

EMS

Eignungstest für das Medizinstudium 2008

Bericht 14 über die Durchführung und Ergebnisse 2008

avec un résumé en français

K.-D. Hänsgen und B. Spicher

Bericht 14 (2008)



Z T D
L = C
C T D

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

Hängsen, Klaus-Dieter; Spicher, Benjamin (2008):

EMS Eignungstest für das Medizinstudium 2008;

Berichte des Zentrums für Testentwicklung, Nummer 14, im Auftrag der Schweizerischen Universitätskonferenz (SUK);

Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik, 2008.

Unter Mitarbeit von P. Bergem, M. Bernasconi, A. Debrot und M. Strazzeri.

In Zusammenarbeit mit der Rektorenkonferenz der Schweizer Universitäten (CRUS), welche das Anmelde- und Zulassungsverfahren zum Medizinstudium durchführt und für diesen Bericht statistische Angaben zur Verfügung stellt; K. Wechsler; Chr. Winzenried.

Redaktion: Tanja Läng

Die Mitglieder des Beirates EMS der Schweizerischen Universitätskonferenz 2008:

Dr. M. Weiss (Vorsitzende, SUK Bern ab 1.4.2008); Dr. N. Ischi (Vorsitzender, SUK Bern bis 31.3.2008); Prof. Dr. M. Dierich (Univ. Innsbruck); Prof. Dr. P. Eggli (Univ. Bern); Dr. R. Hofer (Univ. Bern); Dr. C. Holzbaur (Med. Univ. Wien); Prof. Dr. M. Kleinmann (Univ. Zürich); Prof. Dr. R. Mallinger (Univ. Wien); Prof. Dr. J.-P. Montani (Univ. Freiburg); Prof. Dr. K. Opwis (Univ. Basel); Prof. Dr. M. Perrez (Univ. Freiburg); Prof. Dr. J.-L. Reymond (Univ. Bern); Prof. Dr. U. M. Spornitz (Univ. Basel); Dr. M. Stauffacher (CRUS Bern); Prof. Dr. G. Trost (ITB GmbH Bonn Deutschland); K. Wechsler (CRUS Bern); Prof. Dr. D. Wolfer (Univ. Zürich)

Inhaltsverzeichnis

1	ZUSAMMENFASSUNG	5
1.1	Statistik.....	5
1.2	Organisation und Ablauf in der Schweiz.....	5
1.3	Testgüte.....	6
1.4	Gleichbehandlung der Sprachgruppen	6
1.5	Differenzierung nach Alter	6
1.6	Differenzierung nach Geschlecht.....	7
1.7	Fazit.....	7
2	RESUME	7
2.1	Statistique	7
2.2	Organisation et déroulement en Suisse	8
2.3	Qualité du test.....	8
2.4	Egalité de traitement entre groupes linguistiques	9
2.5	Différences entre classes d'âge	9
2.6	Différences entre femmes et hommes.....	9
2.7	Conclusion	9
3	NUMERUS CLAUSUS (NC) UND MEDIZINSTUDIUM	10
4	ANMELDUNG ZUM MEDIZINSTUDIUM UND TEST 2008	13
4.1	Anmeldungen, Testantritte und Zulassungen	13
4.2	Grösse der Testlokale	20
4.3	Testorte, Wunschuniversitäten und Anreise	21
4.4	Wunschuniversität und Testort nach Wohnkanton.....	22
4.5	Testabsolvierung nach Alter und Geschlecht	26
4.6	Übernahme des Testergebnisses aus Vorjahren	31
4.7	Erneute Testteilnahme	32
4.8	Sprachgruppen	35
4.9	Alter und Maturitätsjahr	37
5	BESCHREIBUNG DES VERWENDETEN EIGNUNGSTESTS	38
5.1	Aufbau des Tests.....	38
5.2	Berechnung der Werte	39
5.3	Mittlerer Rangplatz der Untertests	40
6	TESTANWENDUNG IN DER SCHWEIZ 2008	42
6.1	Verteilungsprüfung.....	42
6.2	Vergleich der Testfassungen 1998 bis 2008	45
6.3	Äquivalenz der Sprachversionen	47
6.3.1	Sprachvergleich für die Untertests.....	49
6.3.2	Darstellung des Korrekturverfahrens	51
6.3.3	Effekte der Korrektur.....	59
6.4	Vergleichbarkeit der Testlokale.....	63
6.5	Vergleich für die Geschlechter.....	65

6.6	Vergleiche für Altersgruppen	69
6.7	Vergleiche nach Wunschuniversitäten	71
7	ERGEBNISSE ZUR TESTGÜTE	73
7.1	Zuverlässigkeit	73
7.2	Binnenstruktur.....	75
7.3	Item-Trennschärfen.....	80
7.4	Konzentriertes und sorgfältiges Arbeiten.....	85
8	TEILNAHMEN AM EMS NACH KANTONEN SEIT 1998.....	90
9	BEISPIELAUFGABEN FÜR DIE UNTERTESTS	91
9.1	Untertest: Quantitative und formale Probleme.....	91
9.2	Untertest: Schlauchfiguren	91
9.3	Untertest: Textverständnis.....	92
9.4	Untertest: Planen und Organisieren.....	93
9.5	Untertest: Konzentriertes und sorgfältiges Arbeiten.....	95
9.6	Untertest: Medizinisch-naturwissenschaftliches Grundverständnis	95
9.7	Untertest: Figuren lernen	96
9.8	Untertest: Fakten lernen	97
9.9	Untertest: Muster zuordnen	97
9.10	Untertest: Diagramme und Tabellen.....	98
10	LITERATUR	99
10.1	Originaltest zur Information und Vorbereitung	100
10.2	Frühere Berichte des ZTD	101

1 Zusammenfassung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2008 für das Medizinstudium in der Schweiz (EMS) und die Zulassung zum Studium der Medizin ab Herbstsemester 2008. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen 1998 bis 2007 herangezogen.

1.1 Statistik

Im Jahr 2008 wurde der EMS in der **Schweiz** zum elften Mal durchgeführt. Ein Numerus clausus ist wiederum für Humanmedizin (seit 1998), Veterinärmedizin (seit 1999) und Zahnmedizin (seit 2004) und dort für Personen notwendig, die sich an den Universitäten Basel, Bern, Freiburg oder Zürich angemeldet haben. An der Universität Zürich gilt der NC auch für die Chiropraktik (seit 2008).

Im Februar 2008 waren mit den vorliegenden Anmeldungen die Kapazitäten dieser Universitäten in Humanmedizin zu 336% in Veterinärmedizin zu 243% und in Zahnmedizin zu 139% ausgelastet.

Insgesamt 2066 Personen haben sich bis zur gesetzten Frist im Mai 2008 bei der CRUS zum EMS-Test angemeldet (Vorjahr: 1932), 80 Personen übernehmen ihr Testergebnis aus dem Vorjahr (gleiche Zahl wie 2007).

Mit gültigem Ergebnis haben den EMS 1961 Personen beendet (Vorjahr: 1831). Vier Personen haben den Test vor dem Nachmittagsteil abgebrochen (Vorjahr: 2 Personen), 101 Personen haben ihre Testanmeldung zurückgezogen bzw. sind nicht zum Test erschienen (Vorjahr: 99).

In **Österreich** wurde in Zusammenarbeit mit dem ZTD erneut an den Medizinischen Universitäten Innsbruck und Wien für Human- und Zahnmedizin am gleichen Tag der gleiche Test eingesetzt. 7671 Personen nutzten die Internetanmeldung, 6396 Personen erschienen zur persönlichen Anmeldung und für diese musste Testlogistik bereitgestellt werden. 4934 Personen nahmen tatsächlich am EMS teil (77% der persönlich Angemeldeten). Sie bewerben sich um 1200 Plätze, wobei 75% österreichischen Bildungsinländern, 20% EU-Bürgern und 5% Nicht-EU-Bürgern zur Verfügung stehen.

In diesem Bericht werden die Gütekennwerte dieses Testeinsatzes in Österreich mit berücksichtigt. Weitere inhaltliche Auswertungen und Evaluationen liegen in den Händen der österreichischen Partner und werden in einem eigenen Bericht dargestellt.

1.2 Organisation und Ablauf in der Schweiz

Der Test fand am 4.7.2008 gleichzeitig an neun Testorten in drei Sprachen (Deutsch: Aarau-Suhr, Basel, Bern, Chur, Luzern, St. Gallen, Zürich; Französisch: Freiburg; Italienisch: Bellinzona) statt. Zu beachten ist, dass die Wahl des Testortes unabhängig vom zukünftigen Studienort möglich ist. Insgesamt waren 21 Testlokale eingerichtet. Dank einer engagierten Mitwirkung aller Beteiligten kann die Testabnahme erneut als erfolgreich eingeschätzt werden.

1.3 Testgüte

Die Zuverlässigkeitskennwerte des Testwertes liegen auch für 2008 im gewohnt hohen Bereich und entsprechen z.B. der für Eignungsdiagnostik massgeblichen DIN 33430 bzw. gleichlautenden ÖNORM D 4000, die je nach Verfahrensklasse 0.70 bis 0.85 fordern. Als Reliabilität nach der Testhalbierungsmethode wird in der Schweiz und Österreich einheitlich ein Koeffizient von 0.92 ermittelt, die Konsistenz des Testprofils liegt immer noch bei 0.82 bzw. 0.83 (hier würden zu hohe Werte allerdings dafür sprechen, dass Untertests weggelassen werden könnten).

Auch die Trennschärfen der Items weisen nach, dass die jeweils richtigen Antworten bevorzugt von Personen gegeben werden, die höhere Leistungen erreichen. Aus diesem Grunde können 2008 alle Aufgaben gewertet werden. Die Faktorenstrukturen sind über die Jahre und zwischen den beteiligten Ländern ebenfalls stabil.

Im Jahr 2008 wurden auch weitere Analysen zur **Prognosegüte** des EMS für Studienerfolg durchgeführt:

- Für **Bern** wurde nachgewiesen, dass sich die Erfolgsquote für die 2. Vorprüfung dem Wert von 90% aller Studienanfänger nähert. Auch für die 2. Vorprüfung ergibt sich noch eine Korrelation von 0.45 zwischen EMS und schriftlichem Teil bzw. 0.35 für die kombiniert schriftlich-mündliche „Parcoursprüfung“ OSPE. Die – realistische – Erhöhung der Absolventenquote auf 80-90% aller Studienanfänger ist eine wichtige Möglichkeit, mehr Ärzte auszubilden. Dies hat den Vorteil, dass die Universitäten im ersten Studienabschnitt über ausreichende Betreuungskapazitäten verfügen, um die Studienreformen fortzuführen.
- In **Wien** ergeben sich Korrelationen zwischen der ersten Prüfung SIP-1 von 0.53, wobei die Prognosekraft des EMS für den Studienerfolg für Männer (0.50) dem für Frauen (0.53) entspricht. Der Test „funktioniert“ in Österreich also ebenso wie in Deutschland und der Schweiz.

1.4 Gleichbehandlung der Sprachgruppen

Wie in jedem Jahr seit 1999 wurden durch das sogenannte DIF-Verfahren diejenigen Items von sprachabhängigen Untertests identifiziert, die in der französischen und italienischen Sprachgruppe vom allgemeinen Trend der Ergebnisse in allen Aufgaben abweichen. Weil zumindest ein Risiko besteht, dass testbedingte Ursachen dafür mitverantwortlich sein können, erfolgte eine Korrektur dieses Unterschiedes. Durchschnittlich werden den französischsprachigen Teilnehmern 0.78 Punkte, den italienischsprachigen Teilnehmern 0.74 Punkte gutgeschrieben. Der Bonus beträgt maximal 1 Punkt, im Vergleich zu den Vorjahren ist diese Korrektur sehr gering. Es kann auch erneut nachgewiesen werden, dass die Unterschiede in den **nicht sprachabhängigen** Untertests **grösser** sind als diejenigen in den **sprachabhängigen** Untertests. Eine sprachbedingte Benachteiligung kann daher auch in diesem Jahr ausgeschlossen werden. Die vorhandenen Unterschiede sind wohl vor allem auf eine unterschiedliche Zusammensetzung der drei Sprachgruppen zurückzuführen, die französische und die italienische Sprachgruppe wählen beispielsweise auch andere Studienorte ohne Numerus clausus.

1.5 Differenzierung nach Alter

Auch in diesem Jahr zeigen sich die bekannten Unterschiede für die ältesten Jahrgänge: Wenn sie ihre Maturitätsprüfung spät abgelegt haben, zeigen sich tendenziell schlechtere Ergebnisse als bei früher Maturitätsprüfung. Jüngere Gruppen erreichen in diesem Jahr bessere Testleistungen. Wir erinnern an die Ergebnisse der Evaluationsstudien, in welchen

dieses Ergebnis auch mit entsprechenden Unterschieden beim Studienerfolg verbunden war – sich im Test also ein tatsächlicher Unterschied abbildet.

1.6 Differenzierung nach Geschlecht

Es bewerben sich weiterhin mehr Frauen (64%) um ein Medizinstudium. Besonders deutlich fällt der Unterschied im Fach Veterinärmedizin aus, bei welchem 83% Bewerberinnen sind. Chancengleichheit definiert sich auch hier nicht über gleiche Mittelwerte beim Testwert, sondern über die korrekte Vorhersage des Studienerfolges für beide Gruppen. Dies wurde in den früheren Evaluationsstudien nachgewiesen. Die Quotenunterschiede weisen auf unterschiedliche Repräsentativität der Bewerberinnen und Bewerber für ihre jeweilige Geschlechterkohorte hin und allein deshalb wären identische Testwerte nicht zu erwarten. Von 1998 bis 2007 betrug dieser Unterschied immer zwischen 1.4 und 2.2 Punkten zugunsten der Männer. Der Unterschied von 1.7 Punkten im Jahr 2008 liegt im Rahmen der Vorjahre; es kann daher auch für 2008 von einer Chancengleichheit ausgegangen werden.

1.7 Fazit

Auch die elfte Testsession des EMS verlief in der Schweiz erfolgreich – zeitgleich mit dem dritten Testeinsatz in Österreich. Der Test erweist sich weiterhin als geeignetes Instrument für die Zulassung zum Medizinstudium, indem er die Eignung mit ausreichend hoher Güte vorhersagbar macht.

2 Résumé

Le présent rapport expose les résultats concernant le test d'aptitudes aux études de médecine en Suisse (AMS) ainsi que les admissions aux études de médecine au semestre d'automne 2008. Il fait également intervenir des données comparatives des éditions antérieures du test (1998-2007).

2.1 Statistique

Cette année, l'AMS a été exécuté pour la onzième fois en **Suisse**. Un numerus clausus (NC) a de nouveau été nécessaire pour la médecine humaine (NC depuis 1998), la médecine vétérinaire (NC depuis 1999) et la médecine dentaire (NC depuis 2004) pour les personnes s'étant inscrites aux universités de Bâle, Berne, Fribourg ou Zurich. A l'Université de Zurich, le NC s'applique également à la chiropractie (depuis 2008).

En février 2008, avec le total des inscriptions, les capacités de ces universités étaient dépassées de 336% en médecine humaine, de 243% en médecine vétérinaire et de 139% en médecine dentaire.

Au total, 2066 (2007: 1932) personnes se sont inscrites auprès de la CRUS au test AMS jusqu'au délai accordé (mai 2008); 80 personnes ont repris leur résultat de l'année précédente (même nombre qu'en 2007).

1961 personnes ont effectué le test (2007: 1831). 4 personnes (2007: 2) ont interrompu le test pendant la session du matin, 101 (2007: 99) ont retiré leur inscription ou ne se sont pas présentées au test.

En **Autriche**, on a de nouveau appliqué en collaboration avec le ZTD aux Universités de médecine d'Innsbruck et de Vienne le même test le même jour pour la médecine humaine et dentaire. 7671 personnes ont utilisé l'inscription par Internet, 6396 personnes se sont présentées pour une inscription personnelle et pour celles-ci, une logistique de test a dû être mise à disposition. 4934 personnes ont effectivement pris part à l'AMS (77% des inscriptions personnelles).

Elles ont sollicité 1200 places, dont le 75% étaient disponibles pour des personnes scolarisées en Autriche, 20% pour des citoyens de l'UE et 5% pour des personnes n'appartenant pas à l'UE.

Nous tenons également compte dans ce rapport des caractéristiques de qualité de ce test en Autriche. D'autres analyses et évaluations du contenu sont de la responsabilité des partenaires autrichiens et seront présentés dans un propre rapport.

2.2 Organisation et déroulement en Suisse

Le test s'est déroulé le 4.7.2008 simultanément en neuf sites différents et en trois langues (allemand: Aarau-Suhr, Bâle, Berne, Coire, Lucerne, Saint-Gall, Zurich; français: Fribourg; italien: Bellinzone). Il est à noter qu'il est possible de choisir le lieu du test indépendamment du futur lieu d'études désiré. On a équipé au total 21 locaux de test. Grâce à la collaboration engagée de toutes les personnes impliquées, le test peut à nouveau être considéré comme réussi.

2.3 Qualité du test

Les indices de fiabilité sont, en 2008 comme pour les années précédentes, élevés et correspondent par exemple aux normes déterminantes pour le diagnostic d'aptitude DIN 33430, respectivement ÖNORM D 4000, qui exigent selon la classe de procédé 0.70 à 0.85. On détermine uniformément comme fiabilité selon la méthode de partition pairs-impairs, en Suisse et en Autriche, un coefficient de 0.92, la consistance du profil de test se situe toujours et encore à 0.82, respectivement 0.83 (de trop hautes valeurs signifieraient toutefois que des sous-tests pourraient être abandonnés).

La sélectivité des items prouve que les réponses justes ont chaque fois été données de préférence par les personnes qui atteignent les meilleures performances. Pour cette raison, tous les problèmes ont pu être évalués en 2008. Les structures factorielles restent également stables d'année en année entre les pays participants.

En 2008, d'autres analyses ont été exécutées aussi sur la **qualité du pronostic** de l'AMS pour le succès aux études:

- Pour **Berne**, on a prouvé que le taux de réussite pour le 2^{ème} examen propédeutique approche la valeur de 90% de tous ceux qui ont commencé les études. Aussi pour le 2^{ème} examen propédeutique, on obtient encore une corrélation de 0.45 entre l'AMS et la partie écrite respectivement de 0.35 pour l' „examen de parcours“ OSPE combiné écrit-oral. L'augmentation – réaliste – du taux de diplômés à 80-90% de tous ceux qui ont commencé les études offre une possibilité importante afin de former davantage de médecins. Cela présente l'avantage que les universités disposent dans la première partie des études de capacités de soutien suffisantes afin de poursuivre les réformes des études.
- A **Vienne**, on obtient des corrélations entre le premier examen SIP-1 de 0.53, le pouvoir prédictif de l'AMS pour le succès aux études correspondant pour les hommes (0.50) à celui des femmes (0.53). Le test „fonctionne“ donc en Autriche comme en Allemagne et en Suisse.

2.4 Egalité de traitement entre groupes linguistiques

Comme chaque année depuis 1999, grâce à la procédure appelée DIF, on a pu identifier les items des sous-tests à forte composante linguistique qui s'écartent des tendances générales des résultats dans tous les exercices dans les groupes francophones ou italophones. Comme il existe un risque que des facteurs liés au test soient responsables de ces différences, elles sont corrigées. En moyenne, les participants francophones sont crédités de 0.78 point, les participants italophones de 0.74 point. Le bonus atteint au maximum 1 point, en comparaison des années précédentes, cette correction est très modeste. On peut aussi à nouveau prouver que les différences sont plus grandes dans les sous-tests à faible composante linguistique que dans ceux à forte composante linguistique. Cette année aussi, un préjudice lié à la version linguistique du test peut donc être exclu. Les différences existantes doivent plutôt être attribuées à une composition différente des trois groupes linguistiques : les francophones et les italophones peuvent par exemple choisir d'autres lieux d'études sans Numerus clausus.

2.5 Différences entre classes d'âge

On trouve cette année aussi les différences connues pour le groupe des plus âgés: si ils ont obtenus plus tard leur maturité, ils ont des résultats en tendance plus faibles que ceux qui ont obtenus leur maturité plus tôt. Les groupes des plus jeunes obtiennent cette année de meilleures performances au test. Nous rappelons les résultats des études d'évaluation, dans lesquels ce résultat était aussi lié à des différences correspondantes pour le succès aux études – il se reproduit donc effectivement une différence dans le test.

2.6 Différences entre femmes et hommes

Il y a toujours plus de femmes (64%) qui se présentent aux études de médecine. Cette différence est particulièrement marquée dans la branche de médecine vétérinaire, où 83% sont des participantes. L'égalité des chances ne se définit ici pas non plus par des moyennes égales au test, mais par la prédiction correcte du succès aux études pour les deux groupes. Cela a été confirmé dans des études d'évaluation antérieures. Les différences dans les quotas indiquent que la représentativité des candidats masculins et féminins est différente ; rien que pour cette raison, on ne devrait pas s'attendre à des scores au test identiques. De 1998 à 2007, cette différence s'est toujours élevée à des valeurs comprises entre 1.4 et 2.2 points en faveur des hommes. La différence de 1.7 points de 2008 se situe dans la même tranche que les années précédentes ; on peut donc présumer que l'égalité des chances est donnée en 2008 aussi.

2.7 Conclusion

La onzième session de l'AMS s'est aussi déroulée avec succès en Suisse – en même temps que le troisième test en Autriche. Le test se révèle toujours comme un instrument adéquat pour l'admission aux études de médecine, en prédisant l'aptitude avec une qualité suffisamment bonne.

3 Numerus clausus (NC) und Medizinstudium

Die Abbildungen 1 und 2 verdeutlichen die Anmeldeverläufe 1991 bis 2008. Für die Universitäten mit NC stieg im Jahr 2007 die Zahl der Anmeldungen überproportional – kehrt 2008 zum langjährigen Trend zurück. Ursache war ein zunächst nur 2007 eingesetztes, vereinfachtes elektronisches Anmeldeverfahren. 2008 wurde zur konventionellen Anmeldung zurückgekehrt. Die Zahl der Bewerber mit gültigem EMS steigt dagegen kontinuierlich, hier findet sich dieser „Knick“ nicht (siehe Seite 18). Diese vermehrten Rückzüge fanden also bis zur Anmeldung zum Test statt.

Auch in der Westschweiz steigt die Zahl der Anmeldungen leicht an. Hier ist aber bekannt, dass nach dem ersten Jahr eine verschärfte Prüfung stattfinden muss, um die Studierendenzahlen mit den Klinikkapazitäten in Einklang zu bringen. Die Wahrscheinlichkeit zum Weiterstudieren nach dem ersten Jahr liegt nunmehr um 50%. Es scheint so, dass die wahrgenommene Attraktivität der Studienorte davon beeinflusst wird.

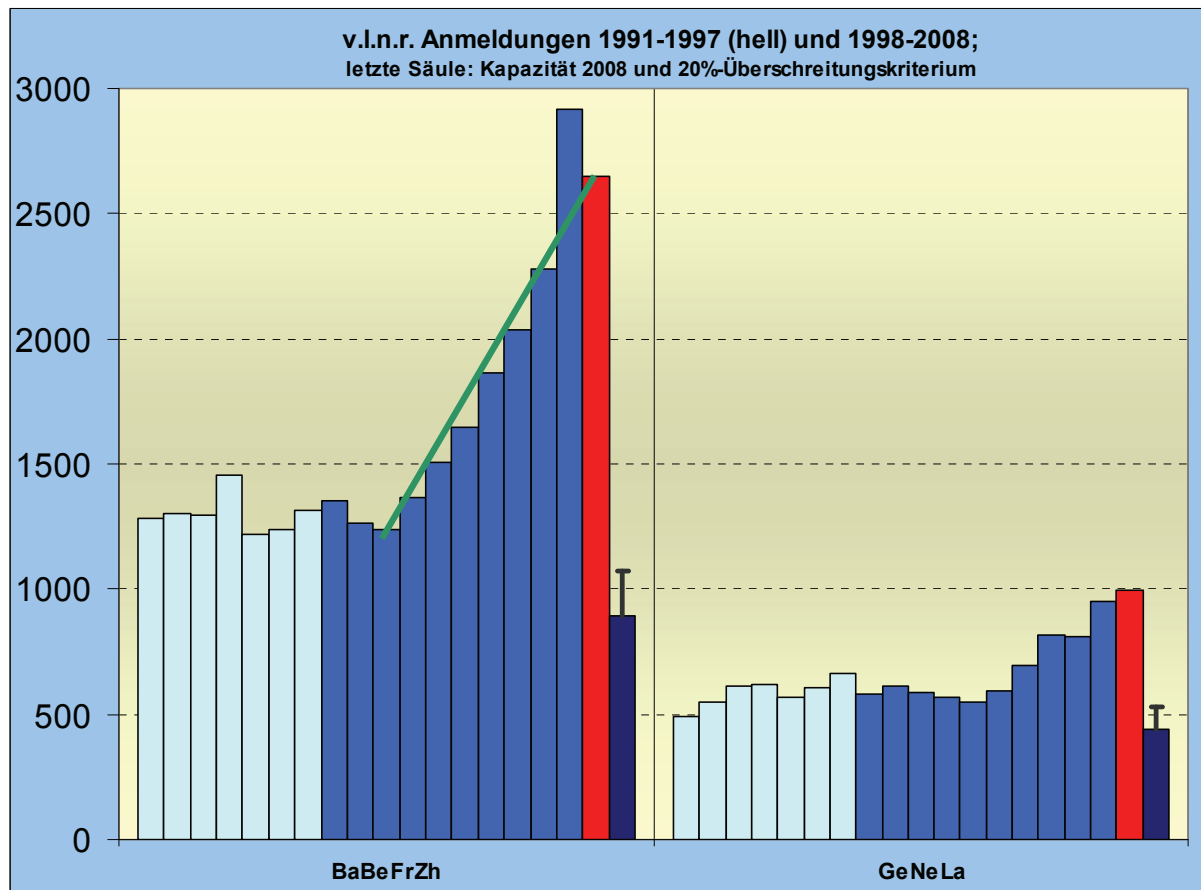


Abbildung 1: Anmeldungen zum Medizinstudium 1991 bis 2008 nach Gruppen (**B**asel, **B**ern, **F**reiburg, **Z**ürich vs. **G**enf, **N**euenburg, **L**ausanne). Hellere Säulen: Jahre ohne NC; dunklere Säulen: Jahre mit NC; letzte Säule: Kapazität 2008 und Markierung 20%-Überschreitungskriterium als Grenze für NC.

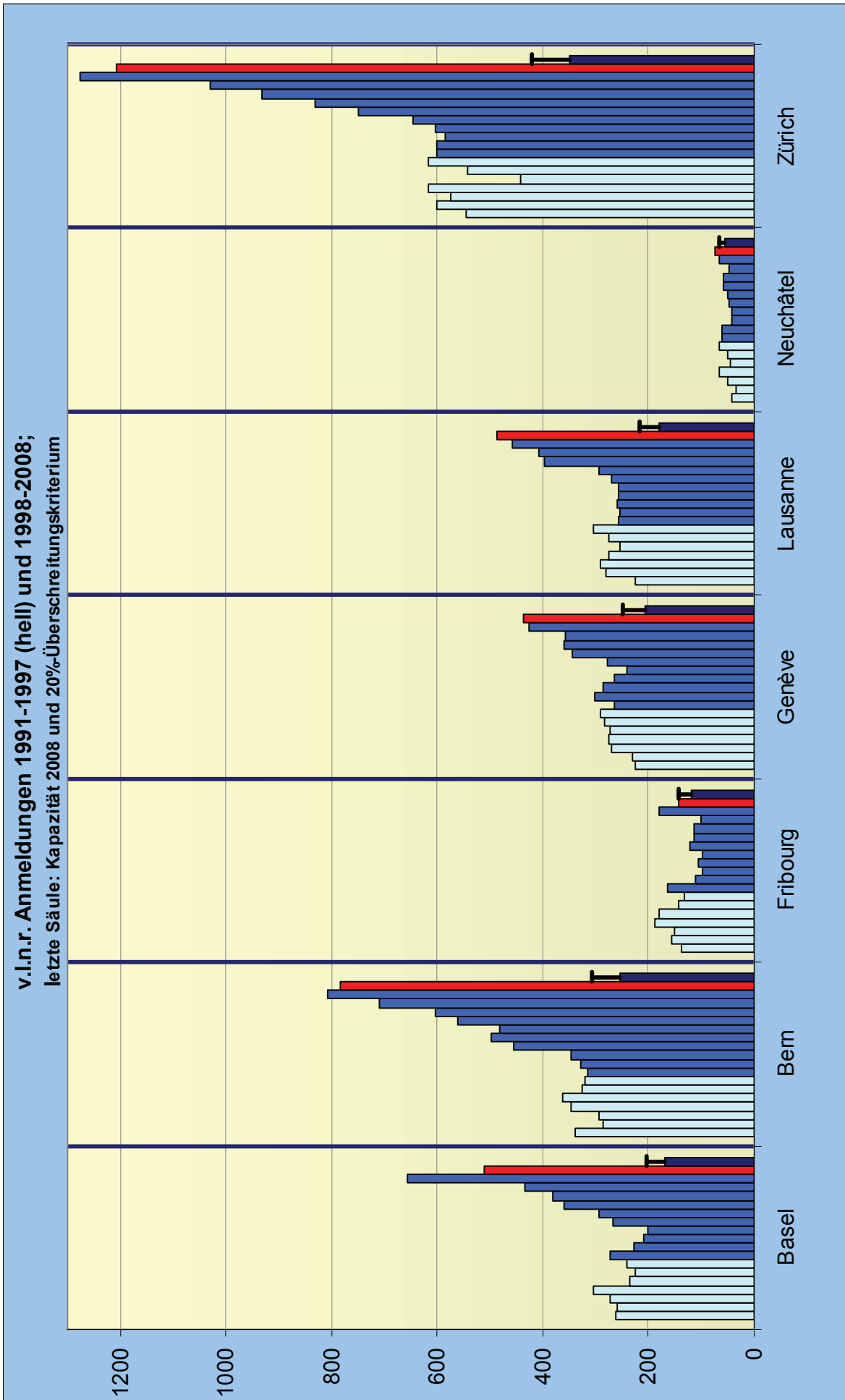


Abbildung 2: Anmeldezahlen pro Universität 1991 bis 2008 – hellere Säulen: Jahre ohne NC; dunklere Säulen: Jahre mit NC; letzte Säule: Kapazität 2008 und Markierung 20%-Überschreitungskriterium als Grenze für NC.

Verglichen mit dem Index aller Studieneintritte an Universitäten und Hochschulen liegt der Anstieg der Anmeldungen für Humanmedizin weiter deutlich höher. Der „Knick“ der Februar-Anmeldungen ist wie beschrieben auf den einmaligen Wechsel des Anmeldeverfahrens zur elektronischen Form zurückzuführen. Die Zahl der Personen mit gültigem EMS-Ergebnis steigt weiter ungebremst. Die öffentliche Diskussion geht von einem steigenden Ärztebedarf aus, Medizinalberufe etablieren sich im Unterschied zu vielen anderen als perspektivreich und sicher.

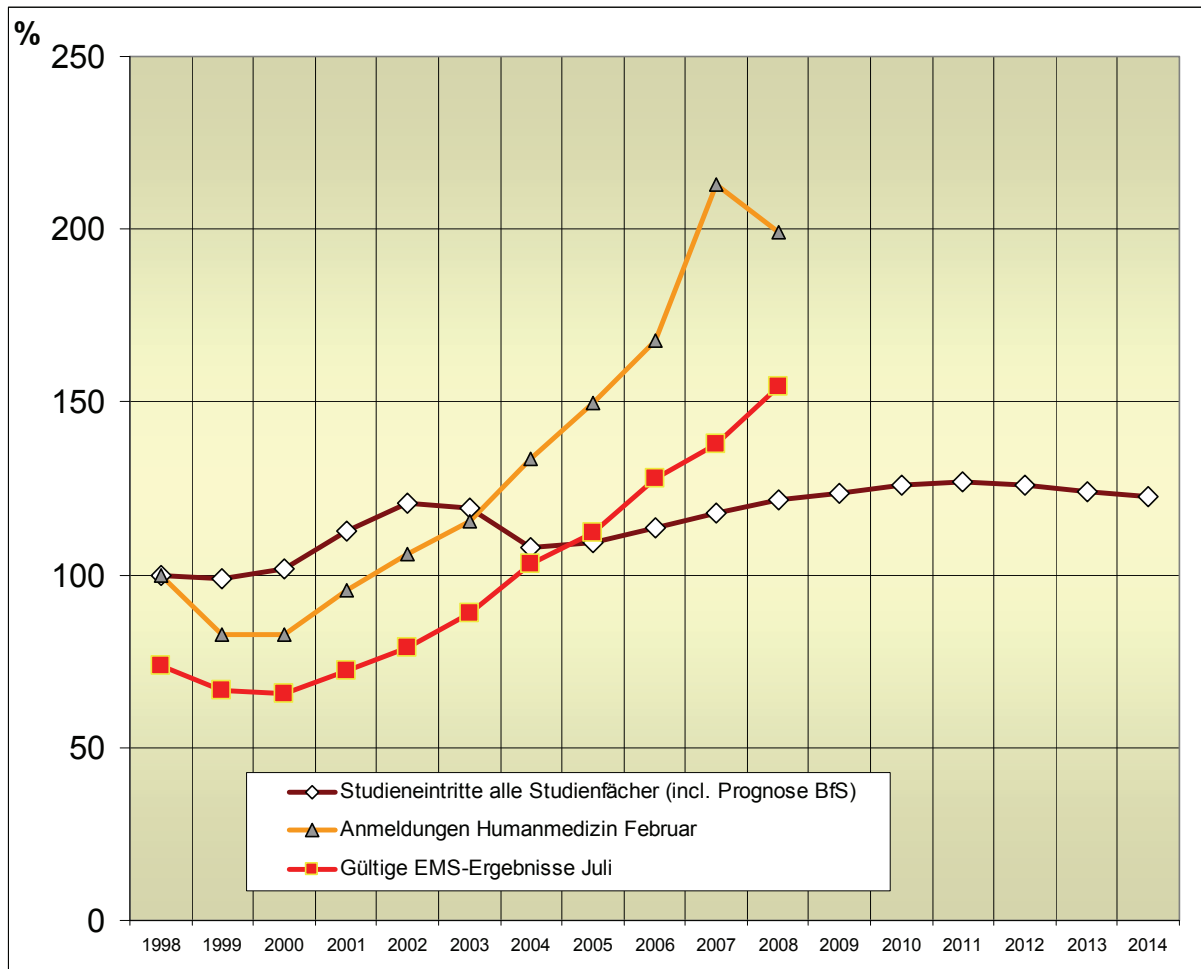


Abbildung 3: **Alle** Studieneintritte an Universitäten und Hochschulen (indexiert, 1998 = 100%, Stand September 2005, Szenario „neutral HE-A-2004“) im Vergleich zu **Medizin**-Studienanmeldungen Februar und Bewerbungen mit gültigem EMS (indexiert, Anmeldungen 1998 = 100%) für Humanmedizin.

4 Anmeldung zum Medizinstudium und Test 2008

4.1 Anmeldungen, Testantritte und Zulassungen

	Humanmedizin (und seit 2008 Chiropraktik)										
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Kapazitäten (Univ. mit NC)	620	546	546	583	648	598	546	546	546	583	603
Anmeldungen CRUS im Februar	1020	845	844	974	1080	1177	1360	1525	1712	2171	2029
... in % Kapazität	186	155	154	167	185	197	249	279	314	372	336
Anmeldungen zum EMS im Mai	797	717	678	752	827	917	1083	1182	1356	1452	1612
+ Übernahme Ergebnis Vorjahr	0	5	33	27	40	47	38	41	45	41	52
= Total Bewerbungen Mai	797	722	711	779	867	964	1121	1223	1401	1493	1664
... in % zu Kapazität	129	132	130	134	134	161	205	224	257	256	276
Absolvierung EMS	750	673	637	709	770	871	1026	1107	1263	1374	1535*
Total Bewerbungen Juli	750	678	670	736	805	907	1051	1143	1302	1405	1576*
Rückzugsquote zwischen Februar und Juli (%)	26.5	19.8	20.6	24.4	25.5	22.9	22.7	25.0	23.9	34.8	22.3
Zugeweilte Studienplätze mit „Überbuchung“	669	631	652	672	758	705	624	640	653	685	726^m
Abgewiesene Bewerbungen	81	42	14	58	47	202	426	503	649	720	850^m
% Bewerbungen, die Studienplatz erhalten	89	94	98	92	94	78	59	56	50	49	45^m

Tabelle 1: Disziplinspezifische Statistiken für die am NC beteiligten Universitäten Basel, Bern, Freiburg und Zürich für Humanmedizin

* 11 Personen wollen aus einem begonnenen Studium der Zahnmedizin in ein höheres Semester Humanmedizin wechseln. Sie nehmen am Test teil, zählen aber nicht als Bewerbungen um die Studienplätze 2008. Vergleichbare Fälle wurden auch in den Vorjahren entsprechend herausgerechnet; 20 Plätze für Chiropraktik (29 Bewerbungen) sind mit eingerechnet.

^m Stand Oktober gemäss Modell

„Modell“ heisst in den Tabellen 1 bis 3, dass mehr Personen einen Studienplatz erhalten als Kapazitäten vorhanden sind. Diese „Überbuchung“ beruht auf den Erfahrungen der Vorjahre und berücksichtigt alle Nichtantritte trotz Zulassung (z.B. weil bei Umleitungen der Studienort nicht zusagt). Ziel der Überbuchungen ist es, die Kapazitäten möglichst genau auszulasten und keine Nachrückverfahren notwendig werden zu lassen. Durch das Vorziehen des Semesterbeginns wird dies in Zukunft besonders wichtig. In den letzten Jahren konnten weitere Zuteilungsrunden zum Auslasten der Kapazität weitgehend vermieden werden. Diese würden die Zahl der abgewiesenen Personen verringern. Im Jahr 2007 gelang dies auf 0.5% genau, die Kapazität auszulasten.

	Veterinärmedizin									
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Kapazitäten (Univ. mit NC)	150	150	165	185	170	150	150	150	150	150
Anmeldungen CRUS im Februar	231	213	217	254	271	273	318	348	411	365
... in % zu Kapazität	153	142	132	137	159	182	212	232	274	243
Anmeldungen zum EMS im Mai	194	176	179	206	225	207	257	279	301	290
+ Übernahme Ergebnis Vorjahr	0	5	5	3	11	13	8	10	7	15
= Total Bewerbungen Mai	194	181	184	209	236	220	265	289	308	305
... in % zu Kapazität	129	121	112	113	139	147	177	193	205	203
Absolvierung EMS	182	164	171	203	217	196	244	268	288	272
Total Bewerbungen Juli	182	167	176	204	228	209	252	278	295	287
Rückzugsquote zwischen Februar und Juli (%)	21.2	20.7	18.9	19.3	15.9	23.4	20.8	20.1	28.2	21.3
Zugeweilte Studienplätze mit „Überbuchung“	160	167	176	202	194	175	180	170	173	173^m
Abgewiesene Bewerbungen	22	0	0	2	34	34	72	108	122	114^m
% Bewerbungen, die Studienplatz erhalten	88	100	100	99	85	84	71	61	59	60^m

Tabelle 2: Disziplinspezifische Statistiken für Veterinärmedizin und die hier am NC beteiligten Universitäten Bern und Zürich (in Freiburg und Basel wird diese Studienrichtung nicht angeboten); ^m Stand Oktober gemäss Modell.

	Zahnmedizin				
	2004	2005	2006	2007	2008
Kapazitäten (Univ. mit NC)	142	142	142	142	142
Anmeldungen CRUS im Februar	208	190	217	267	197
... in % zu Kapazität	146	134	153	188	138
Anmeldungen zum EMS im Mai	163	159	168	179	164
+ Übernahme Ergebnis Vorjahr	1	11	20	32	13
= Total Bewerbungen Mai	164	170	188	212	177
... in % zu Kapazität	115	120	132	149	124
Absolvierung EMS	151	150	157	169	154
Total Bewerbungen Juli	151	161	174	201	167
Rückzugsquote zwischen Februar und Juli (%)	26.9	15.2	19.3	24.7	15,2
Zugewiesene Studienplätze mit „Überbuchung“	151	161	172	165	167^m
Abgewiesene Bewerbungen	0	0	2	36	0^m
% Bewerbungen, die Studienplatz erhalten	100	100	99	82	99^m

Tabelle 3: **Disziplinspezifische Statistiken für die am NC beteiligten Universitäten Basel, Bern, Freiburg und Zürich für Zahnmedizin;**

^m Stand Oktober gemäss Modell.

Für die Zulassung zur **Chiropraktik** ergibt sich ein Problem: Diese Personen müssen das Kriterium der Humanmedizin erfüllen, bilden selbst aber nur eine sehr kleine Kohorte. Der Mittelwertunterschied zur Humanmedizin ist nicht signifikant (Testwert und mittlerer Rangplatz) – Kandidaten für Chiropraktik sind im Mittel nicht weniger geeignet als diejenigen für Humanmedizin. Es gilt quasi automatisch die gleiche Zulassungsquote wie für Humanmedizin – obwohl 20 Plätze für 27 Kandidaten eigentlich bereitstehen. Will man die 20 zur Verfügung stehenden Plätze punktgenau vergeben, muss das Zulassungsverfahren 2 Kriterien erfüllen:

- separater Zulassungswert für Chiropraktik
- Gewährleistung, dass kein späterer Wechsel zur Humanmedizin möglich ist, wenn dieses Zulassungskriterium nicht ebenfalls erfüllt wurde (da dies sich ansonsten als „leichterer“ Zugang zum Studium der Humanmedizin etablieren würde).

