
EMS Eignungstest für das Medizinstudium in der Schweiz 1998

Bericht über die Durchführung und Ergebnisse

avec un résumé en français

K.-D. Hänsgen und B. Spicher

Bericht 4 (1998)



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

EMS Eignungstest für das Medizinstudium in der Schweiz 1998

Bericht über die Durchführung und Ergebnisse

Berichte des Zentrums für Testentwicklung, Nummer 4

im Auftrag der Schweizerischen Hochschulkonferenz

Fribourg: Zentrum für Testentwicklung und Diagnostik 1998

Hänsgen, Klaus-Dieter; Spicher, Benjamin

unter Mitarbeit von Sabine Dobler, Sébastien Simonet,

Nicolas Burger, Steve Fürst, Maurizio Strazzeri

Redaktion: Judith Berger

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

Druck: Cric-Print, Fribourg/Schweiz

Inhaltsverzeichnis

ZUSAMMENFASSUNG	5
RÉSUMÉ	9
VORWORT	12
NUMERUS CLAUSUS UND MEDIZINSTUDIUM	17
ANMELDUNG ZU MEDIZINSTUDIUM UND TEST 1998	20
Anmeldungen, Testantritte und Zulassungen	20
Grösse der Testlokale	23
Testorte und Wunschuniversitäten	23
Testorte und Wohnkantone/Wohnorte bei AusländerInnen	25
TestteilnehmerInnen nach Alter und Geschlecht	26
Sprachgruppen nach Alter und Geschlecht	28
Alter und Maturitätsjahr	29
BESCHREIBUNG DES VERWENDETEN EIGNUNGSTESTS	30
Aufbau des Tests	30
Beispielaufgaben für die Untertests.....	31
1. Untertest: Muster zuordnen.....	32
2. Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	32
3. Untertest: Schlauchfiguren.....	33
4. Untertest: Quantitative und formale Probleme	34
5. Untertest: Konzentriertes und sorgfältiges Arbeiten.....	35
6. Untertest: Figuren lernen.....	35
7. Untertest: Fakten lernen.....	36
8. Untertest: Textverständnis.....	36
9. Untertest: Diagramme und Tabellen	37
Berechnung der Werte	39
Vorhersage des Studienerfolges in Deutschland	40
TESTANWENDUNG IN DER SCHWEIZ 1998	42
Organisation der Testabnahme	42
Verteilungsprüfung	42
Äquivalenz der Testformen 1 und 2	45
Äquivalenz der Sprachversionen	46
<i>Ursachen für unterschiedliche Testleistungen</i>	46
<i>Sprachvergleich für die Untertests</i>	49
<i>Analyse der Sprachversionen auf Itemebene</i>	52
<i>Gleichbehandlung der Sprachgruppen</i>	64
Vergleichbarkeit der Testorte bzw. Testlokale.....	68
Vergleich der Geschlechter	71
Vergleich der Kantone innerhalb der Deutschschweiz	73
Vergleiche für Altersgruppen und Zeitpunkt des Maturitätsabschlusses	76
Vergleiche der Wunschuniversitäten.....	81

ERGEBNISSE ZUR TESTGÜTE	82
Zuverlässigkeit	82
Binnenstruktur.....	83
Item-Trennschärfen.....	88
Konzentriertes und sorgfältiges Arbeiten	91
TYPENANALYSE	94
GENAUIGKEIT DER BELEGERFASSUNG	96
Einlesen der Teile A und B (Multiple Choice)	97
Einlesen Teil K (Konzentriertes und sorgfältiges Arbeiten).....	99
LITERATUR.....	101
Originaltests zur Information und Vorbereitung.....	102
Frühere Berichte des ZTD.....	103
ANHANG ZUM BERICHT	105
Verteilung der Punktwerte	106
Testformen	111
Sprachgruppen	112
Testorte.....	115
Geschlecht.....	118
Kantone innerhalb der Deutschschweiz.....	121
Alter und Maturitätsabschluss.....	124

Zusammenfassung

Organisation und Zulassung

1. Der Eignungstest für das Medizinstudium in der Schweiz (EMS) war 1998 für Personen obligatorisch, die sich in **Basel, Bern, Fribourg** oder **Zürich** für das Fach **Humanmedizin** angemeldet haben. Einer Kapazität von 620 Studienplätzen stand die Zahl von 1020 Bewerberinnen und Bewerbern für dieses Fach gegenüber (Anmeldung bei der Schweizerischen Hochschulkonferenz am 15.2.1998). Entsprechend der gesetzlichen Grundlagen der genannten Kantone wurde der Numerus Clausus (NC) als Notmassnahme lediglich für das Fach Humanmedizin angewendet, weil aufgrund der Bewerbungen in den Fächern Veterinär- und Zahnmedizin 1998 nicht mit wesentlichen Überschreitungen der Kapazitäten gerechnet werden muss.
2. Es war für die Kandidatinnen und Kandidaten, für die der NC zur Anwendung kommt, eine Anmeldung zum Test erforderlich. Insgesamt **797 Personen** hatten sich bis zum Stichtag 19.5.1998 bei der Schweizerischen Hochschulkonferenz zum Test angemeldet.
3. Der Test fand **am 3.7.1998 mit 750 Personen** zur gleichen Zeit und unter gleichen Bedingungen an insgesamt 7 Testorten in 15 Testabnahmestellen statt. Zu beachten ist, dass die Wahl des Testortes unabhängig vom zukünftigen Studienort möglich war. Entsprechend der Empfehlung wurde am häufigsten der dem Wohnort nächste Testort gewählt.
4. Vor dem Testtermin haben 31 Personen der 797 die Anmeldung wieder zurückgezogen, 16 Personen sind ohne vorherige Abmeldung nicht zum Test angetreten. Die **Rückzugsquote von insgesamt 26%** gegenüber den 1020 Anmeldungen zum Studium im Februar 1998 entspricht in etwa derjenigen, die in den Vorjahren bereits zu verzeichnen war. Geht man davon aus, dass die Anmeldung **zum Test** und damit die Bereitschaft zur Testabnahme bereits eine differenziertere Motivation zur Studienaufnahme voraussetzt, würde die Zahl von 750 Personen in etwa derjenigen entsprechen, die in den Vorjahren ein Studium begonnen hätte.
5. **750 Personen** haben den EMS absolviert – alle Personen, die zum Test angetreten sind, haben den Test auch vollständig bearbeitet und mit gültigen Ergebnissen beendet. Von den TestleiterInnen wurde durchgängig eine sehr hohe Leistungsbereitschaft und Motivation aller Bewerberinnen und Bewerber bei der Testabnahme berichtet.
6. 31 Personen absolvierten den Test in **italienischer** (Bellinzona), 39 Personen in **französischer** (Fribourg) und 680 Personen in **deutscher** Sprache (Basel, Bern, Chur, St. Gallen, Zürich).
7. Es gab keine Störungen des Testablaufes oder Pannen bei der Organisation. Die Bedingungen der **Gleichbehandlung sind gegeben** und Benachteiligungen bei der Testabnahme – beispielsweise durch einen irregulären Ablauf in einzelnen Testlokalen - traten nicht auf.
8. Die Rate von insgesamt nur drei Verwarnungen (keine war schwerwiegend) und keinem Ausschluss vom Test ist sehr gering und spricht für einen sehr **disziplinierten Testablauf**.
9. Von den 750 Bewerberinnen und Bewerbern mit gültigen Testergebnissen erhielten 613 in der ersten Runde einen Studienplatz. Somit konnten **81,7%** der Testbesten berücksichtigt werden.

10. **3,9% der Personen**, denen in der ersten Runde ein Studienplatz zugewiesen wurde, haben ihre Bewerbung nicht bei den Universitäten bestätigt. Diese geringe Rückzugsquote nach dem Test stützt ebenfalls die Hypothese, dass die Entscheidung über die Aufnahme eines Medizinstudiums im wesentlichen vor der Testanmeldung getroffen wird. Eine zusätzliche dissuasive Wirkung durch den Test kann daher in diesem Jahr nicht festgestellt werden. Es wurde eine Warteliste gebildet, die auf der Höhe des Testwertes beruhte. Die mangels Bestätigung frei werdenden Plätze wurden in nachfolgenden Runden dann Personen aus der Warteliste zugewiesen.

Testgüte

11. Die **Zuverlässigkeit** des Tests ist mit der in Deutschland nachgewiesenen vergleichbar bezüglich der inneren Konsistenzen, der Reliabilitäten nach der Testhalbierungsmethode und der Profil-Reliabilitäten. Die Koeffizienten konnten fast identisch repliziert werden.
12. Die **Binnenstruktur** der Untertests folgt einer Dreifaktorenstruktur (logisches Denken, Wahrnehmung, Gedächtnis), die ebenso in Deutschland identifiziert worden ist. Dabei scheint die Struktur hierarchisch – es finden sich auch Belege für einen Generalfaktor allgemeiner Fähigkeiten.
13. Die Ergebnisse zeigen, dass die Bedingungen in den einzelnen **Testorten** vergleichbar sind. Auch die beiden parallelen Testformen 1 und 2 weisen keine unterschiedlichen Ergebnisse auf und können gemeinsam ausgewertet werden.

Unterschiede für Gruppen

14. Die Testwerte und die Zulassungsraten für **Männer und Frauen** unterscheiden sich nicht signifikant. Eine wichtige Voraussetzung für die Gleichbehandlung der Geschlechter bei der Zulassung zum Studium ist gegeben. Die Faktorenstruktur der Untertests (Drei-Faktoren-Lösung) stimmt für Männer und Frauen gut überein, was ebenfalls für gleiche qualitative Leistungskomponenten spricht.
15. Bezüglich des **Alters** lässt sich für jüngere Bewerberinnen und Bewerber ein relativer Vorteil bei der Zulassung nachweisen. Es dürfte im übrigen auch der Realität während des Studiums entsprechen, dass Jüngere die Prüfungen früher und mit besseren Noten bestehen. Bei den Älteren sind diejenigen besser, die eine Mittelschulausbildung sofort nach der Primar- bzw. Sekundarstufe aufgenommen und ihre Maturitätsprüfung damit zum frühestmöglichen Zeitpunkt abgelegt haben. Die frühe Wahl des Bildungsweges „Maturität“ ist für die Leistung im Test offenbar bestimmender als das „Noch-Trainiertsein“ durch einen kürzlich abgelegten Schulabschluss. Dies ist auch ein erneuter Hinweis auf eine bedingte Trainierbarkeit des Tests.
16. Es gibt keine Hinweise, dass **jüngere Personen aus unterprivilegierten Schichten** benachteiligt wären, wie vermutet wurde. Dies kann für solche Länder zutreffen, wo der Zugang zum Hochschulstudium durch längere Wartezeiten „ausgesessen“ werden kann - was dann natürlich auch von finanziellen Möglichkeiten abhängt. Beim in der Schweiz angewandten NC wird die Wartezeit nicht berücksichtigt. Die geringe Trainierbarkeit und Abhängigkeit von Faktenwissen macht den Test ebenfalls unabhängiger von sozialer Schichtzugehörigkeit als etwa Eignungsgespräche oder Wissenstests, wo durch intensiveres Training, welches Zeit und Geld kostet, bessere Leistungen erreicht werden könnten.

17. Zwischen den **Sprachgruppen** zeigen sich Leistungsunterschiede. Um Gleichbehandlung zu gewährleisten, müssen **gleiche Testwerte gleichen Schwierigkeitsstufen** entsprechen. Die Adaptation des Tests in eine französisch- und eine italienischsprachige Fassung erfolgte nach den international üblichen Regeln. Sprachbedingte und damit testbedingte Ursachen für die vorhandenen Unterschiede sind wahrscheinlich - auch aufgrund der Erfahrungen aus den USA und Israel waren solche Unterschiede zu erwarten. Dies ist kein Mangel der Übersetzungsqualität. Besonders bei der Übertragung sprachlich komplexer Aufgaben ist aufgrund unterschiedlicher Eigenschaften der Sprachen eine identische Schwierigkeit für die Sprachformen eher unwahrscheinlich. Es ist grundsätzlich **notwendig**, solche **testbedingten Unterschiede auszugleichen**.
18. In den sehr **sprachabhängigen Untertests** sind die **Unterschiede** geringfügig deutlicher als in den weniger sprachabhängigen – allerdings unterscheiden sich auch weniger sprachabhängige Untertests signifikant zwischen den Sprachgruppen. Neben den testbedingten Ursachen müssen kulturelle Unterschiede und/oder eine unterschiedliche Repräsentativität der Testteilnehmenden für die gesamte Sprachgruppe mit in Betracht gezogen werden. Die Gründe für unterschiedliche Repräsentativität liegen auch in den möglichen Alternativen, das Studium an einer anderen französischsprachigen Universität der Schweiz bzw. in Italien aufzunehmen, welche für zwei Sprachgruppen bestehen. Es ist nicht bekannt, wie sich das auf die Zusammensetzung der verschiedenen Sprachgruppen auswirkt. Die faktorenanalytisch ermittelten Unterschiede in den Strukturen der Testvariablen für die drei Sprachgruppen sprechen für kulturell bedingte Unterschiede.
19. Weil die französisch- und italienischsprachige Gruppe in diesem Jahr sehr klein war und die Gefahr einer „Überrepräsentation“ aufgrund höherer Maturitätsquoten dadurch praktisch nicht bestand, erfolgte die **Herstellung vergleichbarer Schwierigkeiten** durch separate Normierung für die drei Sprachformen (Berechnung des Testwertes aufgrund von Mittelwert und Standardabweichung der jeweiligen Sprachgruppe). Ohne diesen Schritt wären schlechtere Leistungen der französisch- und italienischsprachigen Gruppe möglicherweise *zumindest* in Teilen auf den Test zurückzuführen gewesen. Das wäre ein Verstoß gegen das Prinzip der Gleichbehandlung gewesen. Der Ausschluss dieses Risikos ist allerdings damit verbunden, dass kulturell bedingte und auf mangelnde Vergleichbarkeit der Repräsentativität der Sprachgruppen zurückzuführende Leistungsunterschiede mit ausgeglichen werden mussten. Aufgrund der geringen Gruppengrößen können diese empirisch nicht weiter differenziert werden.
20. **Qualifiziertere Verfahren** zur Herstellung einer vergleichbaren Schwierigkeit der Testwerte scheitern an einer zu geringen Zahl zur Verfügung stehender Aufgaben (Elimination der sich unterscheidenden Aufgaben pro Untertest), am Aufwand (Voruntersuchungen an grossen Stichproben mit Zweisprachlern zur Berechnung von „Ankerwerten“ als Vergleichsmassstab sind notwendig) oder an der Stichprobengrösse (Identifikation der sich unterscheidenden Items und Korrektur nur dieser Items mit Lineargleichungen aufgrund von „Ankerwerten“ der sich nicht unterscheidenden Items). Letzteres wird mittelfristig als eine machbare Alternative gesehen, wenn sich die Gruppengrößen verändern oder über mehrere Jahre die Stabilität der Unterschiede verfolgt werden kann.
21. **Ein machbares Verfahren** zur Herstellung gleicher Schwierigkeit, welches differenzierter als die 1998 gewählte Lösung ist, wäre eine Korrektur auf Untertestebene: Korrektur (1) nur sich statistisch signifikant unterscheidender Untertests, die (2) zu den 5 sprachlich anspruchsvollen Untertests gehören. Es würde die mittlere Differenz der „schlechteren“ Sprachgruppen zur „besten“ der jeweiligen

„schlechteren“ Sprachgruppe als Punktwert im jeweiligen Untertest als Schwierigkeitskorrektur zugerechnet. Die Ergebnisse in nichtsprachlichen Tests würden belassen wie sie sind, nicht alle Leistungsunterschiede würden dadurch eliminiert. Die Standardisierung des Punktwertes zum Testwert erfolgt dann aufgrund der Gesamt-Verteilung. Dieses Verfahren wird vom Beirat „Eignungstest“ der Schweizerischen Hochschulkonferenz favorisiert.

22. **Eine machbare Erweiterung** dieses Verfahrens könnte darin bestehen, diesen Ausgleich nur auf diejenigen Aufgaben (Items) zu begrenzen, die sich tatsächlich signifikant zwischen den Sprachgruppen unterscheiden (eine bestimmte Abweichung von der Regressionsgeraden im Schwierigkeitsdiagramm wird überschritten, s.u.). Die Bestimmung des Signifikanzkriteriums wäre möglich – würde aber durch die geringen Stichprobengrößen erschwert.

Regionale Unterschiede

23. Es gibt keine Hinweise dafür, dass sich **Regionen** der Deutschschweiz bezüglich des Testwertes unterscheiden. In der Deutschschweiz zeigen weder eine kantonspezifische Auswertung (für solche mit mehr als 40 Teilnehmern) noch eine Gruppierung der Kantone nach der **Maturitätsquote** bedeutsame Unterschiede der Testleistungen. Nach der Herstellung der Äquivalenz der Sprachversionen gilt dies auch für alle Regionen.
24. Die Zulassungsquoten unterscheiden sich nicht signifikant für die Gruppen nach der ersten Wahl der Universität (Wunschuniversität).

Schlussfolgerungen

Die **Verwendbarkeit** des Tests als Zulassungskriterium für das Medizinstudium in der Schweiz wurde nunmehr auch praktisch unter Beweis gestellt. Für die Zukunft sind vor allem zur weiteren Optimierung der Vergleichbarkeit unterschiedlicher Sprachversionen Entscheidungen zu treffen, die nun auf einer genaueren Analyse der Ergebnisse von 1998 beruhen können. Die verblüffend genaue Replizierbarkeit der aus Deutschland bekannten Gütekriterien des Tests in der Schweiz lässt hoffen, dass auch die guten Prognosewerte für den Studienerfolg in der Schweiz gelten. Damit wäre das Ziel erreicht, dass diejenigen bevorzugt einen Studienplatz erhalten, die bessere Chancen für ein Bestehen der Prüfungen in kürzerer Zeit haben. Die vorhandenen Ausbildungskapazitäten würden effektiv genutzt und folglich würde das Studium effizienter absolviert als unter den jetzigen Bedingungen. Die Verbesserung der Studienbedingungen wird sich auch auf die Entwicklung der Persönlichkeit hin zu einer guten Ärztin bzw. einem guten Arzt förderlich auswirken.

Résumé

Organisation et admission

1. En 1998, le test d'aptitudes pour les études de médecine en Suisse (EMS) était obligatoire pour les personnes qui s'étaient inscrites aux études de médecine humaine à **Bâle, Berne, Fribourg ou Zurich**. 1020 candidats et candidates se sont inscrits dans cette branche alors que la capacité s'élève à 620 places d'études (inscription auprès de la Conférence universitaire suisse jusqu'au 15.2.1998). Conformément aux bases légales des cantons concernés, le numerus clausus a été mis en oeuvre comme mesure d'urgence, uniquement pour la discipline de la médecine humaine, car d'après le nombre de candidatures aux études de médecine dentaire et vétérinaire en 1998, un dépassement important des capacités d'accueil de ces disciplines n'était pas à craindre.
2. Le 19.5.1998, jour de référence, un total de **797** personnes étaient inscrites au test auprès de la Conférence universitaire suisse.
3. 750 personnes ont effectué le test le 3.7.1998, simultanément et dans des conditions semblables, dans 7 villes et dans 15 locaux différents. Il faut noter que le choix du lieu de test restait indépendant du futur lieu d'études. Les participants au test ont le plus souvent choisi le lieu en fonction de sa proximité de leur domicile.
4. 31 personnes parmi les 797 inscrites se sont retirées avant le jour du test, le 3.7.1998; 16 personnes ne se sont pas présentées au test, sans avoir retiré auparavant leur inscription. Le **taux de retraits de 26%** sur l'ensemble des 1020 inscriptions aux études en février 1998 est comparable à ceux des années précédentes. Si l'on part du principe que le fait de s'inscrire au test et d'accepter de le passer présuppose déjà une motivation particulière pour entreprendre ces études, le chiffre de 750 personnes correspondrait approximativement au nombre de personnes qui auraient entrepris ces études les années précédentes.
5. **750 personnes** ont effectué l'EMS - toutes ces personnes l'ont effectué entièrement et l'ont accompli avec des résultats valables. Les examinateurs responsables ont constaté une très forte motivation chez tous les candidats pendant le test.
6. 31 personnes ont effectué le test en **italien** (Bellinzone), 39 personnes en **français** (Fribourg), et 680 personnes l'ont passé en **allemand** (Bâle, Berne, Coire, St-Gall, Zurich).
7. L'organisation et le déroulement du test n'ont rencontré aucun problème. Les conditions de **l'égalité de traitement ont donc été remplies** et personne n'a été désavantagé durant le test - en le passant par exemple dans des conditions non conformes dans un local particulier.
8. Le petit nombre d'avertissements, trois seulement (et aucun n'était sévère), et l'absence d'exclusion du test permettent de conclure que **le test s'est déroulé de manière très disciplinée**.
9. Parmi les 750 candidats qui ont obtenu un résultat valable au test, 613 (**81,7% de participants au test**) ont obtenu une place d'études lors de la première série d'attributions.
10. **3,9 % des personnes** qui ont obtenu une place d'études lors de la première série d'attributions n'ont pas confirmé leur inscription à l'université. Ce faible taux de retraits après le test étaye l'hypothèse que la décision d'entreprendre des études de médecine est prise

en principe avant l'inscription au test. On n'a donc pas constaté cette année d'effet dissuasif suscité par le test. Au cours des séries suivantes d'attributions, les places d'études dont l'utilisation n'a pas été confirmée auprès de l'université ont été octroyées aux personnes sur la liste d'attente, dans l'ordre décroissant des résultats obtenus au test.

Qualité du test

11. La **fiabilité** du test pour la Suisse est comparable à celle qui prévaut en Allemagne, en ce qui concerne la consistance interne, la fiabilité selon la méthode „split-half“ et la fiabilité des profils. Les coefficients ont pu être reproduits de manière quasi identique.
12. La **structure interne** des sous-tests se fonde sur trois facteurs (raisonnement logique, perception, mémoire), qui ont également été identifiés en Allemagne. Cette structure apparaît comme hiérarchique - il existe des preuves que le test comporte aussi un paramètre global mesurant les capacités générales.
13. Les résultats démontrent que les conditions de travail ont été comparables dans les différents **locaux**. Les deux formes pseudoparallèles (1 et 2) du test ne comportent également pas de différences dans les résultats et peuvent être analysées en commun.

Différences entre divers groupes

14. Les résultats obtenus au test et les taux d'admission **des hommes et des femmes** ne diffèrent pas de manière significative. Une condition importante pour l'égalité de traitement des sexes dans l'admission aux études est donc remplie. La structure en facteurs du sous-test (solution en trois facteurs) correspond bien, autant pour les femmes que pour les hommes, ce qui suggère également des composantes qualitatives de performance identiques.
15. En ce qui concerne **l'âge**, on constate que les jeunes candidats ont plus de chances d'être admis aux études que les candidats plus âgés. Cela devrait d'ailleurs aussi correspondre à une réalité durant les études: les jeunes étudiants réussissent leurs examens plus tôt et avec de meilleures notes. Parmi les candidats plus âgés, ceux qui ont suivi des études secondaires directement après l'école obligatoire et qui ont donc passé leur examen de maturité au plus tôt obtiennent de meilleurs résultats. Pour ce groupe de personnes, on peut considérer que le choix d'effectuer dans leur jeunesse une formation scolaire menant à la maturité est plus décisif pour leur performance au test que le fait d'être "encore entraîné" par un examen scolaire passé récemment. Cela démontre à nouveau les limites des possibilités d'entraînement au test.
16. Rien n'indique que **des personnes jeunes issues de couches sociales défavorisées** soient désavantagées, comme cela a pu être supposé. Cela peut être le cas pour des pays où l'accès aux études supérieures peut être soumis à de longues périodes d'attente: la possibilité d'attendre "son tour" dépend alors naturellement aussi des ressources financières individuelles. Le NC mis en oeuvre en Suisse ne prend pas en considération de telles périodes d'attente. La possibilité limitée de s'entraîner au test et sa relative indépendance d'un savoir acquis rendent les résultats obtenus par le candidat moins dépendants de son appartenance sociale que s'il avait effectué des entretiens d'aptitudes ou des tests portant sur des connaissances acquises, où une préparation importante permet d'améliorer les performances.

17. Les **groupes linguistiques** ont accompli des performances inégales. Il est alors indispensable que des **résultats identiques** obtenus dans différentes versions linguistiques correspondent à **un même degré de difficulté**, afin d'assurer l'égalité de traitement. L'adaptation du test en français et en italien s'est faite selon les règles internationales habituelles. Il est probable que les divergences observées ont leur origine dans la langue, et donc dans le test lui-même - les expériences faites aux USA et en Israël vont aussi dans ce sens. Cela n'est pas dû à un manque de qualité de la traduction. Lors de la traduction de questions qui ont une formulation complexe, il est peu probable que le degré de difficulté reste identique dans les différentes versions, étant donné les propriétés différentes de chaque langue. Il est en principe nécessaire de compenser de telles différences inhérentes au test.
18. Les **différences** sont légèrement plus marquées dans les sous-tests où la langue joue un rôle majeur - toutefois certains sous-tests peu dépendants de la formulation révèlent également des divergences significatives entre les groupes linguistiques. En plus des causes inhérentes au test, il faut prendre en considération des différences de nature culturelle et/ou une représentativité différente des participants au test par rapport à leur groupe linguistique dans son ensemble. A l'origine des différences dans la représentativité, il y a probablement aussi les alternatives qui restent ouvertes à ces deux groupes linguistiques: celles d'entreprendre des études en Italie ou dans une autre université francophone en Suisse. Leur impact sur la composition des différents groupes linguistiques n'est pas connu. Les divergences entre les trois groupes linguistiques dans les structures des variables du test, divergences qui ressortent de l'analyse des facteurs, vont dans le sens d'une origine culturelle de ces différences.
19. Comme les groupes francophone et italoophone étaient très petits cette année - ce qui élimina le danger d'une "surreprésentation" en raison de leur taux de maturités plus élevé - la réalisation **de degrés comparables de difficulté** s'est faite par la standardisation séparée des résultats des trois groupes linguistiques (calcul du résultat au test sur la base de la moyenne et de l'écart type de chaque groupe). Sans une telle démarche, les performances inférieures des groupes francophone et italoophone pourraient au moins en partie être imputées au test lui-même, ce qui aurait été un manquement au principe de l'égalité de traitement. En excluant ce risque, on corrige toutefois également des divergences dans la performance qui sont de nature culturelle ou qui sont liées à une représentativité différente des groupes linguistiques. En raison de la petite taille des échantillons, on ne peut pas différencier ces divers facteurs de manière empirique.
20. Les **procédures plus qualifiées** pour établir un même niveau de difficulté des résultats échouent en raison du nombre insuffisant de questions à disposition (élimination des questions qui se différencient dans chaque sous-test), des efforts trop importants que cela impliquerait (analyses préliminaires basées sur des échantillons importants avec des personnes bilingues pour calculer des "valeurs de référence" comme points de comparaison), ou en raison de la taille des échantillons (identification des items qui présentent des divergences et correction de ceux-ci uniquement, par des équations linéaires basées sur les "valeurs de référence" d'items qui ne se différencient pas selon la version linguistique). Cette dernière procédure est envisagée à moyen terme comme une possibilité, lorsque les échantillons seront plus grands ou lorsque la stabilité des différences pourra être observée sur plusieurs années.

21. **Une procédure possible** et plus précise que la solution choisie en 1998 pour rétablir un même niveau de difficulté consisterait en une correction au niveau des sous-tests: correction (1) uniquement des sous-tests révélant une divergence significative, qui (2) font partie des 5 sous-tests où la langue revêt une importance particulière. Les groupes linguistiques "moins bons" se verraient bonifier, sous la forme de nombre de points pour chaque sous-test concerné, la moyenne de la différence des résultats de leurs groupes par rapport au groupe "meilleur". On maintiendrait tels quels les résultats obtenus dans des tests où la langue a moins d'importance, et toutes les différences de performance ne seraient pas éliminées. La standardisation du nombre de points, qui permet d'obtenir le résultat au test, se ferait ensuite en fonction de leur répartition générale. Le comité consultatif "test d'aptitudes" mandaté par la Conférence universitaire suisse favorise cette procédure.
22. **Un élargissement possible** de cette procédure consisterait à limiter la correction aux seules questions (items), qui se différencient effectivement de manière significative d'un groupe linguistique à l'autre (dépassement d'un certain écart de la droite de régression dans le diagramme des difficultés).

Différences régionales

23. Aucun élément indique qu'il existerait des divergences entre **régions** de Suisse alémanique quant aux résultats obtenus au test. En Suisse allemande, ni une analyse spécifique par canton (pour ceux qui ont plus de 40 participants), ni un groupement des cantons selon leur **taux de maturités** ne révèlent des différences significatives. Une fois l'équivalence des versions linguistiques réalisée, cela est valable pour toutes les régions.
24. Les taux d'admission ne se différencient pas de manière significative entre des groupes formés sur la base du choix de l'université (université de premier choix).

Conclusion

L'**utilité** du test comme critère d'admission aux études de médecine en Suisse a donc aussi été prouvée dans la pratique. A l'avenir, il s'agira avant tout de prendre des décisions qui pourront se fonder sur une analyse plus détaillée des résultats, afin d'optimiser les possibilités de comparaison des versions linguistiques. Comme les critères de qualité du test qui prévalent en Allemagne ont pu être reproduits en Suisse, on peut espérer que sa valeur pronostique en ce qui concerne la réussite aux études soit également valable pour la Suisse. Le but serait alors atteint: les personnes qui ont le plus de chances de réussir les examens dans les meilleurs délais obtiennent en priorité une place d'études. Les capacités de formation existantes seraient exploitées de manière efficace et les études se dérouleraient par conséquent de façon plus efficiente que sous les conditions actuelles. Cette amélioration des conditions d'études aura également un impact favorable sur le développement de la personnalité des étudiants pour qu'ils deviennent de bons médecins.

Vorwort

Nach drei Jahren der Vorbereitung kam der Eignungstest für das Medizinstudium in der Schweiz 1998 erstmalig zum Einsatz. Die gesetzlichen Grundlagen für einen Numerus

Clausus wurden in allen Hochschulkantonen nunmehr geschaffen und beispielsweise 1997 mit einer Zweidrittelmehrheit im Kanton Bern und 1998 mit einer Dreiviertelmehrheit im Kanton Zürich vom Volk gutgeheissen.

Wie in den Vorjahren übersteigt 1998 die Nachfrage nach Studienplätzen die Kapazitäten für das Medizinstudium in der Schweiz deutlich. Die Anwendung des Numerus Clausus (NC) wird für die an einer Zulassungsbeschränkung beteiligten Universitäten jährlich neu geprüft und erfolgt als Notmassnahme angepasst an die tatsächliche Kapazitätsüberschreitung: Im Fach "Humanmedizin" wurden die Kapazitäten 1998 für die Universitäten Basel, Bern, Fribourg und Zürich von insgesamt 620 Plätzen mit 1020 Anmeldungen deutlich überschritten und auch bei Unterstellung der bisher üblichen Rückzugsquoten hätten die Zahlen der Studienanfängerinnen und -anfänger bereits vorhandene Probleme weiter verschärft. So erreicht beispielsweise in diesem Jahr in Zürich allein die Zahl der Repetentinnen und Repetenten der 1. Vorprüfung die Kapazität, die für das neue Studienjahr insgesamt zur Verfügung steht.

In den Fächern Zahn- und Veterinärmedizin war die Situation bezüglich der Aufnahmekapazitäten dagegen zumindest in diesem Jahr weniger problematisch, und es wurden dort alle Bewerberinnen und Bewerber zugelassen.

Die Universitäten Genf, Lausanne und Neuenburg sind darauf angewiesen, das Problem der gleichfalls vorhandenen Kapazitätsüberschreitung durch eine verschärfte inneruniversitäre Selektion zu lösen. Die Propädeutikumsprüfungen müssen hier entsprechend anspruchsvoll gestaltet werden, wodurch sich faktisch einem NC ähnliche Selektionswirkungen ergeben. Ein Nachteil dieser Lösung ist, dass in den ersten Jahren die Kapazitäten weiter überschritten werden, was sich ungünstig auf die Studienbedingungen auswirkt und z.B. die Umsetzung individualisierter Studienformen während dieser Zeiträume erschwert. Die allen Studierenden vorab bekannte Notwendigkeit zur verschärften intra-universitären Selektion fördert ein Konkurrenzdenken zwischen den Studierenden, welches für die Persönlichkeitsentwicklung - für teamorientiertes ärztliches Wirken etwa - nicht vorteilhaft sein kann.

Wichtige Veränderungen durch den voruniversitären NC sollen vor allem darin bestehen, die Studienbedingungen bereits im Grundstudium zu verbessern und im Rahmen der Studienreform neue Formen des Studiums zu ermöglichen, beispielsweise problemorientierten Unterricht in Kleingruppen.

Die statistischen Prognosen legen nahe, dass sich die Situation nicht von selbst entspannen wird. Eine steigende Maturitätsquote führt langfristig auch zu einer verstärkten Nachfrage nach Studienplätzen auf dem Gebiet der Medizin. Zusätzlich führt die Verkürzung der Maturitätsdauer von 13 auf 12 Jahre während der Übergangsphase zu Mehrbelastungen, weil sich in einigen Kantonen im Übergangsjahr zwei Jahrgänge um die Plätze eines Studientermins bewerben werden.

In der Diskussion über das beste Auswahlkriterium für den Numerus Clausus besteht bezüglich der **Studieneignung** weitgehend Konsens: Es sollen diejenigen bevorzugt eine Chance zum Medizinstudium erhalten, die dieses mit grosser Wahrscheinlichkeit auch erfolgreich beenden. Bei über 30% Studienabbrüchen ist diese Entscheidung gut begründet. Neben den Kosten für die Universitäten bzw. die Kantone muss man hier auch die letztlich verlorene Zeit für die einzelnen Personen mit berücksichtigen - wenn teilweise mehrere Jahre in eine abgebrochene Ausbildung investiert worden sind.

Die als NC-Kriterium ebenfalls sehr erwünschte **Berufseignung** hat drei Probleme: (1) Sie ist zum Zeitpunkt vor dem Studium noch kaum differenziert - es ist weithin erst die Aufgabe des Studiums, diese zu entwickeln; (2) es existieren keine eindeutigen und wissenschaftlich gesicherten Kriterien, die festlegen, was diese Berufseignung eigentlich

ausmacht. Die Medizin mit ihren zahlreichen Teilgebieten wird bezüglich der Eignung auch kaum auf ein einziges „Merkmalsmuster“ zu reduzieren sein - und (3) entsprechend gibt es auch keine Messverfahren, die Berufseignung zuverlässig genug erfassen, um damit die Zulassung zum Studium gerecht zu regeln. Zweifellos ist die Studieneignung aber ein wichtiger Bestandteil der Berufseignung: nur nach erfolgreichem Absolvieren des Studiums kann der Beruf der Medizinerin/des Mediziners ergriffen werden. Auch nach dem Studium bleiben effektive Techniken zur Wissensaneignung in einem so dynamischen Gebiet wie der Medizin gefragt.

Bei der Prüfung dieser Studieneignung bleibt der **Eignungstest** hinsichtlich vieler Kriterien die optimale Lösung. Er prognostiziert nachweislich den Studienerfolg von allen Kriterien am besten. Sowohl bezüglich der Prüfungsnoten als auch bei der Studiendauer sind die aufgrund der Testleistungen Zugelassenen besser als diejenigen, die z.B. aufgrund einer Warteliste oder von Eignungsgesprächen zugelassen worden sind (Trost 1998). Der Test ist ökonomisch vertretbar - nicht nur wegen der Durchführungskosten, sondern auch weil er im Gegensatz zu Eignungsgesprächen keine Verbesserungsmöglichkeiten durch aufwendiges Training bringt. Dies ist für die soziale Gleichbehandlung ein grosses Plus.

Der Test wurde aus Deutschland übernommen. Seit Februar 1986 wurden dort bei 12 Anwendungen des Tests insgesamt 300'727 Personen getestet (Pro Anwendung durchschnittlich 25'000 Personen, vgl. Trost 1998). Vom August 1980 bis August 1985 fanden darüber hinaus 11 Probeläufe des Tests statt (an mindestens je 5'000 Personen), bevor die Testergebnisse für die Studienzulassung verwendet worden sind. Nur drei Jahre Vorlauf in der Schweiz sind dagegen tatsächlich kurz zu nennen. Am Institut für Test- und Begabungsforschung in Bonn, welches mit der Testentwicklung beauftragt war, haben alle Aufgaben einen relativ aufwendigen Entwicklungs- und Revisionsprozess hinter sich. Daran waren zahlreiche Experten (Lehrbeauftragte aus den Fächern des medizinischen Grundstudiums, Gymnasiallehrer, Psychologen) beteiligt. Die hier verwendeten Aufgaben sind an mindestens zwei grösseren Stichproben bereits empirisch überprüft worden, sodass ihre Güte zumindest in Deutschland sehr gut abgesichert ist. Die deutsche Version wurde für die Schweiz sprachlich überarbeitet. Die Übersetzungen ins Französische und Italienische erfolgten nach den international üblichen Standards.

Man kann vom Test vor allem eine **Gleichbehandlung** erwarten: Die *Test-Info*, eine umfangreiche Broschüre mit Hinweisen für die Vorbereitung auf den Eignungstests, der genauen Erklärung seines Ablaufes sowie mit Beispielaufgaben und Lösungshinweisen für alle Untertests wird allen TeilnehmerInnen zur Verfügung gestellt. Es hat sich in mehreren Untersuchungen gezeigt, dass die dort empfohlene Vorbereitung ausreicht und vor allem hinsichtlich der Kosten aufwendigere Trainingsverfahren keinen nennenswerten Zugewinn bringen (vgl. Hofer u.a. 1997). Veröffentlicht wurden auch Vollversionen des Tests für die Vorbereitung in allen drei Sprachen, in denen die Testabnahme möglich ist. Die Durchführung findet überall unter gleichen Bedingungen statt, alle Zeiten und alle Instruktionstexte des Testleiters sind genau vorgegeben. Die Auswertung erfolgt maschinell mit einem hohen Sicherheitsstandard und Grenzfälle beim Markieren werden „von Hand“ kontrolliert. Die Berechnungsvorschriften für die Testwerte sind objektiv.

Man kann vom Test allerdings nicht erwarten, dass er „blind“ gegenüber Leistungsunterschieden ist, die in der Gesellschaft objektiv vorhanden sind. Er wird sie im Idealfall genauso abbilden müssen, wie sie tatsächlich bestehen. Er darf sie allerdings auch nicht überzeichnen. Dabei werden möglicherweise auch Fakten zutage gefördert, die eine entsprechende Diskussion anregen. Bisher existierten kaum ausreichend objektive Massstäbe für Vergleiche der Leistungsfähigkeit dieser Altersgruppe innerhalb der Schweiz - beispielsweise zwischen Geschlechtern, Kantonen oder Universitäten. Die Ergebnisse

belegen allerdings auch, dass viele der in den Diskussionen vorweggenommenen Unterschiede gar nicht in dem „befürchteten“ Ausmass vorhanden sind.

Die Vergleiche der Zulassungsquoten in diesem Bericht beziehen sich auf die **erste Runde** der Zulassung unmittelbar nach der Testdurchführung. Veränderungen, die sich in weiteren Runden im September 1998 ergeben haben, werden nicht berücksichtigt und sind bezüglich der Auswirkungen auf die Statistiken marginal. Es erfolgten Nachnominierungen von zunächst Abgelehnten auf Plätze von Personen, die ihre Anmeldung bei der jeweiligen Universität nicht bestätigt haben – dies geschah ebenfalls aufgrund der Rangreihe der Testwerte.

Welchen Einfluss hat die Tatsache, dass der Tests in Deutschland jetzt abgeschafft worden ist? Man muss zunächst festhalten, dass es auch in Deutschland beim Numerus Clausus für das Fach Medizin (wie für zehn weitere Fächer) bleibt und die Zulassung aufgrund des allgemeinen Auswahlverfahrens nach den Kriterien „Abiturnote“ und „Wartezeit“ erfolgt. Neben der sinkenden Bewerberquote waren es vor allem Kostengründe, die zur Abschaffung des Tests beitrugen. Immerhin hat es bereits im ersten Jahr des neuen Zulassungsverfahrens eine erneute Zunahme der Studienplatznachfrage bei Medizin um 28,9% (ZVS Info vom 2.9.1998) gegeben und einige Bundesländer erwägen nun ihrerseits, Tests und andere standardisierte Eignungsabklärungen wieder einzuführen. Dies soll vor allem für die neu eingeführten Quoten der direkt durch die jeweiligen Universitäten Zuzulassenden erfolgen – vgl. Deidesheimer Kreis (1997).

Die Chance der Schweiz besteht darin, einen sehr aufwendig entwickelten Test mit wissenschaftlich nachgewiesener Prognosegüte des Studienerfolges zu vergleichsweise moderaten Kosten nachnutzen zu können. Ihr bleiben allerdings wenige Alternativen, weil die Schulnoten hier viel weniger vergleichbar sind als in Deutschland. Auch dort müssen die notwendigen Durchschnittsnoten wegen der vorhandenen Unterschiede der Notenvergabe länderspezifisch festgelegt werden, was einer Quotierung sehr nahekommt. Eine Einführung von Quoten in der Schweiz würde an der Zahl und an der strukturellen Unterschiedlichkeit der Kantone wahrscheinlich schnell scheitern.

Zwischen Test und Abiturnote wurden in Deutschland langjährig Korrelationen von „nur“ 0.39 bis 0.44 festgestellt (Trost 1994). Der Überlappungsbereich beider Kriterien von nur 15 bis 19 Prozent wird von Trost als Beleg für hinreichend unterschiedliche Fähigkeits- bzw. Leistungsaspekte von Abitur und Test gewertet.

