.68 - .71	.73	.75	.69 - .74	.75	.70
Figuren lernen	.69 - .74	.72	.70	.72 - .75	.72	.70
Fakten lernen	.62 - .70	.68	.72	.68 - .73	.70	.70
Diagramme und Tabellen	.67 - .71	.67	.75	.68 - .72	.66	.62

Tabelle 55. Zuverlässigkeit 1999 und 1998; Vergleich der Skalenkonsistenzen für die Untertests, Ergebnisse für Deutschland aus Trost et al. (1997), S. 26 ff.

Die **Reliabilität des Punktwertes** insgesamt – geschätzt nach der Testhalbierungsmethode (gerade vs. ungerade Itemnummer) auf der Basis aller neun Untertests liegt mit 0.91 (1998: 0.92) in einem hohen Bereich. Hier liegt der Vergleichswert für Deutschland zwischen 0.91 und 0.93 (gemessen für rund 3000 Teilnehmer).

Die nachfolgende Tabelle 56 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 (1998: 0.83).

Untertest	Konsistenz ohne den jeweiligen Untertest 1998	Konsistenz ohne den jeweiligen Untertest 1999
Muster zuordnen	.82	.80
Med.-nat. Grundverständnis	.79	.78
Schlauchfiguren	.80	.79
Quant. u. form. Probleme	.80	.79
Textverständnis	.80	.79
Figuren lernen	.81	.80
Fakten lernen	.82	.81
Diagramme und Tabellen	.80	.78
Konzentr. und sorgf. Arbeiten	.82	.81
Konsistenz des Testprofils	.83	.81

Tabelle 56. 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 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 (Troost 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 „perceptual speed“ bei Thurstone identifiziert),

- die Merkfähigkeit (als „memory“ bei Thurstone) - wobei verbales und figurales Gedächtnis als die beiden „Gegensätze“ gleichermaßen berücksichtigt sind.

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:

	Punkt- wert	Muster zuordnen	Schlauch- figuren	Figuren lernen	Konzentr. u. sorgfält. Arb.	Med.- naturwiss. Grund- verst.	Quantit. und form. Probleme	Textver- ständnis	Fakten lernen	Diagram- me und Tabellen
Punktwert	1.000	.558	.654	.642	.561	.725	.683	.662	.583	.718
Muster zuordnen	.558	1.000	.397	.326	.390	.239	.262	.213	.210	.315
Schlauch- figuren	.654	.397	1.000	.377	.379	.349	.344	.317	.210	.393
Figuren lernen	.642	.326	.377	1.000	.413	.310	.248	.288	.407	.324
Konzentr. u. sorgfält. Arb.	.561	.390	.379	.413	1.000	.218	.193	.131	.306	.231
Med.- naturwiss. Grundverst.	.725	.239	.349	.310	.218	1.000	.523	.621	.347	.581
Quantit. und form. Probleme	.683	.262	.344	.248	.193	.523	1.000	.481	.272	.590
Text- verständnis	.662	.213	.317	.288	.131	.621	.481	1.000	.299	.502
Fakten lernen	.583	.210	.210	.407	.306	.347	.272	.299	1.000	.265
Diagramme und Tabellen	.718	.315	.393	.324	.231	.581	.590	.502	.265	1.000

Tabelle 57. Korrelationen zwischen Punktwerten und Untertests

- Alle Untertests laden signifikant im ersten unrotierten Faktor. Dies ist als Hinweis auf einen Generalfaktor der Leistungsfähigkeit zu interpretieren; es ist gemeinsame Varianz zwischen allen Untertests vorhanden. Mit einem Anteil von 40,7% an der aufgeklärten Varianz ist er genauso bedeutsam wie 1998.
- 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 „Werkzeugfunktionen“ 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 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.

Unrotierte Lösung							Rotierte Lösungen					
Faktor	Eigenwert		% Varianz		Cumul. %		Eigenwert		% Varianz		Cumul %	
	1998	1999	1998	1999	1998	1999	1998	1999	1998	1999	1998	1999
1	3.829	3.670	42.546	40.782	42.546	40.782	Zweifaktorenlösung					
2	1.380	1.364	15.330	15.158	57.876	55.940	2.865	2.723	31.831	30.254	31.831	30.254
3	.856	.914	9.510	10.160	67.386	66.100	2.344	2.312	26.045	25.687	57.876	55.940
4	.627	.639	6.969	7.099	74.354	73.199	Dreifaktorenlösung					
5	.564	.585	6.269	6.503	80.623	79.703	2.807	2.683	31.191	29.806	31.191	29.806
6	.511	.535	5.677	5.946	86.300	85.649	1.858	1.811	20.644	20.122	51.835	49.928
7	.489	.527	5.436	5.856	91.737	91.505	1.399	1.455	15.550	16.172	67.385	66.100
8	.392	.418	4.355	4.644	96.092	96.150						
9	.352	.347	3.908	3.850	100.000	100.000						

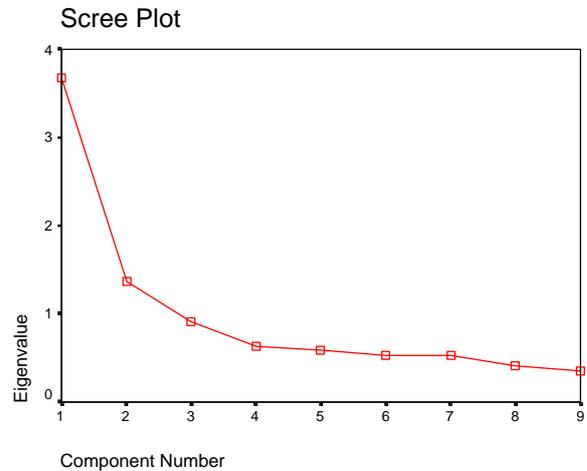
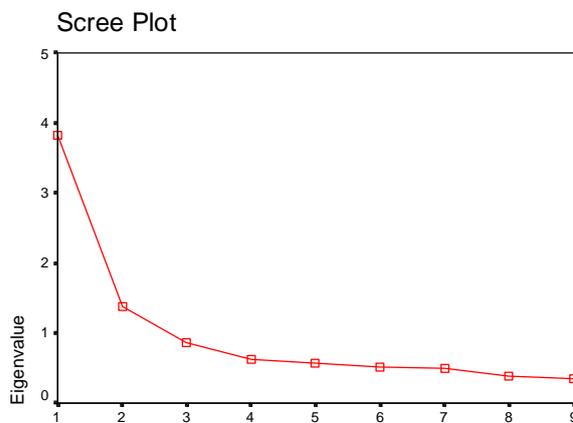


Abbildung 44. Scree-Plot (Eigenwerteverlauf) der Faktoren 1 bis 9 für die Untertests des EMS 1998 (links) und 1999 (rechts).

