

EMS 2014 – Bericht 21



Eignungstest für das Medizinstudium in der Schweiz

Bericht über Durchführung und Ergebnisse
avec un résumé en français

K.-D. Hänsgen und B. Spicher



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

Hänsgen, Klaus-Dieter; Spicher, Benjamin (2014)
EMS Eignungstest für das Medizinstudium 2014
Berichte des Zentrums für Testentwicklung, Nr. 21,
im Auftrag der Schweizerischen Universitätskonferenz (SUK)

Unter Mitarbeit von P. Bergem, M. Bernasconi,
T. Cruchaud, Y. de Zordo und M. Strazzeri

In Zusammenarbeit mit der Rektorenkonferenz der Schweizer
Universitäten (CRUS), die das Anmelde- und
Zuteilungsverfahren der Plätze zum Medizinstudium
durchführt und für diesen Bericht statistische Angaben zur
Verfügung stellt:

J. Bregy, Chr. Winzenried

Redaktion: Tanja Cruchaud



Die Mitglieder des Beirates EMS der Schweizerischen
Universitätskonferenz 2014:

V. Clerc (Vorsitzende, SUK Bern), J. Bregy (CRUS Bern), Prof. Dr. F. Eberle (Univ. Zürich);
Dr. R. Hofer (Univ. Bern); Prof. Dr. M. Kaufmann (Univ. Basel); Prof. Dr. P. Klumb (Univ. Freiburg);
Prof. Dr. T. Lutz (Univ. Zürich); Prof. Dr. J.-P. Montani (Univ. Freiburg); Prof. Dr. K. Opwis (Univ. Basel);
Prof. Dr. S. Rohr (Univ. Bern); Prof. Dr. G. Trost (ITB GmbH Bonn Deutschland);
Dr. E. Werlen (CRUS Bern); Prof. Dr. D. Wolfer (Univ. Zürich);
Dr. A. Zimmerhofer (ITB GmbH Bonn Deutschland); Prof. Dr. N. Zitzmann (Univ. Basel)

Avec un résumé en français

Mit Zusatzbericht EMS 2014 – Nichtwertung von 14 Aufgaben

Inhaltsverzeichnis

1	Zusammenfassung	5
2	Résumé	7
2.1	Statistique	7
2.2	Organisation et déroulement du test en Suisse	7
2.3	Qualité du test	7
2.4	Egalité de traitement des groupes linguistiques	8
2.5	Différences selon le sexe et la classe d'âge	8
2.6	En conclusion	8
3	Zur aktuellen Situation	9
3.1	Erhöhung der Ausbildungskapazitäten und Kosten	9
3.2	Ärztmangel oder Strukturproblem?	10
3.3	Weiterentwicklung des Zulassungsverfahrens	10
4	Numerus clausus (NC) und Medizinstudium	13
5	Anmeldung zum Medizinstudium und Test 2014	16
5.1	Anmeldestatistik Humanmedizin	16
5.2	Anmeldestatistik Veterinärmedizin	18
5.3	Anmeldestatistik Zahnmedizin	19
5.4	Grösse der Testlokale	25
5.5	Testorte und Wunschuniversitäten	26
5.6	Wunschuniversität und Testort nach Wohnkanton	27
5.7	Teilnahmen am EMS (NC) nach Kantonen seit 1998	31
5.8	Testabsolvierung nach Alter und Geschlecht	32
5.9	Übernahme des Testergebnisses aus dem Vorjahr	38
5.10	Sprachgruppen	40
5.11	Alter und Maturitätsjahr	42
6	Beschreibung des verwendeten Eignungstests	43
6.1	Aufbau des Tests	43
6.2	Berechnung der Werte 2014	44
6.3	Mittlerer Rangplatz der Aufgabengruppen	46
7	Testanwendung in der Schweiz 2014	49
7.1	Verteilungsprüfung	49
7.2	Äquivalenz der Sprachversionen	50
7.3	Vergleichbarkeit der Testlokale	66
7.4	Vergleich für die Geschlechter	68
7.5	Vergleiche für Altersgruppen	70
7.6	Vergleiche nach Wunschuniversitäten	72
8	Ergebnisse zur Testgüte	74
8.1	Zuverlässigkeit	74
8.2	Binnenstruktur	76
8.3	Aufgaben-Trennschärfen	80
8.4	Konzentriertes und sorgfältiges Arbeiten	83
9	Beispielaufgaben für die Aufgabengruppen	87
10	Literatur	96
	Anhang: Zusatzbericht EMS 2014 - Nichtwertung von 14 Aufgaben	96

1 Zusammenfassung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2014 für das Medizinstudium in der Schweiz (EMS) und die Zulassung zum Studium der Medizin ab Herbstsemester 2014. Für bestimmte Fragestellungen werden Vergleichsdaten der Testanwendungen 1998 bis 2013 herangezogen.

1.1 Statistik

Ein Numerus clausus (NC) ist 2014 wiederum für Humanmedizin (seit 1998), Veterinärmedizin (seit 1999) und Zahnmedizin (seit 2004) notwendig – dort für Personen, die sich an den Universitäten Basel, Bern, Freiburg oder Zürich angemeldet haben. An der Universität Zürich gilt der NC auch für Chiropraktik (seit 2008).

Die Kapazitäten der Universitäten mit NC für **Humanmedizin** (inklusive Chiropraktik) wurden mit den Anmeldungen im **Februar 2014** zu **417%** (Vorjahr 459%) ausgelastet. Gegenüber dem Vorjahr stieg hier die Ausbildungskapazität von 713 auf 793, die Zahl der Anmeldungen von 3270 auf 3310. Gegenüber 2006 hat die Kapazität in den Universitäten mit NC nunmehr bereits **um 45%** zugenommen (von 546 auf 793). In **Veterinärmedizin** sind bezogen auf die Anmeldungen **325%** (Vorjahr 319%), in **Zahnmedizin 262%** (Vorjahr 285%) Auslastung der Kapazitäten vorhanden.

Insgesamt **3329 Personen** (Vorjahr 3196) haben sich für diese Universitäten bis zur gesetzten Frist im **Mai 2014** bei der CRUS zum EMS-Test angemeldet. **85 Personen** (Vorjahr 123) übernehmen ihr Testergebnis aus dem Jahr 2013.

Mit gültigem Ergebnis haben den EMS **3173 Personen** beendet (Vorjahr 3028). 4 Personen haben den Test vorzeitig abgebrochen (Vorjahr 7 Personen). 152 Personen haben ihre Testanmeldung zurückgezogen bzw. sind nicht zum Test erschienen (Vorjahr 161).

Während der Auswertungsphase sind Unregelmäßigkeiten entdeckt worden, die Korrekturmassnahmen erforderlich machten. 14 der insgesamt 198 Fragen waren bereits vor dem Testtag im Umlauf. Die Schweizerische Universitätskonferenz (SUK), die Rektorenkonferenz der Schweizer Universitäten (CRUS) und das Zentrum für Testentwicklung der Universität Freiburg (ZTD) haben beschlossen, diese 14 Fragen nicht in die Bewertung einzubeziehen. Damit wird gewährleistet, dass die Zulassung unter gleichen Bedingungen für alle auf der Basis der Eignung erfolgt. Diese Massnahme wurde aus Sicherheitsgründen getroffen, obwohl statistisch keine relevanten Auswirkungen der Vorkenntnisse einzelner Aufgaben auf die Testresultate der Kohorte 2014 nachgewiesen werden konnten. Die Kandidatinnen und Kandidaten wurden mit Brief vom 5. August 2014 über die Korrekturmassnahme informiert.

1.2 Organisation und Ablauf in der Schweiz

Der Test fand am 4.7.2014 gleichzeitig **an 9 Testorten** und in 30 Testlokalen in drei Sprachen (Deutsch: Aarau-Suhr, Basel, Bern, Chur, Luzern, St. Gallen, Zürich; Französisch: Freiburg; Italienisch: Bellinzona) statt. Die Wahl des Testortes war unabhängig vom zukünftigen Studienort möglich.

1.3 Testgüte

Die Zuverlässigkeitskennwerte (Reliabilitäten) erreichen auch 2014 das gewohnt hohe Niveau. Mit 0.92 liegt die Reliabilität gemessen nach der Testhalbierungsmethode auf dem für vergleichbare Tests sehr hohen Niveau der Vorjahre. Auch die Konsistenz spricht mit 0.82 für zuverlässige Ergebnisse.

Die Trennschärfen (Korrelation der einzelnen Aufgabe mit der zugehörigen Gruppe) zeigen durchwegs positive Werte.

Auch die Faktorenstruktur der einzelnen Aufgabengruppen zeigt die aus den vergangenen Jahren bekannt hohe Konstanz. Dies ist ein Beleg dafür, dass die Übernahme der Testergebnisse ins Folgejahr gerechtfertigt ist.

1.4 Gleichbehandlung der Sprachgruppen

Sprachunterschiede wurden wiederum mit dem DIF-Verfahren geprüft und entsprechende mögliche testbedingte Unterschiede korrigiert. Bei 5 sprachabhängigen Aufgabengruppen in der französisch- und nur 2 Aufgabengruppen in der italienischsprachigen Adaptation traten signifikante Unterschiede im Vergleich zur deutschsprachigen Version auf, was eine Voraussetzung für die Anwendung des DIF-Verfahrens ist. Die Zahl der zu korrigierenden Aufgaben war wiederum sehr gering. Dies spricht für äquivalente Testversionen, bei welchen die vorhandenen Unterschiede nicht testbedingt sind. Die Kontrolle zeigt, dass wie in nahezu allen Vorjahren die Leistungsunterschiede in den sprachabhängigen Aufgabengruppen erneut weniger gross sind (mit und ohne Korrektur) als in den nicht sprachabhängigen Tests.

1.5 Differenzierung nach Geschlecht und Alter

Unterschiede zwischen den Geschlechtern zeigen sich beim Test-Prozentrang über alle Sprachgruppen und Disziplinen in einer Differenz von 1.8 zugunsten der Männer. In Humanmedizin beträgt der Unterschied für die deutschsprachige Gruppe 1.0. Diese gehört zu den bisher geringsten Differenzen. Bei der Bewertung des Unterschiedes ist auch zu beachten, dass sich 61% Frauen und 39% Männer für Humanmedizin bewerben – dieses Verhältnis ist weder repräsentativ für die entsprechende Kohorte der Gesamtbevölkerung noch der Maturanden und lässt keine Gleichheit der Mittelwerte erwarten.

Auch in diesem Jahr zeigen sich die bekannten Unterschiede für die Gruppe der Testteilnehmenden mit den ältesten Jahrgängen: Wenn sie ihre Maturitätsprüfung spät abgelegt haben, zeigen sich tendenziell schlechtere Ergebnisse als bei früher Maturitätsprüfung. Jüngere Gruppen erreichen in diesem Jahr erneut 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 Fazit

Der EMS 2014 wurde wiederum erfolgreich absolviert, die Ergebnisse der Evaluation sprechen bezüglich der Gütekriterien für eine zu den Vorjahren vergleichbare Qualität des Tests.

2 Résumé

Le présent rapport informe sur les résultats concernant le test d'aptitudes aux études de médecine en Suisse (AMS) et les admissions aux études en question au semestre d'automne 2014. Pour certaines problématiques, des données comparatives des éditions antérieures du test (1998-2013) y figurent également.

2.1 Statistique

Cette année, un **numerus clausus (NC)** est à nouveau 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, de Berne, de Fribourg ou de Zurich. A l'Université de Zurich, le NC s'applique également à la **chiropractie** (depuis 2008).

En **février 2014**, les inscriptions reçues pour la **médecine humaine** (y compris la chiropractie) correspondaient à un dépassement des capacités de ces universités de **417 %** (459 % l'année passée). Par rapport à 2013, les capacités de formation dans ce domaine sont passées de 713 à 793 et le nombre d'inscriptions de 3270 à 3310. Depuis 2006, les capacités des universités avec NC ont progressé de **45 %** (passant de 546 à 793). Cette année, le dépassement des capacités s'élevait par ailleurs à **325 %** (319 % l'année précédente) en **médecine vétérinaire** et à **262 %** (285 % l'année passée) en **médecine dentaire**.

3329 personnes au total (2013 : 3196) se sont inscrites auprès de la CRUS au test AMS jusqu'au délai fixé (mai 2014) ; **85 personnes** ont repris leur résultat de l'année précédente (2013 : 123).

3173 personnes ont terminé l'AMS avec des résultats valables (2013 : 3028). 4 personnes ont interrompu le test avant la fin de la session (2013 : 7). 152 personnes ont retiré leur inscription ou ne se sont pas présentées au test (2013 : 161).

Des irrégularités constatées durant la phase d'évaluation ont rendu nécessaires la prise de mesures correctives. En effet, il s'est avéré que 14 des 198 questions que comprend le test étaient déjà en circulation avant le jour du test. La Conférence universitaire suisse (CUS), la Conférence des recteurs des universités suisses (CRUS) et le Centre pour le développement de tests et le diagnostic (CTD) de l'Université de Fribourg ont par conséquent décidé de ne pas tenir compte de ces 14 questions. Cette mesure permet d'assurer que l'admission aux études s'effectue sur la base de conditions identiques pour tous et en fonction des aptitudes. Elle a été prise pour des raisons de sécurité, même si aucune évidence statistique ne permet de démontrer un avantage lié à la connaissance des exercices diffusés. Les candidat-e-s au test ont été informé-e-s des mesures correctrices par courrier du 5 août.

2.2 Organisation et déroulement du test en Suisse

Le test s'est déroulé le 4 juillet 2014, en même temps sur **9 sites** différents, en 30 espaces et en trois langues (français : Fribourg ; allemand : Aarau-Suhr, Bâle, Berne, Coire, Lucerne, St-Gall, Zurich ; italien : Bellinzone). Il était possible de choisir le lieu du test indépendamment du futur lieu d'études.

2.3 Qualité du test

En 2014 également, les indices de fiabilité se situent au niveau habituellement élevé. Mesurée selon la méthode de partition pairs-impairs, la fiabilité se situe, avec des indices de 0.92 au niveau de

l'année précédente. Les indices de consistance du profil de test, avec des valeurs de 0.82 témoignent également de la fiabilité des résultats.

Les sélectivités (corrélation de l'exercice particulier avec le groupe d'exercices correspondant) montrent des valeurs tout à fait positives.

La structure factorielle des différents groupes d'exercices montre la bonne constance constatée déjà au cours des années passées, ce qui atteste que la reprise des résultats du test l'année suivante est justifiée.

2.4 Egalité de traitement des groupes linguistiques

Des différences de langue ont à nouveau été examinées à l'aide de la procédure appelée DIF et d'éventuelles différences liées au test ont été corrigées. Par rapport à la version allemande, des différences significatives sont apparues dans 5 groupes d'exercices à forte composante linguistique dans l'adaptation de langue française et dans 2 groupes d'exercices dans celle de langue italienne, ce qui est une condition pour l'application de la procédure DIF. Le nombre d'exercices à corriger était une fois encore très petit. Cela atteste de la forte équivalence des versions du test dans lesquelles les différences existantes ne sont pas liées au test. Le contrôle montre, comme lors de presque toutes les années précédentes, que les différences de prestations dans les différents sous-tests à forte composante linguistique sont moins grandes (avec et sans correction) que dans les sous-tests sans composante linguistique.

2.5 Différences selon le sexe et la classe d'âge

Les différences entre les sexes en ce qui concerne les résultats du test pour tous les groupes linguistiques et toutes les disciplines se manifestent par une différence de 1.8 en faveur des hommes. En médecine humaine, la différence pour le groupe de langue allemande s'élève aussi à 1.0. Il s'agit d'une des plus petites différences constatées jusqu'ici. Dans cette évaluation des différences, il importe de relever également que 61 % des candidats pour la médecine humaine sont des femmes et 39 % des hommes – cette relation n'est représentative ni de la cohorte correspondante de la population globale, ni de celle des bacheliers ; elle ne permet pas de s'attendre à une égalité des moyennes.

On trouve cette année aussi les différences connues pour le groupe des participants au test les plus âgés : ceux qui ont obtenu leur maturité plus tard ont des résultats en tendance plus faibles que ceux qui ont obtenu leur maturité plus tôt. Les groupes des plus jeunes obtiennent à nouveau cette année de meilleures performances au test. Nous rappelons les résultats des études d'évaluation dans lesquelles ce résultat était aussi lié à des différences correspondantes pour le succès aux études. Le test présente donc effectivement une différence à cet égard.

2.6 En conclusion

L'AMS 2014 a été réalisé avec succès et, concernant les critères de qualité, les résultats de l'évaluation permettent de conclure que la qualité du test de cette année est comparable à celle des années précédentes.

3 Zur aktuellen Situation

Die Diskussionen zu Medizinstudium und die „Versorgungsdebatte“, besonders zur Humanmedizin, werden weiterhin intensiv geführt. Drei Fragen stehen im Mittelpunkt der Diskussionen:

3.1 Erhöhung der Ausbildungskapazitäten und Kosten

Die Forderung nach Erhöhung der Absolventenzahlen in Medizin bedeutet, dass mehr Anfänger zugelassen werden – diese auch eine wirkliche Chance haben müssen, ihr Studium „aus eigener Kraft“ erfolgreich abzuschliessen und kein verschärfter „inneruniversitärer NC“ nach dem ersten Jahr zur Reduktion der Überschreitungen notwendig ist. Vor allem die Klinikkapazitäten und Betreuungsverhältnisse müssen für eine qualitativ hochstehende Ausbildung ausreichend sein und sind nicht beliebig ausbaubar sowie mit höheren Kosten verbunden. Wie die Zahlen zeigen, geschieht einiges:

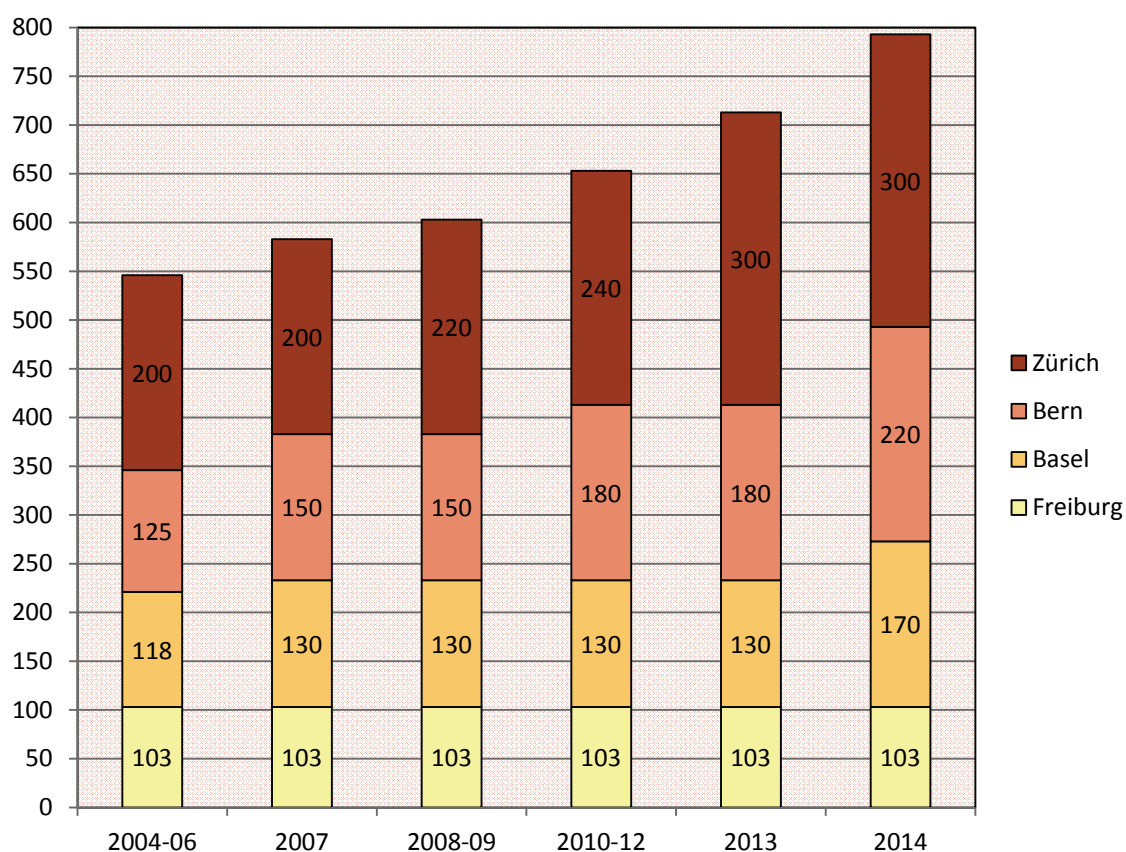


Abbildung 1: Entwicklung der Kapazitäten für Humanmedizin an Universitäten mit NC seit 2004.