Dass die Testleistung langfristig die bessere Prognose, vielleicht auch hinsichtlich des Berufserfolges, liefert, wurde für die Medizin noch nicht speziell untersucht. In einer Studie der Wirtschaft konnte Hossiep (1998) unlängst empirisch nachweisen, dass die Prognosekraft für Berufserfolg von Eignungstests gegenüber Schulnoten mit der Länge des Vorhersagezeitraums immer deutlicher überwiegt. Wenn Eignungstests vor allem das **Potential** erfassen, sich in einer neuen, zukünftigen Situation zurechtzufinden, dann ist dies nicht überraschend. Schulnoten sind dagegen stärker abhängig vom bereits erworbenen Wissen, quasi von den *vergangenen* Bildungsanstrengungen – die sich im Verlaufe des Lebens z.B. hinsichtlich Intensität, Motivation oder Erfolg ändern können.

Der Bericht wird, wie wir hoffen, in ganz verschiedenen Fachkreisen Aufmerksamkeit finden. Es ist eine Gratwanderung, in jedem Teil sowohl allgemeinverständlich zu bleiben, als auch dem Experten die notwendige Information zur Bewertung der Ergebnisse zu geben. Deshalb wurden den Auswertungskapiteln die Diskussionen von Problemen und Ergebnissen in der Regel vorangestellt. Es folgen dann die ausführlichen Ergebnisdarstellungen und schliesslich die notwendigen statistischen Prüfungen. Jeder Leserin und jedem Leser bleibt so überlassen, wie weit man in die Tiefen der Statistik folgen will.

Einem Trend folgend, haben wir die Ergebnisse der statistischen Prüfungen als SPSS-Pivot-Tabellen weitgehend direkt in das Dokument übernommen. So wird es den Experten ermöglicht, die Signifikanzeinschätzungen nachzuvollziehen.

Unser Dank gilt allen Koordinatoren der Testorte, den TestleiterInnen und StellvertreterInnen, die für die erfolgreiche Durchführung des Eignungstests massgeblich die Verantwortung tragen.

Wir danken besonders Herrn Dr. G. Trost und den Kollegen des Instituts für Test- und Begabungsforschung (ITB) in Bonn für die langjährige Unterstützung bei der Adaptation des deutschen TMS für die Schweiz.

Der Universität Fribourg - ihrem Rektorat, Herrn Prof. M. Perrez als Ansprechpartner des Psychologischen Instituts, unseren Kolleginnen und Kollegen im Institut und den „services administratifs“ sind wir für die grosszügige Unterstützung beim Aufbau und der Entwicklung des ZTD zu grossem Dank verpflichtet.

Herrn Dr. N. Ischi und Herrn K. Wechsler von der Schweizerischen Hochschulkonferenz als dem Auftraggeber für den Eignungstest sowie allen Mitgliedern des Beirates "Eignungstest" möchten wir schliesslich für die Kritik und Hilfe danken, ohne die der Bericht nicht hätte entstehen können. Besonders möchten wir Herrn Prof. U. Schallberger (Zürich) und Herrn Prof. R. Bloch (Bern) für die psychometrische und fachliche Mithilfe danken.

Fribourg, November 1998

Klaus-D. Hänsgen, B. Spicher

Numerus Clausus und Medizinstudium

Numerus Clausus (NC) bedeutet Beschränkung der Zulassung zum Hochschulstudium. Die Zahl der Zulassungen wird nach den **Kapazitäten der Bildungseinrichtungen** und unabhängig von der Bewerberzahl festgelegt - es werden nur entsprechend viele Bewerberinnen und Bewerber zum Studium tatsächlich zugelassen, wie Plätze vorhanden sind. Die Beschränkung erfolgt bereits vor dem Studium – deshalb sind für die Umsetzung eines NC objektive und nachvollziehbare **Kriterien notwendig**, nach denen die Bewerberinnen und Bewerber ausgewählt werden können. Die Überschreitung der Ausbildungskapazitäten bedeutet in der Regel Überlastung der Universitäten, Sinken der Ausbildungsqualität oder Unmöglichkeit von Reformprojekten (z.B. Einführung eines problemorientierten Studiums in Kleingruppen).

In der Mehrzahl der entwickelten Industrieländer ist ein NC in einigen oder allen Fächern notwendig, weil die Ausbildungskapazitäten mit dem Drang nach Bildung nicht Schritt halten können (vgl. den Überblick für Europa von Ebach und Trost 1997 bzw. die im Bericht 4 des ZTD vorab erschienene Zusammenfassung von Trost 1996). Überall wächst die Zahl der Zugangsberechtigten zu Universitäten, die Anzahl der Studienplätze kann vor allem aus Kostengründen damit nicht Schritt halten. Der Numerus Clausus ist allerdings in erster Linie ein bildungspolitisches Instrument: die „Grenze“ ist nicht die Kapazität des Arbeitsmarktes. Die Wechselwirkung mit dem Arbeitsmarkt ist eher langfristig und indirekt, z.B. durch Beziehungen zwischen Ausbildungsnachfrage und medizinischer Versorgungsdichte.

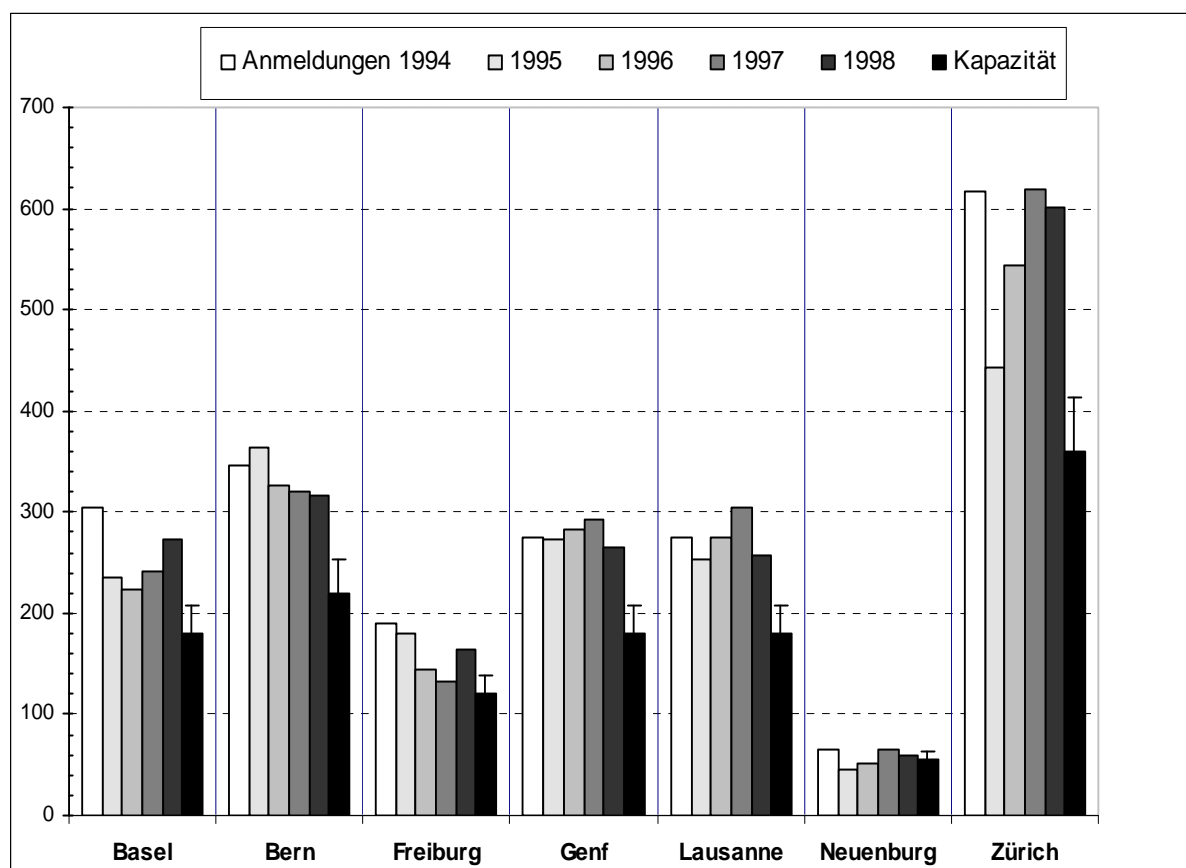


Abbildung 1: Anmeldungen zum Medizinstudium 1994 bis 1998, Anfängerprognose und Kapazitäten 1998 (letzte Säule mit Markierung des 15%-Überschreitungskriteriums als Grenze für die Notwendigkeit einer Zulassungsbeschränkung)

Die Attraktivität der Schweiz als Ausbildungsplatz und die Nachfrage nach in der Schweiz ausgebildeten Absolventinnen und Absolventen muss vor dem Hintergrund der weiteren Globalisierung gesehen werden: Bildung ist ein wichtiges Exportgut der Schweiz, welches es zu erhalten gilt. Eine Überlastung der Universitäten ist für die dafür zwingend notwendige Ausbildungsqualität sicher nicht förderlich.

Betrachtet man die Anmeldungen zum Studium 1994 bis 1998 für die sieben Universitäten, die Medizin im Grundstudium anbieten (Abbildung 1), wird eine beständige Überschreitung der Kapazitäten deutlich. Vor allem die Universität Zürich hat in den letzten drei Jahren die Kapazitäten deutlich überschreiten müssen. In dieser Abbildung sind Anmeldungen für die Teildisziplinen Human-, Veterinär- und Zahnmedizin zusammengefasst – erst ab 1998 wird dies bekanntlich unterschieden. Die Trends innerhalb der einzelnen Universitäten sind dabei aber nicht gleich. In Bern hat es beispielsweise von 1994 zu 1995 entgegen dem allgemeinen Trend eine weitere Steigerung gegeben.

Fasst man die Universitäten zusammen, die 1998 einen NC anwenden (Abbildung 2), wird für diese eine stetige Zunahme der Anmeldezahlen seit 1995 deutlich. 1994 war das Jahr, in welchem der bisherige Anmelderekord liegt. Vor allem Zürich und Basel verzeichneten in

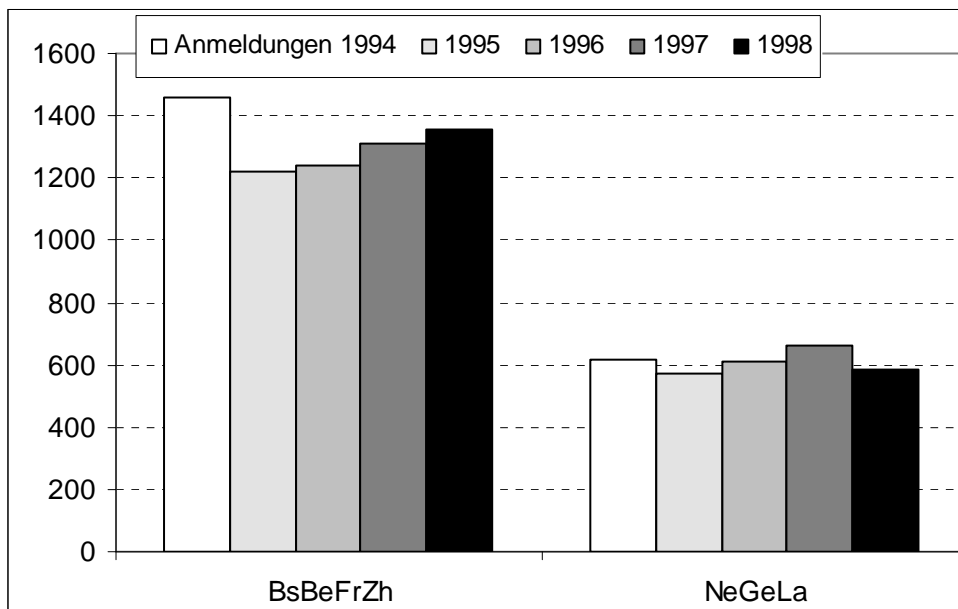


Abbildung 2: Anmeldungen zum Medizinstudium 1994 bis 1998 nach Gruppen (**B**asel, **B**ern **F**ribourg, **Z**ürich vs. **N**euenburg, **G**enf, **L**ausanne)

diesem Jahr eine starke Nachfrage. Daraufhin wurde versucht, durch gezieltere Information zum Medizinerberuf und das Öffentlichmachen der Probleme im Zusammenhang mit der Überlastung von Hochschulen einen „dissuasiven“ Einfluss auf das Bewerberverhalten auszuüben. Auch Diskussionen zur generellen Möglichkeit von NC und Test werden seit 1995 geführt, die eine dissuasive Wirkung haben könnten.

Die Universität Zürich hat nun dieses „Rekordniveau“ seit 2 Jahren wieder erreicht und in Basel geht der Trend auch mindestens seit diesem Jahr wieder in diese Richtung. In den Universitäten Neuenburg, Genf und Lausanne kehrt sich der Trend 1998 erstmals wieder um.

Die Abbildung 3 verdeutlicht, dass sich bis zum Jahr 2002 die Situation kaum von selbst entspannen dürfte: Durch Verkürzungen der Maturitätsdauer werden in einzelnen Kantonen über die nächsten 5 Jahre verteilt jeweils zwei Jahrgänge die Maturität im gleichen Jahr abschliessen und ein Studium beginnen wollen. Auch das Niveau der Studienanfänger nach diesem Zeitraum des Überganges wird den Prognosen nach noch deutlich über dem heutigen Niveau liegen.

Es wird natürlich von mehreren Faktoren abhängen, wie sich die Nachfrage nach Studienplätzen im Fach Medizin entwickelt. Da die Ärztedichte und das Niveau der Gesundheitskosten in der Schweiz bereits jetzt im europäischen Massstab relativ hoch sind (vgl. Hänsgen u. M. 1995b), würde der generelle Verzicht auf die Möglichkeit eines Numerus Clausus ein hohes Risiko für diesen Bereich beinhalten. Aus Sicht der Kantone gehen bildungs- und berufspolitische Aspekte spätestens hier ineinander über: die Medizinausbildung ist eine der teuersten Studienrichtungen, welche gleichzeitig aber auch eine Berufsausbildung darstellt. Es entsteht ein hoher Finanzbedarf für die Kantone (übrigens auch für Nichthochschulkantone durch die an die Hochschulkantone zu zahlenden Deckungsbeiträge). Ist man bereit, den bildungspolitisch notwendigen Beitrag eines freieren Zuganges zum Medizinstudium zu tragen, muss man sehr wahrscheinlich auch mit hohen Folgekosten im Gesundheitsbereich durch eine dann weiter erhöhte Ärztedichte rechnen. Oder man muss dann dort zu restriktiven Massnahmen wie Niederlassungsbeschränkungen oder Kostendeckelungen greifen, wie sie in Deutschland beispielsweise notwendig sind. Der wesentliche Nachteil dieser Lösung: die Betroffenen z.B. arbeitslose Mediziner, haben eine lange und teure Ausbildung hinter sich, die sich dann auch in der individuellen Bilanz nicht rechnet.

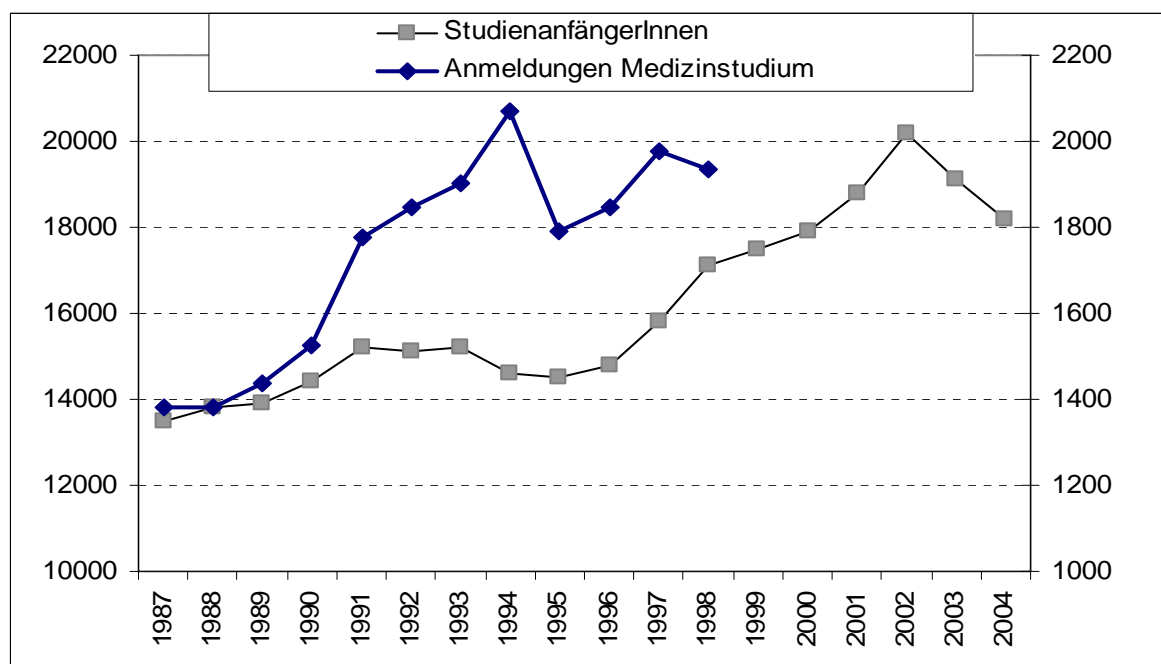


Abbildung 3: Linke Y-Achse: Prognose StudienanfängerInnen 1997 bis 2004 (aus: Heft 15 "Bildung und Wissenschaft" Wieviele Studierende an Schweizerischen Hochschulen?) Rechte Y-Achse: Anmeldungen zum Medizinstudium bei der Schweizerischen Hochschulkonferenz

Anmeldung zu Medizinstudium und Test 1998

Anmeldungen, Testantritte und Zulassungen

1020 Personen hätten 1998 aufgrund der Voranmeldungen (Frist bis 15.2.98) einen Test absolvieren müssen, um einen der 620 Studienplätze für Humanmedizin zu erhalten. Zum Test angemeldet haben sich 797 Personen (Frist bis 19.5.1998). Vor dem Testtermin haben 31 Personen die Anmeldung wieder zurückgezogen (diese Möglichkeit wurde seitens der SHK bis 2 Wochen vor dem Testtermin eingeräumt). 16 Personen sind ohne vorherige Abmeldung nicht zum Test angetreten. 750 Personen sind zum Test angetreten und haben ihn mit gültigen Ergebnissen beendet.

Nicht von vornherein zu erwarten war, dass alle Personen den Test mit gültigen Ergebnissen beenden. In Deutschland wie auch beim Probelauf in Fribourg sind solche Fälle des „Aufgebens“ aufgetreten. Dies deutet auf eine hohe Motivation und Anstrengungsbereitschaft bei der Testabnahme unter "Ernstfallbedingungen" in der Schweiz hin.

Bereits in früheren Jahren haben sich ca. 20% der Kandidatinnen und Kandidaten zum Medizinstudium angemeldet, ohne das Studium dann tatsächlich zu beginnen. Diese "präventiven" Anmeldungen werden vermutlich von Personen vorgenommen, die sich bezüglich des Studienwunsches noch nicht sicher sind. Die Anmeldung ist bekanntlich ab November des Vorjahres, also fast ein Jahr vor Studienantritt, möglich.

Die Anmeldung zum Test, die notwendigen Anstrengungen zur Vorbereitung (Hinweise der Test-Info) und wahrscheinlich auch die Notwendigkeit der Einzahlung der Teilnahmegebühr sind zusätzliche Hürden, die eine gereifere Studienmotivation und Anstrengungsbereitschaft voraussetzen sollten. Deshalb wurde der Rückzug in diesem Jahr vor allem im Zeitraum vor der Testabsolvierung erwartet.

In Tabelle 1 sind die Zahlen der Anmeldung im Verlauf dargestellt. 26% der Personen haben - gegenüber der Anmeldung zum Studium - den Test nicht absolviert. Dies entspricht in etwa der Rückzugsquote, die in den Vorjahren bereits zwischen Anmeldung und Studienantritt zu verzeichnen war. Geht man davon aus, dass die Anmeldung zum Test und die Bereitschaft zur Testabnahme bereits eine differenziertere Motivation zur Studienaufnahme voraussetzt, würde die Zahl von 750 Personen mit gültigen Testergebnissen am ehesten derjenigen entsprechen, die in den Vorjahren ein Studium begonnen hätte. Eine zusätzliche dissuasive Wirkung allein durch den Test könnte unter diesen Bedingungen nicht festgestellt werden.

Die Zulassung zum Studium erfolgte bis Ende September in mehreren Runden: Zunächst wurde für alle Personen mit einem Testwert von mindestens 91 ein Studienplatz vergeben. Auf die 613 in der ersten Runde vergebenen Plätze haben sich 589 Personen bei den Universitäten gemeldet. Nur 3,9% der Personen, denen ein Studienplatz in der ersten Runde zugewiesen wurde, haben ihren Studienplatz nicht bei den Hochschulen bestätigt. Damit steht fest, dass Rückzüge im wesentlichen bereits vor der Testanmeldung erfolgen. Ein ausreichendes Testergebnis wird dazu motivieren, das Studium dann auch zu beginnen.

In nachfolgenden Runden wurden die jeweils freien Plätze an die vom Testwert her Besten der „Warteliste“ vergeben: allen Personen mit einem Testwert ab 89 konnte so ebenfalls ein Studienplatz zugewiesen werden.

Hervorzuheben ist, dass in der ersten Runde nur 29 Umleitungen notwendig waren, d.h. der erstgenannte Studienort („Wunschuniversität“) konnte bei 95,3% der Bewerberinnen und Bewerber berücksichtigt werden.

	Basel	Bern	Fribourg	Zürich	Total
Kapazitäten	152	125	103	240	620
Anmeldung zum Studium bis 19.2.98	235	211	134	440	1020
Anmeldung zum Test Stand 19.5.98	180	169	108	340	797
Anmeldung zum Test Stand 2.7.98 früh	171	166	104	325	766
Antritt und erfolgreiche Absolvierung Test	167	163	101	319	750
Rückzugsquote in Prozent (bezogen auf Studienanmeldung)	29%	23%	25%	27%	26%

Tabelle 1: TestteilnehmerInnen nach gewünschtem Studienort, 2 BewerberInnen, die ihr Studium in Neuenburg beginnen und in Bern bzw. Zürich fortsetzen wollen, haben am Test teilgenommen und werden hier beim Studienort Bern bzw. Zürich gezählt.

Die Abbildung 4 zeigt, dass die Rückzüge sich auf alle Universitäten verteilen. Bis auf die Universität Fribourg bleiben alle anderen über der Kapazitätsgrenze.

Die folgende Kreisdarstellung (Abbildung 5) veranschaulicht den Verlauf von der Anmel-

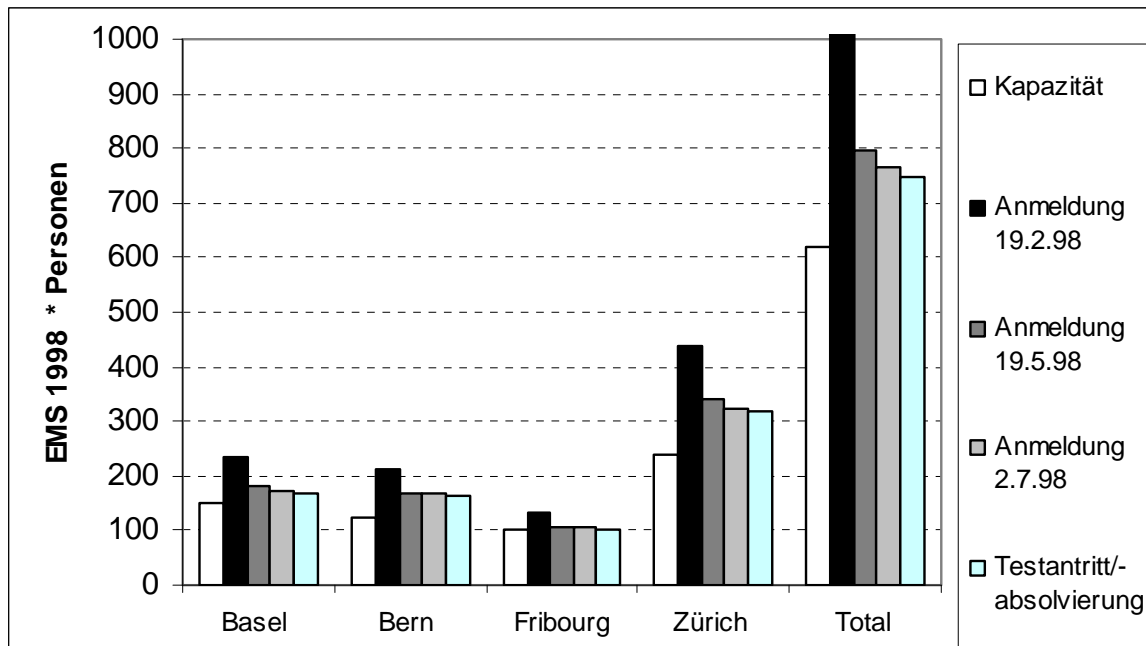


Abbildung 4: Anmeldeverlauf nach Wunschuniversitäten

Abbildung 5: Anteile der verschiedenen KandidatInnengruppen im Verlauf des Zulassungsverfahrens 1998, 1020 Anmeldungen als Basis.

Abbildung 5: Anteile der verschiedenen KandidatInnengruppen im Verlauf des Zulassungsverfahrens 1998, 1020 Anmeldungen als Basis

Grösse der Testlokale

Der Test fand in insgesamt 15 „Testlokalen“ (Hörsäle, Schulaulen, Veranstaltungsräume) statt. In Abbildung 6 sind die Anmeldezahlen und die Zahl der Personen dargestellt, die den Test erfolgreich beendet haben. Die Rückzüge verteilen sich unsystematisch auf die Testorte.

Während in Basel mehrere kleine Testlokale verwendet worden sind, hat sich Bern für zwei grössere Säle entschieden. Zürich hat ein gemischtes Konzept verwendet. Nach Analyse der Protokolle und der Berichte der Testleiter sind die Tests auch in grösseren Sälen problemlos durchzuführen - für eine ressourcenschonendere Testdurchführung werden zukünftig grössere Testlokale dort empfohlen, wo dies möglich ist. Es ist organisatorisch machbar, die Testabnahme in geeigneten Räumen auch mit 100 Personen durchzuführen.

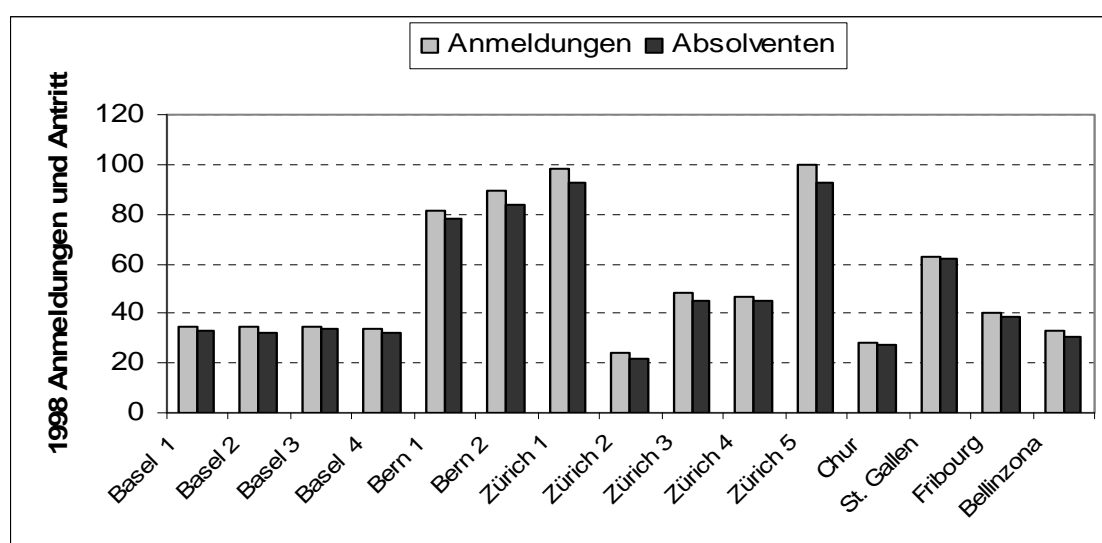


Abbildung 6: Testlokale: Anmeldungen und Absolvierung des Tests

Bei den Testorten mit nur einem Testlokal wurden teilweise sehr wenige Kandidatinnen und Kandidaten getestet. Dennoch scheint die gewählte Lösung ein Optimum bezüglich Anreiseweg, der Sprache des Tests (in einem Testlokal ist der Test jeweils nur in einer Sprache absolvierbar) und dem Organisationsaufwand zu sein.

Testorte und Wunschuniversitäten

Die Kandidatinnen und Kandidaten konnten eine Universität angeben, an welcher sie ihr Studium bevorzugt beginnen wollten. Diese wird nachfolgend mit „Wunschuniversität“ bezeichnet. Es war den Kandidatinnen und Kandidaten freigestellt, sich beim wohnortnächsten Testort anzumelden. Die Wahl des Testortes hatte keinen Einfluss auf die Zulassung. In Diskussionen vor der Testdurchführung wurde die Zahl der Angemeldeten zum Test in den Hochschulorten nicht selten mit der Zahl der Bewerberinnen und Bewerber für die jeweilige Universität fälschlicherweise gleichgesetzt. Aus Tabelle 2 wird ersichtlich, dass in Basel, Bern und Zürich zu etwa 3/4 Übereinstimmung besteht. Fribourg hat ein relativ weites "Einzugsgebiet". Bei Fribourg ist aber zu beachten, dass an diesem Testort der

Test in französischer Sprache abgelegt wurde - die deutschsprachigen Bewerber (auch aus dem Kanton selbst) also auf einen anderen Testort ausweichen mussten. Nur 8 Personen haben den Test in französischer Sprache absolviert, um sich dann an einer andern Universität der Deutschschweiz zu bewerben.

Wunsch-univ.	Testorte							Total
	Basel	Bern	Zürich	Chur	St.Gallen	Fribourg	Bellinzona	
Basel	126 75.4%	13 7.8%	8 4.8%	3 1.8%	7 4.2%		10 6.0%	167 100.0%
Bern		124 76.1%	14 8.6%	8 4.9%	8 4.9%	5 3.1%	4 2.5%	163 100.0%
Fribourg	5 5.0%	20 19.8%	31 30.7%	1 1.0%	11 10.9%	31 30.7%	2 2.0%	101 100.0%
Zürich		5 1.6%	245 76.8%	15 4.7%	36 11.3%	3 .9%	15 4.7%	319 100.0%
Total	131 17.5%	162 21.6%	298 39.7%	27 3.6%	62 8.3%	39 5.2%	31 4.1%	750 100.0%

Tabelle 2: Testorte und Wunschuniversität (erste Wahl)

Die nachfolgende Abbildung 7 macht durch die "Streckung" aller Universitäten auf 100% noch einmal deutlich, dass die Testorte Basel, Bern und Zürich in ihrem Verhältnis der "lokalen" Bewerber zu den übrigen etwa vergleichbar sind.

Testorte und Wohnkantone/Wohnorte bei AusländerInnen

Für differenziertere Analysen steht die nachfolgende Tabelle 3 zur Verfügung. Hier sind die Testorte im Verhältnis zum Wohnkanton tabelliert. Angegeben sind auch die Maturitätsquoten für 1997 (Jahresmitte).

Wohnkanton	Maturitätsquote	Testort							Summe
		Basel	Bern	Zürich	Chur	St.Gallen	Fribourg	Bellinzona	
AG	16,2	20	7	39					66
AI	13,7					2			2
AR	16,0					4			4
BE	12,9		93	2	1		4		100
BL	23,3	49	1						50
BS	20,7	38	1			1			40
FR	21,4		10			1	23		34
GE	29,6						1		1
GL	14,1			3	1				4
GR	13,2		1	1	17				19
LU	12,2	6	15	21		1			43
NE	24,9						2		2
NW	10,2	1	4	1					6
OW	12,2	1		2					3
SG	12,8		2	11	4	47			64
SH	16,6			7					7
SO	13,6	9	14	3					26
SZ	12,7			21					21
TG	9,8			16		6			22
TI	29,8							30	30
UR	13,0			3					3
VD	20,7						3		3
VS	19,4	5	11				6		22
ZG	16,4		1	5					6
ZH	18,7	1		162				1	164
Ausland		1	2	1	4				8
		131	162	298	27	62	39	31	750

Tabelle 3: Wohnkanton und Testorte, Maturitätsquote (Anteil Maturitäten an der 19jährigen Wohnbevölkerung) 1997 nach Bundesamt für Statistik Bern

687 Personen sind Schweizerbürgerinnen und Schweizerbürger. Die übrigen 63 verteilen sich auf insgesamt 26 Nationalitäten.

TestteilnehmerInnen nach Alter und Geschlecht

Die Alters- und Geschlechtsverteilung wird in der nachfolgenden Tabelle 4 dargestellt. 55% Frauen und 45% Männer haben sich insgesamt beworben. Der Anteil Frauen ist vor allem im Altersbereich "unmittelbar nach der Matura" höher. Männer sind in den Gruppen der Älteren häufiger vertreten. Die Altersverteilungen unterscheiden sich für Männer und Frauen signifikant auf 5%-Niveau (Tabelle 5).

	Altersklassen nach Geburtsjahr			Gesamt
	1952-1974	1975-1977	1978-1980	
Männlich (m)	40 11.9% m	115 34.2% m	181 53.9% m	336 44.8% v. Gesamt
Weiblich (w)	38 9.2% w	106 25.6% w	270 65.2% w	414 55.2% v. Gesamt
Gesamt	78 10.4% v. Gesamt	221 29.5% v. Gesamt	451 60.1% v. Gesamt	750

Tabelle 4: TestabsolventInnen nach Alter und Geschlecht

	Wert	Df	Asymp. Signifikanz (2-seitig)
Pearson Chi-Square	9.977	2	.007
Likelihood Ratio	9.970	2	.007

Tabelle 5: Prüfung auf Homogenität der Verteilung Alter x Geschlecht; signifikanter Unterschied.

Die Abbildung 8 veranschaulicht den Anteil der drei Altersgruppen für die Geschlechter grafisch: Es überwiegt der Frauenanteil, der sich vor allem aus der jüngeren Gruppe rekrutiert. Die beiden älteren Gruppen stimmen in etwa überein.

Abbildung 8: Anteil nach Alter (Geburtsjahr) und Geschlecht

Sprachgruppen nach Alter und Geschlecht

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

Schlüsselt man die Alters- und Geschlechtsverteilung nach den Sprachgruppen auf (Tabelle 6), wird deutlich, dass in der französisch- und der italienischsprachigen Gruppe die Anteile der Jüngeren wesentlich grösser sind. Es gibt dort kaum Personen beim Test, die ihr Studium nicht unmittelbar zum frühestmöglichen Zeitpunkt nach dem Ablegen der Maturitätsprüfung beginnen wollen.

SPRACHE		Altersklassen			Total
		1952-1974	1975-1977	1978-1980	
Deutsch	männlich	38 12.6%	110 36.5%	153 50.8%	301 44,3%
	weiblich	38 10.0%	102 26.9%	239 63.1%	379 55,7%
	gesamt	76 11.2%	212 31.2%	392 57.6%	680
Französisch	männlich		5 33.3%	10 66.7%	15 38,5%
	weiblich		3 12.5%	21 87.5%	24 61,5%
	gesamt		8 20.5%	31 79.5%	39
Italienisch	männlich	2 10.0%		18 90.0%	20 64,5%
	weiblich		1 9.1%	10 90.9%	11 35,5%
	gesamt	2 6.5%	1 3.2%	28 90.3%	31

Tabelle 6: Alter und Geschlecht aufgeschlüsselt nach Sprachgruppen

Alter und Maturitätsjahr

Die folgende Tabelle 7 macht deutlich, dass es unter den Älteren zwei Gruppen gibt: Personen die ihre Maturitätsprüfung unmittelbar vor dem Bewerbungszeitpunkt absolviert haben und solche, die nach der Maturitätsprüfung zum frühestmöglichen Zeitpunkt noch einer anderen Tätigkeit nachgegangen sind.

Geburtsjahr	Maturitätsjahr																			
	1972	1978	1981	1982	1983	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	
1952															1					1
1954	1																			1
1957																		1		1
1959		1																		1
1961			1											1				1		3
1962			2	1																3
1963				1	1															2
1964					1												1		1	3
1965																1			1	2
1966						1	1											1	2	5
1967								1												1
1968																		1	3	4
1969									2	1	1			1					2	7
1970									1	1									2	4
1971												2				1		1	2	6
1972												6		1					5	12
1973													1	2				1	6	10
1974														2	1	2		1	6	12
1975															2	4	4	5	1	16
1976																2	13	12	31	58
1977																	6	55	86	147
1978																1	2	43	318	364
1979																		4	80	84
1980																			3	3
	1	1	3	2	2	1	1	1	3	2	1	8	1	7	4	10	27	126	549	750

Tabelle 7: Geburtsjahr und Maturitätsjahr

Beschreibung des verwendeten Eignungstests

Aufbau des Tests

Neun verschiedene Untertests werden durchgeführt und zu einem Gesamtwert verrechnet. Die folgende Tabelle 8 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 8: Struktur und Ablauf des EMS, * Aufgabenzahl und in Klammern die Zahl der gewerteten Aufgaben pro Untertest und für Gesamt, wenn zusätzliche nicht gewertete Einstreuaufgaben verwendet worden sind, Erläuterungen siehe Text.

Die Struktur des deutschen TMS wurde für 12 Testtermine so beibehalten (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 bewährten Abläufe und damit möglicherweise auch die Güte-Kennwerte nicht zu verändern.

Die Erprobung neuer Aufgaben für die in Tabelle 8 gekennzeichneten Untertests erfolgte in Deutschland im Rahmen sogenannter „**Einstreuaufgaben**“. Erst 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 acht verschiedene Testformen bei jedem Durchführungstermin zusammengestellt worden sind, konnten jeweils 32 neue Aufgaben pro Untertest an ausreichend grossen Stichproben erprobt werden. Maximal 20 wurden davon wirklich gebraucht.

Damit die Testlänge des EMS mit derjenigen der deutschen Version vergleichbar ist, wurden auch in der Schweizer Version pro Untertest 4 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 8 zusätzliche Aufgaben pro Untertest erprobt werden konnten. Auf deren Ergebnisse wird im Rahmen dieses Berichtes nicht eingegangen. 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 10 Jahre angewendet. Dabei können einzelne Aufgaben aus verschiedenen 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 bereits erfüllt.

Beispielaufgaben für die Untertests

Nachfolgend wird pro Untertest eine Beispielaufgabe dargestellt. Dies kann lediglich die Prinzipien der Aufgabenstruktur verdeutlichen – die Aufgaben unterscheiden sich innerhalb jedes Untertests bezüglich des Schwierigkeitsgrades beträchtlich. 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 (s. Literaturverzeichnis).

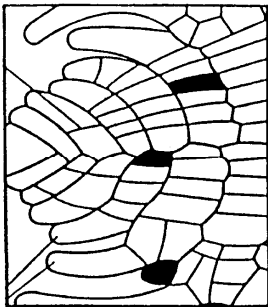
Bei der Beurteilung der Aufgaben wird die **Nähe zu Studienanforderungen** deutlich: an der Entwicklung haben neben Gymnasiallehrern, Medizinerinnen und Psychologen auch Lehrbeauftragte des Grundstudiums Medizin mitgearbeitet. Die Struktur der Untertests ist auf detaillierte Anforderungsanalysen eines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

1. Untertest: Muster zuordnen

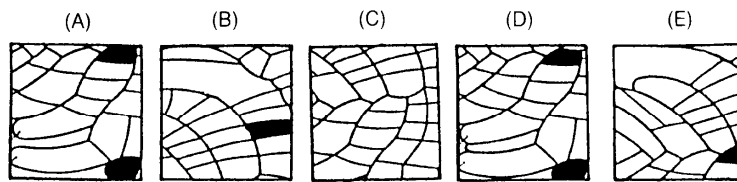
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wiederzuerkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

Ein Beispiel dazu:

Muster



Musterausschnitte



In den meisten Aufgaben dieser Art heben sich die vier nicht deckungsgleichen Musterausschnitte dadurch vom Muster ab, dass Details entweder hinzugefügt, oder weggelassen sind. Zugleich stellt dieser Untertest Anforderungen an die Schnelligkeit der Bearbeitung.

In durchschnittlich 55 Sekunden je Aufgabe muss die Testperson die richtige Lösung herausgefunden haben, dass beispielsweise in der obigen Aufgabe nur der Musterausschnitt (A) deckungsgleich mit einem Teil des Musters ist, und zwar in dessen unterem Bereich, etwa in der Mitte.

2. Untertest: Medizinisch-naturwissenschaftliches Grundverständnis

Hier wird das Verständnis für Fragen der Medizin und der Naturwissenschaften geprüft. Der Text könnte so in einem Lehrbuch stehen. Wichtig für das Verständnis dieser Textpassage ist, ob daraus bestimmte logische Schlüsse gezogen werden können. Alle Fakten, die für die Beantwortung der Aufgabe notwendig sind, stehen im Text – spezielles medizinisches Vorwissen ist nicht erforderlich. Dieses wichtige Prinzip findet sich bei allen Untertests und ist verantwortlich für die geringe Trainierbarkeit der Aufgabenlösung.

Im Kindesalter kann das Zentrum für Sprache, Spracherwerb und Sprachverständnis noch in der linken oder in der rechten Hälfte (Hemisphäre) des Gehirns in einem umschriebenen Hirnrindengebiet (sog. Sprachregion) angelegt werden. Spätestens im zwölften Lebensjahr sind die sprachlichen Fähigkeiten jedoch fest in einer der beiden Hemisphären verankert, und zwar bei den Rechtshändern in der Regel links, bei den Linkshändern in der Mehrzahl ebenfalls links, zum Teil aber auch rechts; die korrespondierende Region der Gegenseite hat zu diesem Zeitpunkt bereits andere Funktionen fest übernommen. Welche der nachfolgenden Aussagen lässt bzw. lassen sich aus diesen Informationen ableiten?