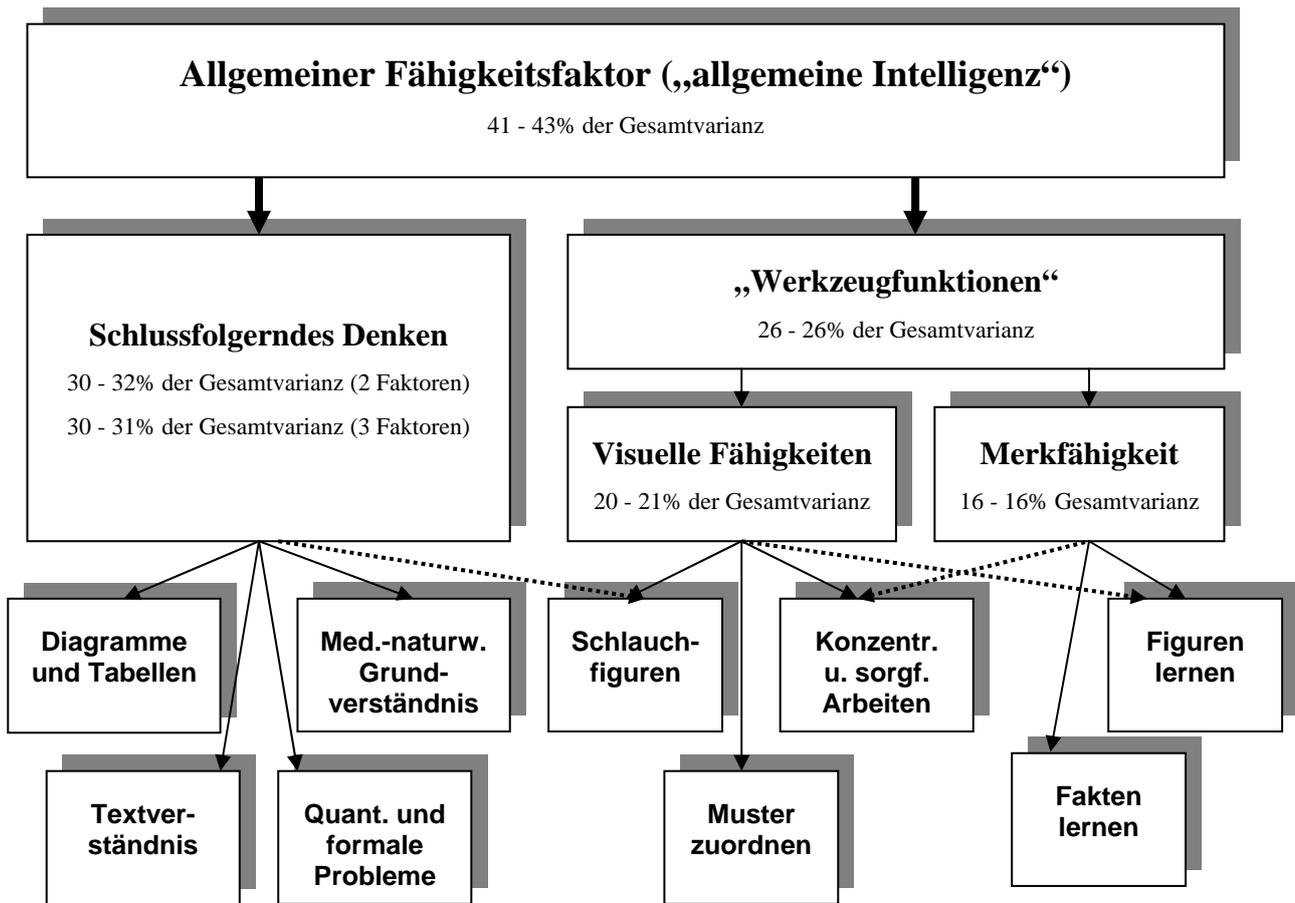


Abbildung 45. Struktur der Untertests des EMS, hierarchische Faktorenstruktur (1999 erster Wert und 1998 zweiter Wert der Varianzaufklärung durch den Faktor).

Im „Scree-Plot“ 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 eins – der Varianzanteil des Faktors ist kleiner als der eines einzelnen Untertests. Die Abbildung 44 zur Struktur der Untertests fasst die Ergebnisse der drei nachfolgenden Analysen zusammen. Die Ladungsmatrix des ersten Faktors der unrotierten Lösung zeigt substanzielle Ladungen für alle Untertests. Ein „Generalfaktormodell“ ist indiziert. Der allgemeine Fähigkeitsfaktor ist sehr varianzstark.

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

Tabelle 58. Ladungen im ersten Faktor der unrotierten Lösung 1998 und 1999 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor).

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

Tabelle 59. 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
	1998	1999	1998	1999	1998	1999	1998	1999			
Diagramme und Tabellen	.83	.76		.29			.71	.67	.82		
Medizin.-naturwiss. Grundverständnis	.78	.81			.27	.22	.71	.71	.81		.20
Quantitat. u. formale Probleme	.77	.78	.23	(.17)			.65	.64	.80	.18	
Textverständnis	.76	.80			.27	.17	.67	.67	.79		.20
Muster zuordnen			.83	.77			.72	.62		.81	
Konzentr. u. sorgfält. Arbeiten			.65	.65	.43	.44	.62	.62		.70	.40
Schlauchfiguren	.44	.33	.60	.70			.58	.60	.35	.71	
Fakten lernen					.88	.86	.83	.79	.21		.87
Figuren lernen	.21	(.14)	.50	.42	.51	.66	.56	.63	.13	.47	.64

Tabelle 60. Faktorenanalyse: Varimaxrotierte Dreifaktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42).