		N	Spannweite	Minimum	Maximum	Mittelwert	Standardabweichung
Testwert	CP	27	25	86	111	99.2	6.6
	HM	1497	59	70	129	100.8	9.9
Mittlerer Rangplatz	CP	27	471	304	775	511.4	127.2
	HM	1497	930	42	972	485.4	176.8

Tabelle 4: Kennwerte Chiropraktik (CP) und Humanmedizin (HM) – Testwert und mittlerer Rangplatz

		Test Varianzgleichheit		T-Test für die Mittelwertgleichheit		
		F	Signifikanz	T	df	Sig. (2seitig)
Testwert	Varianzen gleich	6.016	.014	.846	1522	.398
	Varianzen nicht gleich			1.267	28.217	.216
Mittlerer Rangplatz	Varianzen gleich	4.894	.027	-.761	1522	.447
	Varianzen nicht gleich			-1.045	27.844	.305

Tabelle 5: Signifikanzprüfung der Mittelwertunterschiede Chiropraktik vs. Humanmedizin

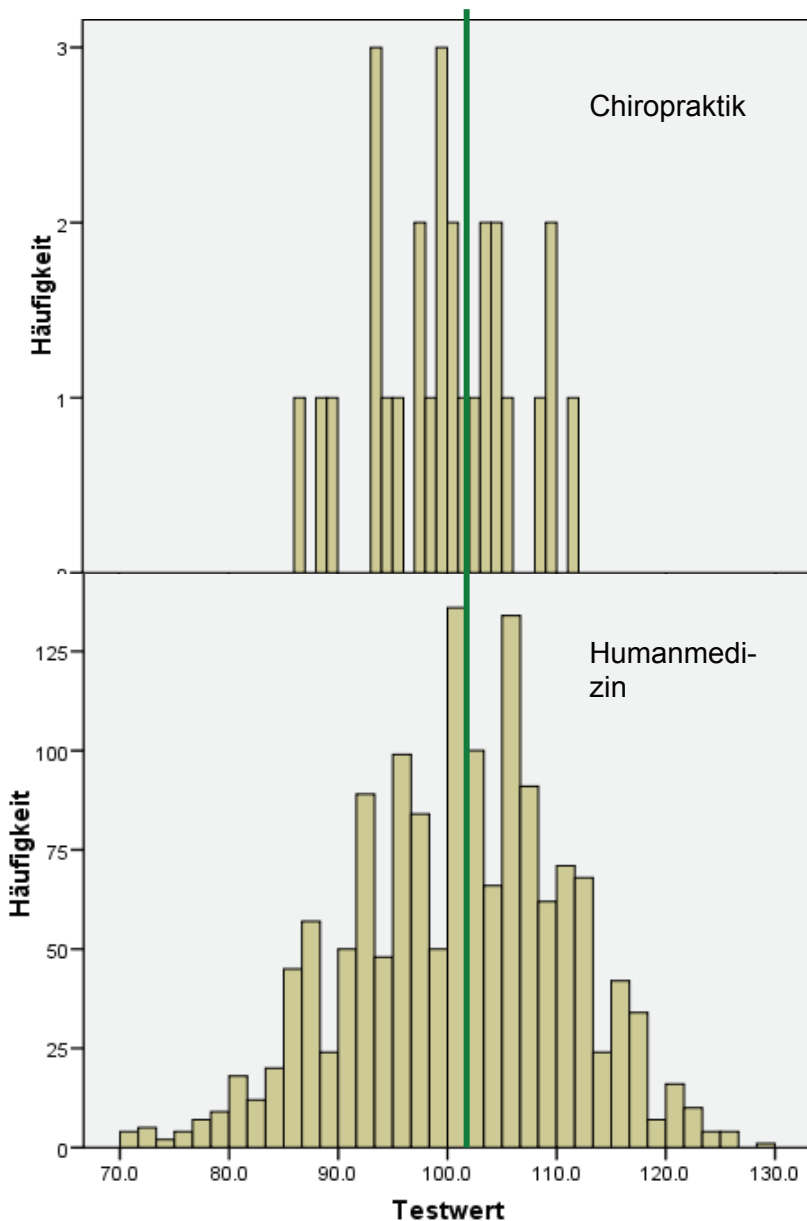


Abbildung 4: Häufigkeitsverteilungen für Bewerbungen Humanmedizin und Chiropraktik. Senkrechte Linie: Zulassungsgrenze

	Studienbeginn 2005			Studienbeginn 2006			Studienbeginn 2007			Studienbeginn 2008		
	Anmel- dungen am 15.2.	Pers. m. gültigem Testwert	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Testwert	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Testwert	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Testwert	Rück- züge in %
Basel HM	350	267	23.7	386	275	28.8	565	340	39.8	443	346	21.9
Bern HM	396	299	24.5	458	376	17.9	515	357	30.7	527	434	17.6
Freiburg HM	105	82	21.9	94	61	35.1	161	107	33.5	126	103	18.3
Zürich HM CP	674	495	26.6	774	590	23.8	930	601	35.4	895 38	665 27	25.7 28.9
Total	1525	1143	25.0	1712	1302	23.9	2171	1405	35.3	2029	1575	22.4
Bern VM	152	121	20.4	187	143	23.5	195	137	29.7	195	156	20.0
Zürich VM	166	131	21.1	161	135	16.1	216	158	26.9	170	131	22.9
Total	318	252	20.8	348	278	20.1	411	295	28.2	365	287	21.4
Basel ZM	32	31	3.1	49	41	16.3	74	49	33.8	54	52	3.7
Bern ZM	56	47	16.1	66	59	10.6	73	62	15.1	48	40	16.7
Freiburg ZM	9	7	22.2	7	6	14.3	14	9	35.7	13	12	7.7
Zürich ZM	93	76	18.3	95	68	28.4	106	81	23.6	82	63	23.2
Total	190	161	15.3	217	174	19.8	267	201	24.7	197	167	15.2

Tabelle 6: Anmeldeverlauf für Universitäten, Human- (HM) incl. Chiropraktik (CP), Veterinär- (VM) und Zahnmedizin (ZM)

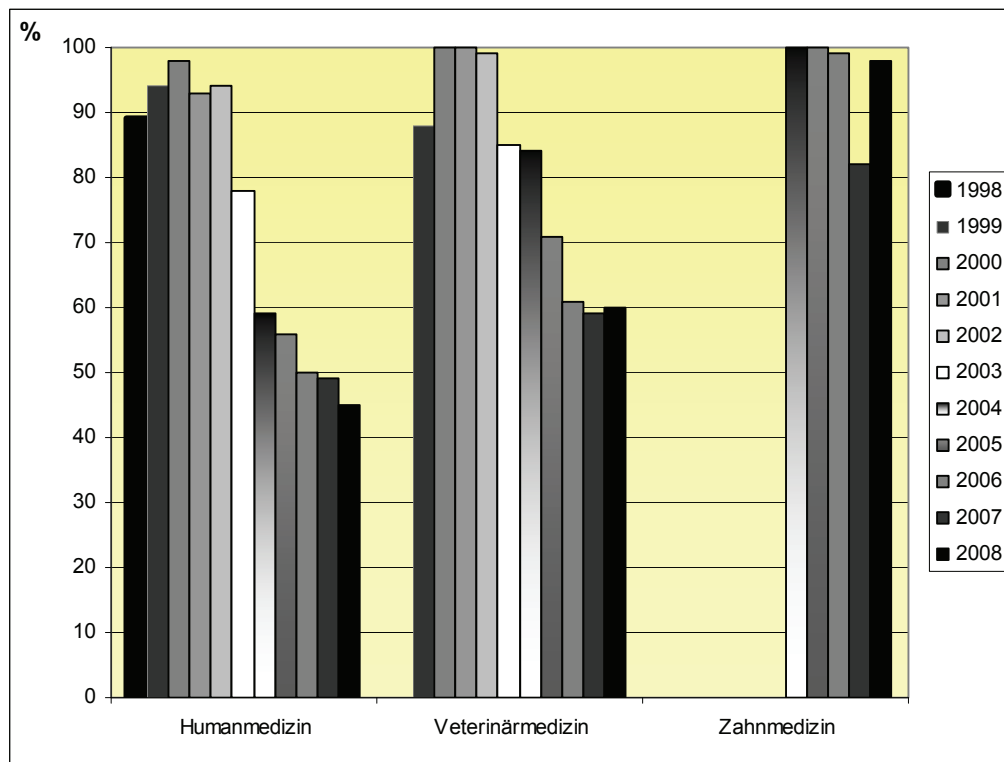


Abbildung 5: Prozent Personen mit gültigem Testwert, denen ein Studienplatz zugewiesen werden kann.

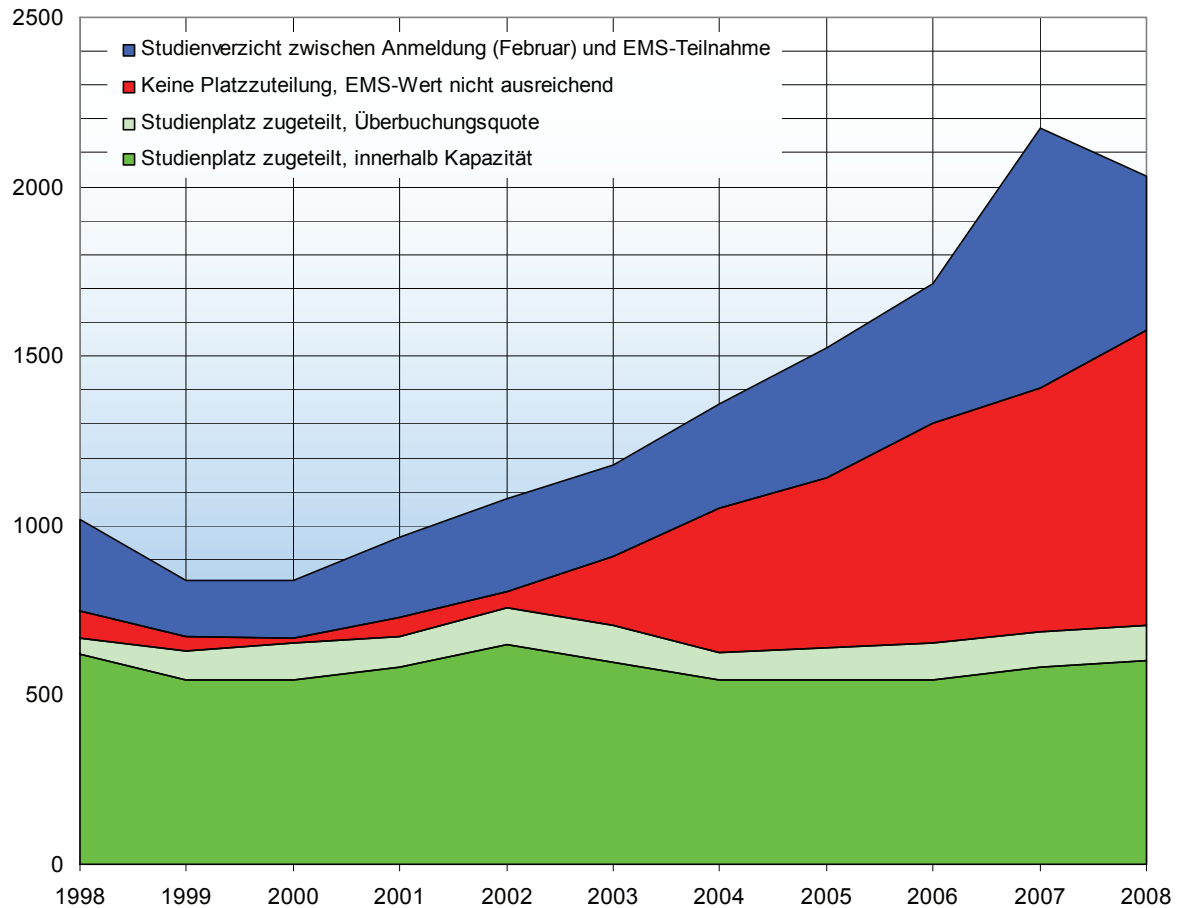


Abbildung 6: Anmeldeverlauf für Humanmedizin 1998 bis 2008 im Vergleich (Universitäten mit NC).

In der Abbildung 5 wird deutlich, dass seit 2002 der mittlere Anstieg der Bewerbungen mit gültigem Testergebnis (ohne Studienverzicht zwischen Februar und Mai) für Humanmedizin konstant ist und pro Jahr etwa der Kapazität einer Universität wie Basel entspricht. Die Zahl der Ablehnungen übersteigt die Zahl der Zulassungen jetzt deutlich.

In Veterinärmedizin ist eine Zunahme seit 2004 zunächst gebremst.

In Zahnmedizin war die dissuasive Wirkung des Testeinsatzes ausreichend, letztlich alle Personen zuzulassen. Hier wird der Testwert vor allem benötigt, um die Umleitungen entsprechend fair vornehmen zu können.

Durch die Überbuchung konnte in den letzten Jahren die Kapazität sehr genau ausgelastet werden (2007 Abweichung 0.5%). Im Jahr 2008 war eine zweite Runde notwendig, um die Kapazität genau auszulasten. Hier muss beobachtet werden, ob mehrere Personen bereits am Test teilnehmen, obwohl sie erst im Folgejahr tatsächlich studieren wollen. Es birgt für diese Personen ein gewisses Risiko, eine Zulassung nicht wahrzunehmen, weil bei steigender Nachfrage die Zulassungsgrenzwerte immer höher werden und nicht garantiert werden kann, dass auch im nächsten Jahr der gleiche Wert für eine Zulassung noch ausreichend ist.

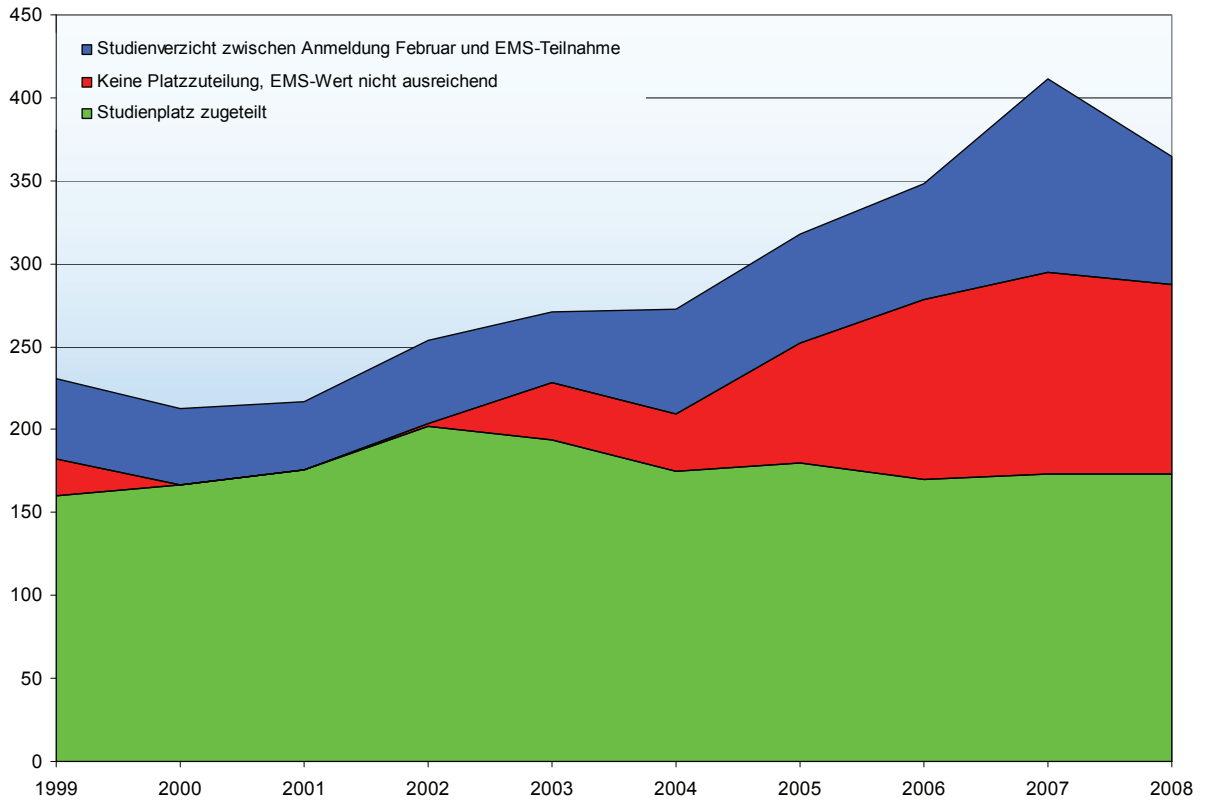


Abbildung 7: Anmeldeverlauf für Veterinärmedizin 1999 bis 2008 im Vergleich (Univ. mit NC).

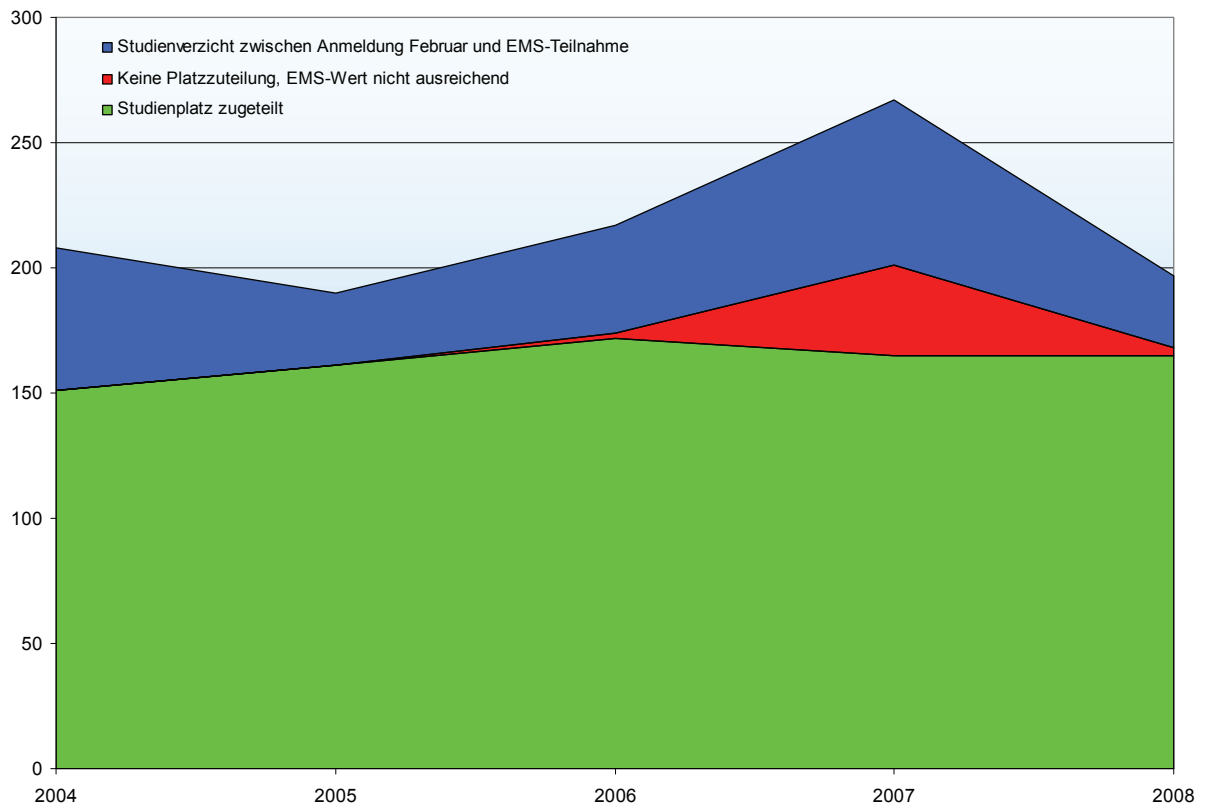


Abbildung 8: Anmeldeverlauf für Zahnmedizin 2004 bis 2008 im Vergleich (Univ. mit NC).

4.2 Grösse der Testlokale

In diesem Jahr erfolgte die Testabnahme in 21 Testlokalen verteilt auf 8 Testorte, Luzern ist neu dazugekommen und erfreute sich einer grossen Nachfrage.

Die Grösse der Testlokale variierte im bisher üblichen Rahmen. Die Zahl der Betreuer war proportional zur Teilnehmerzahl vorgegeben. Es gibt keine Hinweise seitens des Betreuungspersonals oder durch Kandidaten, dass sich unterschiedliche Raumgrössen auf die Ergebnisse auswirken. In den grossen Räumen ist der zur Verfügung stehende Platz teilweise sogar grosszügiger (wenn es sich um grössere Hörsäle handelt).

Interessant ist in diesem Zusammenhang ein Befund aus Innsbruck aus dem Jahre 2006, in welchem sich auch für sehr grosse Testlokale (350 Personen) kein Unterschied beim Testergebnis zu kleineren Testlokalen (70 Personen) ergab. Die in der Schweiz als gross geltenden Testlokale erreichen noch nicht die Grenze, ab der man nachteilige Effekte erwarten müsste.

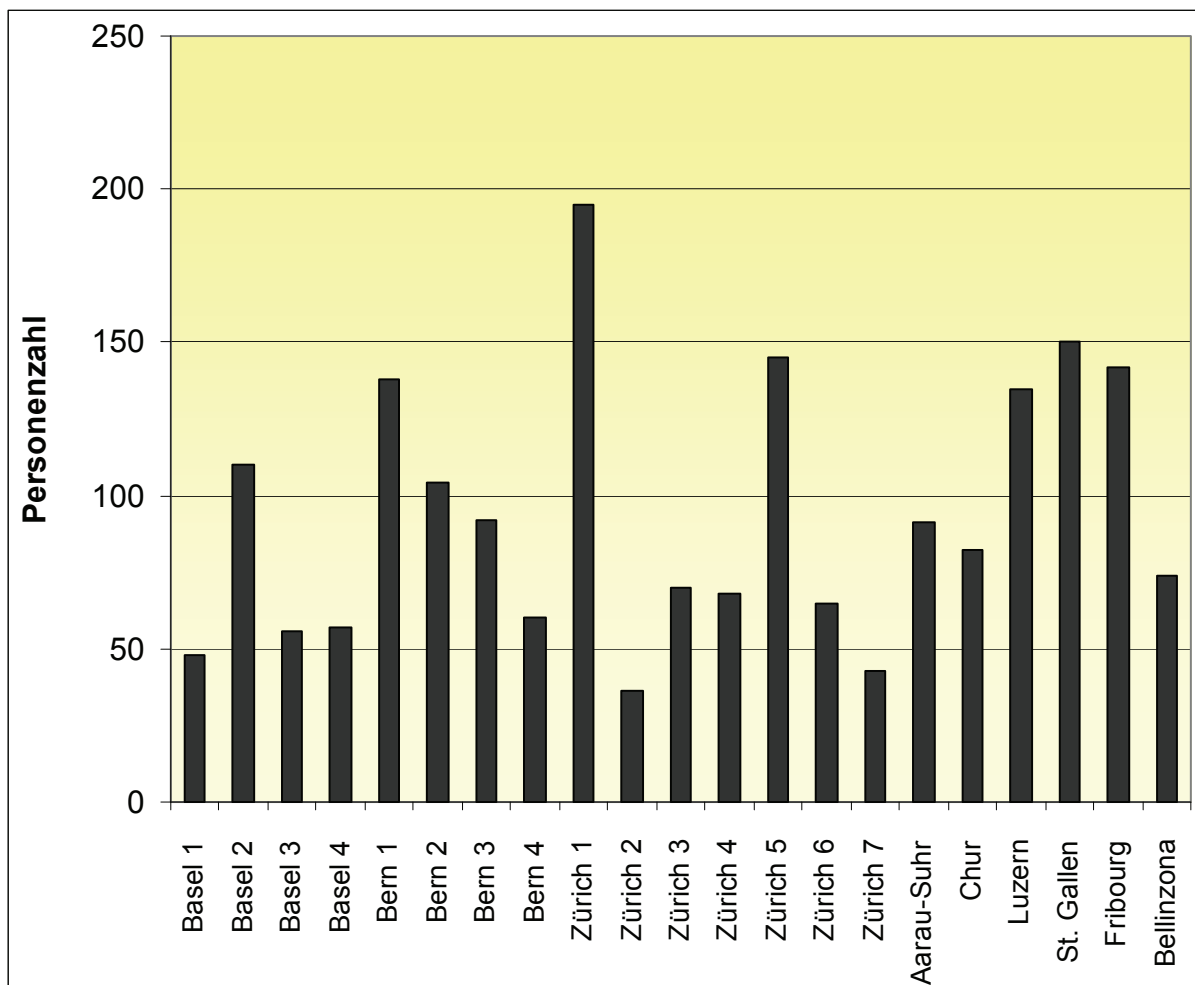


Abbildung 9: Grösse (Zahl der Personen) und Aufteilung der Testlokale 2008.

4.3 Testorte, Wunschuniversitäten und Anreise

Die Wahl des Testortes hat bei der Zulassung keinen Einfluss auf den zukünftigen Studienort. Es wird empfohlen, denjenigen Testort zu wählen, der dem Wohnort am nächsten liegt. Längere Anreisen könnten sich negativ auf die Leistungsfähigkeit auswirken. Da der EMS gewollt auch eine Belastungssituation darstellt, spielen optimale äussere Bedingungen vermutlich auch eine grosse Rolle, um am Testtag die optimal möglichen Leistungen zu erreichen.

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

Testort	Wunschuniversität				Total
	Bern	Basel	Freiburg	Zürich	
BASEL	23 8.5%	222 81.9%	4 1.5%	22 8.1%	271
BELLINZONA	13 17.6%	6 8.1%	14 18.9%	41 55.4%	74
BERN	350 88.8%	11 2.8%	18 4.6%	15 3.8%	394
CHUR	20 24.4%	12 14.6%	3 3.7%	47 57.3%	82
FREIBURG	70 49.3%	4 2.8%	58 40.8%	10 7.0%	142
LUZERN	47 34.8%	42 31.1%	4 3.0%	42 31.1%	135
ST. GALLEN	21 14.0%	30 20.0%	3 2.0%	96 64.0%	150
SUHR	25 27.5%	33 36.3%	0 .0%	33 36.3%	91
ZÜRICH	34 5.5%	30 4.8%	9 1.4%	549 88.3%	622
Total	603	390	113	855	1961

Tabelle 7: Testorte und Wunschuniversität (erste Wahl) 2008, Zeilenprozent: Aufteilung pro Testort auf die Wunschuniversitäten

4.4 Wunschuniversität und Testort nach Wohnkanton

In der folgenden Tabelle sind die Testorte nach Wohnkanton für die einzelnen Disziplinen aufgeschlüsselt. Angegeben sind auch die Vergleichswerte seit 1998 bzw. für die Jahre mit NC.

Wohnkanton/ Wohnort	Humanmedizin 2008: gewünschter Studienort					Total									
	BE	BS	FR	ZH	Total	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998
AG	22	50	0	63	135	124	99	81	93	54	52	42	62	62	66
AI	0	0	0	0	0	3	4	1	4	0	2	1	3	0	2
AR	3	4	0	8	15	12	7	4	8	6	6	8	4	6	4
BE	223	4	8	7	242	204	206	174	179	135	139	148	88	87	101
BL	2	70	0	5	77	81	56	64	54	44	37	41	38	52	50
BS	1	81	2	5	89	79	77	61	48	41	35	31	26	37	40
FR	8	1	48	4	61	46	45	54	42	44	28	24	33	39	34
GE	1	1	0	1	3	2	1	0	0	1	1	1	0	0	1
GL	0	0	1	5	6	1	7	5	1	6	3	3	3	4	4
GR	8	9	2	29	48	44	40	34	41	46	18	30	16	22	19
JU	1	1	1	0	3	5	1	0	3	1	0	0	0	1	0
LU	37	41	5	29	112	98	70	65	53	52	62	42	29	34	43
NE	1	0	2	1	4	1	0	3	1	1	1	0	0	1	2
NW	3	3	0	6	12	12	2	4	5	9	4	3	2	3	6
OW	3	2	0	3	8	5	6	5	5	2	7	3	9	2	3
SG	18	23	3	73	117	111	86	81	55	48	50	41	51	51	64
SH	2	6	1	9	18	15	9	11	6	10	10	15	5	4	7
SO	33	17	0	13	63	37	55	42	42	38	35	21	13	22	26
SZ	6	1	2	18	27	26	17	19	18	11	11	15	16	16	21
TG	4	7	3	24	38	40	42	38	24	22	28	30	20	16	22
TI	6	5	10	23	44	50	45	26	34	30	18	31	21	23	30
UR	3	2	0	3	8	9	8	3	1	3	4	4	2	1	3
VD	1	1	4	2	8	10	7	3	1	2	1	2	2	0	3
VS	12	3	7	3	25	22	20	10	7	5	13	12	11	12	22
ZG	2	1	1	21	25	27	30	20	17	12	16	19	6	4	6
ZH	8	6	1	305	320	283	295	270	263	234	178	130	169	159	165
FL	0	0	0	3	3	5	7	9	5	3	3	5	4	2	3
Übriges Ausland	9	4	2	9	24	22	21	20	16	11	8	7	4	13	3
Total	417	343	103	672	1535	1374	1263	1107	1026	871	770	709	637	673	750

Tabelle 8: Gewünschte Studienorte Humanmedizin nach Wohnkantonen 2008 und Vergleich mit 1998 bis 2007. Das Fürstentum Liechtenstein (FL) wird als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

Wohnkanton/ Wohnort	Veterinärmedizin 2008: gewünschter Studienort			Total								
	Bern	Zürich	Total	2007	2006	2005	2004	2003	2002	2001	2000	1999
AG	5	9	14	16	27	25	11	13	8	7	9	22
AI	0	1	1	1	0	0	1	0	3	0	1	0
AR	0	1	1	1	2	2	2	2	0	2	0	0
BE	38	2	40	35	38	24	29	32	30	39	26	26
BL	7	3	10	15	7	10	12	8	7	4	3	7
BS	4	3	7	7	1	9	4	2	4	7	2	2
FR	17	2	19	10	8	9	6	8	9	3	11	9
GE	6	0	6	12	13	6	2	4	6	9	11	13
GL	0	3	3	2	2	4	0	1	1	0	2	0
GR	3	6	9	9	10	6	10	13	4	6	3	7
JU	5	0	5	8	7	0	4	4	2	1	1	1
LU	7	7	14	16	14	6	11	7	13	5	12	10
NE	8	0	8	6	6	10	4	7	6	3	6	4
NW	1	0	1	3	3	2	1	2	2	0	0	0
OW	1	0	1	1	1	0	0	0	0	0	0	0
SG	3	14	17	20	12	19	14	13	9	7	4	10
SH	0	2	2	6	1	1	1	2	0	3	3	0
SO	6	3	9	6	10	7	7	7	8	4	7	6
SZ	0	4	4	5	6	2	4	2	4	0	2	2
TG	0	10	10	12	12	8	5	11	6	8	3	4
TI	5	12	17	11	8	10	6	9	9	7	6	6
UR	0	0	0	1	1	1	2	1	1	1	0	0
VD	21	1	22	19	25	31	15	15	13	21	15	11
VS	8	0	8	5	7	5	7	4	7	7	7	3
ZG	1	5	6	5	7	3	4	6	3	4	3	6
ZH	1	31	32	50	36	39	29	42	46	19	25	32
FL	1	0	1	0	1	0	0	0	1	2	0	0
Übriges Ausland	2	3	5	6	3	5	1	2	1	2	2	1
Total	150	122	272	288	268	244	196	217	203	171	164	182

Tabelle 9: Gewünschte Studienorte Veterinärmedizin nach Wohnkantonen 2008 und Vergleich mit 1999 bis 2007. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

Wohnkanton/ Wohnort	Zahnmedizin 2008: gewünschter Studienort					Total			
	BE	BS	FR	ZH	Total	2007	2006	2005	2004
AG	2	6	0	3	11	11	14	10	14
AI	0	0	0	2	2	1	0	1	0
AR	0	0	0	0	0	1	2	2	4
BE	19	1	1	1	22	31	32	27	31
BL	0	12	0	0	12	11	5	11	10
BS	0	11	0	0	11	5	7	6	4
FR	0	0	4	0	4	6	3	4	1
GL	0	0	0	0	0	2	1	3	0
GR	0	0	1	3	4	6	5	3	8
JU	0	1	1	0	2	2	0	0	0
LU	3	3	0	5	11	9	10	5	2
NE	0	0	0	0	0	1	1	0	0
NW	1	1	0	0	2	0	0	1	1
OW	0	0	0	1	1	0	0	0	2
SG	2	5	0	10	17	17	13	12	4
SH	1	0	0	2	3	0	0	5	3
SO	3	5	0	2	10	13	10	2	9
SZ	1	0	0	3	4	3	4	1	2
TG	0	0	1	3	4	2	4	9	6
TI	1	0	2	2	5	10	4	3	8
UR	1	0	0	0	1	3	2	2	0
VD	0	0	0	1	1	0	0	0	0
VS	1	0	0	0	1	4	5	4	4
ZG	0	0	0	3	3	2	2	1	1
ZH	0	0	0	15	15	25	28	33	33
FL	0	0	0	4	4	1	1	2	2
Übriges Ausland	1	2	0	1	4	3	4	3	2
Total	36	47	10	61	154	169	157	150	151

Tabelle 10: Gewünschte Studienorte Zahnmedizin nach Wohnkantonen 2008 und Vergleich mit 2004 bis 2007. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

Die nachfolgende Tabelle zeigt, welche Testorte aus den einzelnen Kantonen (für alle Disziplinen zusammengefasst) gewählt werden und dient auch der Kapazitätsplanung für die Testorte.

Wohnkanton	Testort									
	Basel	Bellinzona	Bern	Chur	Freiburg	Luzern	St. Gallen	Suhr	Zürich	Total
AG	16	0	3	0	1	0	0	70	70	160
AI	0	0	0	0	0	0	2	0	1	3
AR	0	0	0	0	0	0	15	0	1	16
BE	2	0	294	0	6	0	1	0	1	304
BL	98	0	0	0	0	0	0	0	1	99
BS	105	0	0	0	0	0	0	0	2	107
FR	1	1	22	0	60	0	0	0	0	84
GE	0	0	0	0	9	0	0	0	0	9
GL	0	0	0	0	0	0	0	0	9	9
GR	0	4	2	52	0	0	0	0	3	61
JU	1	0	1	0	8	0	0	0	0	10
LU	4	0	7	1	0	111	0	8	6	137
NE	0	0	1	0	11	0	0	0	0	12
NW	0	0	0	0	0	13	0	0	2	15
OW	0	0	0	0	0	8	0	0	2	10
SG	0	0	0	19	0	0	106	0	26	151
SH	2	0	1	0	0	0	0	0	20	23
SO	29	0	36	0	0	0	0	10	7	82
SZ	2	0	0	1	0	3	0	0	29	35
TG	3	0	1	0	0	0	20	1	27	52
TI	0	65	0	0	0	0	0	0	1	66
UR	0	0	0	0	0	0	0	0	8	8
VD	0	0	1	0	28	0	0	0	2	31
VS	2	0	16	0	15	0	0	0	1	34
ZG	0	0	1	0	0	0	0	0	33	34
ZH	2	2	2	1	1	0	0	1	359	368
FL	0	0	0	6	0	0	2	0	0	8
Übriges Ausland	4	2	6	2	3	0	4	1	11	33
Total	271	74	394	82	142	135	150	91	622	1961

Tabelle 11: Testorte nach Wohnkantonen 2008. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

4.5 Testabsolvierung nach Alter und Geschlecht

Disziplin	Geburtsjahre	Geschlecht				Gesamt
		männlich		weiblich		
		Anzahl	Anzahl der Zeilen (%)	Anzahl	Anzahl der Zeilen (%)	
Humanmedizin	1957-1984	67	47.5%	74	52.5%	141
	1985-1987	181	47.8%	198	52.2%	379
	1988-1992	351	34.6%	664	65.4%	1015
	Total	599	39.0%	936	61.0%	1535
Veterinärmedizin	1957-1984	6	21.4%	22	78.6%	28
	1985-1987	13	17.6%	61	82.4%	74
	1988-1992	27	15.9%	143	84.1%	170
	Total	46	16.9%	226	83.1%	272
Zahnmedizin	1957-1984	7	50.0%	7	50.0%	14
	1985-1987	32	59.3%	22	40.7%	54
	1988-1992	29	33.7%	57	66.3%	86
	Total	68	44.2%	86	55.5%	154
Gesamt	1957-1984	80	43.7%	103	56.3%	183
	1985-1987	226	44.6%	281	55.4%	507
	1988-1992	407	32.0%	864	68.0%	1271
	Total	713	36.4%	1248	63.6%	1961

Tabelle 12: Testabsolventen nach Geburtsjahr und Geschlecht 2008

Der Männeranteil in Humanmedizin stabilisiert sich weiter bei 40%, der Zuwachs war in diesem Jahr für beide Geschlechter etwa gleich. In Veterinärmedizin bleibt der Geschlechteranteil bei über 80% Frauen. In Zahnmedizin ist das Geschlechterverhältnis wieder auf das Niveau von 2005 und 2006 zurückgekehrt – bei ebenfalls einem Frauenanteil von 60%; ebenso in Humanmedizin.

Betrachtet man gleichzeitig das Alter und das Geschlecht, kommt der Zuwachs bei den Bewerbungen vor allem aus dem Kreis der Jüngeren. Es gibt bisher keine Hinweise darauf, dass sich ein Kreis von „Dauerbewerbern“ bildet.

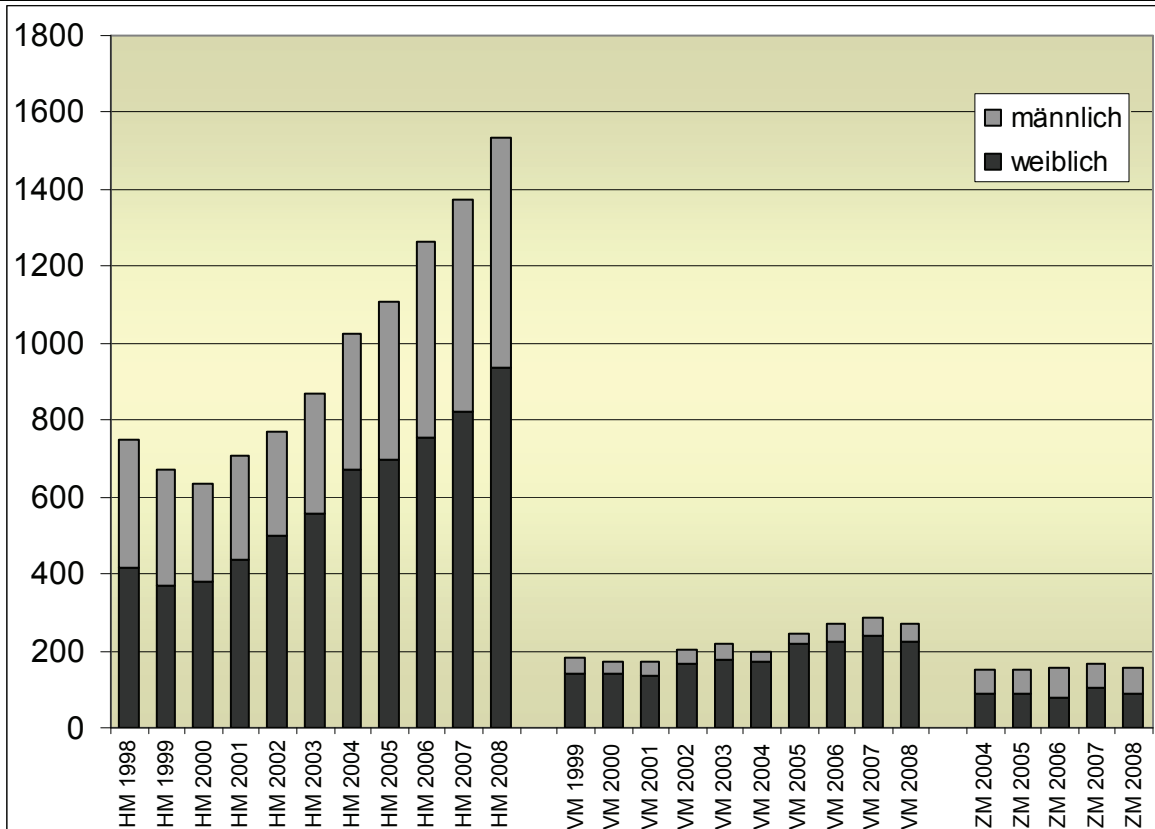


Abbildung 10: Anzahl männlicher und weiblicher Bewerber für Humanmedizin (HM), Veterinärmedizin (VM) und Zahnmedizin (ZM), Jahre mit NC.

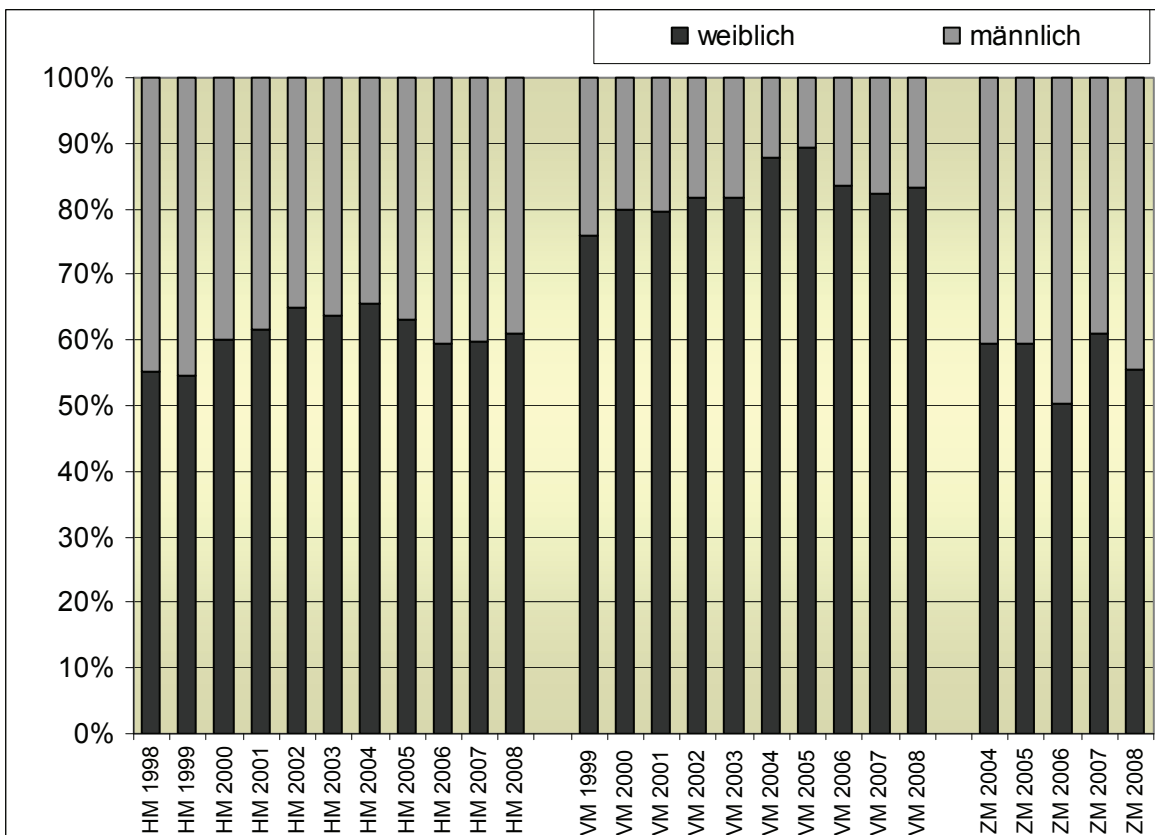


Abbildung 11: Anteil der Geschlechter bei der Bewerbung für Humanmedizin (HM), Veterinärmedizin (VM) und Zahnmedizin (ZM), Jahre mit NC, bezogen auf 100% pro Jahr und Disziplin.