2014 wurden in Basel und Bern je 40 zusätzliche Plätze geschaffen. Insgesamt stehen also 80 Plätze mehr als im Vorjahr zur Verfügung. **Gegenüber 2006 ist dies bereits eine Steigerung der Anfänger-Kapazitäten um 45%.**

In Freiburg werden jetzt die ersten 3 Jahre absolviert (BA), das 3. Jahr wurde neu aufgebaut. Vorher musste man bereits nach dem 2. Jahr an eine andere Universität wechseln. Auch das entlastet nun die anderen Universitäten bei der Planung des BA-Abschnittes.

Für die Erhöhung der Anfängerkapazitäten im Jahre 2014 in Basel um 40 Plätze sind Kosten mitgeteilt worden. Eine Erhöhung des Budgets ist schrittweise in den nächsten 7 Jahren notwendig, für 2014 stehen 700'000.- zur Verfügung, bis 2017 werden 2.8 Mio. Franken jährlich als notwendig angegeben (wobei das Masterstudium, was für diese Kohorte regulär im Herbst 2017 beginnt, eigentlich anderswo teurer eingeschätzt wird). Rechnet man einen Medizinstudienplatz bis zum Masterabschluss mit 500'000.- CHF (es gibt höhere Schätzungen bis 700'000.- CHF), wären **20 Millionen Franken** für diesen Ausbau **pro Jahrgang** als Mehrkosten notwendig – die in den nächsten 7 Jahren mit dem Studienfortschritt der 40 jeweils mehr Zugelassenen sukzessive auch zu effektiven Mehrkosten **pro Jahr** werden (wenn am Ende jährlich einer dieser Jahrgänge das Studium abschliesst).

Die Zahlen verdeutlichen, dass es einerseits nicht um unbedeutende Kosten geht – andererseits eine Erweiterung der Anfängerzahlen auch bis zum Ende durchfinanziert sein muss, wenn man die Studienqualität und -kontinuität für alle Anfänger absichern möchte.

3.2 Ärztemangel oder Strukturproblem?

Richtig ist, dass man die Ausbildungskapazitäten auch am Bedarf ausrichten muss – dabei den demografischen Wandel berücksichtigt und rechtzeitig genügend Jüngere ausbildet, die aus dem Berufsleben ausscheidende Ältere ersetzen. Einige Argumentationen dabei sollte man aber sehr genau prüfen:

- Besteht aktuell wirklich ein Ärztemangel, wenn die Schweiz 24% mehr berufstätige Ärzte pro Einwohner hat als der Durchschnitt der Industrieländer?
- Gibt es stattdessen ein „Strukturproblem“, indem Hausärzte oder Psychiater schlechter bezahlt werden als Spezialärzte und jeder Absolvent sich nach der doch aufwändigen Ausbildung fragt, ob er sich dies antun soll? Sind diese Verhältnisse in den Regionen vergleichbar oder müsste man auch hier die finanziellen Anreize ändern?
- Fallen die Ursachen, dass ausländische Ärzte gern in die Schweiz kommen, durch mehr Ausbildung von Schweizern wirklich weg? Die gute Bezahlung, die guten Arbeitsbedingungen werden bleiben. Entweder man „verelendet“ den Arztberuf oder kann Restriktionen für Ausländer einführen, die in Medizinalberufen arbeiten wollen. Ob „Schweizer zuerst“ aber angesichts der komplizierten Beziehungen mit der EU haltbar ist? Wenn das Gesundheitswesen ein „Angebotsmarkt“ bleibt und die Anbieter selber bestimmen, wie viel Medizin konsumiert wird (durch mehr Tests, mehr Operationen usw.), wäre ein Überangebot für uns alle und die Kosten kontraproduktiv.

Die Medizinversorgung ist ein komplexes „System“, wo Angebot und Nachfrage, tatsächlicher Bedarf und Ausbildungs-Kapazität, Abwanderung von Schweizern und Zustrom von Ausländern, Förderung oder Hemmung durch entsprechende finanzielle Anreize, „Selbstregulation“ und gesellschaftliche Kontrolle so ineinandergreifen, dass man nicht beliebig an einem einzigen Schraubchen drehen sollte. Vielleicht kann man dies national auf sich allein gestellt auch gar nicht mehr planen. Es braucht also eine Strategie.

3.3 Weiterentwicklung des Zulassungsverfahrens

Die richtige **Auswahl** der Personen, um das „beschränkte Gut“ Studienplätze optimal zu nutzen, ist wichtig. Wenn man am Ende auch mehr Ärzte und nicht nur mehr Studienabbrüche will, müssen so viele wie möglich das Studium abschliessen und dann auch im Arztberuf arbeiten. Solange das Medizinstudium anspruchsvoll hinsichtlich der mathematisch-naturwissenschaftlichen Fähigkeiten bleibt, müssen diese ausreichend entwickelt sein.

Kann man unter diesen Bedingungen auf eine Selektion aufgrund der **Studienerfolgsprognose** verzichten? Kann man das Fehlen dieser Fähigkeiten z.B. mit sozialen Kompetenzen ausgleichen: „Menschlich nett aber leider durchgefallen?“ Es gibt gegenläufige Forderungen seitens der Universitäten, den Naturwissenschaftsaspekt als Eignungskriterium stärker zu gewichten, weil der Wissensstand im Mittel abgenommen habe bzw. zu heterogen sei.

Auch ein verschiedentlich behaupteter Gegensatz von **Studien- und Berufserfolg** (der auch als Auswahlkriterium gefordert wird) ist nach unserer Literaturkenntnis nirgends empirisch bewiesen worden – Studienerfolg ist eine Facette des Berufserfolges, weil kognitive Fähigkeiten auch für die spätere ärztliche Berufstätigkeit, für die Anforderungen lebenslangen Lernens bedeutend sind. Hier scheint sich auch ein Klischee zu halten, dass kognitiv fähige Leute „Fachidioten“ seien – dies ist aber wirklich nur ein Klischee ohne empirische Absicherung.

Wenn heute Ärzte „gestresster“ sind, weniger Zeit für die Behandlung des einzelnen Falles haben bzw. sich aus ökonomischen Gründen nehmen, keine Hausbesuche mehr machen, die Anforderungen durch die Komplexität der vielen neuen diagnostischen und therapeutischen Möglichkeiten grösser werden und der Arzt objektiv nicht mehr der liebevolle Onkel oder die liebevolle Tante aus unserer Erinnerung sein kann – und auch nicht mehr 7 Tage pro Woche 10 Stunden arbeiten möchte, muss man die Bewältigbarkeit der beruflichen Anforderungen durch den Arzt verbessern – nicht nur auf Kosten der Freizeit. Denn auch prosoziales Verhalten kann und muss man lernen und förderliche Bedingungen dafür im Studium (und Beruf) schaffen.

Fachbezogene soziale Kompetenzen liegen zum Zeitpunkt vor Studienbeginn ebenso wenig ausgeformt vor wie das Fachwissen. Schon das Verwenden „fachnaher“ Anforderungen beim Prüfen (z.B. Pseudo-Patientengespräche) ist da eigentlich ein Kunstfehler – das richtige Verhalten soll im Studium erlernt werden. Was man prüfen kann, sind Voraussetzungen zum sozialen Lernen, wo sich Menschen auch unterscheiden. Z.B. In Israel (und auch anderswo, wo man dies mit Interviews oder Rollenspielen zu testen versucht) ist das Verfahren zweistufig: Alle müssen zuerst einen Test absolvieren und man prüft nur einen Teil der Personen differenzierter – anders kann man dies auch nicht bezahlen. Es wurde übrigens auch in der Schweiz bei Einführung des NC überlegt, ob man ergebnisabhängig zu Interviews einlädt – hohe Kosten und unklarer Nutzen liessen aber davon Abstand nehmen.

Wollte man die **Maturanote** für die Zulassung verwenden, was auch verschiedentlich überlegt wurde, müsste man den numerischen Durchschnittswert nehmen. Erstens bedeutet der numerisch gleiche Durchschnitt nicht das Gleiche. Kantonsunterschiede, insbesondere bei der Skalierung und Strenge der Notengebung wirken sich aus. Auch die unterschiedlichen Maturitätsquoten in den Kantonen sind zu berücksichtigen. Unterschiedliche Bildungswege (Mittelschule oder dritter Bildungsweg) können ebenfalls nicht immer die gleichen Massstäbe anlegen. So kann eine 6.0 in einem Kanton für einen Bildungsweg durchaus einer 5.8 (oder tiefer) anderswo entsprechen. Wenn man dies (wie z.B. in Deutschland auf der Ebene der Bundesländer) statistisch korrigieren, sprich nivellieren, wollte, scheitert dies auch an der grösseren Zahl der Kantone (mit grösseren Unterschieden) mit kleineren Personenzahlen.

Zweitens könnte bei einer (auch) subjektiven Notenvergabe durch die Lehrpersonen nicht ausgeschlossen werden, dass eine zulassungsentscheidende Rolle irgendwo doch berücksichtigt würde. Vermutlich würde der „Kantönligeist“ mit hineinspielen – wir denken an die Zeiten, wo festgestellte Kantonsunterschiede beim EMS zu teils heftigen Diskussionen hinsichtlich möglicher Benachteiligungen führten.

Verglichen mit Deutschland, wo es durch die langjährige Verwendung der Abiturnote als NC-Kriterium zu einer Häufung von „Einserdurchschnitten“ bei Medizinbewerbungen gekommen ist, streuen die

Schweizer Maturitätsdurchschnitte mehr. 2014 haben in einer freiwilligen Befragung am Ende des EMS Personen mit Wunsch Humanmedizin einen mittleren Maturitätsdurchschnitt 4.8 und eine Standardabweichung 0.42 (n=2271) angegeben. Für die zugelassenen Personen gilt ein Mittelwert 5.0 und eine Standardabweichung 0.39 (n=679). Diese Unterschiede waren auch in Österreich zu beobachten, wo es bis 2006 keinen NC gab und ab diesem Zeitpunkt fast 50% der Bewerbungen „NC-Flüchtlinge“ aus Deutschland waren, die numerisch deutlich bessere Durchschnittsnoten aufwiesen.

Es gibt zwei Voraussetzungen für die (Mit-)Verwendung von Maturitätsnoten als NC-Kriterium:

- (1) Vereinheitlichung der Anforderungen (auch die Lehrinhalte sind ja nicht identisch) sowie
- (2) Vereinheitlichung und Bewertungs-Massstäbe und Objektivierung der Bewertung selbst, etwa durch die Einführung einer Zentralmaturität. Hier könnte nur langfristig etwas an dieser Situation geändert werden, um zumindest eine Mitberücksichtigung bei der Zulassung zu erlauben.

4 Numerus clausus (NC) und Medizinstudium

Auf den ersten Blick scheint 2014 die Nachfrage nach einem Medizinstudium erstmals seit 2001 gegenüber dem Vorjahr nicht anzusteigen. Ist nun ein „Plateau“ erreicht?

Zum einen ist die Kapazitätsüberschreitung in den Universitäten mit NC immer noch so, dass gegenüber dem Stand der Februar-Anmeldungen in Humanmedizin nur jede vierte Person einen Platz erhalten kann und nach dem Test jede dritte. Zum anderen steigt im Fach Humanmedizin die Nachfrage bei den Universitäten mit NC weiter, die Abnahme der Bewerbungen für Zahnmedizin ist vor allem für die „Stagnation“ verantwortlich. Bei Zahnmedizin wurde einiges dafür getan, die Spezifik des Studiums hinsichtlich der Berufs- wie Studienanforderungen bekannter zu machen¹, einschliesslich der Verdeutlichung einer Nichteignung als „Sprungbrett“ in ein Humanmedizinstudium.

Im Fach Humanmedizin wird der nicht unbeträchtliche Kapazitätsausbau durch die gleichzeitige Zunahme des Studieninteresses nivelliert und die Quote Zulassungen/Anmeldungen stagniert in etwa. So wird zumindest der NC nicht weiter verschärft.

Bern und Zürich verzeichnen eine weitere Nachfragesteigerung für alle Medizinstudienplätze zusammen, in Freiburg (nunmehr ist dort ein dreijähriges Vollstudium zum BA möglich) bestätigt sich die stärkere Nachfrage der letzten drei Jahre. In Basel gibt es einen geringfügigen Rückgang der Anmeldungen.

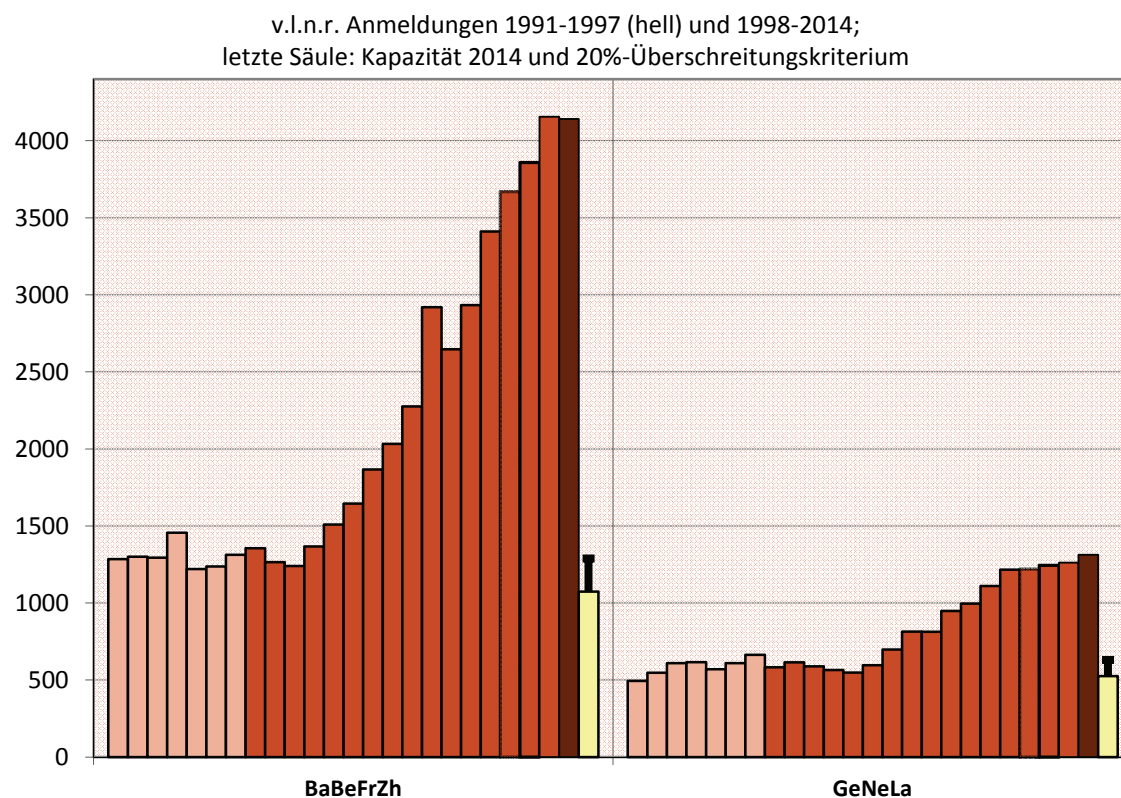


Abbildung 2: Anmeldungen zum Medizinstudium 1991 bis 2014 nach Gruppen (Basel, Bern, Freiburg, Zürich vs. Genf, Neuenburg, Lausanne). Hellere Säulen: Jahre ohne NC; dunklere Säulen: Jahre mit NC; letzte Säule: Kapazität 2014 und Markierung 20%-Überschreitungskriterium als Grenze für NC.

¹ Siehe dazu: <http://unifr.ch/self-assessment/studienberatung.html>

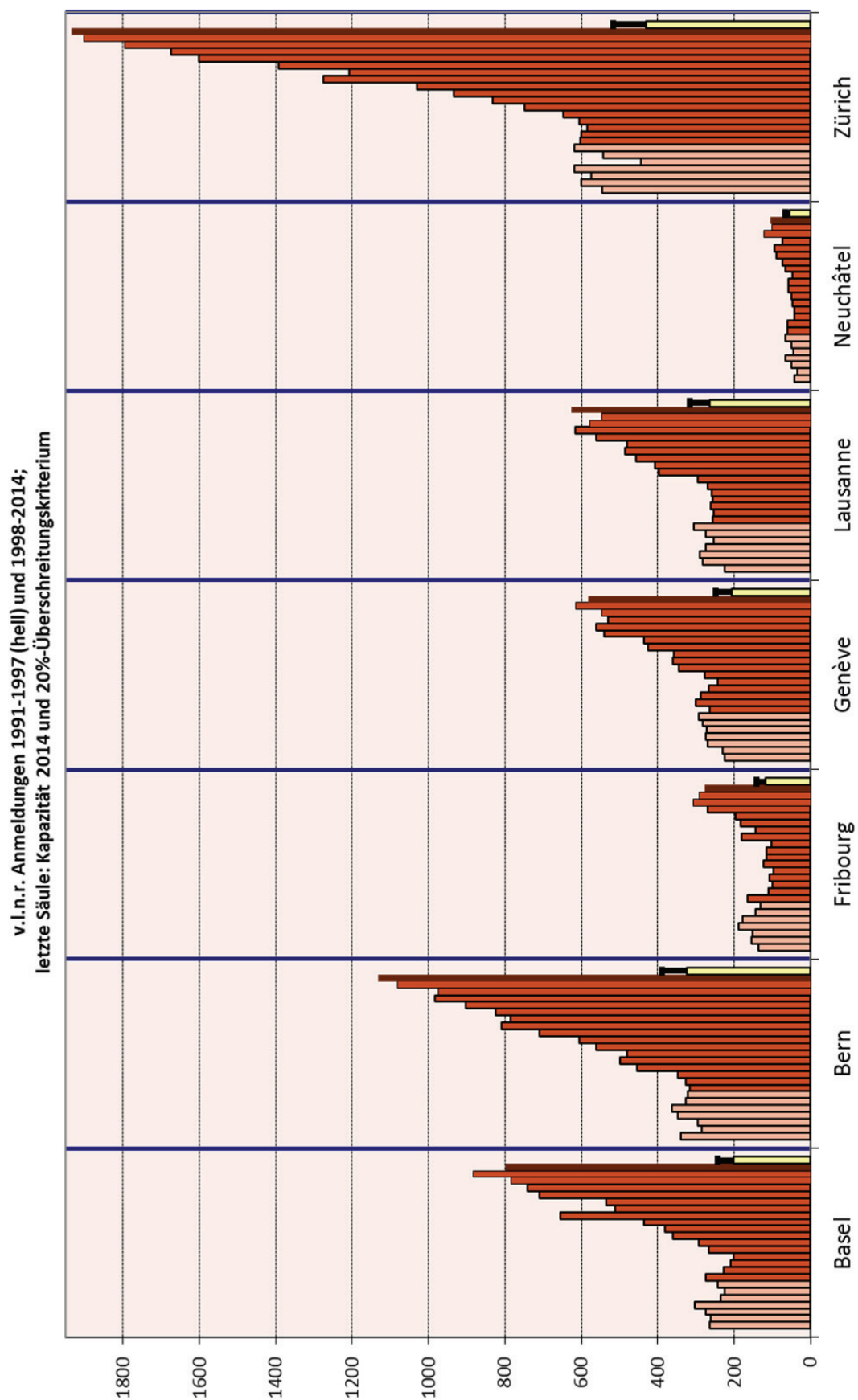


Abbildung 3: Anmeldezahlen pro Universität 1991 bis 2014 – hellere Säulen: Jahre ohne NC; dunklere Säulen: Jahre mit NC; letzte Säule: Kapazität 2014 und Markierung 20%-Überschreitungskriterium als Grenze für NC.