Item-Trennschärfen

Dargestellt sind die Items geordnet nach Itemnummer. Da die Schwierigkeiten mit steigender Itemnummer zunehmen, ist ein "glockenförmiger" Verlauf zu erwarten. In allen Untertests zeigen sich identische Verläufe – der glockenförmige Verlauf ist nur beim ersten Untertest deutlich ausgeprägt. Die Trennschärfen in der Schweiz liegen im Betrag höher, was z.T. auch auf die geringere Personenzahl zurückgeführt werden muss.

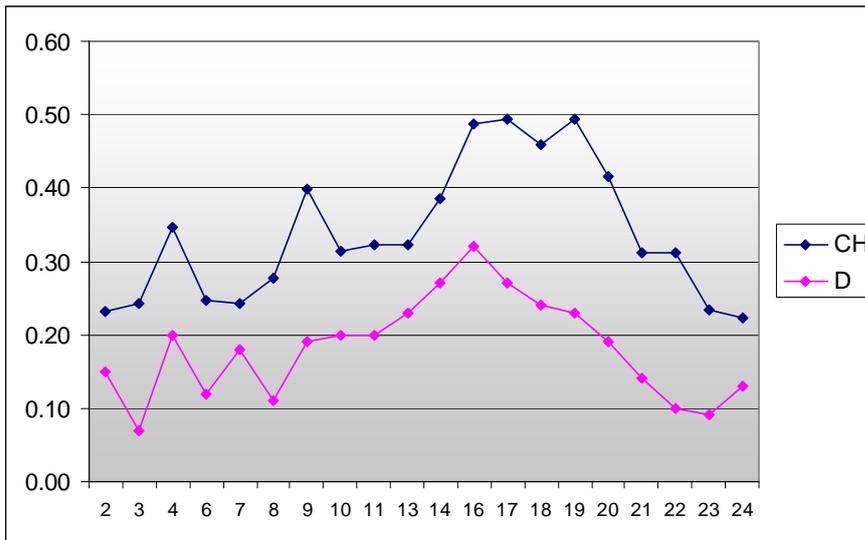


Abbildung 46. Trennschärfen für den Untertest "Muster zuordnen".

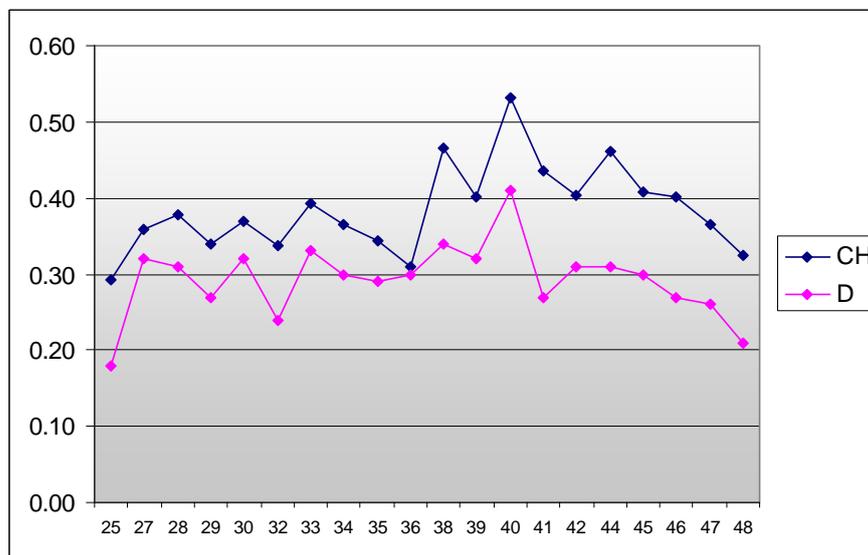


Abbildung 47. Trennschärfen für den Untertest "Medizinisch-naturwissenschaftliches Grundverständnis".

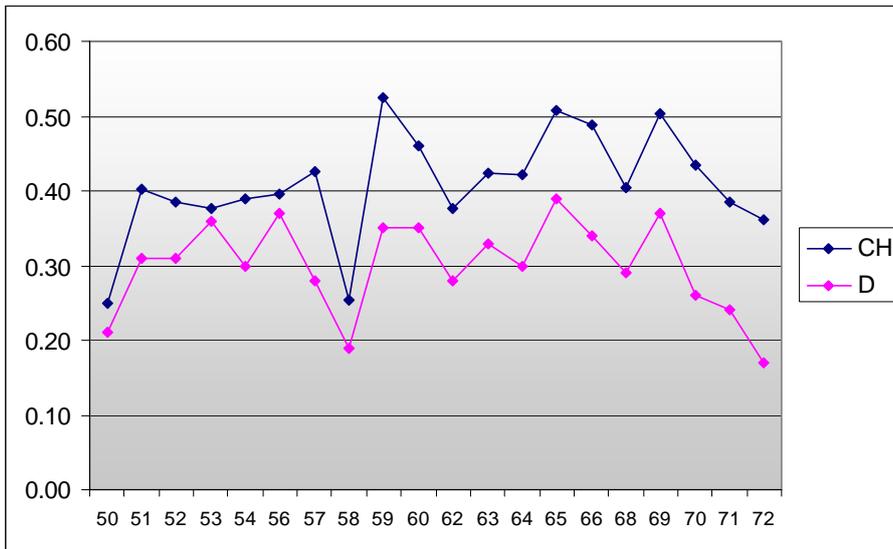


Abbildung 48. Trennschärpen für den Untertest "Schlauchfiguren".

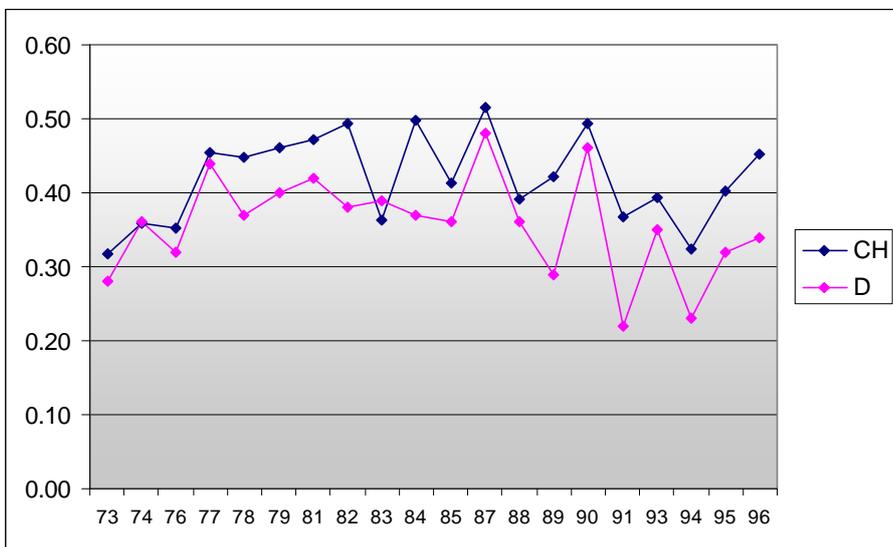


Abbildung 49. Trennschärpen für den Untertest "Quantitative und formale Probleme".