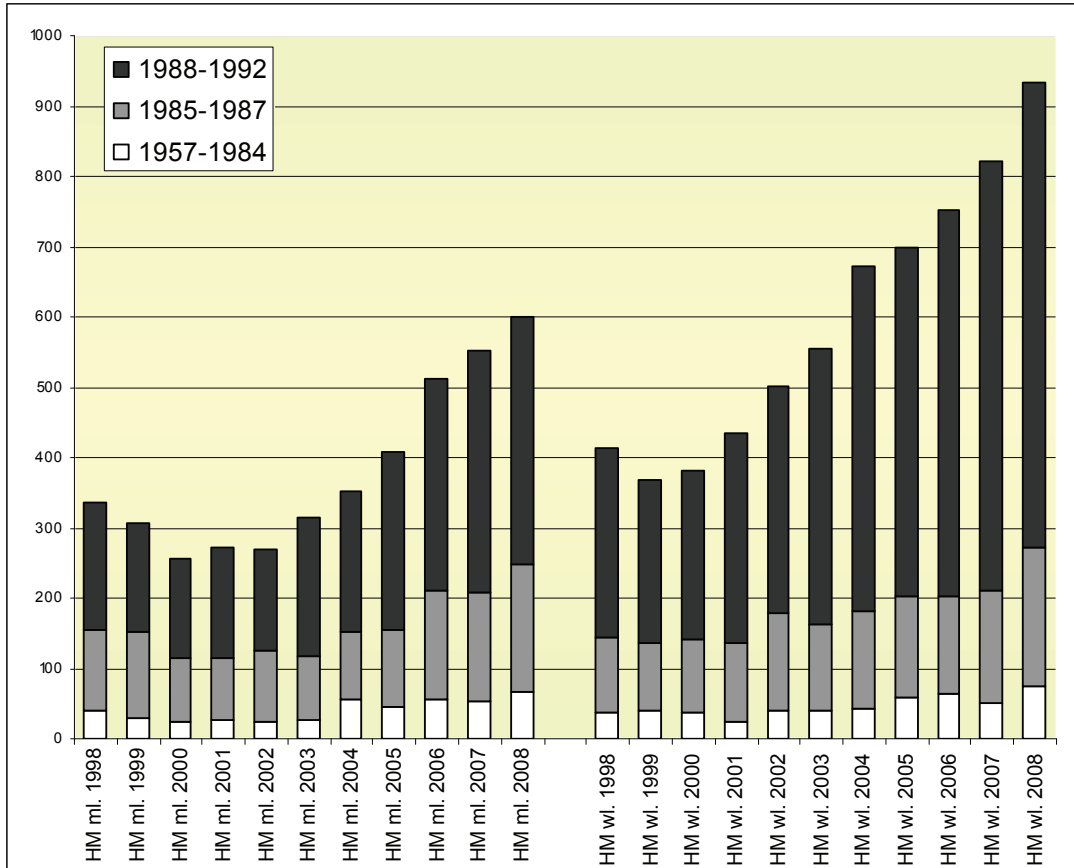


Abbildung 12: Zusammensetzung der Bewerbergruppen nach Geschlecht (mi: männlich, wl: weiblich) und Alter (Geburtsjahre) für Humanmedizin, Absolutzahlen.

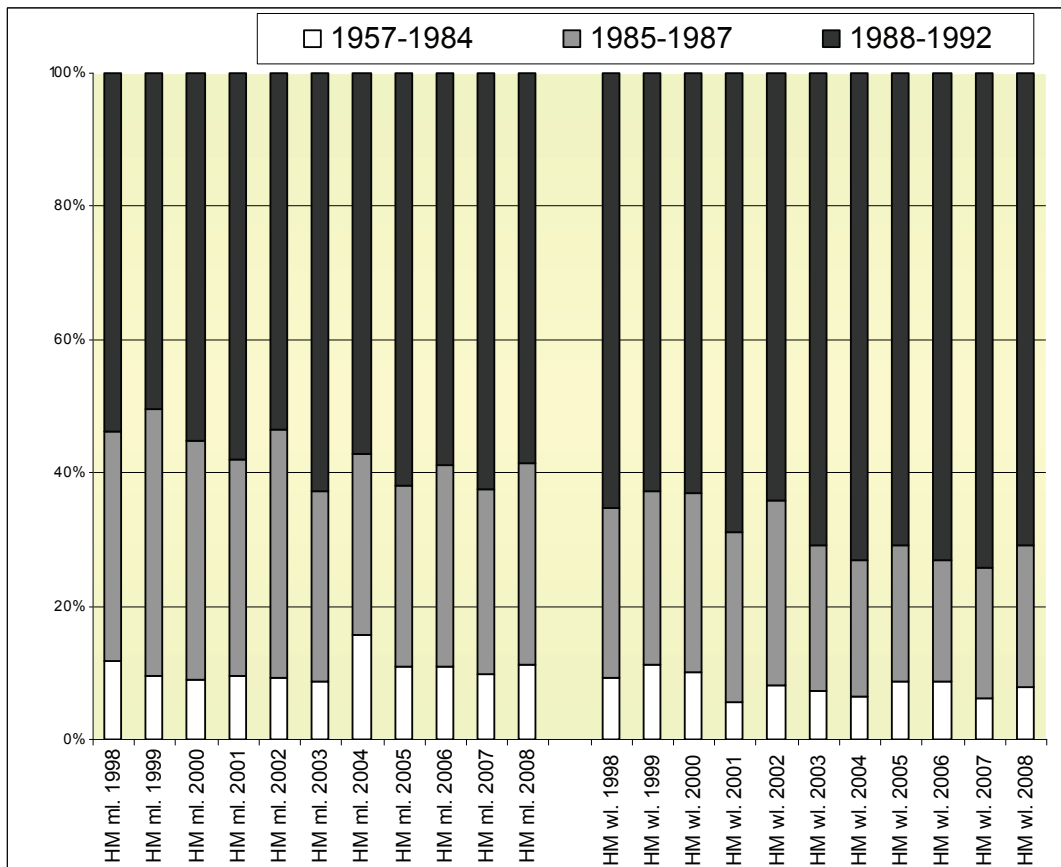


Abbildung 13: Zusammensetzung der Bewerbergruppen nach Geschlecht (mi: männlich, wl: weiblich) und Alter (Geburtsjahre) für Humanmedizin bezogen auf 100% pro Kohorte.

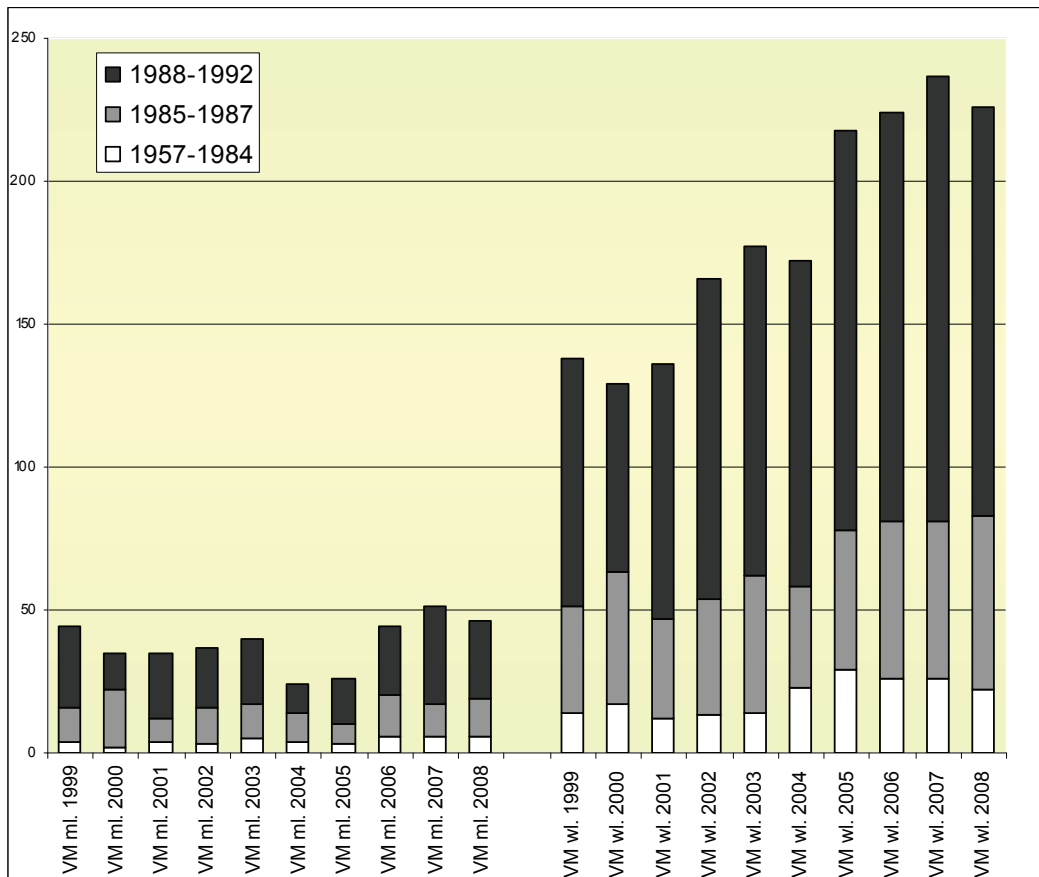


Abbildung 14: Zusammensetzung der Bewerbergruppen nach Geschlecht (ml: männlich, wl: weiblich) und Alter (Geburtsjahre) für Veterinärmedizin, Absolutzahlen.

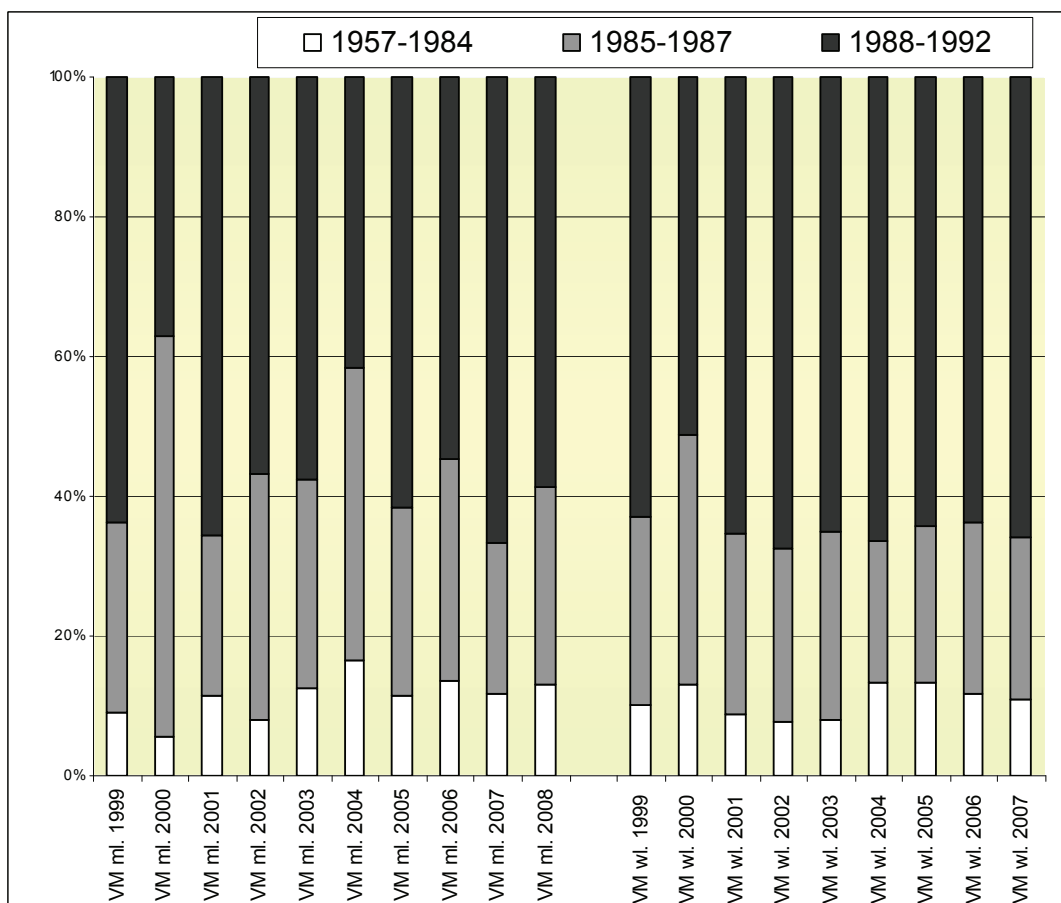


Abbildung 15: Zusammensetzung der Bewerbergruppen nach Geschlecht (ml: männlich, wl: weiblich) und Alter (Geburtsjahre) für Veterinärmedizin bezogen auf 100% pro Kohorte.

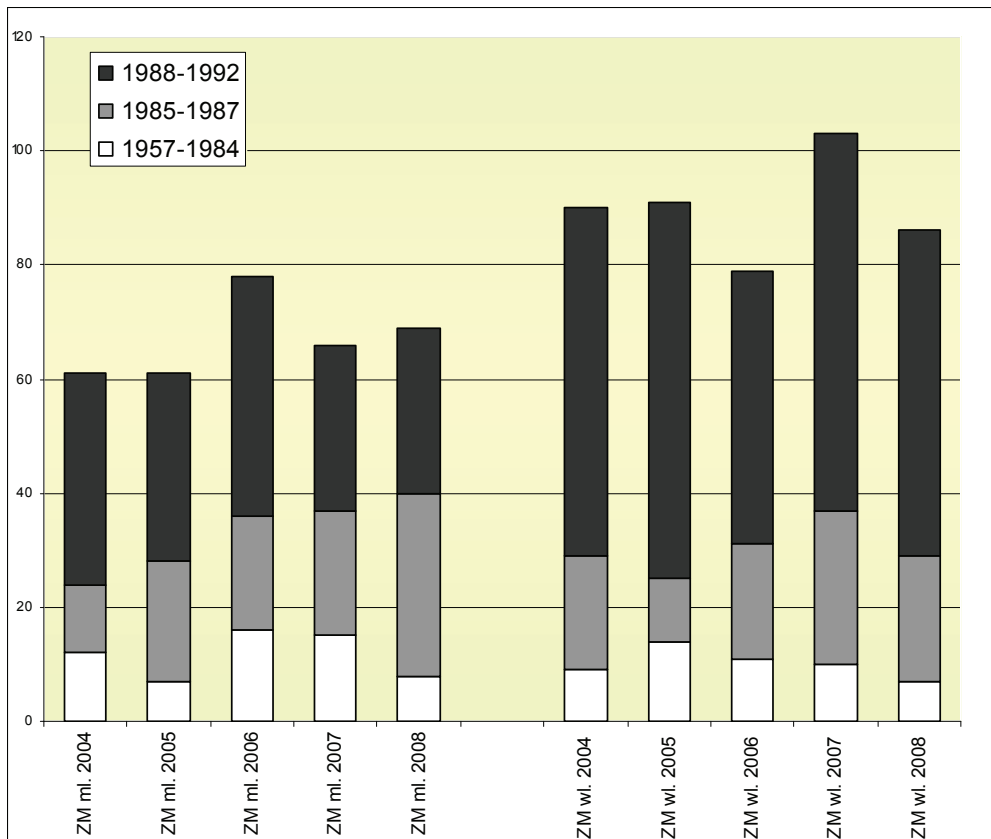


Abbildung 16: Zusammensetzung der Bewerbergruppen nach Geschlecht (ml: männlich, wl: weiblich) und Alter (Geburtsjahre) für Zahnmedizin, Absolutzahlen.

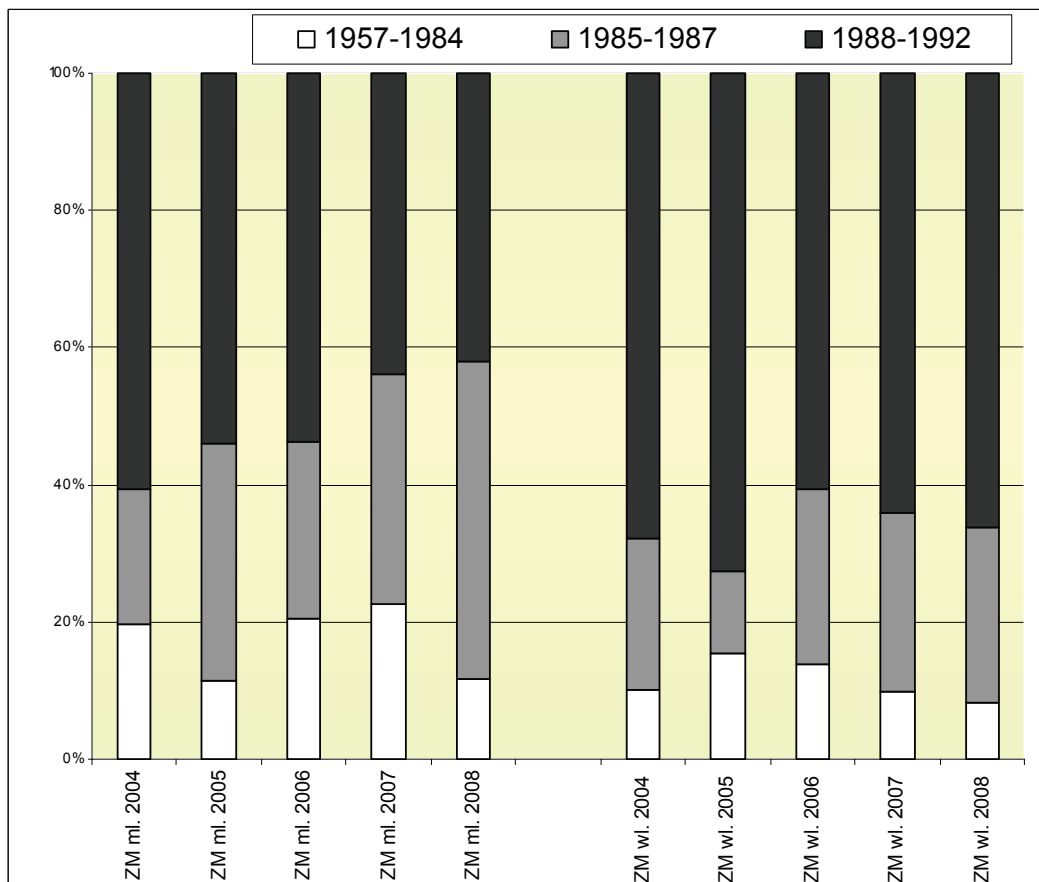


Abbildung 17: Zusammensetzung der Bewerbergruppen nach Geschlecht (ml: männlich, wl: weiblich) und Alter (Geburtsjahre) für Zahnmedizin bezogen auf 100% pro Kohorte.

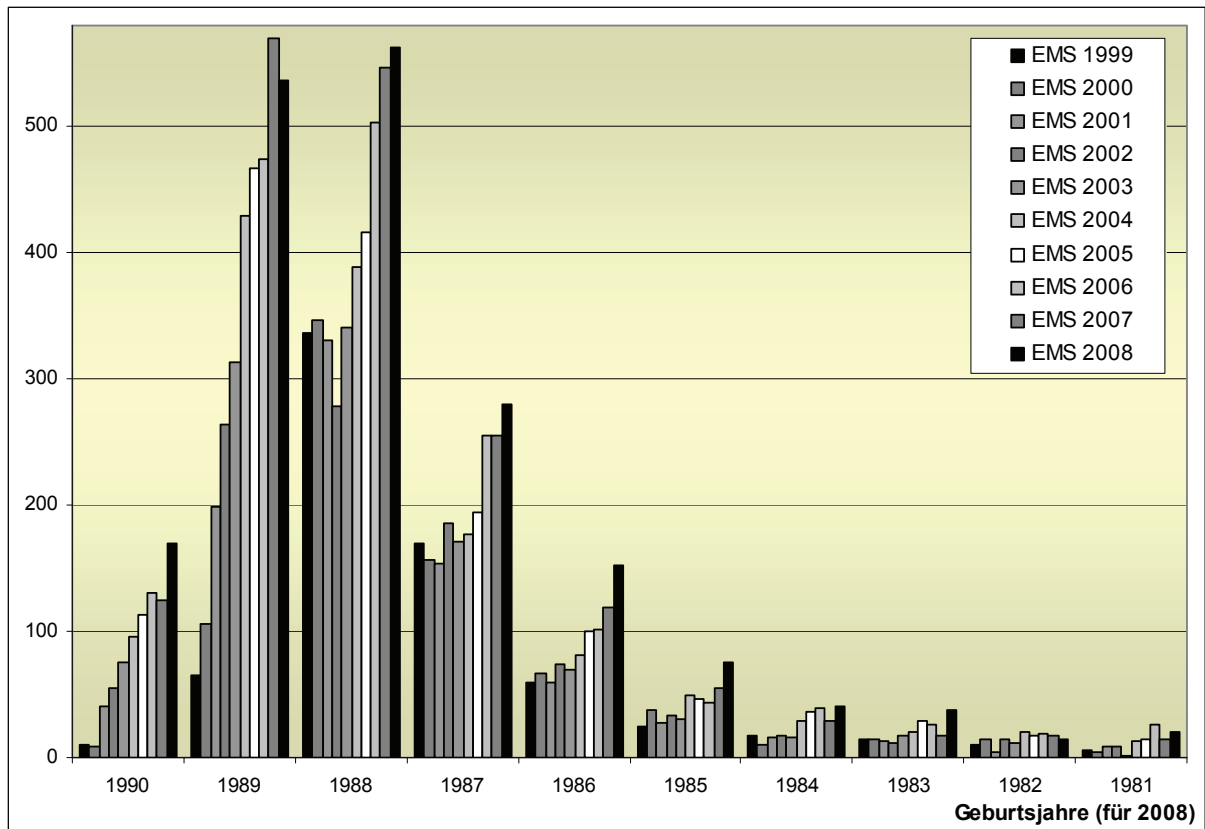


Abbildung 18: Bewerbungen nach Geburtsjahren für 2008 und Entsprechungen für die Jahre 1999 bis 2007 (jeweils um entsprechende Jahre versetzt).

4.6 Übernahme des Testergebnisses aus Vorjahren

Bekanntlich kann man das Ergebnis des Vorjahres in das Folgejahr übertragen lassen oder neu zum Test antreten, wobei dann immer das neue Testergebnis zählt.

Die Zahl der Übernahmen der Testergebnisse aus dem Vorjahr bleibt in etwa konstant (Tabelle 13). Diese Übernahme erfolgt in der Regel durch Personen, die einen für die Zulassung ausreichenden Testwert aufweisen, allerdings noch einmal ein Jahr aussetzen möchten. Das Risiko besteht darin, dass durch die Zunahme der Bewerbungen ein Testwert im Folgejahr nicht mehr für die Zulassung ausreichend sein kann. Vor allem in Humanmedizin kam das vor. Dieses Risiko scheint nun aber ausreichend bekannt zu sein, denn nur für je eine Person bei Human- und bei Veterinärmedizin ist dieser Fall in diesem Jahr aufgetreten.

Deshalb treten viele Personen lieber erneut zum EMS an – vielleicht auch, weil die Wahrscheinlichkeit für eine Verbesserung beim zweiten Testantritt bisher immer ziemlich hoch war. In diesem Jahr hat sich der Anteil von Personen, die sich verschlechtert haben oder gleich geblieben sind allerdings relativ stark erhöht. Eine Ursache für Verbesserungen scheint zu sein, dass man sich auf die erste Teilnahme nicht ausreichend vorbereitet hatte und dann bei der zweiten Testteilnahme besser vorbereitet an den Start geht. Diese geringeren Verbesserungsmöglichkeiten deuten auch auf eine allgemein bessere Vorbereitung auf den Test bereits bei der ersten Testabnahme hin.

Disziplin	Jahr	N	Minimum	Maximum	Mittelwert	Stand.-abw.	Zugelassen 1. Runde
Human- medizin	2000	33	85	119	102.9	7.25	33
	2001	27	94	119	103.0	7.28	27
	2002	40	83	118	104.0	9.16	40
	2003	47	91	121	103.7	8.27	44 (3 nicht)
	2004	37	90	119	105.9	7.70	31 (6 nicht)
	2005	41	73	120	105.7	8.22	35 (6 nicht)
	2006	45	94	128	109.9	7.23	41 (4 nicht)
	2007	33	94	126	111.4	7.7	29 (4 nicht)
	2008	52	100	125	110.2	5,6	51 (1 nicht)
Vete- rinär- medizin	2000	5	97	107	102.0	3.81	5
	2001	5	89	112	102.8	9.52	5
	2002	3	94	109	101.3	7.51	3
	2003	11	86	118	102.9	9.0	10 (1 nicht)
	2004	13	80	106	99.1	7.1	12 (1 nicht)
	2005	8	85	117	105.2	10.93	6 (2 nicht)
	2006	10	90	109	102.1	5.56	8 (2 nicht)
	2007	7	94	112	102.7	6.2	6 (1 nicht)
	2008	15	95	122	104.6	6.7	14 (1 nicht)
Zahn- medizin	2004	1			93.0		1
	2005	11	78	99	91.7	7.10	11
	2006	20	89	117	96.7	6.47	20
	2007	32	89	106	98.2	4.11	31 (1 nicht)
	2008	13	93	103	98.8	3.1	13

Tabelle 13: Testwertstatistiken für Personen mit Übernahme des Testergebnisses aus dem Vorjahr. Zugelassen 1. Runde sind Personen, die sofort einen Platz zugeteilt erhielten (ohne Nachrücker).

4.7 Erneute Testteilnahme

Es ist möglich, erneut zum Test anzutreten und dann jeweils das neuere Ergebnis für die Zulassung zu verwenden. Personen, die sich später als ein Jahr nach der ersten Testteilnahme erneut bewerben, müssen in jedem Falle noch einmal am Test teilnehmen. Vom Testergebnis hängt auch ab, ob ein Studium an der Wunschuniversität aufgenommen werden kann oder ggf. Umleitungen erfolgen. Es treten deshalb auch Personen erneut an, um ihr Ergebnis zu verbessern. Wegen der Verschärfung des NC nimmt die Zahl der erneuten Antritte auch dieses Jahr zu (Tabelle 14) – das Risiko, dass ein knapp über der Zulassungsgrenze liegender Wert in einem Jahr dann im nächsten Jahr nicht mehr reicht, wird grösser. Bisher ist die

Wahrscheinlichkeit eines besseren 2. Testergebnisses deutlich höher. Dieser Unterschied zwischen erster und zweiter Messung hat in diesem Jahr etwas abgenommen (Tabelle 15).

Erste Testteilnahme	Zweite Testteilnahme									
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1998	12	2	1	1	1	1	0	0	9	0
1999		19	3	0	0	2	1	1	2	2
2000			12	0	3	1	1	1	1	0
2001				11	3	1	2	1	0	1
2002					32	13	1	3	1	0
2003						78	21	10	3	2
2004							151	26	15	7
2005								184	34	26
2006									257	58
2007										295
Gesamt	12	21	16	12	39	96	177	226	313	391

Tabelle 14: Wiederholte Testteilnahme: Anzahl der Personen.

Zeitraum	N	1. Testung		2. Testung		Differenz	
		m	s	m	s	m	s
1998-2003	100	86.9	7.8	96.6	9.3	9.7	6.5
2004	96	89.0	7.0	97.0	9.9	8.0	8.2
2005	177	93.7	6.6	101.2	8.4	7.6	9.3
2006	226	92.4	6.9	100.7	8.8	8.3	7.9
2007	313	93.2	7.0	101.9	9.3	8.7	7.9
2008	391	93.9	7.4	100.4	8.5	6.4	7.9
Gesamt	1303	92.5	7.4	100.4	9.0	7.8	8.1

Tabelle 15: Testwertvergleich zwischen erster und zweiter Testteilnahme 1998 bis 2008.

Die Tatsache, dass bei wiederholter Testung bessere Ergebnisse auftreten, spricht für die Empfehlung, im Rahmen der Vorbereitung einmal einen Probelauf unter „ernstfallnahen“ Bedingungen durchzuführen. Ein solcher wird in einigen Kantonen unter geeigneten Bedingungen angeboten. Es wäre denkbar, dass diejenigen Personen bevorzugt ein zweites Mal zum Test antreten, die sich auf die erste Teilnahme nicht ausreichend vorbereitet haben.

		Jahr						gesamt
		1998-2003	2004	2005	2006	2007	2008	
Testwert-differenz	verschlechtert	4	15	30	25	36	64	174
	gleich geblieben	0	3	8	6	8	17	42
	verbessert	96	78	139	195	269	310	1087
Gesamt		100	96	177	226	313	391	1303

Tabelle 16: Zahl der Verbesserungen/Verschlechterungen des Testwerts bei Testwiederholungen.

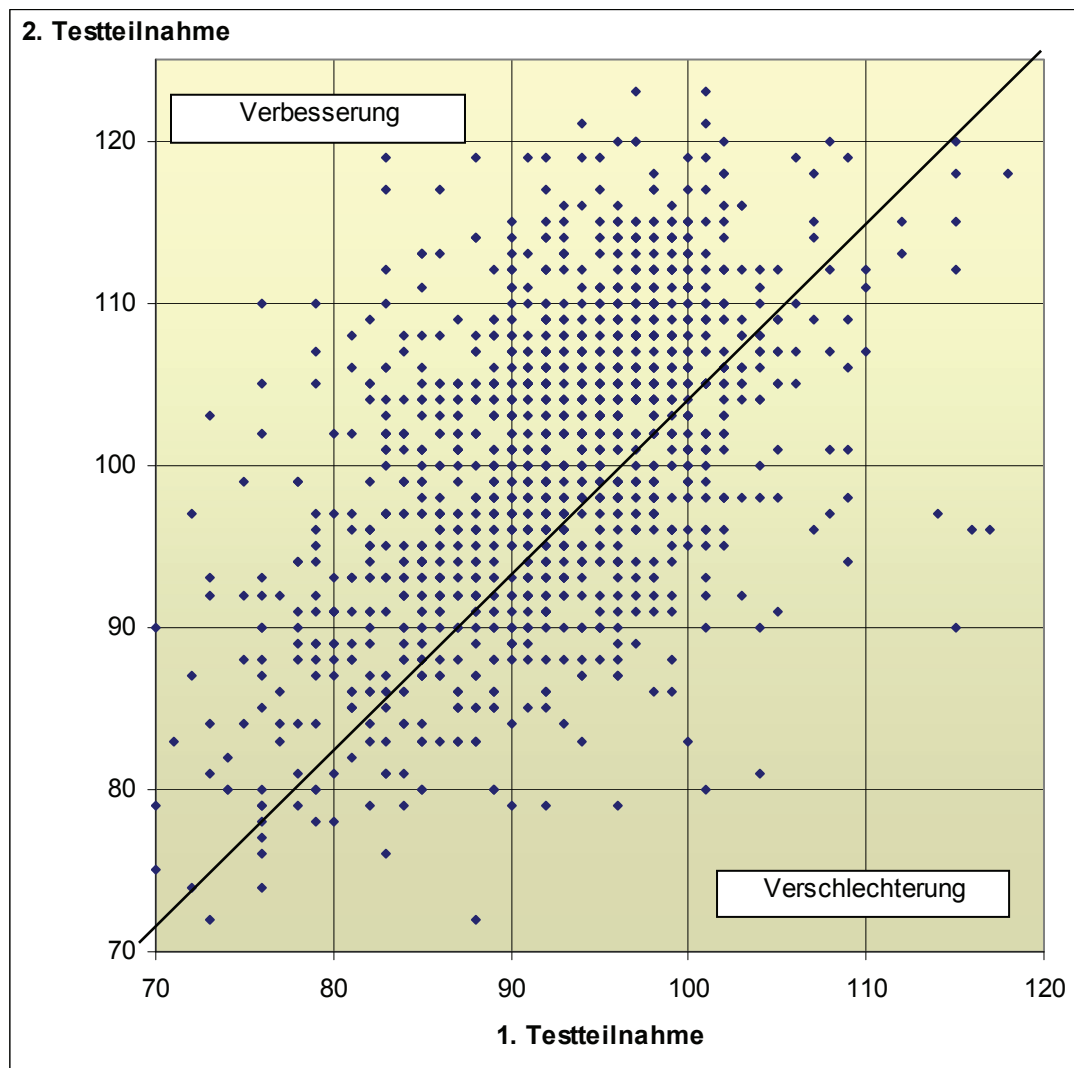


Abbildung 19: Testwerte für Personen, die zweimal am Test teilnahmen (1998 bis 2008).

4.8 Sprachgruppen

Nach wie vor treten vor allem deutschsprachige Kandidaten zum EMS an; für die französisch- und italienischsprachigen Kandidaten existieren Alternativen, das Studium ohne NC in der gewünschten Sprache vollständig zu absolvieren. Die Zahl der französisch- und italienischsprachigen Kandidaten bleibt über die Jahre in etwa vergleichbar.

Der Chancengleichheit der Sprachgruppen wird wie bisher grosse Aufmerksamkeit gewidmet:

- Aufwendiges und angemessenes Verfahren zur Herstellung der Äquivalenz der Sprachversionen (die Evaluation zeigt, dass sprachbedingte Unterschiede nicht wahrscheinlich sind),
- vergleichbares Vorbereitungsmaterial in der jeweiligen Sprache (Test-Info, veröffentlichte Originalversion),
- seit 2005 verstärkte Orientierung auf eine Bearbeitung der veröffentlichten Originalversionen in der jeweiligen Sprache. Das ZTD unterstützt dies durch eine Kurz-Anleitung für die Testleiter dieser Probeläufe, die den Schulen zur Verfügung gestellt wird.

			Geburtsjahre			Total
			1961-1982	1983-1985	1986-1989	
Humanmedizin	deutsch	männlich	65	157	310	532
		weiblich	73	184	623	880
		Total	138	341	933	1412
	französisch	männlich	2	17	19	38
		weiblich	1	10	24	35
		Total	3	27	43	73
	italienisch	männlich	0	7	22	29
		weiblich	0	4	17	21
		Total	0	11	39	50
Veterinärmedizin	deutsch	männlich	5	10	15	30
		weiblich	19	50	91	160
		Total	24	60	106	190
	französisch	männlich	0	3	7	10
		weiblich	3	10	43	56
		Total	3	13	50	66
	italienisch	männlich	1	0	5	6
		weiblich	0	1	9	10
		Total	1	1	14	16
Zahnmedizin	deutsch	männlich	7	29	23	59
		weiblich	7	22	55	84
		Total	14	51	78	143
	französisch	männlich	0	1	0	1
		weiblich	0	0	2	2
		Total	0	1	2	3
	italienisch	männlich	0	2	6	8
		weiblich	0	0	0	0
		Total	0	2	6	8

Tabelle 17: Sprachgruppen der Teilnehmenden 2008 nach Disziplin, Alter und Geschlecht aufgeschlüsselt.

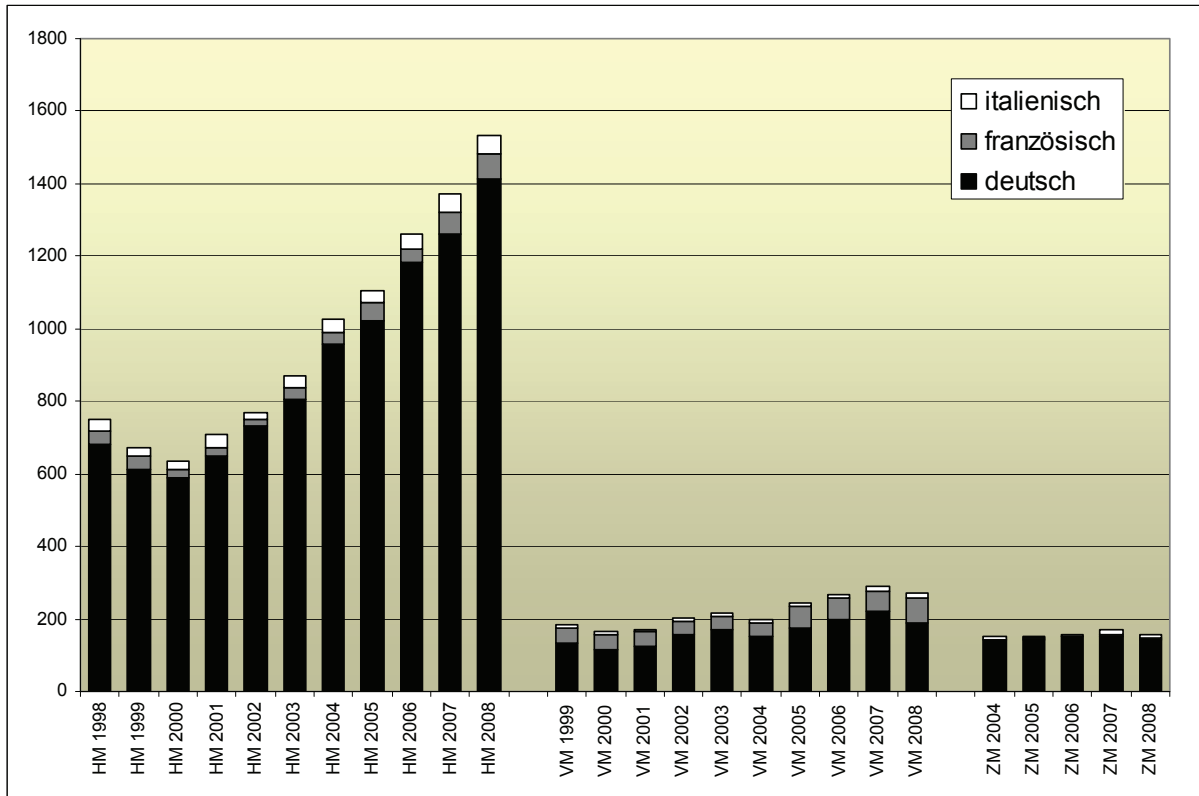


Abbildung 20: Vergleich der Sprachgruppen 1998 bis 2008, Absolutzahlen.

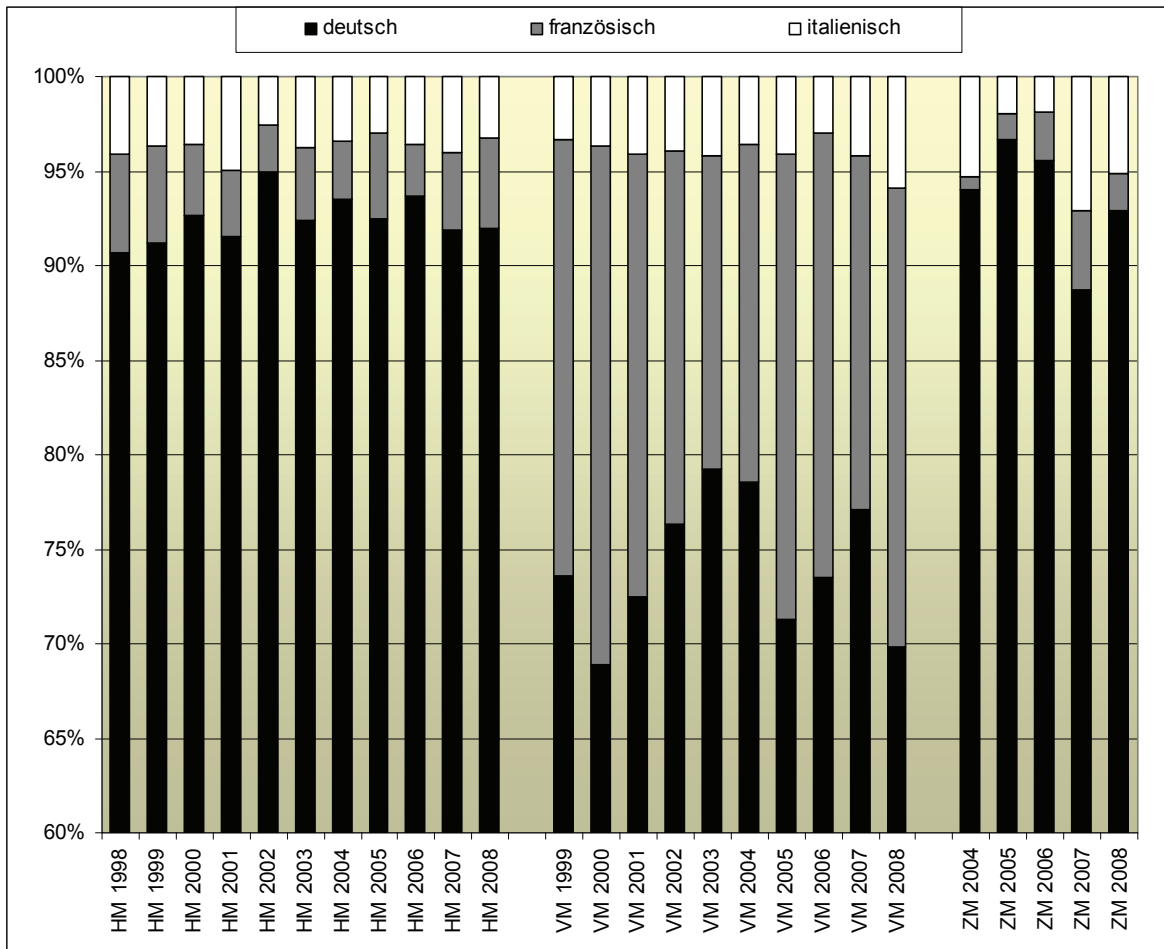


Abbildung 21: Vergleich der Sprachgruppen 1998 bis 2008 Bewerber pro Disziplin und Jahr auf 100% bezogen (dargestellt nur Prozentband zwischen 60 und 100%).

4.9 Alter und Maturitätsjahr

Die Mehrzahl der Personen hat sich gleich nach dem Mittelschulabschluss bzw. im darauf folgenden Jahr zum Medizinstudium beworben.

Die Personen werden später analog der Vorjahre zu Vergleichszwecken in drei Altersgruppen aufgeteilt. Die älteste Gruppe wird wiederum danach unterschieden, ob die Maturitätsprüfung unmittelbar nach der Mittelschule absolviert wurde oder noch eine Wartezeit dazwischen bestand.

Geb.- Jahr	Maturitätsjahr																									
	79	80	81	83	87	88	89	90	92	94	95	96	97	98	99	00	01	02	03	04		05	06	07	08	
59		1																						1	2	
60				1																					1	
61	2	1																							3	
62			1																						1	
63				2																			1		3	
66										1															1	
67													1												1	
68						1																			1	
70							1																		1	
71								1					1										1		3	
72					1				1	1								1					1		5	
73																1	1							2	4	
74										2													1		3	
75											1		1												2	
76										1	1		1	1											4	
77												1	1	1											2	
78											2		2	2		1		1					3	2	11	
79															1		1		2			1			5	
80														1	3	1	2		1	3	1	1	3	2	18	
81																1	7	3				1	2	6	20	
82																	4	1					1	9	15	
83																	1	11	7		2	2	1	5	8	37
84																		2	10		7	4	3	5	10	41
85																					13	21	13	15	13	75
86																					9	26	51	40	26	152
87																			1			4	76	122	77	280
88																							26	326	210	562
89																				1			98	438	537	
90																		1				1	5	152	159	
91																						1		8	9	
92																								2	2	
	2	2	1	3	1	1	1	1	2	4	2	3	4	5	4	4	16	20	21	35	60	173	628	967	1960	

Tabelle 18: Geburtsjahr und Maturitätsjahr. Dicke Linien trennen die Altersgruppen (bei der ältesten Gruppe zusätzlich eine Teilung nach dem Maturitätszeitpunkt).

5 Beschreibung des verwendeten Eignungstests

5.1 Aufbau des Tests

Der EMS besteht seit 2005 aus zehn Untertests, die zu einem Gesamtwert verrechnet werden, „Planen und Organisieren“ ist als neuer Untertest dazugekommen. Im Jahr 2006 wurde die Reihenfolge optimiert, um den Vormittags- und den Nachmittagsteil in der Dauer anzugleichen.

Die Tabelle 19 gibt einen Überblick über den Testaufbau, die Aufgabenzahl und die Dauer der einzelnen Untertests. Aufgabenbeispiele finden Sie im Anhang.

Bezeichnung der Untertests	Aufgaben	Max. Punktzahl	Bearbeitungszeit
Quantitative und formale Probleme	20	20	50 min
Schlauchfiguren	20	20	12 min
Textverständnis	18	18	45 min
Planen und Organisieren	20	20	60 min
Konzentriertes und sorgf. Arbeiten	Blatt mit 1600 Zeichen	20	8 min
	Pause		1 Std.
<i>Lernphase zu den Gedächtnistests</i>			
Figuren lernen	Es werden die Vorlagen zum Einprägen gezeigt		4 min
Fakten lernen			6 min
Medizinisch-naturw. Grundverst.	20	20	50 min
<i>Gedächtnistests:</i>			
Figuren reproduzieren	20	20	5 min
Fakten reproduzieren	20	20	7 min
Muster zuordnen	20	20	18 min
Diagramme und Tabellen	20	20	50 min
Gesamttest	198	198	ca. 5 Std.
Gesamtdauer (inkl. Pause) 9.00 bis ca. 17.00 Uhr			

Tabelle 19: Struktur und Ablauf des EMS 2008.