Bei irreversiblen Hirnrindenverletzungen im Bereich der sogenannten Sprachregion der linken Hemisphäre ...

- I. kommt es bei erwachsenen Linkshändern in der Regel zu keinen wesentlichen Sprachstörungen.
 - II. kommt es bei einem Vorschulkind in der Regel zu einer bleibenden Unfähigkeit, die Muttersprache wieder zu erlernen.
 - III. ist bei zwanzigjährigen Rechtshändern die Fähigkeit, eine Sprache zu erlernen, in der Regel verloren gegangen.
- (A) Nur Ausfall I ist zu erwarten.
 - (B) Nur Ausfall II ist zu erwarten.
 - (C) Nur Ausfall III ist zu erwarten.
 - (D) Nur die Ausfälle I und III sind zu erwarten.
 - (E) Nur die Ausfälle II und III sind zu erwarten.

Bei diesem Aufgabentyp folgen nach der Schilderung des Sachverhalts in der Regel drei oder fünf Aussagen in Form von Behauptungen. Die Testperson muss sich dabei entscheiden, ob sich die Aussagen aus den im Aufgabentext enthaltenen Informationen ableiten lassen. Dazu sind keine speziellen Sachkenntnisse erforderlich. Die korrekte Beurteilung der einzelnen Aussagen setzt das Verstehen des Sachverhalts voraus sowie die Fähigkeit, Schlussfolgerungen aus den im Text enthaltenen Informationen zu ziehen. Konkret lässt sich die Aufgabe, unter Berücksichtigung des unterstrichenen Textes, folgendermassen lösen:

- I. Da bei der Mehrzahl der Linkshänder die Sprachregion in der linken Hemisphäre liegt, müssen sie also mit einer Sprachstörung rechnen, weshalb Aussage I falsch ist.
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind immer noch die Möglichkeit, die Muttersprache wieder zu erlernen. Die Sprachregion wird dann in der rechten Hälfte der Hemisphäre angelegt. Somit ist Aussage II ebenfalls falsch.
- III. Da spätestens im zwölften Lebensjahr die Sprachregion bei Rechtshändern in der Regel fest in der linken Hälfte des Gehirns liegt, ist bei zwanzigjährigen Rechtshändern zu erwarten, dass sie die Fähigkeit, eine Sprache zu erlernen, verloren haben. Die Aussage III ist darum richtig.

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

3. Untertest: Schlauchfiguren

Die folgenden Aufgaben prüfen das räumliche Vorstellungsvermögen – eine Funktion, die beispielsweise für das Verständnis von Röntgenbildern wichtig ist. Während des Studiums werden zahlreiche eigentlich dreidimensional zu betrachtende Strukturen und Vorgänge in zweidimensionalen Abbildungen vermittelt.

Jede Aufgabe besteht aus zwei Abbildungen eines durchsichtigen Würfels, in dem sich ein, zwei oder drei Kabel befinden. Die erste Abbildung (links) zeigt stets die Vorderansicht des

Würfels; auf dem rechten Bild daneben, wo derselbe Würfel noch einmal abgebildet ist, soll die Testteilnehmerin oder der Testteilnehmer herausfinden, ob die Abbildung von rechts (r), links (l), unten (u), oben (o) oder von hinten ist.

Hier sehen Sie den Würfel von vorne!

Hier sehen Sie den Würfel von ? (oben!)

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.

Wieviele 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 q b b b q q

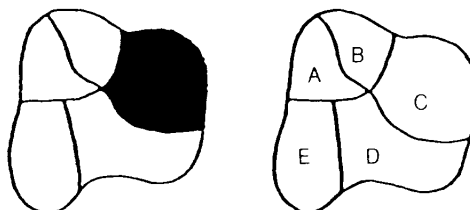
6. Untertest: Figuren lernen

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 4 Minuten Zeit, um sich 20 solcher Figuren einschliesslich der Lage der schwarzen Flächen einzuprägen. Nach ca. 1 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 eingeprägt und behalten werden, die ebenfalls nach der gleichen Zwischenzeit abgefragt werden. Dabei werden 15 Patienten vorgestellt, von denen jeweils der Name, die Altersgruppe, Beruf und Geschlecht, ein weiteres Beschreibungsmerkmal (z.B. Familienstand) sowie die Diagnose erfahren wird. Ein Beispiel für eine derartige Fallbeschreibung ist:

Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

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

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

Diese Freisetzung von T_3 und T_4 ins Blut (Sekretion) wird über die Hirnanhangsdrüse und den Hypothalamus, einen Teil des Zwischenhirns, gesteuert: Das erwähnte Hormon TSH stimuliert ausser der Bildung auch die Sekretion von T_3 und T_4 ; es ist hinsichtlich seiner eigenen Sekretionsrate jedoch abhängig von der Stimulation durch das hypothalamische Hormon TRH. Die TRH-Sekretion wiederum wird z.B. durch Kälte stimuliert, während Wärme hemmend wirken kann. Neben diesen übergeordneten Steuerungsmechanismen existiert noch ein sogenannter Rückkopplungsmechanismus: Eine hohe Konzentration von T_3 und T_4 im Blut hemmt die TSH- und die TRH-Sekretion, eine niedrige Konzentration stimuliert sie. Bei den an der Steuerung der Schilddrüsenhormon-Sekretion beteiligten Arealen von Hirnanhangsdrüse und Hypothalamus können krankheitsbedingte Störungen auftreten, die zu einer Über- oder Unterfunktion der Schilddrüse führen.

Eine der Hauptwirkungen von T_3 und T_4 ist die Beeinflussung des Energieumsatzes durch eine Steigerung des Sauerstoffverbrauchs in stoffwechselaktiven Organen. Entsprechend senkt eine zu niedrige Konzentration der beiden Hormone im Blut (Hypothyreose) den Energieumsatz bzw. die Stoffwechselaktivität unter den normalen Wert, während bei einer zu hohen Konzentration (Hyperthyreose) die Stoffwechselaktivität gesteigert wird. Die Hormone T_3 und T_4 können ebenso wie TSH und TRH für diagnostische und therapeutische Zwecke synthetisch hergestellt werden.

Auf einen solchen Text folgen Fragen, die sich ausschliesslich auf im Text vorhandene Inhalte beziehen; eine Frage mit niedrigem Schwierigkeitsgrad ist zum Beispiel so formuliert:

Welcher der folgenden Vorgänge gehört nicht zu den im Text beschriebenen Schritten, die zur Bildung von T_3 führen?

- (A) Transport von Jod aus den Epithelzellen in die Follikel
- (B) Umwandlung von Jod in Jodid in den Follikeln
- (C) Transport von Jodid aus dem Blut in die Epithelzellen
- (D) Verknüpfung von MIT- und DIT-Resten in den Follikeln
- (E) Verknüpfung von Jod und Tyrosin-Resten in den Follikeln

Für die Beantwortung dieser Frage ist das Verständnis der im obigen Text unterstrichenen Stellen wichtig (im Original sind selbstverständlich keine Hervorhebungen). Der Text sagt nichts über eine Umwandlung von Jod in Jodid in den Follikeln aus, und auch der umgekehrte Prozess, die Umwandlung von Jodid in Jod, findet nicht in den Follikeln statt, sondern in den Epithelzellen. Somit gehört der Vorgang (B) nicht zu den vom Text beschriebenen Schritten, die zur Bildung von T_3 führen.

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.

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“). Beim Test „Konzentriertes und sorgfältiges Arbeiten“ müssen insgesamt 1200 Zeichen der Reihe nach bearbeitet werden. Es können in der zur Verfügung stehenden Zeit 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 wird, um mit den anderen Tests gleichgewichtig zum Punktwert addiert zu werden. 600 Richtige wären das Maximum und entsprechen 20 Punkten.

Alle Punkte der Untertests werden zu einer Summe addiert (Punktwert, vgl. Abbildung 9). 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 als die entsprechende Person.

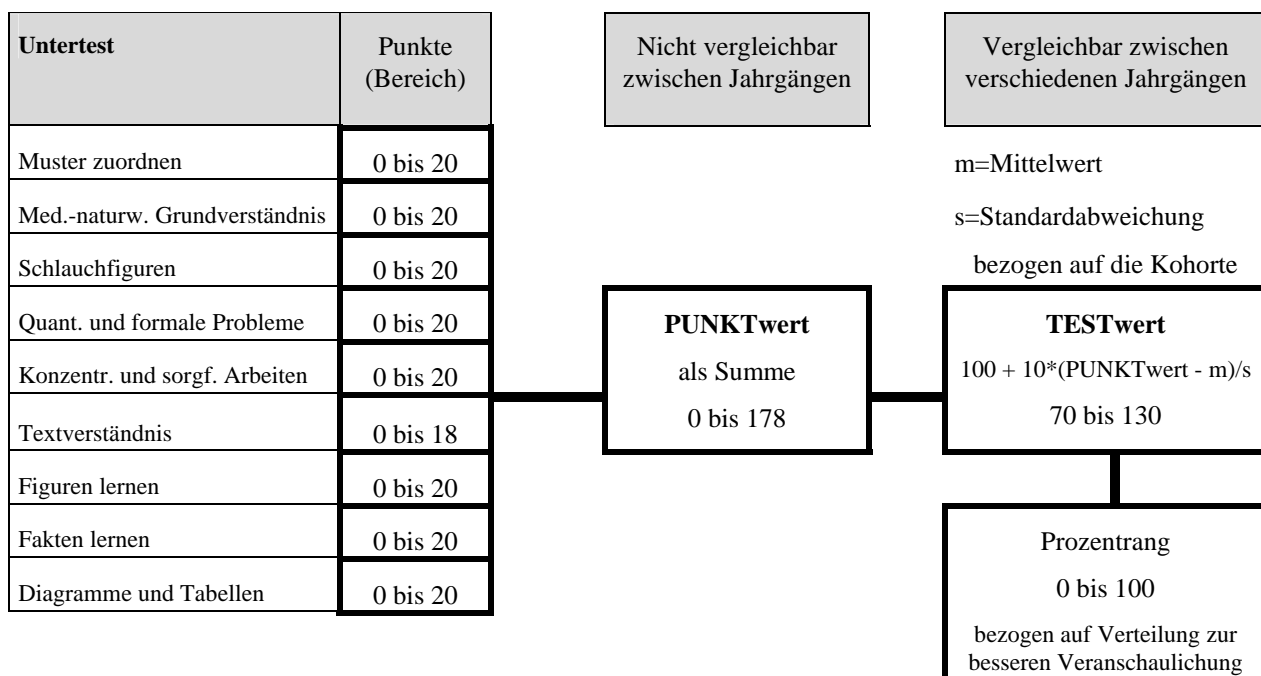


Abbildung 9: Punktwerte der einzelnen Untertests und ihre Zusammenführung über den Punktwert zum Testwert

Vorhersage des Studienerfolges in Deutschland

Der Test wurde in Deutschland über 10 Jahre angewendet. In dieser Zeit fanden mehrere Evaluationen statt, um seine Gütekriterien zu überprüfen (vgl. Bericht 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. 20% 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.

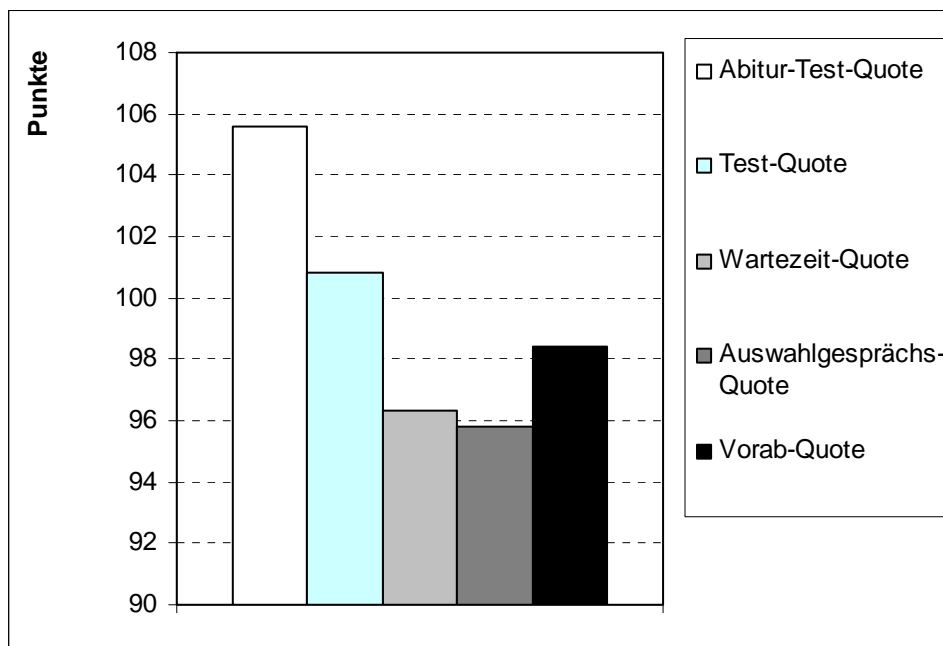


Abbildung 10: Durchschnittliche Punkte für schriftliche Prüfungsleistungen während des Studiums nach der Zulassung; aus Trost (1994, S. 144)

In der schriftlichen Prüfung (Abbildung 10) 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 9) verdeutlichen, dass vor allem die schriftliche Prüfungsnote durch den Test sehr gut vorhergesagt werden kann, der Test seiner Bezeichnung als „Probestudium“ 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 (Trost u.a. 1998).

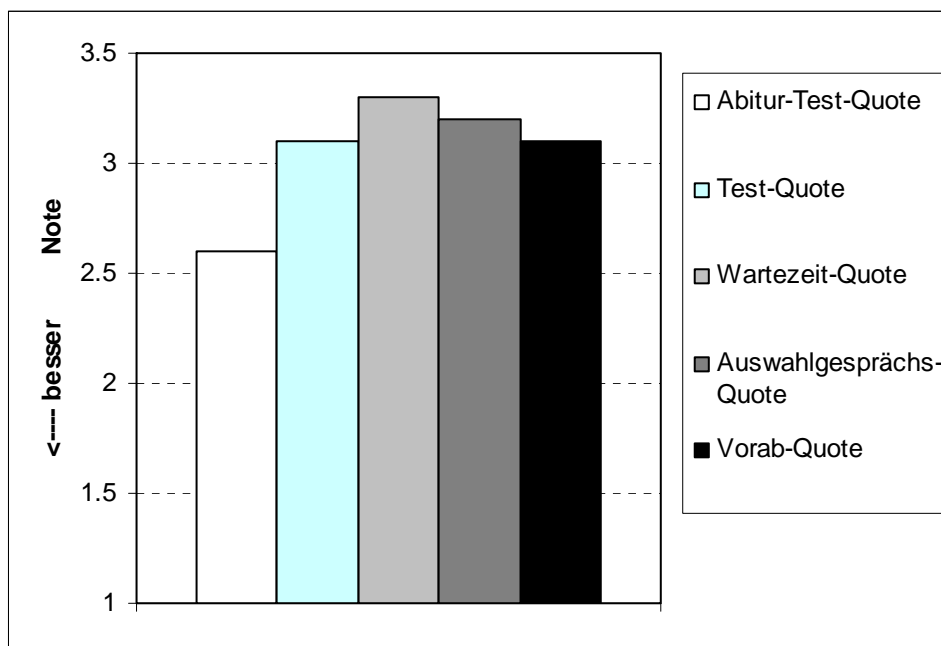


Abbildung 11: Durchschnittliche Note für mündliche Prüfungsleistungen während des Studiums nach der Zulassung; aus Trost (1994, S. 144)

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 9: Korrelationen zwischen den Prädiktorvariablen (TMS-Ergebnisse) und den Ergebnissen der ärztlichen Vorprüfungen; aus Trost (1994)

Testanwendung in der Schweiz 1998

Organisation der Testabnahme

Die Testabnahme fand einheitlich am 3.7.1998 statt. Pro Testlokal war eine Testleiterin/ein Testleiter für den Ablauf verantwortlich. Eine Stellvertreterin/ein Stellvertreter und pro 50 Teilnehmer eine Assistentin/ein Assistent standen diesen für die Beaufsichtigung zur Seite. TestleiterInnen und StellvertreterInnen wurden vorher auf einer eintägigen Veranstaltung geschult und insbesondere auf das genaue Einhalten des Ablaufes und einheitliche Reagieren auf Besonderheiten vorbereitet. 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 eingehalten. Die Identität der Teilnehmerinnen und Teilnehmer wurde anhand eines gültigen Personaldokumentes überprüft. Eine telefonische Hotline war für die Testleiter zum ZTD und zur Hochschulkonferenz eingerichtet.

Aus den Protokollen ergeben sich keine Hinweise, welche die Gültigkeit der Testabnahme einschränken. Drei Verwarnungen betrafen leichte Verstösse gegen Regeln zum Testablauf.

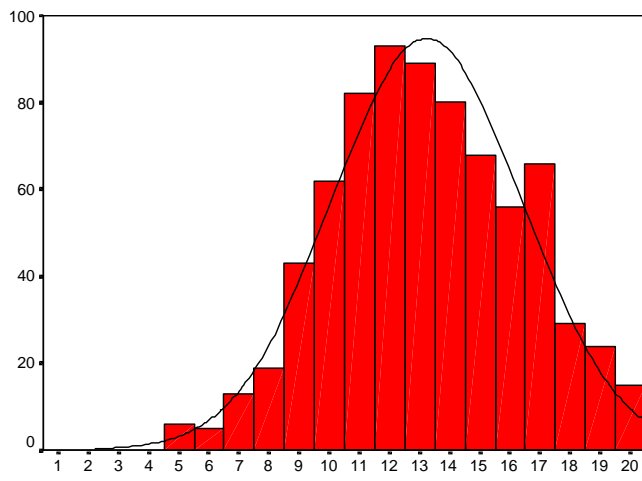
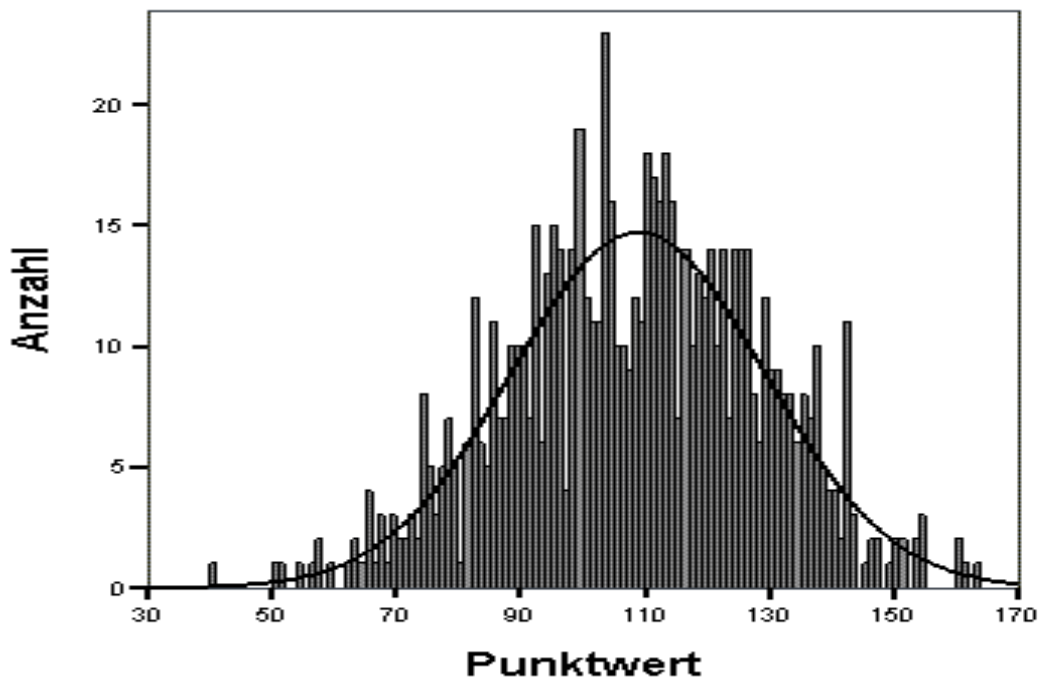
Die formalen Bedingungen an allen Testorten waren vergleichbar, und es traten keine Benachteiligungen auf.

Verteilungsprüfung

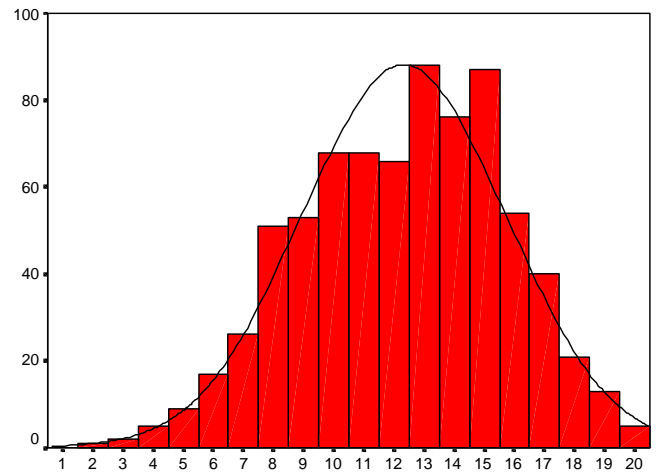
Die Verteilung der Testwerte ist nicht nur entscheidend für die Anwendbarkeit statistischer Prüfverfahren. Die Personen müssen sich ausreichend voneinander unterscheiden lassen, besonders im Bereich des kritischen Zulassungswertes. Der Test darf nicht zu leicht oder zu schwer sein. Idealerweise ist der Test so ausgelegt, dass mittlere Punktzahlen genau durch 50% der Personen erreicht werden und die Punkte um diesen Wert streuen.

Die Normalverteilung kann für den Punktwert angenommen werden (Anhang Tabelle A 5). Fünf Untertests entsprechen einer Poisson-Verteilung (Beispiel: "Konzentriertes und sorgfältiges Arbeiten"). Die Prüfungen für den Testwert können parametrisch erfolgen, bei den Untertests ist die Stichprobengrösse ein Argument dafür, ebenfalls eine parametrische Prüfung vorzunehmen. Alle zentralen Hypothesen wurden auch parameterfrei geprüft, wobei sich keine Unterschiede zur parametrischen Prüfung ergaben.

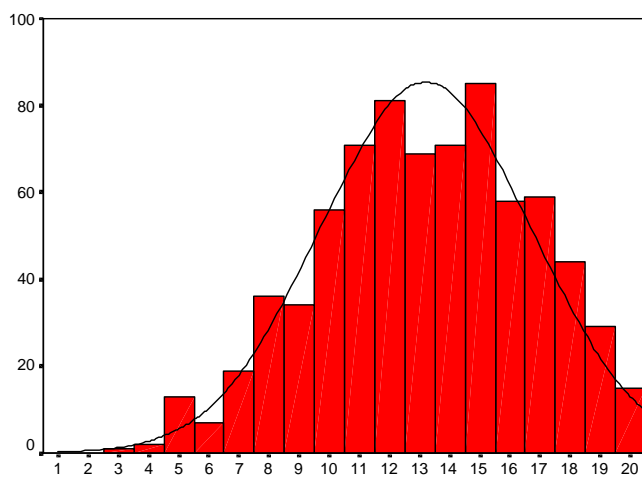
Für die Punkteverteilungen, die einer Poisson-Verteilung folgen, sei angemerkt: Die Seite der Verteilung zum höheren Punktwert hin weist grössere Häufigkeiten auf. Dies kann darauf hindeuten, dass das Niveau der individuellen Vorbereitung anhand der Test-Info stärker variiert. Beim "Konzentrierten und sorgfältigen Arbeiten" wird dies am deutlichsten. Eine Häufung tritt bei den höchsten Punktzahlen auf. Aus Deutschland ist bekannt, dass dieser Test in seiner zeitlichen Dauer über die Jahre verkürzt werden musste, da sich immer häufiger die Punktwerte 20 oder 19 ergaben und der Test dann nicht mehr optimal differenzierte. Dies könnte u.a. auf eine Subgruppe der sehr gut Trainierten hindeuten, die eine hohe Punktzahl erreichten. Eine Verkürzung der Testzeit scheint angebracht.



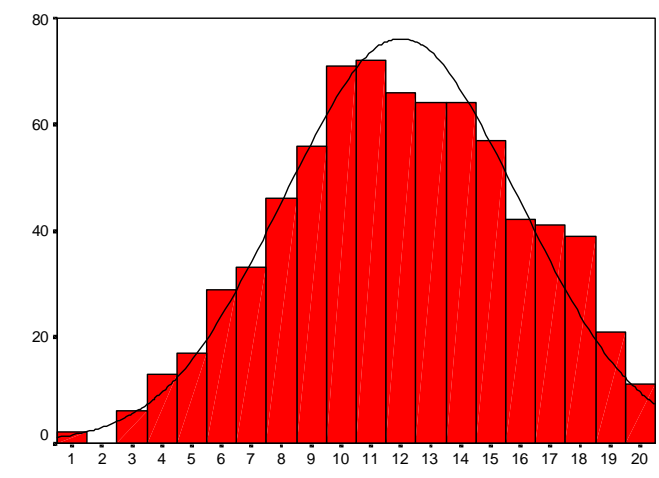
Muster zuordnen



Med.-naturwiss. Grundverständnis

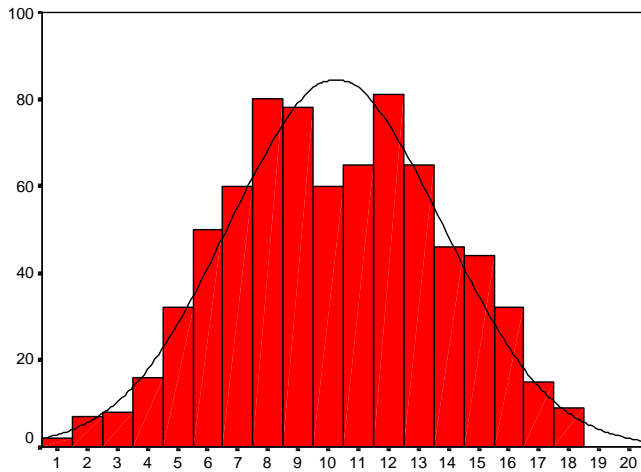


Schlauchfiguren

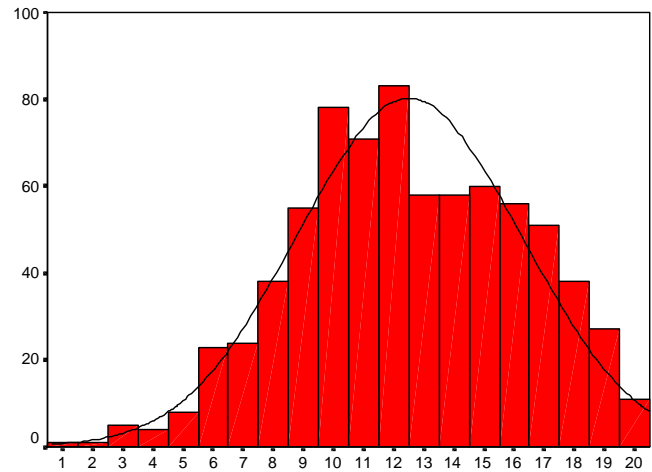


Quantitative und Formale Probleme

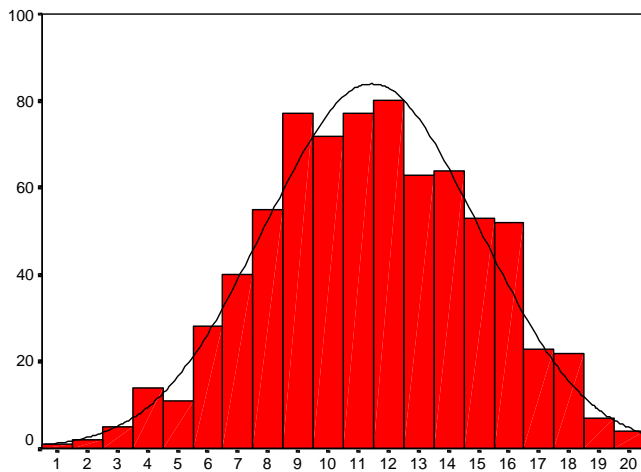
Abbildung 12: Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 1-4



Textverständnis



Figuren lernen



Fakten lernen

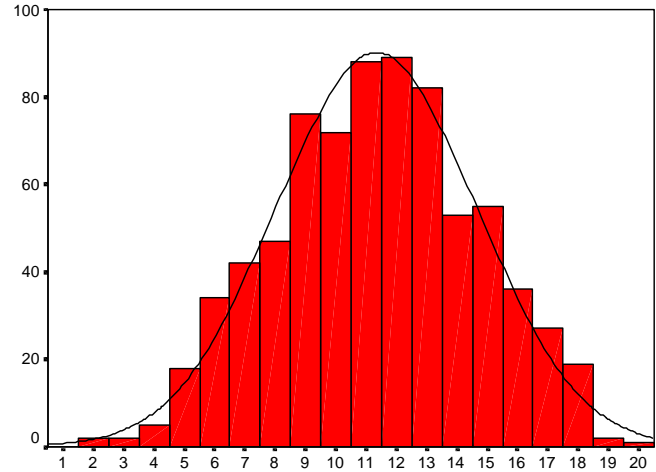
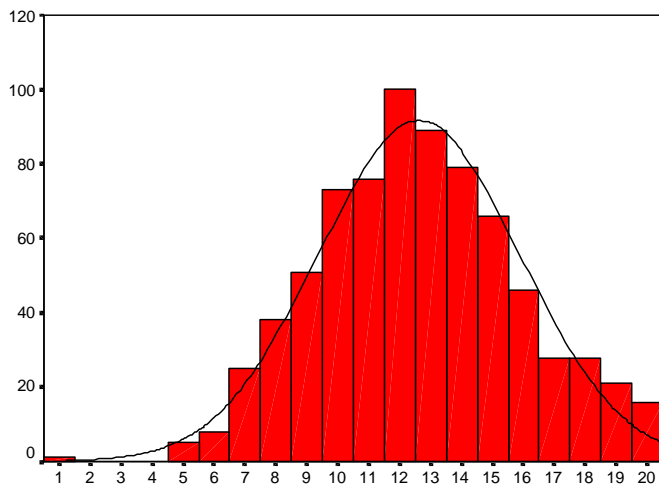


Diagramme und Tabellen



Konzentriertes und sorgfältiges Arbeiten

Abbildung 13: Häufigkeitsverteilungen für Punktwert und die Punktzahlen der Untertests 5-9 (Wert 0 wird nicht dargestellt, dieser kommt insgesamt nur einmal vor – siehe auch Anhang)

Beim Test „Konzentriertes und sorgfältiges Arbeiten“ tritt einmal der Punktwert 0 auf. Diese Person hat entgegen der Instruktion relativ „wahllos“ die zutreffenden Zeichen durchgestrichen – dies über den gesamten Bogen verteilt. Dadurch entsteht eine hohe Zahl bearbeiteter Zeichen – von der aber so viele übersehene Zeichen abgezogen werden müssen, dass praktisch nichts mehr übrig bleibt. Das Verstehen der Instruktion gehört aber zum Test mit dazu.

Äquivalenz der Testformen 1 und 2

Mit der Verwendung von „pseudoparallelen“ Testformen soll vor allem ein „Abschreiben“ der Ergebnisse beim Nachbarn verhindert werden. Die Testleiter wurden so orientiert, dass auf benachbarten Plätzen unterschiedliche Formen bearbeitet werden und auch in der Reihe davor die Formen jeweils so verteilt werden, dass kein Abschreiben möglich ist. In der Praxis wurden die Räume so ausgewählt, dass mehr als ein Platz und auch mehr als eine Sitzreihe zwischen den Personen als Abstand bestand.

Beide verwendeten Testformen unterscheiden sich bezüglich der Einstreuaufgaben, von denen verschiedene in Form 1 und 2 verwendet werden. Ansonsten werden die gleichen Aufgaben für die Berechnung des Punktwertes in beiden Formen verwendet. Einige Aufgaben wurden paarweise vertauscht, was theoretisch einen Einfluss auf die Ergebnisse haben könnte.

In Deutschland wurden 8 Testformen verwendet, um insgesamt mehr Einstreuaufgaben erproben zu können. Sie unterscheiden sich bezüglich der gewerteten Aufgaben gleichfalls nicht.

Die nachfolgende Tabelle 10 enthält Mittelwerte und Standardabweichungen für beide Formen. Die statistische Prüfung (Anhang Tabelle A 6) zeigt, dass keine signifikanten Unterschiede ($p < 5\%$) zwischen den Formen auftreten. Dies gilt im übrigen auch bei separater Betrachtung der Sprachformen, daher werden nachfolgend beide Formen nicht weiter unterschieden.

FORM		Punktwert	Muster zuordnen	Med.-naturwiss. Grundverständnis	Schlauchfiguren	Quant.-formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagramme und Tabellen	Konzentr. u.sorgf. Arbeiten
1	m	108.77	13.31	12.41	13.06	12.17	10.13	12.40	11.22	11.46	12.54
	s	21.07	3.16	3.50	3.57	4.11	3.66	3.62	3.59	3.40	3.40
2	m	109.09	13.14	12.23	13.32	11.83	10.37	12.61	11.66	11.20	12.69
	s	19.59	3.13	3.26	3.41	3.71	3.39	3.81	3.50	3.21	3.10

Tabelle 10: Mittelwerte und Standardabweichungen für die Formen 1 und 2, alle Unterschiede statistisch nicht signifikant (siehe Anhang Tabelle A6).

Äquivalenz der Sprachversionen

Der Test wurde in drei Sprachversionen durchgeführt (deutsch-, französisch-, sowie italienischsprachig). Die in den gesetzlichen Grundlagen geforderte Gleichbehandlung verlangt nicht nur, dass die Durchführungs- und Auswertebedingungen vergleichbar sind:

Auch die erreichten Testwerte müssen für die drei Sprachgruppen so aufeinander bezogen werden, dass gleiche Testwerte gleichen Schwierigkeitsgraden entsprechen.

Nicht immer können die gezählten **Punktwerte** als Zahlenwerte direkt verglichen werden. Dabei muss berücksichtigt werden, dass die gleichen Aufgaben in den einzelnen Sprachgruppen sich **testbedingt** im Schwierigkeitsgrad mehr oder weniger deutlich unterscheiden können. Üblicherweise wird die Lösungswahrscheinlichkeit (angegeben in Prozent) als Mass für die Schwierigkeit verwendet.

Die Nachnutzung eines Tests in einer anderen Sprache ist international weit verbreitet. Deshalb existieren für Übertragungen von Tests in andere Sprachen entsprechende **Kriterien**, vor allem die "*Technical Standards for Translating and Adapting Tests and Establishing Test Score Equivalence*", die von sechs führenden internationalen psychologischen Fachgesellschaften 1994 herausgegeben worden sind. Sie bildeten auch die Grundlage dieser Testanwendung.

Ursachen für unterschiedliche Testleistungen

Die wichtigste Besonderheit der Testübertragung ist, dass man von **Adaptation** spricht und nicht nur von Übersetzung. Das Ziel einer Adaptation ist es, den Test mit gleicher Zuverlässigkeit und Gültigkeit in den verschiedenen Sprachgruppen einzusetzen. Man muss dabei berücksichtigen, dass es in den neun Untertests Aufgaben gibt, die ganz unterschiedlich sprachabhängig sind. Bei den Untertests „Figuren lernen“, „Schlauchfiguren“ und „Konzentriertes und sorgfältiges Arbeiten“ sollten sich keine sprachspezifischen Unterschiede zeigen, wenn man die richtige Übersetzung und das gleiche Verständnis der Instruktionen einmal voraussetzt. Sie sind praktisch weitgehend „sprachfrei“ in den Anforderungen selbst.

In den sprachabhängigen Tests (z.B. "Textverständnis") werden sehr komplexe Informationen in teilweise stark verschachtelten Sätzen dargeboten (siehe Beispiel Seite 57). Es ist ein wesentliches Merkmal des deutschen TMS, dass Fähigkeiten der Sprachbeherrschung bzw. Sprachkompetenz (v.a. logisch-analytische Fähigkeiten) stark gefordert werden – spezifisches Fachwissen aus der Medizin dagegen nicht vorausgesetzt wird. Dabei handelt es sich im übrigen um einen der wichtigsten Vorzüge des Testkonzeptes gegenüber den Wissenstests zur Prüfung der Studieneignung. Wissenstests könnten wesentlich intensiver trainiert werden.

Die Verwendung von **Sprachkompetenz** als Eignungskriterium ist insofern gerechtfertigt, weil deren Komponenten als bedeutsame Faktoren in den meisten Intelligenz- und Fähigkeitskonzepten auch empirisch nachgewiesen werden konnten. Sie entwickelt sich zweifellos in enger Wechselbeziehung mit dem angeeigneten Wissen, hat aber zahlreiche „themenübergreifende“ und eher formale Funktionen. Diese ermöglichen es beispielsweise auch, sich eine bisher unbekannte Thematik zu erarbeiten. Sehr vereinfacht gesagt entscheidet bei der Neuerarbeitung eines Themas dann nicht nur das bereits vorhandene

Faktenwissen. Die Fähigkeit, dieses Wissen effektiv den vorhandenen Quellen entnehmen zu können, sollte dem grundsätzlich gleichwertig sein. In Anforderungsanalysen zur Studierfähigkeit findet sich die Sprachkompetenz deshalb nicht zu Unrecht an bedeutender Stelle. Bei den Aufgaben des Tests wurde Wert darauf gelegt, dass das gesamte zur Lösung notwendige Fachwissen in den Aufgaben enthalten ist. Die Aufgabe der KandidatInnen ist es, die relevante Information in den Texten zu erkennen und mit deren Hilfe bestimmte Fragen zu beantworten.

Die **Lösungsschwierigkeit** dieser Aufgaben wird auch von der Satzstellung, der Kompliziertheit der Verschachtelung von Sätzen und der Wortwahl mitbestimmt. Bereits vermeintlich geringfügige Änderungen im deutschen Text können zu veränderter Aufgabenschwierigkeit führen. Dieser Effekt wird bei der Aufgabenentwicklung im übrigen für Revisionen genutzt. Die tatsächliche Schwierigkeit kann dabei aber kaum vorhergesagt werden – es ist immer eine empirische Überprüfung notwendig.

Die drei Sprachen unterscheiden sich vor allem bezüglich der grammatikalischen Formen zur Satzkonstruktion enorm. Bei der Übersetzung der Aufgaben muss vor allem darauf geachtet werden, dass alle Informationen richtig in die Sprache der Übersetzung transportiert werden. Lehnt man sich zu sehr an die deutsche Satzkonstruktion an, wirkt der Text in der Zielsprache künstlich und meistens sogar unverständlich. Passt man ihn der Grammatik der Zielsprache an, verändern sich automatisch die Gewichtungen der einzelnen Informationen zueinander. Hier musste deshalb im Übersetzungsverfahren immer ein guter Kompromiss gesucht werden.

Man vergleiche dazu die Diskussion und die Beispiele des israelischen Zulassungstests SAT, die Beller (1996) im Bericht 2 des ZTD diskutiert. Es handelt sich dort um grammatisch vergleichsweise einfach strukturierte Items, die in einer russisch- und einer hebräischsprachigen Version dieses Tests angewendet werden. Trotzdem auch dort die gleichen Übersetzungsstandards eingehalten werden, kommt es zu Unterschieden der Schwierigkeit auf Itemebene.

Das gewählte Übersetzungsverfahren beinhaltet:

- Schulung der Übersetzer bezüglich der genannten Probleme. Die italienischsprachige Version wurde von zwei Übersetzern angefertigt. Für die französischsprachige Version stand im ZTD ein Ansprechpartner zur Verfügung, so dass immer Konsultationsmöglichkeiten bestanden. Es war auch bekannt, dass Übersetzer die Texte anders ausrichten, wenn ihnen die Tatsache der Rückübersetzung zur Qualitätskontrolle bekannt ist, deshalb wurden sie darüber nicht informiert.
- Hinzuziehen von Fachexperten und psychologischen Experten für Beratung und Evaluation der Übersetzung (2 pro Sprache) – hier wurde der Text in der Zielsprache optimiert.
- Rückübersetzung ins Deutsche (durch ein zweites Übersetzungsbüro) und Vergleich mit dem Original sowie abschliessende "Konferenz" zum Finden der endgültigen Lösung.

Mit diesen Methoden konnte man recht gut prüfen, ob alle Informationen richtig in die jeweilige Sprachform übertragen worden waren. Ob **die Getesteten** dann diese relevanten Informationen in gleicher Weise und v.a. unter Zeitdruck auch erkennen und verarbeiten können – die Aufgaben tatsächlich gleich schwer bleiben – ist nur empirisch zu beantworten. Laut Standard D.6. der erwähnten Richtlinien ist von einem möglichen „Sprach-Bias“ der Items (als einem für die Frageninhalte spezifischen Unterschied zwischen den Sprachformen) auszugehen.

Die Übertragung in eine andere Sprache kann *eine* Ursache für Schwierigkeitsunterschiede der Aufgaben sein – aufgrund der vorhandenen Erfahrungen sogar eine sehr wahrscheinliche.

Aber nicht alle tatsächlich gemessenen Leistungsunterschiede zwischen den Sprachgruppen können auf diese Sprachbesonderheiten des Tests zurückgeführt werden. Mindestens zwei **weitere Ursachen** für Unterschiede sind möglich:

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 TeilnehmerInnen macht dies wahrscheinlich – die Auswahl kann die "Spitze" oder das "Ende" der Leistungsrangreihe aller MaturandInnen der jeweiligen Sprachgruppe überrepräsentieren, etwa durch unterschiedliches "Wahlverhalten" aufgrund der vorhandenen Alternativen für Studienorte.

Sprachvergleich für die Untertests

In der Tabelle 11 sind Mittelwerte und Standardabweichungen der Untertestwerte und des Punktwertes für die drei Sprachgruppen dargestellt – die Profile mit den Vergleichswerten aus Deutschland werden in der nachfolgenden Abbildung 14 veranschaulicht.