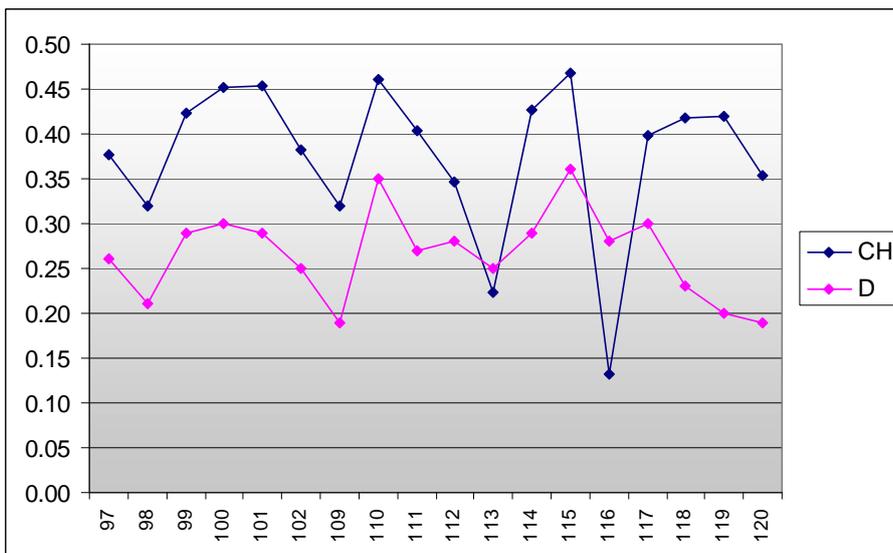


Abbildung 50. Trennschärpen für den Untertest "Textverständnis".

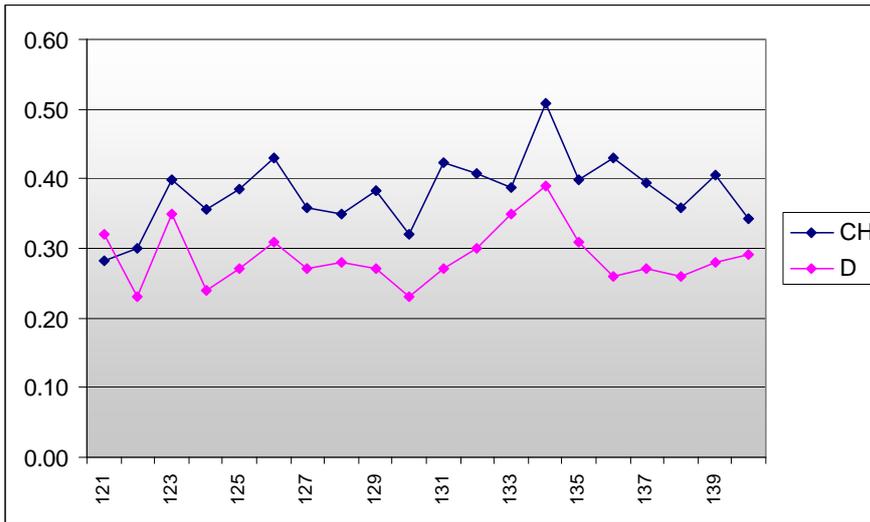


Abbildung 51. Trennschärfen für den Untertest "Figuren lernen".

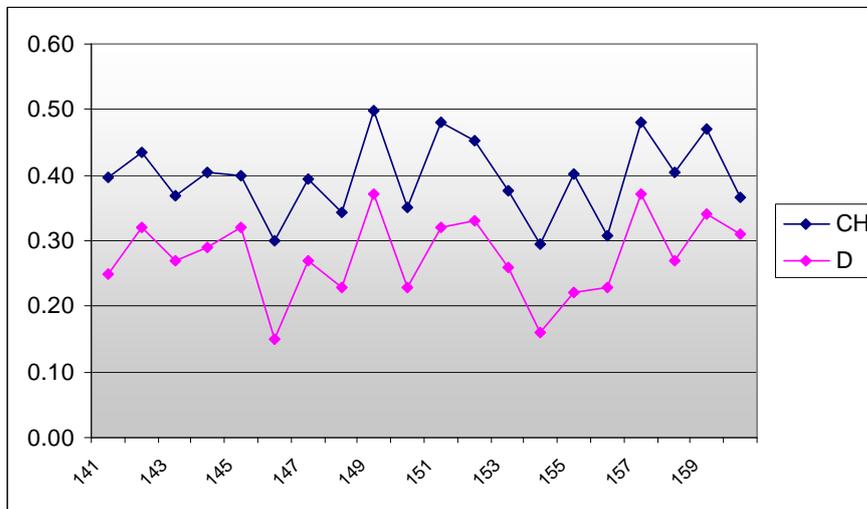


Abbildung 52. Trennschärfen für den Untertest "Fakten lernen".