Beim EMS handelt es sich um eine Adaptation des deutschen Tests für Medizinische Studiengänge (TMS) mit Weiterentwicklungen für die Schweiz. Die Struktur des TMS mit ursprünglich 9 Untertests hat sich in Deutschland im Rahmen von 12 Testeinsätzen bewährt (1986 zwei Testtermine und von 1987 bis 1996 je ein Termin pro Jahr). In der Schweiz wird der Test seit 1998 erfolgreich eingesetzt – bis 2003 mit identischer Struktur wie beim deutschen TMS.

Seit 2004 wird beim Konzentrationstest jährlich eine Vorlage verwendet, die vorher nicht bekannt ist (Zeichen und Durchstreichregel) – nur der Typ der Anforderung und die Zeitdauer bleiben jeweils gleich. Dadurch werden Effekte von exzessivem Üben vermieden und der Test misst wirklich „Konzentriertes und sorgfältiges Arbeiten“.

Im Jahr 2005 wurde die Teststruktur in der Schweiz erstmals revidiert. Aufgrund einer Anforderungsanalyse der neuen Studienbedingungen wurde ein neuer Untertest „Planen und Organisieren“ aufgenommen. Diese Anforderung wurde in erster Priorität umgesetzt. Damit die Testlänge vergleichbar bleibt, wird seitdem in der Schweiz auf nicht gewertete Einstreuaufga-

ben verzichtet. Sie wurden bis 2004 vor allem beibehalten, um die Testanforderung auch hinsichtlich der Länge mit der deutschen TMS-Anforderung identisch zu halten.

Die Vorteile des EMS lassen sich folgendermassen zusammenfassen:

- Auswahl der Testanforderungen aus einer grösseren Menge möglicher Studienanforderungen aufgrund einer Anforderungsanalyse – ständige Anpassung an sich verändernde Bedingungen;
- Wissenschaftlicher Nachweis der Vorhersagbarkeit von Studienerfolg. Dieser erlaubt die gesetzlich geforderte Studienzulassung nach Eignung;
- Konstruktion der Aufgaben durch Experten UND anschliessende empirische Überprüfung, damit die Aufgaben alle Gütekriterien erfüllen, die gewünschte Fähigkeit tatsächlich messen und optimal „schwierig“ sind;
- Beachtung, dass für die Beantwortung der Aufgaben kein spezielles fachliches Vorwissen notwendig ist, sondern tatsächlich die „Studierfähigkeit“ als aktuelle Fähigkeit zur Wissensaneignung und Problemlösung gemessen wird. Dadurch ist der Test auch wenig trainierbar, was sich positiv auf die Sozialverträglichkeit auswirkt (kein zusätzlicher Gewinn durch zusätzliche Trainingskurse nachgewiesen, wenn die empfohlene Vorbereitung erfolgt);
- Es ist ein ökonomisches und genau kapazitätsentsprechendes Zulassungsverfahren möglich, die Rechtsfähigkeit des Verfahrens hat sich bei mehreren Überprüfungen (auch gerichtlich) bestätigt.

Wie wurden die Aufgaben entwickelt? Ausgangspunkt war eine differenzierte Anforderungsanalyse des Medizinstudiums, an der zahlreiche Lehrbeauftragte und Experten mitarbeiteten. Das Ziel war die Vorhersage des Studienerfolges, um ein faires und wissenschaftlich begründetes Zulassungsverfahren zu erhalten. Gemäss diesen Anforderungen wurden einzelne Aufgabengruppen (Untertests) konstruiert. Neun davon erfüllten alle notwendigen Anforderungen. Das Resultat ist also bereits eine Auswahl bewährter Aufgabentypen aus mehreren möglichen Alternativen. Jedes Jahr wurden neue Aufgaben für die Untertests entwickelt und in mehreren Schritten überarbeitet. An dieser Aufgabenentwicklung nahmen zahlreiche Lehrbeauftragte und Experten teil. Die Aufgaben müssen sehr hohe Qualitätsstandards erfüllen, u.a. (1) müssen sie jedes Jahr die Studieneignung gleich zuverlässig messen, (2) muss das Schwierigkeitsspektrum aller Aufgaben annähernd vergleichbar sein, (3) darf kein spezielles Fachwissen vorausgesetzt werden, um die Trainierbarkeit des Tests gering zu halten und (4) muss eine eindeutige, richtige Lösung existieren.

Die Erprobung neuer Aufgaben erfolgte in Deutschland im Rahmen so genannter „**Einstreuaufgaben**“. Nur bei ausreichender Bewährung wurden solche Aufgaben in nachfolgenden Testversionen für die Werteberechnung verwendet. Im Unterschied zu vielen „Übungsaufgaben“, die im so genannten Trainingsmarkt im Umlauf sind, sind die echten EMS-Aufgaben empirisch geprüft, so dass sie bezüglich Lösungseindeutigkeit und Schwierigkeit optimal sind.

5.2 Berechnung der Werte

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

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

0 und 20 transformiert werden, um mit den anderen Tests gleichgewichtig zum Punktwert addiert zu werden.

Alle Punkte der Untertests werden zu einer Summe addiert (**Punktwert**, vgl. Abbildung 22). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Jahre vergleichbar ist. Deshalb findet eine Standardisierung auf den Mittelwert und die Standardabweichung der jeweiligen Testform statt. Dieser **Testwert** liegt zwischen 70 und 130 (der Mittelwert ist 100) und kann in einen **Prozentrangwert** umgerechnet werden. Prozentränge lassen sich am einfachsten veranschaulichen, indem angegeben wird: x Prozent aller Teilnehmenden haben einen schlechteren Testwert erreicht als die entsprechende Person.

In diesem Jahr können aufgrund guter Kennwerte alle Aufgaben gewertet werden. Die Berechnung und Verwendung von Gütekriterien für die Itemselektion ist ein wichtiges Merkmal eines Tests. Insofern bevorzugen wir intern die Strategie, im Zweifel eine Aufgabe lieber auszuschließen und die Testgüte damit ausreichend hoch zu halten. Das Standardisierungsverfahren zur Bildung des Testwertes gewährleistet auch bei unterschiedlichen Aufgabenzahlen vergleichbare Testwerte zwischen den Jahren.

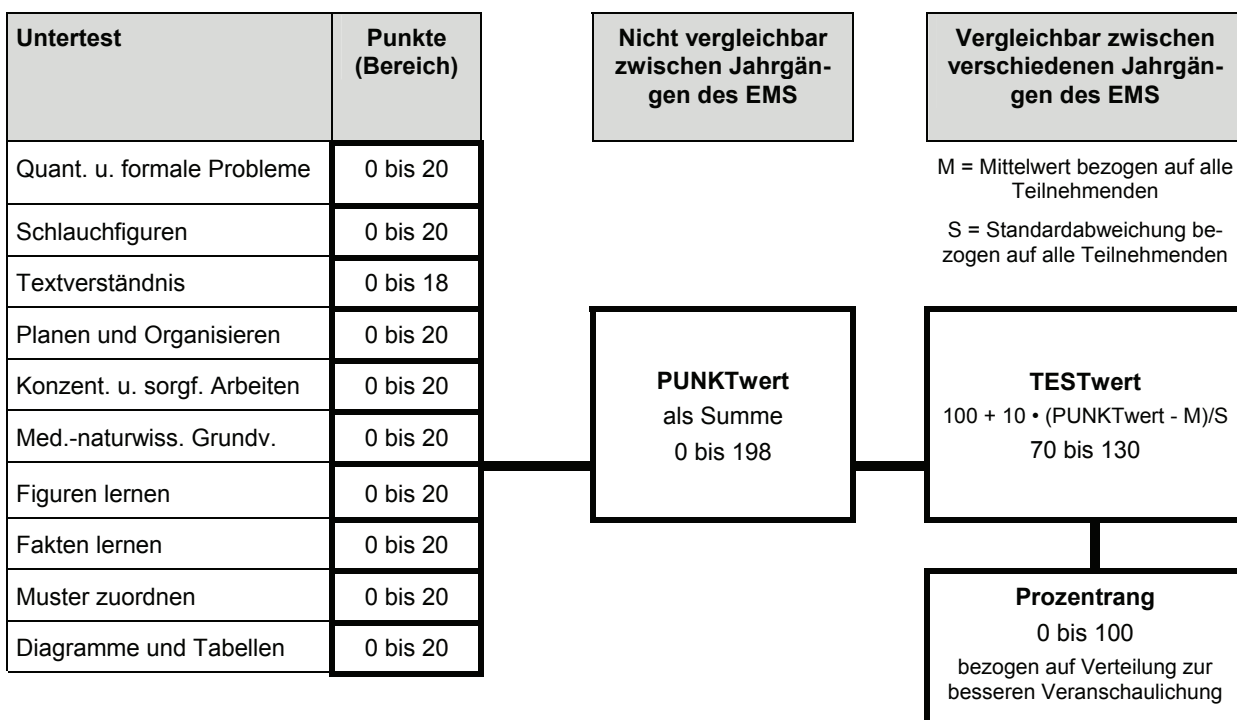


Abbildung 22: Punktwerte der einzelnen Untertests 2008 und ihre Zusammenführung über den Punktwert zum Testwert und Prozentrang.

5.3 Mittlerer Rangplatz der Untertests

Vom Gesetz gefordert ist ein Parameter, welcher die Studieneignung vorhersagt. Dies erfüllt der bisher verwendete Testwert nachweislich. Beim Testwert werden die Punktwerte der einzelnen Untertests addiert und der Gesamtwert wird durch die Standardisierung auf $M = 100$ und $S = 10$ zwischen den Jahren vergleichbar. Das Gewicht der Untertests in diesem Gesamtwert wird letztendlich von der Varianz in diesen Untertests beeinflusst. Da die für die Studieneignung prognoserelevantesten Untertests zugleich die höchste Streuung aufweisen, erhöht dieser Effekt auch die Prognosekraft des Testwertes.

Ziel des EMS ist es, eine kapazitätsentsprechende Vergabe der Studienplätze zu erreichen. Es war vorgekommen, dass mehr Personen den gleichen Testwert aufwiesen als Plätze von der Kapazität her am Ende der Zuteilung noch zur Verfügung standen.

Aus diesem Grunde hat der Beirat EMS beschlossen, ab 2005 weiterhin den Testwert als Zulassungskriterium zu verwenden. Bei **gleichem Testwert** werden die Personen in der Reihenfolge des **mittleren Rangplatzes aller Untertests** berücksichtigt und es werden nur noch so viele Personen zugelassen, wie freie Plätze tatsächlich vorhanden sind. Der mittlere Rangplatz wird ebenfalls auf dem Testbescheid mitgeteilt. In der Praxis gibt es pro Untertest für jeden Punktwert einen Rangplatz-Wert. Der Rangplatz ist umso niedriger, je höher die Punktzahl ist und je weniger Personen insgesamt einen besseren Punktwert erreicht haben. Rangplätze belohnen gute Leistungen in einem schwierigen Untertest stärker, indem die dort vergebenen niedrigen Ränge höheres Gewicht erhalten:

Beispiel: Werden in 2 Untertests maximal 20 Punkte erzielt, gehen immer 40 Punkte in den Punktwert ein. Haben in einem Untertest 100 Personen diesen Wert erreicht, erhalten diese Personen den mittleren Rangplatz 50.5. Haben im anderen Untertest nur 10 Personen diesen Wert erreicht, ist der Rangplatz 5.5. Letzterer wird höher bewertet, weil die 20 Punkte im zweiten Untertest schwerer zu erzielen waren und deshalb wertvoller sind.

Zwecks Vergleichbarkeit der Rangplätze zwischen den Jahren wird ein mittlerer Rangplatz auf der Basis von 1000 Personen verwendet.

- Der Rangplatz bewegt sich 2008 zwischen 42 (Vorjahr 41) und 972 (Vorjahr 962). Niedrige Werte stehen für bessere Leistungen.
- Er korreliert mit dem Testwert mit -0.99 (Vorjahr $-.98$) was praktisch für eine Äquivalenz spricht – siehe auch Abbildung 23.
- Er differenziert allerdings besser zwischen den Personen und lässt eine kapazitätsgenaue Auswahl zu.

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

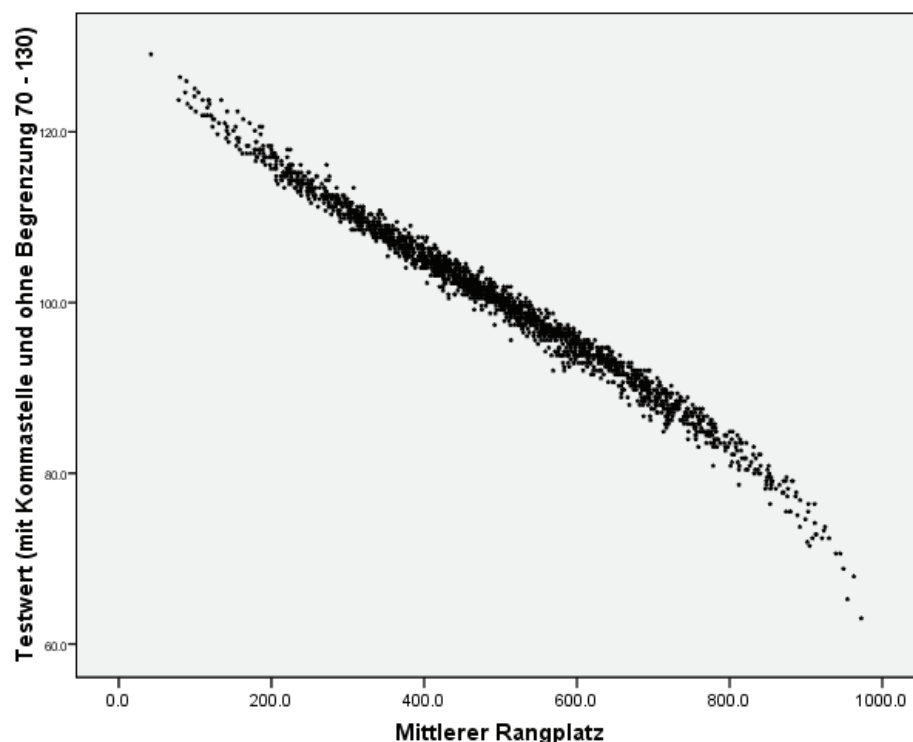


Abbildung 23: Beziehung zwischen Testwert und mittlerem Rangplatz für 2008.

6 Testanwendung in der Schweiz 2008

6.1 Verteilungsprüfung

Die Verteilung des Punktwerts entspricht auch 2008 wiederum einer Normalverteilung. Metrische Verfahren und die Standardisierung zum Testwert durch lineare z-Transformation sind daher zulässig. Die Punktwerte der Einzeltests sind nicht normalverteilt (Tabelle 20).

	Punktwert	Quantitative und formale Probleme	Schlauchfiguren	Textverständnis	Planen und Organisieren	Med.-naturwiss. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme und Tabellen	Konzentr. und sorgf. Arbeiten
Mittelwert	119.29	10.52	14.19	8.76	10.48	11.68	13.74	13.11	13.20	9.86	11.22
Stand.-abweich.	22.29	3.62	3.50	3.27	3.39	3.51	3.72	3.75	3.01	3.16	5.07
Median	121	10	14	9	10	12	14	13	13	10	11
Modalwert	126	10	18	7	10	12	15	13	13	9	10
Spannweite	148	20	18	17	20	20	20	17	20	18	20
25. Perzentil	104	8	12	6	8	9	11	10	11	8	8
75. Perzentil	135	13	17	11	13	14	17	16	15	12	15
K-S: extremste Differenz	0.03	0.07	0.10	0.09	0.08	0.07	0.09	0.07	0.09	0.07	0.07
K-S: Z-Wert	1.31	3.32	4.39	3.77	3.54	3.31	3.82	3.15	3.90	3.24	2.89
K-S: Asymp. Sig.	0.066	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Tabelle 20: Kennwerte der Punktwertskalen und Verteilungsprüfung auf Normalverteilung (K-S: Kolmogorov-Smirnov-Test).

Die in den folgenden Diagrammen dargestellten Verteilungen der Punktwerte für die Untertests zeigen, dass alle Skalen gut bis sehr gut differenzieren. Der Untertest „Schlauchfiguren“ war wiederum der leichteste Untertest (Modalwert 18), der Untertest „Textverständnis“ war auch bei Berücksichtigung des Maximalwerts von 18 (statt 20) Punkten am schwierigsten (Modalwert 7).

Im Untertest „Konzentriertes und sorgfältiges Arbeiten“ sind die Häufungen an beiden Enden der Verteilung auf die Systematik der Umrechnung der Rohwerte in den Normwert zurückzuführen (vgl. die Darstellung dieses Tests ab Seite 85).

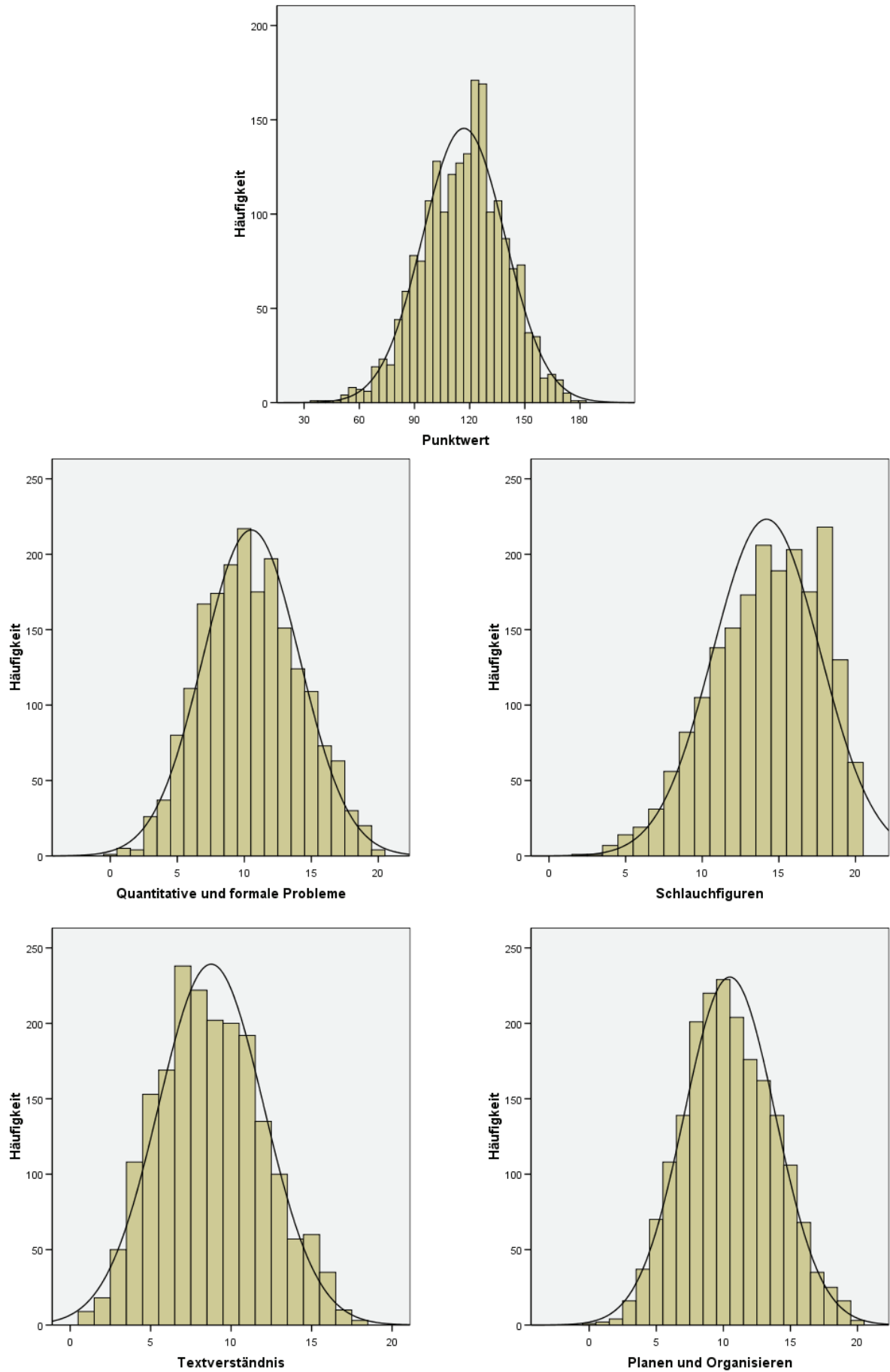


Abbildung 24: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (1).

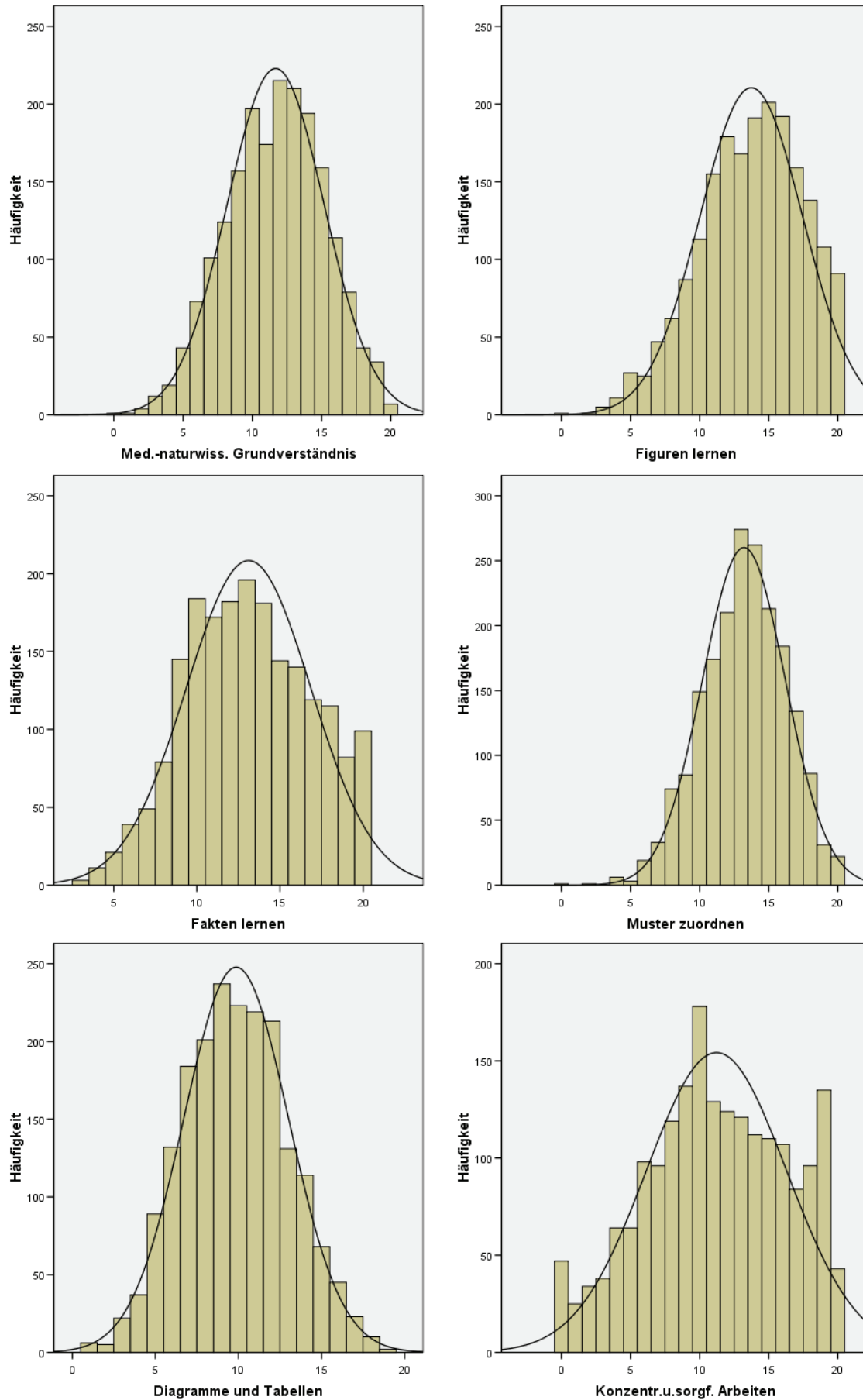


Abbildung 25: Häufigkeitsverteilungen für Punktwert und Punkte der Untertests (2).

6.2 Vergleich der Testfassungen 1998 bis 2008

Durch die Transformation der Punktwerte in Testwerte können diese **Testwerte** aus verschiedenen Jahren direkt miteinander verglichen werden. Dieses Vorgehen wird durch die zwischen den Jahren übereinstimmenden Gütekriterien zusätzlich legitimiert.

Die „absolute“ Schwierigkeit als **Punktwert** unterliegt über die Jahre Schwankungen. Tabelle 21 zeigt den Vergleich für alle bisherigen Testdurchführungen seit 1998 in der Schweiz mit den geschätzten Punktwerten aus Deutschland (Schätzung aufgrund der Item-Schwierigkeiten von Items, die ggf. in einer anderen Reihenfolge und Zusammenstellung angewendet worden sind). Für die Berechnung der Punktwerte in der Schweiz wurden zum Zweck besserer Vergleichbarkeit nur die Ergebnisse der deutschsprachigen Gruppe berücksichtigt:

		2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998
Muster zuordnen	CH	13.3	11.2	11.3	12.7	10.5	10.8	11.3	10.1	10.3	11.6	13.2
	D	11.6	9.3	9.7	10.4	10.3	10.2	10.9	9.7	9.3	11.0	12.2
Med.-naturwiss. Grundverst.	CH	12.0	10.7	10.7	11.3	10.4	10.5	9.9	10.1	10.7	11.3	12.5
	D	11.7	10.8	10.5	11.0	10	10.2	9.5	10.7	10.8	11.2	11.6
Schlauchfiguren	CH	14.5	14.1	13.8	12.3	12.1	12.7	12.1	13.3	13.2	12.9	13.3
	D	10.6	12.3	12.2	10.9	11.8	12	11.8	12.2	12.3	11.9	11.8
Quant. und form. Probleme	CH	10.8	10.9	10.8	10.6	9.9	10.5	11.1	10.9	11.8	11.6	11.9
	D	10.9	10.3	9.9	9.8	9.4	9.6	9.9	9.8	10.2	9.7	9.9
Textverständnis	CH	9.0	9.6	9.1	9.2	8.9	10.2	9.2	8.5	9.0	8.9	10.3
	D	8.4	9.6	8.7	8.9	9.0	8.9	8.6	8.6	9.3	8.9	10.2
Figuren lernen	CH	13.8	11.4	11.7	13.4	13.0	13.6	15.7	10.6	10.2	12.7	12.6
	D	12.5	9.3	9.3	11.5	11.5	10.7	11.8	9.2	9.3	11.8	11.6
Fakten lernen	CH	13.4	11.4	12.0	12.6	12.4	11.4	10.7	10.8	10.1	11.8	11.6
	D	11.4	9.0	9.3	11.1	11.3	10.3	9.7	9.3	8.9	11.0	11.2
Diagramme und Tabellen	CH	10.1	10.5	9.4	10.5	10.7	11.7	10.8	10.3	10.4	10.5	11.4
	D	9.9	9.9	9.6	10.7	10.1	10.7	10.5	10.0	9.7	10.3	11
Total	CH	96.9	89.8	88.8	92.7	87.9	91.4	90.9	84.6	85.7	91.3	96.8
	D	86.9	80.5	79.1	83.7	83.4	82.6	82.7	79.5	79.8	85.8	89.5

Tabelle 21: Mittelwerte (Punkte) der Untertests 1998 bis 2008; deutsche Sprachgruppe Schweiz (Humanmedizin) und Schätzung aufgrund der Item-Schwierigkeitswerte aus Deutschland – für 8 Untertests liegen Werte vor.

Die folgenden Abbildungen veranschaulichen den Vergleich zwischen deutschen und deutschschweizerischen Ergebnissen für die Untertests. Die Differenzen bewegen sich – von wenigen „Ausreissern“ abgesehen – auf weitgehend konstantem Niveau. Die Kurven verlaufen gleichförmig, die relativen Schwierigkeiten stimmen also nach wie vor gut überein. Einziger Ausreisser ist der Untertest „Schlauchfiguren“. Hier muss die weitere Entwicklung beobachtet werden.

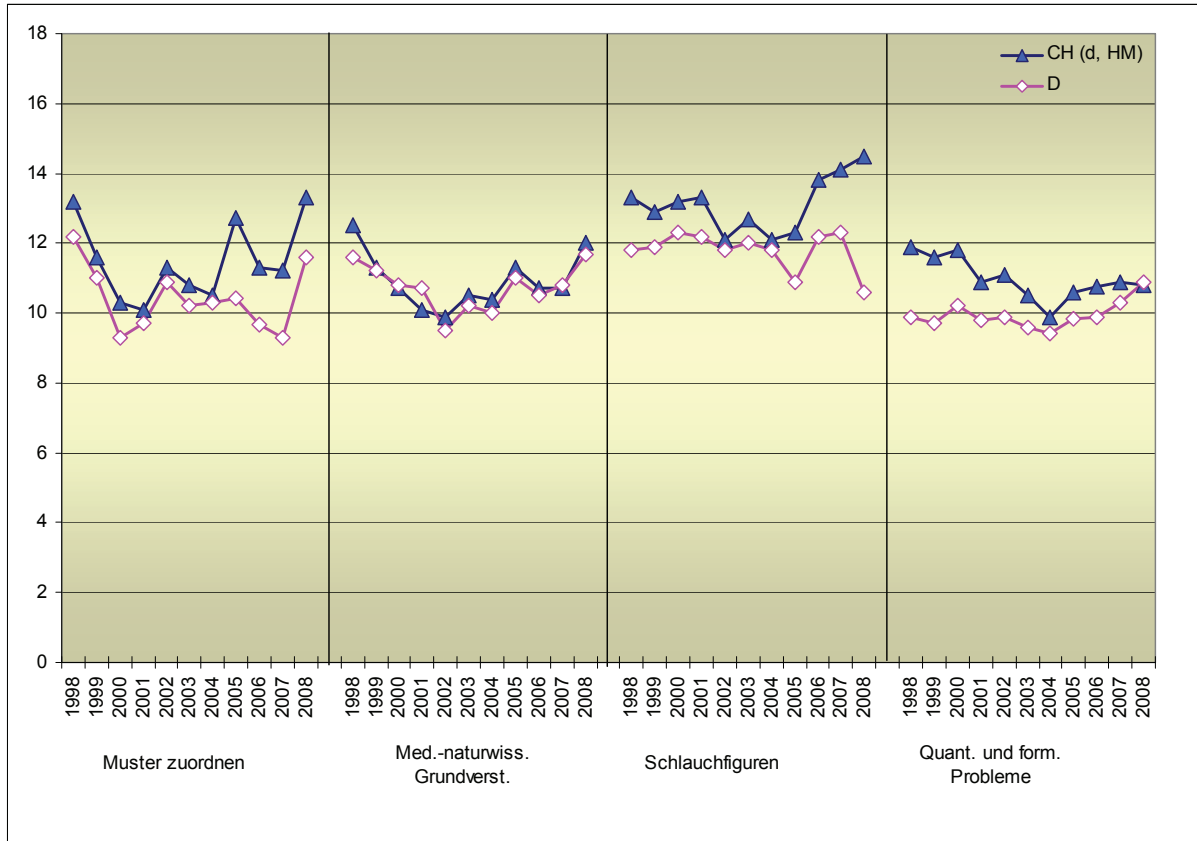


Abbildung 26: Mittelwertvergleiche der Untertests Schweiz (deutschsprachig, Humanmedizin) und Vergleichswerte Deutschland, 1998 bis 2008.

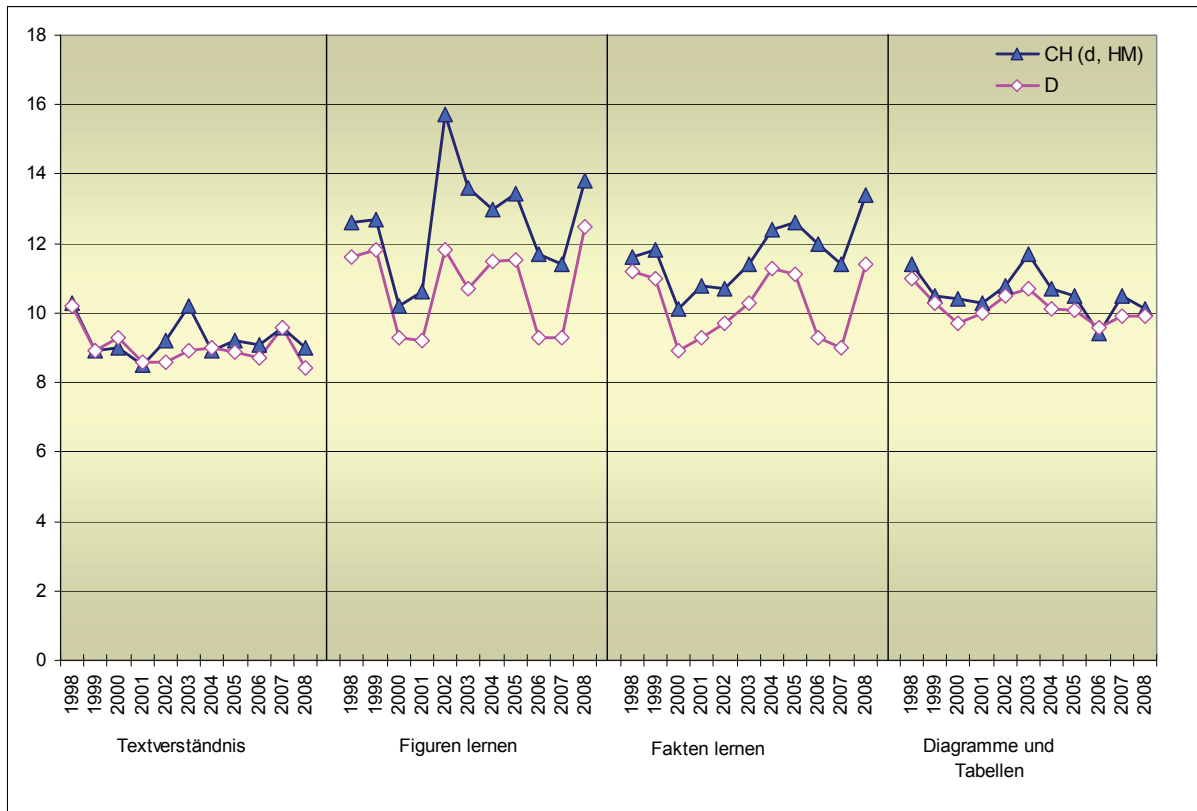


Abbildung 27: Mittelwertvergleiche der Untertests Schweiz (deutschsprachig, Humanmedizin) und Vergleichswerte Deutschland 1998 bis 2008.

6.3 Äquivalenz der Sprachversionen

Der Test wurde nach den gleichen Kriterien wie in den Vorjahren von der deutschen Vorlage adaptiert. Eine ausführliche Beschreibung des Verfahrens kann dem Bericht 5 (1999) entnommen werden. Zielsetzung bleibt, möglichst übereinstimmende Aufgabenschwierigkeiten zwischen den Sprachversionen zu schaffen.

Chancengleichheit als Gleichbehandlung der Sprachgruppen muss gewährleistet sein, die Lösungswahrscheinlichkeit einer Aufgabe muss *bei gleicher Fähigkeit* tatsächlich auch vergleichbar sein. Hervorzuheben ist, dass seit 2005 in französischer und italienischer Sprache die Schlussredaktion des Tests von zweisprachigen Lehrpersonen durchgeführt wird, die an Maturitätsschulen unterrichten. Sie wurden auch dahingehend geschult, dass es vor allem auf eine Übertragung der Schwierigkeit in die jeweilige Sprache ankommt. Da sie zudem mit dem Sprachniveau der Maturitätsstufe durch ihre Tätigkeit gut vertraut sind, gewinnt der Test zusätzlich an Güte.

Die Gleichheit ist nicht allein durch Übersetzung zu beeinflussen, sondern hängt von mehreren Faktoren ab. Bei allen Adaptationen muss mit Unterschieden hinsichtlich der Aufgabenschwierigkeiten zwischen den Sprachversionen gerechnet werden. Abbildung 28 fasst zusammen, welche Ursachen für diese Differenzen verantwortlich sein können.

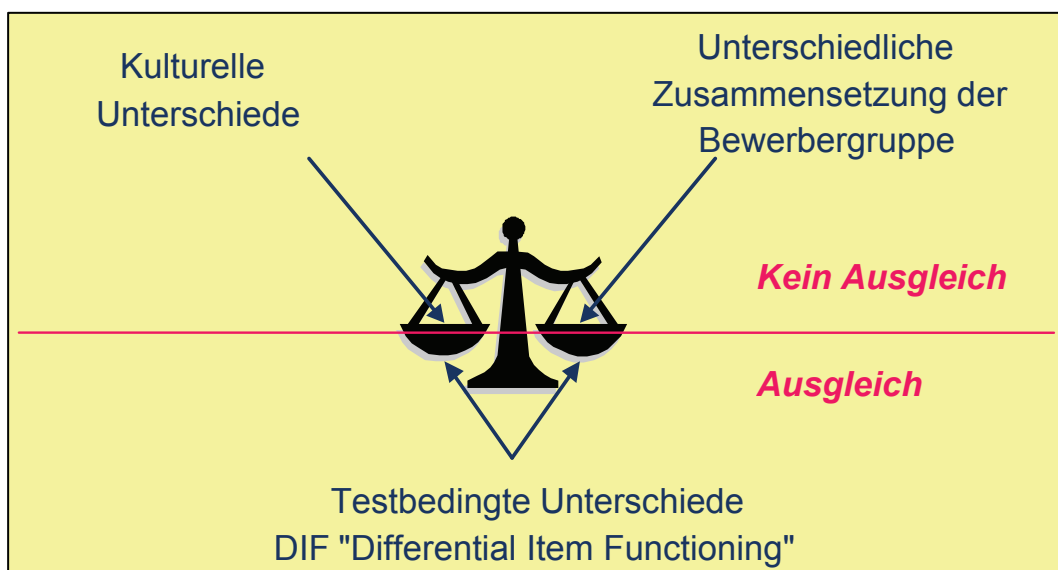


Abbildung 28: Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen.

Zwei Ursachen führen zu „echten“ Personen-Unterschieden, die nicht auf den Test oder seine Adaptation zurückgeführt werden können. Sie müssen durch die Ergebnisse abgebildet werden – hier erfolgt kein Ausgleich.

Testbedingte Unterschiede werden durch die Adaptation verursacht und sollen möglichst vollständig ausgeglichen werden. Beispiele wären eine Vereinfachung oder Erschwerung der Aufgabe bei der Übersetzung (Wortwahl, Satzgliederung o.ä.). Erneut wird zu diesem Zweck das bewährte DIF-Verfahren (Differential Item Functioning) angewendet.

Kulturelle Unterschiede zwischen den Sprachgruppen

- Mögliche generelle Fähigkeitsunterschiede oder unterschiedliche Fähigkeitsprofile (anderes Verhältnis der einzelnen Fähigkeiten zueinander) im Vergleich der Kulturen können vorhanden sein, die möglicherweise auf Unterschiede im Schulsystem zurückzuführen sind. 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. Französisch- und italienischsprachige Personen können den NC durch ein Studium in ihrer Muttersprache an einer anderen Universität vermeiden. Die geringe Zahl der französisch- und italienischsprachigen Teilnehmer macht dies wahrscheinlich: Die Auswahl kann die „Spitze“ oder das „Ende“ der Leistungsrangreihe aller Maturanden der jeweiligen Sprachgruppe überrepräsentieren, etwa durch unterschiedliches „Wahlverhalten“ aufgrund der vorhandenen Alternativen für Studienorte.

Testbedingte Unterschiede

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

Der Beschluss der Schweizerischen Hochschulkonferenz aus dem Jahre 1999 ist weiterhin bindend. Er beinhaltet:

1. Einen Ausgleich nur bei sprachabhängigen Tests vorzunehmen. Dies sind sechs von zehn Untertests. Der Sprachausgleich beschränkt sich demnach auf die stärker sprachabhängigen Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“, „Quantitative und formale Probleme“, „Textverständnis“, „Fakten lernen“ „Diagramme und Tabellen“ sowie „Planen und Organisieren“.
2. Den Ausgleich nur vorzunehmen, wenn signifikante Mittelwertunterschiede im entsprechenden Untertest vorhanden sind. Geprüft werden dazu die Abweichungen der jeweiligen Zielsprache von der deutschen Sprachgruppe.

Adaptationsbedingte Effekte sind nur in den sechs **sprachabhängigen Untertests** zu erwarten, während die vier **sprachunabhängigen Untertests** davon kaum betroffen sein können.

Spricht man von DIF, so muss dieses von systematischen Mittelwertsdifferenzen, welche alle Aufgaben eines Untertests gleichermaßen betreffen, abgegrenzt werden. Die Ergebnisse der letzten Jahre haben gezeigt, dass sich die Sprachgruppen in manchen Untertestscores signifikant unterscheiden. Differenzen kommen nicht nur durch einzelne Items zustande, sondern auch durch eine Verschiebung der Schwierigkeiten aller Items eines Untertests. Es ist unwahrscheinlich, dass diese systematischen Unterschiede testbedingt sind. Testbedingte Unterschiede als sprachliche Besonderheiten zeigen sich in spezifischen Abweichungen einzelner Items. Systematische Unterschiede sind dagegen mit grösserer Wahrscheinlichkeit den Faktoren „Kulturunterschiede“ und „unterschiedliche Repräsentativität“ zuzurechnen.

Es ergibt sich für das Ausgleichsverfahren die nachfolgende Zielstellung:

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

6.3.1 Sprachvergleich für die Untertests

Wie jedes Jahr treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Untertests auf. Die Punktzahlen der deutschsprachigen Teilnehmer liegen mehrheitlich über jenen der übrigen Kandidaten. Im Untertest „Quantitative und formale Probleme“ erzielten allerdings die deutschsprachigen Teilnehmer dieses Jahr die tiefsten Punktzahlen.