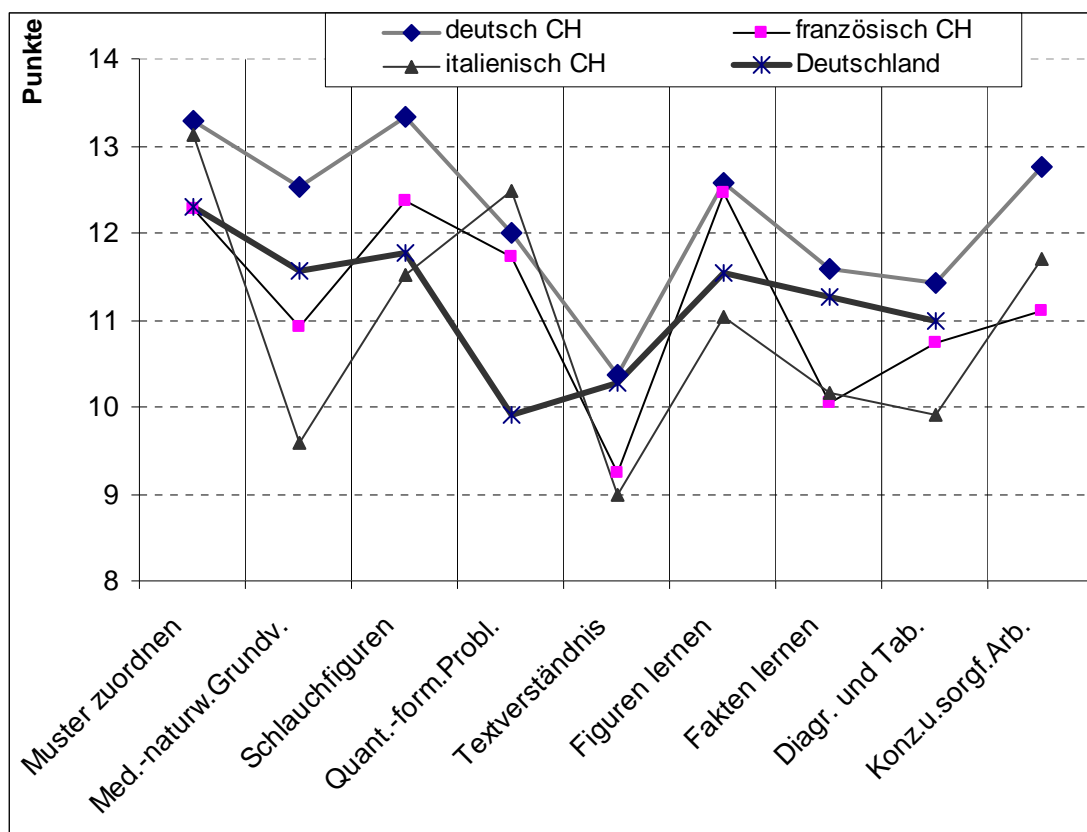


Abbildung 14: Mittelwerte der Untertests für die Sprachgruppen und Vergleich zu Deutschland

Bei nur drei Untertests ("Muster zuordnen", "Quantitative und formale Probleme" sowie "Figuren lernen") sind keine Unterschiede statistisch signifikant. Bei allen anderen Untertests und beim Punktwert bestehen signifikante Unterschiede zwischen den Sprachgruppen. Interessant ist im übrigen auch, dass die deutschsprachige Gruppe der Schweiz bessere Ergebnisse erreicht als die Vergleichsgruppe aus Deutschland.

Bei den multiplen Mittelwertvergleichen (siehe Anhang) zeigt sich in sechs geprüften Merkmalen ein besserer Wert bei der deutschsprachigen Gruppe als bei einer oder beiden anderen Sprachgruppen.

		N	Mittelw.	Standardabw.	Minimum	Maximum
Punktwert ** d > (f, i)	deutsch	680	109.84	20.31	41.00	164.00
	französisch	39	100.87	19.55	52.00	141.00
	italienisch	31	98.51	17.08	64.00	132.00
Muster zuordnen	deutsch	680	13.28	3.15	5.00	20.00
	französisch	39	12.28	3.17	5.00	18.00
	italienisch	31	13.12	2.86	9.00	19.00
Med.-naturwiss. Grundverständnis** d > (f, i)	deutsch	680	12.53	3.35	2.00	20.00
	französisch	39	10.92	2.76	6.00	17.00
	italienisch	31	9.58	3.23	3.00	16.00
Schlauchfiguren** d > i	deutsch	680	13.32	3.49	3.00	20.00
	französisch	39	12.35	3.24	6.00	18.00
	italienisch	31	11.51	3.38	6.00	19.00
Quant. und formale Probleme	deutsch	680	11.99	3.94	1.00	20.00
	französisch	39	11.71	3.56	6.00	19.00
	italienisch	31	12.48	3.99	3.00	20.00
Textverständnis**	deutsch	680	10.37	3.52	1.00	18.00
	französisch	39	9.23	3.30	2.00	16.00
	italienisch	31	9.00	3.55	3.00	18.00
Figuren lernen	deutsch	680	12.57	3.72	2.00	20.00
	französisch	39	12.46	3.91	1.00	19.00
	italienisch	31	11.03	3.33	5.00	18.00
Fakten lernen** d > f	deutsch	680	11.57	3.52	1.00	20.00
	französisch	39	10.05	3.83	3.00	16.00
	italienisch	31	10.16	3.29	4.00	17.00
Diagramme und Tabellen** d > i	deutsch	680	11.43	3.33	2.00	20.00
	französisch	39	10.74	2.88	5.00	16.00
	italienisch	31	9.90	2.93	5.00	17.00
Konzentr.u.sorgf. Arbeiten** d > f	deutsch	680	12.74	3.26	0.00	20.00
	französisch	39	11.10	3.27	5.00	19.00
	italienisch	31	11.70	2.35	7.00	16.00

Tabelle 11: 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.

Welche der oben genannten möglichen Ursachen sind nun für diese Unterschiede verantwortlich? In der nachfolgenden Tabelle 12 werden die Untertests nach der Sprachabhängigkeit geordnet. Die Einordnung beruht auf übereinstimmenden Urteilen von

5 Mitarbeitern des ZTD, die mit der Materie vertraut waren. Bei „Textverständnis“ sind die recht komplizierten Sätze bestimmend, bei „Fakten lernen“ mussten die Namen der Personen in übliche Namen der jeweiligen Sprachregion übertragen werden, wobei immer drei eine Gemeinsamkeit haben mussten. Bei „Diagramme und Tabellen“ wie auch bei „Quantitative und formale Probleme“ scheint die Sprachabhängigkeit deutlich geringer – die Abbildungen und Tabellen wie auch die mathematischen Formeln sind in den jeweiligen Sprachen identisch.

Unterschiede sind in sprachabhängigen sowie weniger sprachabhängigen Tests nachweisbar. Bis auf „Quantitative und formale Probleme“ erreichen die Deutschsprachigen jeweils höhere Werte als die beiden anderen Gruppen. Da beim letztgenannten Test Männer bessere Leistungen erzielen und in der italienischsprachigen Stichprobe der Männeranteil grösser ist als in den beiden anderen, würde man dies auf die Stichprobenszusammensetzung zurückführen können.

Rating nach Sprachabhängigkeit	Testteil	Mittelwerte signifikante Unterschiede	Varianzanalyse „Sprache“ F und Signifikanz (p) Markiert $p < .05$
1	Textverständnis		3,99 .019**
2	Fakten lernen	d > f	5,57 .004**
3	Medizinisch-naturwiss. Grundverständnis	d > (f, i)	15,34 .000**
4	Diagramme und Tabellen	d > i	3,86 .021**
5	Quantitative und formale Probleme		0,33 .714
gering	Konzentriertes und sorgfältiges Arbeiten	d > f	6,05 .002**
gering	Muster zuordnen		1,89 .151
gering	Schlauchfiguren	d > i	5,19 .006**
gering	Figuren lernen		2,57 .077

Tabelle 12: Rating nach Sprachabhängigkeit der Untertests durch Mitarbeiter des Zentrums für Testentwicklung ZTD.

Man kann die signifikanten Unterschiede zwischen den Sprachversionen aber nicht allein auf die Wirkung des Sprachfaktors reduzieren. Unterschiede treten in allen Bereichen auf. Zumindest bleibt aber festzuhalten, dass in vier von fünf sprachabhängigen Tests der Einfluss deutlich vorhanden ist – bei den anderen sind zwei mit Unterschieden und zwei mit keinen Unterschieden vorhanden. Auch bei „Schlauchfiguren“ und dem „Konzentrierten und sorgfältigen Arbeiten“ sind aber Unterschiede nachweisbar.

Dabei ist die Reihenfolge der Sprachgruppen bezüglich der Testleistungen nicht immer die gleiche, was wiederum für unterschiedliche Fähigkeitsprofile in den Sprachgruppen sprechen würde.

Beim **Probelauf** des Tests in Fribourg 1995 an **vergleichbaren Maturand-Innenstichproben** einer Schule in den Sprachen Französisch und Deutsch hatten sich bekanntlich **keine** derartigen Unterschiede zwischen den Sprachgruppen gezeigt (vgl. Bericht 2 des ZTD).

Eine sehr zentrale Ursache der Unterschiede sehen wir in der mangelnden Vergleichbarkeit der Gruppen: Die drei Gruppen sind bezüglich der Zusammensetzung ganz **unterschiedlich repräsentativ** für die MaturandInnen der jeweiligen Sprachgruppe. Bekanntlich ist die MaturandInnenquote im Tessin und in den Kantonen der Westschweiz grösser als in der Deutschschweiz. Die geringen Grössen der französisch- und italienischsprachigen Gruppe stehen dazu aber in keinem definierbaren Verhältnis (Auswahlmöglichkeit der anderen Studienorte Lausanne und Genf, mögliches Studium in Italien). Dies verunmöglicht es auch, den Ursachen weiter empirisch nachzugehen.

Äquivalenz der Tests in den drei Sprachformen bestimmt sich nicht über gleiche Mittelwerte für die Punktwerte, sondern über vergleichbare Schwierigkeiten – so die Forderung. Erforderlich sind allerdings auch

- eine vergleichbare Trennschärfe des Items (Unterscheidung von "Besseren" und "Schlechteren" der Sprachgruppe) im Sinne des Untertestwertes, zu dem es gehört;
- einen gleichen Rangplatz des jeweiligen Items in Rangreihe aller Items eines Untertests bezogen auf die Schwierigkeit der Lösung.

Dem soll auf Itemebene weiter nachgegangen werden:

Analyse der Sprachversionen auf Itemebene

Ein Vergleich zwischen den Sprachversionen auf Itemebene zeigt zunächst, dass innerhalb der Untertests die Itemschwierigkeiten hoch miteinander korrelieren. Aus Tabelle 13 wird ersichtlich, dass bis auf einen Wert alle über 0,84 - die Mehrheit sogar über 0,90 - liegen. Die Itemschwierigkeiten der italienischen Übersetzung des Untertests "Medizinisch-naturwissenschaftliches Grundverständnis" korrelieren mit denjenigen der deutschsprachigen Version mit .74.

Insbesondere die zwar signifikanten aber doch recht geringen Differenzen im sehr sprachabhängigen Untertest "Textverständnis" geben Grund zur Hoffnung, dass auch eine Übersetzung von textintensiven Testteilen grundsätzlich möglich ist. Eine genauere Betrachtung zeigt aber, dass in fast allen Untertests sowohl systematische wie auch itemspezifische Differenzen auftreten. Aufgrund der kleinen Stichproben für die französische und italienische Version ist es praktisch unmöglich, die Ursache der Differenzen zu identifizieren. Überraschend ist, dass neben den eher sprachabhängigen Untertests "Medizinisches-naturwissenschaftliches Grundverständnis" und "Quantitative und formale Probleme" auch der weitgehend sprachunabhängige Untertest "Figuren lernen" am unteren Ende der beobachteten Korrelationen angesiedelt ist.

Untertest	n	deutsch (CH)- französisch	deutsch (CH)- italienisch	deutsch (CH)- Deutschland
Muster zuordnen	20	.97	.97	.99
Med.- nat. Grundverständnis	20	.94	.74	.96
Schlauchfiguren	20	.95	.93	.97
Quant. u. Form. Probleme	20	.92	.85	.94
Textverständnis	18	.91	.93	.97
Figuren lernen	20	.88	.92	.98
Fakten lernen	20	.84	.92	.96
Diagramme und Tabellen	20	.93	.92	.99

Tabelle 13: Untertests: Pearson-Korrelation zwischen den Itemschwierigkeiten der Sprachversionen, n: Anzahl der Items

Eine genauere Analyse ist durch direkte Itembetrachtung möglich. Es ist daran zu erinnern, dass die französisch- und die italienischsprachigen Versionen von nur 39 beziehungsweise 31 Testteilnehmern bearbeitet wurden. Alle nachfolgenden Analysen werden durch diese geringen Stichproben in ihrer Aussagefähigkeit beeinflusst.

Weist ein Item in unterschiedlichen Sprachgruppen verschiedene statistische Kennwerte auf, so spricht man von "differential item functioning" (DIF). Eine Möglichkeit, DIF's zu entdecken, besteht in der Methode "Delta-Plot" oder "transformed item difficulty" (TID) nach Angoff (1972). Dabei werden die Itemschwierigkeiten in einem Diagramm für beide Sprachgruppen gleichzeitig abgebildet. Jeder Punkt repräsentiert ein Item und als Wertepaar die Schwierigkeiten in beiden Gruppen.

Weisen beide Gruppen identische Fähigkeiten auf, so kommen die Punkte in der graphischen Veranschaulichung in einer Anordnung um die Diagonale zu liegen.

Dass es sich um **kein schweizspezifisches** Problem handelt, sollen die Abbildungen 15 und 16 zeigen (Cook 1998): In den USA wurden für einen Hochschulzulassungstest (SAT/PAA) Aufgaben aus dem Spanischen ins Englische und umgekehrt übersetzt. Dabei wurde ein dem hier Dargestellten vergleichbares Übersetzungsverfahren angewendet. Bei den 91 mathematischen Aufgaben stimmen die Schwierigkeiten der Aufgaben in beiden Sprachen recht gut überein. Bei den 142 sprachlichen Items (die auch Wissensfragen beinhalten), zeigt sich hingegen eine sehr deutliche Abweichung. Es wird mit einer relativ grossen Itemmenge gearbeitet – die Items für eine Endform werden dann nach der Übereinstimmung der Schwierigkeiten ausgewählt. Dieses Vorgehen erfordert Voruntersuchungen wie auch genügend Items zur Auswahl.

Auch im SAT in Israel (Beller 1996) werden im nichtverbalen Teil (Quantitative Reasoning) des Tests diejenigen Items nicht verrechnet, die deutliche Schwierigkeitsunterschiede zeigen. Der Test wird immerhin in 6 verschiedenen Sprachen angewendet.

Im verbalen Teil (Verbal Reasoning) treten grössere Differenzen auf („This is the most problematic area, because the meaning of verbal items may be drastically altered by translation“, Beller 1996, S. 20). Hier scheinen die Unterschiede so gross zu sein, dass ein Eliminieren der Items zu wenige für eine Verrechnung übriglassen würde. Deshalb werden in diesem Teil des Tests Verrechnungsverfahren eingesetzt, um die Items vergleichbar zu machen.

Es folgen Schwierigkeitsdiagramme (Abbildungen 17 – 24), wo für 8 Untertests die Gruppe der Deutschschweiz (X-Achse) jeweils mit beiden anderen Sprachgruppen und den Ergebnissen aus Deutschland verglichen wird. In diesen Diagrammen sind die Items in ein Koordinatensystem für jeweils zwei Itemschwierigkeiten eingetragen. Auf die übliche Standardisierung, vgl. z.B. Angoff (1993) (13 + 5 z) wurde vorerst verzichtet, um systematische wie itemspezifische Unterschiede aus einem Diagramm ablesen zu können: Die Diagonale entspricht identischen Itemschwierigkeiten. Der Wertebereich bis 110% musste gewählt werden, um im Grafikprogramm die oberen Itemnummern noch anzuzeigen – er ist natürlich nur zwischen 0% und 100% definiert.

Es ist möglich, dass ein systematischer Unterschied zwischen beiden Gruppen besteht, der alle Items gleichermassen betrifft. Dafür ist eher ein genereller Fähigkeitsunterschied zwischen den Gruppen anzunehmen und in den Abbildungen wird dies als Regressionsgerade der Schwierigkeitswerte-Verteilungen in beiden Sprachgruppen eingezeichnet. Eine Entfernung der Regressionsgeraden von der Diagonale zeigt solche systematischen Unterschiede an. Die Regressionsgerade muss nicht parallel zur Diagonalen sein, sondern der systematische Unterschied kann sich seinerseits mit steigender Schwierigkeit systematisch verändern.

Einzelne Punkte, welche weiter **von der Regressionsgeraden** entfernt sind, repräsentieren Items, welche einen „Item-Bias“ aufweisen – d.h. die Schwierigkeiten unterscheiden sich in beiden Sprachgruppen ganz spezifisch für das jeweilige Item. In diesem Fall muss in Bezug auf das betreffende Item von sprach- oder kulturspezifischen Ursachen (zum Beispiel Übersetzung, unterschiedlicher Vertrautheitsgrad) ausgegangen werden.

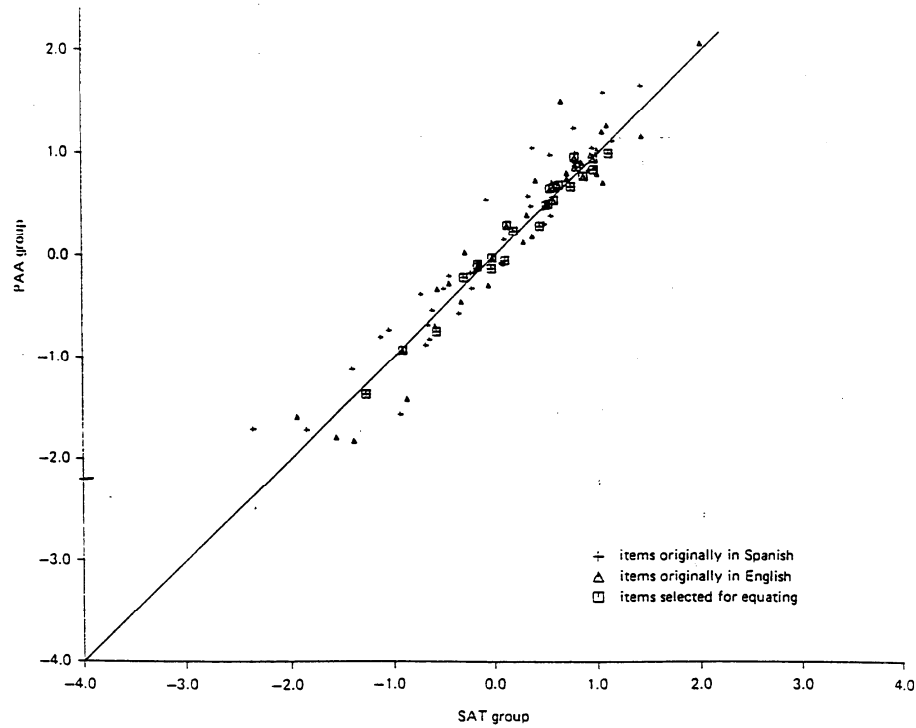


Abbildung 16: b-Plot (standardisierte Itemschwierigkeit) für 191 mathematische Items, SAT (englisch) vs. PAA (spanisch)

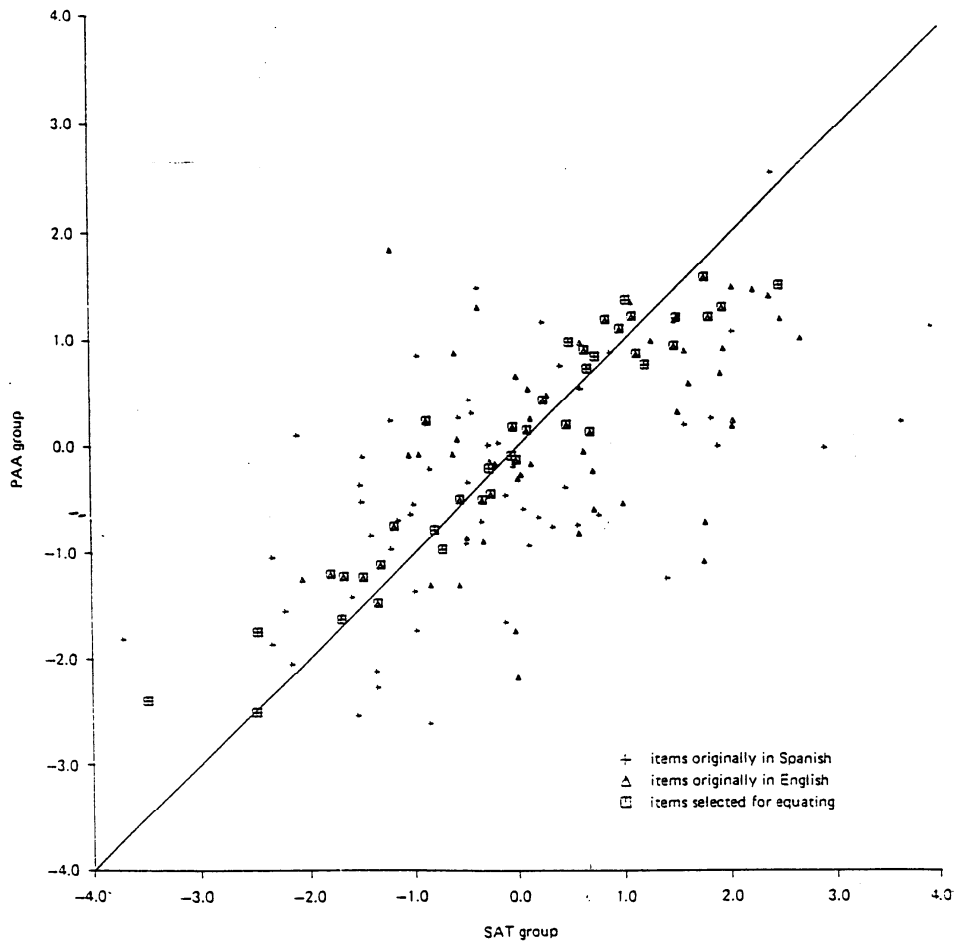


Abbildung 15: b-Plot für 142 verbale Items, jeweils getestet in einer englisch (SAT)- und spanischsprachigen (PAA) Stichprobe. aus: L.L.Cook (1998)

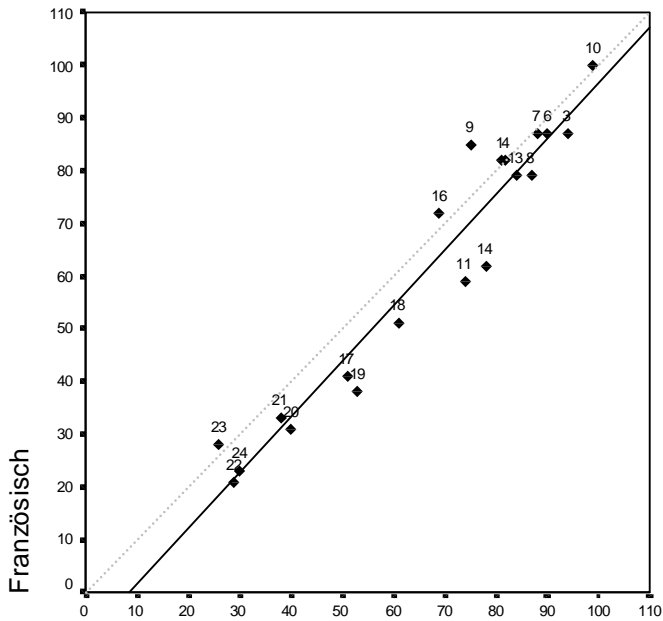
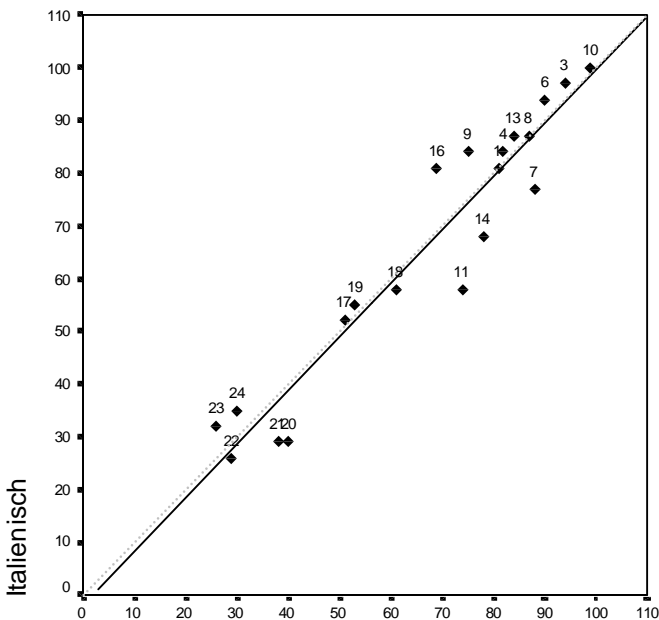


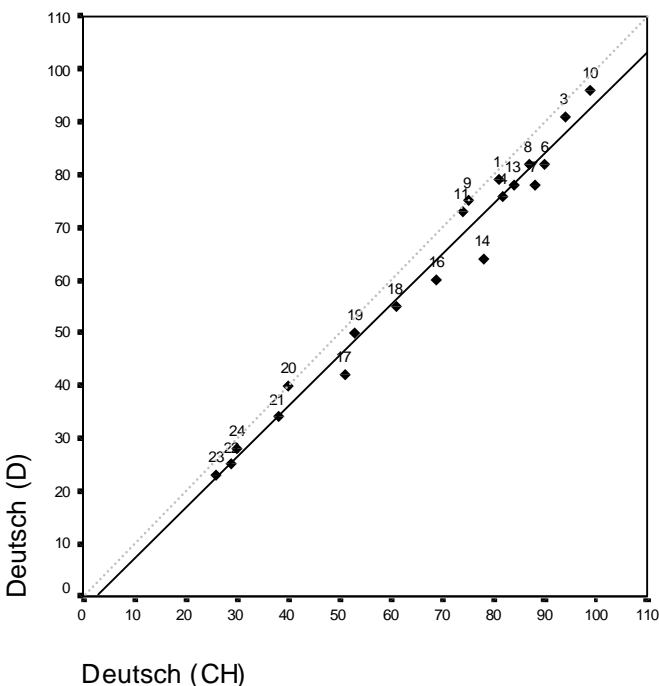
Abbildung 17: Muster zuordnen

Die Anforderung des Untertests "Muster zuordnen" ist bis auf die Instruktion sprachunabhängig. Alle Vergleiche zwischen den Sprachgruppen zeigen geringe systematische Unterschiede und recht gute Übereinstimmungen auf Item-Ebene.

Aufgaben mit identischen Schwierigkeiten in beiden Gruppen liegen auf der gestrichelten Linie (Diagonale), d.h. ein gleicher Prozentsatz Personen hat die Aufgabe richtig beantwortet.



Die durchgezogene Linie als Regressionsgerade beider Schwierigkeitsverteilungen berücksichtigt systematische Unterschiede zwischen beiden Gruppen: Weicht diese parallel zur Diagonale ab, ist ein systematischer Unterschied der Aufgabenschwierigkeiten anzunehmen, der eher auf generelle Fähigkeitsunterschiede beider Gruppen hinweist. Es sind die Leistungen derjenigen Gruppe dann besser, in deren Richtung die Regressionsgerade von der Diagonalen verschoben ist.



Für einzelne Aufgaben, die von der Regressionsgeraden abweichen, ist unterschiedliches "Funktionieren" (DIF) in den Sprachgruppen wahrscheinlich - dies wären die deutlichsten Belege für beim Test liegende Ursachen von Leistungsunterschieden. Wegen der geringen Stichprobenzahlen muss auf die statistische Prüfung dieser Abweichung verzichtet werden.

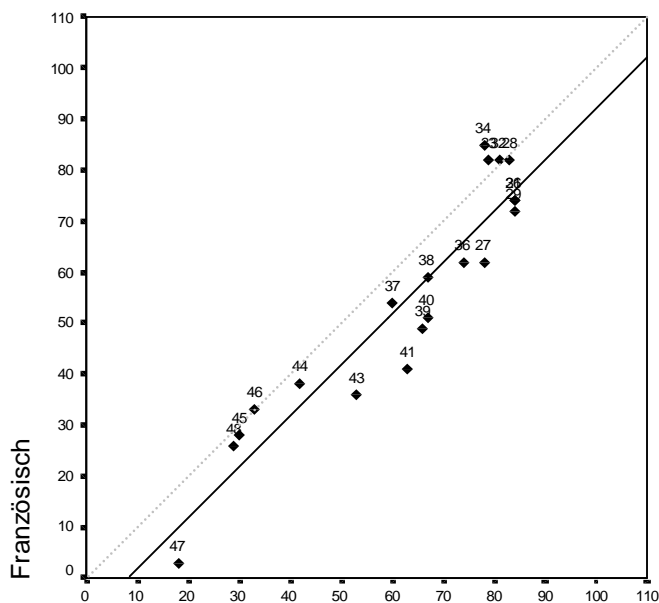
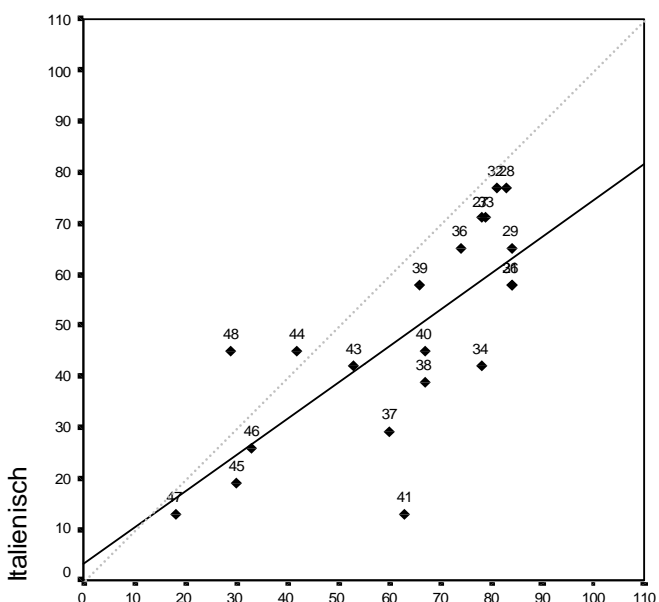


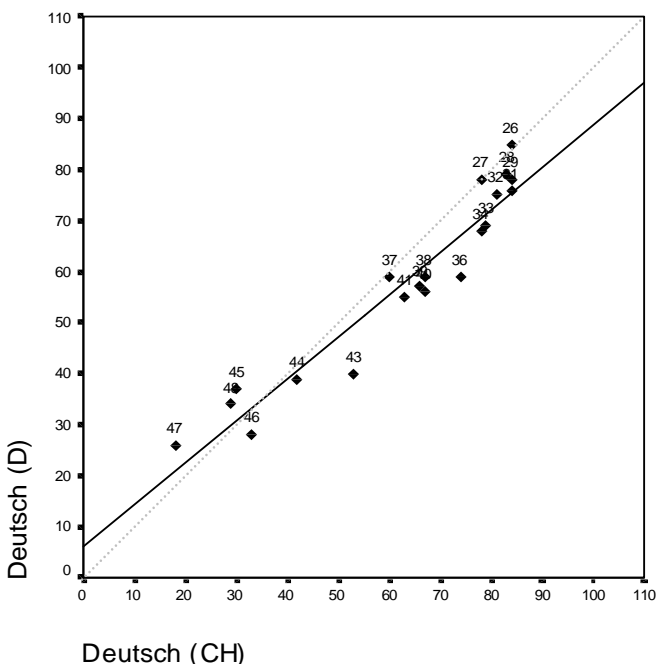
Abbildung 18: Medizinisch-naturwissenschaftliches Grundverständnis

Der Vergleich zwischen französischer und deutscher Version zeigt eine Parallelverschiebung der Regressionsgeraden. Es weichen im Vergleich zur nächsten Darstellung allerdings keine Items extrem von der Regressionsgeraden ab - "auffälligste" Items sind die Fragen 34 und 41.



Das Diagramm für die italienische Version dieses Untertests weist am deutlichsten von allen auf mögliche Sprachprobleme hin. Insbesondere die Fragen 48, 41 und 37, eventuell auch 34 und 38, weisen deutlich unterschiedliche und unsystematisch variierende Item-Schwierigkeiten auf.

Mittlerweile hat es eine Konsultation zu diesen Items mit den Übersetzern gegeben, um mögliche Ursachen zu finden. Die Richtigkeit der Übersetzung wurde bestätigt – Ursachen für diese Unterschiede wurden aus der Formulierung nicht erkannt. Die Aufgaben erscheinen gleich schwer.



Für ein präziseres Korrekturverfahren würde man eine Teilsumme nur für die abweichenden Items bilden, die Differenz zwischen beiden Sprachgruppen berechnen und der italienischsprachigen Gruppe gutschreiben. Dann blieben die generellen Unterschiede und kleinere, nicht signifikante Unterschiede erhalten (siehe Tabelle 16).

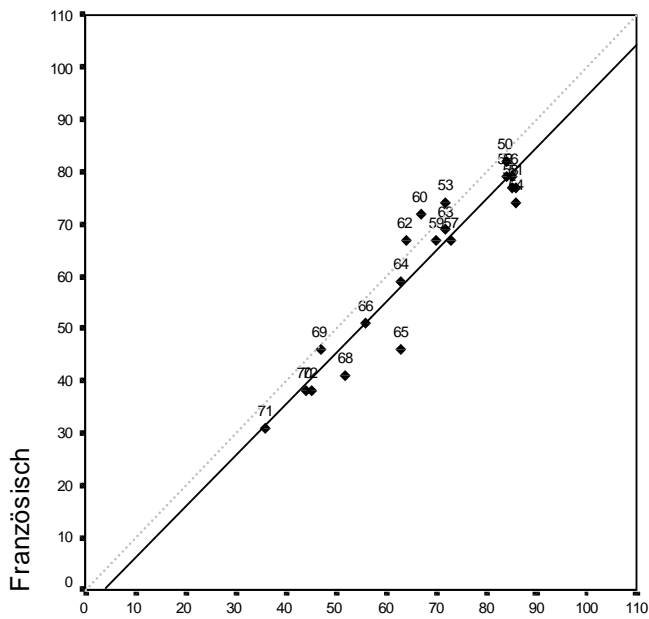
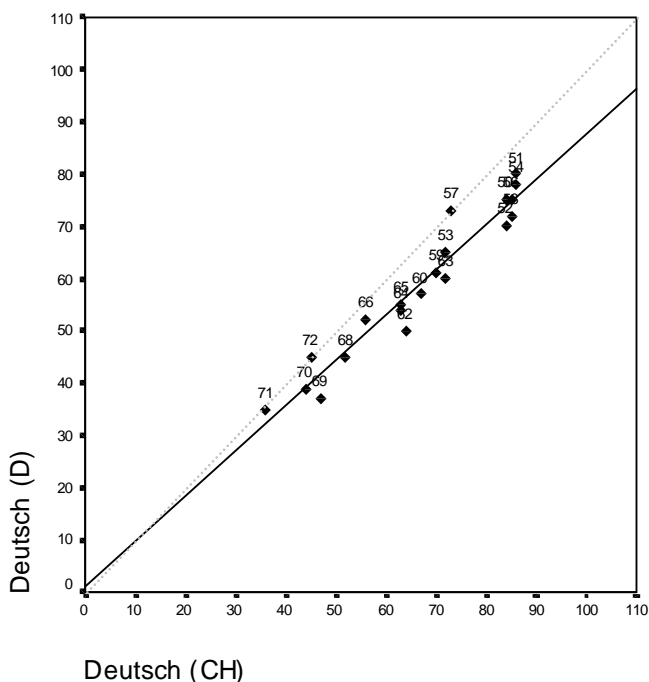
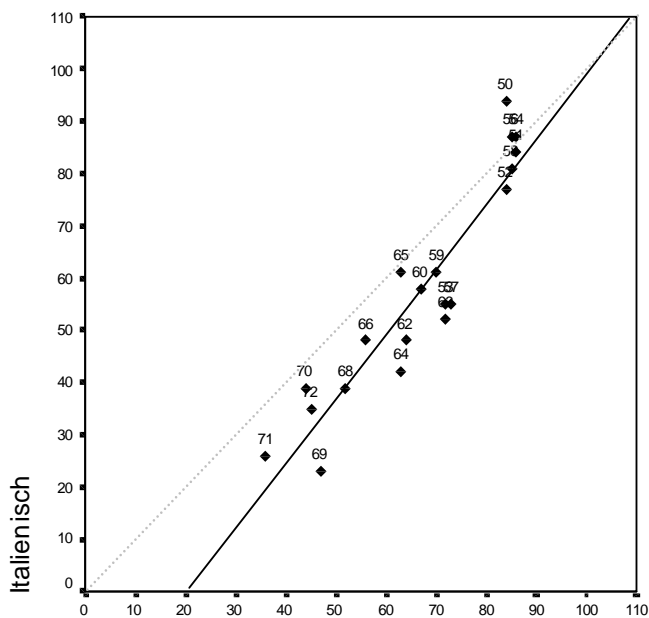


Abbildung 19: Schlauchfiguren

Dieser Untertest ist wiederum sprachunabhängig. Die einzelnen Items liegen recht eng an der Regressionsgeraden. Insbesondere bezüglich der italienischen Version zeigt sich eine systematische Verschiebung im Bereich der schwereren (letzten) Items. Die Regressionsgerade folgt hier nicht der Diagonalen. Der Leistungsunterschied zwischen beiden Gruppen ist umso deutlicher, je schwieriger die Aufgaben sind. Leichtere Aufgaben werden dagegen fast gleich gut gelöst. Dies ist ein Hinweis darauf, wie komplex die Problematik ist. Ursachen für dieses Phänomen können beim Test, in kulturellen Unterschieden oder der Stichprobenszusammensetzung liegen: Sprachprobleme könnten sich in komplexen Sätzen stärker auswirken – genauso gut können kulturelle Unterschiede ein unterschiedliches Herangehen an verschieden komplexe Aufgaben begründen. Schliesslich könnten in einer Gruppe die Leistungsbesseren weniger stark vertreten sein, welche die schwereren Aufgaben lösen würden.



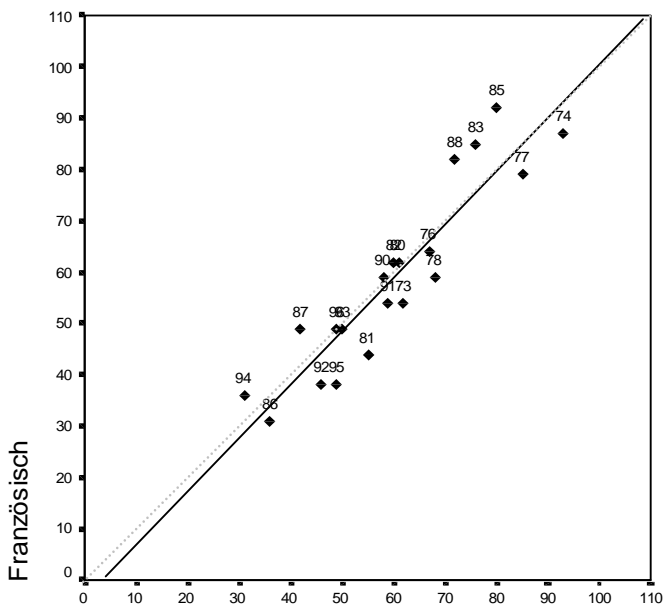
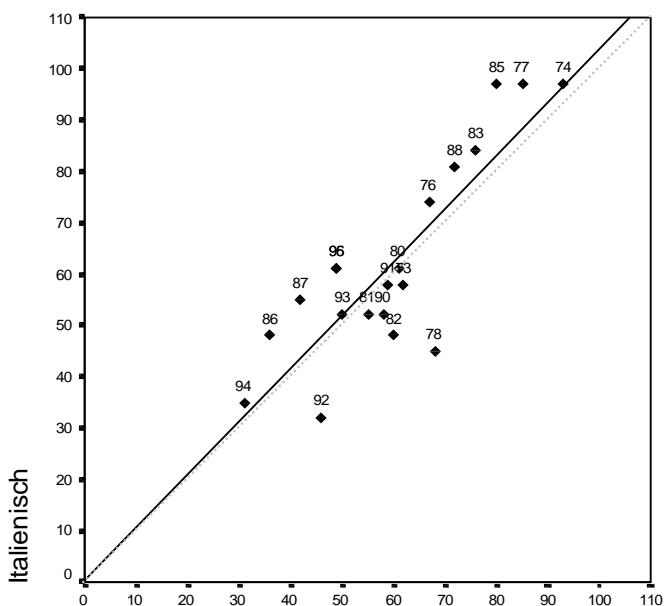
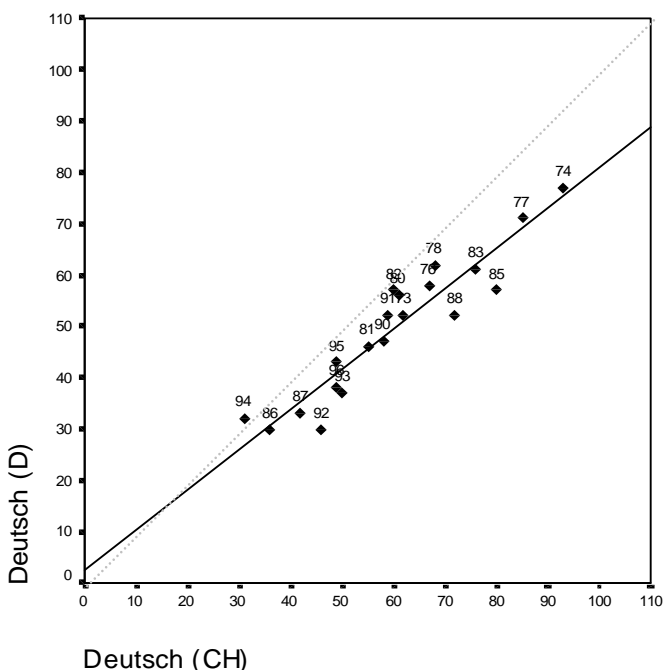


Abbildung 20: Quantitative und formale Probleme

Dieser Untertest ist verglichen mit "Textverständnis" weniger sprachabhängig, weil die Aufgaben vor allem aus identischen mathematischen Formeln bestehen. Zwischen der deutschen und französischen Version dieses Tests zeigt sich denn auch eine gute Übereinstimmung.