Besonders die folgende Abbildung zeigt, dass die Nachfrage nach Medizin-Studienplätzen einem eigenen Trend folgt, der nicht an demografische Entwicklungen oder die Entwicklung der sonstigen Studierendenzahlen gekoppelt ist. Dieser Trend deckt sich mit Erfahrungen in Deutschland und Österreich, wo es eine vergleichbare Versorgungsdebatte gibt.

Der immer noch hohe Status des Berufes, eine Zukunftssicherheit der Berufswahl (befördert durch die Diskussion, dass ein Ärztemangel besteht), die verbesserten Arbeitsbedingungen machen die Berufswahl attraktiv – und vielleicht attraktiver als die Wahl anderer Fächer.

Nahezu alle entwickelten Industrieländer müssen zu einem NC greifen, weil die Nachfrage das Studienplatzangebot deutlich übersteigt. Der NC in der Schweiz ist mit 3 Kandidaten auf einen Platz dabei im internationalen Massstab sogar noch einer der mildesten, es gibt keine „besondere Härte“ in der Schweiz, eher noch weiteres Steigerungspotential der Nachfrage.

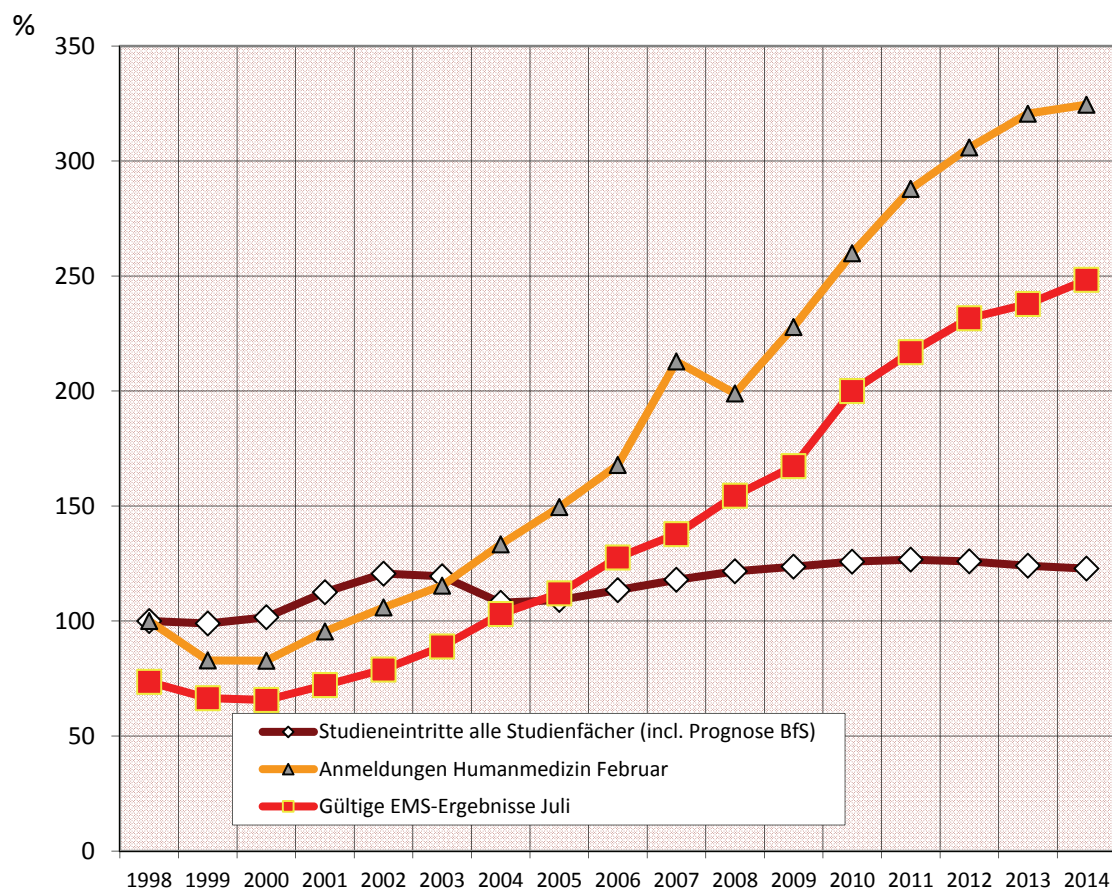


Abbildung 4: Alle Studieneintritte an Universitäten und Hochschulen (indexiert, 1998 = 100%) im Vergleich zu Medizin-Studienanmeldungen Februar und Bewerbungen mit gültigem EMS (indexiert, Anmeldungen 1998 = 100%) für Humanmedizin.

5 Anmeldung zum Medizinstudium und Test 2014

5.1 Anmeldestatistik Humanmedizin

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

	Humanmedizin (seit 2008 incl. Chiropraktik)												
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Kapazitäten (Univ. mit NC)	648	598	546	546	546	583	603	603	653	653	653	713	793
Anmeldungen im Februar	1080	1177	1360	1525	1712	2171	2029	2324	2651	2936	3120	3270	3310
... in % zu Kapazität	185	197	249	279	314	372	336	385	406	450	478	459	417
Anmeldungen zum EMS Mai	827	917	1083	1182	1356	1452	1612	1765	2109	2295	2461	2547	2653
Absolvierung EMS	770	871	1026	1107	1263	1374	1535	1664	2016	2172	2337	2403	2525
Übernahme Ergebnis Vorjahr	40	47	38	41	45	41	52	66	54	65	54	60	42
Bewerbungen Juli	805	907	1051	1143	1302	1405	1576	1707	2040	2212	2363	2426	2533*
Rückzugsquote zwischen Februar und Juli (%)	25.5	22.9	22.7	25.0	23.9	34.8	22.3	26.6	23.0	24.6	24.3	25.8	23.5
Zugeweilte Studienplätze mit „Überbuchung“	758	705	624	640	653	685	726	707	763	745	745	805	872^m
Abgewiesene Bewerbungen	47	202	426	503	649	720	850	1000	1277	1467	1618	1621	1661^m
% Bewerbungen, die Studienplatz erhalten	94	78	59	56	50	49	45	41	37	34	32	33	34^m

* 34 Personen wollen aus einem Studium der Zahnmedizin in ein höheres Semester Humanmedizin wechseln. Sie nahmen am Test teil, zählen aber nicht als Bewerbungen um die Studienplätze 2014. Vergleichbare Fälle wurden auch in den Vorjahren entsprechend herausgerechnet. 20 Plätze für Chiropraktik (77 Bewerbungen im Februar, 66 Testantritte, ein Übertrag aus 2013) sind mit eingerechnet. m Stand Oktober gemäss Modell. (Zahlen 1998 bis 2002 siehe Berichte des ZTD bis 2011.)

„Modell“ heisst, dass etwas 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 früh genau auszulasten (auch weil das Semester früh beginnt) und keine Nachrückverfahren notwendig werden zu lassen.

Betrachtet man den Anmeldeverlauf über die Jahre, nimmt vor allem die Quote der Personen weiter zu, die bei gültigem EMS-Ergebnis nicht zugelassen werden können. Die Zahl der Rückzüge zwischen Februar (Anmeldung Studium) und der Testteilnahme bleibt prozentual in etwa gleich.

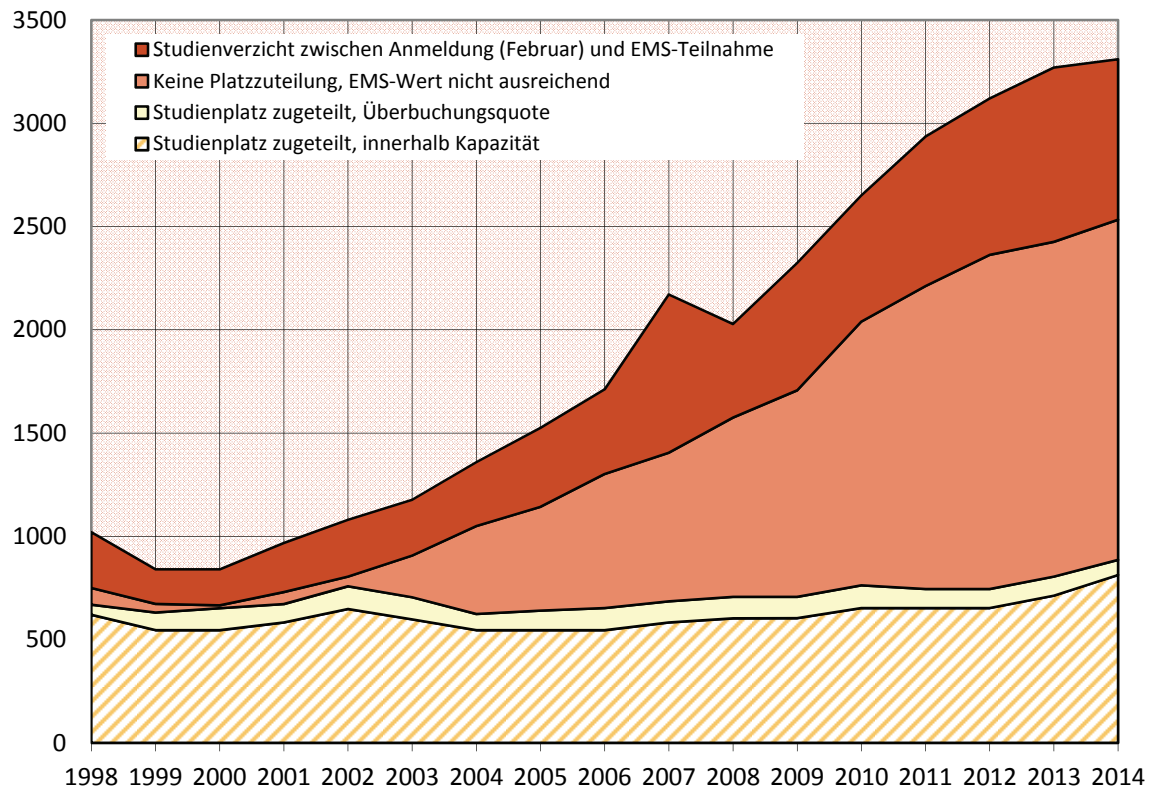


Abbildung 5: Anmeldeverlauf Humanmedizin. Seit 2002 nehmen die Anmeldezahlen und die Zahl der Testteilnahmen kontinuierlich zu. Die Zunahme der Anmeldezahlen 2007 war auf ein vereinfachtes nur elektronisches Anmeldeverfahren zurückzuführen, danach musste die Anmeldung zusätzlich wieder schriftlich erfolgen.

Der Kapazitätsausbau in den letzten beiden Jahren hat dafür gesorgt, dass die Zulassungsquote in etwa gleich bleibt. Die Nachfrage nimmt also ebenfalls ungefähr um das Dreifache des Anstiegs der Studienplätze im ersten Jahr zu. Dennoch ist der Numerus clausus im Vergleich zu anderen Ländern vergleichsweise mild. In Deutschland sind es 5 bzw. 12 (Winter- bzw. Sommersemester), in Österreich 6. Alle europäischen Länder mussten den Zugang zum Medizinstudium beschränken. Für Schweizer ist es im Vergleich zu vielen anderen Ländern leichter, Arzt zu werden.

5.2 Anmeldestatistik Veterinärmedizin

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.

	Veterinärmedizin												
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Kapazitäten (Univ. mit NC)	185	170	150	150	150	150	150	150	150	150	150	150	150
Anmeldungen im Februar	254	271	273	318	348	411	365	371	432	408	406	478	487
... in % zu Kapazität	137	159	182	212	232	274	243	247	288	272	271	319	325
Anmeldungen zum EMS Mai	206	225	207	257	279	301	290	294	360	308	327	365	403
Absolvierung EMS	203	217	196	244	268	288	272	281	342	295	317	350	384
Übernahme Ergebnis Vorjahr	3	11	13	8	10	7	15	16	11	16	8	24	17
Bewerbungen Juli	204	228	209	252	278	295	287	297	353	311	325	374	401
Rückzugsquote zwischen Februar und Juli (%)	19.3	15.9	23.4	20.8	20.1	28.2	21.3	19.9	18.0	23.8	20.0	21.8	17.7
Zugeteilte Studienplätze mit „Überbuchung“	202	194	175	180	170	173	173	173	173	173	173	173	173^m
Abgewiesene Bewerbungen	2	34	34	72	108	122	114	124	180	138	152	201	228^m
% Bewerbungen, die Studienplatz erhalten	99	85	84	71	61	59	60	58	49	56	53	46	43^m

m: Stand Oktober gemäss Modell.

Im Fach Veterinärmedizin hat es auch 2014 einen neuen Anmelderekord zu Studium und Test gegeben, auch dieses Studium hat offenbar an Attraktivität gewonnen. Die Rückzugsquote liegt im Bereich der Vorjahre – erneut liegt die Chance auf Zuteilung eines Platzes für die Personen mit absolviertem Eignungstest aber unter 50%.

5.3 Anmeldestatistik Zahnmedizin

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

	Zahnmedizin										
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Kapazitäten (Univ. mit NC)	142	142	142	142	142	142	142	142	139	139	131
Anmeldungen im Februar	208	190	217	267	197	239	315	324	332	396	343
... in % zu Kapazität	146	134	153	188	138	168	222	228	239	285	262
Anmeldungen zum EMS Mai	163	159	168	179	164	181	233	250	257	284*	273
Absolvierung EMS	151	150	157	169	154	174	220	239	248	275	264
Übernahme Ergebnis Vorjahr	1	11	20	32	13	31	37	37	25	39*	26
Bewerbungen Juli	151	161	174	201	167	205	257	276	273	313	290
Rückzugsquote zwischen Februar und Juli (%)	26.9	15.2	19.3	24.7	15.2	14.3	18.4	14.8	17.8	21.0	15.5
Zugeteilte Studienplätze mit „Überbuchung“	151	161	172	165	167	170	165	159	159 ^m	161	150^m
Abgewiesene Bewerbungen	0	0	2	36	0	35	92	117	114 ^m	152	140^m
% Bewerbungen, die Studienplatz erhalten	100	100	99	82	99	83	64	58	58 ^m	51	52^m

m Stand Juli gemäss Modell; *: eine Person wollte in einem höheren Semester von Veterinärmedizin nach Zahnmedizin wechseln, zählte nicht als Bewerbung für das laufende Jahr.

Da die Zahl der beantragten Studienwechsel nach dem ersten Jahr in die Humanmedizin auch 2014 nach wie vor hoch ist, muss betont werden, dass Zahnmedizin kein „Parkstudium“ für Humanmedizin ist. Dieser Personenkreis muss sich im Falle eines Disziplinwechsels erneut dem EMS stellen und wegen der bisher immer steigenden Nachfrage im Folgejahr dann auch einen höheren Zulassungswert erreichen als im Jahr der ersten Medizinbewerbung. Damit im System der Zulassung die Zahnmedizin-Studienplätze ausgelastet werden, kann der Studienwechsel wie bisher nicht bereits vor der

Studienaufnahme beantragt werden („erst einmal Zahnmedizin anmelden, wenn der EMS-Wert reicht, dann schnell zu Humanmedizin wechseln“).

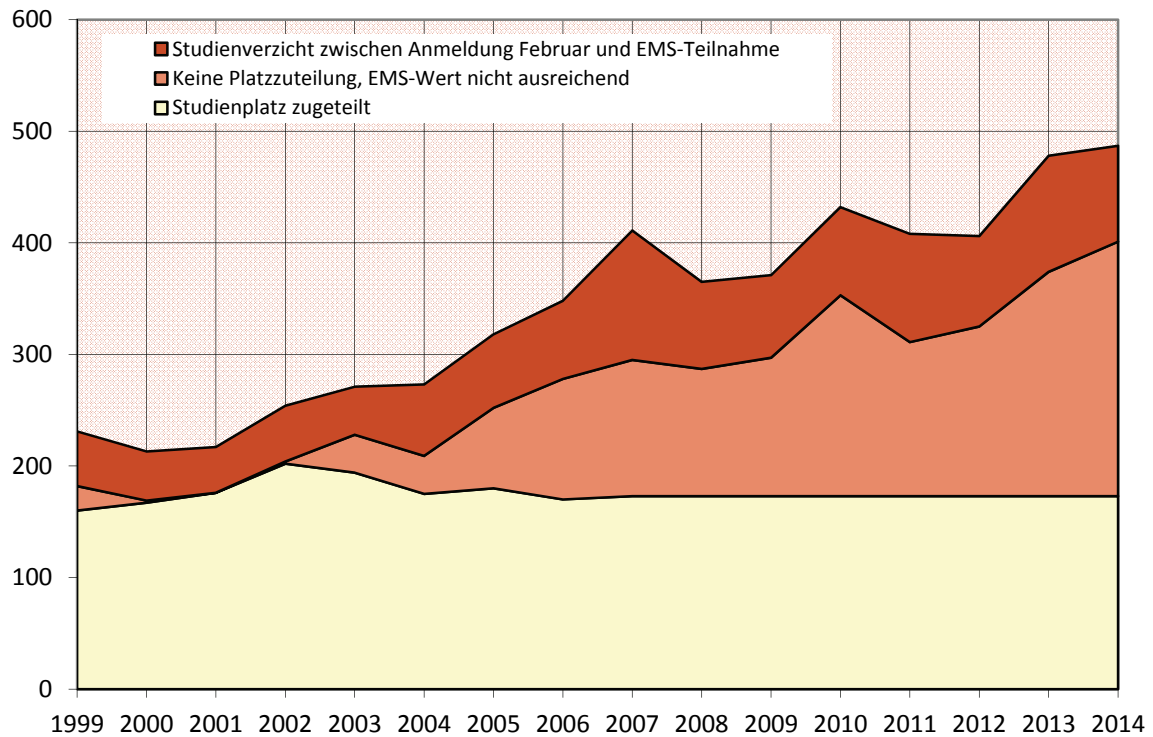


Abbildung 6: Anmeldeverlauf für Veterinärmedizin 1999 bis 2014 im Vergleich (Univ. mit NC)

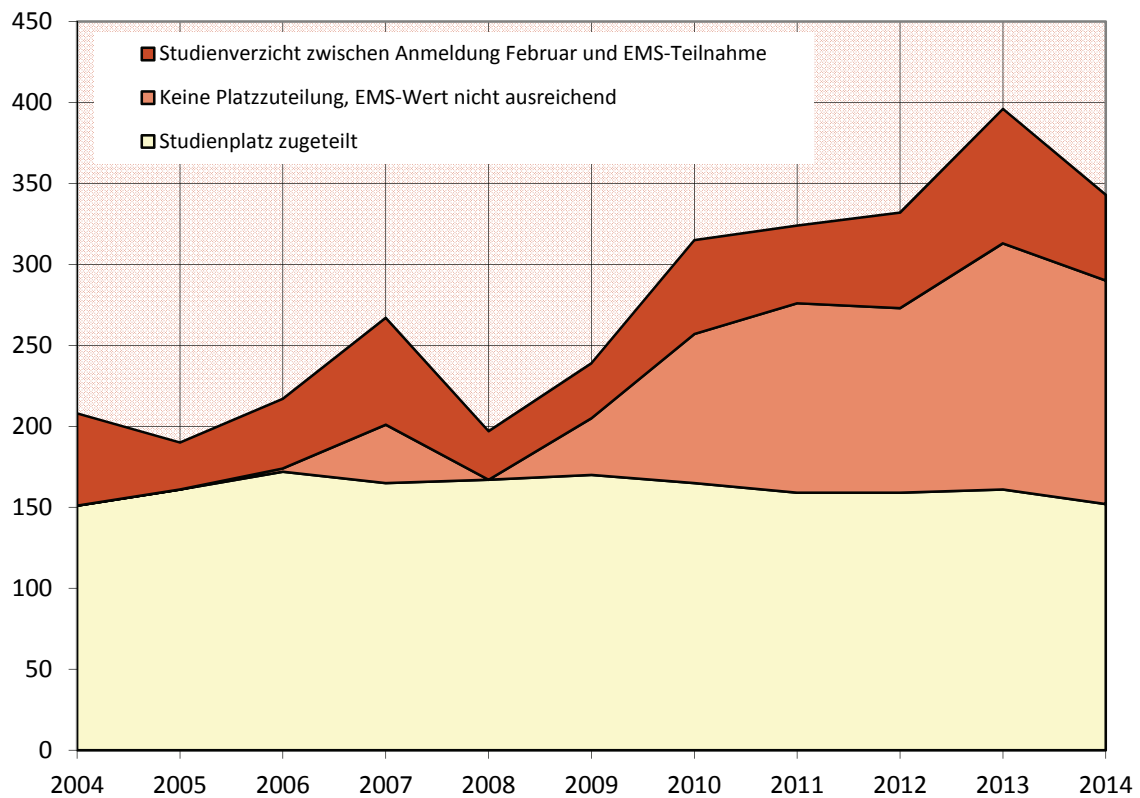


Abbildung 7: Anmeldeverlauf für Zahnmedizin 2004 bis 2014 im Vergleich (Univ. mit NC)

So sollte gewährleistet bleiben, dass sich nur Personen für Zahnmedizin bewerben, die ein wirkliches Interesse an diesem Studium bzw. dem Beruf haben. Bei Zahnmedizin lag der Verdacht nahe, dass wegen des niedrigeren notwendigen Test-Prozentranges für die Zulassung erst einmal Zahnmedizin begonnen wird, dann aber ein Umstieg versucht wird, der frühestens nach einem Jahr möglich wäre. Die Bedingungen für einen Disziplinwechsel ab dem 2. Studienjahr (in alle möglichen Richtungen) sind an den Universitäten Basel, Bern, Freiburg und Zürich in den jeweiligen kantonalen Verordnungen klar geregelt. Der Studienwechsel ist kein „Automatismus“. Es bedarf immer zweier Voraussetzungen:

- Es muss der Zulassungswert für Humanmedizin erreicht werden, was in der Regel eine erneute Testteilnahme erfordert. Da der Zulassungswert aufgrund der Nachfrageentwicklung von Jahr zu Jahr steigt, wird die Hürde also höher als im Jahr der Erstbewerbung. Nur rund 1/3 der Kandidaten erfüllen pro Jahr diese Bedingung.
- Es muss auch Kapazität in Humanmedizin an der gewünschten Universität vorhanden sein, die Universität muss den Wechsel genehmigen (was wegen fehlender Kapazitäten nicht immer möglich ist und durch den Druck zur maximalen Auslastung der Kapazitäten auch eher die Ausnahme sein sollte).