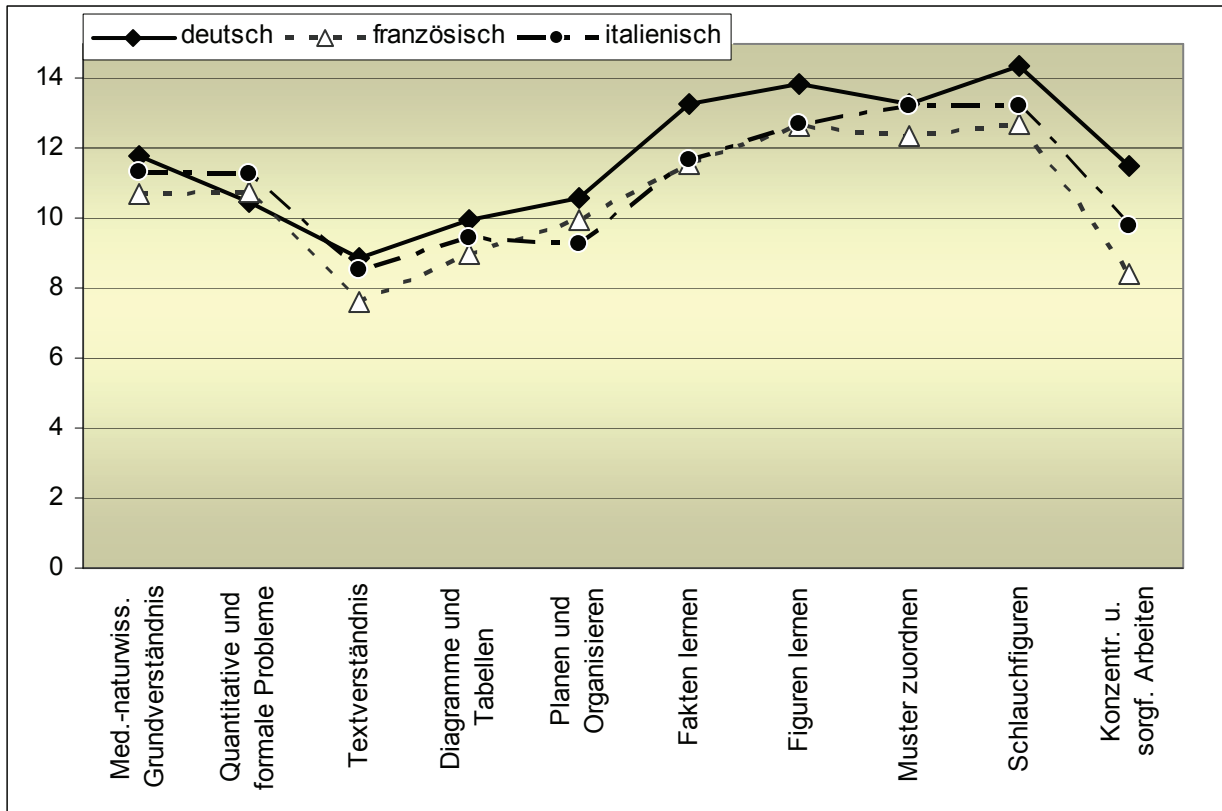


Abbildung 29: Mittelwerte der Untertests für die Sprachgruppen 2008 (unkorrigiert).

Die Abbildung 30 zeigt den Vergleich zwischen deutsch- und französischsprachigen Testteilnehmern für alle Testanwendungen seit 1998. In Abbildung 31 ist der Vergleich zwischen deutsch- und italienischsprachigen Personen dargestellt. Die senkrechten Linien bezeichnen die Spannweite (Minimum bis Maximum) der Differenzen über alle Jahrgänge seit 1998, die Verbindungslinie steht für die Differenzen aus dem aktuellen Jahrgang. Höhere Differenzwerte sprechen für bessere Ergebnisse der deutschen Sprachgruppe.

Die Differenzen zwischen den Sprachgruppen bewegen sich tendenziell im unteren Bereich der gewohnten Größenordnungen.

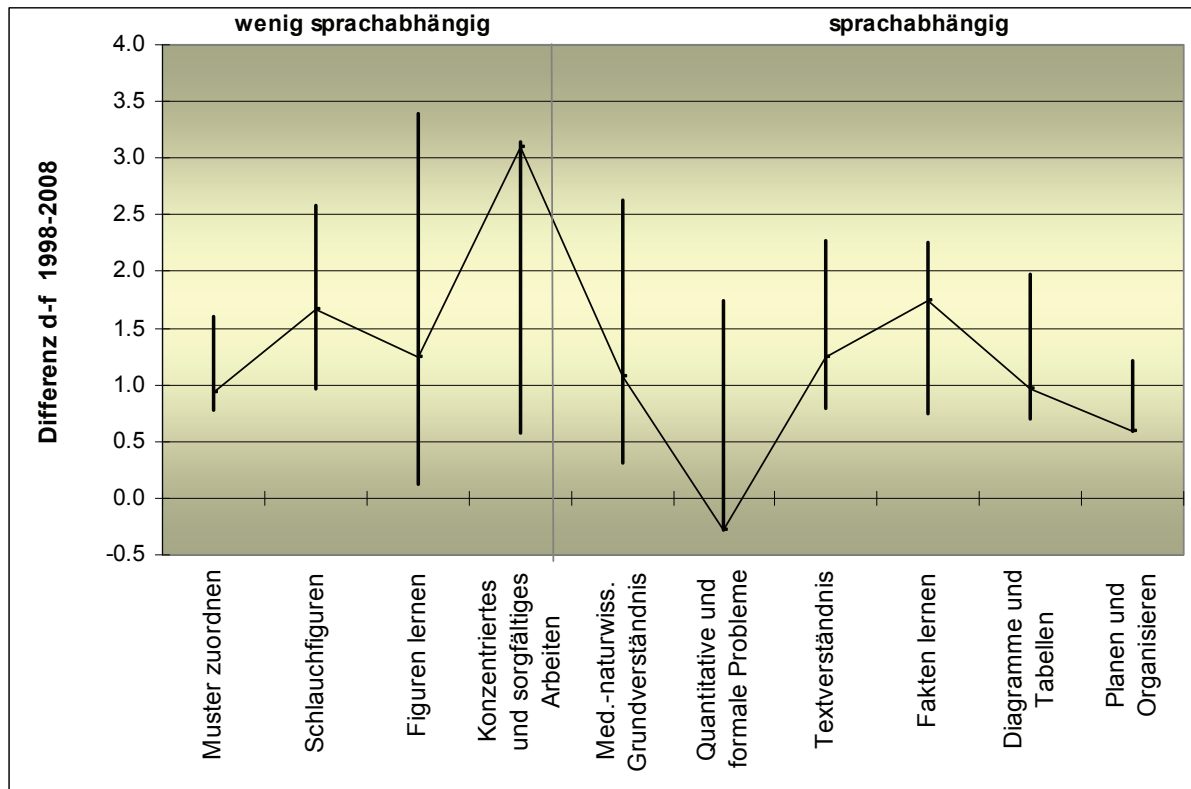


Abbildung 30: Differenz Testwert **deutsch-französisch** 2008 (unkorrigiert) und Spannweite dieser Differenz (min-max) der Jahre 1998 bis 2008 (vertikale Linien).

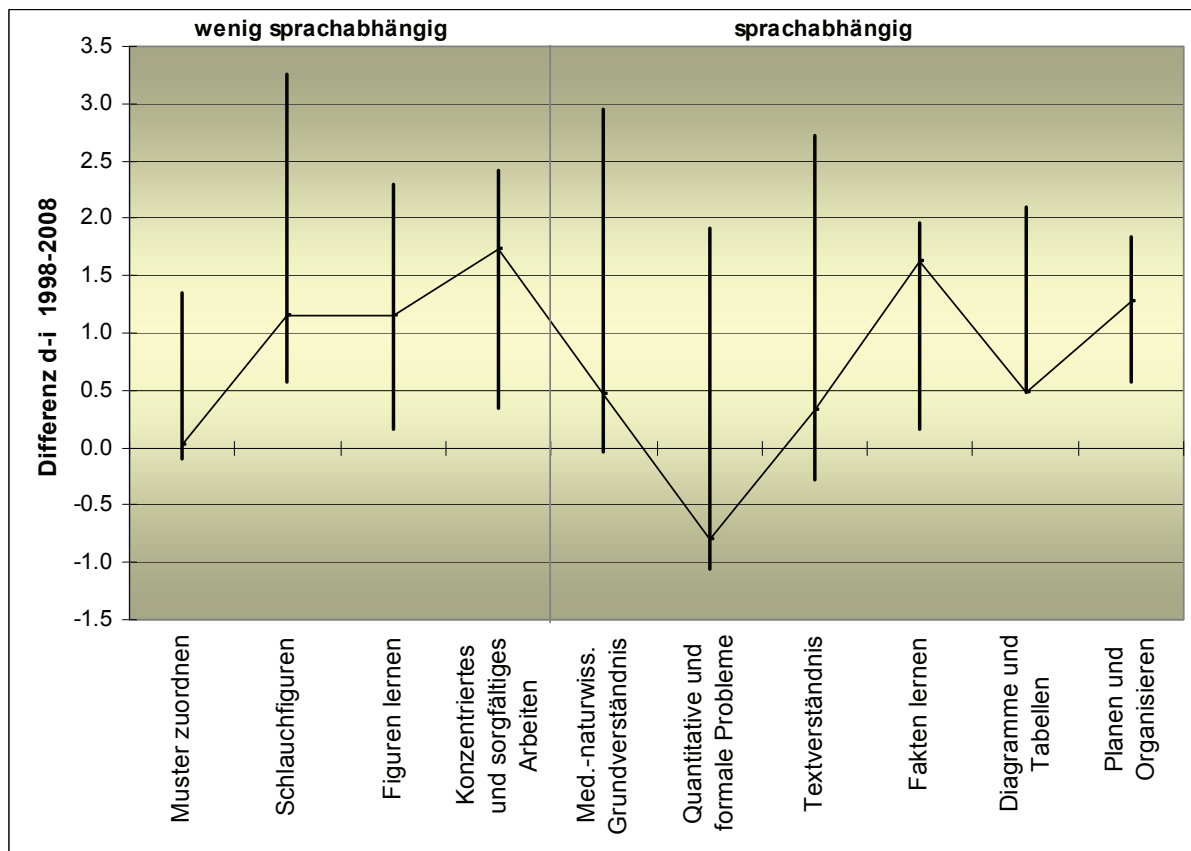


Abbildung 31: Differenz Testwert **deutsch-italienisch** 2008 (unkorrigiert) und Spannweite dieser Differenz (min-max) der Jahre 1998 bis 2008 (vertikale Linien).

6.3.2 Darstellung des Korrekturverfahrens

6.3.2.1 Identifikation und Ausgleich testbedingter Unterschiede geschehen wie folgt:

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

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

6.3.2.2 Welche Untertests ausgleichen?

Die folgenden Tabellen zeigen die Sprachvergleiche für die sprachabhängigen Untertests. Referenz ist die deutschsprachige Testversion. Die französischsprachigen Teilnehmer (Tabelle 22) erzielten in fünf von sechs sprachabhängigen Untertests signifikant geringere Punktwerte als die deutschsprachigen Teilnehmer. Für die italienischsprachigen Kandidaten (Tabelle 23) weisen zwei sprachabhängige Untertests signifikante Differenzen auf.

Untertest	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	1745	11.78	3.50	.000	Ja
	f	142	10.71	3.42		
Quantitative und formale Probleme	d	1745	10.47	3.63	.360	Nein
	f	142	10.76	3.45		
Textverständnis	d	1745	8.86	3.27	.000	Ja
	f	142	7.61	3.05		
Fakten lernen	d	1745	13.29	3.74	.000	Ja
	f	142	11.55	3.60		
Diagramme und Tabellen	d	1745	9.95	3.17	.001	Ja
	f	142	8.99	2.97		
Planen und Organisieren	d	1745	10.57	3.39	.048	Ja
	f	142	9.99	3.18		

Tabelle 22: Signifikanzprüfung deutsche vs. französische Sprachform.

Untertest	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	1745	11.78	3.50	.263	Nein
	i	74	11.31	3.71		
Quantitative und formale Probleme	d	1745	10.47	3.62	.064	Nein
	i	74	11.27	3.76		
Textverständnis	d	1745	8.86	3.27	.405	Nein
	i	74	8.54	3.36		
Fakten lernen	d	1745	13.29	3.74	.000	Ja
	i	74	11.68	3.37		
Diagramme und Tabellen	d	1745	9.95	3.17	.205	Nein
	i	74	9.47	2.89		
Planen und Organisieren	d	1745	10.57	3.39	.002	Ja
	i	74	9.30	3.57		

Tabelle 23: Signifikanzprüfung deutsche vs. italienische Sprachform.

Mit der Bestimmung der auszugleichenden Untertests ist noch nicht bekannt, ob es sich bei den Differenzen um in der Stichprobe „real“ vorhandene Differenzen oder um mögliche adaptionsbedingte Unterschiede handelt. Zu diesem Zweck ist eine weitere Analyse der Aufgaben der betreffenden Untertests notwendig.

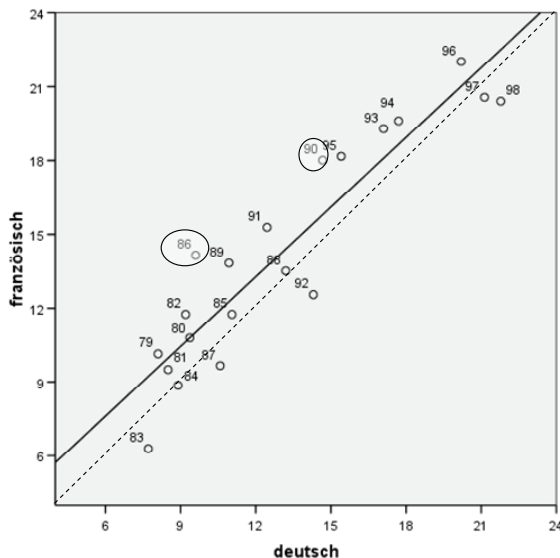
6.3.2.3 Identifikation von DIF-Items und Bestimmung des Korrekturwertes

Eine Möglichkeit zur Identifikation von DIF, auch bei kleinen Stichprobengrößen, ist die Methode „Delta-Plot“. Bei diesem Verfahren werden die zu vergleichenden Item-Schwierigkeiten z-standardisiert und anschliessend in „ Δ -Werte“ transformiert. Die Transformation erfolgt über die Formel:

$$\Delta = 13 - 4z.$$

Dies bedeutet, dass Δ einen Mittelwert von 13 und eine Standardabweichung von 4 aufweist. Hohe Werte stehen für „schwierige“ (von weniger Probanden gelöste) Items.

Die aus den Delta-Werten abgeleitete Regressionsgerade $Y = AX + B$ beschreibt die Beziehung zwischen den interessierenden Sprachversionen. Eine graphische Darstellung der Delta-Werte (Delta-Plot) würde im Idealfall eine ellipsenförmige Anordnung von Punkten entlang der Diagonalen ergeben. Dies würde bedeuten, dass sowohl die Itemschwierigkeiten wie auch deren Reihenfolge in beiden Sprachversionen vergleichbar sind.



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

Abbildung 32: Beispiel für einen Delta-Plot.

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

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

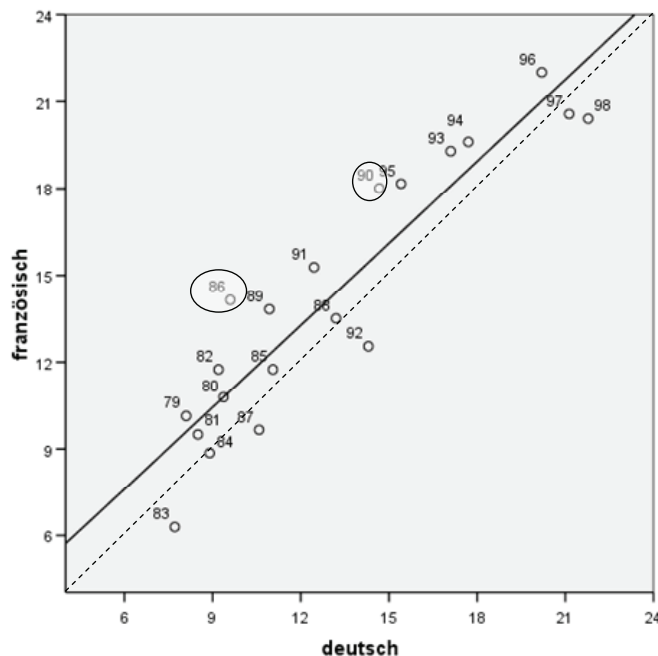
$$D_i = \frac{AX_i - Y_i + B}{\sqrt{A^2 + 1}}$$

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

In der Folge sollen innerhalb der auszugleichenden Untertests die kritischen Items identifiziert werden. Da, wie oben erwähnt, keine symmetrische DIF-Behandlung verwendet wird, ist die Richtung der Abweichung der Regressionsgeraden von der Diagonalen massgebend. Es werden also jene Items ausgeglichen, **welche mindestens 1.5 Punkte** (Longford, Holland & Thayer, 1993) von der Regressionsgeraden in entgegengesetzter Richtung zur Diagonalen abweichen. Als Sprachausgleich wird der benachteiligten Sprachgruppe die Differenz der betreffenden Itemschwierigkeit zur deutschen Itemschwierigkeit gutgeschrieben (also die Dif-

ferenz der Lösungswahrscheinlichkeit). Dies betrifft nur jene Personen, welche das fragliche Item nicht korrekt beantwortet haben.

6.3.2.4 Medizinisch-naturwissenschaftliches Grundverständnis



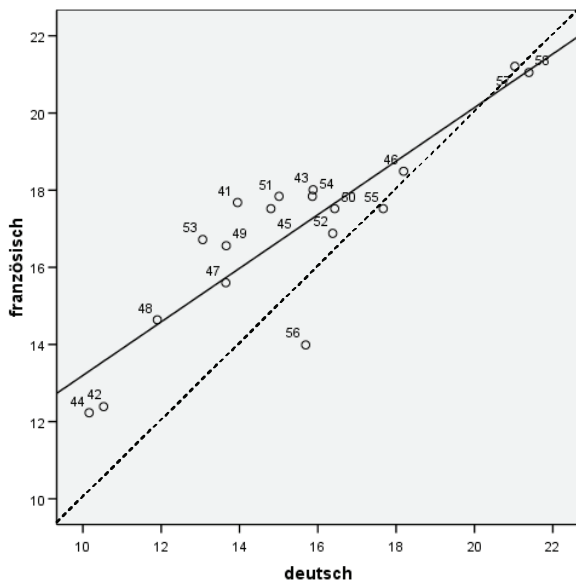
Im Untertest „medizinisch-naturwissenschaftliches Grundverständnis“ fallen zwei Items auf, welche die kritische Distanz zur Regressionsgeraden überschreiten. Für die Fragen 86 und 90 erhalten die französischsprachigen Teilnehmer entsprechende Bonuspunkte, falls sie diese Aufgaben nicht korrekt gelöst haben.

Abbildung 33: Delta-Plot für „Medizinisch-naturwissenschaftliches Grundverständnis“ (deutsch-französisch).

Item	Schwierigkeit			Δ -Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
79	0.81	0.72	0.64	8.10	10.14	12.04	-0.40			
80	0.75	0.69	0.68	9.37	10.79	11.12	0.00			
81	0.79	0.75	0.70	8.50	9.50	10.50	0.34			
82	0.76	0.65	0.74	9.20	11.75	9.58	-0.81			
83	0.83	0.89	0.91	7.71	6.29	5.88	2.13			
84	0.77	0.77	0.80	8.90	8.86	8.34	1.08			
85	0.68	0.65	0.62	11.05	11.75	12.35	0.46			
86	0.74	0.54	0.54	9.60	14.15	14.19	-2.29		.20	
87	0.70	0.74	0.73	10.58	9.66	9.88	1.65			
88	0.58	0.57	0.72	13.20	13.51	10.19	0.65			
89	0.68	0.56	0.68	10.93	13.83	11.12	-1.14			
90	0.52	0.37	0.41	14.67	18.01	17.27	-1.61		.15	
91	0.62	0.49	0.50	12.45	15.28	15.12	-1.15			
92	0.54	0.61	0.65	14.30	12.55	11.73	2.11			
93	0.41	0.32	0.39	17.10	19.29	17.58	-0.88			
94	0.39	0.30	0.32	17.70	19.61	19.12	-0.70			
95	0.49	0.37	0.43	15.41	18.17	16.66	-1.22			
96	0.28	0.20	0.23	20.20	22.02	21.27	-0.73			
97	0.24	0.26	0.38	21.13	20.57	17.89	0.96			
98	0.21	0.27	0.26	21.78	20.41	20.66	1.52			

Tabelle 24: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“ (d: deutsch; f: französisch; i: italienisch).

6.3.2.5 Textverständnis



Der Untertest „Textverständnis“ weist zwar eine signifikante Differenz zwischen deutsch- und französischsprachiger Teilnehmergruppe auf, jedoch zeigen sich auf Ebene der einzelnen Items keine Abweichungen, welche auf einzelne, ausgleichende Differenzen hinweisen. Insgesamt waren für diese Sprachgruppe vor allem die Fragen im Mittelbereich schwieriger zu beantworten.

Abbildung 34: Delta-Plot für „Textverständnis“ (deutsch-französisch).

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
41	0.55	0.39	0.46	13.95	17.68	16.04	-1.43			
42	0.70	0.62	0.58	10.53	12.39	13.27	0.96			
43	0.47	0.37	0.38	15.88	18.01	17.89	-0.59			
44	0.72	0.63	0.70	10.16	12.23	10.50	0.89			
45	0.51	0.39	0.43	14.80	17.52	16.66	-0.82			
46	0.37	0.35	0.34	18.19	18.49	18.81	0.33			
47	0.56	0.48	0.58	13.65	15.60	13.27	0.11			
48	0.64	0.52	0.59	11.90	14.64	12.96	-0.10			
49	0.56	0.44	0.41	13.66	16.56	17.27	-0.67			
50	0.44	0.39	0.55	16.43	17.52	13.89	0.12			
51	0.50	0.38	0.51	15.01	17.84	14.81	-0.96			
52	0.44	0.42	0.47	16.38	16.88	15.73	0.61			
53	0.59	0.43	0.73	13.06	16.72	9.88	-1.15			
54	0.47	0.38	0.39	15.86	17.84	17.58	-0.48			
55	0.39	0.39	0.50	17.67	17.52	15.12	0.82			
56	0.48	0.55	0.46	15.69	13.99	16.04	2.59			
57	0.24	0.23	0.19	21.03	21.21	22.20	-0.29			
58	0.22	0.24	0.26	21.39	21.05	20.66	0.05			

Tabelle 25: DIF-Analyse „Textverständnis“ (d: deutsch; f: französisch; i: italienisch).

6.3.2.6 Fakten lernen

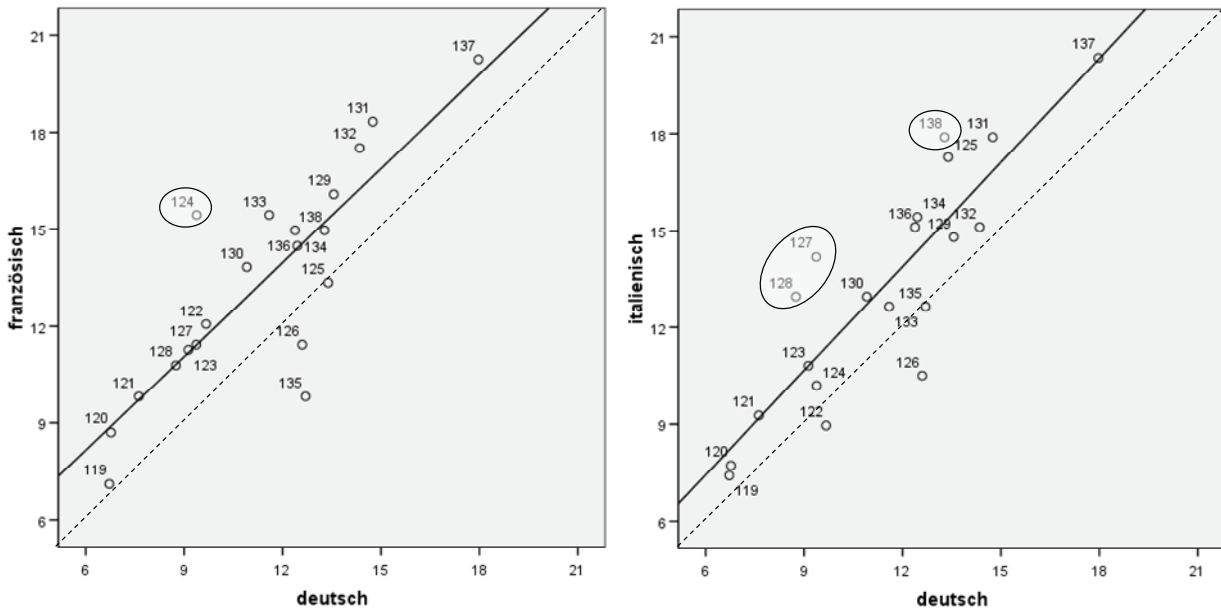


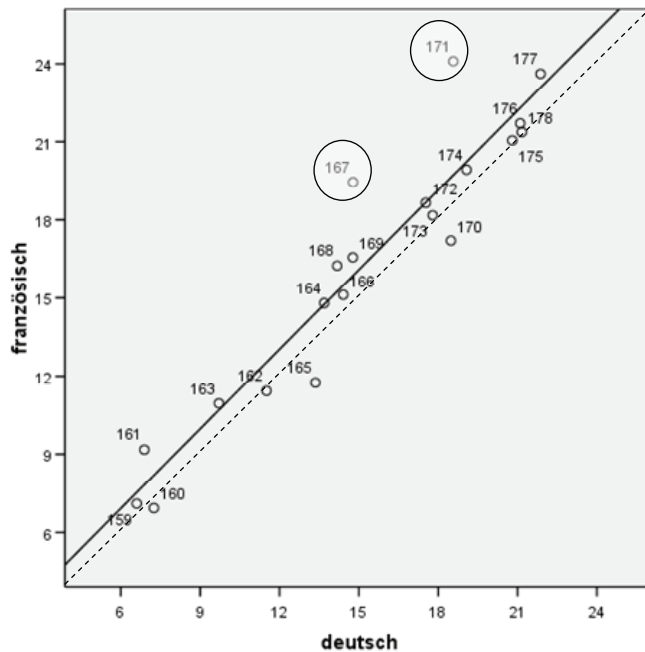
Abbildung 35: Delta-Plot für „Fakten lernen“ (deutsch-französisch und deutsch-italienisch).

Für die französischsprachigen Teilnehmer fällt eine Aufgabe in den „auszugleichenden Bereich“ (124), für die italienischsprachige Gruppe sind es die Items 127, 128 und 138. Ansonsten ist für beide Gruppen eine praktisch parallele Verschiebung der Regressionsgeraden zu beobachten.

Item	Schwierigkeit			Δ -Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
119	0.87	0.85	0.84	6.72	7.10	7.42	1.25	0.54		
120	0.87	0.78	0.82	6.77	8.70	7.73	0.14	0.37		
121	0.83	0.73	0.76	7.62	9.82	9.27	-0.08	-0.06		
122	0.74	0.63	0.77	9.67	12.07	8.96	-0.26	1.65		
123	0.76	0.67	0.69	9.13	11.27	10.81	-0.06	0.00		
124	0.75	0.49	0.72	9.38	15.44	10.19	-2.88	0.60	.27	
125	0.58	0.58	0.41	13.39	13.35	17.27	1.41	-1.29		
126	0.61	0.66	0.70	12.60	11.43	10.50	2.24	2.75		
127	0.75	0.66	0.54	9.37	11.43	14.19	-0.01	-2.13		.21
128	0.78	0.69	0.59	8.75	10.79	12.96	0.02	-1.74		.18
129	0.57	0.46	0.51	13.56	16.08	14.81	-0.43	0.51		
130	0.68	0.56	0.59	10.91	13.83	12.96	-0.67	-0.17		
131	0.52	0.36	0.38	14.75	18.33	17.89	-1.22	-0.72		
132	0.53	0.39	0.50	14.35	17.52	15.12	-0.91	0.89		
133	0.66	0.49	0.61	11.59	15.44	12.65	-1.34	0.54		
134	0.62	0.53	0.49	12.45	14.48	15.43	-0.05	-0.72		
135	0.61	0.73	0.61	12.70	9.82	12.65	3.46	1.35		
136	0.62	0.51	0.50	12.38	14.96	15.12	-0.44	-0.56		
137	0.37	0.27	0.27	17.97	20.25	20.35	-0.35	-0.03		
138	0.58	0.51	0.38	13.28	14.96	17.89	0.18	-1.79		.20

Tabelle 26: DIF-Analyse „Fakten lernen“ (d: deutsch; f: französisch; i: italienisch).

6.3.2.7 Diagramme und Tabellen



Mit zwei Ausnahmen fällt die Differenz zwischen den Ergebnissen der deutschen und französischen Sprachgruppe vergleichsweise gering aus: Für die abweichenden Items 167 und 171 werden Bonuspunkte gutgeschrieben.

Abbildung 36: Delta-Plot für „Diagramme und Tabellen“ (deutsch-französisch).

Item	Schwierigkeit			Δ-Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
159	0.87	0.85	0.89	6.60	7.10	6.19	0.30			
160	0.85	0.86	0.89	7.25	6.93	6.19	0.88			
161	0.86	0.76	0.80	6.89	9.18	8.34	-0.95			
162	0.66	0.66	0.69	11.51	11.43	10.81	0.77			
163	0.74	0.68	0.68	9.71	10.95	11.12	-0.18			
164	0.56	0.51	0.68	13.69	14.80	11.12	-0.03			
165	0.58	0.65	0.62	13.36	11.75	12.35	1.87			
166	0.53	0.50	0.39	14.41	15.12	17.58	0.25			
167	0.51	0.31	0.46	14.78	19.45	16.04	-2.51		.20	
168	0.54	0.45	0.45	14.18	16.24	16.35	-0.69			
169	0.52	0.44	0.32	14.77	16.56	19.12	-0.50			
170	0.35	0.41	0.41	18.48	17.20	17.27	1.70			
171	0.35	0.11	0.18	18.57	24.10	22.51	-3.07		.24	
172	0.39	0.35	0.35	17.53	18.65	18.50	0.01			
173	0.38	0.37	0.32	17.79	18.17	19.12	0.53			
174	0.33	0.29	0.41	19.07	19.93	17.27	0.21			
175	0.25	0.24	0.23	20.80	21.05	21.27	0.66			
176	0.24	0.21	0.18	21.10	21.70	22.51	0.42			
177	0.20	0.13	0.27	21.87	23.62	20.35	-0.37			
178	0.23	0.23	0.27	21.16	21.37	20.35	0.69			

Tabelle 27: DIF-Analyse „Diagramme und Tabellen“ (d: deutsch; f: französisch; i: italienisch).

6.3.2.8 Planen und Organisieren

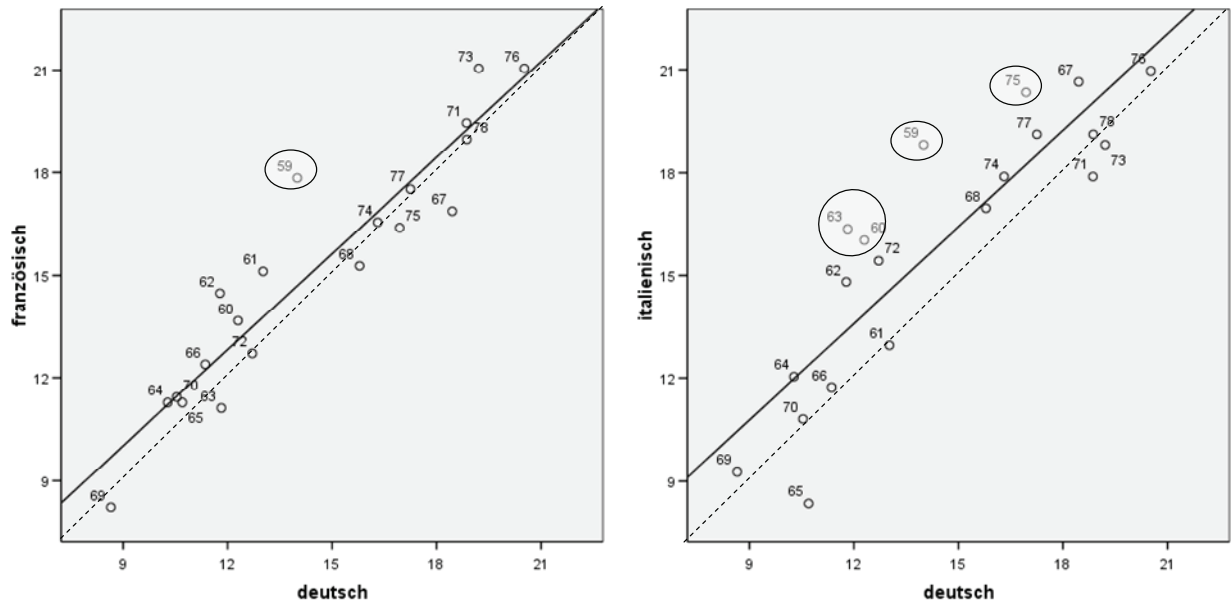


Abbildung 37: Delta-Plot für „Planen und Organisieren“ (deutsch-französisch und deutsch-italienisch).

In der französischsprachigen Gruppe weist ein Item (59) DIF auf, in der italienischsprachigen sind es vier (59, 60, 63 und 75).

Item	Schwierigkeit			Δ -Wert			Differenz		Bonus	
	d	f	i	d	f	i	f	i	Bonus f	Bonus i
59	0.55	0.38	0.34	14.00	17.84	18.81	-2.30	-2.43	.17	.21
60	0.62	0.56	0.46	12.30	13.67	16.04	-0.41	-1.57		.16
61	0.59	0.50	0.59	13.02	15.12	12.96	-0.98	1.16		
62	0.65	0.53	0.51	11.78	14.48	14.81	-1.36	-1.03		
63	0.64	0.68	0.45	11.82	11.11	16.35	1.13	-2.13		.20
64	0.71	0.67	0.64	10.28	11.27	12.04	-0.04	-0.04		
65	0.69	0.67	0.80	10.70	11.27	8.34	0.24	2.94		
66	0.66	0.62	0.65	11.36	12.39	11.73	-0.12	0.92		
67	0.35	0.42	0.26	18.45	16.88	20.66	1.45	-0.73		
68	0.47	0.49	0.42	15.79	15.28	16.96	0.80	0.14		
69	0.78	0.80	0.76	8.65	8.22	9.27	1.07	0.86		
70	0.70	0.66	0.69	10.54	11.43	10.81	0.02	1.03		
71	0.34	0.31	0.38	18.86	19.45	17.89	-0.15	1.57		
72	0.61	0.61	0.49	12.71	12.71	15.43	0.56	-0.85		
73	0.32	0.24	0.34	19.21	21.05	18.81	-1.08	1.14		
74	0.45	0.44	0.38	16.31	16.56	17.89	0.22	-0.17		
75	0.42	0.44	0.27	16.94	16.40	20.35	0.76	-1.54		.15
76	0.26	0.24	0.24	20.52	21.05	20.97	-0.19	0.46		
77	0.41	0.39	0.32	17.25	17.52	19.12	0.16	-0.43		
78	0.34	0.33	0.32	18.87	18.97	19.12	0.21	0.68		

Tabelle 28: DIF-Analyse „Planen und Organisieren“ (d: deutsch; f: französisch; i: italienisch).

6.3.3 Effekte der Korrektur

Die durch den Ausgleich resultierenden Bonuspunkte können Tabelle 29 entnommen werden. Durchschnittlich werden den französischsprachigen Teilnehmern 0.78 Punkte, den italienischsprachigen Teilnehmern 0.74 Punkte gutgeschrieben. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Der Bonus beträgt also jeweils 0 bis 1 Punkt. Die Auswirkungen der Korrektur sind damit wie in den Vorjahren gering.

		n	Minimum	Maximum	m	s
Franz.	Planen und Organisieren	142	.00	.17	.11	.082
	Med. naturw. Grundverständnis	142	.00	.35	.19	.131
	Fakten lernen	142	.00	.27	.14	.135
	Diagramme und Tabellen	142	.00	.44	.35	.123
	GESAMT	142	.00	1.23	.78	.261
	Planen und Organisieren	74	.00	.72	.45	.221
Ital.	Fakten lernen	74	.00	.59	.29	.204
	GESAMT	74	.00	1.31	.74	.308

Tabelle 29: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen.

Die Mittelwertsdifferenzen für jeden Untertest **nach** dem Sprachausgleich sind in Tabelle 30 dargestellt. Die Ergebnisse multipler Mittelwertvergleiche können der ersten Spalte entnommen werden. Differenzen zwischen den Teilnehmern sind dort beispielsweise in der Form $D > (F, I)$ angegeben. Das Beispiel würde bedeuten, dass die deutschsprachigen Kandidaten signifikant höhere Werte erzielt haben als die beiden anderen Sprachgruppen.

In Abbildung 41 und 42 sind die Ausmasse der (ungerundeten) Bonuspunkte für die bisherigen Testdurchführungen dargestellt (1998 wurde kein DIF-Verfahren angewendet). Die senkrechten Linien bezeichnen die Spannweite in jedem Jahr (Minimum bis Maximum), der jeweilige Mittelwert wird durch die Markierung gekennzeichnet.

		n	Mittelwert	Stand.abw.	Minimum	Maximum
Testwert ** $D > (I, F)$	deutsch	1745	100.49	9.96	70	129
	französisch	142	95.40	9.08	71	124
	italienisch	74	97.58	9.55	76	122
Muster zuordnen ** $(D, I) > F$	deutsch	1745	13.27	3.00	0	20
	französisch	142	12.35	2.85	4	18
	italienisch	74	13.24	3.21	4	20
Med.-naturwiss. Grund- verständnis ** $D > F$	deutsch	1745	11.78	3.50	1	20
	französisch	142	10.90	3.33	4.35	19
	italienisch	74	11.31	3.71	0	18
Schlauchfiguren ** $D > (I, F)$	deutsch	1745	14.35	3.47	3	20
	französisch	142	12.68	3.53	4	19
	italienisch	74	13.20	3.35	2	20
Quantitative und formale Probleme <i>homogen</i>	deutsch	1745	10.47	3.62	0	20
	französisch	142	10.76	3.45	1	19
	italienisch	74	11.27	3.76	2	19
Textverständnis ** $D > F$	deutsch	1745	8.86	3.27	1	18
	französisch	142	7.61	3.05	2	18
	italienisch	74	8.54	3.36	2	16
Figuren lernen ** $D > (I, F)$	deutsch	1745	13.87	3.74	0	20
	französisch	142	12.63	3.29	4	20
	italienisch	74	12.72	3.46	4	20
Fakten lernen ** $D > (I, F)$	deutsch	1745	13.29	3.74	3	20
	französisch	142	11.69	3.55	4.27	20
	italienisch	74	11.97	3.25	3.38	19
Diagramme und Tabellen <i>homogen</i>	deutsch	1745	9.95	3.17	1	19
	französisch	142	9.35	2.93	2.44	17.24
	italienisch	74	9.47	2.89	3	16
Konzentriertes und sorgf. Arbeiten ** $D > (I, F)$	deutsch	1745	11.51	5.04	0	20
	französisch	142	8.41	4.66	0	19
	italienisch	74	9.77	4.68	0	19
Planen und Organisieren <i>homogen</i>	deutsch	1745	10.57	3.39	0	20
	französisch	142	10.09	3.15	2.17	18
	italienisch	74	9.74	3.43	1.72	18

Tabelle 30: Testwert und Punktwerte der Untertests für die Sprachgruppen. **/*: signifikanter Einfluss des Faktors „Sprache“ bei varianzanalytischer Prüfung (1%-, bzw. 5%-Niveau); darunter sind signifikante Unterschiede bei multiplem Mittelwertvergleich angegeben. In französischer und italienischer Sprachgruppe korrigierte Punktwerte.

Der Vergleich für die französischsprachige Gruppe zeigt, dass 2008 nur ein vergleichsweise geringer Bonuszuschlag erfolgt. Für die italienischsprachige Version war 2007 ein Anstieg zu beobachten, 2008 entspricht der Wert jedoch wieder den Vorjahren.

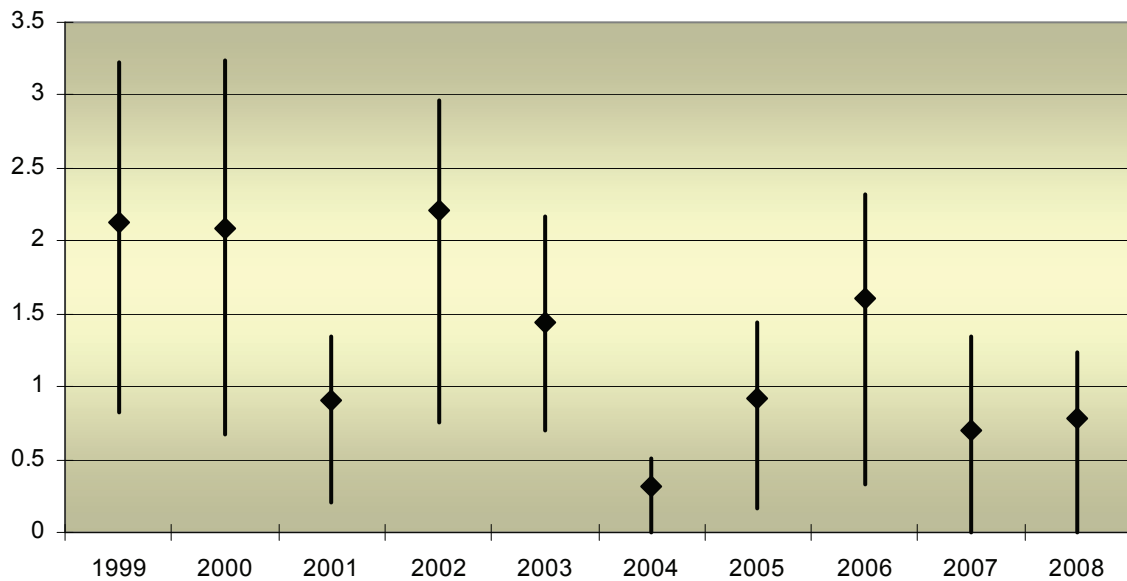


Abbildung 38: Bonuspunkte (ungerundet) über die Jahre für die französischsprachigen Kandidaten (Minimum – Mittelwert – Maximum).

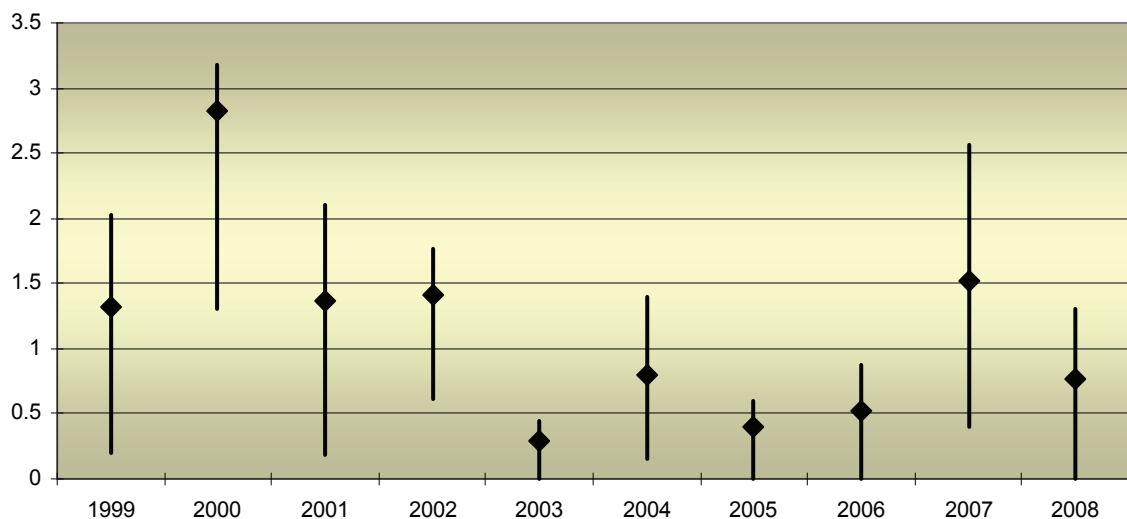


Abbildung 39: Bonuspunkte (ungerundet) über die Jahre für die italienischsprachigen Kandidaten (Minimum – Mittelwert – Maximum).