Auch zwischen der deutschen und italienischen Version zeigt sich kein systematischer Unterschied. Auffallend sind hier allerdings die Items 85, 78. und 92.



Bezüglich der deutschen Originalversion kann beobachtet werden, dass in der Schweiz insbesondere bei den leichteren Aufgaben bessere Resultate erzielt wurden. Mit zunehmender Schwierigkeit gleichen sich die Schwierigkeitsstufen der Aufgaben an.

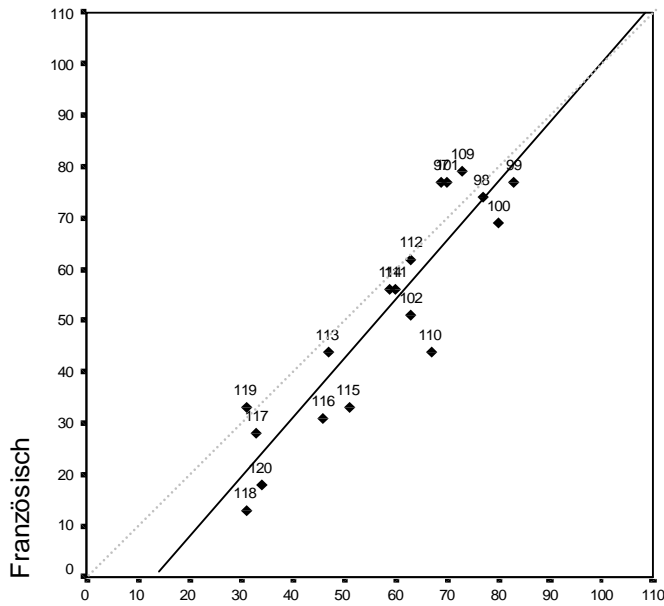
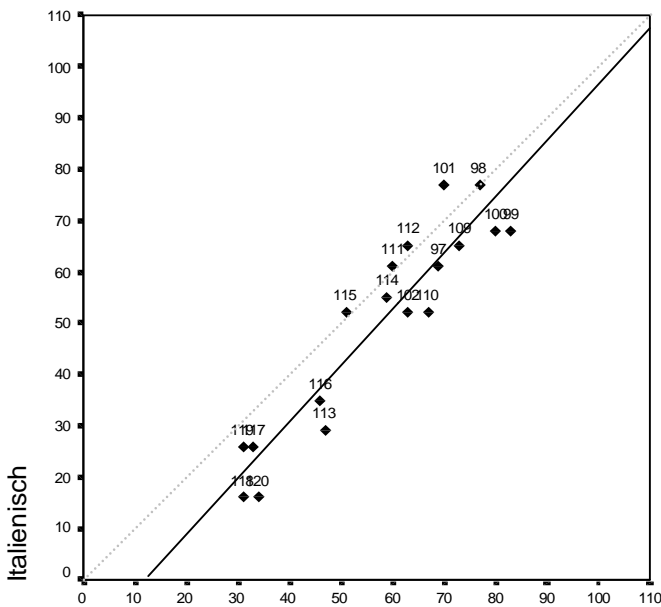
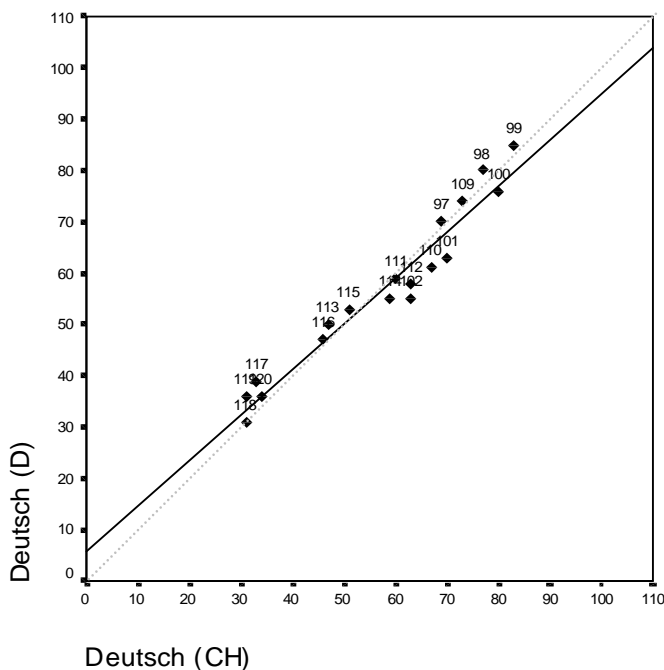


Abbildung 21: Textverständnis

In diesem Untertest findet eine leichte systematische Verschiebung im Bereich der mittleren und schwierigeren Items statt, während die leichteren Items weitgehend vergleichbare Werte aufweisen. Möglicher "Ausreisser" ist hier das Item 110 .



Eine ähnliche systematische Verschiebung wie oben zeigt sich auch bei der italienischen Übersetzung.



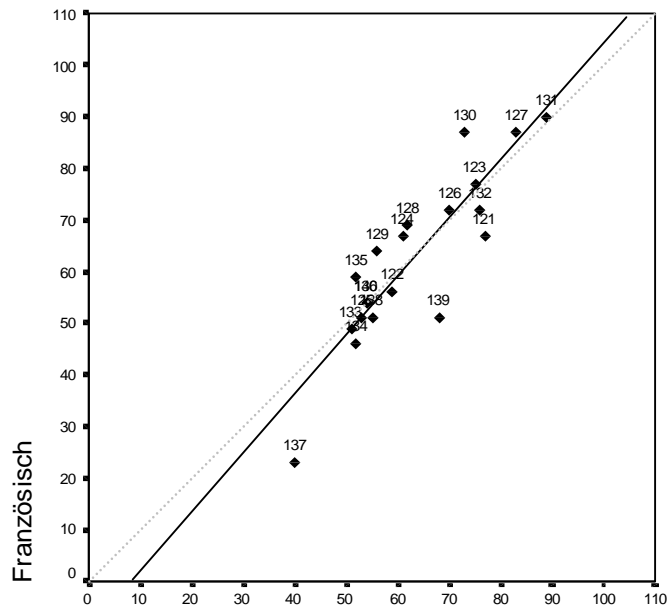
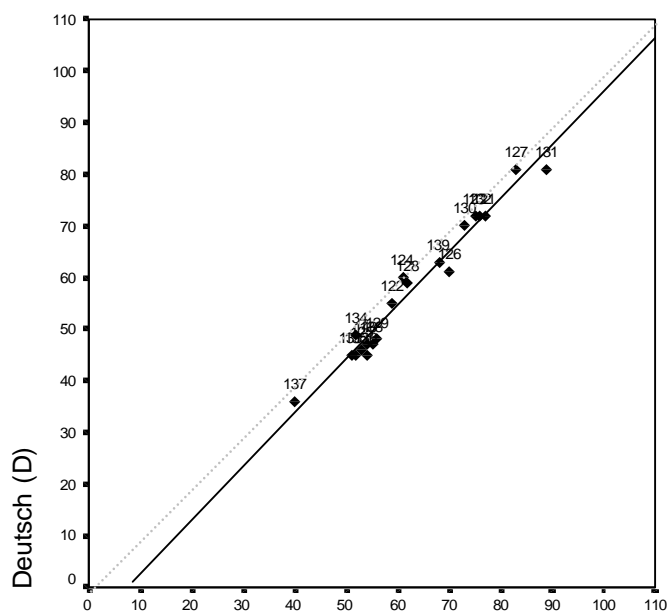
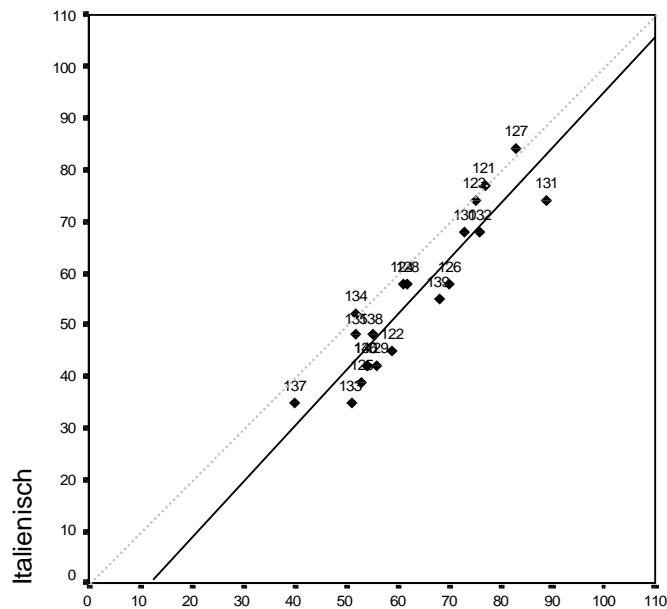


Abbildung 22: Figuren lernen

Der Untertest "Figuren lernen" ist wiederum sprachunabhängig. Die italienische und die deutsche Version weisen leichte systematische Verschiebungen auf. Die französische Version weist einige Abweichungen auf Item-Ebene auf, deren Ursprung aufgrund der Sprachunabhängigkeit des Untertests und der guten systematischen Übereinstimmung wohl vor allem in der geringen Stichprobengröße vermutet werden darf.



Deutsch (CH)

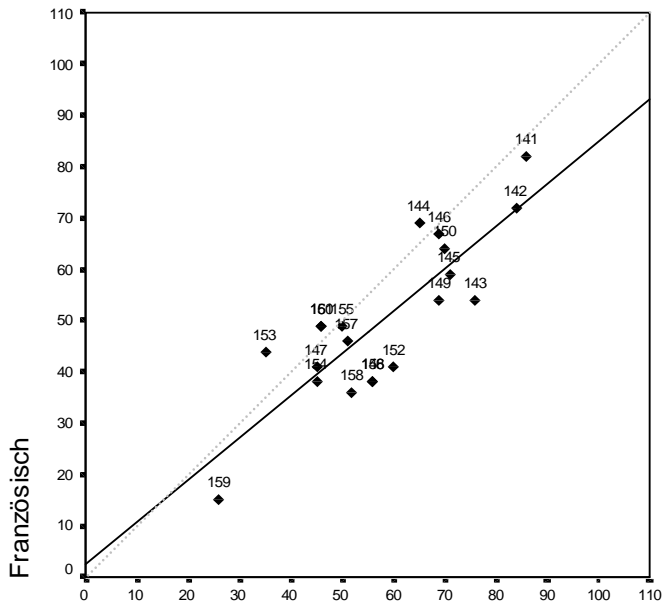
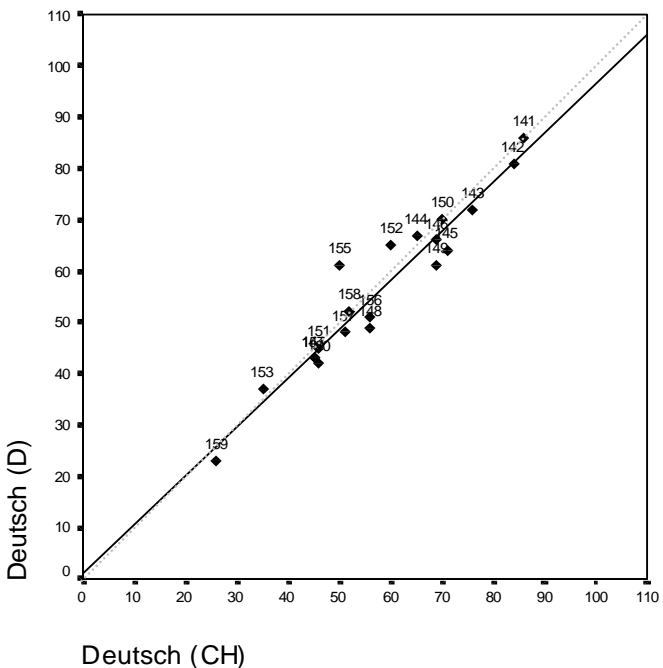
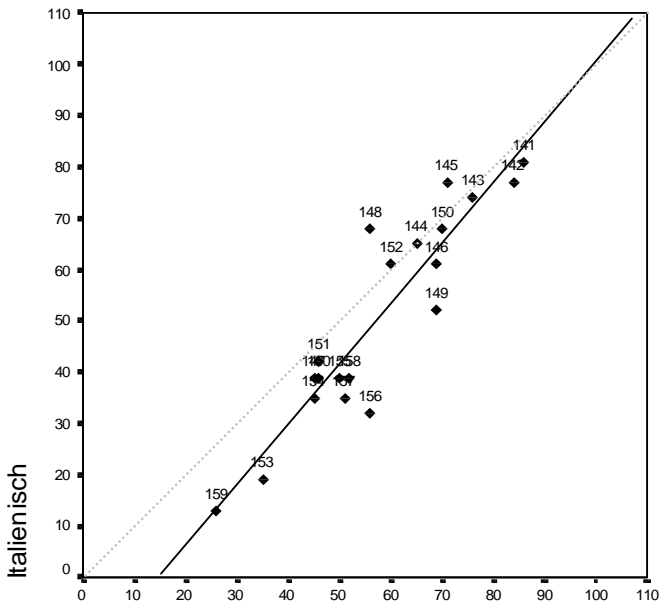


Abbildung 23: Fakten lernen

Der wesentliche Eingriff bei der Übersetzung dieses Untertests bestand in der Analoisierung der zu lernenden Personenamen (drei gehören zu einem Bereichs-Oberbegriff). Auch hier werfen die teilweise doch recht deutlichen Differenzen auf Itemebene die Frage auf, welcher Anteil der beobachteten Abweichungen auf die Stichprobengrößen zurückzuführen ist.



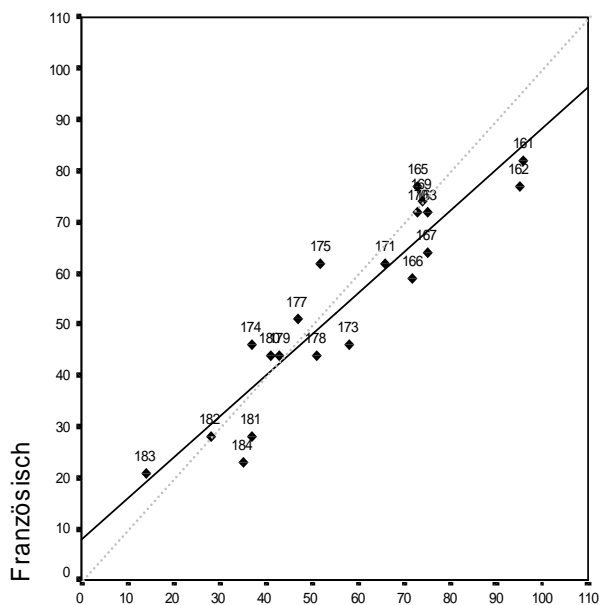
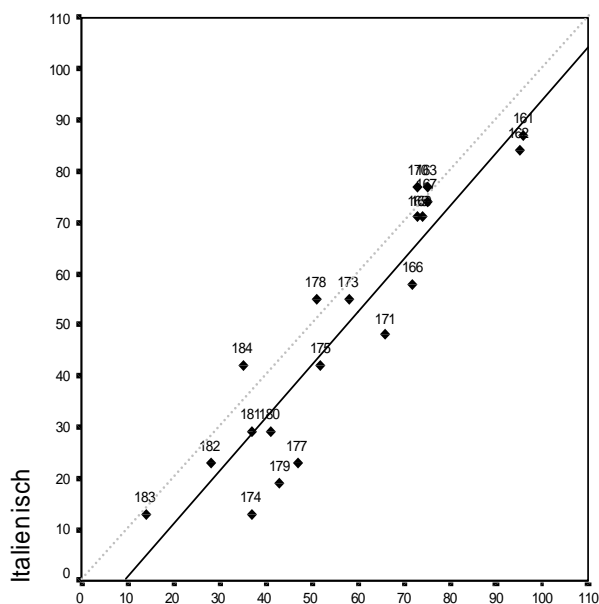
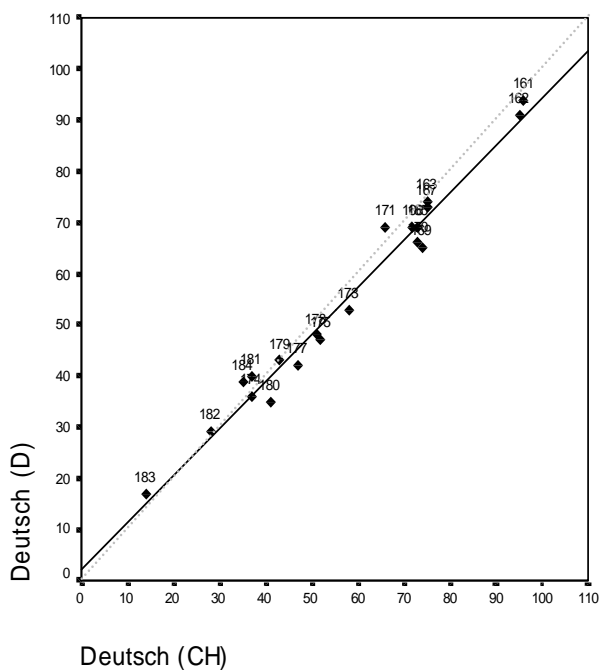


Abbildung 24: Diagramme und Tabellen

Das Diagramm für diesen Untertest zeigt einige Items, welche stärker von der Regressionsgeraden abweichen (184, 175, 165).



Die italienische Übersetzung zeigt eine deutliche systematische Verschiebung. Auch hier fällt Item 184 auf. Ausserdem liegen im Bereich der schwierigeren Fragen die Items 174, 179 und 177 deutlich unterhalb der Regressionsgeraden.



Deutsch (CH)

Ein weiterer Hinweis für DIF's liegt in unterschiedlichen Rangreihenfolgen von Itemschwierigkeiten. Üblicherweise werden leichtere Aufgaben am Anfang, schwerere eher am Ende gestellt. Die KandidatInnen können so ihr Zeitbudget in den Sprachgruppen vergleichbar gestalten. Die Kurve für die deutsche Originalversion ist nahezu monoton steigend. Durch die Abweichungen, besonders in der italienischsprachigen Stichprobe, kann auch das Zeitverhalten verändert sein. Zu schwere Aufgaben am Anfang können zu viel Zeit kosten und dann steht nicht mehr genügend Arbeitszeit für die nachfolgenden Aufgaben zur Verfügung bzw. eine andere Budgeteinteilung wirkt sich auf die nachfolgenden Aufgaben aus. Die Abbildung 25 zeigt die Sprachvergleiche für den "problematischsten" Untertest "Medizinisch-naturwissenschaftliches Grundverständnis".

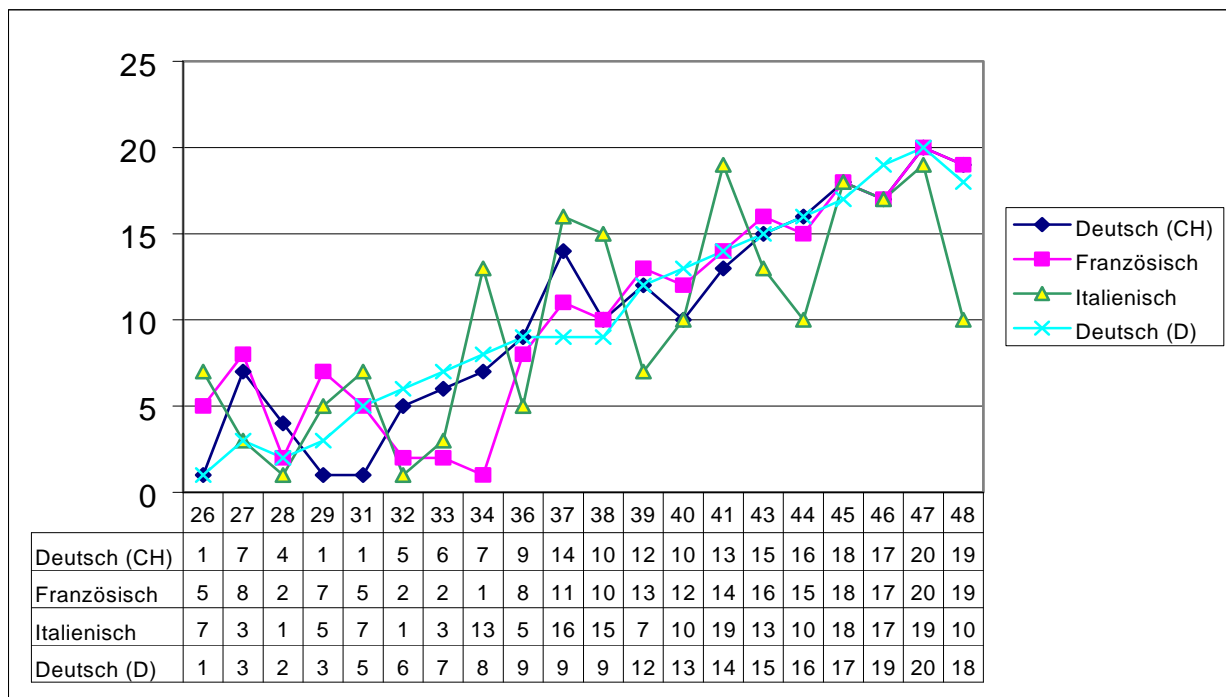


Abbildung 25: Rangreihe der Itemschwierigkeiten für den Untertest "Medizinisch-naturwissenschaftliches Grundverständnis"

Die Itemnummern sind mit dem Rangplatz der entsprechenden Itemschwierigkeit für jede Sprachversion dargestellt. Die graphische Darstellung bestätigt, dass insbesondere die italienische Version ein sehr spezifisches Profil auf der Item-Ebene aufweist.

Gleichbehandlung der Sprachgruppen

Fassen wir zunächst zusammen:

1. Es liegen Hinweise vor, dass die Zahlen-Absolutwerte der Punkte teilweise **nicht** vergleichbaren Schwierigkeitsstufen zwischen den Sprachgruppen entsprechen und dass aus Gründen der Gleichbehandlung die Herstellung von Äquivalenz geboten ist.
2. Die differenzierte statistische Berechnung dieser Äquivalenz und ein Ausgleich auf Itemebene wird durch die geringen Stichprobengrößen und die unterschiedliche Repräsentativität für die jeweiligen Sprachgruppen erschwert bzw. verunmöglicht.

3. Die in den Zahlen vorhandenen Unterschiede lassen sich nicht allein auf Sprachunterschiede reduzieren. Es gibt deutliche Hinweise, dass auch Leistungsunterschiede vorhanden sind. Diese müssen allerdings bis zum Vorliegen repräsentativer Befunde in Zusammenhang mit Punkt 2 gesehen werden.

Eine **Vergleichbarkeit der Absolutzahlen für die Punktwerte** zu erreichen ist grundsätzlich möglich: Der "Sprach-Bias" wird gesondert bestimmt und dann herausgerechnet. Eine entsprechende Untersuchung erfordert beispielsweise, dass zweisprachige Personen jeweils Teile des Tests in den verschiedenen Sprachen beantworten. Unter der Annahme einer stabilen Fähigkeit jeder Person kann man dann über einen Versuchsplan den systematischen Sprachunterschied der Items bestimmen und verrechnen. Das erscheint zu aufwendig - schon weil jedes Jahr ein neuer Test verwendet wird. Dabei ist auch damit zu rechnen, dass einzelne Items "eliminiert" werden müssen, was zu einer nicht vertretbaren Verkürzung des Tests führt (es stehen nur je 20 Aufgaben pro Untertest zur Verfügung).

Andere mögliche Ausgleichsrechnungen beruhen auf bestimmten statistischen Verfahren, "DIF" (Different Item Functioning) abzuschätzen und auszugleichen (vgl. Beller 1996). Über Lineargleichungen werden aus den sich nicht unterscheidenden Items „Ankerwerte“ bestimmt und Korrekturen für die unterschiedlichen Items berechnet. Verunmöglicht wird diese Alternative in unserem Fall aber durch die **geringen Grössen** der französisch- und italienischsprachigen Gruppe.

Langfristig könnte es Bedeutung erlangen, die Leistungen genauer aufeinander zu beziehen, wenn im Rahmen einer gesamtschweizerischen Lösung repräsentative Bewerber-Innengruppen den Test absolvieren, unterschiedliche BewerberInnenquoten (aufgrund der unterschiedlichen Maturitätsquoten) eine Sprachgruppe aber bevorteilen würden. In diesem Falle können aber aufgrund der dann vorhandenen grösseren Stichproben differenziertere Berechnungsverfahren für den Spracheinfluss vorgenommen werden, die nur den Sprach-Bias ausgleichen - andere Unterschiede aber in den Daten belassen.

Machbare Korrekturverfahren zur Herstellung von Gleichbehandlung für die nahe Zukunft (unter der Voraussetzung, dass sich die Stichprobenzahlen nicht ändern) könnten sein:

- nur die sprachabhängigen Tests zu korrigieren, indem bei vorhandenem signifikantem Unterschied die Differenz zur besten Sprachgruppe den beiden anderen Sprachgruppen gutgeschrieben wird. Andere Unterschiede bleiben erhalten. Das würde aber Konsequenzen für das Verrechnungsverfahren haben und vom bisher in der Test-Info angekündigten Verfahren abweichen.
- nur die Items zu korrigieren, die bezüglich der Schwierigkeit einen bestimmten Schwierigkeitsunterschied übersteigen. Dies ist zweifellos ein sehr aufwendiges Verfahren, welches aufgrund seiner statistischen Voraussetzungen anspruchsvoll bezüglich der politischen Vermittelbarkeit ist. Für dieses Verfahren wäre noch zu entscheiden, ob die Stichprobengrössen ausreichend sind oder die Abweichungen auf Itemebene bei Stichproben von um 30 Personen ohnehin so grossen Zufallsschwankungen unterliegen, dass **nur** auf Skalenebene eine hinreichende Verlässlichkeit für Korrekturen erreicht wird.

Durch die für 1998 gewählte Alternative einer getrennten Normierung der Sprachversionen - als das übliche Verfahren für diesen Fall - wird die **faktische Gleichwertigkeit** des Tests in den Sprachgruppen erreicht. Mögliche Unterschiede, die auf den Test zurückzuführen sind, werden eliminiert. Besteht ein Risiko, eine Sprachgruppe aufgrund einer Mitwirkung von Testursachen als "zu schlecht" oder "zu gut" zu bewerten und ist dieses Risiko

bezüglich seiner Ursachen derart komplex und nicht abschätzbar, ist die Herstellung der Gleichbehandlung angezeigt.

Letztlich war **eine politische Entscheidung** notwendig, um das Verfahren der getrennten Normierung der Sprachgruppen anzuwenden und die hier dargestellten Gründe *aufgrund der tatsächlichen Ergebnisse* als stichhaltig anzuerkennen. Der von der Schweizerischen Hochschulkonferenz eingesetzte **Beirat "Eignungstest"** des Zentrums für Testentwicklung hat diesen Fall aufgrund der Umstände als wahrscheinlich vorhergesehen. Er konnte deshalb - nach Vorliegen der konkreten Ergebnisse – diese getrennte Normierung mehrheitlich als angemessen ansehen.

Bereits in der Test-Info 1998 (für die KandidatInnen) wurde ausgeführt: "Unterschiede zwischen den Sprachformen sind durch die getrennte Berechnung der Testwerte für die Bewerberinnen und Bewerber jeder einzelnen Sprachform hingegen ausgeglichen" (S.8). Wir haben uns auch auf die Ergebnisse der beiden durchgeführten wissenschaftlichen Symposien gestützt, wo vor allem Erfahrungen bei mehrsprachigen Tests aus den USA und Israel analysiert werden konnten.

Die Zulassungsraten entsprechen in diesem Fall genau den Verhältnissen zwischen den Sprachgruppen bei der Bewerbung. Dadurch ist auch verglichen mit dem **bisherigen Verfahren** "ohne Zulassungsbegrenzung" der geringste Unterschied vorhanden: alle Sprachgruppen werden prozentual gleich vom NC "betroffen". Tabelle 14 und Tabelle 15 zeigen, dass sich die Zulassungsquoten auf der Basis des gewählten Vorgehens statistisch nicht signifikant unterscheiden.

	zugelassen	nicht zugelassen	Total
deutsch	558 82.1%	122 17.9%	680
französisch	33 84.6%	6 15.4%	39
italienisch	24 77.4%	7 22.6%	31
Total	615 82.0%	135 18.0%	750

Tabelle 14: Zulassung nach Sprache des Tests, Zulassungsquoten für die drei Sprachgruppen

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.623	2	.732
Likelihood Ratio	.605	2	.739
Linear-by-Linear Association	.150	1	.699

Tabelle 15: Prüfung Homogenität Zulassung nach Sprache des Tests, keine signifikanten Unterschiede nachweisbar

In der folgenden Tabelle 16 werden alle Möglichkeiten zusammengestellt, die wir für die Testsession 1999 und nachfolgend sehen würden, um die Gleichbehandlung der Sprachgruppen zu garantieren:

Mögliche Alternativen	
<p>Sprachgruppen einheitlich normieren</p> <p>Alle Testwerte beziehen sich auf den gleichen Mittelwert und die gleiche Standardabweichung</p>	<p>Vorteil: Auch gleiche Punktwerte bedeuten in jeder Sprachgruppe das Gleiche;</p> <p>Nachteile: Testbedingte Schwierigkeitsunterschiede führen zu Ungleichbehandlung der Sprachgruppen (gleiche Punktwerte müssen in den Sprachgruppen nicht gleichen Schwierigkeiten entsprechen);</p> <p>zusätzliches Problem: Dieses Vorgehen war für 1998 nicht angekündigt.</p>
<p>Sprachgruppen separat normieren</p> <p>(Jede Stichprobe wird als repräsentativ für die jeweilige Sprachgruppe angenommen und ihre Mittelwerte bzw. Streuungen werden zur Standardisierung herangezogen.)</p>	<p>Vorteile: Gleiche Testwerte entsprechen einer gleichen Schwierigkeit; Gleichbehandlung der Sprachgruppen – jede Sprachgruppe erhält praktisch die gleiche Zulassungsquote; mögliche testbedingte Unterschiede werden ausgeglichen;</p> <p>Nachteile: Kulturelle Unterschiede und Einflüsse unterschiedlicher Maturitätsquoten werden mit ausgeglichen (wegen der geringen Personenzahl aber wenig bedeutsam); gleiche Punktzahlen bedeuten nicht das Gleiche in den Sprachgruppen.</p>
<p>Korrektur nur derjenigen Untertests, die sprachabhängig sind (Textverständnis, Fakten lernen, Medizinisch-naturwiss. Grundverständnis, Diagramme und Tabellen, Quantitative und formale Probleme) und sich signifikant unterscheiden (keine Korrektur bei nicht sprachabhängigen Tests</p> <p>Mittlere Differenzen der anderen Sprachgruppen zur „besten Leistung“ werden errechnet; Addition der jeweiligen mittleren Differenz zur jeweiligen „schlechteren“ Sprachgruppe.</p>	<p>Vorteile: Mögliche testbedingte Ungleichbehandlung wird ausgeschlossen; Leistungsunterschiede in nichtsprachlichen Untertests bleiben erhalten;</p> <p>Nachteile: Gleiche Punktzahlen bedeuten nicht das Gleiche in den Sprachgruppen;</p> <p>Voraussetzung: Rechtzeitige Ankündigung in der Test-Info.</p>
<p>Variante: wie oben, aber nur Korrektur der Items, die signifikant von der Regressionsgeraden der Schwierigkeiten in beiden Sprachgruppen abweichen.</p> <p>Pro Sprachgruppe F und I die Items bestimmen, die sich pro Untertest in der Schwierigkeit unterscheiden.</p> <p>Nur für diese wird eine Teilsumme berechnet und die Differenz der „schlechteren“ Gruppe gutgeschrieben.</p>	<p>Bewertung wie oben, aber es werden nur die tatsächlich stark unterschiedlichen Items korrigiert.</p> <p>Für diese Items ist der Verdacht am meisten begründet, dass es sich um testbedingte Unterschiede handelt.</p> <p>Vergleiche dazu Seite 57</p>
<p>Korrektur der Untertests aufgrund von Linargleichungsmodellen (Berechnung von Ankerwerten aufgrund der sich nicht unterscheidenden Items und Berechnung eines Korrekturwertes für jedes sich signifikant unterscheidende Item).</p>	<p>Psychometrisch exaktestes Verfahren, erfordert aber grössere Stichproben, mindestens 100 Personen</p>
<p>Items vorher überprüfen an je einer zweisprachigen Stichprobe und/oder an drei getrennten Stichproben;</p> <p>Schlechte Items eliminieren bzw. Verrechnungen empirisch bestimmen.</p>	<p>Problem: geringe Itemzahl (Test nicht weiter verkürzbar); Risiko vorheriger Testbekanntheit; Kosten (v.a. dies wäre jedes Jahr neu notwendig)</p>

Tabelle 16: Verfahren zur Gleichbehandlung

Vergleichbarkeit der Testorte bzw. Testlokale

Gleiche Durchführungsbedingungen in den einzelnen Testorten sind eine wichtige Voraussetzung für Chancengleichheit. Dazu gehören beispielsweise die genaue Einhaltung von Zeiten, das wörtlich genaue Verlesen 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. Die Richtigkeit dieser Aussage lässt sich anhand der Daten überprüfen: wenn systematische Unterschiede zwischen Testlokalen bzw. Testorten auftreten, könnte dies auch an unterschiedlichen Bedingungen liegen (beispielsweise wenn tatsächlich mehr oder auch weniger Zeit für die Bearbeitung einer Aufgabe eingeräumt worden wäre).

In der Tabelle 17 werden die Testwerte aller Testlokale gegenübergestellt. Für Fribourg und Bellinzona ist die Standardisierung auf die sprachgruppenspezifischen Mittelwerte und Standardabweichungen erfolgt. Die Varianzen sind für alle Testlokale homogen (Tabelle 18).

Testlokal	N	Mittelw.	Standardabw.	Minimum	Maximum
Basel 1	33	100.4	8.7	81.00	126.00
Basel 2	32	98.2	9.5	78.00	116.00
Basel 3	34	98.0	10.0	71.00	122.00
Basel 4	32	98.9	10.1	74.00	114.00
Bern 1	78	100.1	10.0	80.00	121.00
Bern 2	84	98.3	11.2	73.00	122.00
Zürich 1	93	100.6	10.3	77.00	125.00
Zürich 2	22	100.9	8.3	82.00	116.00
Zürich 3	45	102.1	10.7	66.00	118.00
Zürich 4	45	99.5	10.5	78.00	127.00
Zürich 5	93	101.0	9.3	74.00	122.00
Chur	27	97.5	9.7	78.00	114.00
St.Gallen	62	101.5	8.5	84.00	122.00
Fribourg	39	100.0	10.0	75.00	121.00
Bellinzona	31	100.0	10.0	80.00	120.00
Total	750	100.0	9.9	66.00	127.00

Tabelle 17: Testwert (standardisiert) für die einzelnen Testlokale

	Levene Statistik	Fg1	Fg2	Signifikanz
TESTWERT	.928	14	735	.528

Tabelle 18: Prüfung auf Varianzhomogenität für Punktwert und Testwert zwischen Testlokalen

		Quadr.-summe	Fg	Mittl. Quadr.su.	F	Sig.
TESTWERT	Zwischen Gruppen	1186.032	14	84.717	.850	.614
	Innerhalb Gruppen	73232.668	735	99.636		
	Total	74418.700	749			

Tabelle 19: Einfache Varianzanalyse für den Faktor "Testlokale" für Punktwert und Testwert (standardisiert)

Auch eine varianzanalytische Prüfung findet keine signifikanten Unterschiede zwischen den Testlokalen bezüglich des Testwertes (Tabelle 19).

Die Prüfung der einzelnen Untertests zeigt für alle Testlokale homogene Mittelwertsgruppen. Es werden hier nur die Ergebnisse der deutschsprachigen Testlokale miteinander verglichen, weil mögliche Unterschiede im Verhältnis zur französisch- und italienischsprachigen Gruppe auf der Ebene des Testwertes ohnehin ausgeglichen werden.

Testlokal	N	Homogene Subgruppen für alpha = .05	
		1	2
Basel 2	32	11.5	
Basel 3	34	12.3	12.3
Basel 4	32	12.4	12.4
Zürich 1	93	12.4	12.4
Bern 1	78	12.5	12.5
Bern 2	84	12.6	12.6
Chur	27	12.7	12.7
Zürich 4	45	12.7	12.7
Basel 1	33	12.8	12.8
Zürich 5	93	12.8	12.8
Zürich 3	45	12.9	12.9
Zürich 2	22	13.4	13.4
St. Gallen	62		13.9
Sig.		.254	.641

Tabelle 20: Rangreihe der Mittelwerte beim Untertest „Konzentriertes und sorgfältiges Arbeiten“. Der multiple Mittelwertvergleich liefert 2 homogene Subgruppen, die sich bezüglich Zugehörigkeit von Basel 2 und Sankt Gallen unterscheiden.

Lediglich beim Untertest „Konzentriertes und sorgfältiges Arbeiten“, dessen Ergebnisse wir darstellen, ist der Unterschied zwischen dem besten Testlokal und demjenigen mit dem geringsten Wert signifikant (Tabelle 20). Die Differenz von 0,8 Punkte von Basel 2 zum in der Rangreihe nächsten Testlokal ist halb so gross wie die Spannweite von 1,6 aller übrigen Werte. Die „Standardabweichung Gesamt“ beträgt 3,2 – diejenigen der einzelnen Testlokale sind nicht geringer (zwischen 2,4 und 3,7). Deshalb ist dies nicht als Hinweis für eine mögliche Abweichung bei den Durchführungsbedingungen zu interpretieren.

Testorte 1998	N	Homogene Subgruppen für alpha = .05 1
Chur	27	97.5926
Basel	131	98.9389
Bern	162	99.1852
Fribourg	39	100.0256
Bellinzona	31	100.0645
Zürich	298	100.8523
St.Gallen	62	101.5161
Sig.		.377

Tabelle 21: Multiple Mittelwertvergleiche für den Testwert, alle Testorte

Auch der Vergleich der Zulassungsquoten zwischen den Testorten zeigt keine signifikanten Unterschiede (Tabellen 21 bis 23). Es kann von einer Chancengleichheit zwischen den Testorten ausgegangen werden. Da die Zulassung aufgrund der Verteilung der Testwerte erfolgt, ist diese Analyse für die Bewertung des Tests am wichtigsten.

	zugelassen	nicht zugelassen	Total
Basel	107 81.7%	24 18.3%	131 100.0%
Bern	123 75.9%	39 24.1%	162 100.0%
Zürich	253 84.9%	45 15.1%	298 100.0%
Chur	21 77.8%	6 22.2%	27 100.0%
St.Gallen	54 87.1%	8 12.9%	62 100.0%
Fribourg	33 84.6%	6 15.4%	39 100.0%
Bellinzona	24 77.4%	7 22.6%	31 100.0%
Total	615 82.0%	135 18.0%	750 100.0%

Tabelle 22: Zulassung nach Testorten

	Wert	Fg	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.794	6	.254
Likelihood Ratio	7.629	6	.267
Linear-by-Linear Association	.657	1	.418

Tabelle 23: Prüfung auf Homogenität der Zulassung nach Testorten: keine signifikanten Unterschiede

Vergleich der Geschlechter

Im Vorfeld des Tests wurde die Forderung nach Gleichbehandlung der Geschlechter mehrfach thematisiert. Dies betrifft vor allem Fragen der Testfairness: es darf kein Geschlecht durch den Test benachteiligt sein. Die Definition, was Testfairness ist, spielt dabei allerdings mit. 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 ermittelt, 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 etwa um 2 Punkte tiefere Werte als Männer. Relativ gesehen 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 muss dieser Unterschied sich im Test wiederfinden.

Im Probelauf des Tests (Hofer u.a. 1995) konnte dieser Befund so nicht bestätigt werden. Frauen erreichten, vor allem in der deutschsprachigen Stichprobe, bessere Ergebnisse als Männer.

Bei der Bewertung der nun vorliegenden Ergebnisse muss man auch berücksichtigen, dass sich mehr Frauen zum Studium beworben haben als Männer. Damit könnten sich unterschiedliche Quoteneffekte auf die Testleistungen auswirken. Anhand der konkreten Ergebnisse (Tabelle 24) kann festgestellt werden, dass sich die Testwerte bei einer Streuung von 20 Punkten (Varianzhomogenität ist gegeben, vgl. Tabelle 25) nur um 1,6 Punkte unterscheiden. Dieser **Unterschied ist statistisch nicht signifikant** (t-Test), er liegt eindeutig im Bereich des Zufalls.

GESCHLECHT	N	Mittelwert	Standardabweichung
männlich	336	109.8	20.8
weiblich	414	108.2	19.8

Tabelle 24: Mittelwerte und Standardabweichung des Punktwertes für die Geschlechter

Levene's Test auf Varianzhomogenität (F)	Sig.	t-test Mittelwertunterschiede (t)	df	Sig. (2-tailed)
.719	.397	1.131	748	.259

Tabelle 25: Prüfung auf Varianzhomogenität und Mittelwertunterschiede für Geschlechter: keine signifikanten Unterschiede anzunehmen.