Darauf hinzuweisen ist auch, dass ein Studium der Zahnmedizin auch spezifische Studienanforderungen stellt, unter anderem sehr gut entwickelte manipulative Fähigkeiten und Feinmotorik. Dies wird gemeinsam mit den Universitäten in Zukunft noch besser herausgearbeitet – denn wenn man diesen Aspekt bei der Studienbewerbung nicht berücksichtigt, sind Frustrationen vorprogrammiert. Deshalb ist Zahnmedizin als „Sprungbrett“ in die Humanmedizin nicht geeignet und im Rahmen des zusätzlichen Basler Self-Assessments www.sazamed.ch wird für Zahnmedizin ein eigenes Anforderungsbild vermittelt.

Tabelle 4: EMS-Teilnahme wegen Antrag Studienwechsel Zahn- nach Humanmedizin (in ein höheres Studienjahr).

	Wechselwunsch	davon BS	davon BE	davon FR	davon ZH
2014	37 (20)	17 (6)	15 (11)	2 (2)	3 (1)
2013	37 (14)	16 (6)	8 (4)	3 (1)	10 (3)
2012	28 (7)	15 (3)	7 (2)		6 (2)
2011	25 (8)	14 (1)	6 (4)	1 (0)	4 (3)
2010	30 (10)	20 (5)	9 (5)		1 (0)
2009	23 (8)	16 (4)	1 (0)	2 (1)	4 (3)
2008	11 (4)	9 (3)	2 (1)		
2007	10 (5)	10 (5)			
2006	7 (2)	6 (2)		1 (0)	

In Klammern Zahl der Personen, die das Zulassungskriterium erfüllt haben – die allerdings nur dann wechseln dürfen, wenn auch Kapazitäten in Humanmedizin vorhanden sind.

Für die Zulassung zur **Chiropraktik** ergibt sich ebenfalls eine Besonderheit. Diese Personen bilden selbst nur eine sehr kleine Kohorte, müssen aber das Kriterium der Humanmedizin erfüllen. Der vorhandene Mittelwertunterschied zur Humanmedizin beim Testwert hat in diesem Jahr weiter abgenommen.

Damit ein späterer Wechsel zur Humanmedizin immer noch möglich ist, wird diese Besonderheit gewünscht, da sich dies ansonsten als „leichterer“ Zugang zum Studium der Humanmedizin etablieren würde.

Auf die 20 für die Chiropraktik reservierten Plätze erfolgen 2014 19 Zulassungen (18 in 2013). Dies ist ein Zeichen dafür, dass sich in den letzten beiden Jahren mehr leistungsstarke Personen für dieses Fach interessieren. Der verbleibende Platz wird an eine Person vergeben, deren Wunsch ein Studium der Humanmedizin ist.

Tabelle 5: Kennwerte Chiropraktik (CP) und Humanmedizin ohne CP (HM) – Testwert und mittlerer Rangplatz für die Testteilnahmen 2011 bis 2014. Für 2014 wurde der Testwert nach dem Verfahren vor 2014 zu Vergleichszwecken berechnet.

		2012			2013			2014		
		n	m	s	n	m	s	n	m	s
Testwert (klassisch)	CP	48	98.3	8.5	52	99.7	10.0	66	100.1	8.3
	HM	2289	100.8	10.0	2351	100.8	10.0	2459	100.7	10.0
Mittlerer Rangplatz	CP	48	533	157	52	504	182	66	502	155
	HM	2289	485	182	2351	485	180	2459	487	182

Tabelle 6: Zulassung Stand Juli 2014 (66 EMS-Teilnahmen 2014 und 1 Übernahme Testergebnis aus 2013).

	Chiropraktik	Humanmedizin ohne CP
Zugelassen	19 (28%); 20 Plätze vorhanden	872 (34%)
Nicht zugelassen	48 (72%)	1661 (66%)
Total	67	2533

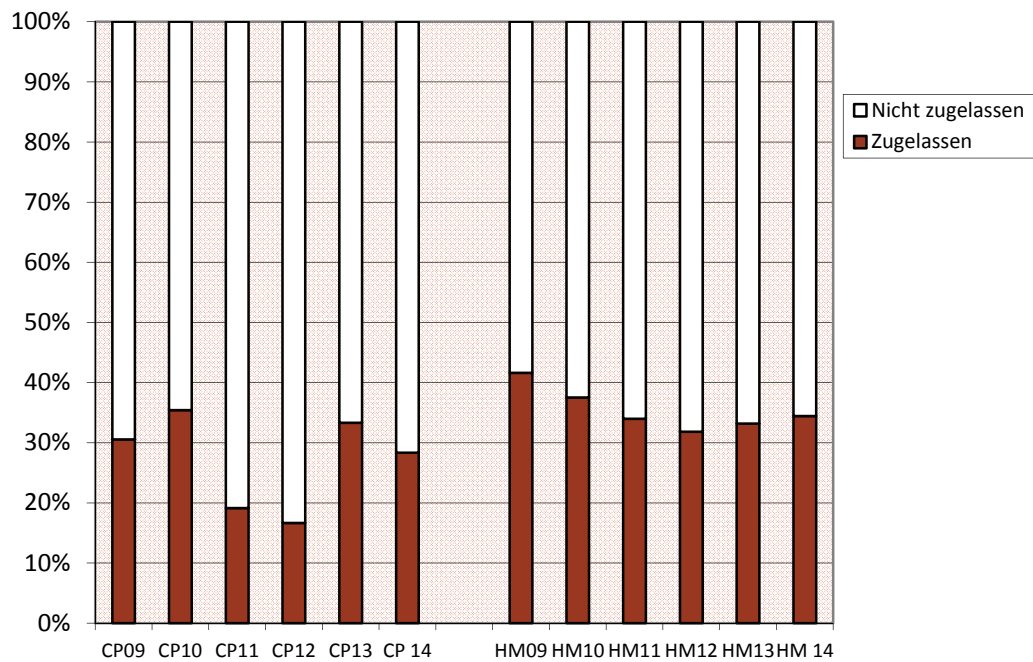


Abbildung 8: Aufteilung für Bewerbungen Humanmedizin und Chiropraktik nach der Zulassung.

Tabelle 7: Anmeldeverlauf für Human- (HM) inkl. Chiropraktik (CP), Veterinär- (VM) und Zahnmedizin (ZM). Jeweils ohne Personen, die Wechsel in ein höheres Studienjahr beantragt haben.

	Studienbeginn 2011			Studienbeginn 2012			Studienbeginn 2013			Studienbeginn 2014		
	Anmel- dungen am 15.2.	Pers. m. gültigem Test	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Test	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Test	Rück- züge in %	Anmel- dungen am 15.2.	Pers. m. gültigem Test	Rück- züge in %
Basel HM	659	481	27.0	700	517	26.1	780	577	26.0	717	549	23.4
Bern HM	691	534	22.7	680	533	21.6	697	516	26.0	747	593	20.6
Freiburg HM	243	191	21.4	293	221	24.6	266	200	24.8	257	203	21.0
Zürich HM	1264	960	24.1	1378	995	27.8	1454	1079	25.8	1478	1121	24.2
Zürich CP	67	46	31.3	69	48	30.4	73	54	26.0	77	67	13.0
Total	2924	2212	24.2	3120	2314	25.8	3270	2426	25.8	3276	2533	22.7
Bern VM	206	156	24.3	218	175	19.7	267	205	23.2	274	225	17.9
Zürich VM	215	155	27.9	188	142	24.5	211	169	19.9	213	176	17.4
Total	421	311	26.1	406	317	21.9	478	374	21.8	487	401	17.7
Basel ZM	83	73	12.0	85	66	22.4	100	77	23.0	63	58	7.9
Bern ZM	87	75	13.8	76	57	25.0	116	90	22.4	96	76	20.8
Freiburg ZM	26	8	69.2	14	14	0.0	25	18	28.0	19	16	15.8
Zürich ZM	129	120	7.0	158	111	29.7	155	128	17.4	165	140	15.2
Total	325	276	15.1	333	248	25.3	396	313	21.0	343	290	15.5

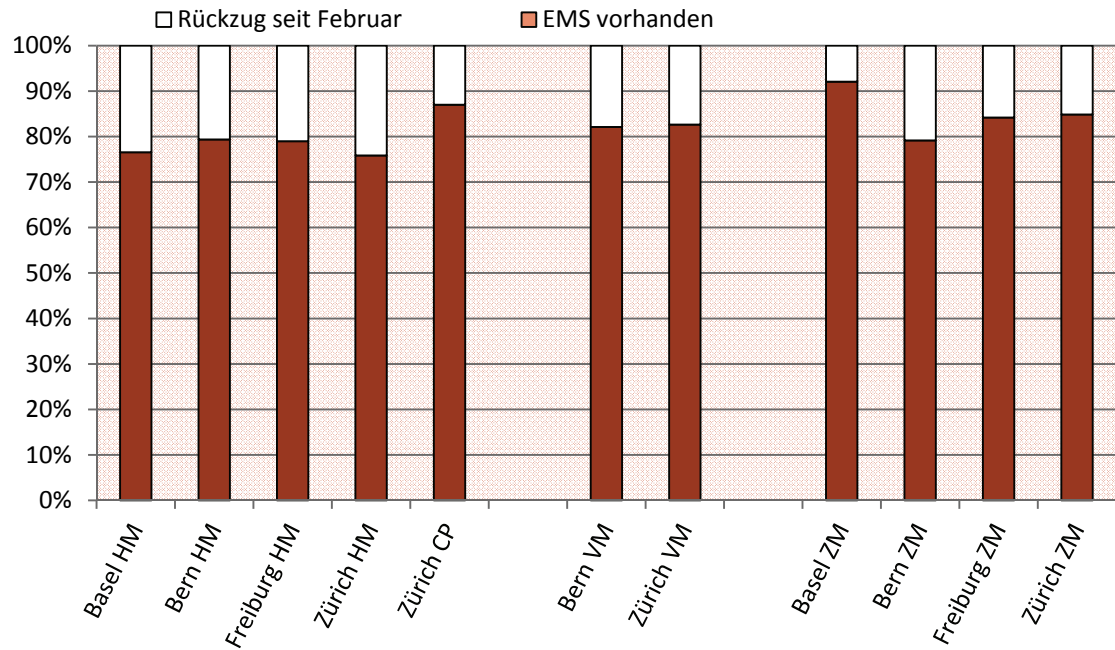


Abbildung 9: Rückzüge nach Disziplin und Universität zwischen Februar und EMS für 2014.

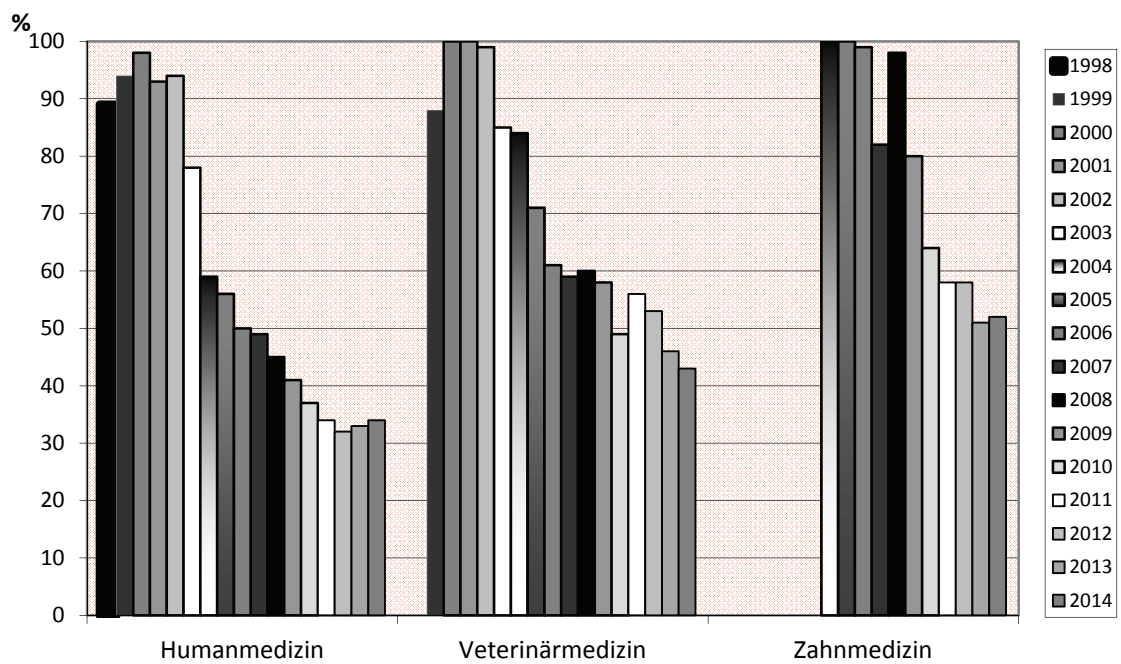


Abbildung 10: Prozent Personen mit gültigem Test-Prozentrang, denen ein Studienplatz zugewiesen werden kann.

5.4 Grösse der Testlokale

In diesem Jahr erfolgte die Testabnahme wiederum in 30 Testlokalen verteilt auf 9 Testorte. Die Grösse der Testlokale variierte im bisher üblichen Rahmen. Die Zahl der Betreuer war proportional zur Teilnehmerzahl vorgegeben. Wiederum wurde in einem grösseren Testlokal in Zürich auch das „Sektorkonzept“ angewendet, d.h. ein zentraler Testleiter für Zeitnahme und Instruktionsverlesung, alle anderen administrativen Aufgaben in getrennten Sektoren.

Es gibt keine Hinweise seitens des Betreuungspersonals oder der 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).

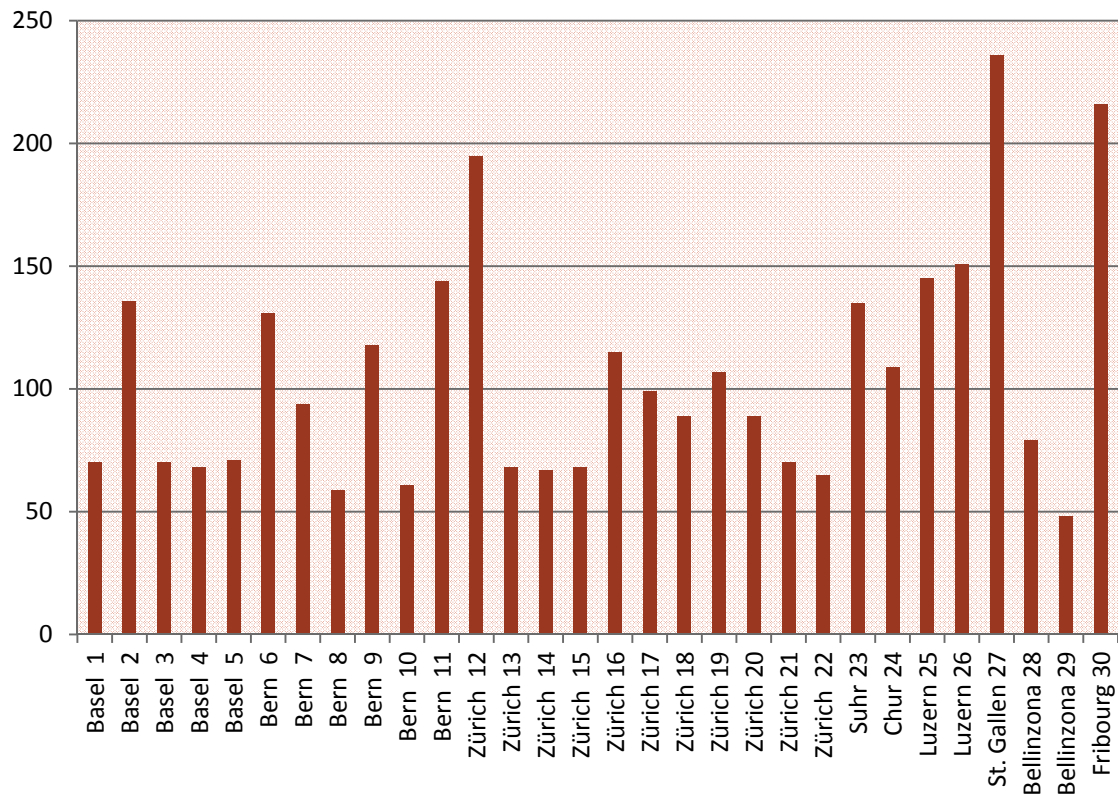


Abbildung 11: Grösse (Zahl der Personen) und Aufteilung der Testlokale 2014.

5.5 Testorte und Wunschuniversitäten

Tabelle 8: Testorte und Wunschuniversität (erste Wahl) 2014; Aufteilung pro Testort auf die Wunschuniversitäten.