Es muss sichergestellt sein, dass Unterschiede zwischen den Sprachgruppen nicht im Test selber begründet liegen. Dies kann durch eine Analyse des Verhältnisses zwischen sprachabhängigen und wenig sprachabhängigen Untertests belegt werden. Die Differenzen zwischen der deutschsprachigen Referenzgruppe und den anderssprachigen Teilnehmern werden nach folgender Formel berechnet:

$$\text{Delta } z = \frac{m_{\text{Deutsch}} - m_{\text{Sprachgruppe}}}{s_{\text{Deutsch}}}$$

m ist der Mittelwert und s die Standardabweichung der entsprechenden Sprachgruppe. Die resultierenden Werte „Delta z“ drücken die Abweichung der jeweiligen Sprachgruppe von der deutschen Sprachgruppe in Standardabweichungen aus. 0.5 hiesse, dass die betroffene Sprachgruppe um eine halbe Standardabweichung geringere Punktzahlen erzielt als die deutschsprachige Referenzgruppe.

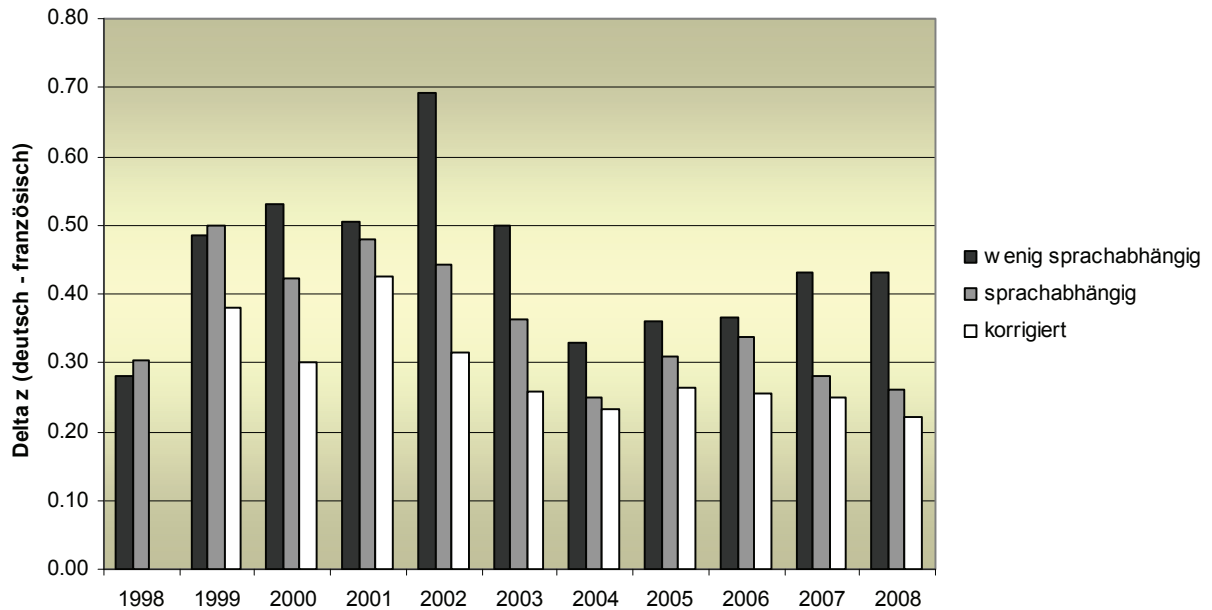


Abbildung 40: Abweichung (Delta z) für den Testwert zwischen deutscher und französischer Sprachgruppe. Für sprachabhängige Untertests sind korrigierte und unkorrigierte Werte dargestellt.

Erneut sind in der französischen Gruppe die Unterschiede in den wenig sprachabhängigen Untertests bereits vor dem Ausgleich grösser als jene der sprachabhängigen Untertests. Die Korrektur verringert diese zusätzlich, so dass eine Benachteiligung ausgeschlossen werden kann. Für die italienischsprachige Gruppe wird die gleiche Konstellation beobachtet.

Die Leistungsunterschiede zwischen den Sprachgruppen werden also deutlicher von den weniger sprachabhängigen Untertests bestimmt. Da die Faktorenanalysen auch 2008 wieder einen gemeinsamen Faktor für alle Untertests mit fast 40% Varianzaufklärung identifiziert haben, ist die Erwartung begründet, dass sich die Unterschiede in den wenig sprachabhängigen Untertests auch in den sprachabhängigen Untertests zeigen.

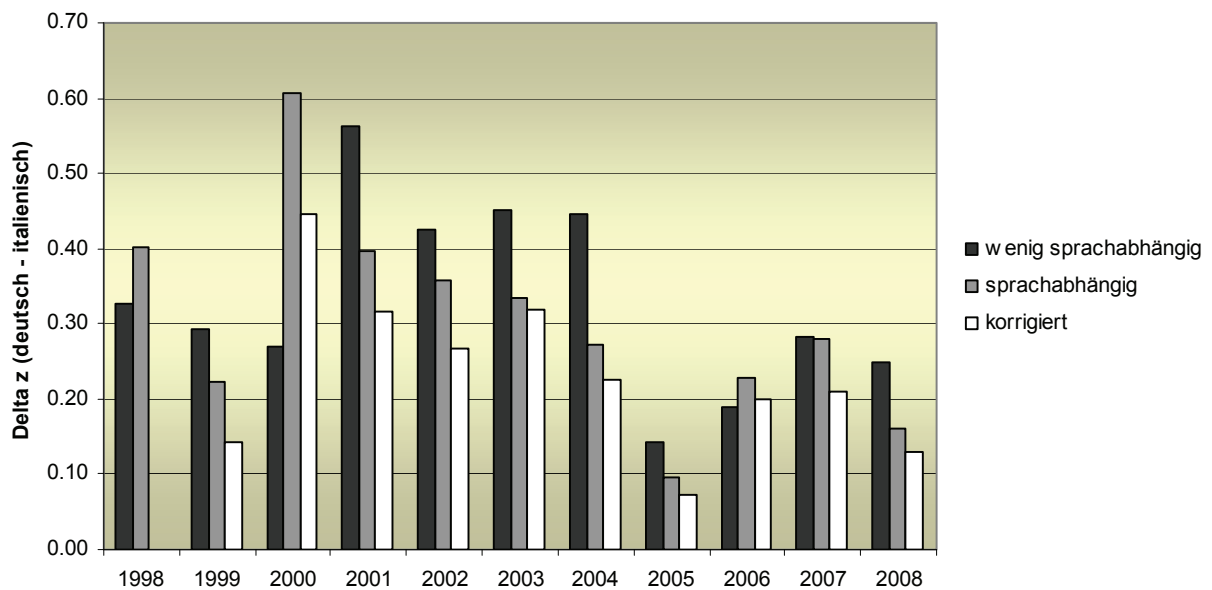


Abbildung 41: Abweichung (Delta z) für den Testwert zwischen deutscher und italienischer Sprachgruppe. Für sprachabhängige Untertests sind korrigierte und unkorrigierte Werte dargestellt.

6.4 Vergleichbarkeit der Testlokale

Die Durchführungsbedingungen an den einzelnen Testorten sind standardisiert. Aus den Protokollen der Testabnahme ergeben sich keine Hinweise, die als Einschränkungen dieser Chancengleichheit zu bewerten wären. Insbesondere die protokollierten Zeiten sind der Vorgabe entsprechend. Testwert und die Werte von drei Untertests unterscheiden sich signifikant zwischen den Testlokalen. Diese Unterschiede sind allerdings mit regionalen Unterschieden konfundiert.

In Tabelle 32 zeigt sich, dass nur das Testlokal Basel 1 signifikant tiefere Ergebnisse in den Untertests „Quantitative und formale Probleme“ sowie „Planen und Organisieren“ aufweist. Diese liegen jedoch im Rahmen der zwischen den Testorten gefundenen Differenzen.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testwert	<i>Zwischen Gruppen</i>	2400.6	18	133.367	1.350	0.147
	<i>Innerhalb Gruppen</i>	170565.5	1726	98.821		
	<i>Total</i>	172966.1	1744			
Muster zuordnen	<i>Zwischen Gruppen</i>	161.5	18	8.970	0.996	0.461
	<i>Innerhalb Gruppen</i>	15549.3	1726	9.009		
	<i>Total</i>	15710.8	1744			
Med.-naturwiss. Grundverständnis	<i>Zwischen Gruppen</i>	298.9	18	16.605	1.362	0.141
	<i>Innerhalb Gruppen</i>	21041.9	1726	12.191		
	<i>Total</i>	21340.8	1744			
Schlauchfiguren	<i>Zwischen Gruppen</i>	154.7	18	8.597	0.710	0.804
	<i>Innerhalb Gruppen</i>	20904.1	1726	12.111		
	<i>Total</i>	21058.8	1744			
Quant. und formale Probleme	<i>Zwischen Gruppen</i>	887.6	18	49.313	3.864	0.000
	<i>Innerhalb Gruppen</i>	22029.2	1726	12.763		
	<i>Total</i>	22916.8	1744			
Textverständnis	<i>Zwischen Gruppen</i>	153.5	18	8.527	0.797	0.705
	<i>Innerhalb Gruppen</i>	18458.0	1726	10.694		
	<i>Total</i>	18611.5	1744			
Figuren lernen	<i>Zwischen Gruppen</i>	195.3	18	10.852	0.774	0.733
	<i>Innerhalb Gruppen</i>	24199.9	1726	14.021		
	<i>Total</i>	24395.2	1744			
Fakten lernen	<i>Zwischen Gruppen</i>	240.8	18	13.380	0.956	0.509
	<i>Innerhalb Gruppen</i>	24150.9	1726	13.992		
	<i>Total</i>	24391.8	1744			
Diagramme und Tabellen	<i>Zwischen Gruppen</i>	156.2	18	8.679	0.861	0.627
	<i>Innerhalb Gruppen</i>	17392.2	1726	10.077		
	<i>Total</i>	17548.5	1744			
Konzentr. u. sorgf. Arbeiten	<i>Zwischen Gruppen</i>	734.8	18	40.821	1.616	0.049
	<i>Innerhalb Gruppen</i>	43593.4	1726	25.257		
	<i>Total</i>	44328.1	1744			
Planen und Organisieren	<i>Zwischen Gruppen</i>	351.4	18	19.522	1.712	0.031
	<i>Innerhalb Gruppen</i>	19678.4	1726	11.401		
	<i>Total</i>	20029.8	1744			

Tabelle 31: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den (deutschsprachigen) Testlokalen.

	Testwert	Muster zuordnen	Med.-naturwiss. Grundverständnis	Schlauchfiguren	Quant. und formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagramme und Tabellen	Konzentr. u. sorgf. Arbeiten	Planen und Organisieren	Lokalgrösse	
Testlokal	1	97.42	13.02	10.44	14.17	9.23	8.56	13.81	12.56	9.31	10.52	9.00	48
	2	99.46	13.05	11.51	14.23	9.37	8.66	13.95	13.45	9.78	11.61	10.03	110
	3	100.46	13.52	11.61	14.46	10.39	8.41	13.75	13.48	9.38	11.77	11.20	56
	4	98.07	12.89	11.28	14.14	9.46	8.56	13.61	12.96	9.32	10.16	10.07	57
	5	99.17	13.22	11.89	14.28	9.65	8.73	13.75	12.98	9.96	10.33	10.13	138
	6	101.37	13.38	12.05	14.48	10.87	8.93	13.90	13.08	10.36	12.03	10.87	104
	7	101.23	13.77	11.33	15.01	10.43	8.78	13.68	13.35	10.38	11.86	10.98	92
	8	99.72	13.30	12.07	13.80	9.45	9.43	13.43	13.30	9.67	11.53	10.12	60
	9	99.98	13.07	11.49	14.55	10.56	8.61	13.81	13.04	10.08	10.94	10.56	195
	10	100.33	13.86	11.50	14.08	11.06	8.83	13.86	13.42	9.47	10.89	10.08	36
	11	101.40	14.01	11.81	14.33	10.99	8.90	13.84	13.56	9.96	12.01	10.64	70
	12	100.59	12.99	11.49	14.24	10.56	9.01	13.66	13.16	9.85	12.47	10.66	68
	13	101.14	13.41	12.12	14.55	10.59	8.91	14.09	13.50	10.11	11.74	10.48	145
	14	102.86	13.80	12.35	15.14	11.23	8.78	14.26	14.28	10.00	12.60	10.88	65
	15	102.19	13.12	12.40	13.98	10.67	9.67	15.09	13.47	10.63	11.09	11.65	43
	16	102.37	13.11	12.62	14.42	12.16	9.55	14.27	13.88	10.21	11.02	11.07	91
	17	100.91	13.48	11.54	14.12	10.62	8.62	14.48	13.56	9.63	12.15	10.73	82
	18	99.81	13.15	12.01	14.01	10.04	8.72	13.32	12.67	10.04	11.93	10.63	135
	19	100.96	12.89	11.73	14.20	11.24	9.14	13.75	13.50	9.83	11.87	10.81	150

Tabelle 32: Darstellung der detaillierten Untertestergebnisse nach Testlokalen (deutschsprachig). Markiert: Mittelwerte für Testwert und die Punktwerte der Untertests im multiplen Mittelwertsvergleich nach Tukey, die signifikant tiefer als in mindestens einem anderen Testlokal liegen.

6.5 Vergleich für die Geschlechter

Die Forderung nach Gleichbehandlung der Geschlechter konnte in den bisherigen Testdurchführungen jeweils als erfüllt betrachtet werden. Chancengleichheit bedeutet, dass **bei gleicher Eignung die gleichen Chancen auf eine Zulassung** bestehen. Beide Teilgruppen sind unterschiedlich repräsentativ in der Bewerberkohorte vertreten.

Der Männeranteil in der Humanmedizin sank zwischen 1999 bis 2002 immer weiter ab. 2003 war erstmals wieder ein geringer Anstieg des Anteils an Kandidaten zu verzeichnen, der aber 2004 weiter abfiel. 2008 liegt der Männeranteil bei 36%.

Die mittleren Testwerte (in Klammer die Standardabweichung) nach Geschlecht, Sprache und Disziplin können Tabelle 33 entnommen werden.

		Sprache			Disziplin			Gesamt
		deutsch	französisch	italienisch	HM	VM	ZM	
Geschlecht	männlich	101.6	97.3	96.8	102.0	100.5	96.7	101.4
		(10.3)	(8.7)	(9.6)	(9.7)	(10.8)	(8.8)	(9.8)
		N=621	N=49	N=43	N=552	N=51	N=66	N=669
	weiblich	99.9	94.4	98.7	100.2	97.5	95.1	99.2
		(9.7)	(9.2)	(9.5)	(9.9)	(9.9)	(9.8)	(10.0)
		N=1124	N=93	N=31	N=822	N=237	N=103	N=1162

Tabelle 33: Statistiken für den Testwert (1. Zeile Mittelwert, 2. Zeile Standardabweichung, 3. Zeile Personenzahl) nach Geschlechtern für Sprache und Disziplin.

Von 1998 bis 2007 betrug der Unterschied beim Testwert in der Schweiz für die zahlenmässig grosse deutsche Sprachgruppe zwischen 1.4 und 2.2 Punkten zugunsten der Männer (1.6, 1.5, 1.4, 2.2, 1.3, 1.4, 2.0, 1.6, 1.9, 2.0). Der Unterschied von 1.7 Testwertpunkten im Jahr 2008 liegt also im Rahmen der Vorjahre.

		Humanmedizin			Veterinärmedizin			Zahnmedizin		
		Deutsch	Franz.	Italienisch	Deutsch	Franz.	Italienisch	Deutsch	Franz.	Italienisch
männlich	m	102.4	98.4	97.7	100.4	92.8	101.8	95.7	102.00	89.6
	s	10.3	7.9	9.1	7.4	10.6	6.9	10.4		10.3
	n	532	38	29	30	10	6	59	1	8
weiblich	m	100.3	96.5	96.7	99.4	93.1	102.9	95.8	93.5	
	s	9.6	8.7	8.8	9.7	9.4	10.0	9.3	5.0	
	n	880	35	21	160	56	10	84	2	0

Tabelle 34: Kombinierte Statistiken für den Testwert nach Geschlechtern, Sprache und Disziplin. *Kursiv: Wegen zu geringer Personenzahl nicht für Vergleichszwecke heranzuziehen.*

Sprache	Disziplin	Levene-Test Gleichheit der Varianzen			t-Test Gleichheit des Mittelwerts		
			F	Sig.	t	df	Sig.
Deutsch	Humanmedizin	Varianzen inhomogen	4.83	.028	3.67	1062	.000
	Veterinärmedizin	Varianzen inhomogen	4.69	.031	.62	49	.538
	Zahnmedizin	Varianzen homogen	.333	.565	-.07	141	.943
Französisch	Humanmedizin	Varianzen homogen	.081	.777	1.01	71	.316
Italienisch	Humanmedizin	Varianzen homogen	.120	.731	0.38	48	.706

Tabelle 35: Prüfung zur Varianzhomogenität und der Mittelwertunterschiede (t-Test) für Testwert zwischen Geschlechtern. Französische Sprachgruppe Zahnmedizin und italienische Sprachgruppe Veterinär- und Zahnmedizin wegen Fallzahl nicht berechnet.

Untertest	Geschlecht	m	s	Levene-Test (Varianz)			t-Test Mittelwert		
					F	Sig.	t	df	Sig.
Muster zuordnen	männlich	13.07	3.02	<i>homogen</i>	.01	.910	-1.43	1959	.152
	weiblich	13.28	3.00						
Med.-naturwiss. Grundverst.	männlich	12.08	3.74	<i>Nicht homogen</i>	10.91	.001	3.66	1355	.000
	weiblich	11.46	3.36						
Schlauchfiguren	männlich	15.04	3.26	<i>Nicht homogen</i>	11.33	.001	8.47	1585	.000
	weiblich	13.70	3.55						
Quantitative und formale Probleme	männlich	11.59	3.58	<i>homogen</i>	1.33	.248	10.08	1959	.000
	weiblich	9.92	3.50						
Textverständnis	männlich	9.23	3.28	<i>homogen</i>	.58	.446	4.81	1959	.000
	weiblich	8.49	3.23						
Figuren lernen	männlich	13.46	3.86	<i>Nicht homogen</i>	5.18	.023	-2.50	1406	.012
	weiblich	13.90	3.62						
Fakten lernen	männlich	12.71	3.86	<i>homogen</i>	2.82	.093	-3.51	1959	.000
	weiblich	13.33	3.67						
Diagramme und Tabellen	männlich	10.55	3.27	<i>Nicht homogen</i>	5.10	.024	7.28	1388	.000
	weiblich	9.47	3.02						
Konzent. u. sorgfältiges Arbeiten	männlich	10.79	5.21	<i>homogen</i>	2.21	.038	-2.83	1959	.005
	weiblich	11.46	4.98						
Planen und Organisieren	männlich	10.51	3.49	<i>homogen</i>	2.20	.139	.28	1959	.780
	weiblich	10.46	3.33						

Tabelle 36: Geschlechtsspezifische Mittelwerte (m), Standardabweichungen (s) und Ergebnisse der Prüfungen auf Varianzhomogenität und Mittelwertsunterschiede.

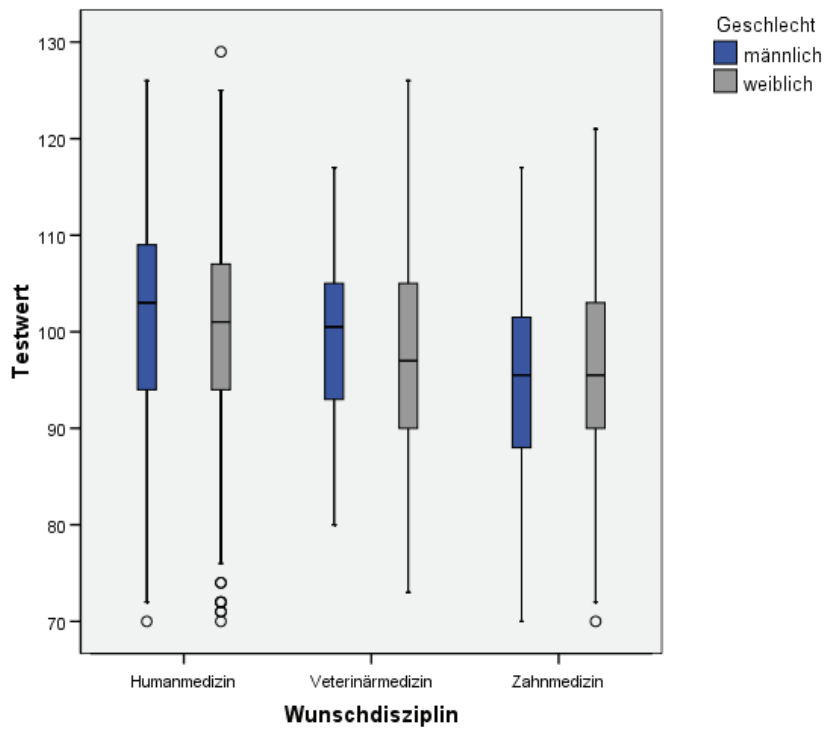


Abbildung 42: Boxplot für den Testwert nach Geschlecht und Wunschdisziplin (HM: Humanmedizin, VM: Veterinärmedizin, ZM: Zahnmedizin).

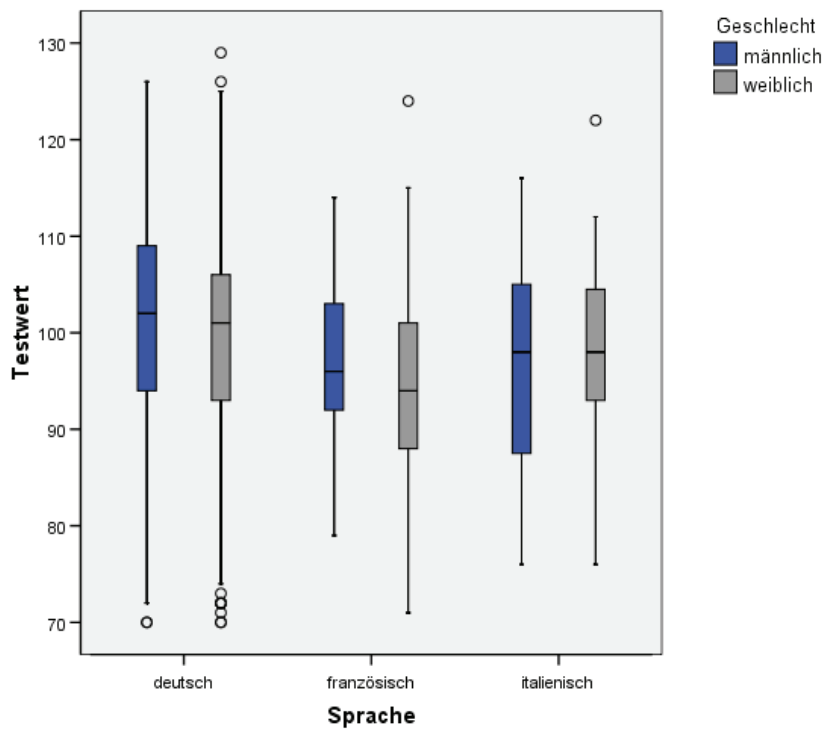


Abbildung 43: Boxplot für den Testwert nach Geschlecht, getrennt für die drei Sprachgruppen.

Die nachfolgenden Darstellungen verdeutlichen die Vergleichbarkeit der diesjährigen Ergebnisse mit früheren Jahrgängen (positive Werte bedeuten höhere Punktzahlen der männlichen Bewerber). Die seit 2003 andauernde Tendenz im Untertest „Konzentriertes und sorgfältiges

Arbeiten“ zu höheren Punktzahlen der männlichen Testteilnehmer wurde 2008 zumindest unterbrochen.

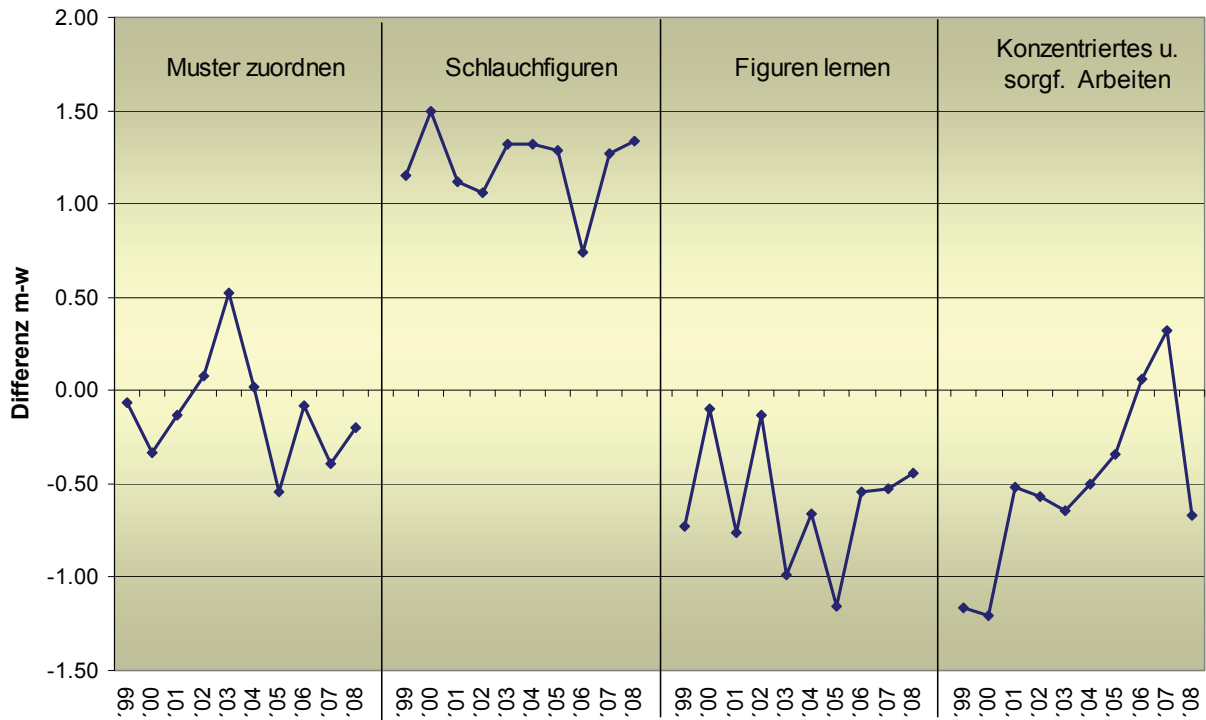


Abbildung 44: Differenzen für Punktwerte der wenig sprachabhängigen Untertests, geschlechtsspezifisch.

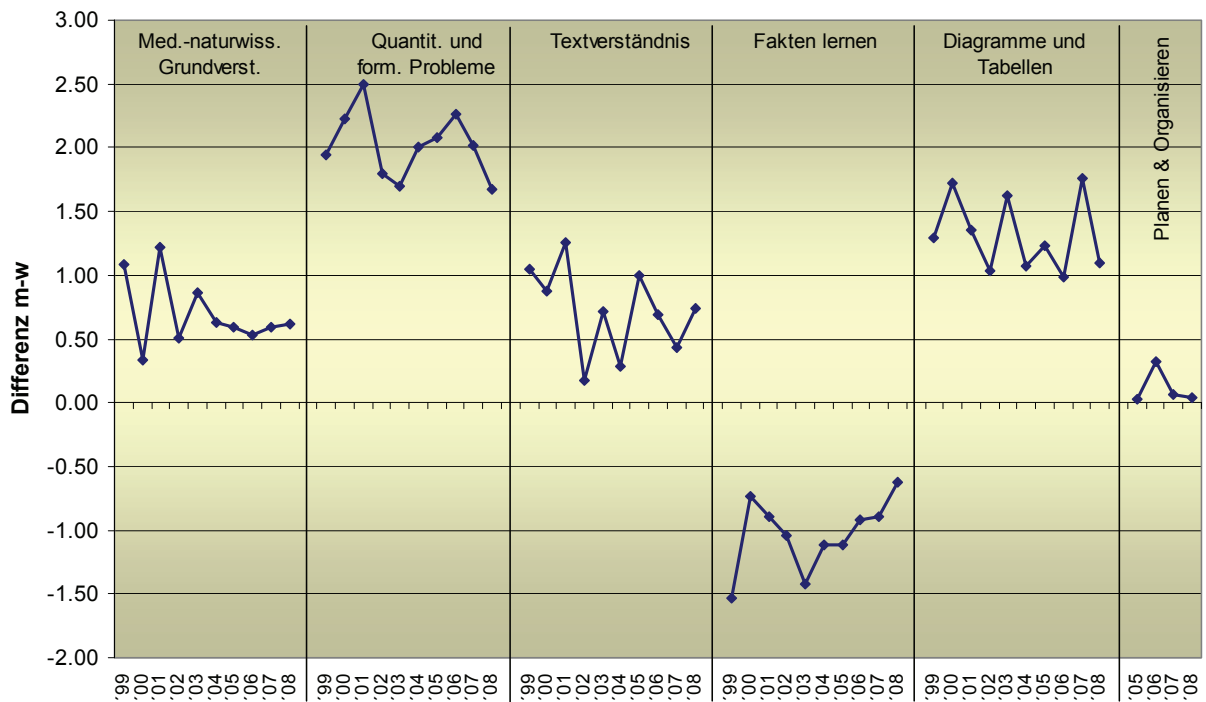


Abbildung 45: Differenzen für Punktwerte der sprachabhängigen Untertests, geschlechtsspezifisch.

6.6 Vergleiche für Altersgruppen

Für die Überprüfung der Alterseffekte wurden die Gruppen wie folgt gebildet. Die älteste Gruppe wurde nahe dem Median des Maturitätsalters geteilt.

Maturitätsjahr	Geburtsjahr			Total
	1957-1984	1985-1987	1988-1991	
1979	2			2
1980	2			2
1981	1			1
1983	3			3
1987	1			1
1988	1			1
1989	1			1
1990	1			1
1992	2			2
1994	4			4
1995	2			2
1996	3			3
1997	4			4
1998	5			5
1999	4			4
2000	4			4
2001	15			15
2002	19		1	20
2003	20	1	0	21
2004	12	22	1	35
2005	7	51	2	60
2006	7	140	26	173
2007	21	177	430	628
2008	41	116	811	968
	182	507	1271	1960

Tabelle 37: Gruppenbildung für das Jahr der Maturitätsprüfung bezogen auf die Geburtsjahre.

Wiederum zeigen die älteren Jahrgänge mit später Maturität tendenziell die niedrigsten Testwerte. Die ältesten Jahrgänge mit früher Maturität unterscheiden sich aber nicht signifikant von der zweitjüngsten Altersgruppe. Erneut sind es die Untertests „Quantitative und formale Probleme“ und „Diagramme und Tabellen“ in denen diese Teilnehmer besonders gute oder sogar höchste Ergebnisse erreichen.

	Geburtsjahr und Maturität	Mittelwert	Standard-abw.	Quadr.-summe	df	MQ	F	Sig.	Homogene Gruppen	
Testwert	1957-1984, frühe Maturität	97.31	12.81	5023.43	3	1674.48	17.27	.000	■	
	1957-1984, späte Maturität	95.33	9.42							■
	1985-1987	98.82	9.23							
	1988-1992	101.06	9.92							
Muster zuordnen	1957-1984, frühe Maturität	12.34	3.47	138.74	3	46.25	5.15	.002	■	
	1957-1984, späte Maturität	12.40	2.90							■
	1985-1987	13.30	2.85							
	1988-1992	13.29	3.03							
Med.-naturwiss. Grundverständnis	1957-1984, frühe Maturität	11.70	4.19	341.90	3	113.97	9.38	.000	■	
	1957-1984, späte Maturität	10.72	3.68							■
	1985-1987	11.17	3.47							
	1988-1992	11.97	3.43							
Schlauchfiguren	1957-1984, frühe Maturität	13.59	4.16	109.22	3	36.41	2.97	.031	■	
	1957-1984, späte Maturität	13.49	3.73							■
	1985-1987	14.43	3.34							
	1988-1992	14.19	3.50							
Quantitative und formale Probleme	1957-1984, frühe Maturität	10.97	4.23	395.90	3	131.97	10.21	.000	■	
	1957-1984, späte Maturität	9.42	3.43							■
	1985-1987	9.98	3.51							
	1988-1992	10.81	3.60							
Textverständnis	1957-1984, frühe Maturität	8.24	3.28	396.35	3	132.12	12.58	.000	■	
	1957-1984, späte Maturität	7.79	3.08							■
	1985-1987	8.24	3.08							
	1988-1992	9.09	3.31							
Figuren lernen	1957-1984, frühe Maturität	12.23	4.21	257.15	3	85.72	6.26	.000	■	
	1957-1984, späte Maturität	12.96	3.98							■
	1985-1987	13.81	3.60							
	1988-1992	13.87	3.69							
Fakten lernen	1957-1984, frühe Maturität	12.64	4.83	206.00	3	68.67	4.91	.002	■	
	1957-1984, späte Maturität	11.98	3.68							■
	1985-1987	12.97	3.87							
	1988-1992	13.29	3.62							
Diagramme und Tabellen	1957-1984, frühe Maturität	9.85	3.59	286.08	3	95.36	9.69	.000	■	
	1957-1984, späte Maturität	8.85	2.79							■
	1985-1987	9.44	2.94							
	1988-1992	10.12	3.21							
Konzentriertes und sorgfältiges Arbeiten	1957-1984, frühe Maturität	9.47	5.83	900.82	3	300.27	11.89	.000	■	
	1957-1984, späte Maturität	9.43	5.13							■
	1985-1987	10.79	5.20							
	1988-1992	11.65	4.89							
Planen und Organisieren	1957-1984, frühe Maturität	9.62	3.29	504.50	3	168.17	14.95	.000	■	
	1957-1984, späte Maturität	9.23	3.15							■
	1985-1987	9.99	3.29							
	1988-1992	10.83	3.40							

Tabelle 38: Testwert und Punktwerte für die drei Altersgruppen (nach Geburtsjahren) – für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung; varianzanalytische Prüfung der Unterschiede. In den letzten Spalten sind die bezüglich des Mittelwertes homogenen Gruppen dargestellt. Schattierungen in der gleichen **Spalte** bedeuten, dass sich die grau markierten Gruppen NICHT voneinander unterscheiden.

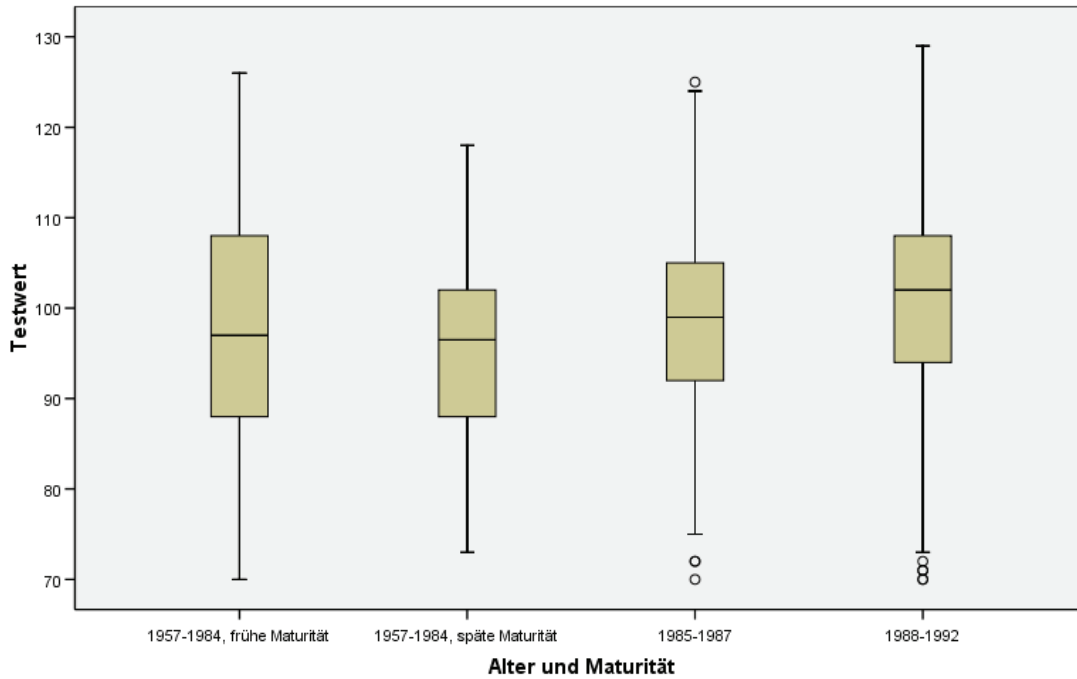


Abbildung 46: Boxplot für Testwert der Altersgruppen (Geburtsjahre) und Maturität.

6.7 Vergleiche nach Wunschuniversitäten

Die Leistungen der Kandidaten unterscheiden sich 2008 nicht signifikant hinsichtlich der bevorzugten Universitäten.

	Wunsch-universität	n	Mittelwert	Standard-abw.	Quadrat-summe	df	MQ	F	Sig.
Gesamt-stichprobe	Bern	603	99.78	10.152	351.559	3	117.19	1.18	.317
	Basel	390	99.83	9.825					
	Freiburg	113	98.81	9.148					
	Zürich	855	100.42	10.022					
Deutsch-sprachig, Humanmedizin	Bern	403	100.93	10.028	188.18	3	62.73	.64	.592
	Basel	334	100.64	9.560					
	Freiburg	36	102.36	8.945					
	Zürich	639	101.37	10.115					

Tabelle 39: Testwerte nach Wunschuniversitäten für Gesamtstichprobe und Humanmedizin, deutsch

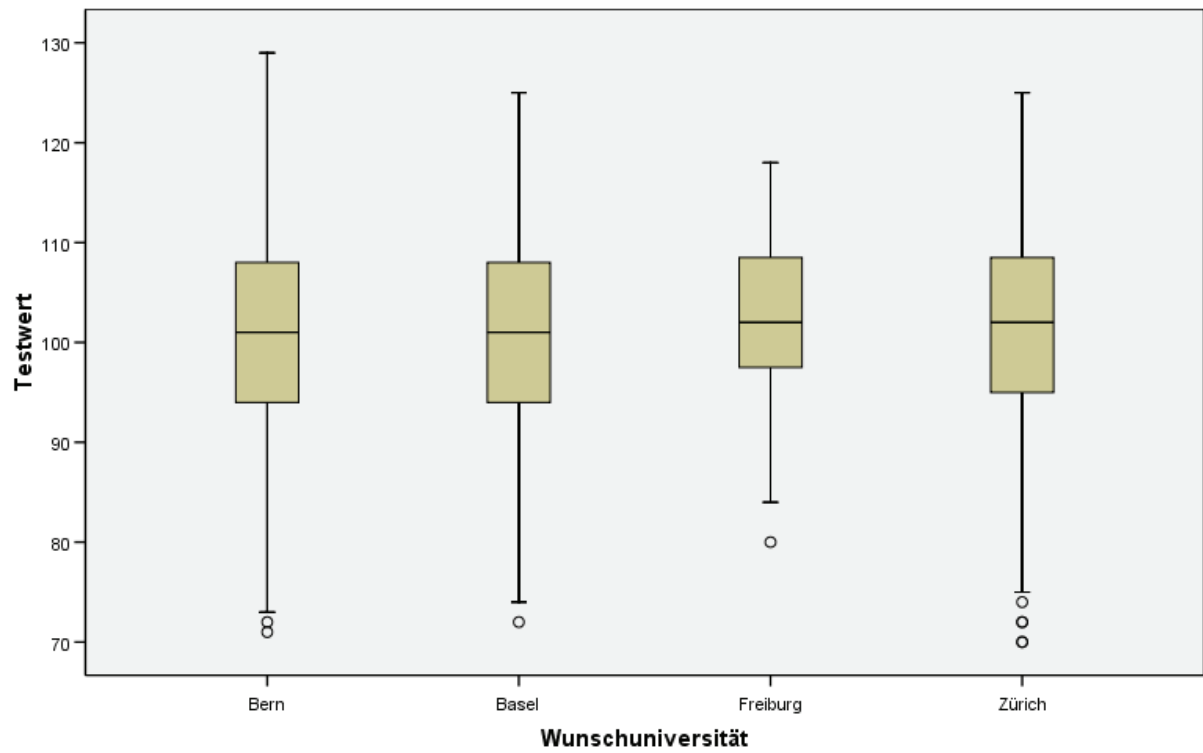


Abbildung 47: Boxplot für Testwerte nach Wunschuniversitäten (HM, deutsch)

7 Ergebnisse zur Testgüte

7.1 Zuverlässigkeit

Die Reliabilität kann anhand zweier Koeffizienten berechnet werden. Die Reliabilitätschätzung nach der Testhalbierungsmethode (Teilung nach gerad- und ungeradzahligem Aufgaben) ist eine der gebräuchlichsten Zuverlässigkeitsschätzungen. Wegen der Stichprobengrößen werden nur die Ergebnisse des deutschsprachigen Tests der Schweiz verwendet. Die internen Konsistenzen (Cronbach Alpha) schätzen die Messgenauigkeit anhand der Korrelationen jeder Aufgabe mit allen anderen des entsprechenden Untertests.

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

Die Zuverlässigkeitswerte des Punktwertes liegen 2008 mit 0.92 im gewohnt hohen Bereich der Vorjahre. Auch die Konsistenzen der Testprofile liegen mit 0.82/0.83 im üblichen Bereich. Hierbei ist zu beachten, dass zu hohe Werte für sehr gleichartige Untertests (mit der Frage, ob man einzelne weglassen kann) sprechen würden, sehr niedrige Werte für eine heterogene Testbatterie, die nicht ohne weiteres zu einem Testwert zusammengefasst werden dürfte. Der Bereich um 0.80 scheint deshalb optimal, weil vergleichbare Werte auch in den Jahren mit einer erfolgreichen Evaluation des Zusammenhanges von Studienerfolg und Eignungstest gefunden worden sind. Dieser Wertebereich wird in den Folgejahren eingehalten. Es bestehen somit keine Einwände seitens der Zuverlässigkeit der Messung, den Punktwert bzw. Testwert für die Eignungsmessung zu verwenden.

	Reliabilität nach Testhalbierungsmethode											
	DE	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
DE / CH	.91 - .93	.92	.91	.90	.91	.91	.92	.91	.92	.90	.90	.92
AT										.92	.93	.92

Tabelle 40: Reliabilität des Punktwertes (Split Half) für Deutschland, Schweiz und Österreich.

	Konsistenz des Testprofiles										
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
CH	.83	.81	.81	.80	.80	.81	.78	.82	.79	.80	.82
AT									.83	.84	.83

Tabelle 41: Konsistenz des Testprofils für Schweiz und Österreich.

Die Zuverlässigkeitswerte der Untertestwerte wurden untersucht, um Bewertungen der Untertests vorzunehmen – sie sind für die Güte des Testwertes selbst nur indirekt bedeutsam. Hier müssten lediglich Eingriffe erfolgen, wenn der Testwert selbst eine zu geringe Zuverlässigkeit aufweisen würde (was aber ausgeschlossen werden kann).

Insgesamt werden mit den Vorjahren vergleichbare Kennwerte erreicht.