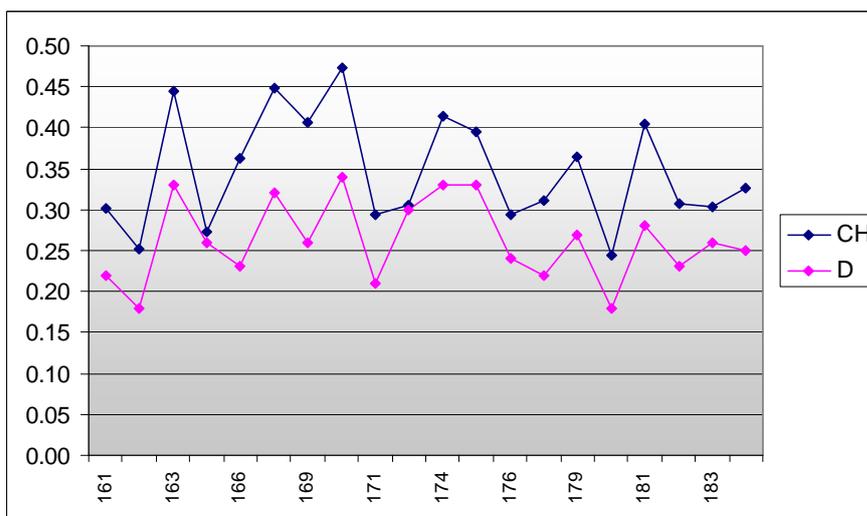


Abbildung 53. Trennschärfen für den Untertest "Diagramme und Tabellen".

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.

Die Leistungen in der deutschen Sprachgruppe haben sich gegenüber 1998 nicht verändert. Auch in der italienischen Sprachgruppe werden 1998 nur sieben Zeichen weniger beantwortet (bei einer Standardabweichung von 76 Zeichen) und im Mittel nur ein Fehler mehr gemacht. Nur in der französischsprachigen Gruppe werden 1999 37 Zeichen weniger als 1998 (über 1/3 der Standardabweichung) richtig bearbeitet und sieben Fehler mehr gemacht. Wie bei den Vergleichen der Sprachgruppen diskutiert, scheinen hierfür vor allem Veränderungen der Stichprobe durch verändertes Studienwahlverhalten verantwortlich zu sein.

Insgesamt gibt es aufgrund der Ergebnisse (noch) 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.

	deutsch		französisch		italienisch	
	M	S	M	S	M	S
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
... übersehen	10.02	11.96	18.84	60.80	6.52	7.50
... fälschlich durchgestrichen	2.69	3.38	2.14	3.58	2.71	2.25
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 61. Kennwerte für den Untertest „Konzentriertes und sorgfältiges Arbeiten“. * zwei Personen haben die Instruktionen nicht verstanden (228 resp. 505 Fehler).

	RICHTIG	FALSCH	... übersehen	... fälschlich markiert
RICHTIG	1.000 .	-.075 .028	-.103 .003	.151 .000
FALSCH	-.075 .028	1.000 .	.990 .000	.397 .000
... übersehen	-.103 .003	.990 .000	1.000 .	.261 .000
... fälschlich markiert	.151 .000	.397 .000	.261 .000	1.000 .

Tabelle 62. Interkorrelationen der Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“ und Signifikanzniveaus (jeweils darunter).

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., 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

Es steht in allen drei Sprachen eine vollständige Originalversion zur Verfügung. Diese wurde bei früheren Testterminen in Deutschland verwendet. Vom ZTD wurden französisch- und italienischsprachige Adaptationen dieses veröffentlichten Originaltests entwickelt.

Institut für Test- und Begabungsforschung (Hrsg.) (1995). Test für medizinische Studiengänge (Aktualisierte Originalversion 2). Herausgegeben im Auftrag der Kultusminister der Länder der BRD. 4. Auflage. Göttingen: Hogrefe.

Centre pour le développement de tests et le diagnostic, Université de Fribourg (Suisse) en collaboration avec l'Institut für Test- und Begabungsforschung, Bonn, Allemagne (Editeur) (1996). Le test d'aptitudes pour les études de médecine. Adaptation française de la version originale dans son intégralité. Göttingen: Hogrefe.

Centre pour le développement de tests et le diagnostic, Université de Fribourg (Svizzera) in collaborazione con l'Institut für Test- und Begabungsforschung, Bonn, Germania (Editore) (1996). Il test attitudinale per lo studio della medicina. Adattamento italiano di una versione originale completa. Göttingen: Hogrefe.

Frühere Berichte des ZTD

Die nachfolgenden Berichte des ZTD zur Vorbereitung des Eignungstests und zum 1998 erstmalig durchgeführten Eignungstest sind im Internet unter www.unifr.ch/ztd verfügbar.

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> K.-D. Hänsgen und B. Spicher 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> 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> 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-Dieter Hänsgen</p>	<p>Le test d'aptitudes pour les études de médecine en Suisse - Epreuve d'essai Rainer Hofer, Daniel Ruefli & Klaus-Dieter Hänsgen</p>

Anhang zum Bericht 1999

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	1	5	11	20	39	58	84	98	110	104	105	81	55	43	23	11	4	1
Med.-nat. Grundverst			3	8	19	32	35	58	60	81	94	94	82	75	55	59	38	23	20	15	4
Schlauchfiguren		1	1	3	8	10	29	34	45	52	61	75	76	66	70	93	77	65	54	24	11
Quant./form. Prob.		2	4	11	10	22	38	37	70	75	81	82	71	84	63	55	47	38	42	13	10
Textverständnis		5	6	16	59	69	88	94	104	86	73	67	67	47	26	25	15	6	2		
Figuren lernen			1	4	8	12	17	40	47	60	76	64	89	84	80	74	64	52	47	26	10
Fakten lernen		1	1	6	15	25	42	40	49	68	79	81	87	83	77	65	45	41	23	14	13
Diagramm/Tabellen		2	3	9	14	28	51	47	79	112	106	98	87	86	56	33	21	14	4	5	
Konzentr.u sorgfält.Ar.	2				1	6	17	21	60	75	82	103	102	85	87	59	49	35	25	24	22
Deutsche Sprachgr.	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	8	16	32	45	72	81	102	96	93	71	49	40	21	11	4	1
Med.-nat. Grundverst			3	6	10	22	30	47	50	71	80	85	75	64	53	56	37	22	19	14	4
Schlauchfiguren		1		1	5	8	25	28	41	47	50	64	66	55	60	81	74	63	47	22	10
Quant./form.Probl.		2	4	11	7	17	29	33	60	66	72	72	63	76	52	50	43	33	37	12	9
Textverständnis		4	6	12	47	58	76	80	90	71	62	61	62	45	26	25	15	6	2		
Figuren lernen				2	7	8	13	34	41	50	66	55	76	78	76	62	57	46	45	23	9
Fakten lernen		1	1	5	13	18	36	32	43	59	63	70	77	74	69	59	42	39	21	13	13
Diagramm. u. Tabellen		1	3	7	10	19	38	40	61	100	92	89	79	82	55	32	20	12	3	5	
Konzentr./sorgf. Arb.					1	5	8	16	44	58	73	95	89	76	80	53	46	34	25	24	21
Französ. Sprachgr.	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	3	3	5	8	8	14	7	5	8	5	3	3	2			
Med.-nat. Grundverst				2	7	10	3	9	8	7	9	5	4	6	2	2		1	1		
Schlauchfiguren			1	2	2	2	3	3	4	3	7	6	8	9	7	8	2	1	5	2	1
Quant. u. form. Probl.					3	4	9	4	10	5	4	7	5	6	6	4	3	1	4	1	
Textverständnis				3	10	9	7	8	12	11	7	4	5								
Figuren lernen			1	1	1	4	2	3	4	5	10	7	10	4	3	9	5	2	2	2	1
Fakten lernen				1	2	5	5	6	3	7	11	7	8	7	4	4	2	2	2		
Diagramm u. Tabellen		1		2	4	7	8	5	14	8	7	5	7	3	1	1	1	1	1		
Konzentr./sorgf. Arb.	2					1	7	4	15	14	7	4	4	7	5	2	2	1			1