Disziplin	Testort	Wunschuniversität				
		Bern	Basel	Freiburg	Zürich	Gesamt
Humanmedizin	Basel	3	321	4	16	344
	Bellinzona	8	18	33	36	95
	Bern	404	16	39	28	487
	Chur	21	14	8	50	93
	Fribourg	7	4	91	15	117
	Luzern	90	68	5	73	236
	St. Gallen	31	37	5	113	186
	Suhr	15	55	5	34	109
	Zürich	20	28	10	800	858
	Gesamt	599	561	200	1165	2525
Veterinärmedizin	Basel	21				31
	Bellinzona	8				24
	Bern	66				67
	Chur	6				12
	Fribourg	88				94
	Luzern	10				28
	St. Gallen	3				25
	Suhr	6				16
	Zürich	5				87
	Gesamt	213				384
Zahnmedizin	Basel		35	2	3	40
	Bellinzona		1	2	5	8
	Bern	50	3			53
	Chur	1			3	4
	Fribourg			5		5
	Luzern	11	7	1	13	32
	St. Gallen		2	2	21	25
	Suhr	3	3		4	10
	Zürich	4	4	1	78	87
	Gesamt	69	55	13	127	264
Gesamt	Basel	24	356	6	29	415
	Bellinzona	16	19	35	57	127
	Bern	520	19	39	29	607
	Chur	28	14	8	59	109
	Fribourg	95	4	96	21	216
	Luzern	111	75	6	104	296
	St. Gallen	34	39	7	156	236
	Suhr	24	58	5	48	135
	Zürich	29	32	11	960	1032
	Gesamt	881	616	213	1463	3173

5.6 Wunschuniversität und Testort nach Wohnkanton

In der folgenden Tabelle sind die Testorte nach Wohnkanton für die einzelnen Disziplinen aufgeschlüsselt. Angegeben sind auch Vergleichswerte seit 2003 für die Jahre mit NC.

Tabelle 9: Gewünschte Studienorte Humanmedizin nach Wohnkantonen 2014 und Vergleich mit 2003 bis 2013 (1998 bis 2002 siehe Bericht 17 des ZTD).

Wohnkanton/ Wohnort	Humanmedizin 2014 gewünschter Studienort					Total									
	BE	BS	FR	ZH	Total	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004
AG	25	85	4	112	226	198	185	172	161	121	135	124	99	81	93
AI	4	1	0	3	8	12	12	8	5	2	0	3	4	1	4
AR	4	7	0	12	23	26	19	23	24	12	15	12	7	4	8
BE	328	11	14	17	370	309	352	323	283	279	242	204	206	174	179
BL	0	123	1	5	129	142	138	113	110	92	77	81	56	64	54
BS	0	120	1	5	126	131	132	115	116	73	89	79	77	61	48
FR	12	0	89	5	106	78	94	96	75	67	61	46	45	54	42
GE	0	0	1	2	3	3	3	4	4	4	3	2	1	0	0
GL	2	1	1	8	12	15	8	11	9	9	6	1	7	5	1
GR	19	6	5	41	71	72	68	64	72	52	48	44	40	34	41
JU	0	1	3	1	5	5	3	10	5	7	3	5	1	0	3
LU	65	51	4	42	162	147	134	133	138	111	112	98	70	65	53
NE	3	0	3	1	7	7	5	4	1	1	4	1	0	3	1
NW	5	5	0	2	12	15	13	7	4	8	12	12	2	4	5
OW	9	2	0	3	14	7	10	9	10	6	8	5	6	5	5
SG	24	28	6	109	167	137	150	139	132	113	117	111	86	81	55
SH	1	4	0	21	26	30	32	21	19	10	18	15	9	11	6
SO	39	40	3	16	98	86	72	75	77	58	63	37	55	42	42
SZ	9	2	1	41	53	41	36	29	34	25	27	26	17	19	18
TG	4	6	1	46	57	70	69	66	62	53	38	40	42	38	24
TI	8	18	33	34	93	108	97	86	70	75	44	50	45	26	34
UR	2	2	1	4	9	10	8	9	16	12	8	9	8	3	1
VD	2	1	12	5	20	15	15	16	14	9	8	10	7	3	1
VS	12	3	8	7	30	39	39	43	36	22	25	22	20	10	7
ZG	6	8	1	36	51	54	49	50	34	26	25	27	30	20	17
ZH	7	9	8	562	586	599	538	481	452	383	320	283	295	270	263
FL	3	5	0	6	14	16	9	4	13	5	3	5	7	9	5
Übriges Ausland	6	22	0	19	47	31	47	61	40	29	24	22	21	20	16
Total	599	561	200	1165	2525	2403	2337	2172	2016	1664	1535	1374	1263	1107	1026

Das Fürstentum Liechtenstein (FL) wird als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

Tabelle 10: Gewünschte Studienorte Veterinärmedizin nach Wohnkantonen 2014 und Vergleich mit 2003 bis 2013.

Wohnkanton/ Wohnort	Veterinärmedizin 2014 gewünschter Studienort			Total									
	BE	ZH	Total	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004
AG	5	19	24	24	21	21	24	11	14	16	27	25	11
AI	0	1	1	0	1	0	0	1	1	1	0	0	1
AR	0	4	4	5	2	2	1	1	1	1	2	2	2
BE	54	1	55	51	45	41	62	32	40	35	38	24	29
BL	13	4	17	15	17	10	14	4	10	15	7	10	12
BS	3	4	7	5	5	4	7	8	7	7	1	9	4
FR	17	0	17	26	10	17	17	17	19	10	8	9	6
GE	16	2	18	12	12	12	16	15	6	12	13	6	2
GL	3	2	5	4	2	3	2	1	3	2	2	4	0
GR	4	7	11	12	12	14	10	12	9	9	10	6	10
JU	3	1	4	4	2	1	1	2	5	8	7	0	4
LU	7	9	16	17	19	20	28	20	14	16	14	6	11
NE	11	0	11	8	7	3	3	13	8	6	6	10	4
NW	1	1	2	2	1	0	2	0	1	3	3	2	1
OW	3	1	4	2	1	2	1	1	1	1	1	0	0
SG	2	13	15	16	16	14	13	18	17	20	12	19	14
SH	0	5	5	2	2	4	2	1	2	6	1	1	1
SO	11	2	13	12	7	9	8	10	9	6	10	7	7
SZ	2	9	11	12	12	10	3	4	4	5	6	2	4
TG	1	9	10	10	7	4	7	8	10	12	12	8	5
TI	9	15	24	12	17	17	17	8	17	11	8	10	6
UR	0	0	0	0	4	0	0	1	0	1	1	1	2
VD	30	2	32	27	26	27	29	27	22	19	25	31	15
VS	16	0	16	10	12	7	17	11	8	5	7	5	7
ZG	0	4	4	5	5	2	3	2	6	5	7	3	4
ZH	1	53	54	48	45	44	50	49	32	50	36	39	29
FL	0	1	1	2	1	2	1	1	1	0	1	0	0
Übriges Ausland	1	2	3	7	6	5	4	3	5	6	3	5	1
Total	213	171	384	350	317	295	342	281	272	288	268	244	196

Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt. (1999 bis 2002 siehe Bericht 17 des ZTD).

Tabelle 11: Gewünschte Studienorte Zahnmedizin nach Wohnkantonen 2014 und Vergleich mit 2004 (Einführung des NC für Zahnmedizin) bis 2013.

Wohn- kanton/ Wohnort	Zahnmedizin 2014 gewünschter Studienort					Total									
	BE	BS	FR	ZH	Total	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004
AG	5	4	0	14	23	26	33	23	16	13	11	11	14	10	14
AI	0	0	1	0	1	2	0	1	0	0	2	1	0	1	0
AR	0	0	0	2	2	2	1	3	5	3	0	1	2	2	4
BE	41	1	1	0	43	48	38	36	31	41	22	31	32	27	31
BL	0	16	0	1	17	17	12	19	10	10	12	11	5	11	10
BS	0	14	0	0	14	23	22	10	10	7	11	5	7	6	4
FR	0	2	2	0	4	4	8	7	7	5	4	6	3	4	1
GE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GL	0	0	0	2	2	2	3	3	0	0	0	2	1	3	0
GR	1	0	0	2	3	6	6	5	6	6	4	6	5	3	8
JU	0	1	0	0	1	0	0	0	2	0	2	2	0	0	0
LU	8	6	1	7	22	14	13	15	19	7	11	9	10	5	2
NE	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0
NW	0	1	0	1	2	4	1	1	2	0	2	0	0	1	1
OW	1	0	0	1	2	1	0	1	1	0	1	0	0	0	2
SG	0	1	3	17	21	24	24	12	13	18	17	17	13	12	4
SH	0	0	0	4	4	1	2	0	0	1	3	0	0	5	3
SO	6	4	0	2	12	15	15	15	14	10	10	13	10	2	9
SZ	1	1	0	6	8	5	3	4	8	2	4	3	4	1	2
TG	0	0	0	5	5	5	5	2	9	5	4	2	4	9	6
TI	0	1	2	5	8	4	9	8	7	2	5	10	4	3	8
UR	0	0	0	1	1	1	3	2	2	0	1	3	2	2	0
VD	0	0	2	0	2	1	0	2	2	0	1	0	0	0	0
VS	1	0	1	0	2	3	0	0	2	2	1	4	5	4	4
ZG	1	0	0	6	7	3	2	2	1	3	3	2	2	1	1
ZH	2	3	0	48	53	58	43	56	44	36	15	25	28	33	33
FL	1	0	0	1	2	4	0	3	1	1	4	1	1	2	2
Übriges Ausland	1	0	0	2	3	2	5	8	8	2	4	3	4	3	2
Total	69	55	13	127	264	275	248	239	220	174	154	169	157	150	151

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 der Kapazitätsplanung für die Testorte.

Tabelle 12: Testorte nach Wohnkantonen 2014.

Wohn- kanton/ Wohnort	Testort									
	Basel	Bellin- zona	Bern	Chur	Freiburg	Luzern	St. Gallen	Suhr	Zürich	Total
AG	40	1	4			5		116	107	273
AI							10			10
AR							27		2	29
BE	3		446	1	9	3		1	5	468
BL	157							1	5	163
BS	147									147
FR		1	43		83					127
GE					21					21
GL			1	4					14	19
GR		2	2	72		2			7	85
JU	1				9					10
LU	5		9			179			7	200
NE			2		16					18
NW						15			1	16
OW						20				20
SG	2		2	17			147		35	203
SH	1						1		33	35
SO	34		60			2		16	11	123
SZ			1			28			43	72
TG	2					1	41		28	72
TI	1	119		1		1			3	125
UR			2			7			1	10
VD			3		51					54
VS	1		21		24				2	48
ZG						33			29	62
ZH	2	3	4	1	1		3	1	678	693
FL	1		1	13					2	17
Übriges Ausland	18	1	6		2		7		19	53
Total	415	127	607	109	216	296	236	135	1032	3173

Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

5.7 Teilnahmen am EMS (NC) nach Kantonen seit 1998

Bisher fanden 30207 Testteilnahmen in der Schweiz im Rahmen eines NC statt, bei denen ein gültiges EMS-Ergebnis erzielt wurde. Seit 1998 unterliegt Humanmedizin, seit 1999 Veterinärmedizin und erst seit 2004 Zahnmedizin einem NC an den Universitäten Basel, Bern, Freiburg und Zürich².

Tabelle 13: Gesamtzahl der Teilnehmer am EMS 1998 bis 2014 – nur Univ. mit NC.

Wohnkanton/ Wohnort	Disziplin			Geschlecht		Sprache			Total
	Human- medizin	Veterinär- medizin	Zahn- medizin	männlich	weiblich	deutsch	franzö- sisch	italie- nisch	
AG	1933	278	194	877	1528	2400	2	3	2405
AI	67	10	8	43	42	85			85
AR	207	27	25	92	167	259			259
BE	3619	605	380	1657	2947	4525	78	1	4604
BL	1318	160	134	597	1015	1606	6		1612
BS	1257	81	119	555	902	1454		3	1457
FR	966	196	53	445	770	422	790	3	1215
GE	31	167	0	40	158	6	192		198
GL	107	32	16	44	111	154	1		155
GR	757	148	58	356	607	919		44	963
JU	49	47	7	39	64	9	94		103
LU	1485	228	127	681	1159	1837	1	2	1840
NE	39	105	3	33	114	12	135		147
NW	121	21	14	65	91	156			156
OW	111	14	8	42	91	133			133
SG	1593	217	175	706	1279	1985			1985
SH	248	35	19	111	191	302			302
SO	860	130	125	448	667	1113	2		1115
SZ	415	83	44	168	374	540	1	1	542
TG	697	125	56	275	603	878			878
TI	881	184	68	490	643	24	1	1108	1133
UR	110	13	17	52	88	140			140
VD	128	355	8	112	379	28	463		491
VS	368	133	27	198	330	323	205		528
ZG	446	68	27	220	321	539	1	1	541
ZH	5505	640	424	2345	4224	6523	17	29	6569
FL	110	14	21	52	93	145			145
Ausland	404	58	44	210	296	460	35	11	506
Total	23832	4174	2201	10953	19254	26977	2024	1206	30207

Das Fürstentum Liechtenstein (FL) ist als Mitglied der Interkantonalen Universitätsvereinbarung aufgeführt.

² Gemäss des Beschlusses der Schweizerischen Universitätskonferenz von 2007 werden keine Testergebnisse mehr nach Kantonen aufgeschlüsselt veröffentlicht.

5.8 Testabsolvierung nach Alter und Geschlecht

Tabelle 14: Testabsolventen nach Geburtsjahr und Geschlecht 2014.

Disziplin	Geburtsjahre	Geschlecht				Total
		männlich		weiblich		
		Anzahl	Anzahl der Zeilen (%)	Anzahl	Anzahl der Zeilen (%)	
Human- medizin	ab 1994	570	35.4%	1038	64.6%	1608
	1991 - 1993	338	47.1%	380	52.9%	718
	bis 1990	81	40.7%	118	59.3%	199
	Gesamt	989	39.2%	1536	60.8%	2525
...davon Chiropraktik	ab 1994	6	14.6%	35	85.4%	41
	1991 - 1993	7	38.9%	11	61.1%	18
	bis 1990	5	71.4%	2	28.6%	7
	Gesamt	18	27.3%	48	72.7%	66
Veterinär- medizin	ab 1994	31	13.0%	207	87.0%	238
	1991 - 1993	24	21.6%	87	78.4%	111
	bis 1990	8	22.9%	27	77.1%	35
	Gesamt	63	16.4%	321	83.6%	384
Zahnmedizin	ab 1994	37	30.1%	86	69.9%	123
	1991 - 1993	40	40.4%	59	59.6%	99
	bis 1990	21	50.0%	21	50.0%	42
	Gesamt	98	37.1%	166	62.9%	264
Total	ab 1994	638	32.4%	1331	67.6%	1969
	1991 - 1993	402	43.3%	526	56.7%	928
	bis 1990	110	39.9%	166	60.1%	276
	Gesamt	1150	36.2%	2023	63.8%	3173

In **Humanmedizin** beträgt das Geschlechterverhältnis weiterhin rund 60 zu 40 zugunsten der Frauen. Seit 6 Jahren ist dieses Verhältnis aber konstant (vergleiche Abbildung 13). Dies bedeutet, dass der Zuwachs der Nachfrage sich aus beiden Geschlechtern gleichartig rekrutiert. Die vereinzelt zu findende Interpretation, dass vor allem das Interesse von Frauen für die Zunahme der Bewerbungszahlen verantwortlich ist, trifft also weiterhin nicht zu.

In **Veterinärmedizin** hat sich das Verhältnis ebenfalls stabilisiert – etwas über 80% Frauen interessieren sich für dieses Studium. Hier hat man allerdings die Sorge, dass für eine Berufsausübung in der Grosstierhaltung zu wenige Absolventen ausgebildet werden – Frauen wählen diese Richtung weniger häufig. Gegen eine „Bevorzugung“ der Männer bei der Zulassung, um dies auszugleichen, spricht das Prinzip der Gleichbehandlung. Bei zu niedrigem Test-Prozentrang wäre zudem die Studien-Erfolgsprognose ungünstiger.

Für **Zahnmedizin** sind die Verhältnisse mit Humanmedizin weiterhin vergleichbar, hier entspricht die Nachfrage der Frauen derjenigen des Vorjahres.

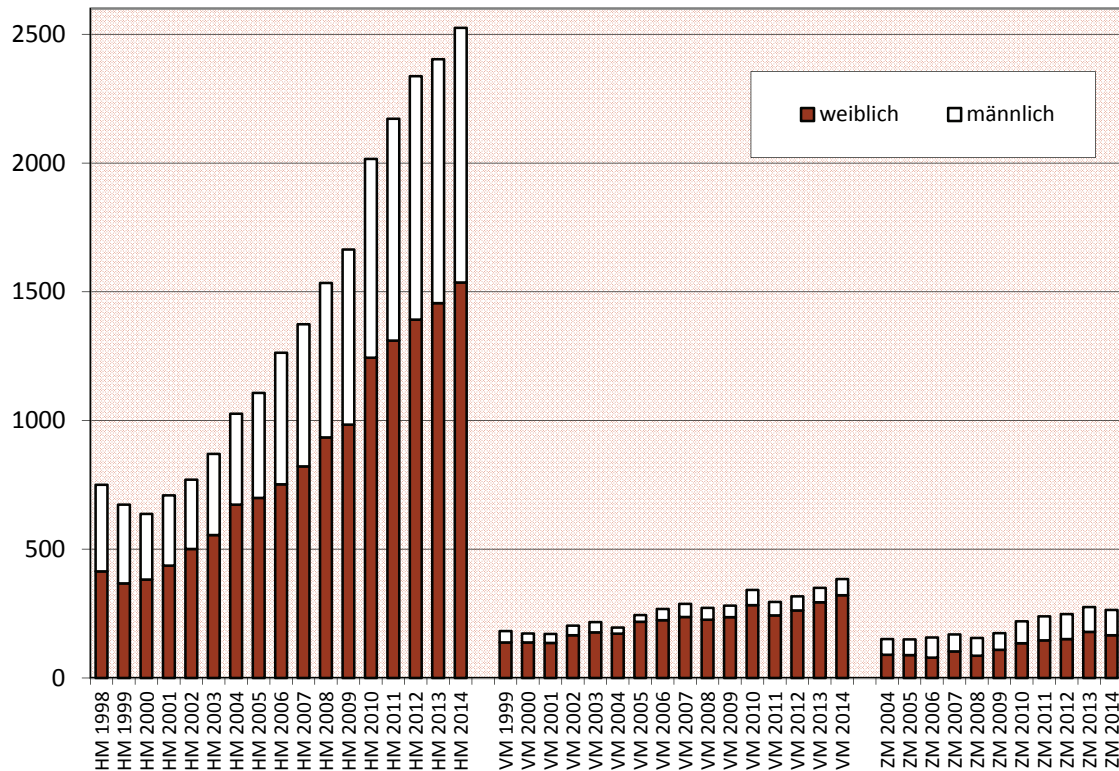


Abbildung 12: Anzahl männlicher und weiblicher Bewerber für Humanmedizin (HM, ab 2008 inklusive Chiropraktik), Veterinärmedizin (VM) und Zahnmedizin (ZM), Jahre mit NC.