Untertest	Reliabilität nach Testhalbierungsmethode											
	Deutschland	Schweiz (d)										
		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Muster zuordnen	.67 - .73	.72	.61	.62	.59	.56	.55	.57	.70	.65	.62	.66
Med.-nat. Grundverständnis	.65 - .72	.72	.71	.60	.69	.57	.73	.67	.73	.74	.65	.71
Schlauchfiguren	.75 - .82	.75	.74	.74	.72	.67	.76	.74	.73	.71	.75	.74
Quant. u. form. Probleme	.70 - .75	.78	.75	.74	.74	.71	.68	.72	.76	.71	.72	.71
Textverständnis	.69 - .74	.75	.70	.75	.75	.71	.78	.70	.75	.68	.69	.69
Figuren lernen	.72 - .75	.72	.70	.62	.62	.81	.78	.73	.74	.68	.64	.78
Fakten lernen	.68 - .73	.70	.70	.61	.72	.69	.64	.72	.73	.75	.72	.76
Diagramme und Tabellen	.68 - .72	.66	.62	.61	.63	.64	.62	.68	.71	.61	.70	.65
Planen und Organisieren									.63	.42	.68	.64

Tabelle 42: Zuverlässigkeit 1998 bis 2008; Vergleich der Reliabilitäten nach der Testhalbierungsmethode für die Untertests, Ergebnisse Deutschland aus Trost et al. (1997), S. 26 ff.

Untertest	Innere Konsistenz (Cronbachs Alpha)											
	Deutschland	Schweiz (d)										
		1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Muster zuordnen	.64 - .69	.69	.60	.53	.56	.57	.53	.54	.67	.60	.55	.61
Med.-nat. Grundverständnis	.64 - .73	.70	.70	.61	.66	.57	.72	.66	.72	.71	.65	.70
Schlauchfiguren	.76 - .80	.71	.74	.71	.73	.68	.74	.72	.72	.70	.73	.74
Quant. u. form. Probleme	.69 - .71	.76	.75	.76	.74	.71	.67	.71	.76	.71	.70	.70
Textverständnis	.68 - .71	.73	.75	.74	.70	.68	.78	.67	.73	.66	.69	.64
Figuren lernen	.69 - .74	.72	.70	.62	.64	.82	.77	.71	.74	.66	.63	.74
Fakten lernen	.62 - .70	.68	.72	.64	.69	.70	.73	.72	.72	.73	.73	.74
Diagramme und Tabellen	.67 - .71	.67	.75	.63	.61	.62	.70	.66	.70	.61	.69	.62
Planen und Organisieren									.62	.42	.63	.64

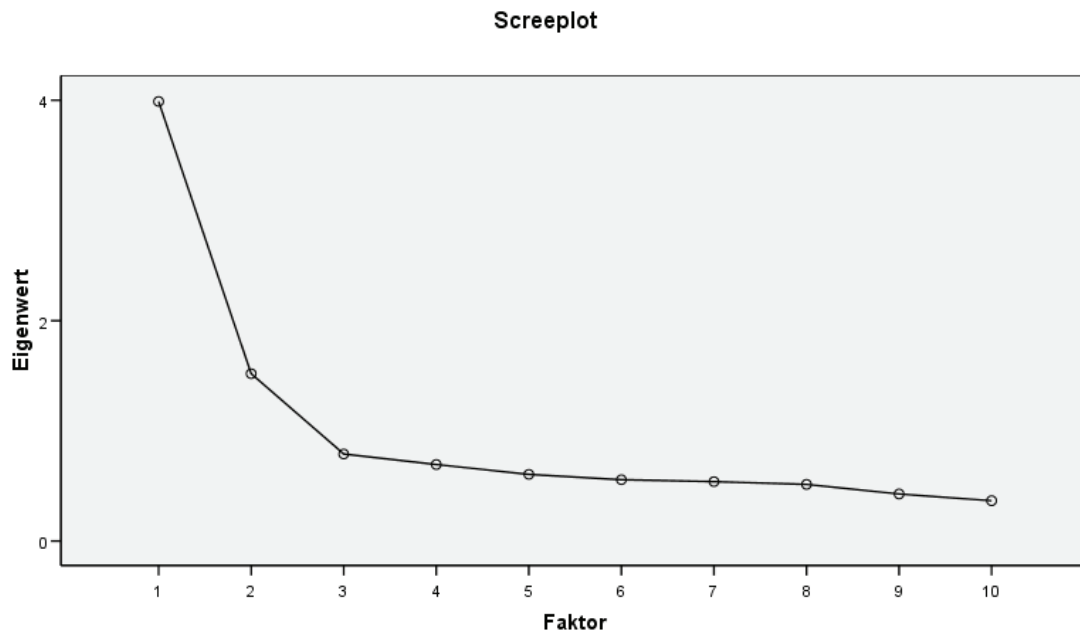
Tabelle 43: Zuverlässigkeit 1998 bis 2008; Vergleich der Skalenkonsistenzen für die Untertests, Ergebnisse für Deutschland aus Trost et al. (1997), S. 26 ff.

7.2 Binnenstruktur

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

	Muster zuordnen	Med.-naturwis. Grundv.	Schlauchfiguren	Quant. u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Planen und Organisieren	Punkt-wert CH	Punkt-wert AT
Muster zuordnen		0.23	0.42	0.23	0.19	0.34	0.32	0.27	0.34	0.21	0.56	0.58
Med.-naturw. Grundverst.	0.27		0.32	0.53	0.60	0.26	0.27	0.61	0.20	0.47	0.70	0.73
Schlauchfiguren	0.43	0.38		0.32	0.28	0.40	0.33	0.32	0.33	0.25	0.63	0.67
Quant. u. formale Probl.	0.26	0.56	0.39		0.46	0.20	0.22	0.54	0.19	0.44	0.65	0.68
Textverständnis	0.21	0.61	0.32	0.49		0.23	0.27	0.52	0.16	0.44	0.64	0.64
Figuren lernen	0.37	0.28	0.41	0.23	0.19		0.47	0.26	0.35	0.21	0.61	0.60
Fakten lernen	0.31	0.33	0.31	0.24	0.26	0.44		0.24	0.29	0.23	0.59	0.58
Diagramme und Tabellen	0.31	0.62	0.39	0.58	0.54	0.26	0.29		0.21	0.50	0.69	0.72
Konzent. u. sorgfält. Arbeiten	0.37	0.19	0.33	0.20	0.16	0.33	0.27	0.20		0.20	0.58	0.55
Planen und Organisieren	0.24	0.50	0.31	0.46	0.43	0.22	0.25	0.53	0.21		0.62	0.63

Tabelle 44: Korrelationen zwischen Punktwerten der Untertests CH (Schweiz, über der Diagonale) und AT (Österreich, unter der Diagonale) sowie mit dem Gesamtwert.



Die Abbildung 48: Scree-Plot (Eigenwerteverlauf) der Faktoren 1 bis 10 für die Untertests des EMS 2008.

Die Korrelationen der einzelnen Untertests mit der Gesamtpunktzahl bleiben über die Jahre recht stabil. Der Scree-Plot stellt den Eigenwerteverlauf der Faktoren grafisch dar. Die Ein-Faktorenlösung bleibt deutlich hervorgehoben.

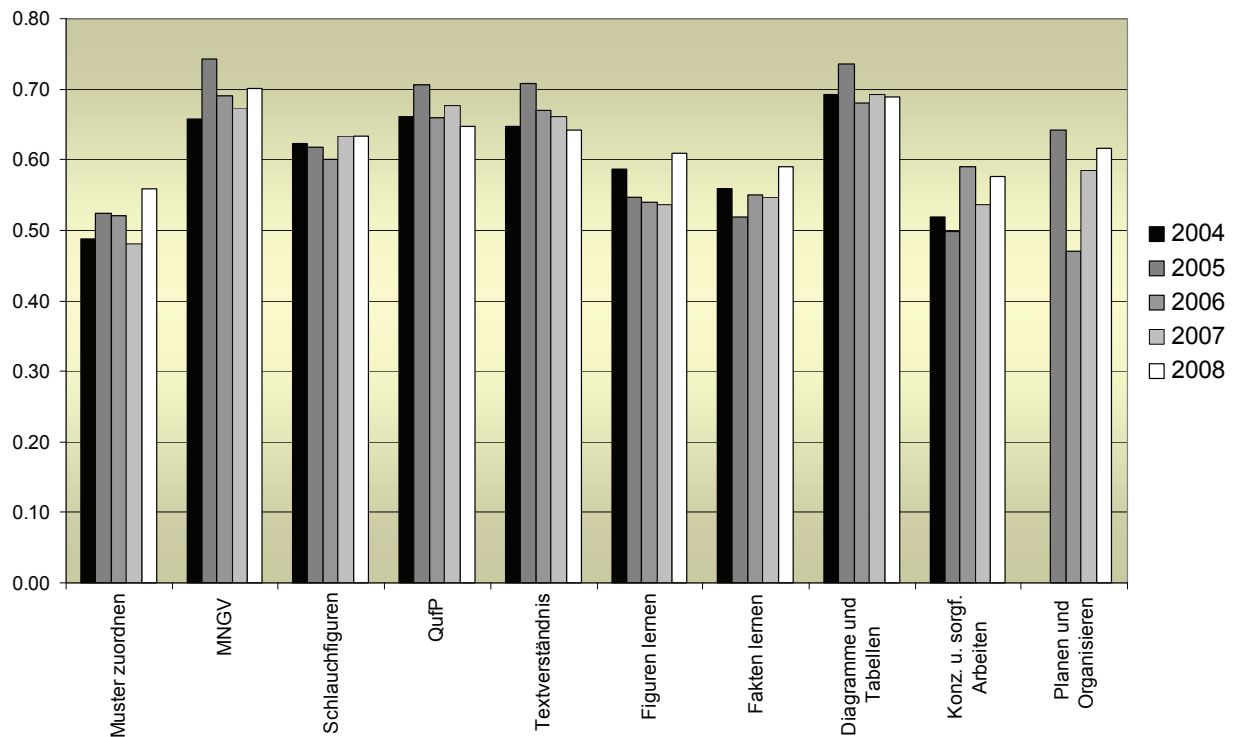


Abbildung 49: Korrelationen zwischen Punktwerten und Untertests für die letzten fünf Jahre.

Unrotierte Lösung															
Faktor	Eigenwert					% Varianz					Kumuliert %				
	2008	2007	2006	2005	2004	2008	2007	2006	2005	2004	2008	2007	2006	2005	2004
1	3.99	3.73	3.65	4.04	3.36	39.89	37.32	36.50	40.38	37.3	39.9	37.32	36.50	40.38	37.3
2	1.52	1.60	1.61	1.48	1.39	15.19	15.98	16.10	14.77	15.4	55.1	53.29	52.61	55.14	52.7
3	0.79	0.84	0.85	0.91	0.9	7.90	8.35	8.54	9.15	10.1	63	61.65	61.15	64.29	62.8
4	0.69	0.72	0.77	0.71	0.82	6.94	7.22	7.72	7.09	9.06	69.9	68.87	68.87	71.38	71.9
5	0.61	0.65	0.65	0.61	0.62	6.05	6.50	6.49	6.06	6.9	76	75.37	75.37	77.44	78.8
6	0.56	0.63	0.61	0.58	0.52	5.57	6.32	6.08	5.82	6.36	81.5	81.69	81.45	83.26	85.1
7	0.54	0.57	0.56	0.53	0.54	5.39	5.75	5.55	5.26	6.03	86.9	87.43	87.00	88.52	91.1
8	0.51	0.49	0.52	0.47	0.47	5.14	4.90	5.19	4.69	4.62	92.1	92.33	92.19	93.21	95.8
9	0.43	0.41	0.41	0.37	0.38	4.27	4.12	4.11	3.74	4.24	96.3	96.45	96.30	96.95	100
10	0.37	0.35	0.37	0.30		3.66	3.55	3.70	3.05		100	100	100	100	

Tabelle 45: Faktorenanalyse: Verlauf der unrotierten Lösung (Schweiz 2004-2008).

Varimax-rotierte Lösungen														
Eigenwert					% Varianz					Kumuliert %				
2008	2007	2006	2005	2004	2008	2007	2006	2005	2004	2008	2007	2006	2005	2004
Zwei-Faktorenlösung														
3.05	3.05	2.91	3.26	2.63	30.5	30.54	29.12	32.58	29.26	30.5	30.54	29.12	32.58	29.26
2.46	2.28	2.35	2.27	2.11	24.6	22.75	23.49	22.70	23.48	55.1	53.29	52.61	55.28	52.74
Drei-Faktorenlösung														
3.03	3.05	2.91	3.27	2.56	30.3	30.46	29.05	32.74	28.39	30.3	30.46	29.05	32.74	28.39
1.81	1.68	1.81	1.73	1.93	18.1	16.80	18.13	17.31	21.43	48.4	47.26	47.18	50.05	49.81
1.46	1.44	1.40	1.43	1.17	14.6	14.39	13.97	14.33	12.97	63	61.65	61.15	64.38	62.79

Tabelle 46: Varianzanteile der einzelnen Faktorenlösungen Schweiz 2004-2008 (rotierte Lösungen).

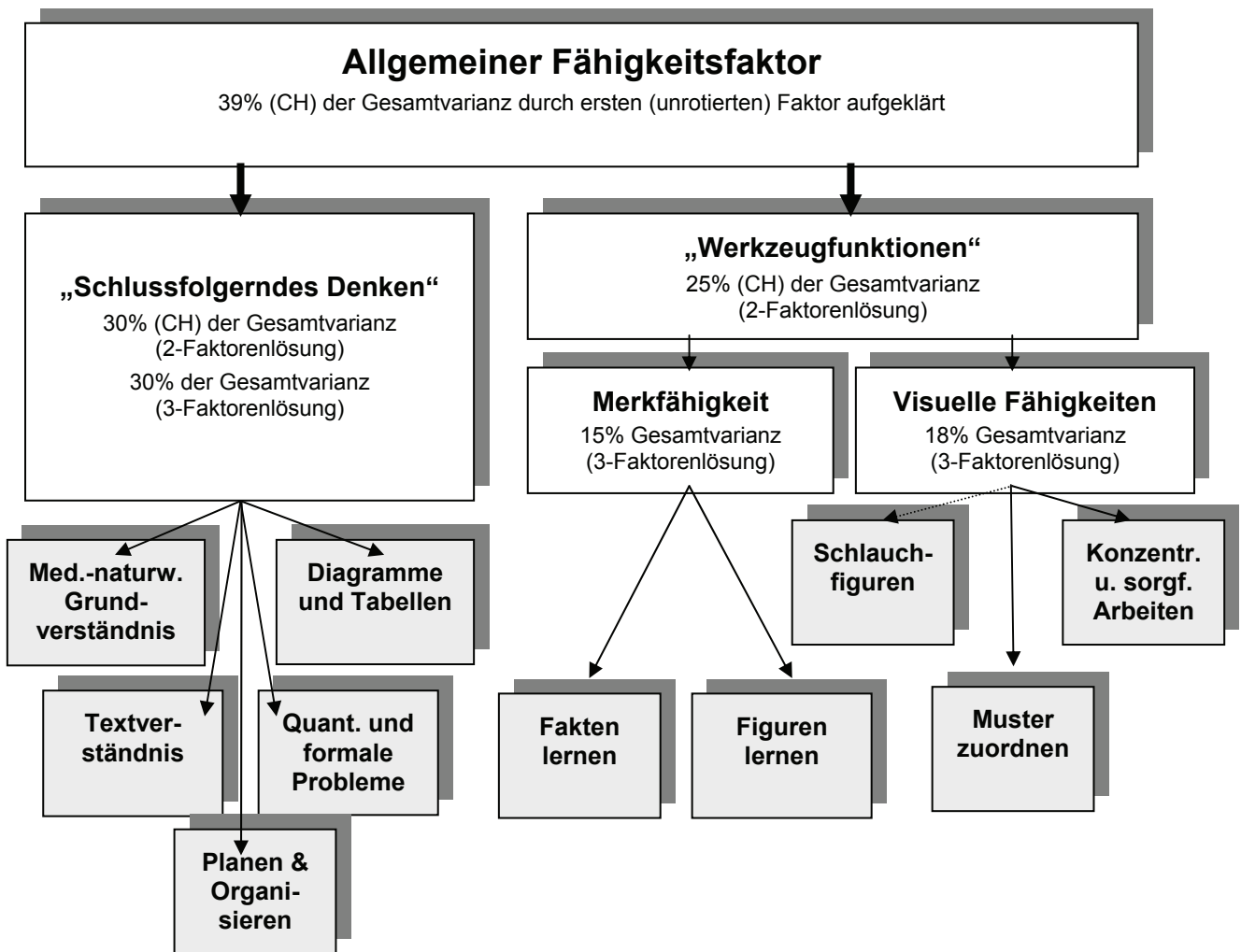


Abbildung 50: Struktur der Untertests des EMS, hierarchische Faktorenstruktur 2008 CH (Schweiz) und AT (Österreich).

Die Struktur des EMS bleibt weiterhin stabil – trotz Einführung eines neuen Untertests („Planen und Organisieren“, 2005) und der laufenden Neufassung der Instruktionen zum Untertest „Konzentriertes und sorgfältiges Arbeiten“.

Die Leistungen im gesamten Test werden in Form eines „allgemeinen Fähigkeitsfaktors“ am stärksten durch die Untertests „medizinisch-naturwissenschaftliches Grundverständnis“, „Diagramme und Tabellen“, „Textverständnis“ sowie „Quantitative und formale Probleme“ charakterisiert.

In der Zweifaktorenlösung stehen die Untertests „Diagramme und Tabellen“ und „medizinisch-naturwissenschaftliches Grundverständnis“ für den Faktor „Schlussfolgerndes Denken“ und der Untertest „Figuren lernen“ als typischer Repräsentant für den Faktor „Werkzeugfunktionen“ hervor.

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

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Muster zuordnen	.56	.53	.57	.57	.71	.48	.45	.46	.47	.43	.53
Med.-naturwiss. Grundverständnis	.76	.76	.74	.74	.7	.74	.71	.79	.74	.74	.75
Schlauchfiguren	.69	.64	.65	.62	.69	.61	.61	.58	.57	.60	.61
Quant. und formale Probleme	.7	.7	.72	.66	.68	.73	.7	.75	.71	.73	.68
Textverständnis	.72	.7	.71	.72	.62	.77	.7	.76	.72	.72	.69
Figuren lernen	.62	.59	.53	.56	.62	.55	.54	.48	.47	.48	.56
Fakten lernen	.53	.52	.55	.52	.57	.51	.51	.47	.48	.48	.55
Diagramme und Tabellen	.72	.75	.71	.72	.56	.76	.74	.78	.74	.76	.74
Konzentr. u. sorgf. Arbeiten	.47	.51	.55	.49	.54	.48	.44	.41	.52	.43	.48
Planen und Organisieren								.68	.51	.62	.65

Tabelle 47: Ladungen im ersten Faktor der unrotierten Lösung 1998 bis 2007 („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor).

	Faktor 1						Faktor 2						Kommunalitäten (h ²)					
	03	04	05	06	07	08	03	04	05	06	07	08	03	04	05	06	07	08
Diagramme und Tabellen	.83	.79	.81	.80	.82	.79						.20	.73	.66	.70	.66	.69	.66
Med.-naturwiss. Grundverständnis	.86	.80	.82	.77	.79	.81							.76	.65	.71	.64	.64	.69
Textverständnis	.83	.70	.81	.77	.74	.76							.74	.63	.68	.62	.58	.60
Quant. und formale Probleme	.77	.70	.78	.79	.79	.74	.20						.65	.60	.64	.64	.64	.58
Planen und Organisieren			.72	.61	.69	.70									.54	.37	.49	.52
Konzentr. u. sorgf. Arbeiten							.79	.33	.51	.62	.64	.66	.62	.20	.28	.41	.41	.45
Figuren lernen							.78	.74	.74	.73	.69	.74	.65	.56	.56	.54	.49	.57
Muster zuordnen						.15	.58	.70	.74	.66	.67	.68	.62	.47	.56	.45	.46	.48
Fakten lernen	.26					.18	.50	.67	.62	.64	.64	.66	.78	.47	.40	.42	.44	.47
Schlauchfiguren	.26		.29		.29	.27	.66	.67	.62	.68	.65	.65	.60	.52	.47	.50	.50	.50

Tabelle 48: Ladungen und Kommunalitäten der Zwei-Faktorenlösung 2003 bis 2008, varimaxrotiert.

Faktor	Schweiz								Deutschland		
	1		2		3		h ²		1	2	3
	2007	2008	2007	2008	2007	2008	2007	2008			
Diagramme und Tabellen	.82	.79		.21			.67	.67	.82		
Medizin.-naturwiss. Grundverständnis	.79	.81					.69	.69	.81		.2
Quantitat. u. formale Probleme	.79	.74		.22			.64	.59	.80	.18	
Textverständnis	.73	.76				.23	.66	.63	.79		.2
Planen und Organisieren	.69	.70					.42	.52	-	-	-
Muster zuordnen			.76	.78			.69	.65		.81	
Konzentr. u. sorgfält. Arbeiten			.74	.69		.21	.63	.52		.70	.4
Schlauchfiguren		.26	.61	.67	.28	.21	.56	.56	.35	.71	
Fakten lernen					.83	.84	.73	.77	.21		.87
Figuren lernen				.35	.76	.74	.63	.69	.13	.47	.64

Tabelle 49: Faktorenanalyse: Varimaxrotierte Drei-Faktorenlösung, Schweiz 2007 und 2008 sowie Vergleichswerte Deutschland (deutsche Daten nach Blum, 1996, in Trost et al., 1998, S. 42).

7.3 Item-Trennschärfen

Die Item-Trennschärfen sind die Korrelationen des Punktwertes für den jeweiligen Untertest mit den zugeordneten Items. Positive Korrelationen weisen darauf hin, dass die Leistungsbesten im jeweiligen Untertest auch beim entsprechenden Item die richtige Lösung bevorzugt gewählt haben. Zu beachten ist, dass bei sehr leichten und sehr schwierigen Items wegen der geringeren Antwortvarianz auch die Trennschärfe in der Regel geringer ausfallen wird. Negative Trennschärfen würden auf Items hinweisen, die missverständlich formuliert sind oder keine eindeutige Lösung haben – die Leistungsbesten eine andere als die vorgegebene Lösung gewählt haben. Entsprechend der festgelegten Auswertedirektive des EMS werden solche Items von der Auswertung nachträglich ausgeschlossen und nicht gewertet, um Risiken der Fehlbewertung auszuschliessen. Dieser Fall traf 2008 nicht ein.

Unterschiede zu den (theoretischen) Kennwerten aus Deutschland können auch durch die Positionierung der Items innerhalb des Untertests mitbegründet sein. So werden etwa Items am Ende eines Untertests weniger häufig gelöst (aufgrund fehlender Zeit), was neben ihrer Schwierigkeit auch die Trennschärfe beeinflusst.

Die nachfolgenden Abbildungen zeigen die Trennschärfen aller Items.

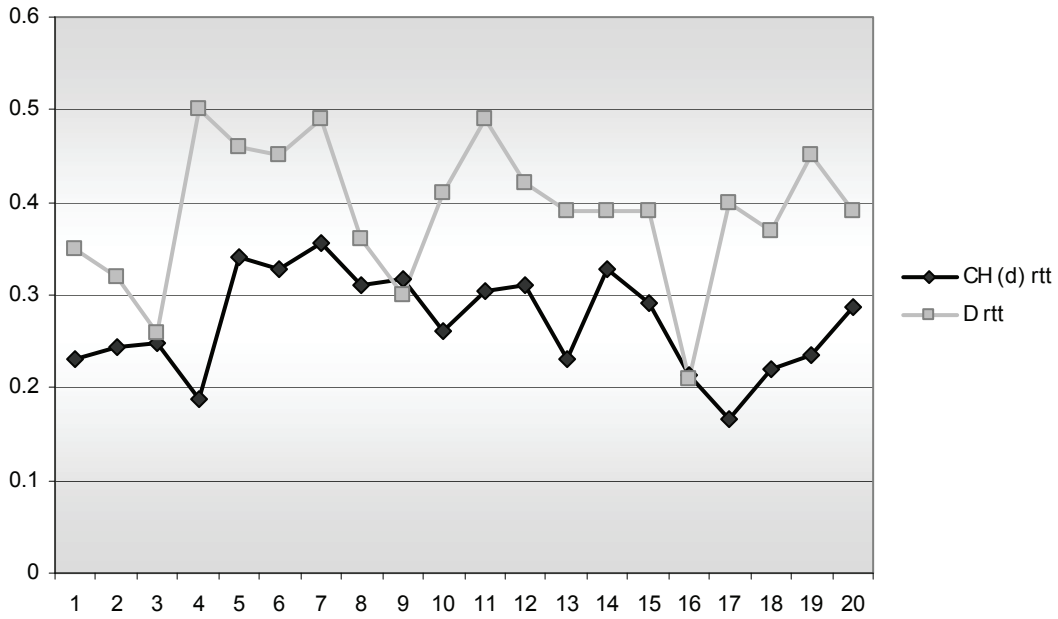


Abbildung 51: Trennschärfen für den Untertest „Quantitative und formale Probleme“.

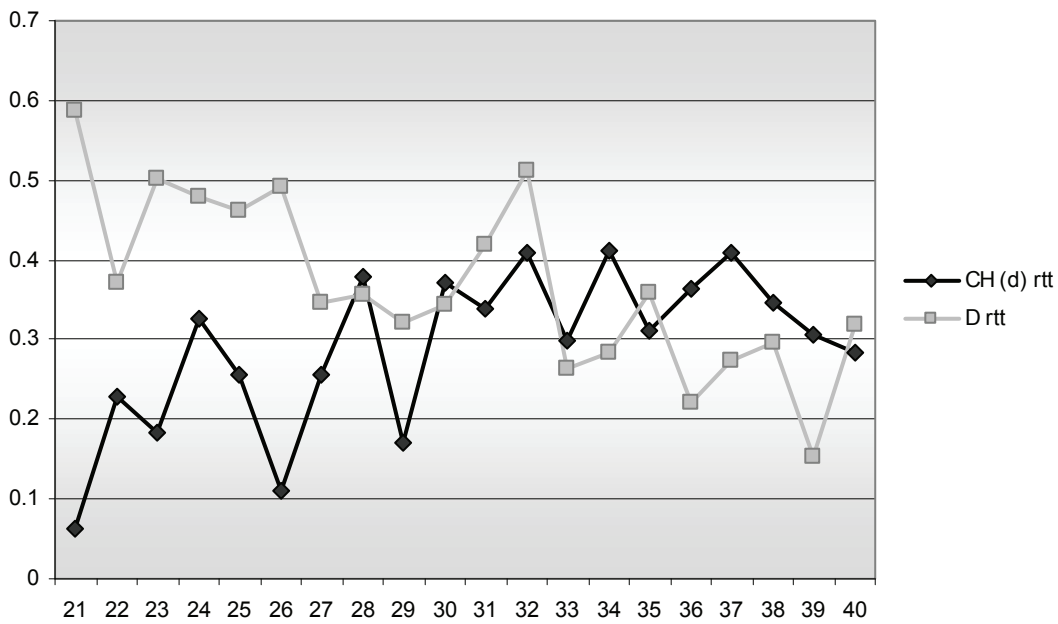


Abbildung 52: Trennschärfen für den Untertest „Schlauchfiguren“.

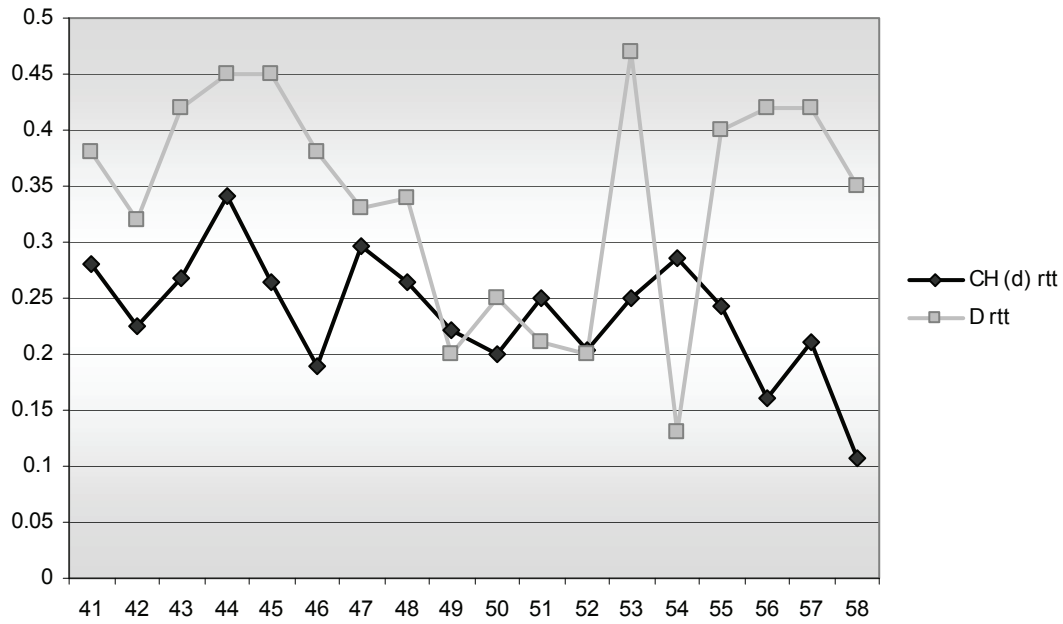


Abbildung 53: Trennschärfen für den Untertest „Textverständnis“.

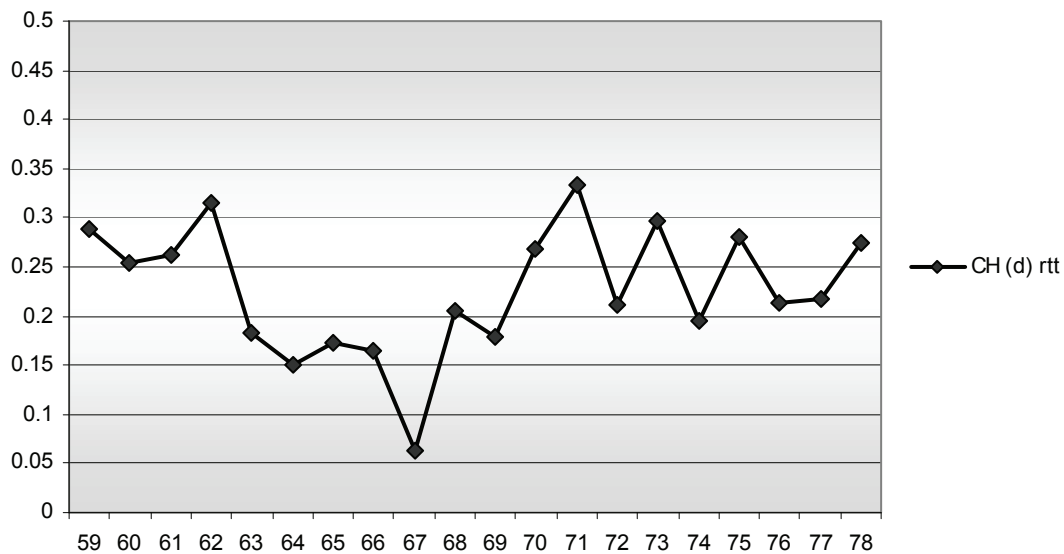


Abbildung 54: Trennschärfen für den Untertest „Planen und Organisieren“ (keine deutschen Vergleichswerte).

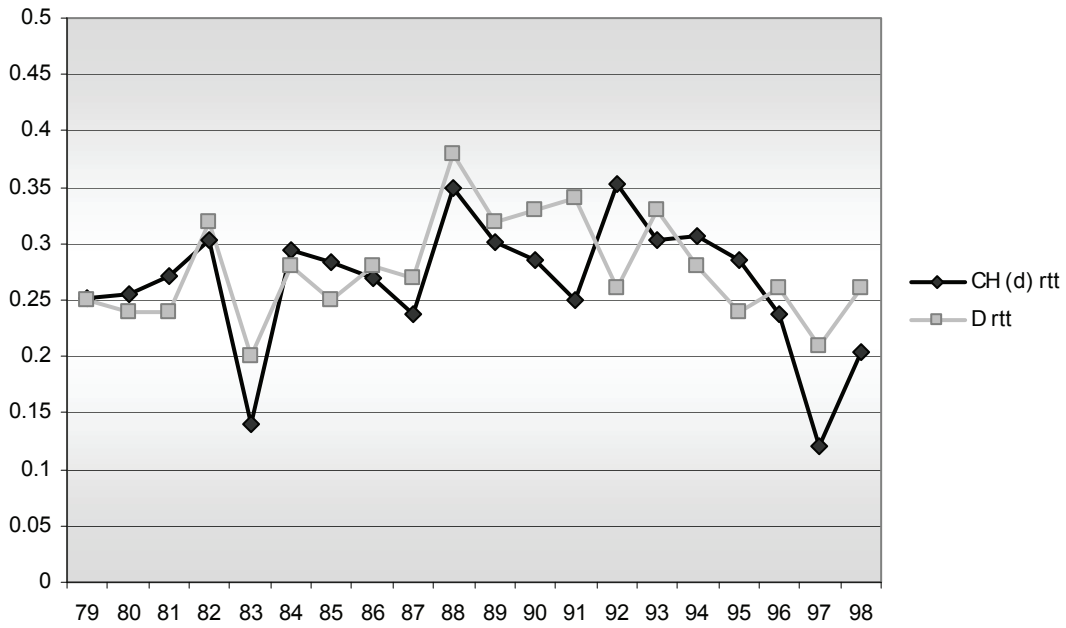


Abbildung 55: Trennschärfen für den Untertest „Medizinisch-naturwissenschaftliches Grundverständnis“.

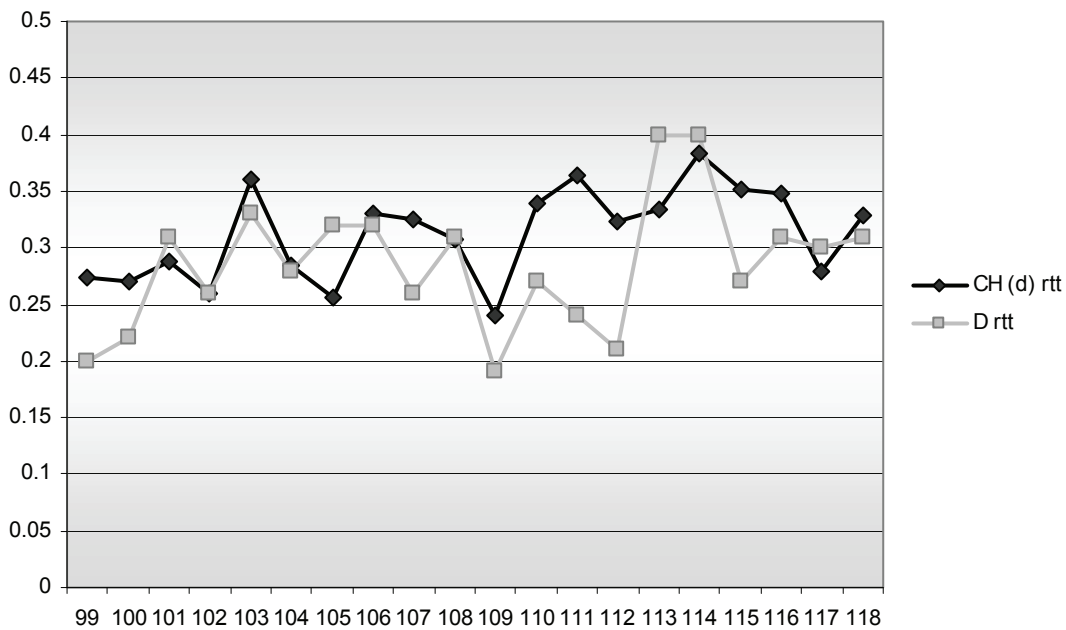


Abbildung 56: Trennschärfen für den Untertest „Figuren lernen“.

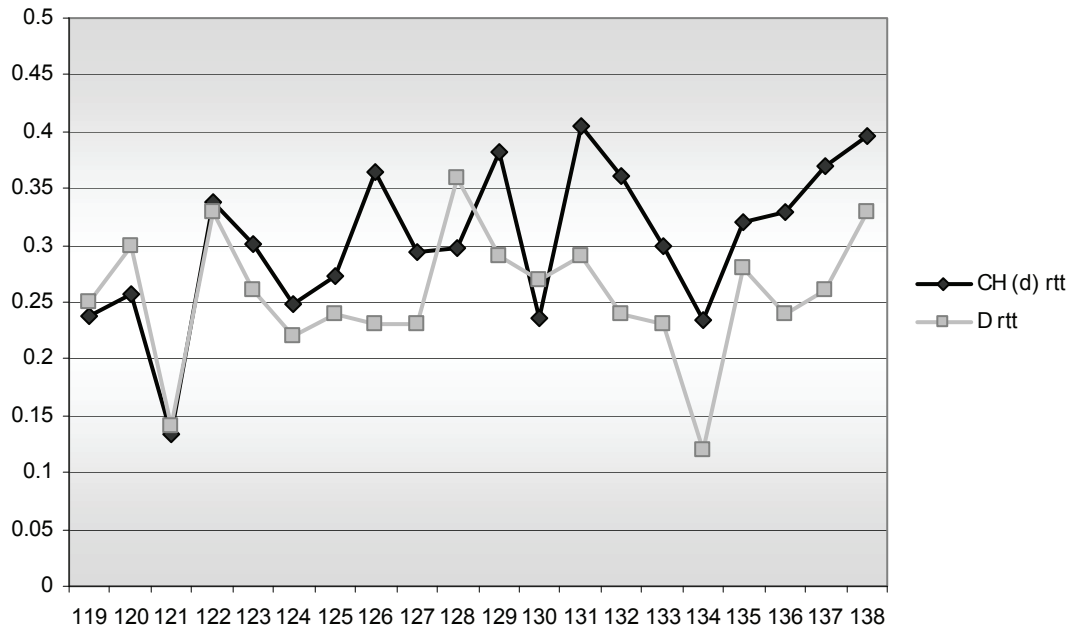


Abbildung 57: Trennschärpen für den Untertest „Fakten lernen“.

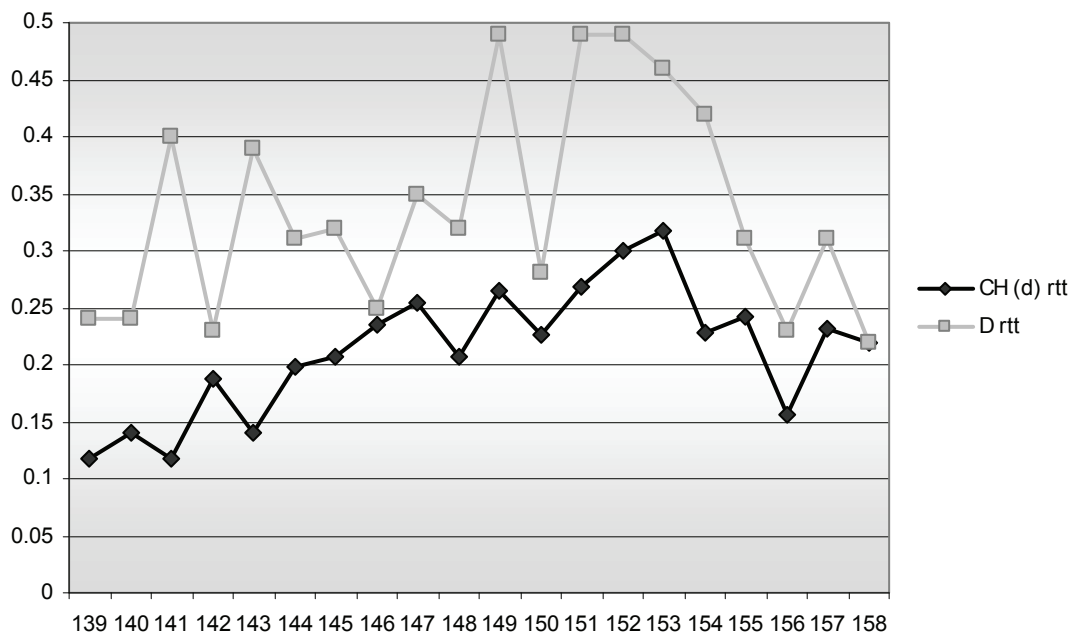


Abbildung 58: Trennschärpen für den Untertest „Muster zuordnen“.

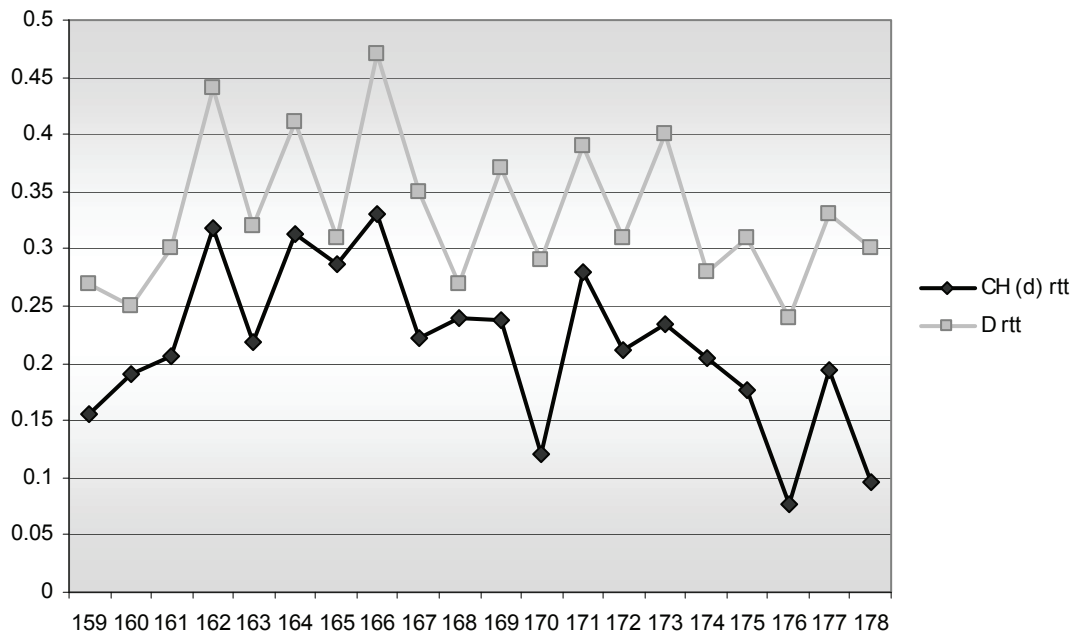


Abbildung 59: Trennschärfen für den Untertest „Diagramme und Tabellen“.

7.4 Konzentriertes und sorgfältiges Arbeiten

Dieser Untertest wurde 2004 revidiert und jährlich mit einer variierenden Anforderung (Zeichen sowie Durchstreichregel) durchgeführt, die vorher nicht bekannt ist. Es hatte sich früher gezeigt, dass durch exzessives Üben (40 bis teilweise 80mal) bei diesem Test ein so hoher Automatisierungsgrad erreicht wurde, dass nahezu alle Zeichen bearbeitet werden konnten (vgl. Bericht 9 des ZTD und Vorbereitungsreport 2003). Dieses Ergebnis ist nicht im Sinne der Messintention dieses Tests: Es soll kein „Fleisstest“ sein, sondern tatsächlich Konzentration und Sorgfalt gemessen werden.

Aus der Literatur ist bekannt, dass hochspezialisierte Automatisierungen stark bedingungsabhängig sind – ändert sich etwas, müssen diese Automatismen neu antrainiert werden (wie etwa bei der Fließbandarbeit). Indem nun jedes Jahr Zeichen und Regel vorher nicht bekannt sein werden, wirken sich Lerneffekte durch exzessives Üben nicht mehr so deutlich aus. In den Jahren 2004 bis 2006 findet sich eine glockenförmige Verteilung des Punktwertes – im Unterschied zu stark rechtsschiefen Verteilungen in den Vorjahren.