Auch die Zulassungsquoten selbst (Tabelle 26) unterscheiden sich nicht signifikant – hier spielt bekanntlich neben Mittelwert und Streuung auch die Art der Verteilung eine Rolle. Die drei Prozent Differenz könnten auch teilweise auf die Auswirkung unterschiedlicher BewerberInnenquoten zurückzuführen sein - sie bleiben aber wiederum ganz eindeutig im Bereich der Zufallsschwankungen.

	zugelassen	Nicht zugelassen	Total
männlich	281 83.6%	55 16.4%	336 100.0%
weiblich	334 80.7%	80 19.3%	414 100.0%
Total	615 82.0%	135 18.0%	750 100.0%

Tabelle 26: Zulassungsquoten nach Geschlecht

	Wert	Fg	Asymp. Sig. (2-seitig)	Exact Sig. (2-seitig)
Pearson Chi-Square	1.097	1	.295	
Fisher's Exact Test				.339

Tabelle 27: Prüfung der Homogenität Zulassung (Geschlecht) keine signif. Unterschiede

In den einzelnen Untertests sind, wie sich aufgrund der bisherigen Befunde leicht vermuten liess, Geschlechtsunterschiede nachweisbar (Abbildung 26). Es gibt Tests mit besseren Werten für die Frauen und solche mit besseren Werten für Männer. Die Werte sind im Anhang (Tabelle A 14) aufgeführt. Der Test als Ganzes gleicht die Unterschiede in den Untertests aus – die gleiche Gesamtleistung kann auf mehreren Wegen, aufgrund unterschiedlicher Kombinationen der Fähigkeiten erreicht werden. Es ist ein besonderer Vorzug des Tests, dass das Testprofil bezüglich der Geschlechter ausbalanciert ist. Dieser Befund bestätigt sich auch für die Ergebnisse der Schweiz.

Die Faktorenanalysen (s.u.) zeigen, dass die einzelnen Untertestleistungen für Männer und Frauen die gleiche Struktur aufweisen. Dies spricht gegen qualitative Unterschiede dieser Fähigkeitsprofile - sie sind eher quantitative Abstufungen.

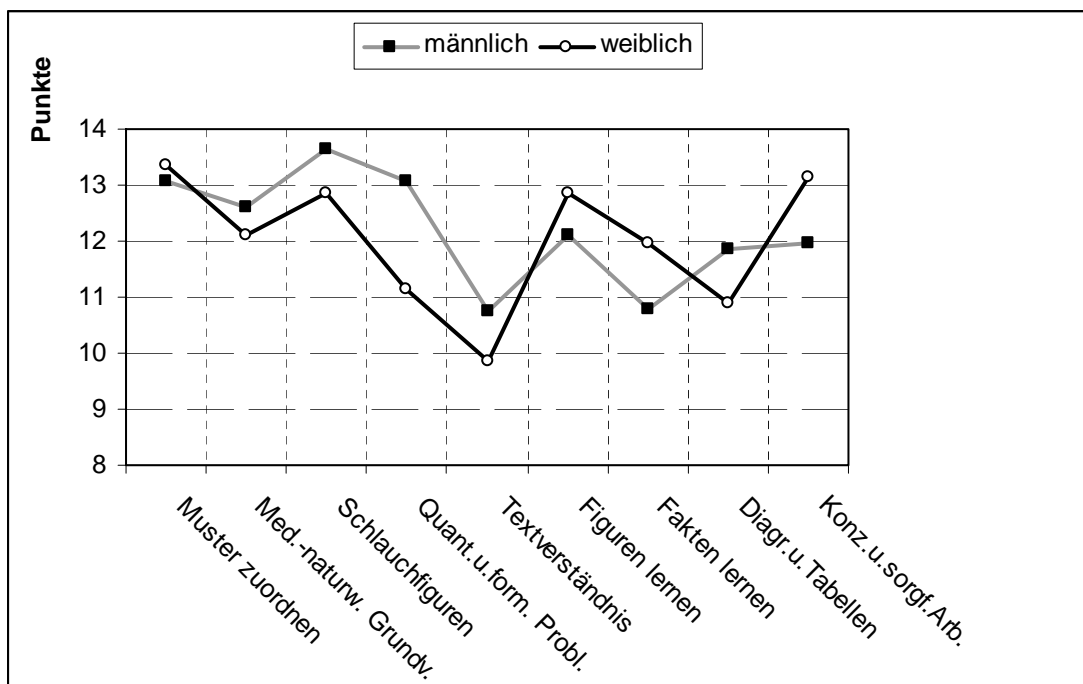


Abbildung 26: Mittelwerte-Profil der Untertests für die Geschlechter

Vergleich der Kantone innerhalb der Deutschschweiz

Beim Vergleich der Testergebnisse zwischen den Kantonen ist zu beachten, dass **nicht die Qualität** der Maturitätsausbildung verglichen wird. Wenn der Test **weniger abhängig vom Vorwissen** ist, so gilt das natürlich auch für das innerhalb der Schulausbildung angeeignete Wissen. Psychologisch gemessene Leistungen werden üblicherweise danach unterschieden, wie stark sie bildungsabhängig sind. Erst eine Differenz zwischen bildungsabhängigen und weniger abhängigen Leistungsmessungen erlaubt Rückschlüsse, ob die Ressourcen oder Potentiale einer Person im Bildungsprozess adäquat gefordert bzw. umgesetzt worden sind. Für die Evaluation der verschiedenen Bildungssysteme wären dann z.B. Wissenstests als ergänzende Vergleichsgrösse notwendig.

Dagegen wird im Eignungstest vor allem gemessen, inwieweit sich die getesteten Personen hinsichtlich ihrer **Fähigkeiten zur Wissensaneignung** unterscheiden. Diese können üblicherweise auch als Voraussetzungen guter Schulleistungen gelten. Vorstellbar wäre, dass langfristig über die gesamte Schulzeit die Fähigkeiten zur Wissensaneignung trainiert werden und so die Qualität der Schulausbildung einen Einfluss auf diese Fähigkeiten hat. Die später dargestellten Befunde zur Altersabhängigkeit sprechen in gewisser Weise dagegen: es ist für die Testleistungen der "Älteren" kein Vorteil, wenn sie ihren Schulabschluss erst unmittelbar vor der Studienbewerbung erworben haben und noch "im Training" sind.

In den deutschen Ergebnissen wird generell eine signifikante Korrelation zwischen Abiturnote und Testergebnis um 0.40 (Trost 1995 S. IV) nachgewiesen. Sogenannte „Verwirbelungseffekte“, also im Einzelfall nicht übereinstimmende Ergebnisse gegenüber einer nur notengesteuerten Auswahl, betrug 1994/95 beispielsweise bei den aufgrund der

Testquote Zugelassenen um 90%, bei der aufgrund der Abitur-Test-Quote Zugelassenen zwischen 25 und 34 %. (Trost 1995, S. 200). Im Jahr 1993/94 wurden im Studiengang Zahnmedizin über die Testquote dank guter Testwerte 80 von 171 Bewerbern mit der Abitur-Note 3 und schlechter zugelassen.

Eine Hypothese zur Erklärung der Unterschiede zwischen den Sprachregionen war die unterschiedliche Maturitätsquote. Auch innerhalb der Deutschschweiz unterscheiden sich die einzelnen Kantone bezüglich der Maturitätsquote als Anteil der Maturitäten an der 19jährigen ständigen Wohnbevölkerung deutlich. Wie unterscheiden sich die Kandidatinnen und Kandidaten aus Kantonen mit unterschiedlicher Maturitätsquote? In den Diskussionen findet man die Meinung, dass bei steigender Quote auch mehr Personen die Maturität ablegen, deren Fähigkeiten „eigentlich“ dafür nicht ausreichen würden.

hoch (> 18%)			mittel (13 – 18%)			niedrig (< 13%)		
Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl	Kanton	Maturitäts- quote	Anzahl
BL	23,3	50	SH	16,6	7	BE	12,9	96
BS	20,7	40	ZG	16,4	6	SG	12,8	64
ZH	18,7	163	AG	16,2	66	SZ	12,7	21
			AR	16,0	4	LU	12,2	43
			GL	14,1	4	OW	12,2	3
			AI	13,7	2	NW	10,2	6
			SO	13,6	26	TG	9,8	22
			GR	13,2	19			
			UR	13,0	3			
Summe		253	Summe		137	Summe		255

Tabelle 28: Gruppierung der Kantone nach der Maturitätsquote

Zur Überprüfung dieser Hypothese wurden drei Gruppen der Kantone nach der Maturitätsquote gebildet. Dabei wurden nur die Personen dieser Kantone berücksichtigt, die den Test in deutscher Sprache abgelegt haben (Tabelle 28). Die Aussage 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 (Tabelle 29, Prüfergebnisse siehe Anhang Tabelle A 17).

Um die Frage regionaler Unterschiede weiter zu untersuchen, wurden die Kantone mit mehr als 40 Teilnehmerinnen und Teilnehmern am Test verglichen (wiederum nur deutsche Sprachgruppe). Es zeigt sich, dass nur bei „Textverständnis“ und „Konzentriertem und sorgfältigem Arbeiten“ signifikante Unterschiede nachweisbar sind – der Testwert als eigentliches Zulassungskriterium unterscheidet sich nicht signifikant. Beim „Textverständnis“ sind alle Mittelwerte homogen. Beim „Konzentrierten und sorgfältigen Arbeiten“ werden zwei homogene Gruppen gefunden, die sich bezüglich Luzern und Basel bzw. St. Gallen unterscheiden.

Maturitäts- quote	Punkt- wert	Test- wert	Muster zuord- nen	Med.- naturw. Grund- verst.	Schlauch - figuren	Quant. und formale Probl.	Textver- ständ- nis	Figuren lernen	Fakten lernen	Diagr.u. Tabel- len	Konz. u. sorgf. Arbei- ten
----------------------	----------------	---------------	-------------------------	--------------------------------------	--------------------------	------------------------------------	---------------------------	-------------------	------------------	---------------------------	-------------------------------------

hoch	m	110.1	100.1	13.1	12.6	13.5	12.0	10.5	12.5	11.5	11.4	12.5
	s	21.2	10.4	3.0	3.4	3.6	3.8	3.7	3.8	3.6	3.2	3.3
mittel	m	111.5	100.8	13.4	12.7	13.6	12.4	10.6	12.4	11.6	11.3	13.0
	s	17.3	8.4	2.9	3.1	3.1	3.6	3.1	3.7	3.3	3.2	3.2
niedrig	m	110.7	100.4	13.5	12.5	13.1	12.0	10.1	12.9	11.7	11.6	12.9
	s	20.3	10.0	3.2	3.2	3.5	4.1	3.4	3.5	3.4	3.4	3.2

Tabelle 29: Mittelwerte und Standardabweichungen für drei Gruppen der Kantone nach der Maturitätsquote

Bezogen auf den **Testwert** als eigentlichem Zulassungskriterium bleibt festzuhalten, dass keine Fähigkeitsunterschiede zwischen den Personen der deutschschweizer Kantone nachweisbar sind, die sich auf das Zulassungsverfahren auswirken.

Kanton		Punkt- wert	Test- wert	Muster zuord- nen	Med.- naturw. Grund- verst.	Schlauch - figuren	Quant. und formale Probl.	Textver- ständ- nis	Figuren lernen	Fakten lernen	Diagr.u. Tabel- len	Konz. u. sorgf. Arbei- ten
AG	m	112.7	101.4	13.4	12.8	13.2	12.7	10.8	12.6	12.2	11.5	13.0
	s	16.1	7.9	3.0	3.1	2.7	3.6	2.9	4.0	3.2	3.2	2.9
BE	m	110.2	100.3	13.3	12.6	13.4	11.8	10.3	12.4	11.5	11.5	12.9
	s	22.4	10.9	3.1	3.5	3.8	4.2	3.3	3.5	3.8	3.4	3.2
BL	m	108.0	99.1	13.1	12.5	12.8	11.4	10.6	12.1	11.3	11.5	12.4
	s	18.9	9.3	2.9	3.2	3.6	3.6	3.3	3.9	3.8	3.2	3.0
BS	m	104.3	97.2	12.7	11.8	13.6	10.9	9.0	12.0	10.9	11.3	11.7
	s	21.4	10.5	3.0	3.8	3.6	3.9	3.9	3.4	3.6	3.1	3.0
LU	m	109.3	99.7	13.7	12.0	13.0	12.6	9.5	13.4	11.6	11.6	11.6
	s	17.7	8.7	3.1	2.7	3.0	4.3	3.1	3.1	2.9	3.3	2.8
SG	m	112.0	101.0	13.6	12.28	13.4	11.9	10.3	13.2	11.7	11.3	13.8
	s	18.2	8.9	3.3	2.7	3.1	3.7	3.3	4.0	3.3	3.1	3.1
ZH	m	112.2	101.2	13.3	12.9	13.8	12.5	10.9	12.7	11.7	11.4	12.8
	s	21.5	10.5	3.1	3.3	3.6	3.7	3.7	3.9	3.5	3.3	3.3

Tabelle 30: Vergleich der Kantone mit mehr als 40 TeilnehmerInnen für Testwert und die einzelnen Untertests

Für die beiden in der Varianzanalyse bezüglich der Kantone signifikanten Untertests seien nachfolgend die Mittelwerte dargestellt (Tabellen 31 und 32). Sie wurden mittels multiplen Mittelwertvergleich hinsichtlich der Signifikanz geprüft. Eine Interpretation der Ursachen ist aufgrund fehlender Vergleichsdaten kaum möglich.

Kanton	N	Homogene Subgruppe für alpha = .05
BS	40	9.0
LU	43	9.5
BE	100	10.3
SG	64	10.3
BL	50	10.6
AG	66	10.8
ZH	164	10.9
Sig.		.054

Tabelle 31: Mittelwerte für "Textverständnis", keine Mittelwertunterschiede beim multiplen Mittelwertvergleich nach Tukey trotz signifikantem Unterschied zwischen den Kantonen in der Varianzanalyse

Kanton	N	Homogene Subgruppen für alpha = .05	
LU	43	11.6	
BS	40	11.7	
BL	50	12.4	12.4
ZH	164	12.8	12.8
BE	100	12.9	12.9
AG	66	13.0	13.0
SG	64		13.8
Sig.		.217	.154

Tabelle 32: Mittelwerte für "Konzentriertes und sorgfältiges Arbeiten", zwei homogene Mittelwertgruppen beim multiplen Mittelwertvergleich nach Tukey

Vergleiche für Altersgruppen und Zeitpunkt des Maturitätsabschlusses

Zur Altersabhängigkeit vom Studienerfolg gibt es zwei gegensätzliche Tendenzen, die sich aus psychologischen Befunden ableiten lassen: 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 idealerweise so ergänzen, dass die effektive Leistungsfähigkeit mit steigendem Alter zumindest gleichbleiben kann. Es hängt natürlich von der Art der Prüfung der Studienanforderungen ab, ob eine Tendenz bevorteilt wird: bei Aufgaben, die eher auf Auswendiglernen hinauslaufen, sollten Jüngere besser abschneiden – bei dem, was man problemorientiertes Studium nennt, sollten Ältere zumindest nicht schlechter sein. Für beide Tendenzen finden sich im EMS entsprechende Tests. 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: die Geburtsjahre 1978-1980, die sich im 19. Lebensjahr unmittelbar nach der Maturitätsprüfung bewerben, die Jahre 1975-1977, die dies 1-2 Jahre später tun und die Jahre 1952-1974, die dies mehrere Jahre später tun (Tabelle 33).

	Häufigkeit	kumulative Prozent		Häufigkeit	kumulative Prozent
1952	1	.1	1969	7	4.5
1954	1	.3	1970	4	5.1
1957	1	.4	1971	6	5.9
1959	1	.5	1972	12	7.5
1961	3	.9	1973	10	8.8
1962	3	1.3	1974	12	10.4
1963	2	1.6	1975	16	12.5
1964	3	2.0	1976	58	20.3
1965	2	2.3	1977	147	39.9
1966	5	2.9	1978	364	88.4
1967	1	3.1	1979	84	99.6
1968	4	3.6	1980	3	100.0

Tabelle 33: Aufteilung der KandidatInnen nach Geburtsjahren

Die **älteste Gruppe** lässt sich nach dem Alter beim Erlangen der Maturität 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 34). 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 bei 22 Jahren angesetzt. Beide Gruppen unterscheiden sich nicht bezüglich des gegenwärtigen Alters: $29,4 \pm 5,1$ Jahre (frühere Maturität) vs. $28,7 \pm 4,9$ Jahre (spätere Maturität).

Alter bei Abschluss der Matura	Geburtsjahr			Alter bei Abschluss der Matura	Geburtsjahr		
	1952-1974	1975-1977	1978-1980		1952-1974	1975-1977	1978-1980
17			1	28	2		
18	2		9	29	3		
19	18	10	123	30	3		
20	11	72	318	31	2		
21	4	102		32	4		
22		36		33	1		
23	1	1		34	1		
24	9			36	1		
25	6			40	1		
26	6			42	1		
27	2						

Tabelle 34: Differenzierung der Altersgruppen nach dem Alter bei Abschluss der Maturitätsprüfung

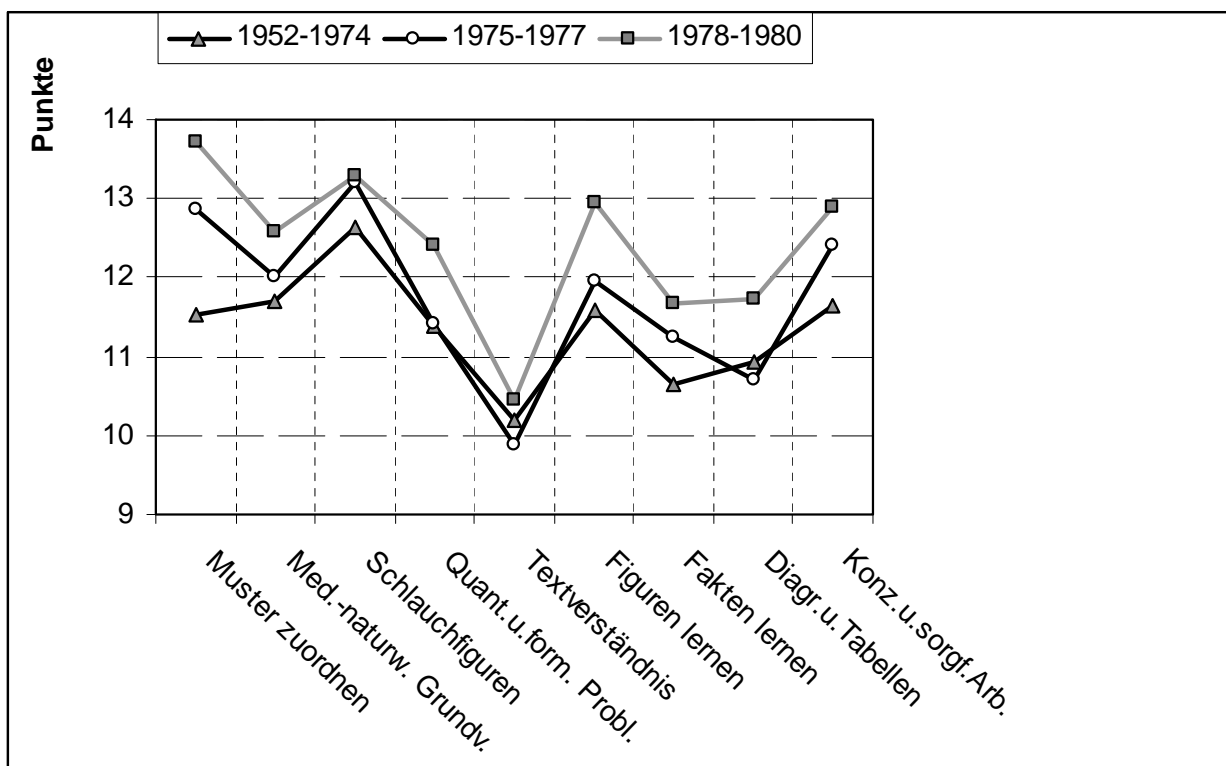


Abbildung 27: Mittelwerte der Untertests für Altersgruppen (nach Geburtsjahr)

Die Leistungen der drei Altersgruppen sind in Abbildung 27 bzw. Tabelle 35 dargestellt, die Ergebnisse der statistischen Prüfungen befinden sich im Anhang. Bis auf zwei Untertests sind die Unterschiede signifikant – die Mittelwerte der jüngeren Gruppe sind bei Punkt- und Testwert signifikant am besten, die beiden anderen Altersgruppen unterscheiden sich hier nicht signifikant. Es lässt sich ein Vorteil der jüngeren Bewerberinnen und Bewerber nachweisen, die sich sofort nach einer frühestmöglich bestandenen Maturität zum Studium bewerben. Dies dürfte 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. Die Unterschiede zeigen sich hier nicht in den „Werkzeugfunktionen“ Aufmerksamkeit und Gedächtnis, sondern dem logischen Denken (Quantitative und formale Probleme, Diagramme und Tabellen). Diese Fragen müssten im Zusammenhang mit den Entscheidungen, eine Maturitätsprüfung abzulegen, weiter differenziert werden. Die Leistungsbesten haben sich offenbar auch früh für eine Maturität entschieden – diese Unterschiede gleichen sich später nicht aus.

Alter bei Maturität		Punkt-wert	Test-wert	Muster-zuordnen	Med.-naturw. Grundv.	Schlauch-figuren	Quant formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagramme und Tabellen	Konz.u. sorgf. Arbeiten
1978-1980 (1)	m	111.6	101.5	13.7	12.5	13.2	12.4	10.4	12.9	11.6	11.7	12.8
	s	19.7	9.5	3.1	3.3	3.48	3.8	3.5	3.6	3.5	3.2	3.1
1975-1977 (2)	m	105.6	98.1	12.8	12.0	13.2	11.4	9.8	11.9	11.2	10.7	12.4
	s	20.9	10.3	2.9	3.4	3.4	3.9	3.6	3.7	3.6	3.4	3.4
1952-1974 (3)	m	102.2	96.5	11.5	11.6	12.6	11.3	10.2	11.5	10.6	10.9	11.6
	s	19.3	9.5	2.9	3.2	3.6	4.2	3.2	3.6	3.0	3.0	3.4
sig VA		**	**	**	**		**		**	**	**	**
sig mm		1 – 23	1 – 23	1-2-3					1-23	12-23	12-23	12-23
Differenzierung der Gruppe 1952-1974 nach Alter bei Erlangen der Maturität												
bis 21 Jahre	m	106.6	98.8	11.3	12.4	13.2	12.6	11.0	11.7	10.7	12.0	11.4
	s	18.9	9.2	3.0	3.4	3.1	4.6	3.3	3.9	3.1	2.7	3.2
23-42 Jahre	m	98.6	94.5	11.6	11.0	12.1	10.3	9.5	11.4	10.5	10.0	11.7
	s	19.0	9.3	2.9	3.0	3.9	3.6	3.1	3.5	3.0	3.0	3.6
sig VA		**	**				**	*			**	

Tabelle 35: Mittelwerte und Standardabweichungen für Altersgruppen, Prüfung der Unterschiede auf Signifikanz; Varianzanalyse (sig. VA) und multipler Mittelwertvergleich nach Tukey (sig. mm) – signifikant unterschiedliche Gruppen sind durch Gedankenstrich getrennt.

Die Zulassungsquoten für die drei Altersgruppen unterscheiden sich signifikant (Tabelle 36). 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 (1952-1974 geboren) nicht chancenlos ist, wird durch den Vergleich der Zulassungsquoten gezeigt. Die Personen, die ihre Maturitätsprüfung früh abgelegt haben, erreichen die Zulassungsquote der mittleren Altersgruppe (77%). 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.

	nicht zugelassen	zugelassen	Total
1952 – 1974	21 26,9 %	57 73,1%	78 100.0%
1975 – 1977	50 22,6 %	171 77.4%	39 100.0%
1978 – 1980	55 12.2%	396 87.8%	451 100.0%
Total	135 18,0%	615 82,0%	750 100.0%
1952 – 1974 frühe Maturität	8 22.9%	27 77,1%	35 100%
1952 – 1974 späte Maturität	13 30.2%	30 69.8%	43 100 %

Tabelle 36: Zulassung nach Geburtsjahr – unten Differenzierung der älteren Gruppe nach dem Zeitpunkt des Ablegens der Maturitätsprüfung

	Wert	Fg	Asymp. Sig. (2-seitig)
Pearson Chi-Square	17.924	2	.000
Likelihood Ratio	17.389	2	.000
Linear-by-Linear Association	17.090	1	.000

Tabelle 37: Homogenitätsprüfung Zulassung nach Alter: signifikanter Unterschied

Vergleiche der 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 BewerberInnen an allen Universitäten unterscheiden sich bezüglich der Studieneignung nicht (Tabellen 38 und 39).

	zugelassen	nicht zugelassen	Total
Basel	138 82.6%	29 17.4%	167 100.0%
Bern	131 80.4%	32 19.6%	163 100.0%
Fribourg	78 77.2%	23 22.8%	101 100.0%
Zürich	268 84.0%	51 16.0%	319 100.0%
Total	615 82.0%	135 18.0%	750 100.0%

Tabelle 38: Zulassungsquoten für Gruppen nach Wunschuniversität

	Wert	df	Asymp. Sig. (2-seitig)
Pearson Chi-Square	2.773	3	.428
Likelihood Ratio	2.700	3	.440
Linear-by-Linear Association	.296	1	.586
N of Valid Cases	750		

Tabelle 39: Prüfung Homogenität Zulassung nach Wunschuniversität: keine signifikanten Unterschiede

Ergebnisse zur Testgüte

Zuverlässigkeit

Die Zuverlässigkeit (Messgenauigkeit 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
Muster zuordnen	.64 - .69	.69	.67 - .73	.72
Med.-nat. Grundverständnis	.64 - .73	.70	.65 - .72	.72
Schlauchfiguren	.76 - .80	.71	.75 - .82	.75
Quant. u. Form. Probleme	.69 - .78	.76	.70 - .75	.78
Textverständnis	.68 - .71	.73	.69 - .74	.75
Figuren lernen	.69 - .74	.72	.72 - .75	.72
Fakten lernen	.62 - .70	.68	.68 - .73	.70
Diagramme und Tabellen	.67 - .71	.67	.68 - .72	.66

Tabelle 40: Zuverlässigkeit; 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 **0,92** in einem hohen Bereich. Hier liegt der Vergleichswert für Deutschland zwischen 0,91 und 0,93 (gemessen für rund 3000 Teilnehmer).

Die nachfolgende für StatistikerInnen interessante Tabelle 41 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 .83.

Untertest	Konsistenz ohne den jeweiligen Untertest
Muster zuordnen	.82
Med.-nat. Grundverständnis	.79
Schlauchfiguren	.80
Quant. u. Form. Probleme	.80
Textverständnis	.80
Figuren lernen	.81
Fakten lernen	.82
Diagramme und Tabellen	.80
Konzentr. und sorgf. Arbeiten	.82

Tabelle 41: Innere Konsistenzen, wenn der jeweilige Untertest weggelassen worden wäre; Die Konsistenz (standardized alpha) des Testprofils beträgt 0,83.

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:

		Punktwert	Muster zuordnen	Med.-naturwiss. Grundverständnis	Schlauchfiguren	Quant.u. formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagramme u. Tabellen	Konzentr.u. sorgf. Arbeiten
Punktwert	R	1.000	.578 .000	.733 .000	.692 .000	.690 .000	.699 .000	.650 .000	.571 .000	.692 .000	.523 .000
Muster zuordnen	R	.578 .000	1.000	.249 .000	.404 .000	.268 .000	.236 .000	.383 .000	.217 .000	.287 .000	.401 .000
Med.-naturwiss. Grundverständnis	R	.733 .000	.249 .000	1.000	.426 .000	.522 .000	.625 .000	.305 .000	.334 .000	.592 .000	.217 .000
Schlauchfiguren	R	.692 .000	.404 .000	.426 .000	1.000	.420 .000	.359 .000	.408 .000	.275 .000	.407 .000	.338 .000
Quant.u. formale Probleme	R	.690 .000	.268 .000	.522 .000	.420 .000	1.000	.495 .000	.297 .000	.201 .000	.588 .000	.156 .000
Textverständnis	R	.699 .000	.236 .000	.625 .000	.359 .000	.495 .000	1.000	.302 .000	.308 .000	.549 .000	.182 .000
Figuren lernen	R	.650 .000	.383 .000	.305 .000	.408 .000	.297 .000	.302 .000	1.000	.414 .000	.281 .000	.368 .000
Fakten lernen	R	.571 .000	.217 .000	.334 .000	.275 .000	.201 .000	.308 .000	.414 .000	1.000	.226 .000	.341 .000
Diagramme und Tabellen	R	.692 .000	.287 .000	.592 .000	.407 .000	.588 .000	.549 .000	.281 .000	.226 .000	1.000	.106 .004
Konzentr. sorgf. Arbeiten	R	.523 .000	.401 .000	.217 .000	.338 .000	.156 .000	.182 .000	.368 .000	.341 .000	.106 .004	1.000

Tabelle 42: 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 42,5% an der aufgeklärten Varianz ist er sehr bedeutsam.
- 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 15% Varianz mehr aufgeklärt als bei der Ein-Faktorenlösung.

- Die Drei-Faktorenlösung entspricht dem aus Deutschland bekannten Ergebnis, wobei 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 7 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 %
1	3.829	42.546	42.546	Zweifaktorenlösung		
2	1.380	15.330	57.876	2.865	31.831	31.831
3	.856	9.510	67.386	2.344	26.045	57.876
4	.627	6.969	74.354	Dreifaktorenlösung		
5	.564	6.269	80.623	2.807	31.191	31.191
6	.511	5.677	86.300	1.858	20.644	51.835
7	.489	5.436	91.737	1.399	15.550	67.385
8	.392	4.355	96.092			
9	.352	3.908	100.000			

Tabelle 43: Varianzaufklärung der Faktoren: unrotierte Lösung (links) und Zwei- bzw. Dreifaktorenlösung (rechts)

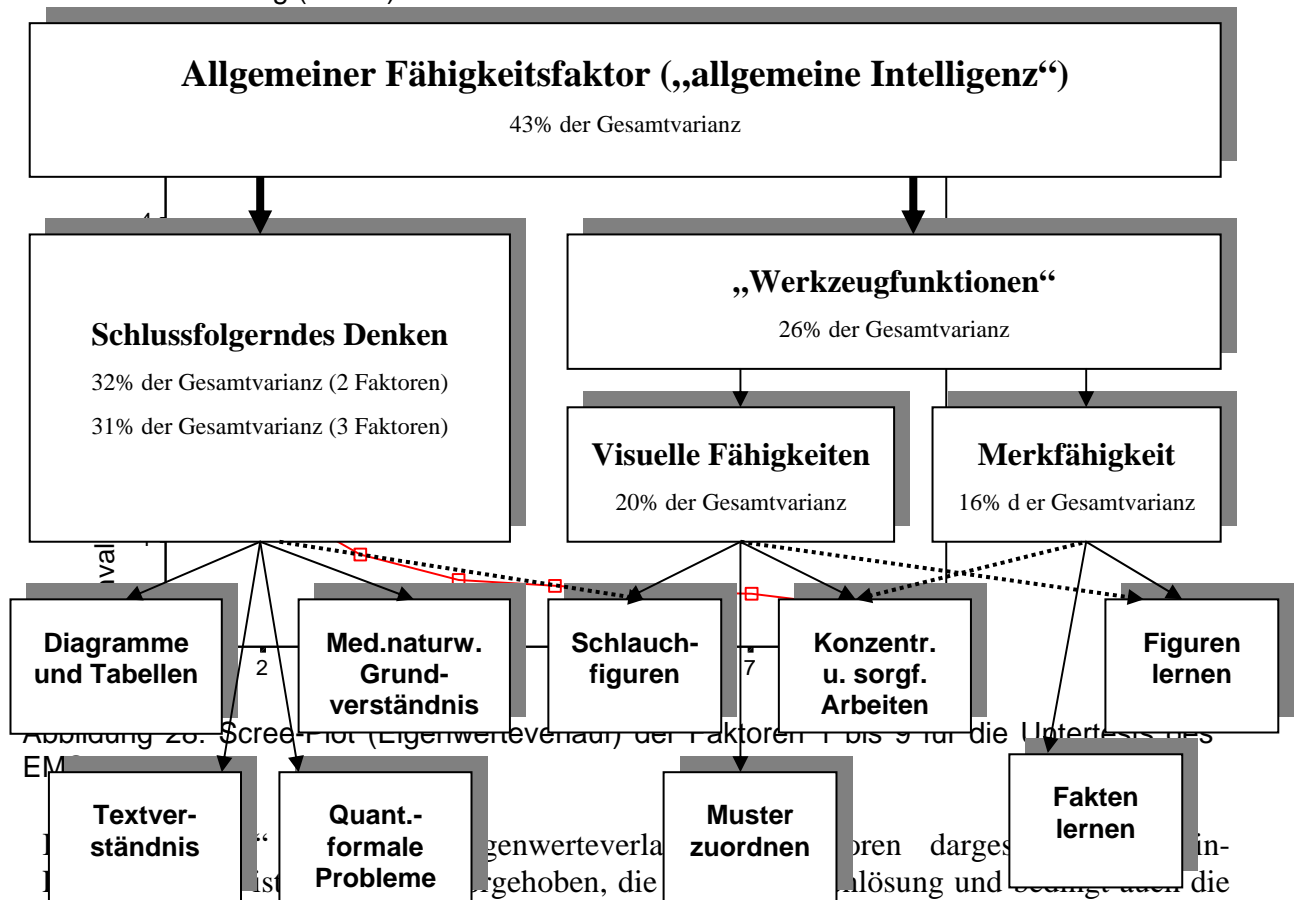


Abbildung 29: Struktur der Untertests des EMS, hierarchische Faktorenstruktur

kleiner als der eines einzelnen Untertests. Die Abbildung 29 zur Struktur der Untertests fasst die Ergebnisse der drei nachfolgenden Analysen zusammen. Die Ladungsmatrix des ersten Faktors der unrotierten Lösung zeigt substantielle Ladungen für alle Untertests. Ein „Generalfaktormodell“ ist indiziert. Der allgemeine Fähigkeitsfaktor ist sehr varianzstark.

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

Tabelle 44: Ladungen im ersten Faktor der unrotierten Lösung („Generalfaktor“)

	1	2	Kommun.
Diagramme und Tabellen	.83		.71
Med.-naturwiss. Grundverständnis	.79	.22	.68
Textverständnis	.77		.64
Quant.-formale Probleme	.77		.62
Konzentr.u.sorgf. Arbeiten		.78	.62
Figuren lernen	.23	.70	.55
Muster zuordnen		.66	.48
Fakten lernen	.20	.59	.40
Schlauchfiguren	.45	.54	.50

Tabelle 45: 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.

	Schweiz Gesamt				Deutschland		
	1	2	3	h ²	1	2	3
Diagramme und Tabellen	.83			.71	.82		
Medizin.-naturwiss. Grundverständnis	.78		.27	.71	.81		.20

Quantitat. u.formale Probleme	.77	.23		.65	.80	.18	
Textverständnis	.76		.27	.67	.79		.20
Muster zuordnen		.83		.72		.81	
Konzentr. u. sorgfält. Arbeiten		.65	.43	.62		.70	.40
Schlauchfiguren	.44	.60		.58	.35	.71	
Fakten lernen			.88	.83	.21		.87
Figuren lernen	.21	.50	.51	.56	.13	.47	.64

Tabelle 46: Faktorenanalyse: Varimaxrotierte Dreifaktorenlösung, Schweiz und Deutschland (deutsche Daten nach Blum 1996, in: Trost et al. 1998, S. 42)

Zwischen den Sprachgruppen haben sich Unterschiede gezeigt, die sprachliche und nicht-sprachliche Leistungen betreffen. Es wurden Belege für die Hypothese gefunden, dass es kulturelle Unterschiede im Leistungsbereich zwischen den Sprachgruppen gibt. Zwischen den Geschlechtern gibt es Unterschiede in den Einzeltests – der Testwert unterscheidet sich nicht. Liegt dem Unterschied jeweils die gleiche Binnenstruktur der Untertests zugrunde? Es wurden jeweils Dreifaktorenlösungen berechnet. Bei der französisch- und italienischsprachigen Gruppe macht sich die geringe Stichprobengröße bemerkbar - weil nur 9 Variablen verwendet werden, ist die Faktorenanalyse aber formal ohne Einschränkungen anwendbar.

	Männer				Frauen			
	1	2	3		1	2	3	
Diagramme und Tabellen	.812	.261		.73	.837			.72
Med.naturw.Grundv.	.796		.247	.70	.788		.255	.70
Textver-ständnis	.782		.283	.70	.736		.308	.64
Quant.u.for-mal.Probleme	.733	.233		.60	.773	.272		.67
Muster zuordnen	.228	.849		.77		.792		.65
Konzentr. u. sorgf. Arbeit.		.648	.495	.67		.717	.239	.57
Schlauchfiguren	.413	.605	.164	.56	.410	.625		.57
Fakten lernen	.235		.870	.82	.179		.895	.85
Figuren lernen	.180	.519	.526	.58	.253	.495	.498	.56
Eigenwerte Varianz in %	2,75	1,92	1,46	68,2%	2,74	1,90	1,30	66,0%

Tabelle 47: Faktorenanalyse: varimaxrotierte Dreifaktorenlösung für Männer und Frauen

Die Ladungsmuster und Varianzaufklärungen unterscheiden sich **nicht zwischen den Geschlechtern**, es zeigt sich eine übereinstimmende Struktur der Untertests. Die Verwendung der gleichen Berechnungsgrundlagen für den Testwert ist gerechtfertigt.

Die Ladungsmuster der **französisch- und italienischsprachigen Gruppe** weichen demgegenüber von der deutschsprachigen in einzelnen Variablen ab. „Textverständnis“ lädt in der französischen Gruppe mit „Fakten lernen“ und „Muster zuordnen“ in einem Faktor, der dadurch kein reiner Gedächtnisfaktor ist. Figuren lernen lädt in einem etwas veränderten Faktor „Schlussfolgerndes Denken“. Auch in der italienischen Sprachgruppe sind die Faktoren 2 und 3 gegenüber der deutschen Version verändert. Dieser Nachweis qualitativer

struktureller Unterschiede ist eine weitere Begründung für eine separate Standardisierung der Sprachgruppen (Tabelle 48).

	Deutsch				Französisch				Italienisch			
	1	2	3	h ²	1	2	3	h ²	1	2	3	h ²
Diagr. u. Tabellen	<u>.83</u>	.15		.72	<u>.82</u>			.71	<u>.83</u>	-.15	.29	.80
Med.naturw.Grundv.	<u>.79</u>		.25	.70	<u>.83</u>		.24	.77	<u>.63</u>	.39	-.19	.59
Textverständnis	<u>.78</u>		.25	.67	.28		<u>.79</u>	.71	<u>.79</u>	.17		.67
Quant.u.formal.Probl.	<u>.77</u>	.24		.66	<u>.76</u>		.16	.63	<u>.78</u>	.17		.65
Muster zuordnen	.17	<u>.83</u>		.72	.26	<u>.51</u>	<u>.55</u>	.64			<u>.80</u>	.67
Konzr. u. sorgf. Arbeit.		<u>.65</u>	.44	.62		<u>.91</u>		.84		<u>.84</u>	.20	.76
Schlauch-figuren	.44	<u>.60</u>		.57	.48	<u>.67</u>	.15	.71	.37	<u>.65</u>	.32	.68
Fakten lernen	.17		<u>.89</u>	.83		.21	<u>.82</u>	.72	.50	<u>.72</u>		.79
Figuren lernen	.22	.47	<u>.56</u>	.59	<u>.48</u>	.38	.37	.52	-.16	.16	<u>.75</u>	.62
Eigenw. Varianz in %	2,84	1,81	<u>1,44</u>	68%	<u>2,78</u>	1,72	1,85	71%	2,79	1,94	<u>1,50</u>	69%

Tabelle 48: Faktorenanalyse: varimaxrotierte Drei-Faktorenlösung für Sprachgruppen: unterstrichen: erwartet aufgrund der Struktur in der deutschsprachigen Gruppe

Item-Trennschärfen

Als weitere Vergleichsmöglichkeit können die Trennschärfen der Items herangezogen werden. Es handelt sich dabei um die Korrelationen der Items mit dem zugehörigen Untertest-Wert. Hier soll wegen der geringen Stichprobengrößen in der französisch- und italienischsprachigen Gruppe nur die Originalversion im Vergleich zur deutschen Version betrachtet werden (Abbildungen 30 bis 37).

Dargestellt sind die Items geordnet nach der Schweizer Version (nach Trennschärfen in absteigender Reihenfolge). Inhaltlich sind unterschiedliche Trennschärfen schwieriger zu interpretieren als Differenzen in der Itemschwierigkeit. Auch sie liefern aber Hinweise für unterschiedliche Itemfunktionen zwischen den Sprachversionen.