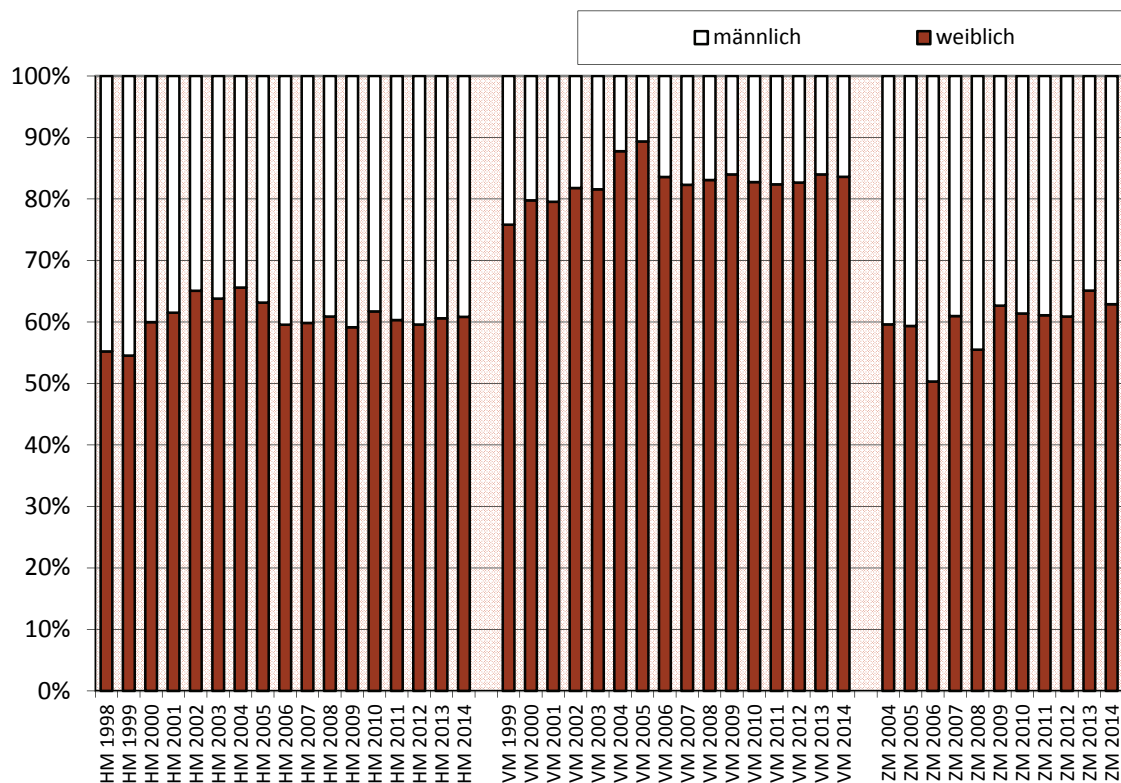


Abbildung 13: Anteil der Geschlechter bei der Bewerbung für Humanmedizin (HM, ab 2008 inklusive Chiropraktik), Veterinärmedizin (VM) und Zahnmedizin (ZM), Jahre mit NC, bezogen auf 100% pro Jahr und Disziplin.

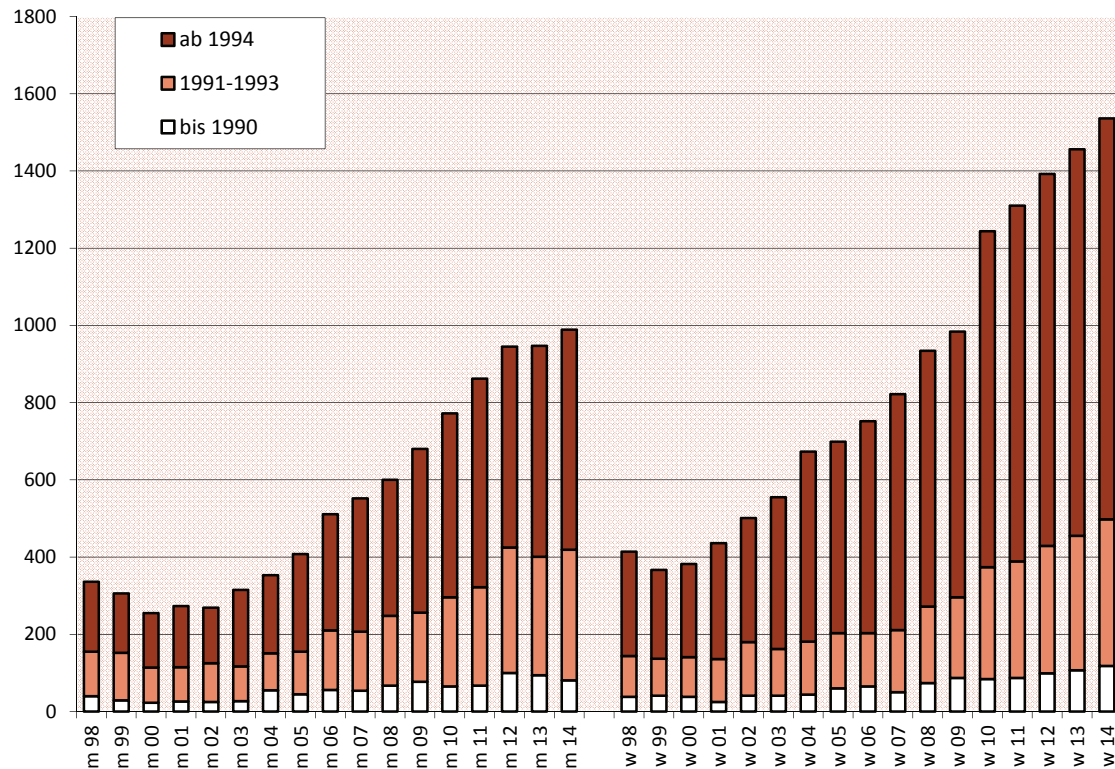


Abbildung 14: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Humanmedizin inkl. Chiropraktik, Absolutzahlen.

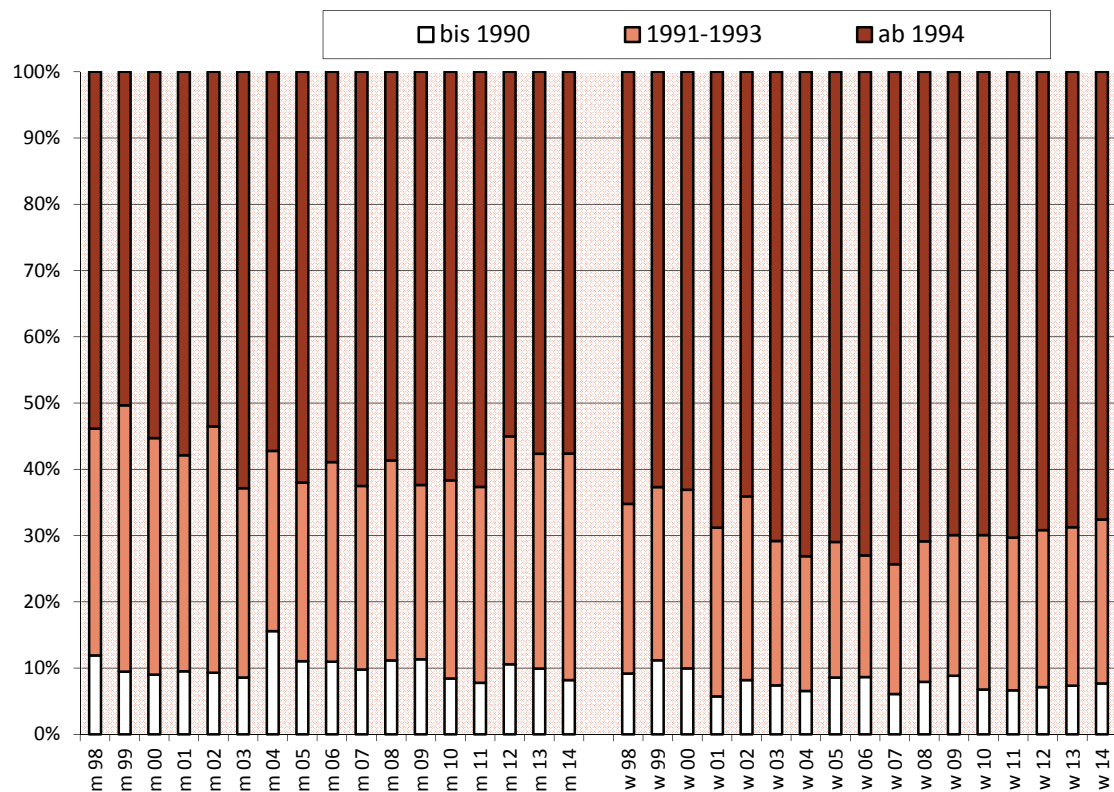


Abbildung 15: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Humanmedizin inkl. Chiropraktik bezogen auf 100% pro Kohorte.

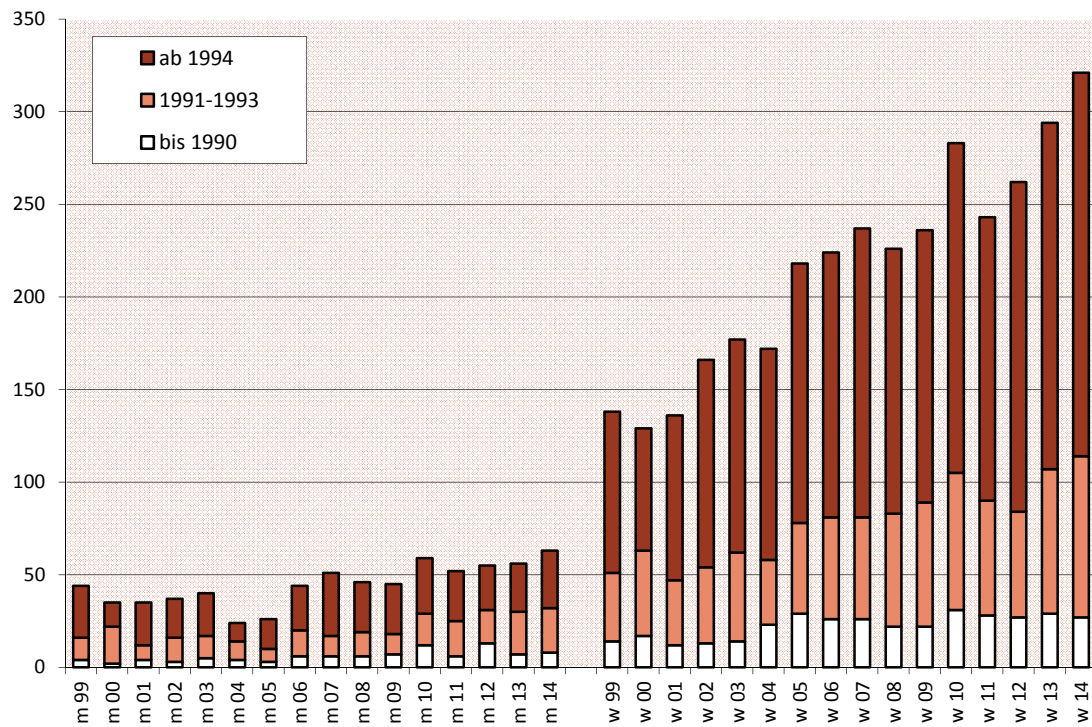


Abbildung 16: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Veterinärmedizin, Absolutzahlen.

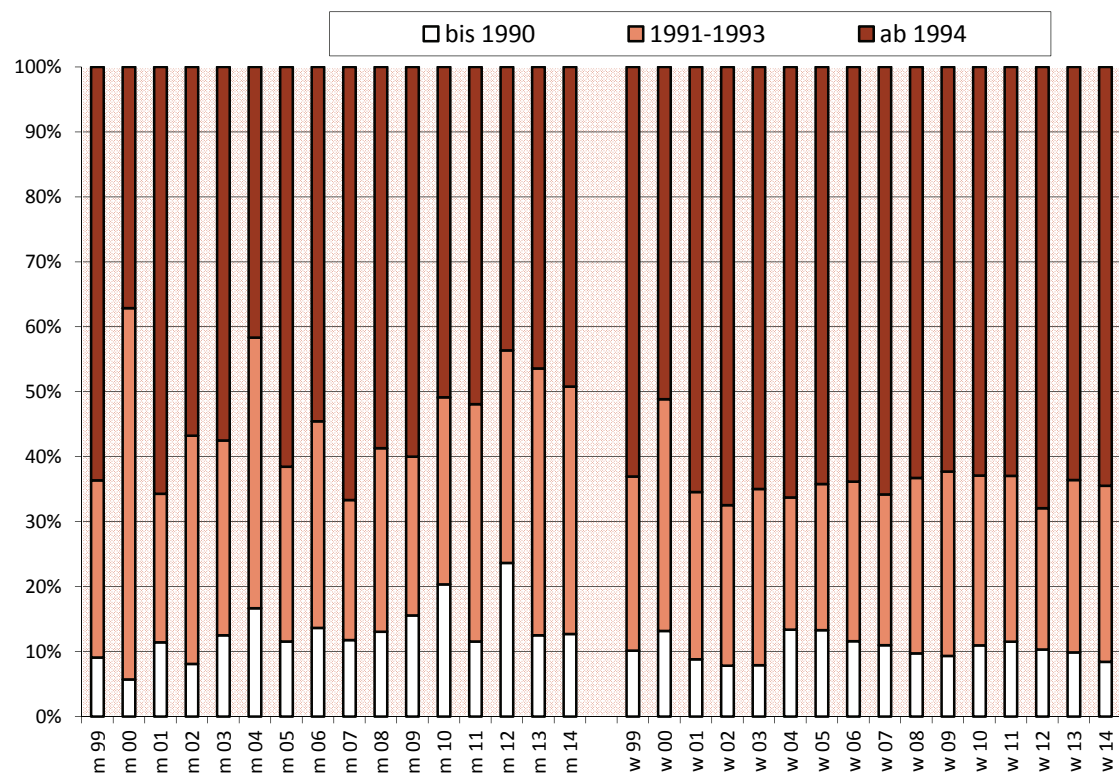


Abbildung 17: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Veterinärmedizin bezogen auf 100% pro Kohorte.

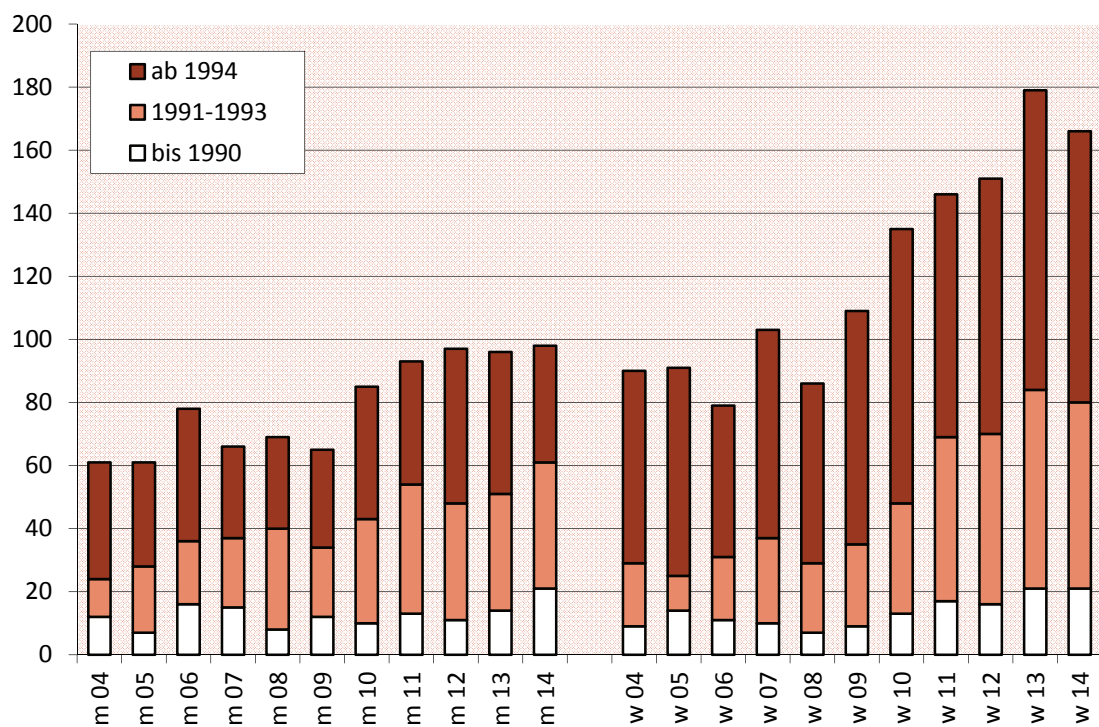


Abbildung 18: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Zahnmedizin, Absolutzahlen.

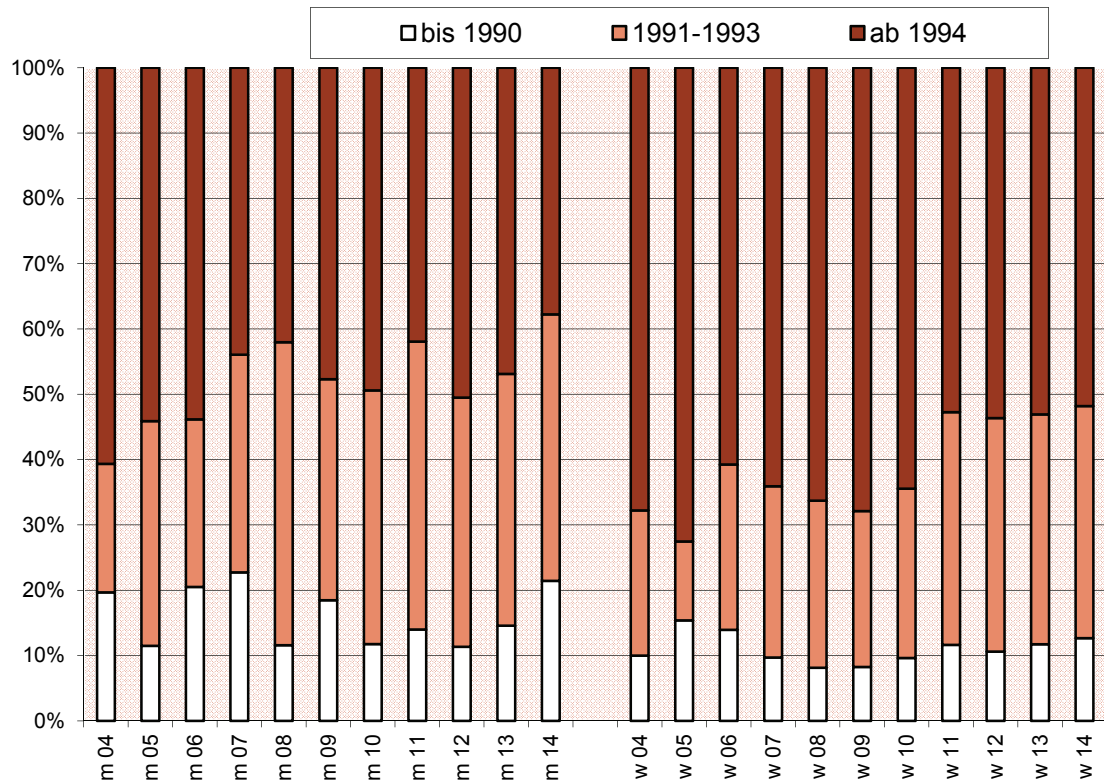


Abbildung 19: Zusammensetzung der Bewerbergruppen nach Geschlecht (m: männlich, w: weiblich) und Alter (Geburtsjahre) für Zahnmedizin bezogen auf 100% pro Kohorte.

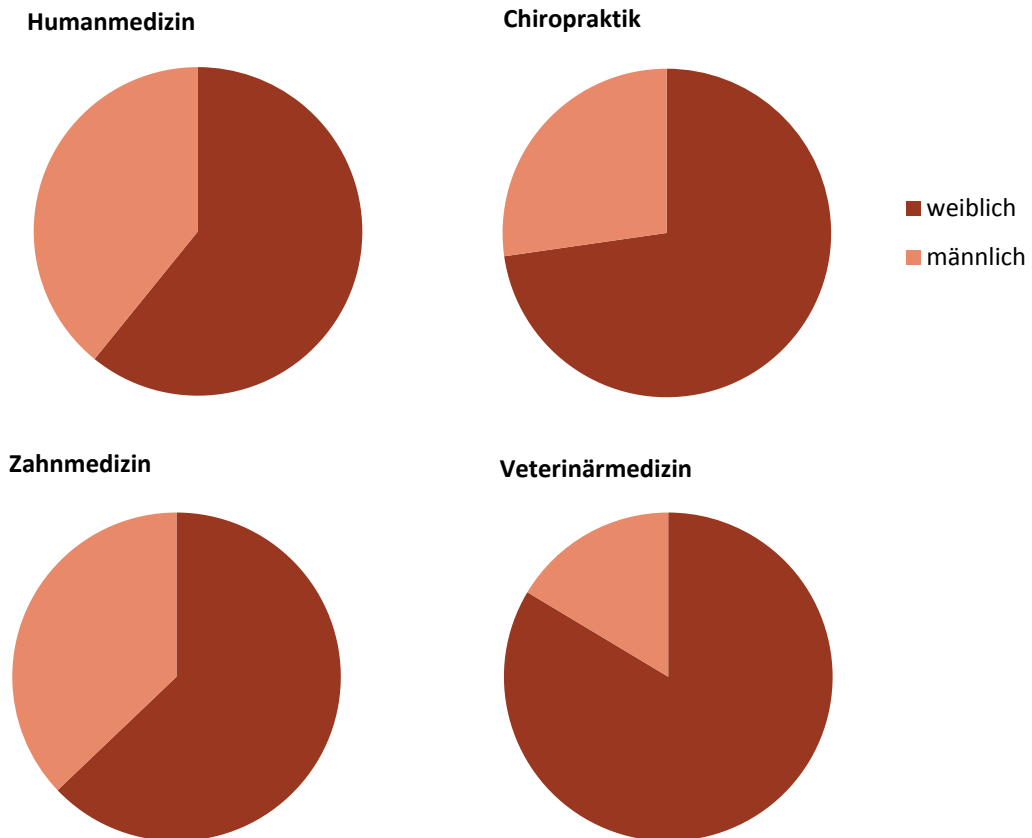


Abbildung 20: Verhältnisse der Geschlechter nach Disziplin im Vergleich 2014.