Ab 2006 wurde auch in der Test-Info eine Form bereitgestellt, die formal der neuen Version entspricht. Die trotzdem weiter empfohlene Vorbereitung mit der veröffentlichten Originalversion soll dem Zweck dienen, Erfahrungen zu sammeln, wie man 8 Minuten zusammenhängend konzentriert arbeiten kann und wie man die Markierungen vornehmen muss, damit sie regelkonform sind.

Die Zahl der Zeichen auf dem Arbeitsbogen wurde von 1200 auf 1600 erhöht. Dies war möglich, weil die neue Einlesetechnik keine Markierungszonen um das eigentliche Zeichen mehr verlangt – dadurch wurde der Bogen insgesamt sogar übersichtlicher. „Deckeneffekte“, d.h. dass alle Zeichen richtig gelöst werden, sind dadurch unwahrscheinlich und der Test differenziert über den gesamten Leistungsbereich gut. Die Sorgfalt beim Markieren (nicht in Nachbarzeichen markieren, deutlich markieren) gehört bei diesem Test mit zur Anforderung – entsprechende Regelverstöße werden als Fehler gewertet.

2008 bestand der Untertest aus den Zeichen „f“ und „t“. Es sollte jedes „f“ markiert werden, wenn direkt danach ein „t“ folgte.

	Beschreibung	Wertebereich
Richtige	Target richtig markiert (f vor t)	0...400
Fehler I	Target nicht markiert, „übersehen“	0...400
Fehler IIa	Nontarget markiert, „falscher Alarm“: (f markiert vor t)	0...400
Fehler IIb	Distraktor markiert, „falscher Alarm“: t markiert	0...800

Der Punktwert (PW) berechnet sich wie folgt:

$$\text{PW} = \text{Richtige} - \text{Fehler I} - \text{Fehler IIa} - \text{Fehler IIb}$$

Der Rohwert Richtige minus Fehler beträgt im Mittel 312.6 mit einer Standardabweichung von 68.7. Werden mehr Fehler als Richtige gemacht, entstehen negative Werte, die auf Null gesetzt werden.

Die Standardisierung zum Punktwert von 0 bis 20 erfolgt nach folgender Regel: Aus der Verteilung der Rohwerte erhalten die unteren 2.5% der Personen den Punktwert 0, die oberen 2.5% den Punktwert 20. Der Wertebereich für die mittleren 95% der Personen wird in 19 gleiche Abschnitte geteilt, sodass sich die nachfolgende Umrechnungstabelle ergibt. Dieses Verfahren gewährleistet, dass auch unterschiedlich schwierige Konzentrationstests zu einer vergleichbaren Bepunktung über die einzelnen Jahre führen und der differenzierte Wertebereich nicht durch Ausreisser eingeschränkt wird.

Punktwert 0-20	Von Rohwert	Bis Rohwert
0	0	203
1	204	213
2	214	223
3	224	233
4	234	243
5	244	254
6	255	264
7	265	274
8	275	284
9	285	294
10	295	304
11	305	314
12	315	324
13	325	334
14	335	344
15	345	355
16	356	365
17	366	375
18	376	385
19	386	395
20	396	400

Tabelle 50: Umrechnung Rohwert in Punktwert „Konzentriertes und sorgfältiges Arbeiten“ 2008

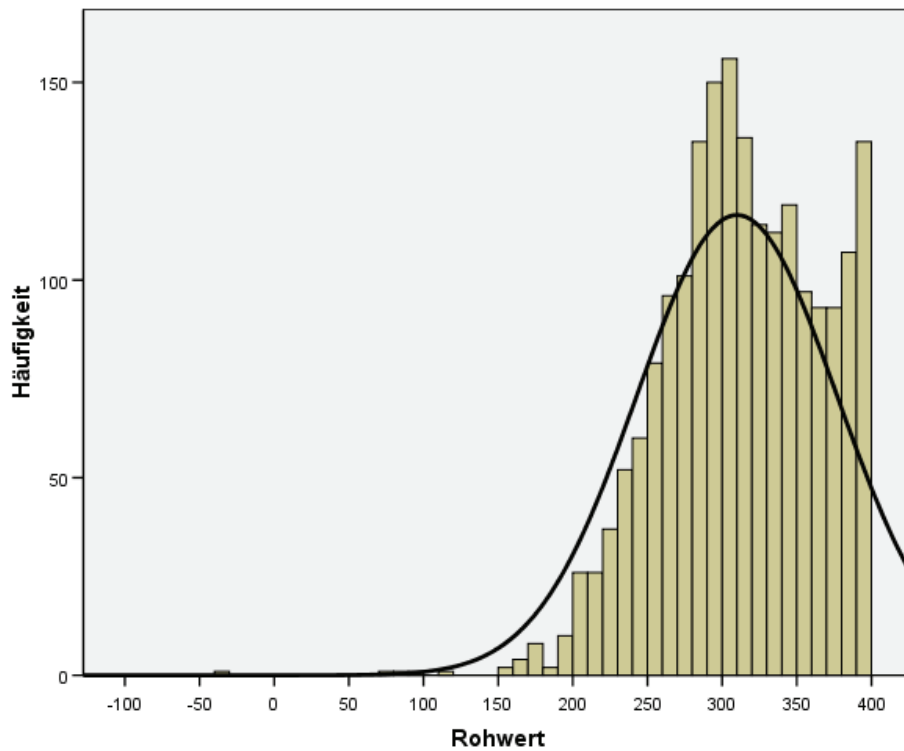


Abbildung 60: Verteilung des Rohwertes für „Konzentriertes und sorgfältiges Arbeiten 2008“, negative Werte nicht Null gesetzt. 6 Ausreisser liegen unter -100 (nicht dargestellt).

		Mittelwert	Standardabweichung	Minimum	Maximum
Deutsch	Richtige	324.27	52.32	6	400
	Fehler	11.69	33.87	0	753
	... Auslassungen (I)	5.83	18.34	0	384
	... Fehlmarkierungen (II)	5.87	17.28	0	369
	letztes bearbeitetes Zeichen	1317.99	203.91	555	1600
Französisch	Richtige	290.49	49.04	171	400
	Fehler	7.28	7.31	0	36
	... Auslassungen (I)	3.20	5.16	0	32
	... Fehlmarkierungen (II)	4.08	3.77	0	25
	letztes bearbeitetes Zeichen	1172.93	200.63	689	1597
Italienisch	Richtige	305.80	47.20	196	399
	Fehler	7.74	6.58	0	30
	... Auslassungen (I)	3.05	3.85	0	20
	... Fehlmarkierungen (II)	4.69	4.10	0	17
	letztes bearbeitetes Zeichen	1233.51	187.82	790	1597
Gesamt	Richtige	321.12	52.70	6	400
	Fehler	11.22	32.06	0	753
	... Auslassungen (I)	5.53	17.39	0	384
	... Fehlmarkierungen (II)	5.69	16.36	0	369
	letztes bearbeitetes Zeichen	1304.30	206.91	555	1600

Tabelle 51: Sprachspezifische Statistiken für Parameter des Untertests „Konzentriertes und sorgfältiges Arbeiten“ und Gesamt.

	Fehler gesamt	Auslassungen (I)	Fehlmarkierungen (II)
Richtige	-.21(**)	-.21(**)	-.19(**)
Fehler gesamt		.95(**)	.95(**)
Auslassungen (I)	.95(**)		.80(**)

Tabelle 52: Korrelationen der Parameter im Konzentrationstest; **Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

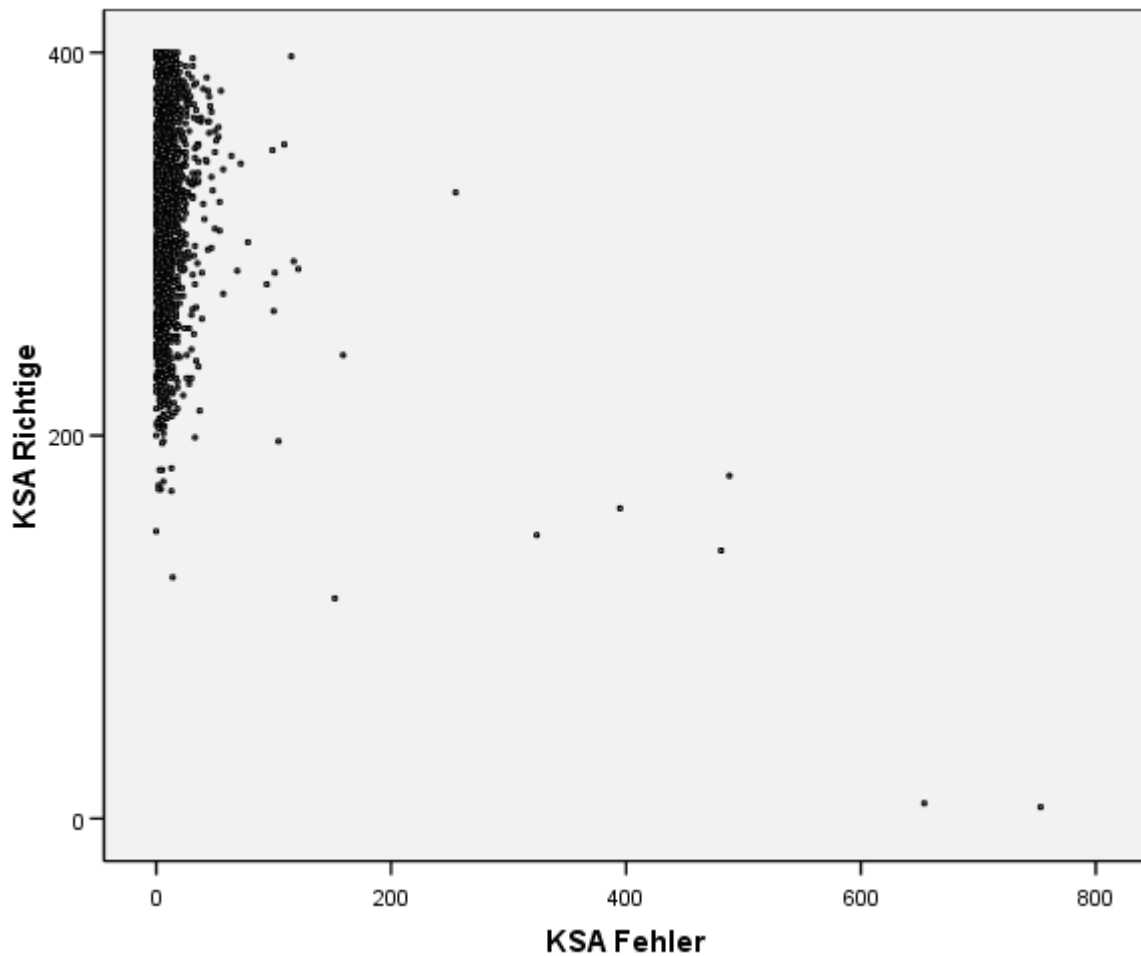


Abbildung 61: Scatterplot für alle Personen – Mengenleistung (Richtige) und Genauigkeit (niedrige Fehlerzahl).

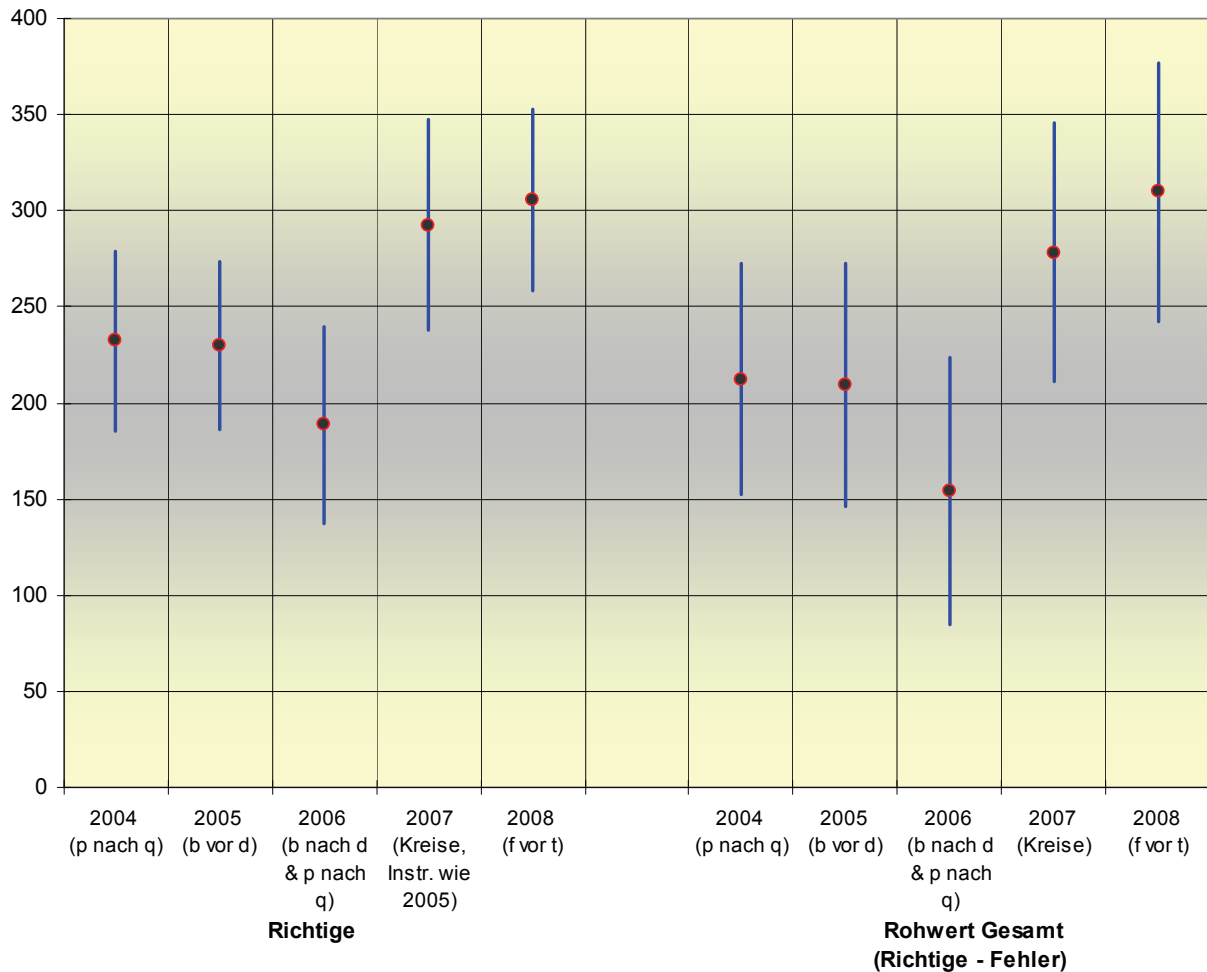


Abbildung 62: Leistungen „Konzentriertes und sorgfältiges Arbeiten“ ($m \pm s$) seit Einführung variierender Instruktionen und Zeichen (2004)

Die Ausgabe 2008 des Untertests „Konzentriertes und sorgfältiges Arbeiten“ stellt eine der leichtesten unter den bisher verwendeten Varianten dar. Dies ist einerseits auf die einfache Instruktion (gegenüber 2006) aber auch auf die relativ gute Unterscheidbarkeit der verwendeten Zeichen zurückzuführen. Die Streuung der Ergebnisse bleibt mit den Vorjahren vergleichbar – damit ist die Differenzierung zwischen den Kandidaten gewährleistet. Zur Vermeidung eines Deckeneffekts sollte der Untertest aber in Zukunft wieder leicht schwieriger gestaltet werden.

8 Teilnahmen am EMS nach Kantonen seit 1998

Seit 1998 haben bisher 13'701 Personen am EMS teilgenommen. Zu beachten ist, dass seit 1998 Humanmedizin, seit 1999 Veterinärmedizin und erst seit 2004 Zahnmedizin einem NC an den Universitäten Basel, Bern, Freiburg und Zürich unterliegen¹.

Wohnkanton/ Wohnort	Disziplin			Geschlecht		Sprache			Total
	Human- medizin	Veterinär- medizin	Zahn- medizin	männlich	weiblich	deutsch	franzö- sisch	italie- nisch	
AG	870	153	60	395	688	1082	1	0	1083
AI	20	7	4	10	21	31	0	0	31
AR	80	12	9	41	60	101	0	0	101
BE	1703	319	143	768	1397	2130	34	1	2165
BL	594	83	49	266	460	722	4	0	726
BS	564	45	33	239	403	641	0	1	642
FR	450	92	18	220	340	199	360	1	560
GE	10	82	0	19	73	6	86	0	92
GL	43	15	6	17	47	64	0	0	64
GR	358	77	26	173	288	441	0	20	461
JU	14	33	4	16	35	4	47	0	51
LU	660	108	37	281	524	804	1	0	805
NE	14	60	2	17	59	5	71	0	76
NW	62	14	4	26	54	80	0	0	80
OW	55	3	3	18	43	61	0	0	61
SG	755	125	63	331	612	943	0	0	943
SH	110	19	11	44	96	140	0	0	140
SO	394	71	44	198	311	508	1	0	509
SZ	197	31	14	73	169	241	0	1	242
TG	320	79	25	128	296	424	0	0	424
TI	352	89	30	208	263	8	1	462	471
UR	46	8	8	17	45	62	0	0	62
VD	39	187	1	65	162	14	213	0	227
VS	159	60	18	81	156	163	74	0	237
ZG	182	47	9	88	150	236	1	1	238
ZH	2466	350	134	1055	1895	2925	9	16	2950
FL	49	6	10	19	46	65	0	0	65
Ausland	149	30	16	81	114	174	16	5	195
Total	10715	2205	781	4894	8807	12274	919	508	13701

Tabelle 53: Gesamtzahl der Teilnehmer am EMS 1998 bis 2008. Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

¹Aufgrund eines Beschlusses der Schweizerischen Universitätskonferenz von 2007 werden keine Testergebnisse mehr nach Kantonen aufgeschlüsselt veröffentlicht, weil dies vereinzelt für kantonale Bildungsvergleiche benutzt worden ist, obwohl die Repräsentativität für diesen Zweck nicht gegeben ist und darauf ausdrücklich hingewiesen wurde.

9 Beispielaufgaben für die Untertests

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

9.1 Untertest: Quantitative und formale Probleme

Mit Hilfe dieses Untertests wird die Fähigkeit überprüft, im Rahmen medizinischer und naturwissenschaftlicher Fragestellungen mit Zahlen, Grössen, Einheiten und Formeln richtig umzugehen. Diese Anforderung dürfte für mehrere Fächer des Grundlagenstudiums der Medizin bedeutsam sein.

Zum Beispiel:

Eine Broteinheit (BE) ist definiert als diejenige Nahrungsmenge in Gramm, die 12 Gramm Kohlenhydrate enthält. Bei der Verbrennung von 1 g Kohlenhydraten im Organismus werden 16 Kilojoule (kJ) an Energie frei. Ein Patient, der auf Diät gesetzt ist, soll pro Tag 4800 kJ zu sich nehmen, ein Fünftel davon in Kohlehydraten.

Wie viele BE sind dies täglich?

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

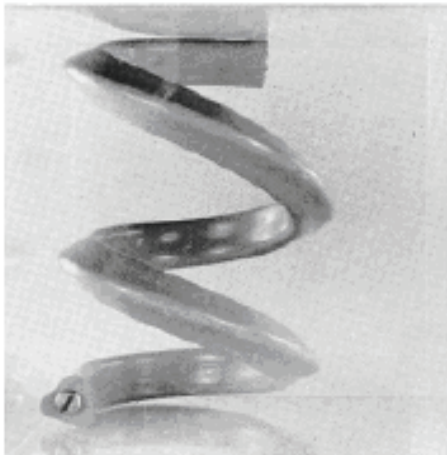
Bei solchen Fragen werden die Kenntnisse der Mittelstufen-Mathematik, nicht jedoch Lerninhalte vorausgesetzt. Der Patient soll ein Fünftel von 4800 kJ in Kohlehydraten zu sich nehmen, das sind also 960 kJ. Dividiert man diese Zahl durch 16, so erhält man die Anzahl g Kohlehydrate, nämlich 60 g, die es 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.

9.2 Untertest: Schlauchfiguren

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

Jede Aufgabe besteht aus zwei Abbildungen eines durchsichtigen Würfels, in dem sich ein, zwei oder drei Kabel befinden. Die erste Abbildung (links) zeigt stets die Vorderansicht des Würfels; auf dem rechten Bild daneben, in welchem derselbe Würfel noch einmal abgebildet ist, soll die Testteilnehmerin oder der Testteilnehmer herausfinden, ob die Abbildung die Ansicht von rechts (r), links (l), unten (u), oben (o) oder von hinten (h) zeigt.

9)



Hier sehen Sie den Würfel von vorne!



Hier sehen Sie den Würfel von ...?
(hinten!)

- (A): r
- (B): l
- (C): u
- (D): o
- (E): h

9.3 Untertest: Textverständnis

Mit Hilfe dieses Untertests wird die Fähigkeit geprüft, umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten. Die Texte sind inhaltlich und grammatikalisch anspruchsvoll – sie können unter Nutzung von Notizen und Unterstreichungen erarbeitet werden. Die Abfrage erfolgt wiederum über die Auswahl einer richtigen oder falschen Aussage aus fünf vorgegebenen Aussagen. Diese Texte waren vor allem beim Übersetzen anspruchsvoll – zur Schwierigkeit gehören nicht nur die Inhalte, sondern auch die Satzstruktur.

Ein Beispiel:

Zu den Aufgaben der Schilddrüse gehören Bildung, Speicherung und Freisetzung der jodhaltigen Hormone Trijodthyronin (T_3) und Thyroxin (T_4). In der Schilddrüse befinden sich zahlreiche Hohlräume, Follikel genannt, deren Wände von einer Schicht 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 Jodatom eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schliesslich T_4 , während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T_3 hervorgeht. T_3 und T_4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt.

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

Konzentration stimuliert sie. Bei den an der Steuerung der Schilddrüsenhormon-Sekretion beteiligten Arealen von Hirnanhangsdrüse und Hypothalamus können krankheitsbedingte Störungen auftreten, die zu einer Über- oder Unterfunktion der Schilddrüse führen.

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

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

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

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

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

9.4 Untertest: Planen und Organisieren

Der Untertest „Planen und Organisieren“ ist seit 1995 Bestandteil des EMS und prüft Fähigkeiten, die für eine effiziente Selbstorganisation im Studium wichtig sind. Er umfasst verschiedene von einander unabhängige Szenarien, auf die sich jeweils mehrere Aufgaben beziehen. Auch bei diesem Untertest wird kein Wissen vorausgesetzt. Es werden Aufgabenstellungen in Form so genannter Szenarien vorgegeben. Zu diesen Szenarien werden dann genau wie bei anderen Untertests einige Fragen gestellt. Bitte beachten Sie, dass im „richtigen“ EMS sehr vielfältige Szenarien zur Anwendung kommen, die auch ganz anderen Schemen folgen können. Das nachfolgende Szenario soll das Prinzip verdeutlichen.

Szenario: Semesterarbeit

Sie haben die Aufgabe, im kommenden Wintersemester eine Semesterarbeit zu einem vorgegebenen Thema zu schreiben. Nach gründlichem Studium der einschlägigen Literatur sollen Sie in Ihrer Arbeit zunächst einen Überblick über die wichtigsten vorliegenden Ergebnisse zum Thema geben und dann einen eigenen Lösungsansatz entwickeln.

Die folgenden Termine liegen bereits fest:

- 11.10., 17.00 Uhr: Sie erhalten das Thema Ihrer Semesterarbeit, Literaturhinweise und weitere Instruktionen.
- 17.10.: Beginn der Vorlesungszeit
- 19.12. bis 1.1.: Weihnachtsferien (vorlesungsfreie Zeit)
- 11.2.: Ende der Vorlesungszeit
- 7.3., 8.00 Uhr: Abgabe der Semesterarbeit

- Während der Vorlesungswochen: Jede Woche von Montag bis Donnerstag Lehrveranstaltungen; an diesen Tagen können Sie sich nicht der Semesterarbeit widmen.
- In der ersten Hälfte der Weihnachtsferien (16. bis 24.12.): Urlaubsreise
- Zwei volle Wochen im unmittelbaren Anschluss an die Vorlesungszeit: Prüfungsvorbereitung (1 Woche) und Teilnahme an mündlichen Prüfungen (1 Woche)

Die folgenden Arbeitsschritte bzw. -phasen müssen Sie einplanen (sie sind hier nicht in chronologischer Abfolge aufgeführt):

- Zwei Treffen mit dem Betreuer Ihrer Arbeit: Eines unmittelbar vor dem Schreiben der Semesterarbeit, eines vor der abschliessenden Überarbeitung. Dauer: Je ein voller Arbeitstag. (Der Betreuer ist nur dienstags und freitags verfügbar, und zwar von der dritten bis zur neunten Vorlesungswoche sowie nach dem Ende der Vorlesungszeit.)
- Literatursuche und -beschaffung (Dauer: 5 Tage)
- Erstellung des Konzepts der Semesterarbeit; die Literatur muss hier bereits verarbeitet sein (Dauer: 3 Tage)
- Korrekturlesen durch einen Freund (Überprüfung auf Tippfehler, unklare Formulierungen etc.) (Dauer: 3 Tage)
- Lesen und Zusammenfassen der Literatur (Dauer: 10 Tage)
- Schreiben der Semesterarbeit (Dauer: 18 Tage)
- Abschliessendes Überarbeiten der Semesterarbeit (Dauer: 4 Tage)
- Kopieren der Arbeit (unmittelbar vor der Abgabe) (Dauer: 1 Tag)
- Ausführen der Korrekturvorschläge Ihres Freundes, bevor Sie die Arbeit mit dem Betreuer durchsprechen (Dauer: 1 Tag)
- Reserve für Unvorhergesehenes (Dauer: 1 Tag)

Eine Arbeitsphase muss stets beendet sein, bevor die nächste beginnt (Ausnahme: Korrekturlesen).

Ferner müssen Sie einplanen:

- An den Wochenenden arbeiten Sie – wie auch Ihr Freund – nur samstags (im Urlaub natürlich gar nicht).
- Am zweiten Weihnachtstag arbeiten Sie nicht.

Welche der folgenden Aussagen über die Semesterarbeit trifft bzw. treffen zu?

- Noch vor den Weihnachtsferien können Sie mit dem Schreiben der Semesterarbeit beginnen.
- Insgesamt werden Sie sich in 18 Wochen jeweils mindestens einen Tag lang mit der Semesterarbeit beschäftigen.

- Nur Aussage I trifft zu.
- Nur Aussage II trifft zu.
- Beide Aussagen treffen zu.
- Keine der beiden Aussagen trifft zu.

Welche der folgenden Aussagen über die Einbeziehung Ihres Freundes trifft bzw. treffen zu?

- I. Der günstigste Zeitraum, in dem Ihr Freund Ihre Arbeit Korrektur lesen kann, ist die erste Hälfte der letzten Vorlesungswoche.
- II. Angenommen, Ihr Freund teilt Ihnen mit, er habe leider nur in der Zeit vom 25. bis zum 27. Februar Gelegenheit zum Korrekturlesen; dann nützt Ihnen sein Angebot für den ursprünglich geplanten Zweck dieses Korrekturvorgangs nicht mehr.

- (A) Nur Aussage I trifft zu.
- (B) Nur Aussage II trifft zu.
- (C) Beide Aussagen treffen zu.
- (D) Keine der beiden Aussagen trifft zu.

9.5 Untertest: Konzentriertes und sorgfältiges Arbeiten

Nachfolgend eine Beispielinstruktion aus der Test-Info:

Mit diesem Test soll Ihre Fähigkeit, rasch, sorgfältig und konzentriert zu arbeiten erfasst werden. Sie sehen nachfolgend ein Blatt mit 40 Zeilen, die aus je 40 Buchstaben u und m gebildet werden.

**Ihre Aufgabe ist es, zeilenweise jedes u zu markieren,
VOR dem in der Zeile unmittelbar ein m steht:**

m ✗

Sie dürfen kein u markieren, vor dem kein m steht. Sie dürfen natürlich auch kein m markieren. Beides wären Fehler. Nachfolgend sehen Sie ein richtig bearbeitetes Beispiel:

m ✗ m ✗ m ✗ u u m m ✗ m ✗ u u u m ✗ m m

Es werden Zeichen vorgegeben und bestimmte Zeichen sind zu markieren. Dies kann ein Merkmal eines einzelnen Zeichens sein (ein b mit 2 Strichen) oder eine Zeichenfolge (wenn ein p auf ein q folgt). Bei diesem Test hatte sich gezeigt, dass exzessives Üben zu verbesserten Leistungen führt. Da ein „Fleisstest“ nicht intendiert ist, werden seit 2004 die Zeichen und die Regel vor dem Test nicht mehr bekannt gegeben. Übungseffekte haben dadurch nachweislich einen geringen Einfluss.

9.6 Untertest: Medizinisch-naturwissenschaftliches Grundverständnis

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

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

Bei irreversiblen Hirnrindenverletzungen im Bereich der 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.

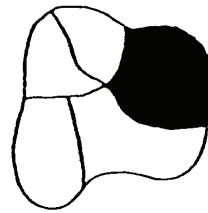
Nach dieser Analyse des Textes ist es offensichtlich, dass die Antwort (C) richtig ist.

9.7 Untertest: Figuren lernen

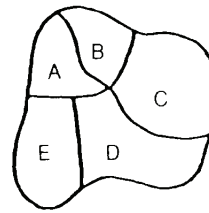
Für beide nachfolgenden Gedächtnistests wird nach der Mittagspause das Material zum Einprägen ausgeteilt. Vor der Abfrage des Gelernten wird der Untertest „Textverständnis“ bearbeitet, damit liegt die Zeit des Behaltens der gelernten Inhalte über einer Stunde. Gedächtnisleistungen sind wichtige Voraussetzungen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

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

Gezeigte Figur zum Einprägen



Gezeigte Figur beim Abfragen



Die Testperson hat vier Minuten Zeit, um sich 20 solcher Figuren einschliesslich der Lage der schwarzen Flächen einzuprägen. Nach ca. einer Stunde muss sie angeben können, welcher Teil der Abbildung geschwärzt war, und dies direkt auf dem Antwortbogen eintragen. Die Lösung ist natürlich C.

9.8 Untertest: Fakten lernen

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

Lemke, 30 Jahre, Dachdecker, ledig, Schädelbasisbruch

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

Der Patient mit dem Schädelbasisbruch ist von Beruf...

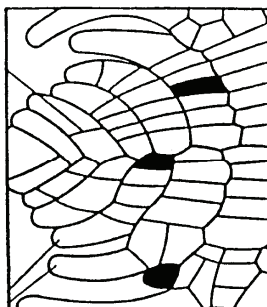
- (A) Installateur
- (B) Lehrer
- (C) Dachdecker
- (D) Handelsvertreter
- (E) Physiker

9.9 Untertest: Muster zuordnen

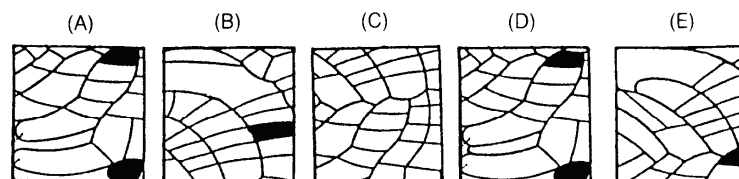
In diesem Untertest wird die Fähigkeit geprüft, Ausschnitte in einem komplexen Bild wieder zu erkennen. Dazu werden pro Aufgabe ein Muster und je fünf Musterausschnitte (A) bis (E) vorgegeben. Die Testteilnehmerin oder der Testteilnehmer soll herausfinden, welcher dieser fünf Musterausschnitte an irgendeiner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

Ein Beispiel dazu:

Muster



Musterausschnitte



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

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

9.10 Untertest: Diagramme und Tabellen

Mit dieser Aufgabengruppe wird die Fähigkeit geprüft, Diagramme und Tabellen richtig zu analysieren und zu interpretieren. In dieser Form werden während des Studiums zahlreiche Zusammenhänge vermittelt. Eine Aufgabe dazu:

Die folgende Tabelle beschreibt die Zusammensetzung und den Energiegehalt von vier verschiedenen Milcharten. Unter Energiegehalt der Milch verstehen wir dabei die Energiemenge, gemessen in Kilojoule (kJ), welche 100 Gramm (g) Milch dem Organismus ihres Konsumenten liefern können.

Milchart	Eiweiss	Fett	Milchzucker	Salze	Energiegehalt
menschliche Muttermilch	1,2 g	4,0 g	7,0 g	0,25 g	294 kJ
Vollmilch	3,5 g	3,5 g	4,5 g	0,75 g	273 kJ
Magermilch	3,3 g	0,5 g	4,5 g	0,75 g	160 kJ
Buttermilch	3,0 g	0,5 g	3,0 g	0,55 g	110 kJ

Welche Aussage lässt sich aus den gegebenen Informationen nicht ableiten?

- (A) Menschliche Muttermilch enthält mehr als doppelt soviel Milchzucker wie Buttermilch.
- (B) Vollmilch enthält im Vergleich zur menschlichen Muttermilch etwa die dreifache Menge an Salzen und Eiweiss.
- (C) Zur Aufnahme der gleichen Energiemenge muss ein Säugling fast dreimal soviel Buttermilch wie Muttermilch trinken.
- (D) Der Unterschied zwischen Magermilch und Vollmilch ist bei der Mehrzahl der aufgeführten Merkmale geringer als der Unterschied zwischen Magermilch und Buttermilch.
- (E) Der Eiweissgehalt der Milch ist für den Energiegehalt von entscheidender Bedeutung.

Wie bei den Untertests „Medizinisch-naturwissenschaftliches Grundverständnis“ und „Textverständnis“ sind auch hier zur Lösung dieser Aufgabe keine speziellen naturwissenschaftlichen, medizinischen oder statistischen Kenntnisse erforderlich. Die richtige Lösung lässt sich allein aus der jeweils graphisch oder tabellarisch dargebotenen Information und dem zugehörigen Aufgabentext ableiten. Aus den angegebenen Werten ist kein systematischer Zusammenhang zwischen Eiweiss- und Energiegehalt ableitbar, so dass die Aussage (E) nicht abgeleitet werden kann.

Diese Beispielaufgaben aus den zehn Untertests zeigen, dass es hier um Problemstellungen geht, die auch aus einem Lehrbuch des Grundstudiums Medizin stammen könnten. In den Aufgabenstellungen sind alle Informationen enthalten, die man zum Lösen benötigt. Das Problem ist zunächst zu erkennen, die Information genau zu analysieren und die richtige Lösung zu finden.

10 Literatur

- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In Holland, P.W., Wainer, H. (Eds.), Differential Item Functioning. Hillsdale, NJ: Erlbaum.
- Beller, M., Gafni, N. (1995). Translated Versions of Israel's interuniversity Psychometric Entrance Test (PET). In T. Oakland & R.K. Hambleton (Eds.), International Perspectives of Academic Assessment, S.207-218. Boston: Kluwer.
- Beller, M. (1996). Translating, equating and validating Scholastic Aptitude Tests: The Israeli Case. In Hänsgen, K.-D., Hofer, R. & Ruefli, D. (Hrsg). Eignungsdiagnostik und Medizinstudium, (S. 14-29), Bericht 2. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.
- Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (Second Edition). Hillsdale, NJ: Erlbaum.
- Cook, L.L. (1998). Can Scores Obtained on Test Given in Different Languages to Examinees of Different Cultures be Equally Valid? ICAP: San Francisco.
- Deidesheimer Kreis (1997). Hochschulzulassung und Studieneignungstests: studienfeldbezogene Verfahren zur Feststellung der Eignung für Numerus clausus und andere Studiengänge. Göttingen, Zürich: Vandenhoeck und Ruprecht.
- Ebach, J., Trost, G. (1997). Admission to Medical Schools in Europe. Lengerich: Pabst.
- Hänsgen, K.-D (2007): Numerus clausus in der Medizin – werden die Richtigen ausgewählt für Studium und Beruf? Schweizerische Ärztezeitung | Bulletin des médecins suisses | Bollettino dei medici svizzeri | 2007;88: 46
- Hänsgen, K.-D. (2007): Wie bilden wir die richtigen Studierenden zu Ärzten aus? Neue Zürcher Zeitung, 18. Oktober 2007
- Hänsgen K.-D., Spicher B. (2002). Numerus clausus: Finden wir mit dem «Eignungstest für das Medizinstudium» die Geeigneten? Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri - 2002; 83(31):1653-1660.
<http://www.saez.ch/pdf/2002/2002-31/2002-31-842.PDF>
- Hänsgen K.-D., Spicher B. (2002). Numerus clausus: Le « test d'aptitudes pour les études de médecine » (AMS) permet-il de trouver les personnes les plus aptes? Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri - 2002; 83 (47) 2562 – 2569. <http://www.saez.ch/pdf/2002/2002-47/2002-47-1144.PDF>
- Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 1: Erfahrungen mit dem EMS als Zulassungskriterium. Schweizerische Ärztezeitung Heft 12, S. 666 – 672.
- Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 2: EMS und Chancengleichheit. Schweizerische Ärztezeitung Heft 13, S. 723-730.

- Hängsen, 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.
- Hängsen, K.-D., Hofer, R. & Ruefli, D. (1995b). Der Eignungstest für das Medizinstudium in der Schweiz. Schweizerische Ärztezeitung, 37, S. 1476 - 1496.
- Longford, Holland & Thayer, (1993). Stability of the MH D-DIF Statistics Across Populations. In P. W. Holland & H. Wainer (Ed.) Differential Item Functioning, S. 171 – 196. Lawrence Erlbaum Associates: New Jersey.
- Oswald, U. (1999). Der Eignungstest 1998 für das Medizinstudium. Schweizerische Ärztezeitung 80, S. 1313 – 1317.
- Trost, G. (Hrsg.) (1994). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1995). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (19. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1996). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (20. Arbeitsbericht). Bonn: ITB.
- Trost, G. (Hrsg.) (1997). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (21. Arbeitsbericht). Bonn: ITB.
- Trost, G., Blum, F., Fay, E., Klieme, E., Maichle, U., Meyer, M. & Nauels, H.-U. (1998). Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse. Bonn: ITB.
- ZVS Press: Pressedienst der Zentralstelle für die Vergabe von Studienplätzen (1998). Deutlicher Anstieg der Bewerbungen für das Medizin-Studium. Dortmund: ZVS.

10.1 Originaltest zur Information und Vorbereitung

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

10.2 Frühere Berichte des ZTD

Bericht 13 (2007)
EMS Eignungstest für das Medizinstudium in der Schweiz 2007 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français (mit Ergebnissen der Befragung der zukünftigen Tätigkeit) K.-D. Hänsgen und B. Spicher</i>
Bericht 12 (2006)
EMS Eignungstest für das Medizinstudium in der Schweiz 2006 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher
Bericht 11 (2005)
EMS Eignungstest für das Medizinstudium in der Schweiz 2005 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> <i>Inkl. einer Befragung über die Studienmotivation; K.-D. Hänsgen und B. Spicher</i>
Bericht 10 (2004)
EMS Eignungstest für das Medizinstudium in der Schweiz 2004 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher
Bericht 9 (2003)
EMS Eignungstest für das Medizinstudium in der Schweiz 2003 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher
Bericht 8 (2002)
EMS Eignungstest für das Medizinstudium in der Schweiz 2002 <i>Bericht über die Durchführung und Ergebnisse, Zusammenfassender Vergleich 1998 bis 2002 - avec un résumé en français; K.-D. Hänsgen und B. Spicher</i>
Bericht 7 (2001)
EMS Eignungstest für das Medizinstudium in der Schweiz 2001 <i>Bericht über die Durchführung und Ergebnisse, Vorhersage von Prüfungserfolg Zusammenfassender Vergleich 1998 bis 2001 - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher
Bericht 6 (2000)
EMS Eignungstest für das Medizinstudium in der Schweiz 2000 <i>Bericht über die Durchführung und Ergebnisse, Evaluation der Vorhersage von Prüfungserfolg. Zusammenfassender Vergleich 1998 bis 2000 - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher
Bericht 5 (1999)
EMS Eignungstest für das Medizinstudium in der Schweiz 1999 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher

Bericht 4 (1998)	
EMS Eignungstest für das Medizinstudium in der Schweiz 1998 <i>Bericht über die Durchführung und Ergebnisse - avec un résumé en français</i> K.-D. Hänsgen und B. Spicher	
Bericht 3 (1997) - Rapport 3 (1997)	
Eignungsprüfung für das Medizinstudium - Kriterien und Testverfahren <i>Bericht über das Internationale Symposium in Bern am 8. November 1996</i> Hrsg. von K.-D. Hänsgen und N. Ischi	L'examen d'aptitude aux études de médecine en Suisse - Critères et procédés d'application du test <i>Rapport sur le Symposium international à Berne</i> Editeurs: K.-D. Hänsgen; N. Ischi
Günter Trost Eignungskriterien bei der Zulassung zum Medizinstudium in Europa: Ergebnisse einer Erhebung Ingemar Wedman & Widar Henriksson The Swedish Scholastic Aptitude Test. Research and main findings John L. Hackett The Medical College Admission Test (MCAT) - its use in U.S. and Canada and some results of validation Piet J. Janssen Admission to the study in medicine in Belgium: two 'different' solutions to the 'same' problem; reflections of a Flemish school psychologist	Klaus-Dieter Hänsgen Der Eignungstest für das Medizinstudium in der Schweiz als Instrument für die Studienzulassung Rainer Hofer & Klaus-Dieter Hänsgen Die Trainierbarkeit von Testleistungen im Zusammenhang mit einem Eignungstest für das Medizinstudium Urs Schallberger Anforderungen an das Zulassungsverfahren für das Medizinstudium in der Schweiz: Leitlinien
Bericht 2 (1996) - Rapport 2 (1996)	
Eignungsdiagnostik und Medizinstudium <i>Symposiumsbericht</i> Hrsg. von K.-D. Hänsgen, R. Hofer und D. Ruefli	Diagnostic d'aptitudes et études de médecine <i>Rapport d'un symposium</i> Edité par K.-D. Hänsgen, R. Hofer et D. Ruefli
Klaus-D. Hänsgen Vorwort: Eignungstests und Medizinstudium Christina Stage Experiences with the Swedish Scholastic Aptitude Test Michal Beller Translating, equating and validating Scholastic Aptitude Tests: The Israeli Case Eckhard Klieme Erfolgsprognose in medizinischen Studiengängen - Zur Validität des Tests für medizinische Studiengänge und anderer Auswahlinstrumente	Günter Trost Testergebnisse versus Schulnoten als Auswahlkriterien: Paternoster-Effekt, Filter-Effekt, Kosten-Nutzen-Effekte und Auswirkungen auf die Fairness der Zulassung Urs Schallberger Nutzen, Fairness, Validität und Akzeptanz von Selektionsverfahren Rainer Hofer, Daniel Ruefli & Klaus-D. Hänsgen Der „Test des Tests“ - Ergebnisse eines Probeaufs des Eignungstests in der Schweiz in deutscher und französischer Sprache
Bericht 1 (1995) - Rapport 1 (1995)	
Der Eignungstest für das Medizinstudium in der Schweiz - ein Probelauf Rainer Hofer, Daniel Ruefli & Klaus-D. Hänsgen	Le test d'aptitudes pour les études de médecine en Suisse - Epreuve d'essai Rainer Hofer, Daniel Ruefli & Klaus-D. Hänsgen

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