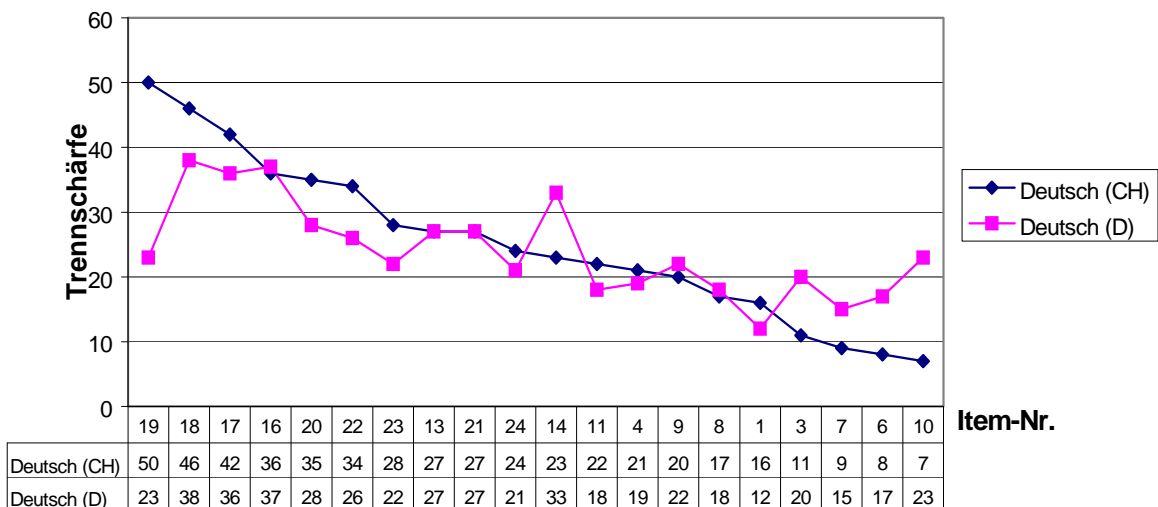


Abbildung 30: Trennschärfen für den Untertest "Muster zuordnen"

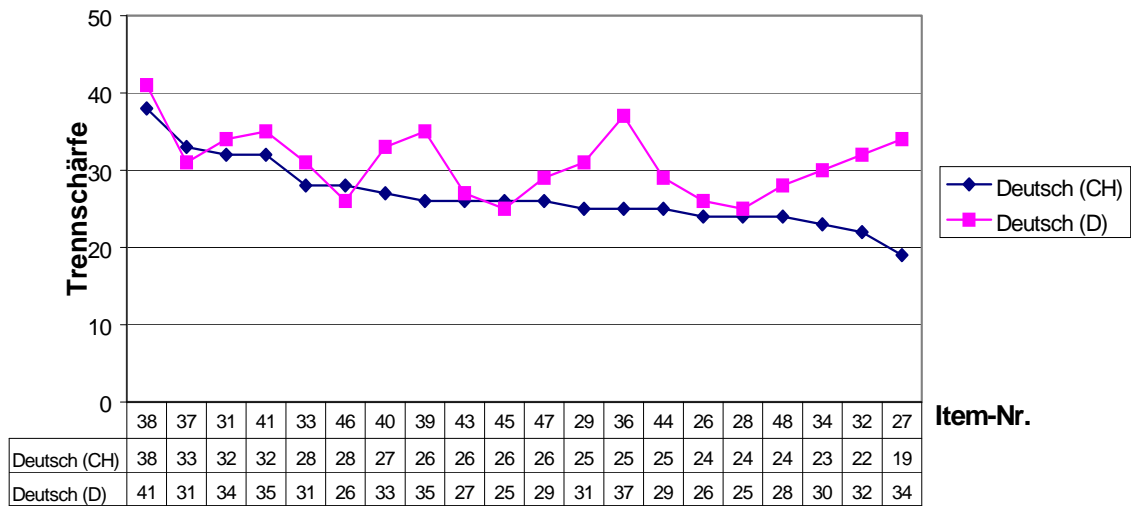


Abbildung 31: Trennschärpen für den Untertest "Medizinisch-naturwissenschaftliches Grundverständnis"

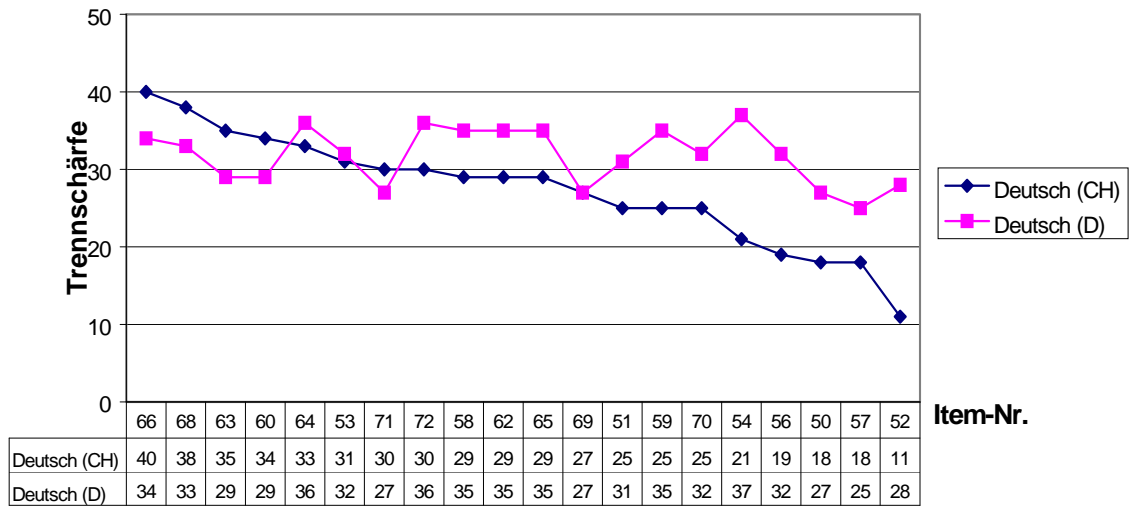


Abbildung 32: Trennschärpen für den Untertest "Schlauchfiguren"

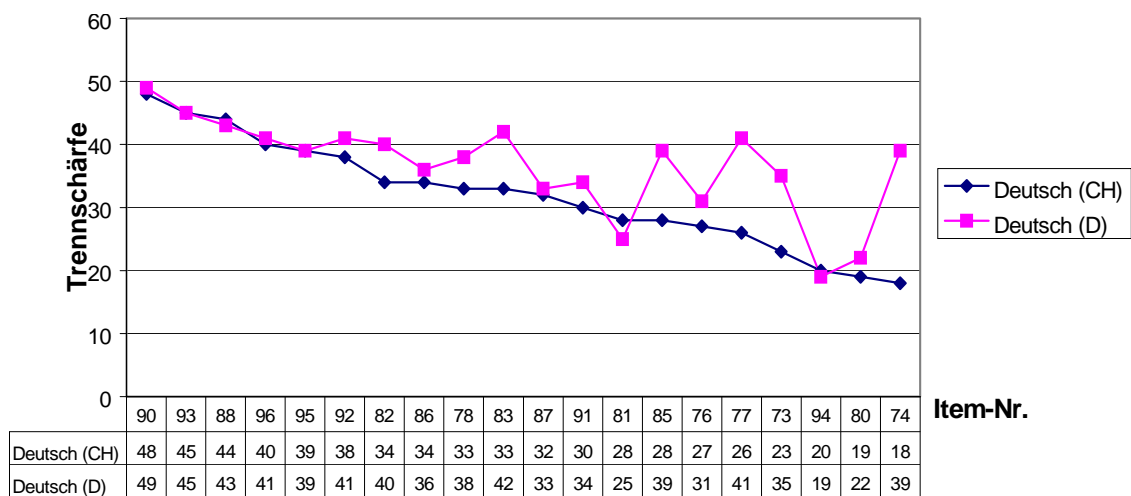


Abbildung 33: Trennschärpen für den Untertest "Quantitative und formale Probleme"

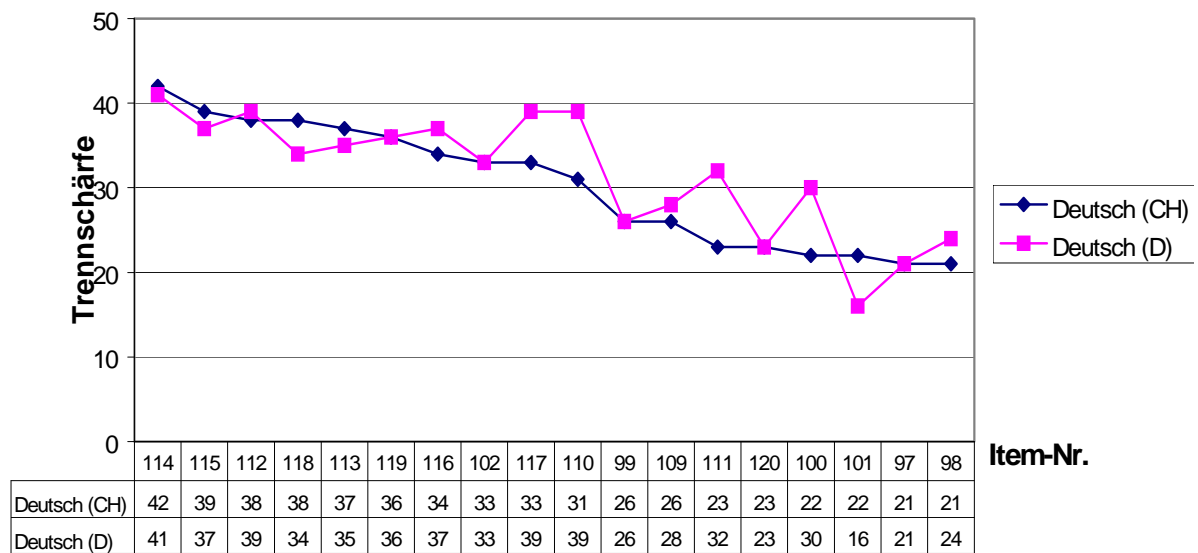
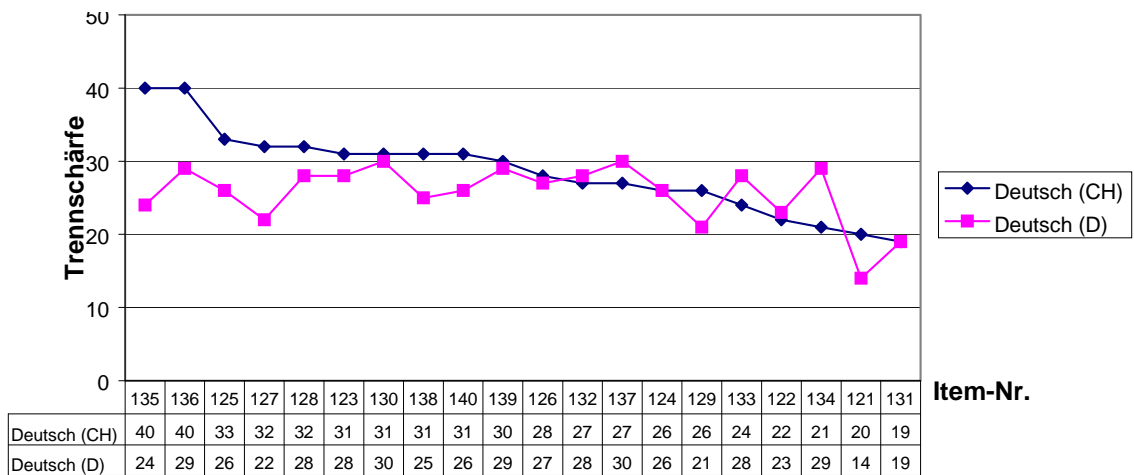


Abbildung 34: Trennschärfe für den Untertest "Textverständnis"

Abbildung 35: Trennschärfe für den Untertest "Figuren lernen"

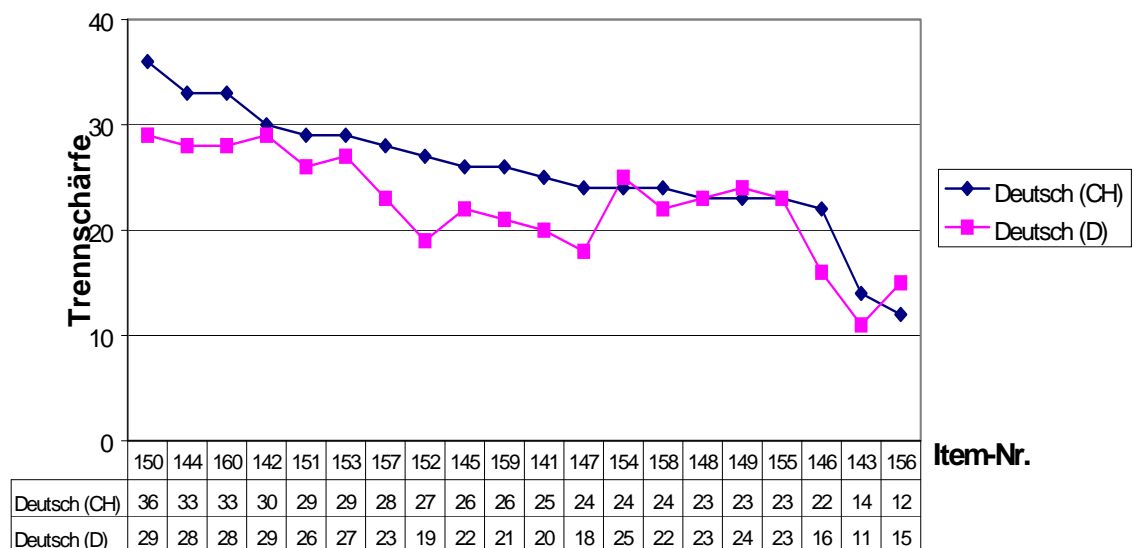


Abbildung 36: Trennschärfen für den Untertest "Fakten lernen"

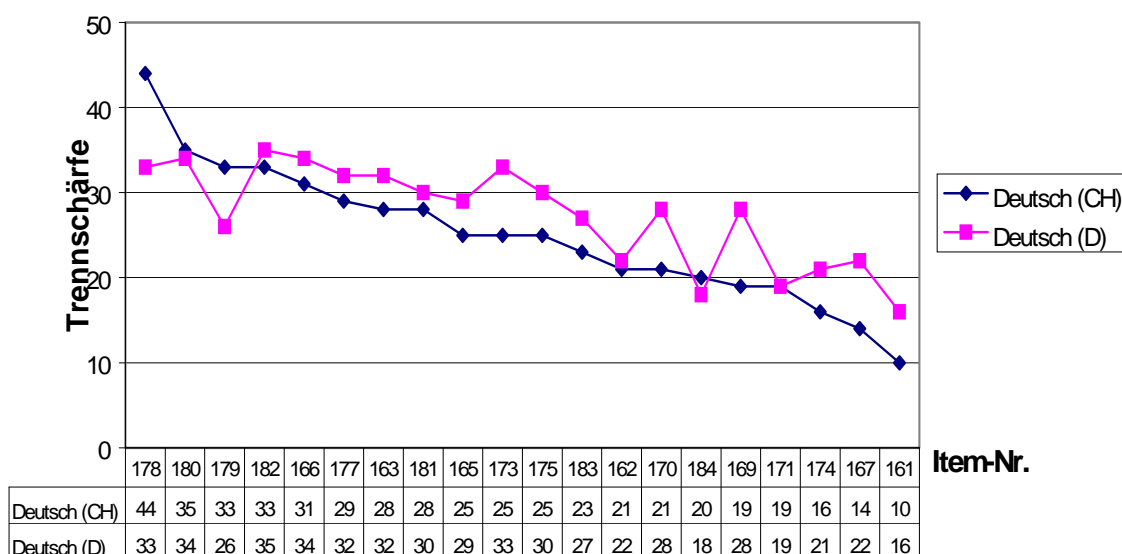


Abbildung 37: Trennschärfen für den Untertest "Diagramme und Tabellen"

Die Darstellungen der Trennschärfekoeffizienten wurden nach der Höhe in der schweizerdeutschen EMS-Version sortiert. Bei einzelnen Items zeigen sich deutliche Abweichungen, die auch für eine unterschiedliches Lösungsverhalten sprechen können. Dennoch liegen die meisten Trennschärfen innerhalb der Bandbreite, die vom deutschen TMS vorgegeben worden sind.

Konzentriertes und sorgfältiges Arbeiten

Hier kann die Zuverlässigkeit nicht nach dem Itemkonzept eingeschätzt werden. Bei diesem Untertest wird die Menge der Richtigen als Differenz bearbeitete Zeichen – 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.

	deutsch		französisch		italienisch	
	M	S	M	S	M	S
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
... übersehen	8.61	16.15	10.15	14.74	5.48	5.95
... fälschlich durchgestrichen	3.16	3.42	3.67	6.47	3.06	5.11

Tabelle 49: Kennwerte für den Untertest „Konzentriertes und sorgfältiges Arbeiten“

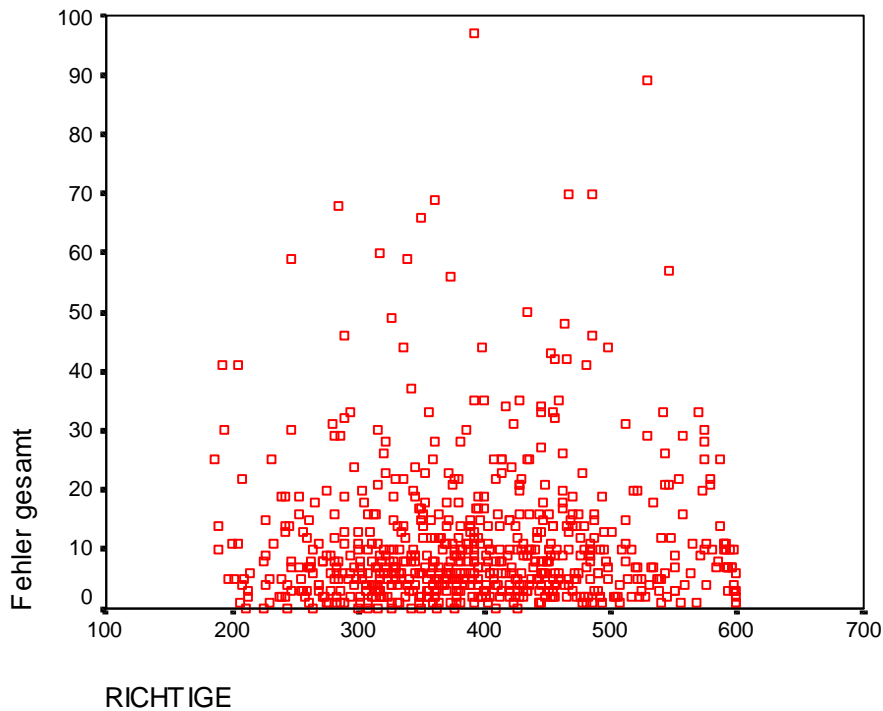


Abbildung 38: Summe der Richtigen im Verhältnis zu den Fehlern

	RICHTIG	FALSCH	übersehen	fälschlich durchgestrichen
RICHTIG	1.000 .	.02 .588	-.02 .628	.16 .000
FALSCH	.02 .588	1.000 .	.97 .000	.35 .000
... übersehen	-.02 .628	.97 .000	1.000 .	.13 .000
... fälschlich durchgestrichen	.16 .000	.35 .000	.13 .000	1.000 .

Tabelle 50: Interkorrelationen der Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“ und Signifikanzniveaus (jeweils darunter)

Menge und Fehler korrelieren nicht signifikant (Tabelle 50). Auch die beiden Fehlertypen korrelieren mit .13 nur schwach. Fälschliches Durchstreichen kommt dabei wesentlich seltener vor als das Übersehen von Zeichen – ein aus anderen Durchstreichtests bekannter Befund für Normalpersonen.

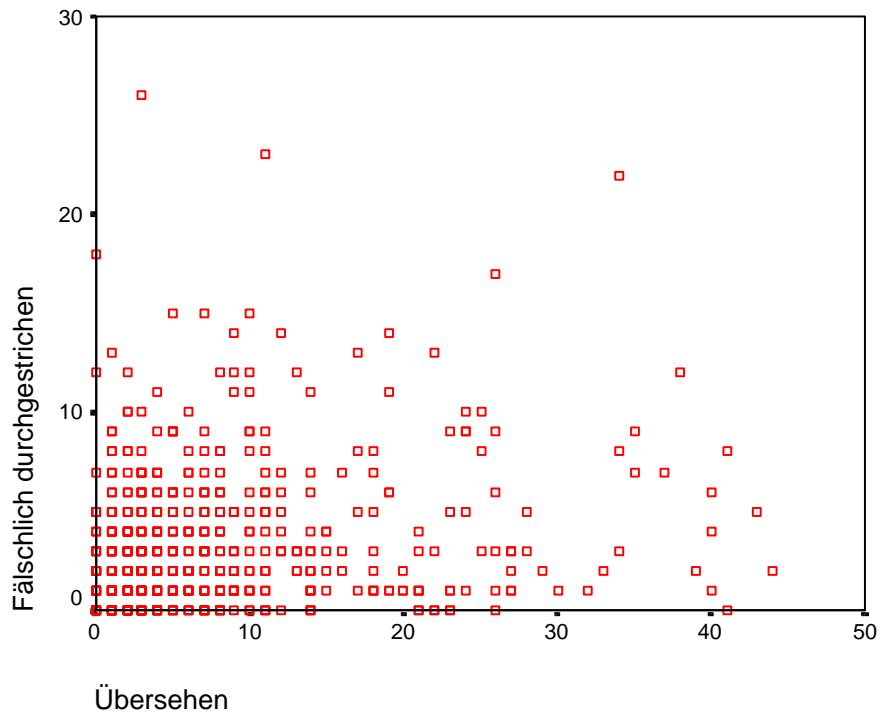


Abbildung 39: Fehlerarten - fälschlich durchgestrichene im Verhältnis zu übersehenen Zeichen

Typenanalyse

Der Frage, ob unterschiedliche Leistungs-Profile identifizierbar sind, soll abschliessend nachgegangen werden. Bleiben diese zukünftig stabil, könnten damit genauere Prognosemodelle für Studienerfolg überprüft werden. Es ist wahrscheinlich, dass die alleinige Betrachtung von Mittelwerten für diesen Zweck zu undifferenziert ist, und es muss auch kein allgemeingültiges „Muster“ der Leistungen geben, welches für alle Personen gleichermaßen gilt. Bei der Bewertung jedes Merkmals ist dann auch entscheidend, in welcher Beziehung die anderen Merkmale zu ihm stehen (z.B. ob die gleiche logische Denkfähigkeit mit einer entsprechenden Wahrnehmungs- oder Gedächtnisleistung verbunden ist oder nicht).

Hier soll zunächst nur gezeigt werden, dass sich verschiedene gut interpretierbare „Muster“ der Merkmale tatsächlich identifizieren lassen. Für die eigentliche Bewertung der Aussagekraft dieser Klassifikation fehlen noch Prognosedaten, die sich aus dem Studienverlauf ergeben werden.

Mittels hierarchischer Clusteranalyse (Euklidische Distanzen, z-Transformation der Untertest-Variablen, Methode von WARD) wurden verschiedene Klassifikationen geprüft. Die 6-Klassen-Lösung ist in allen Clustern ausreichend besetzt, auch vom Fusionskriterium (3938) her bietet sie sich an (weniger Änderung zur 7-Klassenlösung mit 3821). Ein grösserer Sprung besteht zur 5-Klassen-Lösung (4218). Die Untertests sind nach den Bereichen der Dreifaktoren-Lösung geordnet.

Nr	n	Testwert	männl.	weibl.	deutsch	französisch	italienisch
6	90	135.8	11.3%	12.6%	12.6%	7.7%	3.2%
2	114	124.4	21.4%	10.1%	16.0%	5.1%	9.7%
4	175	118.1	18.5%	27.3%	24.4%	12.8%	12.9%
5	97	101.9	19.6%	7.5%	12.1%	25.6%	16.1%
1	120	101.2	10.7%	20.3%	15.3%	23.1%	22.6%
3	154	81.5	18.5%	22.2%	19.6%	25.6%	35.5%
Ges	750	108.9	100.0%	100.0%	100.0%	100.0%	100.0%

Tabelle 51: Testwert und Zusammensetzung der 6-Cluster-Lösung

Ordnet man die 6 Cluster nach dem Testwert (Tabelle 51), zeigt sich zunächst eine Ordnung nach der Profilhöhe bzw. Gesamtleistung. Dies gilt nicht für den Unterschied zwischen Cluster 5 und 1, die sich nur bezüglich der Profilverläufe unterscheiden – die gleiche Leistung kommt hier also auf unterschiedlichem Wege zustande. Eine eindeutige Dominanz von Geschlechtern oder Sprachgruppen in einem Cluster zeigt sich nicht – wenngleich es Regelmäßigkeiten zu geben scheint. In Cluster 2, 4 und 5 dominiert je ein Geschlecht. Die auf der Mittelwertebene gefundenen Belege, dass die Leistungsprofile von Männern und Frauen nicht gleich sein müssen, werden erhärtet. Es gibt allerdings auch nicht **das** geschlechtertypische Profil – es ist vielmehr eine **Summe aus vielen unterschiedlichen Profilen**. Bei den Sprachgruppen macht sich die geringe Stichprobengrösse bemerkbar.

Das eigentlich Interessante sind die Mittelwertprofile, die in Abbildung 40 dargestellt sind. Je zwei Profile (1 und 3, 4 und 5, 2 und 6) zeigen im Faktor „logisches Denken“ (erste vier

Untertests) annähernd vergleichbare Verläufe. Ihr Leistungsunterschied bestimmt sich aus den Untertests des Wahrnehmungs- und Gedächtnisfaktors. Dort sind die Profilverläufe nicht parallel, sondern sie folgen ganz verschiedenen Mustern. Die weitere Aufklärung von Leistungsprofilen scheint wichtig, wenn eine Evaluation des Tests ansteht.

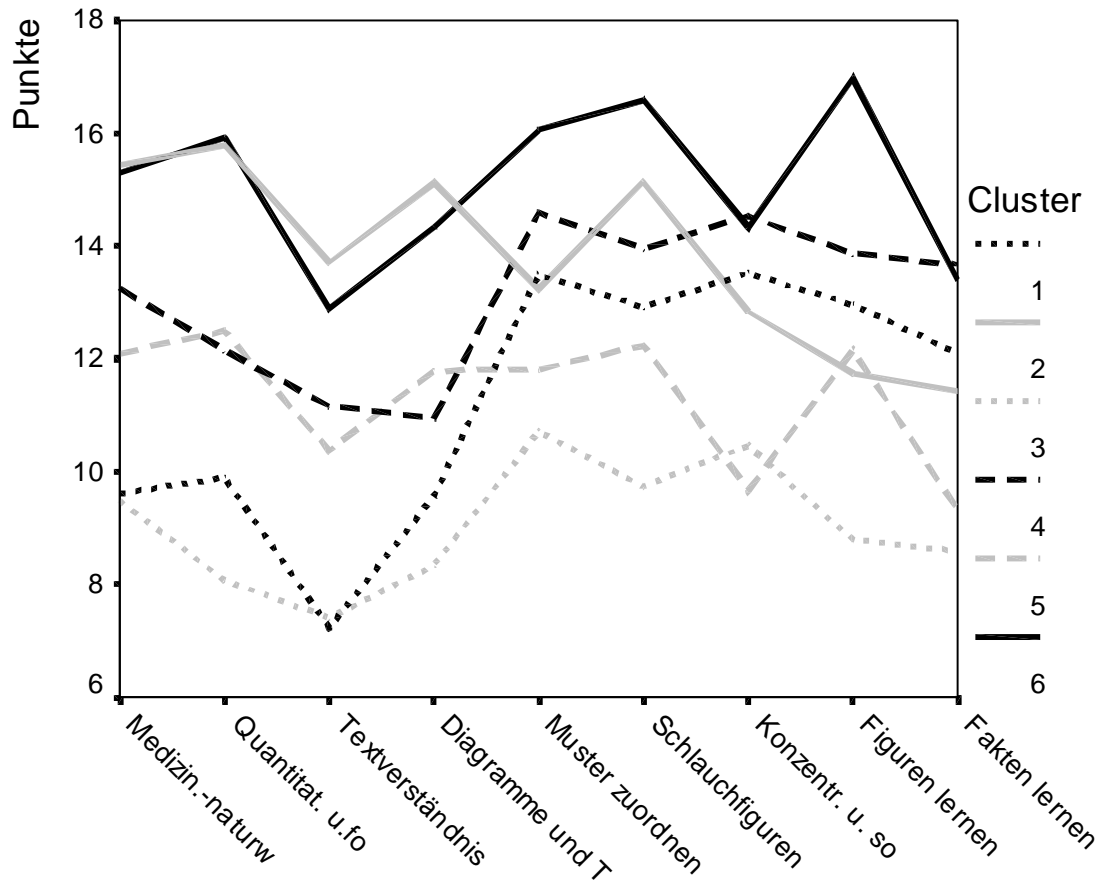


Abbildung 40: Clustermittelwertprofile der Untertests für die 6-Cluster-Lösung

Genauigkeit der Belegerfassung

Trotz ausführlicher Instruktionen ist immer damit zu rechnen, dass Probanden bei der Beantwortung von Fragebogen Fehler unterlaufen, oder dass Störungen anderer Art auftreten, welche ein korrektes maschinelles Einlesen erschweren. Beim Einlesen der Ergebnisse muss gewährleistet sein, dass möglichst keine Fehler auftreten. Dies ist vor allem durch mehrfaches Einlesen und eine Kombination von logischer Datenprüfung und visueller Nachkontrolle der kritischen Fälle zu erreichen.

Wie arbeitet die Datenerfassung? Das Lesegerät prüft, ob der Kontrast zwischen der Markierung im Antwortfeld und der Umgebung einen bestimmten Wert übersteigt. Ist dies gegeben, zählt dies als Antwort. Es besteht die Möglichkeit, den notwendigen Kontrast beim Einlesen zu variieren. Mögliche Fehlerquellen auf Probandenseite können sein:

- zu dünne Markierungen
- zu kurze Markierungen
- zu schwache Markierungen
- Verschmutzungen (z.B. Radierrückstände nach Korrekturen, verschmierte Markierungen, Verschmutzung durch Speisen und Getränke,...)
- Mehrfachmarkierungen (tatsächliche oder aufgrund ungenügenden Radierens)

Bezüglich unkorrekter Markierungen und den Folgen ungenügenden Radierens wurden die Testteilnehmer bereits im Vorfeld durch die "Test-Info" auf mögliche Fehlerquellen aufmerksam gemacht. Es wurde darauf hingewiesen, die Lösungsfelder deutlich zu markieren, und eventuelle Radiespuren möglichst sauber zu entfernen. Dennoch musste angenommen werden, dass etwa hinsichtlich der Stärke der Markierungen grosse Differenzen zwischen den Testteilnehmern bestehen. Ein Einlesen mit sehr hoher Empfindlichkeit hätte wohl die schwachen Markierungen erfasst, gleichzeitig aber Personen benachteiligt, die stark markieren und möglicherweise Korrekturen anbringen würden. Versuche im Vorfeld haben gezeigt, dass starke Markierungen nach Radieren bei hoher Empfindlichkeit des Lesegerätes meist noch als Markierung erkannt werden. Andererseits hätte ein Einlesevorgang bei hoher Empfindlichkeit alle schwächeren Markierungen übersehen, und somit (wie sich zeigte) einen Grossteil der Teilnehmer unkorrekt erfasst.

Die Verwendung einer mittleren Einstellung allein wäre eine unbefriedigende Lösung, weil damit die beschriebenen Probleme nicht gelöst wären, sondern sich die Fehler beim Einlesen nur anders verteilt hätten. Anstatt entweder nur "Stark-" oder "Schwachmarkierer" zu benachteiligen, wären hier einfach die Extreme beider Ausprägungen unkorrekt beurteilt worden, was bei Fragestellungen und Entscheidungen von der Tragweite einer Zulassung zu einer gewünschten Studienrichtung nicht vertretbar ist.

Aus den obengenannten Gründen wurde entschieden, ein dreistufiges Einleseverfahren zu verwenden. Die Fragebögen sollten mit drei unterschiedlichen Einleseempfindlichkeiten erfasst, und die Daten mit Hilfe eines eigens entwickelten Auswerteprogrammes aggregiert werden. Dabei galt der Grundsatz "im Zweifel zu Gunsten des Probanden".

Mit Hilfe "simulierter" Antwortbögen wurden drei Empfindlichkeiten des Lesers (Thresholds) bestimmt. Dabei mussten auf der niedrigsten Einstellung (entspricht der höchsten Empfindlichkeit) auch schwächste Markierungen noch eingelesen werden. Als

mittlere Einstellung wurde die vom Hersteller empfohlene Konfiguration gewählt. Auf der höchsten Einstellung durften auch deutliche Raderspuren nicht als Markierungen eingelesen werden.

Als Mehrfachantworten eingelesene Beantwortungen können speziell auf der niedrigsten Einleseempfindlichkeit auch von Raderspuren oder sonstigen kleinen Verschmutzungen stammen. Um diese nicht unbesehen als "falsch" zu bewerten, wurden die Itembeantwortungen über eine binäre Summierung erfasst. Diese hat den Vorteil, dass jede eingelesene Zahl genau einer Antwortkonfiguration zugeordnet werden kann, und damit auch eindeutig ersichtlich ist, ob eine eventuelle Mehrfachantwort die korrekte Lösung beinhaltet oder nicht. Für jede Itembeantwortung wurde eine Zahl zwischen 1 und 31 eingelesen, wobei die Zahlen 1, 2, 4, 8, 16 für die Markierung von jeweils einer Antwortkategorie stehen, bei allen anderen Zahlen handelt es sich um "Mehrfachantworten" (unbeantwortete Items wurden mit 0 kodiert).

Bereits vor dem Einlesevorgang wurde auf allen Antwortbögen überprüft, ob die personenspezifische EMS-Nummer ausreichend deutlich markiert war, um vom Scanner eingelesen zu werden. Fehlerhaft oder undeutlich markierte Nummern wurden nachbearbeitet. Die ausgefüllten Bogen wurden für die Testlokale pro Testteil je dreimal mit den verschiedenen Thresholds eingelesen. Dabei wurden zusätzlich Testort und Testlokal über den Dateinamen kodiert.

Während des Einlesevorgangs wurde kontrolliert, ob alle EMS-Nummern vollständig durch den Scanner erfasst wurden. Die eingelesenen Daten erscheinen während des Prozesses auf dem Bildschirm. Fehlende Ziffern oder gänzlich fehlende EMS-Nummern sind dabei leicht auszumachen und das entsprechende Antwortblatt kann zur Nachbearbeitung der Daten aussortiert werden. Durch die Kontrolle und Korrektur vor dem Einlesen konnte die Anzahl derartiger Fehler relativ klein gehalten werden. Der häufigste Fehler betraf speziell auf der höchsten Empfindlichkeitsstufe einzelne, nicht erfasste Ziffern. Unvollständig eingelesene EMS-Nummern wurden so testlokalweise korrigiert.

Einlesen der Teile A und B (Multiple Choice)

Bei der Überführung sämtlicher Dateien (geordnet nach Testlokalen, Testteilen und Einlestufen) in eine Gesamtdatei wurden die Personen nach ihren EMS-Nummern sortiert. Eine Variable prüfte die Übereinstimmung aller Codes pro Person. Es zeigte sich, dass einigen wenigen Personen beim Übertragen ihrer EMS-Nummer Fehler unterlaufen waren. Die korrekten Nummern konnten über die Namens- und Testlokalliste ermittelt werden. Der nächste Schritt bestand in dem Vergleich zwischen eingelesener Antwort (pro Einlestufe) und dem Lösungsschlüssel. Dabei wurde jedes Item als R, S, M oder F bewertet, wobei:

R: Richtige Lösung

S: Mehrfachantwort, welche aber das korrekte Item enthält (Summe)

M: Missing

F: Falsche Antwort. Dabei handelt es sich um (nur eine) falsche Lösung, oder um eine Mehrfachantwort, welche die korrekte Lösung nicht beinhaltet

Aus den drei Bewertungen jedes Items wurde entschieden, ob das Item als korrekt oder falsch gewertet, oder ob es manuell überprüft werden musste. Dazu wurden die Bewertungen itemweise gezählt und die Summen als Variablen CR, CS, CM und CF

gekennzeichnet (C=count). Der Tabelle 51 sind die Entscheidungen aufgrund der möglichen Beurteilungs-Kombinationen zu entnehmen. Als richtig wurden die Fälle bewertet, bei denen CR mindestens 2 beträgt - die in den drei Einlesestufen mindestens zweimal als korrekt beantwortet erfasst wurden. Bei dem Fall CR=1, CS=2 muss es sich um deutliche Verschmutzungen handeln, die aber auf der höchsten Einlesestufe ausgefiltert werden. Nach dem Prinzip "Im Zweifel zu Gunsten des Probanden" muss auch diese Kombination als richtige Antwort gelten. Ähnlich verhält es sich mit dem Fall CR=1, CM=2. Dabei handelt es sich um eine schwache (korrekte) Markierung, die nur auf der niedrigsten Einlesestufe erkannt wird.

CR	CS	CM	CF	Entscheidung	CR	CS	CM	CF	Entscheidung
3	0	0	0	<i>Richtig</i>	0	3	0	0	Prüfen
2	1	0	0	<i>Richtig</i>	0	2	1	0	Prüfen
2	0	1	0	<i>Richtig</i>	0	2	0	1	Prüfen
2	0	0	1	<i>Richtig</i> *	0	1	2	0	Prüfen
1	2	0	0	<i>Richtig</i>	0	1	1	1	Prüfen
1	1	1	0	Prüfen	0	1	0	2	Falsch
1	1	0	1	Prüfen *	0	0	2	1	Falsch
1	0	2	0	<i>Richtig</i>	0	0	1	2	Falsch
1	0	1	1	Prüfen *	0	0	0	3	Falsch
1	0	0	2	Prüfen *					

* Diese Kombinationen sind nur der Vollständigkeit halber aufgeführt, aus logischen Überlegungen aber nicht zu erwarten
 CR = Anzahl korrekt; CS = Anzahl Summe korrekt (Mehrfachantwort, aber richtige Lösung enthalten); CM = Anzahl Missings; CF = Anzahl falsch oder Mehrfachantwort ohne richtige Lösung)

Tabelle 52: Itembeurteilungen und Entscheidung

In den Tabellen 52 und 53 sind die Häufigkeiten der zu überprüfenden Items für die Testteile A und B dargestellt. In 77%, bzw. 72% der Fälle konnten eindeutige Entscheidungen erreicht werden. Insgesamt mussten 459 Itembeantwortungen (von total 138'000 eingelesenen) überprüft werden, was einem Anteil von rund 0.3% entspricht. Da in der Mehrzahl der Fälle aber nur ein Fehler pro Antwortblatt auftrat, mussten dennoch 384 Antwortblätter von 1'500 (Teil A und B) hinsichtlich eines oder mehrerer zu überprüfender Items manuell überarbeitet werden, was rund 25% entspricht.

Items zu prüfen	n	%	Kumulative %
0	577	76.9	76.9
1	140	18.7	95.6
2	28	3.7	99.3
3	5	.7	100.0
Total	750	100.0	

Tabelle 53: pro Person zu überprüfende Items (Teil A)

Items zu prüfen	n	%	Kumulative %
0	539	71.9	71.9
1	155	20.7	92.5
2	42	5.6	98.1
3	11	1.5	99.6
4	3	.4	100.0
Total	750	100.0	

Tabelle 54: pro Person zu überprüfende Items (Teil B)

Einlesen Teil K (Konzentriertes und sorgfältiges Arbeiten)

Im Rahmen des „Konzentrationstests“ konnte jeder Proband maximal 1'200 Items bearbeiten. Da bei diesem Test die Bearbeitungszeit ein wesentlicher Faktor war, wurden die Markierungen z.T. nicht ganz so sorgfältig angebracht wie in den Testteilen A und B. Aus diesem Grund wurden hier die Empfindlichkeiten des Lesegerätes verringert, so dass bei höchster Empfindlichkeit auch geringste Markierungen, und bei geringster Empfindlichkeit "durchschnittliche" Markierungen immer noch erkannt wurden. Das Problem von unsauberen Korrekturen konnte hier vernachlässigt werden, da aufgrund des Zeitdruckes kaum radiert wurde.

Die Antwortmöglichkeiten bestanden in diesem Testteil aus 0 (nicht angestrichen) oder 1 (angestrichen). In einem ersten Schritt wurden die eingelesenen Antworten summiert. Aus dieser Summe wurde die bereinigte Antwort abgeleitet.

Tabelle 54 zeigt die Aufschlüsselung der bereinigten Antworten aufgrund der Summierung der drei Einlestufen.

Summe aus drei Einlestufen	Korrekte Lösung (Lösungsschlüssel)	Bereinigte Antwort
0	Irrelevant	0
1	0	0
1	1	1
2	Irrelevant	1
3	Irrelevant	1

Tabelle 55 : Aggregation der Antworten im Konzentrationstest

Es zeigte sich, dass die höchste Empfindlichkeitsstufe anfällig für kleinste Verunreinigungen ist. Wird auf jeder Empfindlichkeitsstufe der Wert 0 eingelesen, so kann davon ausgegangen werden, dass kein Zeichen vorhanden ist. Wird nur auf der höchsten Empfindlichkeitsstufe der Wert 1 eingelesen, so ist nicht auszuschließen, dass es sich hierbei um eine "Phantommarkierung" handelt. Da die Auswertung nach dem Grundsatz "Im Zweifel zu Gunsten des Probanden" erfolgte, wurde in diesem Fall jeweils im Sinne des Antwortschlüssels entschieden. Diese möglicherweise kleine Ungenauigkeit fällt jedoch

durch die Art der Berechnung des Testwertes in diesem Testteil (siehe weiter unten) nicht ins Gewicht.