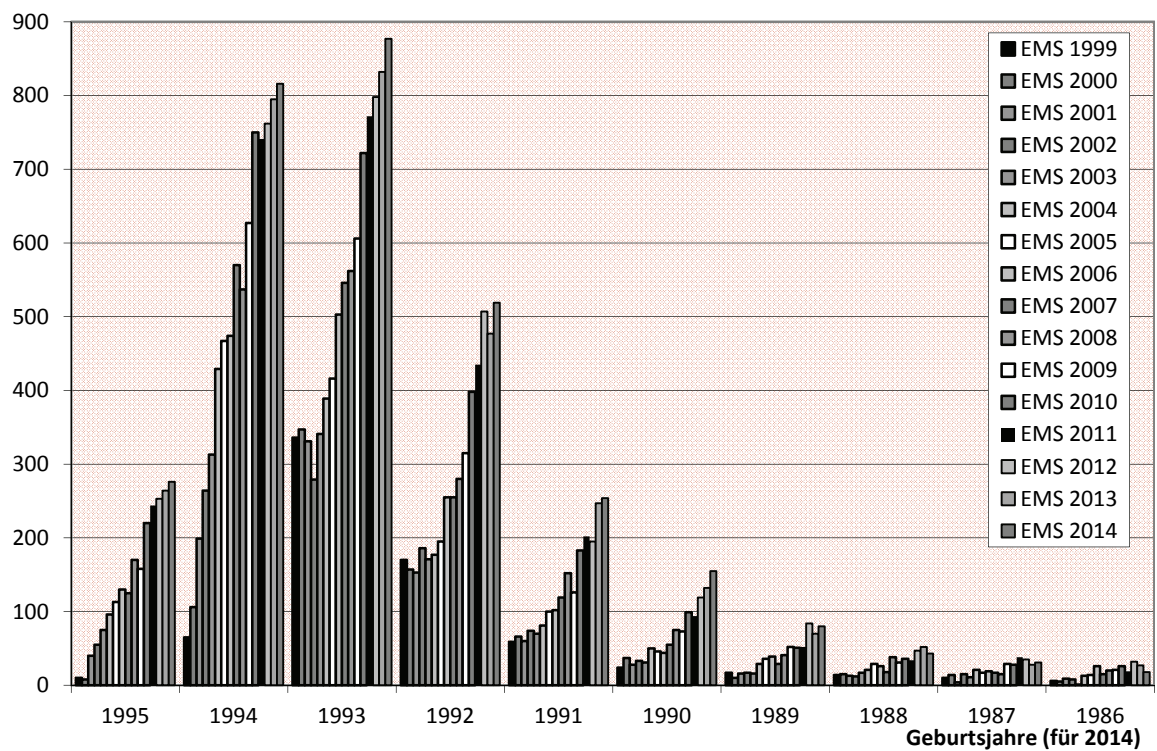


Abbildung 21: Bewerbungen nach Geburtsjahren für 2014 und Entsprechungen für die Jahre 1999 bis 2013 (jeweils um entsprechende Jahre versetzt).

5.9 Übernahme des Testergebnisses aus dem Vorjahr

Bekanntlich kann man das Ergebnis des Vorjahres in das Folgejahr übertragen lassen oder neu zum Test antreten, wobei dann immer das neue Testergebnis zählt. Weil der NC bisher von Jahr zu Jahr strenger wird, kann dabei ein Test-Prozentrang im Folgejahr nicht mehr für die Zulassung ausreichen, obwohl er in diesem Jahr noch genügt. Bei der Entscheidung für diese Variante sollte daher eine gute Reserve vorhanden sein, wenn man sich dafür entscheidet.

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. 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.

In dem Zusammenhang sei noch einmal darauf hingewiesen, dass sich der auf dem Testbescheid mitgeteilte Prozentrang auf alle Teilnehmenden eines Jahrganges, also alle Disziplinen bezieht. Deshalb können diese nur bedingt die Zulassungs-/Ablehnungsquoten pro Disziplin erklären, da zwischen den Disziplinen beträchtliche Unterschiede vorhanden sind und auch die Test-Prozentrangüberträge aus dem Vorjahr noch einzurechnen sind.

Tabelle 15: Test-Prozentrangstatistiken für Personen mit Übernahme des Testergebnisses aus dem Vorjahr.

Disziplin	Jahr	N	Minimum	Maximum	Mittelwert	Stand.-abw.	Zugelassen 1. Runde
Humanmedizin	2010	54	80	121	108.2	6.8	39 (15 nicht)
	2011	65	92	126	111.2	6.0	56 (8 nicht)
	2012	54	101	127	112.1	5.0	49 (5 nicht)
	2013	60	90	129	111.3	6.5	49 (11 nicht)
	2014	42	105	121	112.5	4.1	39 (3 nicht)
Veterinärmedizin	2010	11	101	113	105.1	3.2	11
	2011	16	94	116	103.3	5.5	15 (1 nicht)
	2012	8	99	118	106.0	6.0	8 (0 nicht)
	2013	24	85	116	105.0	6.6	21 (3 nicht)
	2014	17	97	118	106.3	5.4	16 (1 nicht)
Zahnmedizin	2010	37	91	114	100.2	4.4	34 (3 nicht)
	2011	37	93	118	101.3	5.1	33 (4 nicht)
	2012	25	95	108	102.7	2.7	25 (0 nicht)
	2013	39	94	110	102.1	3.3	36 (3 nicht)
	2014	26	94	108	102.9	3.0	25 (1 nicht)

Zugelassen in der 1. Runde sind Personen, die sofort einen Platz zugeteilt erhielten (ohne Nachrücker).

Die Abbildung 22 verdeutlicht, dass beim zweiten Antritt grundsätzlich eine Verbesserung, im Mittel um 20 Punkte möglich ist. Die Abbildung 23 zeigt allerdings, dass abhängig vom im ersten Jahr erzielten Test-Prozentrang differenziertere Unterschiede bei der zweiten Testteilnahme auftreten – was Verbesserungen wie Verschlechterungen betrifft. Wir wissen aus den Befragungen, dass einige Personen sich bei der ersten Testteilnahme wenig informieren und vorbereiten und dadurch niedrigere Werte erreichen. Dieser Personenkreis wird sich in der Regel verbessern. In Bereichen unter dem für Humanmedizin zulassungsrelevanten Bereich (um den Wert 70) sind Verbesserungen immer noch überwiegend, dort gibt es aber auch die meisten Verschlechterungen.

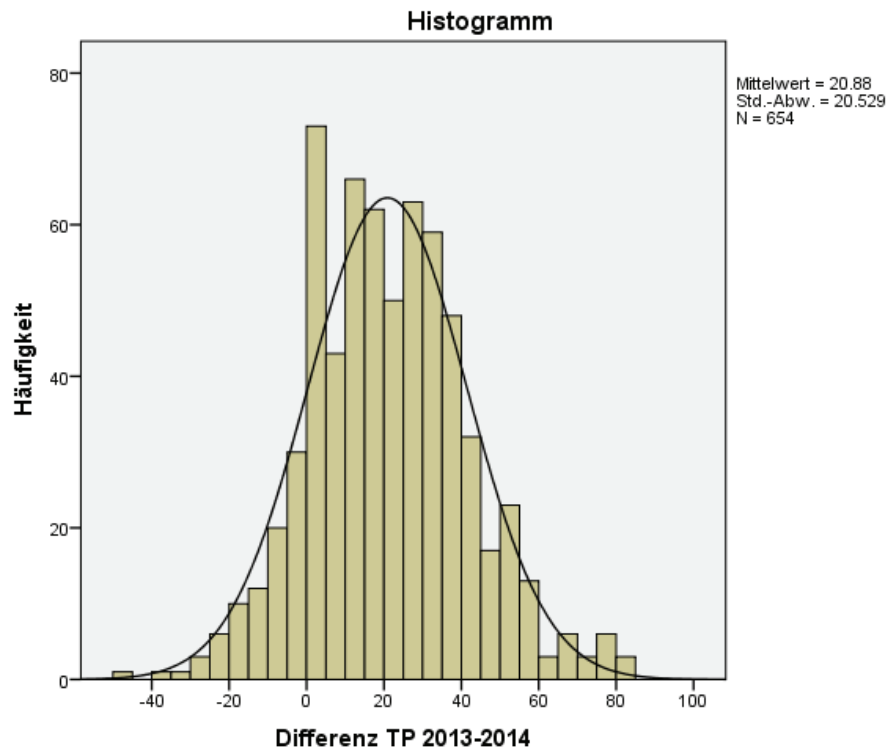


Abbildung 22: Mittlere Differenz beim Test-Prozentrang zweiter minus erster Antritt, positive Werte sind Verbesserungen. Die Korrelation der Test-Prozentränge beider Antritte beträgt 0.61.

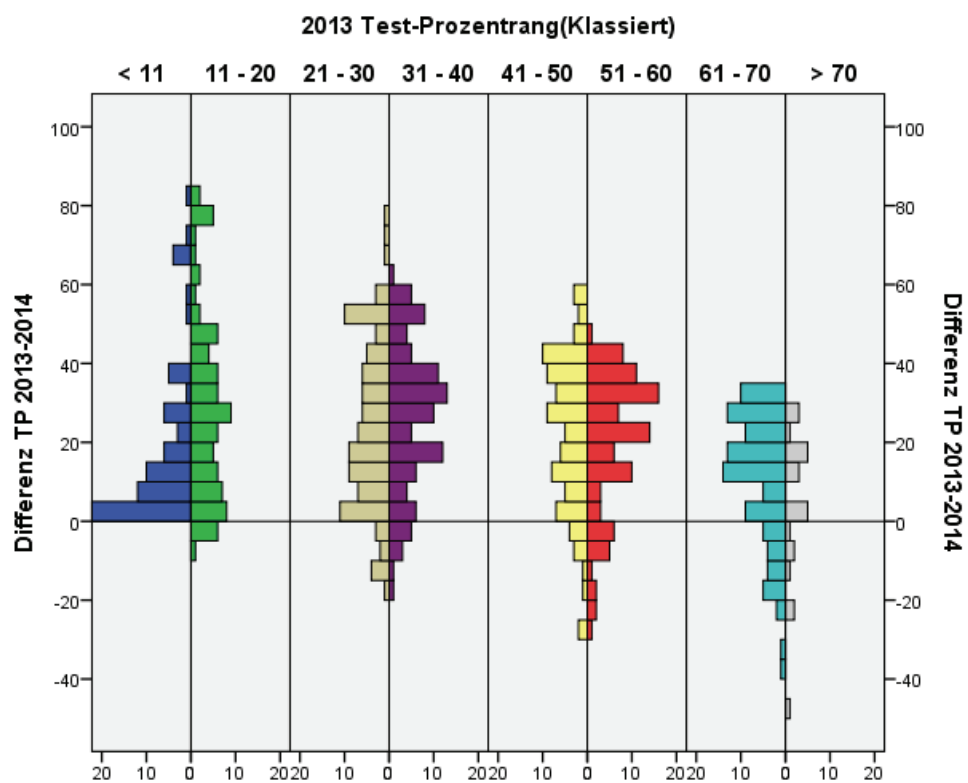


Abbildung 23: Verteilung der Differenzen für Test-Prozentrang 2013 – 2014 (positive Werte sind Verbesserungen) abhängig vom Bereich des Test-Prozentranges bei der ersten Teilnahme 2013.

5.10 Sprachgruppen

Tabelle 16: Sprachgruppen der Teilnehmenden 2014 nach Disziplin, Alter und Geschlecht.

			Geburtsjahre			Total
			ab 1994	1991 - 1993	bis 1990	
Humanmedizin	deutsch	männlich	518	315	77	910
		weiblich	940	349	114	1403
		Total	1458	664	191	2313
	französisch	männlich	24	18	3	45
		weiblich	55	15	2	72
		Total	79	33	5	117
	italienisch	männlich	28	5	1	34
		weiblich	43	16	2	61
		Total	71	21	3	95
... davon Chiropraktik	deutsch	männlich	4	5	5	14
		weiblich	25	9	2	36
		Total	29	14	7	50
	französisch	männlich	2	2	0	4
		weiblich	5	2	0	7
		Total	7	4	0	11
	italienisch	männlich	0	0	0	0
		weiblich	5	0	0	5
		Total	5	0	0	5
Veterinärmedizin	deutsch	männlich	22	15	7	44
		weiblich	129	69	24	222
		Total	151	84	31	266
	französisch	männlich	7	6	1	14
		weiblich	63	16	1	80
		Total	70	22	2	94
	italienisch	männlich	2	3	0	5
		weiblich	15	2	2	19
		Total	17	5	2	24
Zahnmedizin	deutsch	männlich	33	40	21	94
		weiblich	78	59	20	157
		Total	111	99	41	251
	französisch	männlich	2	0	0	2
		weiblich	2	0	1	3
		Total	4	0	1	5
	italienisch	männlich	2	0	0	2
		weiblich	6	0	0	6
		Total	8	0	0	8

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.

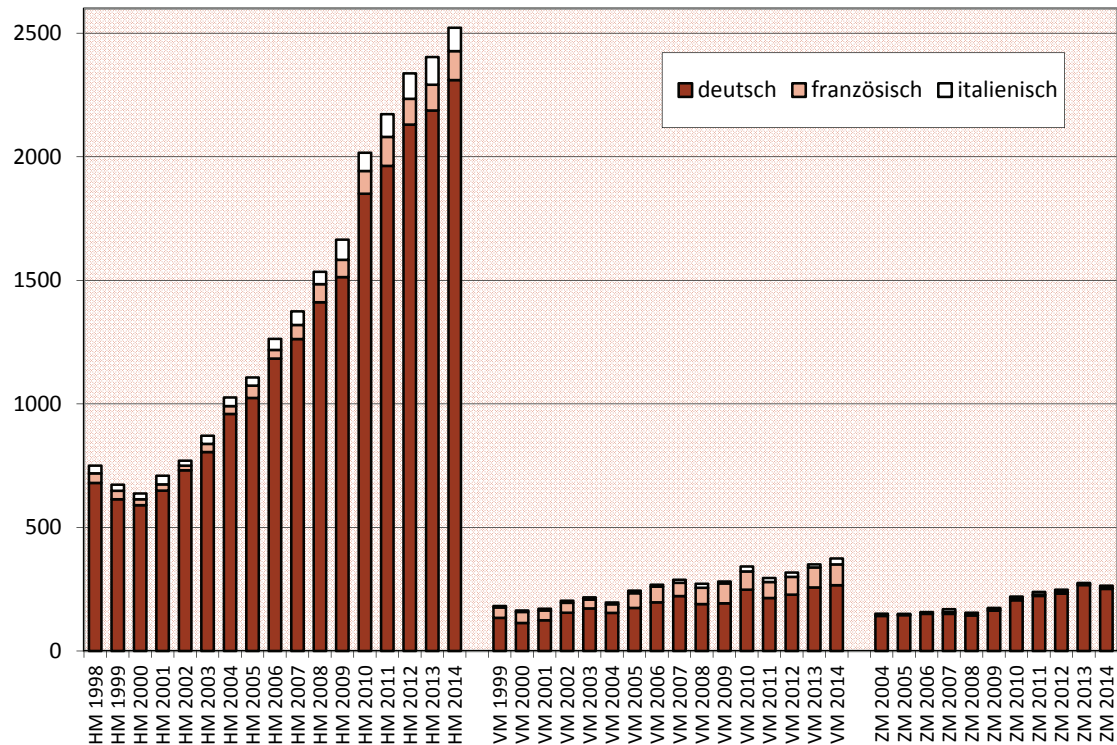


Abbildung 24: Vergleich der Sprachgruppen 1998 bis 2014, Absolutzahlen.

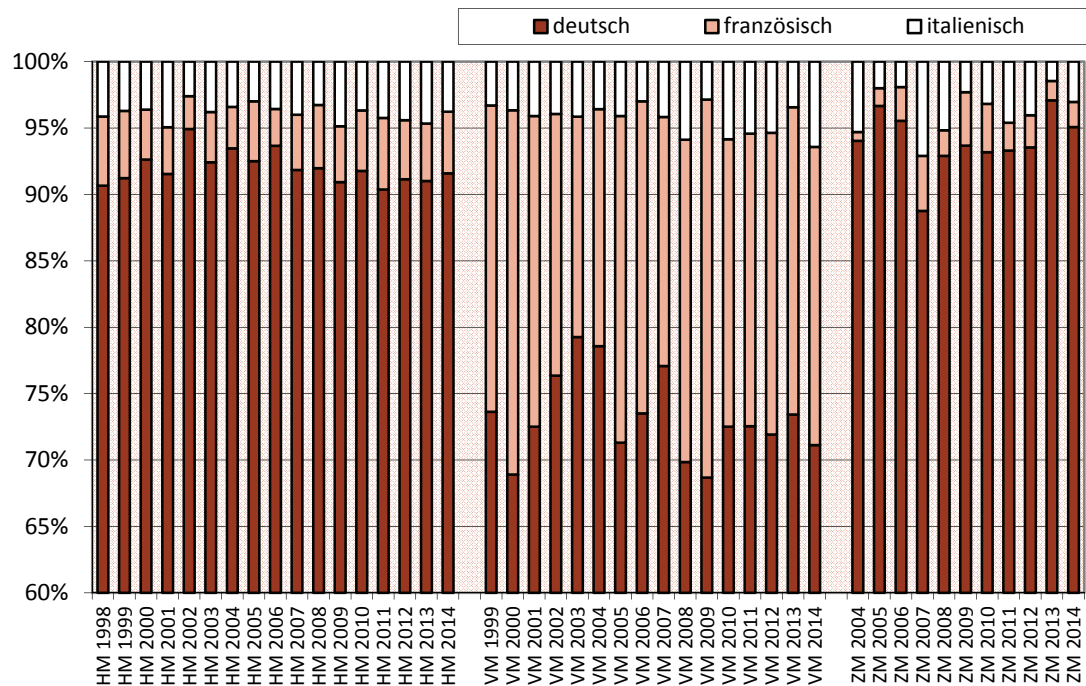


Abbildung 25: Vergleich der Sprachgruppen 1998 bis 2014. Bewerber pro Disziplin und Jahr auf 100% bezogen (dargestellt nur Prozentband zwischen 60 und 100%).

5.11 Alter und Maturitätsjahr

Die Personen werden analog der Vorjahre 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.

Tabelle 17: Geburtsjahr und Maturitätsjahr.