Tabelle A 1. Punktwerteverteilung gesamt, deutsche und französische Sprachgruppe

Italienische Sprachgr.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen							1	2	5	4	3	1	3	4	5	3					
Med.-nat. Grundverst.					2		2	2	2	3	5	4	3	5		1	1				1
Schlauchfiguren					1		1	3		2	4	5	2	2	3	4	1	1	2		
Quant./form .Probl.						1				4	5	3	3	2	5	1	1	4	1		1
Textverständnis		1		1	2	2	5	6	2	4	4	2		2							
Figuren lernen				1			2	3	2	5		2	3	2	1	3	2	4			1
Fakten lernen						2	1	2	3	2	5	4	2	2	4	2	1				1
Diagramme u. Tabellen						2	5	2	4	4	7	4	1	1				1			
Konzentr. u. sorgfält. Arb.							2	1	1	3	2	4	9	2	2	4	1				
Humanmedizin	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen			1		4	8	15	31	37	64	77	90	82	87	61	49	37	16	9	4	1
Med.-nat. Grundverst.			1	5	12	24	27	47	45	64	74	74	62	62	42	48	35	19	17	12	3
Schlauchfiguren		1	1	1	5	8	21	23	34	39	51	57	58	56	54	71	66	51	44	21	11
Quant./form .Probl.		2	4	10	6	15	24	31	49	61	55	67	56	67	51	48	41	33	33	12	8
Textverständnis		5	4	10	42	52	65	74	85	68	58	56	54	39	19	21	15	4	2		
Figuren lernen			1	3	7	10	14	31	36	50	55	50	71	67	66	58	51	42	34	19	8
Fakten lernen		1	1	5	10	16	34	27	42	55	61	65	69	65	59	54	35	29	21	13	11
Diagramme u. Tabellen		2	2	9	11	16	37	40	58	85	85	77	64	73	48	28	18	11	4	5	
Konzentr. u. sorgfält. Arb.	2				1	5	14	18	37	57	62	84	83	56	67	51	40	33	21	21	21
Veterinärmedizin	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	3	5	8	21	20	21	20	22	18	20	6	6	7	2		
Med.-nat. Grundverst.			2	3	7	8	8	11	15	17	20	20	20	13	13	11	3	4	3	3	1
Schlauchfiguren				2	3	2	8	11	11	13	10	18	18	10	16	22	11	14	10	3	
Quant./form .Probl.				1	4	7	14	6	21	14	26	15	15	17	12	7	6	5	9	1	2
Textverständnis			2	6	17	17	23	20	19	18	15	11	13	8	7	4		2			
Figuren lernen				1	1	2	3	9	11	10	21	14	18	17	14	16	13	10	13	7	2
Fakten lernen				1	5	9	8	13	7	13	18	16	18	18	18	11	10	12	2	1	2
Diagramme u. Tabellen			1		3	12	14	7	21	27	21	21	23	13	8	5	3	3			
Konzentr. u. sorgfält. Arb.						1	3	3	23	18	20	19	19	29	20	8	9	2	4	3	1

Tabelle A 2. Punktwerteverteilung italienische Sprachgruppe und für Disziplinen Gesamt

	Humanmedizin		Veterinärmedizin	
	Punktwert	Testwert	Punktwert	Testwert
48	1			
49	1			
51	2			
53	1		1	
54			1	
55			1	
56	2			
57	1			
58	1		1	
59	3			
61	1		1	
62	1		2	
64	1		1	
65	4			
66	1		1	
67	2		2	
68	4		3	
69	3			
70	3		1	
71	8			
72	6		2	
73	6	2	3	
74	2	2		
75	4	1	1	1
76	9			2
77	4	3	1	
78	4	4	2	1
79	6	1	1	1
80	5	1	5	2
81	6	5	3	1
82	11	3		3
83	12	7	3	3
84	10	11	1	1
85	11	12	2	5
86	7	6	6	1
87	5	13	2	1
88	11	10	3	3
89	10	11	4	8
90	6	23	4	3
91	8	21	2	3
92	11	12	5	8
93	19	21	5	7
94	15	14	2	6
95	10	30	3	10
96	10	25	5	5
97	11	21	4	9
98	14	24		3
99	10	29	3	9
100	10	27	3	10
101	19	19	6	8
102	12	21	3	10
103	15	28	7	6

- Fortsetzung siehe Folgeseite -

	Humanmedizin		Veterinärmedizin	
	Punktwert	Testwert	Punktwert	Testwert
104	11	28	2	6
105	8	24	6	8
106	7	24	3	7
107	14	19	7	5
108	15	21	3	3
109	13	20	3	5
110	15	21	1	6
111	13	19	5	3
112	9	12	4	2
113	15	9	4	
114	15	7	5	1
115	9	7	2	2
116	9	10	3	1
117	10	6	2	1
118	14	10	1	2
119	7	7	2	
120	12	7	1	
121	8	6	4	
122	12	2	1	
123	9		5	
124	8	1		
125	11	2	3	
126	6	1	1	
127	6	1	1	
128	3	1		
129	6			
130	2			
131	5	1	1	
132	5		2	
133	2			
134	5			
135	5		1	
136	4			
137	2		1	
138	6		2	
139	4			
140	3			
141	4			
142	5			
143	2			
144	4			
145	2			
147	2			
150	1			
152	1			
153	1			
155	1			
156	1			
158	1			
164	1			