Für die Auswertung dieses Teils musste nun das letzte beantwortete Item bestimmt werden. Ein Vergleich zwischen den bereinigten Antworten (bis zum letzten bearbeiteten Zeichen) und dem Lösungsschlüssel ergibt für jedes Item folgende Beurteilungs-Möglichkeiten (Tabelle 55):

		Bereinigte Antwort	
		0	1
Richtige Lösung	0	Korrekt nicht markiert (d)	Fälschlicherweise markiert (c)
	1	Übersehen (b)	Korrekt markiert (a)

Tabelle 56: Beurteilung der bereinigten Antworten

Der maximale Testwert in diesem Untertest beträgt 20, insgesamt können maximal 600 Zeichen korrekt markiert, also 600 Punkte erreicht werden. Der Testwert für den Konzentrationstest berechnet sich demnach nach der Formel

$$t = \frac{a - (b + c)}{30}$$

Aus der Formel wird ersichtlich, dass in diesem Untertest gegenüber einzelnen unkorrekt erfassten Items bezüglich des Testwertes eine gewisse Toleranz besteht. Dies um so mehr, da dieser gerundet wird. Der kritische Punkt besteht hingegen hier in der Bestimmung des letzten beantworteten Items. Hat jemand tatsächlich den halben Test bearbeitet, es wird aber fälschlicherweise ein Item aus den hinteren Regionen als bearbeitet eingelesen, dann müssten alle dazwischenliegenden, fälschlicherweise nicht markierten Items, als "Übersehen" gewertet werden. Aus diesem Grund wurden jene Fälle manuell kontrolliert, in denen zwischen dem letzten und dem zweitletzten beantworteten Item mindestens zehn unbeantwortete lagen. Es zeigte sich, dass hier tatsächlich vereinzelt Markierungen erkannt wurden wo keine waren (auf der höchsten Empfindlichkeitsstufe). Die Anzahl dieser Fälle war jedoch gering, so dass davon ausgegangen werden kann, dass auch hier die bereinigten Daten die Probanden im Zweifelsfall eher begünstigen als benachteiligen.

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*, 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. Bericht 2* Fribourg: Zentrum für Testentwicklung und Diagnostik.
- 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*. Fribourg.
- Hänsgen, K.-D., Hofer, R., Ruefli, D. (Hrsg.) (1996) *Eignungsdiagnostik und Medizinstudium. Tagungsband. Berichte des ZTD Band 2*. Fribourg.
- 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
- Hofer, R., Ruefli, D., Hänsgen, K.-D.(1995). Der Eignungstest für das Medizinstudium in der Schweiz. Ein Probelauf. *Berichte des ZTD Band 1* Univ. Fribourg
- Hossiep, R. (1997) *Konsequenzen aus neueren Erkenntnissen zur Potentialbeurteilung. Vorabmanuskript (unveröffentlicht) Mai 1998*.
- Trost, G. (Hrsg.) (1994). *Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht)*. ITB: Bonn.
- Trost, G. (Hrsg.) (1995). *Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (19. Arbeitsbericht)*. ITB: Bonn.

Trost, G. (Hrsg.) (1996). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (20. Arbeitsbericht). ITB: Bonn.

Trost, G. (Hrsg.) (1997). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (21. Arbeitsbericht). ITB: Bonn.

Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. und Nauels, H.-U. (1998). Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse. ITB: Bonn

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

Originaltests zur Information und Vorbereitung

Es stehen in allen drei Sprachen vollständige Originalversionen zur Verfügung. Diese wurden bei früheren Testterminen in Deutschland verwendet. Vom ZTD wurden französisch- und italienischsprachige Adaptationen dieser 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 sind im Internet unter www.unifr.ch/ztd verfügbar. Es existiert je eine deutsch- und eine französischsprachige Fassung, falls der Text nicht in englischer Sprache verfasst worden ist.

Bericht 3 (1997) - Rapport 3 (1997)	
<p>Eignungsprüfung für das Medizinstudium -Kriterien und Testverfahren</p> <p><i>Bericht über das Internationale Symposium in Bern am 8. November 1996</i></p> <p>hrsg. von K.-D. Hänsgen und N. Ischi</p>	<p>L'examen d'aptitude aux études de médecine en Suisse - Critères et procédés d'application du test</p> <p><i>Rapport sur le Symposium international à Berne</i></p> <p>Editeurs: K.-D. Hänsgen; N. Ischi</p>
<p>Günter Trost</p> <p>Eignungskriterien bei der Zulassung zum Medizinstudium in Europa: Ergebnisse einer Erhebung in allen europäischen Ländern</p> <p>Ingemar Wedman & Widar Henriksson</p> <p>The Swedish Scholastic Aptitude Test. Research and main findings</p> <p>John L. Hackett</p> <p>The Medical College Admission Test (MCAT) - its use in U.S. and Canada and some results of validation studies</p> <p>Piet J. Janssen</p> <p>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</p> <p>Der Eignungstest für das Medizinstudium in der Schweiz als Instrument für die Beschränkung der Studienzulassung</p> <p>Rainer Hofer & Klaus-Dieter Hänsgen</p> <p>Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium in der Schweiz</p> <p>Urs Schallberger</p> <p>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

Verteilung der Punktwerte

Gesamt	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen						6	5	13	19	43	62	82	93	89	80	68	56	66	29	24	15
Med.-nat. Grundv.			1	2	5	9	17	26	51	53	68	68	66	88	76	87	54	40	21	13	5
Schlauchf.				1	2	13	7	19	36	34	56	71	81	69	71	85	58	59	44	29	15
Qu. u.form. Probleme		2		6	13	17	29	33	46	56	71	72	66	64	64	57	42	41	39	21	11
Text-verständn.		2	7	8	16	32	50	60	80	78	60	65	81	65	46	44	32	15	9		
Figuren lernen		1	1	5	4	8	23	24	38	55	78	71	83	58	58	60	56	51	38	27	11
Fakten lernen		1	2	5	14	11	28	40	55	77	72	77	80	63	64	53	52	23	22	7	4
Diagramme u. Tabellen			2	2	5	18	34	42	47	76	72	88	89	82	53	55	36	27	19	2	1
Konzentr. u. sorgfält. A.	1					5	8	25	38	51	73	76	100	89	79	66	46	28	28	21	16
Deutsch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen						5	3	12	18	42	55	68	80	85	74	63	48	64	26	22	15
Med.-nat. Grundv.			1	1	4	7	15	22	41	45	57	58	62	83	72	83	51	39	21	13	5
Schlauchf.				1	2	13	4	16	30	30	46	65	72	62	66	82	52	54	42	28	15
Qu. u.form. Probleme		2		5	13	17	27	27	43	48	64	67	59	61	53	53	38	39	36	18	10
Text-verständn.		2	6	7	11	28	46	51	75	65	56	60	75	57	45	43	30	15	8		
Figuren lernen			1	5	4	7	21	19	31	52	71	66	77	47	53	56	52	46	35	26	11
Fakten lernen		1	2	3	9	9	26	34	50	68	65	71	75	56	58	49	49	22	22	7	4
Diagramme u. Tabellen			2	2	5	15	28	37	42	67	67	78	80	74	46	55	34	26	19	2	1
Konzentr. u. sorgfält. A.	1					4	7	20	31	45	68	68	86	81	74	61	44	26	28	20	16

Tabelle A 1: Häufigkeitsverteilung der Punkte für die Untertests – Gesamt und deutschsprachige Gruppe

Französisch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen						1	2	1	1		3	8	5	4	5	2	4	1	2		
Med.-nat. Grundv.							1	2	5	5	6	7	3	2	2	3	2	1			
Schlauchf.							1	1	4	1	5	3	7	5	1	3	1	5	2		
Qu. u.form. Probleme							2	4	1	5	4	4	4	1	6	2	2	1	1	2	
Text-verständn.			1		2	3	2	4	3	7	2	5	5	2		1	2				
Figuren lernen		1					2	1	2	2	3	3	4	7	1	3	3	4	2	1	
Fakten lernen				2	3	1	1	3	4	5	1	2	3	6	3	3	2				
Diagramme u. Tabellen						2	2	2	3	4	3	5	7	4	5		2				
Konzentr. u. sorgfält. A.						1	1	3	5	5	2	2	9	3	2	2	1	2		1	
Italienisch	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Muster zuordnen										1	4	6	8		1	3	4	1	1	2	
Med.-nat. Grundv.				1	1	2	1	2	5	3	5	3	1	3	2	1	1				
Schlauchf.							2	2	2	3	5	3	2	2	4		5			1	
Qu. u.form. Probleme				1				2	2	3	3	1	3	2	5	2	2	1	2	1	1
Text-verständn.				1	3	1	2	5	2	6	2		1	6	1				1		
Figuren lernen						1		4	5	1	4	2	2	4	4	1	1	1	1		
Fakten lernen					2	1	1	3	1	4	6	4	2	1	3	1	1	1			
Diagramme u. Tabellen						1	4	3	2	5	2	5	2	4	2			1			
Konzentr. u. sorgfält. A.								2	2	1	3	6	5	5	3	3	1				

Tabelle A 2: Häufigkeitsverteilung der Punkte für die Untertests – französisch- und italienischsprachige Gruppe

	PUNKT-	TEST-	Muster	Mediz.-	Schlauch	Quantit.	Textver-	Figuren	Fakten	Diagram	Konz. u.
--	--------	-------	--------	---------	----------	----------	----------	---------	--------	---------	----------

	WERT	WERT	zuordn.	naturw. Grundv	figuren	u.form. Probl.	ständnis	lernen	lernen	me und Tabellen	sorgfält. Arbeiten
Sprachgruppe Deutsch (n=680)											
m	109.8	100.0	13.2	12.5	13.3	11.9	10.3	12.5	11.5	11.4	12.7
s	20.3	9.9	3.1	3.3	3.4	3.9	3.5	3.	3.5	3.3	3.2
Kurt.	-.176	-.123	-.479	-.406	-.418	-.589	-.613	-.554	-.426	-.458	-.172
Skew.	-.157	-.187	.002	-.243	-.258	-.101	-.035	-.096	-.036	-.055	.085
Sprachgruppe Französisch (n=39)											
m	100.8	100.0	12.2	10.9	12.3	11.7	9.2	12.4	10.0	10.7	11.1
s	19.5	10.0	3.1	2.7	3.2	3.5	3.3	3.9	3.8	2.8	3.2
Kurt.	.043	.086	.012	-.494	-.797	-.658	-.261	.542	-.999	-.569	-.303
Skew.	-.027	-.013	-.356	.477	.093	.306	.035	-.622	-.266	-.300	.354
Sprachgruppe Italienisch (n=31)											
m	98.5	100.0	13.1	9.5	11.5	12.4	9.0	11.0	10.1	9.9	11.7
s	17.0	10.0	2.8	3.2	3.3	3.9	3.5	3.3	3.2	2.9	2.3
Kurt.	-.513	-.494	-.720	-.421	-.702	-.283	-.159	-.786	-.274	-.443	-.372
Skew.	-.161	-.159	.656	-.013	.244	-.154	.408	.230	.051	.270	-.315
Gesamt (n=750)											
m	108.9	100.0	13.2	12.3	13.1	12.0	10.2	12.5	11.4	11.3	12.6
s	20.3	9.9	3.1	3.3	3.4	3.9	3.5	3.7	3.	3.3	3.2
Kurt.	-.200	-.140	-.442	-.454	-.489	-.587	-.608	-.517	-.426	-.465	-.177
Skew.	-.129	-.176	.003	-.196	-.215	-.086	-.011	-.107	-.054	-.036	.103

Tabelle A 3: Verteilungskennwerte für Punktwert, Testwert und die Untertests – gesamt und für Sprachgruppen

Punktwert			
Wert	deutsch	französisch	italienisch
41	1		
51	1		
52		1	
55	1		
57	1		
58	2		
60	1		
63	1		
64	1		1
65	1		
66	4		
67	1		
68	2	1	
69	1		
70	1		2
71	2		
72	2		
73	3		
74	2		
75	6	2	
76	4		1
77	3		
78	5		
79	5	1	1
80	3	1	1
81	1		
82	5		1
83	11	1	
84	5	1	
85	5		
86	11		
87	7		
88	6		1
89	9	1	
90	8	1	1
91	5	5	
92	6		1
93	13	1	1
94	6		
95	11	1	1
96	12		3
97	14		
98	4		
99	12	2	
100	15	2	2
101	11		1
102	10		1
103	10	1	
104	21	1	1
105	16		
106	9		1
107	9	1	
108	8	1	
109	11	1	
110	9		2
111	15	2	1
112	17		
113	14	2	
114	15	2	1
115	14		2
116	7		
117	14		
118	10		
119	13		
120	11		1

Testwert			
Wert	deutsch	französisch	italienisch
66	1		
71	1		
73	1		
74	3		
75	1	1	
77	2		
78	5		
79	3		
80	2		1
81	4		
82	5		
83	10	1	2
84	8		
85	8		
86	6		
87	16	2	1
88	16		
89	13	2	2
90	17		1
91	11	2	
92	19		
93	23		
94	18	2	1
95	27	5	1
96	21	1	1
97	31	1	1
98	25		1
99	17	2	3
100	20	2	
101	32	1	3
102	29	1	1
103	21	1	1
104	24	2	1
105	38	2	
106	21	2	
107	19	2	3
108	28		
109	13		1
110	21		2
111	15	2	
112	14	1	
113	15		1
114	14	1	1
115	7		1
116	13	1	
117	3		
118	3		
119	2		
120	3	1	1
121	2	1	
122	5		
125	2		
126	1		
127	1		

- Fortsetzung Verteilung Punktwert-

121	14		
122	9	1	
123	12	1	1
124	6		1
125	13	1	
126	14		
127	14		
128	8		
129	5	1	
130	12		
131	9		
132	7	1	1
133	8		
134	8		
135	6		
136	8		
137	7		
138	10		
139	4	1	
140	4		
141	3	1	
142	2		
143	11		
144	3		
146	1		
147	2		
148	2		
150	1		
151	2		
152	2		
154	2		
155	3		
161	2		
162	1		
164	1		

Tabelle A 4: Verteilung für Punktwert und Testwert für die Sprachgruppen

	Punkt- wert	Muster zuordnen	Med.- Natur- wiss. Grund- verständ- nis	Schlauch- figuren	Quant.- formale Probleme	Text- verständ- nis	Figuren- lernen	Fakten- lernen	Dia- gramme und Tabellen	Konzen- triertes u.sorgf. Arbeiten
Prüfung auf Normalverteilung										
Extremste Differenz	.030	.082	.091	.084	.061	.083	.076	.065	.065	.078
Kolmogorov- Smirnov Z	.827	2.254	2.491	2.288	1.683	2.268	2.074	1.767	1.776	2.140
Asymp. Sig. (2-seit.)	.501	.000	.000	.000	.007	.000	.000	.004	.004	.000
Prüfung auf Poissonverteilung										
Extremste Differenz	.171	.037	.051	.042	.053	.057	.049	.039	.025	.023
Kolmogorov- Smirnov Z	4.691	1.003	1.392	1.137	1.458	1.556	1.349	1.065	.673	.643
Asymp. Sig. (2-seit.)	.000	.267	.042	.151	.028	.016	.052	.207	.756	.803

Tabelle A 5: Prüfung auf Normal- und Poissonverteilung für die (Kolmogorov-Smirnov Test)

Testformen

Subtest		Quadr.-summe	df	Mittl. Quadr.su.	F	Sig.
Punktwert	Zwischen Gruppen	25.697	1	25.697	.062	.803
	Innerhalb Gruppen	309843.669	748	414.230		
	Total	309869.367	749			
Muster zuordnen	Zwischen Gruppen	5.216	1	5.216	.525	.469
	Innerhalb Gruppen	7430.796	748	9.934		
	Total	7436.012	749			
Med.-naturwiss. Grundverständnis	Zwischen Gruppen	5.694	1	5.694	.496	.481
	Innerhalb Gruppen	8584.925	748	11.477		
	Total	8590.619	749			
Schlauchfiguren	Zwischen Gruppen	12.389	1	12.389	1.013	.315
	Innerhalb Gruppen	9148.405	748	12.230		
	Total	9160.795	749			
Quant.-formale Probleme	Zwischen Gruppen	21.505	1	21.505	1.398	.237
	Innerhalb Gruppen	11503.494	748	15.379		
	Total	11524.999	749			
Textverständnis	Zwischen Gruppen	10.683	1	10.683	.856	.355
	Innerhalb Gruppen	9329.676	748	12.473		
	Total	9340.359	749			
Figuren lernen	Zwischen Gruppen	8.752	1	8.752	.631	.427
	Innerhalb Gruppen	10380.683	748	13.878		
	Total	10389.435	749			
Fakten lernen	Zwischen Gruppen	36.690	1	36.690	2.910	.088
	Innerhalb Gruppen	9430.229	748	12.607		
	Total	9466.919	749			
Diagramme und Tabellen	Zwischen Gruppen	12.117	1	12.117	1.106	.293
	Innerhalb Gruppen	8193.211	748	10.953		
	Total	8205.328	749			
Konzentr.u.sorgf. Arbeiten	Zwischen Gruppen	4.217	1	4.217	.397	.529
	Innerhalb Gruppen	7940.483	748	10.616		
	Total	7944.700	749			

Tabelle A 6: Ergebnisse der Varianzanalyse zur Prüfung des Einflusses der Testform

Sprachgruppen

	Levene Statistic	Fg 1	Fg2	Sig.
Punktwert	.936	2	747	.393
Muster zuordnen	.126	2	747	.882
Med.-naturwiss. Grundverständnis	1.799	2	747	.166
Schlauchfiguren	.323	2	747	.724
Quant.-formale Probleme	.297	2	747	.743
Textverständnis	.611	2	747	.543
Figuren lernen	.272	2	747	.762
Fakten lernen	1.291	2	747	.276
Diagramme und Tabellen	.973	2	747	.379
Konzentr.u.sorgf. Arbeiten	2.331	2	747	.098

Tabelle A 7: Varianzhomogenität besteht zwischen den Sprachgruppen für den Punktwert und die einzelnen Untertests

		Quadr.-summe	df	Mittl. Quadr.su.	F	Sig.
Punktwert	Zwischen Gruppen	6467.867	2	3233.934	7.962	.000
	Innerhalb Gruppen	303401.499	747	406.160		
	Total	309869.367	749			
Muster zuordnen	Zwischen Gruppen	37.550	2	18.775	1.896	.151
	Innerhalb Gruppen	7398.462	747	9.904		
	Total	7436.012	749			
Med.-naturwiss. Grundverständnis	Zwischen Gruppen	338.950	2	169.475	15.342	.000
	Innerhalb Gruppen	8251.669	747	11.046		
	Total	8590.619	749			
Schlauchfiguren	Zwischen Gruppen	125.609	2	62.805	5.192	.006
	Innerhalb Gruppen	9035.185	747	12.095		
	Total	9160.795	749			
Quant.-formale Probleme	Zwischen Gruppen	10.373	2	5.186	.336	.714
	Innerhalb Gruppen	11514.626	747	15.414		
	Total	11524.999	749			
Textverständnis	Zwischen Gruppen	98.824	2	49.412	3.994	.019
	Innerhalb Gruppen	9241.535	747	12.372		
	Total	9340.359	749			
Figuren lernen	Zwischen Gruppen	71.063	2	35.531	2.572	.077
	Innerhalb Gruppen	10318.372	747	13.813		
	Total	10389.435	749			
Fakten lernen	Zwischen Gruppen	139.116	2	69.558	5.570	.004
	Innerhalb Gruppen	9327.803	747	12.487		
	Total	9466.919	749			
Diagramme und Tabellen	Zwischen Gruppen	84.029	2	42.015	3.865	.021
	Innerhalb Gruppen	8121.299	747	10.872		
	Total	8205.328	749			
Konzentr.u.sorgf. Arbeiten	Zwischen Gruppen	126.725	2	63.362	6.054	.002
	Innerhalb Gruppen	7817.975	747	10.466		
	Total	7944.700	749			

Tabelle A 8: Einfache Varianzanalyse: Einfluss der Sprachgruppen auf Punktwert und die Untertests

	zwischen SPRACHE	und SPRACHE	Mittl. Differenz	Std. Error	Sig.
Punktwert	deutsch	französisch	8.9767	3.318	.019
		italienisch	11.3324	3.701	.006
	französisch	italienisch	2.3557	4.849	.878
Muster zuordnen	deutsch	französisch	1.0047	.518	.128
		italienisch	.1577	.578	.960
	französisch	italienisch	-.8470	.757	.503
Med.-naturwiss. Grundverständnis	deutsch	französisch	1.6078	.547	.009
		italienisch	2.9502	.610	.000
	französisch	italienisch	1.3424	.800	.213
Schlauchfiguren	deutsch	französisch	.9631	.573	.212
		italienisch	1.8059	.639	.013
	französisch	italienisch	.8428	.837	.572
Quant.-formale Probleme	deutsch	französisch	.2776	.646	.903
		italienisch	-.4883	.721	.777
	französisch	italienisch	-.7659	.945	.696
Textverständnis	deutsch	französisch	1.1398	.579	.120
		italienisch	1.3706	.646	.085
	französisch	italienisch	.2308	.846	.960
Figuren lernen	deutsch	französisch	.1179	.612	.980
		italienisch	1.5472	.683	.061
	französisch	italienisch	1.4293	.894	.246
Fakten lernen	deutsch	französisch	1.5281	.582	.023
		italienisch	1.4181	.649	.074
	französisch	italienisch	-.1100	.850	.991
Diagramme und Tabellen	deutsch	französisch	.6917	.543	.410
		italienisch	1.5321	.606	.031
	französisch	italienisch	.8404	.793	.540
Konzentr.u.sorgf. Arbeiten	deutsch	französisch	1.6460	.533	.006
		italienisch	1.0389	.594	.187
	französisch	italienisch	-.6071	.778	.715

Tabelle A 9: Multiple Mittelwertvergleiche (Tukey) für die Sprachgruppen, signifikante Mittelwertunterschiede mit $p < .05$

Testorte

		N	Mittelw.	Standardabw.	Minimum	Maximum
Punktwert	Basel	131	107.6336	19.5094	51.00	162.00
	Bern	162	108.1975	21.6711	55.00	155.00
	Zürich	298	111.5503	20.3557	41.00	164.00
	Chur	27	104.8519	19.8101	65.00	139.00
	St.Gallen	62	112.8387	17.4536	77.00	155.00
	Fribourg	39	100.8718	19.5558	52.00	141.00
	Bellinzona	31	98.5161	17.0800	64.00	132.00
	Total	750	108.9133	20.3399	41.00	164.00
Muster zuordnen	Basel	131	12.8321	3.1699	5.00	19.00
	Bern	162	13.1358	3.2255	5.00	20.00
	Zürich	298	13.4664	3.1152	5.00	20.00
	Chur	27	13.5926	2.7493	8.00	19.00
	St.Gallen	62	13.6452	3.2898	7.00	20.00
	Fribourg	39	12.2821	3.1702	5.00	18.00
	Bellinzona	31	13.1290	2.8606	9.00	19.00
	Total	750	13.2280	3.1509	5.00	20.00
Med.-naturwiss. Grundverständnis	Basel	131	12.3130	3.4487	4.00	20.00
	Bern	162	12.3827	3.5036	2.00	19.00
	Zürich	298	12.8054	3.3256	4.00	20.00
	Chur	27	11.7778	3.2502	6.00	18.00
	St.Gallen	62	12.3871	2.9104	7.00	20.00
	Fribourg	39	10.9231	2.7662	6.00	17.00
	Bellinzona	31	9.5806	3.2329	3.00	16.00
	Total	750	12.3253	3.3867	2.00	20.00
Schlauchfiguren	Basel	131	13.3969	3.3665	3.00	20.00
	Bern	162	12.9753	3.8058	5.00	20.00
	Zürich	298	13.4732	3.4613	4.00	20.00
	Chur	27	12.4074	3.1164	7.00	19.00
	St.Gallen	62	13.7419	3.1723	5.00	20.00
	Fribourg	39	12.3590	3.2484	6.00	18.00
	Bellinzona	31	11.5161	3.3850	6.00	19.00
	Total	750	13.1973	3.4972	3.00	20.00

- Fortsetzung -

Quant.-formale Probleme	Basel	131	11.4656	3.7709	3.00	19.00
	Bern	162	11.6852	4.2754	1.00	20.00
	Zürich	298	12.3490	3.9174	1.00	20.00
	Chur	27	11.8519	3.6343	5.00	19.00
	St.Gallen	62	12.2903	3.5408	4.00	20.00
	Fribourg	39	11.7179	3.5685	6.00	19.00
	Bellinzona	31	12.4839	3.9906	3.00	20.00
	Total	750	12.0013	3.9226	1.00	20.00
Textverständnis	Basel	131	9.9771	3.4873	1.00	18.00
	Bern	162	10.1358	3.3316	2.00	17.00
	Zürich	298	10.7148	3.6287	1.00	18.00
	Chur	27	9.7778	3.8062	2.00	16.00
	St.Gallen	62	10.4194	3.4287	3.00	18.00
	Fribourg	39	9.2308	3.3044	2.00	16.00
	Bellinzona	31	9.0000	3.5590	3.00	18.00
	Total	750	10.2547	3.5314	1.00	18.00
Figuren lernen	Basel	131	12.3817	3.8179	3.00	20.00
	Bern	162	12.5679	3.6319	3.00	20.00
	Zürich	298	12.7215	3.6945	2.00	20.00
	Chur	27	11.1852	3.6059	4.00	18.00
	St.Gallen	62	12.9516	3.8854	6.00	20.00
	Fribourg	39	12.4615	3.9125	1.00	19.00
	Bellinzona	31	11.0323	3.3315	5.00	18.00
	Total	750	12.5093	3.7244	1.00	20.00
Fakten lernen	Basel	131	11.6565	3.6242	4.00	20.00
	Bern	162	11.2716	3.7813	2.00	20.00
	Zürich	298	11.7785	3.4257	1.00	20.00
	Chur	27	10.7407	3.9864	2.00	18.00
	St.Gallen	62	11.6290	2.8298	6.00	17.00
	Fribourg	39	10.0513	3.8316	3.00	16.00
	Bellinzona	31	10.1613	3.2974	4.00	17.00
	Total	750	11.4413	3.5552	1.00	20.00

-Fortsetzung -

Diagramme und Tabellen	Basel	131	11.3206	3.2279	2.00	19.00
	Bern	162	11.4198	3.3096	2.00	18.00
	Zürich	298	11.4597	3.4534	3.00	20.00
	Chur	27	10.8148	3.0889	6.00	16.00
	St.Gallen	62	11.8710	3.1751	4.00	18.00
	Fribourg	39	10.7436	2.8811	5.00	16.00
	Bellinzona	31	9.9032	2.9366	5.00	17.00
	Total	750	11.3360	3.3098	2.00	20.00
Konzentr.u.sorgf. Arbeiten	Basel	131	12.2901	3.2117	5.00	20.00
	Bern	162	12.6235	3.2837	6.00	20.00
	Zürich	298	12.7819	3.2757	.00	20.00
	Chur	27	12.7037	3.1356	8.00	19.00
	St.Gallen	62	13.9032	3.1660	7.00	20.00
	Fribourg	39	11.1026	3.2751	5.00	19.00
	Bellinzona	31	11.7097	2.3550	7.00	16.00
	Total	750	12.6200	3.2569	.00	20.00

Tabelle A 10: Mittelwerte und Standardabweichungen für die Untertests nach Testorten

	Levene Statistic	Fg 1	Fg2	Sig.
Punktwert	1.394	6	743	.214
Muster zuordnen	.455	6	743	.842
Med.-naturwiss. Grundverständnis	1.405	6	743	.210
Schlauchfiguren	.881	6	743	.508
Quant.-formale Probleme	.948	6	743	.460
Textverständnis	.502	6	743	.807
Figuren lernen	.306	6	743	.934
Fakten lernen	1.809	6	743	.095
Diagramme und Tabellen	.609	6	743	.723
Konzentr.u.sorgf. Arbeiten	.854	6	743	.528

Tabelle A 11: Prüfung hinsichtlich Varianzhomogenität für die Untertests nach Testorten

Geschlecht

		N	Mittelw.	Standardabw.	Minimum	Maximum
Punktwert	männlich	336	109.8	20.8936	41.00	161.00
	weiblich	414	108.2	19.8724	52.00	164.00
	Total	750	108.9	20.3399	41.00	164.00
Muster zuordnen	männlich	336	13.0	3.1323	5.00	20.00
	weiblich	414	13.3	3.1645	5.00	20.00
	Total	750	13.2	3.1509	5.00	20.00
Med.-naturwiss. Grundverständnis	männlich	336	12.5	3.3132	3.00	20.00
	weiblich	414	12.1	3.4342	2.00	20.00
	Total	750	12.3	3.3867	2.00	20.00
Schlauchfiguren	männlich	336	13.6	3.4743	3.00	20.00
	weiblich	414	12.8	3.4793	4.00	20.00
	Total	750	13.1	3.4972	3.00	20.00
Quant.-formale Probleme	männlich	336	13.0	3.8155	1.00	20.00
	weiblich	414	11.1	3.8033	1.00	20.00
	Total	750	12.0	3.9226	1.00	20.00
Textverständnis	männlich	336	10.7	3.5264	1.00	18.00
	weiblich	414	9.8	3.4869	1.00	18.00
	Total	750	10.2	3.5314	1.00	18.00
Figuren lernen	männlich	336	12.1	3.8920	2.00	20.00
	weiblich	414	12.8	3.5529	1.00	20.00
	Total	750	12.5	3.7244	1.00	20.00
Fakten lernen	männlich	336	10.7	3.5103	1.00	20.00
	weiblich	414	11.9	3.5071	2.00	20.00
	Total	750	11.4	3.5552	1.00	20.00
Diagramme und Tabellen	männlich	336	11.8	3.3255	2.00	20.00
	weiblich	414	10.8	3.2356	2.00	18.00
	Total	750	11.3	3.3098	2.00	20.00
Konzentr.u.sorgf. Arbeiten	männlich	336	11.9	3.3384	5.00	20.00
	weiblich	414	13.1	3.0924	.00	20.00
	Total	750	12.6	3.2569	.00	20.00

Tabelle A 12: Mittelwerte und Standardabweichungen der Untertests getrennt nach Geschlechtern

	Levene Statistic	Fg 1	Fg2	Sig.
Punktwert	.719	1	748	.397
Muster zuordnen	.217	1	748	.642
Med.-naturwiss. Grundverständnis	.693	1	748	.405
Schlauchfiguren	.177	1	748	.674
Quant.-formale Probleme	.005	1	748	.943
Textverständnis	.077	1	748	.781
Figuren lernen	1.779	1	748	.183
Fakten lernen	.050	1	748	.823
Diagramme und Tabellen	.379	1	748	.538
Konzentr.u.sorgf. Arbeiten	2.495	1	748	.115

Tabelle A 13: Prüfung auf Varianzhomogenität, keine signifikanten Unterschiede sind nachweisbar

		Quadr.-summe	df	Mittl. Quadr.su.	F	Sig.
Punktwert	Zwischen Gruppen	528.620	1	528.620	1.278	.259
	Innerhalb Gruppen	309340.747	748	413.557		
	Total	309869.367	749			
Muster zuordnen	Zwischen Gruppen	13.269	1	13.269	1.337	.248
	Innerhalb Gruppen	7422.743	748	9.923		
	Total	7436.012	749			
Med.-naturwiss. Grundverständnis	Zwischen Gruppen	42.408	1	42.408	3.711	.054
	Innerhalb Gruppen	8548.210	748	11.428		
	Total	8590.619	749			
Schlauchfiguren	Zwischen Gruppen	117.614	1	117.614	9.728	.002
	Innerhalb Gruppen	9043.181	748	12.090		
	Total	9160.795	749			
Quant.-formale Probleme	Zwischen Gruppen	673.951	1	673.951	46.458	.000
	Innerhalb Gruppen	10851.048	748	14.507		
	Total	11524.999	749			
Textverständnis	Zwischen Gruppen	152.958	1	152.958	12.453	.000
	Innerhalb Gruppen	9187.401	748	12.283		
	Total	9340.359	749			
Figuren lernen	Zwischen Gruppen	101.397	1	101.397	7.372	.007
	Innerhalb Gruppen	10288.038	748	13.754		
	Total	10389.435	749			
Fakten lernen	Zwischen Gruppen	259.269	1	259.269	21.062	.000
	Innerhalb Gruppen	9207.649	748	12.310		
	Total	9466.919	749			
Diagramme und Tabellen	Zwischen Gruppen	176.839	1	176.839	16.476	.000
	Innerhalb Gruppen	8028.489	748	10.733		
	Total	8205.328	749			
Konzentr.u.sorgf. Arbeiten	Zwischen Gruppen	261.716	1	261.716	25.480	.000
	Innerhalb Gruppen	7682.984	748	10.271		
	Total	7944.700	749			

Tabelle A 14: Signifikante Unterschiede der Untertestleistungen zwischen den Geschlechtern.

Kantone innerhalb der Deutschschweiz

		Sum of Squares	df	Mean Square	F	Sig.
PUNKTWERT	Zwischen	163.422	2	81.711	.202	.817
	Innerhalb	259735.781	642	404.573		
	Total	259899.203	644			
TESTWERT	Zwischen	44.796	2	22.398	.230	.795
	Innerhalb	62584.388	642	97.483		
	Total	62629.184	644			
Muster zuordnen	Zwischen	18.986	2	9.493	.964	.382
	Innerhalb	6320.952	642	9.846		
	Total	6339.938	644			
Medizin.-naturwiss. Grundverständnis	Zwischen	2.428	2	1.214	.111	.895
	Innerhalb	7030.682	642	10.951		
	Total	7033.110	644			
Schlauchfiguren	Zwischen	35.660	2	17.830	1.468	.231
	Innerhalb	7794.985	642	12.142		
	Total	7830.645	644			
Quantitat. u.formale Probleme	Zwischen	18.196	2	9.098	.597	.551
	Innerhalb	9781.395	642	15.236		
	Total	9799.591	644			
Textverständnis	Zwischen	29.414	2	14.707	1.202	.301
	Innerhalb	7852.297	642	12.231		
	Total	7881.712	644			
Figuren lernen	Zwischen	33.868	2	16.934	1.228	.294
	Innerhalb	8855.130	642	13.793		
	Total	8888.998	644			
Fakten lernen	Zwischen	6.116	2	3.058	.250	.779
	Innerhalb	7852.697	642	12.232		
	Total	7858.812	644			
Diagramme und Tabellen	Zwischen	5.233	2	2.617	.237	.789
	Innerhalb	7078.007	642	11.025		
	Total	7083.240	644			
Konzentr. u. sorgfält. Arbeiten	Zwischen	32.482	2	16.241	1.527	.218
	Innerhalb	6828.628	642	10.636		
	Total	6861.110	644			

Tabelle A 15: Prüfung des Einflusses der Maturitätsquote – Varianzanalyse für die drei Kantons-Gruppen nach der Maturitätsquote

		Quadr.-summe	df	Mittl. Quadr.su.	F	Sig.
Punktwert	Zwischen Gruppen	2862.210	6	477.035	1.170	.321
	Innerhalb Gruppen	212011.992	520	407.715		
	Total	214874.201	526			
Muster zuordnen	Zwischen Gruppen	29.937	6	4.990	.504	.805
	Innerhalb Gruppen	5144.382	520	9.893		
	Total	5174.319	526			
Med.-naturwiss. Grundverständnis	Zwischen Gruppen	70.847	6	11.808	1.092	.366
	Innerhalb Gruppen	5623.472	520	10.814		
	Total	5694.319	526			
Schlauchfiguren	Zwischen Gruppen	53.381	6	8.897	.736	.620
	Innerhalb Gruppen	6281.648	520	12.080		
	Total	6335.028	526			
Quant.-formale Probleme	Zwischen Gruppen	150.923	6	25.154	1.641	.134
	Innerhalb Gruppen	7969.403	520	15.326		
	Total	8120.326	526			
Textverständnis	Zwischen Gruppen	165.552	6	27.592	2.289	.034
	Innerhalb Gruppen	6268.964	520	12.056		
	Total	6434.516	526			
Figuren lernen	Zwischen Gruppen	83.397	6	13.900	.972	.444
	Innerhalb Gruppen	7438.417	520	14.305		
	Total	7521.814	526			
Fakten lernen	Zwischen Gruppen	51.643	6	8.607	.681	.665
	Innerhalb Gruppen	6576.566	520	12.647		
	Total	6628.209	526			
Diagramme und Tabellen	Zwischen Gruppen	4.537	6	.756	.069	.999
	Innerhalb Gruppen	5677.106	520	10.918		
	Total	5681.643	526			
Konzentr.u.sorgf. Arbeiten	Zwischen Gruppen	185.452	6	30.909	3.036	.006
	Innerhalb Gruppen	5293.888	520	10.181		
	Total	5479.340	526			

Tabelle A 16: Varianzanalytische Prüfung des Einflusses "Wohnkanton" (mehr als 40 TeilnehmerInnen)

Alter und Maturitätsabschluss

		Quadr.-summe	df	Mittl. Quadr.su.	F	Sig.
Punktwert	Zwischen Gruppen	9219.203	2	4609.602	11.453	.000
	Innerhalb Gruppen	300650.163	747	402.477		
	Total	309869.367	749			
Muster zuordnen	Zwischen Gruppen	361.125	2	180.563	19.065	.000
	Innerhalb Gruppen	7074.887	747	9.471		
	Total	7436.012	749			
Med.-naturwiss. Grundverständnis	Zwischen Gruppen	82.612	2	41.306	3.627	.027
	Innerhalb Gruppen	8508.006	747	11.390		
	Total	8590.619	749			
Schlauchfiguren	Zwischen Gruppen	29.044	2	14.522	1.188	.305
	Innerhalb Gruppen	9131.750	747	12.225		
	Total	9160.795	749			
Quant.-formale Probleme	Zwischen Gruppen	181.020	2	90.510	5.960	.003
	Innerhalb Gruppen	11343.979	747	15.186		
	Total	11524.999	749			
Textverständnis	Zwischen Gruppen	49.561	2	24.780	1.992	.137
	Innerhalb Gruppen	9290.798	747	12.437		
	Total	9340.359	749			
Figuren lernen	Zwischen Gruppen	222.330	2	111.165	8.168	.000
	Innerhalb Gruppen	10167.105	747	13.611		
	Total	10389.435	749			
Fakten lernen	Zwischen Gruppen	81.373	2	40.686	3.238	.040
	Innerhalb Gruppen	9385.546	747	12.564		
	Total	9466.919	749			
Diagramme und Tabellen	Zwischen Gruppen	164.664	2	82.332	7.649	.001
	Innerhalb Gruppen	8040.664	747	10.764		
	Total	8205.328	749			
Konzentr.u.sorgf. Arbeiten	Zwischen Gruppen	121.032	2	60.516	5.778	.003
	Innerhalb Gruppen	7823.668	747	10.473		
	Total	7944.700	749			

Tabelle A 17: Varianzanalyse: Prüfung signifikante Unterschiede zwischen Leistungen für Altersgruppen

	Signifikanz des Mittelwertunterschiedes zwischen ...		Mean Difference (I-J)	Std. Error	Sig.
	Altersgruppe	Altersgruppe			
PUNKTWERT	jünger	mittel	-3.4344	2.642	.395
		älter	-9.4300	2.460	.000
	mittel	älter	-5.9956	1.647	.001
TESTWERT	jünger	mittel	-1.6048	1.289	.427
		älter	-5.0459	1.201	.000
	mittel	älter	-3.4411	.804	.000
Muster zuordnen	jünger	mittel	-1.3250	.405	.003
		älter	-2.1817	.377	.000
	mittel	älter	-.8566	.253	.002
Medizin.-naturwiss. Grundverständnis	jünger	mittel	-.3258	.444	.744
		älter	-.8931	.414	.079
	mittel	älter	-.5673	.277	.101
Schlauchfiguren	jünger	mittel	-.5845	.460	.413
		älter	-.6600	.429	.272
	mittel	älter	-7.5579E-02	.287	.963
Quantitat. u.formale Probleme	jünger	mittel	-1.8100E-02	.513	.999
		älter	-1.0167	.478	.084
	mittel	älter	-.9986	.320	.005
Textverständnis	jünger	mittel	.3318	.464	.755
		älter	-.2450	.432	.838
	mittel	älter	-.5768	.290	.114
Figuren lernen	jünger	mittel	-.3733	.486	.722
		älter	-1.3676	.452	.007
	mittel	älter	-.9943	.303	.003
Fakten lernen	jünger	mittel	-.5905	.467	.415
		älter	-1.0202	.435	.050
	mittel	älter	-.4297	.291	.302
Diagramme und Tabellen	jünger	mittel	.2300	.432	.855
		älter	-.7781	.402	.129
	mittel	älter	-1.0081	.269	.001
Konzentr. u. sorgfält. Arbeiten	jünger	mittel	-.7790	.426	.161
		älter	-1.2676	.397	.004
	mittel	älter	-.4885	.266	.157

Tabelle A 18: Multiple Mittelwertvergleiche für die drei Altersgruppen

		Sum of Squares	df	Mean Square	F	Sig.
PUNKTWERT	Zwischen	1227.907	1	1227.907	3.384	.070
	Innerhalb	27573.939	76	362.815		
	Total	28801.846	77			
TESTWERT	Zwischen	357.479	1	357.479	4.094	.047
	Innerhalb	6636.008	76	87.316		
	Total	6993.487	77			
Muster zuordnen	Zwischen	1.510	1	1.510	.169	.682
	Innerhalb	677.939	76	8.920		
	Total	679.449	77			
Medizin.- naturwiss. Grundverständnis	Zwischen	34.416	1	34.416	3.302	.073
	Innerhalb	792.199	76	10.424		
	Total	826.615	77			
Schlauchfiguren	Zwischen	20.757	1	20.757	1.582	.212
	Innerhalb	997.460	76	13.124		
	Total	1018.218	77			
Quantitat. und formale Probleme	Zwischen	93.782	1	93.782	5.455	.022
	Innerhalb	1306.679	76	17.193		
	Total	1400.462	77			
Textverständnis	Zwischen	40.113	1	40.113	3.905	.052
	Innerhalb	780.605	76	10.271		
	Total	820.718	77			
Figuren lernen	Zwischen	2.402	1	2.402	.176	.676
	Innerhalb	1036.637	76	13.640		
	Total	1039.038	77			
Fakten lernen	Zwischen	.503	1	.503	.052	.820
	Innerhalb	733.151	76	9.647		
	Total	733.654	77			
Diagramme und Tabellen	Zwischen	79.817	1	79.817	9.292	.003
	Innerhalb	652.862	76	8.590		
	Total	732.679	77			
Konzentr. u. sorgfält. Arbeiten	Zwischen	1.858	1	1.858	.155	.694
	Innerhalb	908.360	76	11.952		
	Total	910.218	77			

Tabelle A 19: Varianzanalyse: Prüfung des Einflusses des Maturitätsjahres für die Gruppe der 1952-1974 geborenen KandidatInnen