Tabelle A 3. Punkt- und Testwertverteilung nach Disziplinen

		Levene's Test for Equality of Variances		t-test for Equality of Mean		
		F	Sig.	T	df	Sig. (2-tailed)
Testwert	Equal variances assumed	.934	.334	.864	853	.388
Muster zuordnen	Equal variances not assumed	5.101	.024	.885	853	.376
Schlauchfiguren	Equal variances assumed	1.425	.233	-1.131	853	.258
Figuren lernen	Equal variances assumed	.322	.571	-.440	853	.660
Konzentriertes und sorgfältiges Arbeiten	Equal variances assumed	.403	.526	.903	853	.367
Med.-naturwiss. Grundverständnis	Equal variances assumed	.476	.490	.790	853	.430
Quantitative und formale Probleme	Equal variances assumed	1.087	.298	1.150	853	.250
Textverständnis	Equal variances assumed	1.840	.175	1.354	853	.176
Fakten lernen	Equal variances assumed	1.129	.288	.296	853	.768
Diagramme und Tabellen	Equal variances assumed	.105	.746	1.244	853	.214

Tabelle A 4. T-Test für die beiden Testformen: Varianzhomogenität ist bis auf Muster zuordnen gegeben – keine signifikanten Mittelwertunterschiede sind nachweisbar.

		Quadrat- summe	df	Mittl. quadr. Abweichung	F	Sig.
Testwert	Zwischen Gruppen	3155.081	2	1577.541	16.281	.000
	Innerhalb Gruppen	82553.497	852	96.894		
	Total	85708.578	854			
Muster zuordnen	Zwischen Gruppen	116.944	2	58.472	6.615	.001
	Innerhalb Gruppen	7530.799	852	8.839		
	Total	7647.743	854			
Schlauch- figuren	Zwischen Gruppen	131.725	2	65.862	4.629	.010
	Innerhalb Gruppen	12121.707	852	14.227		
	Total	12253.432	854			
Figuren lernen	Zwischen Gruppen	124.371	2	62.185	4.632	.010
	Innerhalb Gruppen	11438.979	852	13.426		
	Total	11563.350	854			
Konzentriertes und sorgfältiges Arbeiten	Zwischen Gruppen	586.153	2	293.077	26.928	.000
	Innerhalb Gruppen	9272.836	852	10.884		
	Total	9858.989	854			
Medizinisch- naturwiss. Grundverstän.	Zwischen Gruppen	383.779	2	191.890	14.953	.000
	Innerhalb Gruppen	10933.791	852	12.833		
	Total	11317.570	854			
Quantitative und formale Probleme	Zwischen Gruppen	77.058	2	38.529	2.597	.075
	Innerhalb Gruppen	12638.625	852	14.834		
	Total	12715.683	854			
Text- verständnis	Zwischen Gruppen	98.159	2	49.079	4.521	.011
	Innerhalb Gruppen	9248.907	852	10.856		
	Total	9347.066	854			
Fakten lernen	Zwischen Gruppen	65.340	2	32.670	2.336	.097
	Innerhalb Gruppen	11916.597	852	13.987		
	Total	11981.937	854			
Diagramme und Tabellen	Zwischen Gruppen	192.747	2	96.373	9.990	.000
	Innerhalb Gruppen	8219.550	852	9.647		
	Total	8412.297	854			

Tabelle A 5. Varianzanalytische Prüfung des Einflusses für den Faktor Sprache

			Mittelwert Difference (I-J)	Std. Error	Sig.
	(I) Sprache	(J) Sprache			
Testwert	deutsch	französisch	6.5482	1.1851	.000
		italienisch	3.1709	1.8042	.184
	französisch	deutsch	-6.5482	1.1851	.000
		italienisch	-3.3773	2.0977	.241
	italienisch	deutsch	-3.1709	1.8042	.184
		französisch	3.3773	2.0977	.241
Muster zuordnen	deutsch	französisch	1.2508	.3579	.001
		italienisch	.6646	.5449	.442
	französisch	deutsch	-1.2508	.3579	.001
		italienisch	-.5862	.6336	.624
	italienisch	deutsch	-.6646	.5449	.442
		französisch	.5862	.6336	.624
Schlauchfiguren	deutsch	französisch	1.2115	.4541	.021
		italienisch	1.1215	.6914	.236
	französisch	deutsch	-1.2115	.4541	.021
		italienisch	-8.9983E-02	.8038	.993
	italienisch	deutsch	-1.1215	.6914	.236
		französisch	8.998E-02	.8038	.993
Figuren lernen	deutsch	französisch	1.1789	.4411	.021
		italienisch	1.0851	.6716	.239
	französisch	deutsch	-1.1789	.4411	.021
		italienisch	-9.3803E-02	.7809	.992
	italienisch	deutsch	-1.0851	.6716	.239
		französisch	9.380E-02	.7809	.992
Konzentriertes und sorgfältiges Arbeiten	deutsch	französisch	2.8560	.3972	.000
		italienisch	1.1485	.6047	.139
	französisch	deutsch	-2.8560	.3972	.000
		italienisch	-1.7076	.7031	.040
	italienisch	deutsch	-1.1485	.6047	.139
		französisch	1.7076	.7031	.040

- Fortsetzung siehe Folgeseite -

Medizinisch-naturwissenschaftliches Grundverständnis	deutsch	französisch	2.3181	.4313	.000
		italienisch	.8749	.6566	.377
	französisch	deutsch	-2.3181	.4313	.000
		italienisch	-1.4431	.7634	.141
	italienisch	deutsch	-.8749	.6566	.377
		französisch	1.4431	.7634	.141
Quantitative und formale Probleme	deutsch	französisch	.7532	.4637	.235
		italienisch	-1.0571	.7059	.292
	französisch	deutsch	-.7532	.4637	.235
		italienisch	-1.8103	.8208	.070
	italienisch	deutsch	1.0571	.7059	.292
		französisch	1.8103	.8208	.070
Textverständnis	deutsch	französisch	1.1298	.3967	.012
		italienisch	.6855	.6039	.492
	französisch	deutsch	-1.1298	.3967	.012
		italienisch	-.4443	.7021	.802
	italienisch	deutsch	-.6855	.6039	.492
		französisch	.4443	.7021	.802
Fakten lernen	deutsch	französisch	.7900	.4503	.185
		italienisch	.9365	.6855	.359
	französisch	deutsch	-.7900	.4503	.185
		italienisch	.1465	.7970	.982
	italienisch	deutsch	-.9365	.6855	.359
		französisch	-.1465	.7970	.982
Diagramme und Tabellen	deutsch	französisch	1.5942	.3739	.000
		italienisch	.9106	.5693	.246
	französisch	deutsch	-1.5942	.3739	.000
		italienisch	-.6835	.6619	.556
	italienisch	deutsch	-.9106	.5693	.246
		französisch	.6835	.6619	.556

Tabelle A 6. Paarweise Mittelwertvergleiche für die Sprachgruppen

		Quadratsumme	df	Mittl. quadr. Abweichung	F	Sig.
Testwert	Zwischen Gruppen	28.164	2	14.082	.143	.867
	Innerhalb Gruppen	69506.767	706	98.452		
	Total	69534.931	708			
Muster zuordnen	Zwischen Gruppen	9.251	2	4.626	.534	.587
	Innerhalb Gruppen	6120.103	706	8.669		
	Total	6129.354	708			
Schlauchfiguren	Zwischen Gruppen	.497	2	.248	.018	.982
	Innerhalb Gruppen	9752.691	706	13.814		
	Total	9753.188	708			
Figuren lernen	Zwischen Gruppen	3.048	2	1.524	.116	.891
	Innerhalb Gruppen	9281.158	706	13.146		
	Total	9284.206	708			
Konzentriertes und sorgfältiges Arbeiten	Zwischen Gruppen	46.654	2	23.327	2.099	.123
	Innerhalb Gruppen	7845.148	706	11.112		
	Total	7891.803	708			
Medizinisch-naturwissenschaftliches Grundverständnis	Zwischen Gruppen	8.789	2	4.394	.341	.711
	Innerhalb Gruppen	9097.981	706	12.887		
	Total	9106.770	708			
Quantitative und formale Probleme	Zwischen Gruppen	14.831	2	7.415	.491	.612
	Innerhalb Gruppen	10670.523	706	15.114		
	Total	10685.354	708			
Textverständnis	Zwischen Gruppen	5.471	2	2.736	.239	.787
	Innerhalb Gruppen	8072.272	706	11.434		
	Total	8077.743	708			
Fakten lernen	Zwischen Gruppen	11.869	2	5.934	.406	.666
	Innerhalb Gruppen	10314.743	706	14.610		
	Total	10326.612	708			
Diagramme und Tabellen	Zwischen Gruppen	2.204	2	1.102	.116	.890
	Innerhalb Gruppen	6689.049	706	9.475		
	Total	6691.252	708			

Tabelle A 7. Varianzanalytische Prüfung Unterschiede der Kantone (drei Gruppen) nach der Maturitätsquote.

		Quadrat- summe	df	Mittl. quadr. Ab- weichung	F	Sig.
Testwert	Zwischen Gruppen	2447.329	6	407.888	4.076	.001
	Innerhalb Gruppen	57939.62 8	579	100.068		
	Total	60386.95 7	585			
Muster zuordnen	Zwischen Gruppen	105.770	6	17.628	2.085	.053
	Innerhalb Gruppen	4896.341	579	8.457		
	Total	5002.111	585			
Schlauch- figuren	Zwischen Gruppen	71.476	6	11.913	.852	.530
	Innerhalb Gruppen	8099.221	579	13.988		
	Total	8170.696	585			
Figuren lernen	Zwischen Gruppen	155.843	6	25.974	2.019	.061
	Innerhalb Gruppen	7447.897	579	12.863		
	Total	7603.741	585			
Konzentr. u. sorgfält. Arbeiten	Zwischen Gruppen	131.176	6	21.863	1.989	.065
	Innerhalb Gruppen	6363.502	579	10.991		
	Total	6494.677	585			
Med.- naturwiss. Grundverst.	Zwischen Gruppen	220.543	6	36.757	2.933	.008
	Innerhalb Gruppen	7256.612	579	12.533		
	Total	7477.155	585			
Quantit. und form. Probleme	Zwischen Gruppen	397.892	6	66.315	4.371	.000
	Innerhalb Gruppen	8785.202	579	15.173		
	Total	9183.094	585			
Text- verständnis	Zwischen Gruppen	206.876	6	34.479	2.969	.007
	Innerhalb Gruppen	6724.790	579	11.614		
	Total	6931.666	585			
Fakten lernen	Zwischen Gruppen	269.817	6	44.970	3.209	.004
	Innerhalb Gruppen	8114.744	579	14.015		
	Total	8384.561	585			
Diagramme und Tabellen	Zwischen Gruppen	67.155	6	11.193	1.154	.330
	Innerhalb Gruppen	5616.962	579	9.701		
	Total	5684.118	585			

Tabelle A 8. Varianzanalytische Prüfung der Unterschiede zwischen Kantonen mit mehr als 38 Teilnehmer.

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
Muster zuordnen	Equal variances assumed	.494	.484	.652	86	.516
Schlauchfiguren	Equal variances not assumed	3.969	.050	-1.095	83.259	.277
Figuren lernen	Equal variances assumed	1.175	.281	-2.283	86	.025
Konzentr. u. sorgfält. Arbeiten	Equal variances assumed	1.234	.270	-1.977	86	.051
Med.-naturwiss. Grundverständnis	Equal variances assumed	.052	.821	-1.571	86	.120
Quantit. und form. Probleme	Equal variances assumed	.940	.335	-2.019	86	.047
Textverständnis	Equal variances assumed	.249	.619	-.823	86	.413
Fakten lernen	Equal variances assumed	2.087	.152	-.664	86	.508
Diagramme und Tabellen	Equal variances assumed	.278	.599	-1.196	86	.235
Testwert	Equal variances assumed	.313	.577	-2.278	86	.025

Tabelle A 9. Älteste Bewerbergruppe: Signifikante Mittelwertunterschiede zwischen den Untergruppen „Frühe Maturität“ und „Späte Maturität“.