Geb.- jahr	Maturitätsjahr																											
	< 93	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	Ges				
62																				1					1			
63	2																								1			
64			1																				1		2			
67																							1		1			
68	1																								1			
69	1																								1			
70																					1				1			
71	1				1																				2			
72	1																						1		2			
73	1																								1			
75				1																					1			
77						2	1																		3			
78							1																		1			
79								1															1		2			
80								1													1	1	1		4			
81									4		1						1				1		2		9			
82										1	3	1			1			1					2		9			
83										1	2	3				1						1	1		9			
84													4		1				3	3	1	1	2		15			
85												1	4	5			2			2	2		1		17			
86													2	5	4		1				1	3	5		21			
87															7	4		1	1			2	3		18			
88															3	7	3	3	1	2	7		5		31			
89																	6	12	5		5	5	10		43			
90															1		5	11	21	3	11	14	14		80			
91																		10	38	34	22	21	30		155			
92																	1		11	61	60	70	51		254			
93		1						1											1	20	167	233	96		519			
94									1					1		1		1	2		61	520	290		877			
95				1												1			4	1	4	186	619		816			
96																				1	1	12	249		263			
97																							13		13			
Ges.	7	1	1	2	1	2	2	3	5	2	6	5	10	11	17	14	19	39	87	128	345	1069	1398		3173			

Das Geburtsdatum wurde bei Einlass zum Test kontrolliert und ist daher für alle Personen richtig. Einige Personen gaben an, jünger als 18 Jahre zum Zeitpunkt der Maturität gewesen zu sein. Ursache sind wahrscheinlich Fehleingaben bei der elektronischen Anmeldung.

6 Beschreibung des verwendeten Eignungstests

6.1 Aufbau des Tests

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

Die Tabelle 18 gibt einen Überblick über den Testaufbau, die Aufgabenzahl und die Dauer der einzelnen Aufgabengruppen. Aufgabenbeispiele finden Sie im Anhang.

Tabelle 18: Struktur und Ablauf des EMS seit 2005.

Aufgabengruppen	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.
Lernphase zu den Gedächtnistests: Figuren lernen Fakten lernen	Es werden die Vorlagen zum Einprägen gezeigt		4 min 6 min
Medizinisch-naturw. Grundverst.	20	20	50 min
Gedächtnistests: Figuren reproduzieren Fakten reproduzieren	20 20	20 20	5 min 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			

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 Aufgabengruppen 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 eine neue Aufgabengruppe „Planen und Organisieren“ aufgenommen. Diese Anforderung wurde in erster Priorität umgesetzt. Damit die Testlänge vergleichbar bleibt, wird seitdem in der Schweiz auf nicht gewertete Einstreuaufgaben 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. Auf der Grundlage dieser Analysen wurden ca. 50 Bereiche und Aufgabentypen geprüft und die unter vielen Aspekten am besten geeignet erscheinenden 13 für die Erprobungen im Übergangsverfahren des TMS ausgewählt und empirisch erprobt. Von diesen haben vor allem aufgrund der Ergebnisse zur Validität dann 9 Aufgabengruppen Eingang in den TMS gefunden, eine (Planen und Organisieren) kam dann im EMS dazu.

Jedes Jahr wurden neue Aufgaben für die Aufgabengruppen 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 mittels „Einstreuaufgaben“, die mit bearbeitet, aber nicht gewertet wurden. 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.

6.2 Berechnung der Werte 2014

Die CRUS, die für die Anmeldung zum Medizinstudium verantwortlich ist, erhielt am 20. Juli 2014 einen anonymen Brief und eine Anlage mit Trainingsmaterial eines Trainingsanbieters. Moniert wird, dass sich im Trainingsmaterial dieses Anbieters teilweise die gleichen Aufgaben wie im EMS 2014 befinden. Genauere Analysen durch das ZTD, das für die Bereitstellung, die Anwendung und die Auswertung des EMS-Tests verantwortlich ist, haben ergeben, dass 14 Aufgaben im Material des Trainingsanbieters mit 14 Aufgaben des EMS 2014 identisch sind (ein Text „Textverständnis“ mit sechs Fragen, je vier Aufgaben der Aufgabengruppen „Quantitative und formale Probleme“ sowie „Diagramme und Tabellen“). Das sind 7% von insgesamt 198 Aufgaben. Es ist davon auszugehen, dass

die 14 betreffenden Originalaufgaben während einer früheren Testsession in Deutschland beim alten TMS 1996 gestohlen worden sind.

Die rechtlichen Grundlagen des Numerus clausus schreiben die Messung der Eignung unter gleichen Bedingungen für alle Kandidatinnen und Kandidaten vor. Ist diese Gleichheit verletzt, muss sie durch ein geeignetes Mittel und in einem transparenten Verfahren wiederhergestellt werden. Aus diesem Grund haben die SUK, die CRUS und das ZTD beschlossen, dass die 14 bereits vorher bekannten Aufgaben aus der Test-Bewertung ausgeschlossen werden. Dies wird als bestmögliches Korrekturverfahren erachtet: Die Bedingungen sind dadurch bei den gewerteten Aufgaben hinsichtlich des Vorwissens für alle Kandidatinnen und Kandidaten gleich; die Aufgaben, die einigen Kandidierenden schon bekannt waren, wurden eliminiert. Die verkürzte Fassung ist somit die gültige Version des EMS 2014.

Ein etwaiger Vorteil derjenigen Personen, die die gestohlenen Aufgaben kannten, durch einen Zeitgewinn lässt sich weder theoretisch begründen – die Bearbeitungszeit steht bei Fähigkeitstests in keiner linearen Beziehung zur Leistung – noch empirisch anhand der Daten nachweisen. Es ist auch keine Ungerechtigkeit gegenüber jenen Personen entstanden, die theoretisch auf der Basis der Ergebnisse ohne den Ausschluss der bekannt gewordenen Aufgaben zugelassen worden wären. Geeignete Kandidierende beantworten üblicherweise alle Fragen richtig, nicht nur die vorher bekannten. Insofern sind durch die Korrektur die Bedingungen für alle Teilnehmenden und damit die Gleichbehandlung wiederhergestellt worden.

Die 10 Aufgabengruppen des Eignungstests für das Medizinstudium (EMS) ausser dem „Konzentrierten und sorgfältigen Arbeiten“ liefern Punkte als Summe richtig gelöster Aufgaben – zwischen 0 und 20 (bzw. 18 für die Aufgabengruppe „Textverständnis“).

Aufgabengruppe	Punkte (Bereich)	Nicht vergleichbar zwischen Jahrgängen des EMS	Vergleichbar zwischen verschiedenen Jahrgängen des EMS
Quantitative u. formale Probleme	0 bis 16*	PUNKTwert als Summe 0 bis 184*	Test-Prozentrang (TP) (1 – 100) Ganzzahlige kumulierte prozentuale Häufigkeit des jeweiligen Punktwertes über alle Teilnehmer Wenn notwendig zur weiteren Differenzierung:
Schlauchfiguren	0 bis 20		
Textverständnis	0 bis 12*		
Planen und Organisieren	0 bis 20		
Konzentriertes und sorgfältiges Arbeiten	0 bis 20		
Medizinisch-naturwissenschaftliches Grundverständnis	0 bis 20		
Figuren lernen	0 bis 20	Rangplätze der Punktwerte für jede Aufgabengruppe (zwischen 1 und der Maximalzahl der Personen)	Mittlerer Rangplatz (MR) Ganzzahliger Mittelwert der Rangplätze aller 10 Aufgabengruppen und Bezug auf 1000 Personen
Fakten lernen	0 bis 20		
Muster zuordnen	0 bis 20		
Diagramme und Tabellen	0 bis 16*		

Abbildung 26: Punktwerte der einzelnen Aufgabengruppen 2014 und ihre Zusammenführung über den Punktwert zum Test-Prozentrang und Prozentrang. *: 14 Aufgaben wurden nicht gewertet.

Für die Aufgabengruppe „Konzentriertes und sorgfältiges Arbeiten“ müssen insgesamt 1600 Zeichen der Reihe nach bearbeitet werden – davon sind 400 Zielzeichen zu markieren. Es können in der zur Verfügung stehenden Zeit in der Regel nicht alle Zeichen bearbeitet werden. Die Position des letzten markierten Zeichens bestimmt, wie viele Zeichen als bearbeitet gewertet werden. Alle übersehenen und fälschlich markierten Zeichen vor diesem letzten bearbeiteten Zeichen zählen als Fehler und

werden von der Menge der korrekt markierten Zeichen abgezogen. Der resultierende Wert ist der „Rohwert“, der in eine Skala zwischen 0 und 20 transformiert wird, um mit den anderen Aufgabengruppen gleichgewichtig zum Punktwert addiert zu werden. Die schlechtesten 2.5% erhalten den Punktwert 0, die besten 2.5% erhalten den Punktwert 20. Der Wertebereich des Rohwertes für die verbleibenden 95% wird in 19 äquidistante Abschnitte eingeteilt und den Punktwerten 1 - 19 zugeordnet. Obwohl sich die Testanforderung zwischen den Jahren ändert, resultieren so vergleichbare und annähernd normalverteilte Punktwertverteilungen. Alle Punkte der Aufgabengruppen werden zu einer Summe addiert (**Punktwert**). Dieser Wert hat den Nachteil, dass er nicht zwischen Tests verschiedener Jahre vergleichbar ist (diese Vergleichbarkeit ist eine Voraussetzung, dass Ergebnisse ins Folgejahr übertragen werden können).

Aus der Gesamtpunktzahl (Punktwert) werden zwei Werte berechnet:

- **der Test-Prozentrang (TP) und**
- **der mittlere Rangplatz der Aufgabengruppen (MR).**

Der Test-Prozentrang (TP) zeigt an, wie viele Prozent aller teilnehmenden Personen ein schlechteres oder gleich gutes Ergebnis wie Sie erreicht haben. Die Differenz zu 100 zeigt an, wie viele Prozent der Personen ein besseres Testergebnis als Sie erreicht haben.

Er wird direkt aus der kumulierten Häufigkeits-Verteilung der Punktwerte aller Personen berechnet, die in einem Jahr am EMS-Test teilnahmen und ist ganzzahlig. Werte unter 10 werden auf 10 gerundet.

Es kann vorkommen, dass durch die bisher kontinuierlich zunehmenden Anmeldungszahlen ein für die Zulassung ausreichender Wert im Folgejahr nicht mehr ausreichend ist. Dabei ist schwer vorherzusagen, wie sich die Kapazitäten und Anmeldezahlen für das Folgejahr entwickeln werden. Es sollte eine ausreichende Reserve gegenüber dem für die Zulassung im Jahr der Testabsolvierung ausreichenden Wert einkalkuliert werden.

6.3 Mittlerer Rangplatz der Aufgabengruppen

Nur bei gleichem Test-Prozentrang werden die Personen in der Reihenfolge des mittleren Rangplatzes aller Aufgabengruppen (RP) berücksichtigt, wenn nicht mehr alle zugelassen werden können.

Diese Rangplätze werden dann über alle 10 Aufgabengruppen gemittelt. Damit diese mittleren Rangplätze ebenfalls zwischen den Jahren vergleichbar sind, werden sie auf 1000 Personen bezogen umgerechnet. Niedrige mittlere Rangplätze entsprechen jeweils den besseren Leistungen. 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 einer schwierigen Aufgabengruppe stärker, indem dort niedrigere Rangplätze für die besten Leistungen resultieren.

Werden in 2 Aufgabengruppen maximal 20 Punkte erzielt, gehen immer 40 Punkte in den Punktwert ein. Haben in einer Aufgabengruppe 100 Personen diesen Wert erreicht, erhalten diese Personen den mittleren Rangplatz 50.5. Haben in einer anderen Aufgabengruppe nur 10 Personen diesen Wert erreicht, ist der Rangplatz 5.5. Dies wird besser bewertet, weil die 20 Punkte der zweiten Aufgabengruppe schwerer zu erzielen waren, deshalb wertvoller sind.

Zur Herstellung der Vergleichbarkeit der Rangplätze zwischen den Jahren wird der mittlere Rangplatz auf die Basis von 1000 Personen umgerechnet.

- Der Mittlere Rangplatz MR bewegt sich z.B. 2012 zwischen 56 und 972. Niedrige Werte stehen für bessere Leistungen.

- Er korreliert mit dem Test-Prozentrang TP mit 0.98 (trotz der Anhebung aller Werte unter 10 auf 10). Dies entspricht der Korrelation über die Jahre zwischen dem früher verwendeten Testwert und MR zwischen -0.98 und -0.99 was praktisch für eine Äquivalenz spricht.

Beispielrechnung

100 Personen nehmen an einem Test mit 2 Aufgabengruppen teil, in denen man je maximal 20 Punkte erreichen kann.

Aufgabengruppe 1:

- Hier erzielen 10 Kandidaten die maximale Punktzahl 20. Bei gleicher Punktzahl erhält jede Person den durchschnittlichen Rang dieser Personen (die 1 bis 10 belegen).
- Kandidat A hat diese maximale Punktzahl erreicht und belegt damit in dieser Aufgabengruppe zusammen mit den anderen 9 Kandidaten Rang 5.5 (Durchschnitt der Ränge 1 bis 10, also $(1+10)/2=5.5$).
- Kandidat B hat, zusammen mit 14 weiteren Kandidaten, 19 Punkte erzielt. Diese Kandidaten erhalten in dieser Aufgabengruppe den Rang 18 (Durchschnitt der Ränge 11 bis 25, also $(11+25)/2=18$).

Aufgabengruppe 2:

- In dieser Aufgabengruppe erreicht nur Kandidat B die volle Punktzahl von 20 Punkten. Er erhält als einziger den Rang 1.
- Kandidat A erreicht hier 19 Punkte, zusammen mit nur 3 weiteren Kandidaten. Sie alle erhalten in dieser Aufgabengruppe den durchschnittlichen Rang $(2+5)/2=3.5$.

Beide Kandidaten erreichten damit die gleiche Punktzahl von 39 Punkten (und den gleichen Test-Prozentrang).

Der Mittlere Rangplatz für die beiden Kandidaten berechnet sich wie folgt:

- Kandidat A: $(5.5 + 3.5)/2=4.5$
- Kandidat B: $(18 + 1)/2=9.5$

Und auf 1000 Personen:

- Kandidat A: $4.5/100 \times 1000=45$
- Kandidat B: $9.5/100 \times 1000=95$

Hätten 1000 Personen am Test teilgenommen, resultiert für Kandidat A im Mittel über beide Aufgabengruppen der 45. Rang, für Kandidat B der 95. Rang. Kandidat A wird dafür belohnt, in der ersten Aufgabengruppe die maximale, aber vor allem auch in der zweiten Aufgabengruppe zusammen mit nur sehr wenigen Kandidaten eine hohe Punktzahl erreicht zu haben.

Kandidat B erhält durch die geringere Punktzahl in der ersten Aufgabengruppe (in dem viele Personen das gleiche oder höhere Ergebnis erzielt haben) einen höheren (schlechteren) Rang.

Der Unterschied zwischen den Kandidaten ist dadurch gerechtfertigt, dass Kandidat A trotz gleicher Punktzahl über beide Aufgabengruppen mehr Kandidaten übertroffen hat als Kandidat B.

Müsste man zwischen beiden Personen unterscheiden, wäre Kandidat A der bessere.

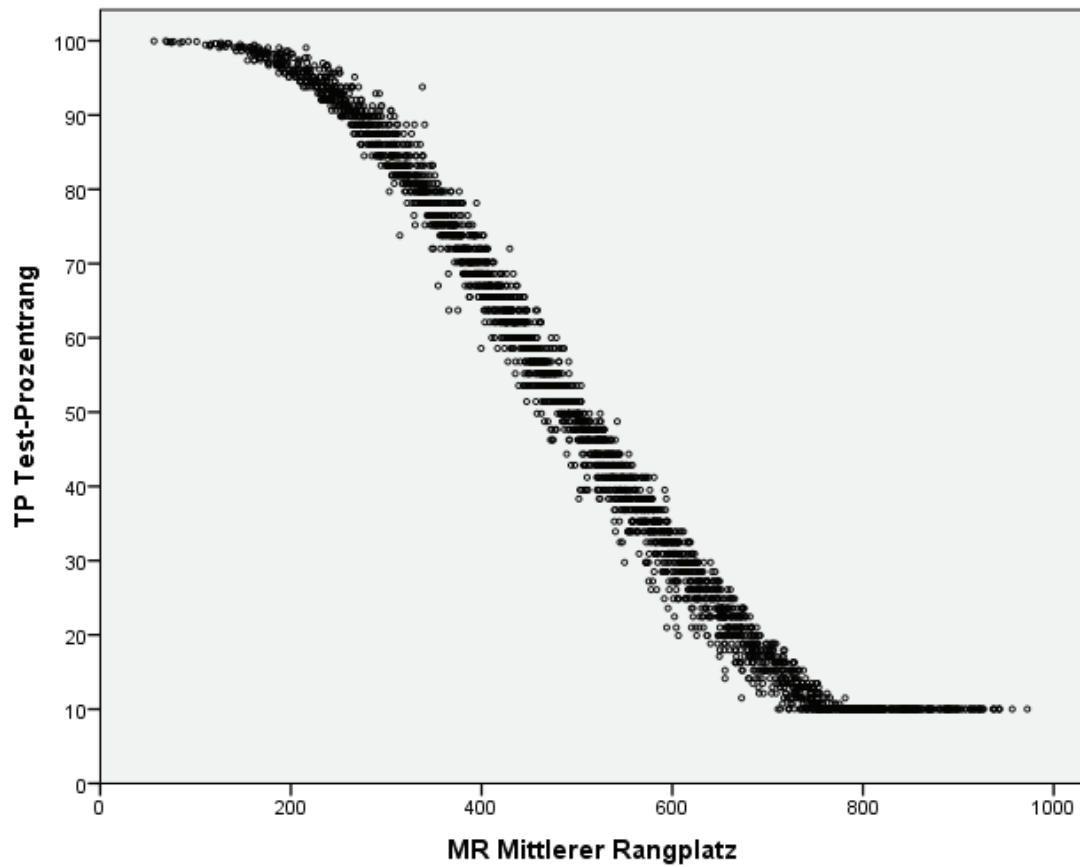


Abbildung 27: Beziehung zwischen Test-Prozentrang und mittlerem Rangplatz für 2014 (Werte kleiner 10 bei TP auf 10 angehoben).

7 Testanwendung in der Schweiz 2014

7.1 Verteilungsprüfung

Die Verteilung des Punktwerts entspricht annähernd einer Normalverteilung. Metrische Prüfverfahren sind daher bedingt aussagefähig. Die Punktwerte der einzelnen Aufgabengruppen sind wie in allen Vorjahren statistisch nicht als normalverteilt anzusehen (Tabelle 19).

Tabelle 19: Kennwerte der Punktwertskalen Gesamt und Aufgabengruppen

	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	113.2	9.8	14.6	6.3	11.3	11.0	14.1	14.7	12.0	8.6	10.7
Median	115.0	10.0	15.0	6.0	11.0	11.0	14.0	15.0	12.0	9.0	11.0
Modalwert	126	11	16	6	11	11	16	20	12	8	10
Stand.-abweich.	21.6	3.1	3.5	2.6	3.5	3.5	3.6	4.1	3.2	3.0	4.5
Spannweite	138	15	18	12	19	19	19	19	20	16	20
25. Perz.	99	8	12	4	9	8	12	12	10	6	8
75. Perz.	129	12	17	8	14	14	17	18	14	11	14
Schiefte	-.36	-.23	-.57	.11	-.17	-.01	-.45	-.48	-.21	-.03	-.17
Kurtosis	-0.20	-0.47	-0.21	-0.60	-0.48	-0.48	-0.23	-0.60	-0.05	-0.45	-0.31

Zu den leichteren Aufgabengruppen gehören „Schlauchfiguren“, „Figuren lernen“ und „Fakten lernen“, was sich mit den langjährigen Ergebnissen deckt. „Konzentriertes und sorgfältiges Arbeiten“ liegt mit 10.7 im Zielbereich einer optimalen Differenzierung. Bei „Quantitative und formale Probleme“ sowie „Diagramme und Tabellen“ liegt durch die Nichtwertung von Aufgaben das Maximum bei 16, bei „Textverständnis“ ist das diesjährige Maximum 12.

7.2 Äquivalenz der Sprachversionen

Sprachrelevante Aufgaben wurden 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.

Die 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.

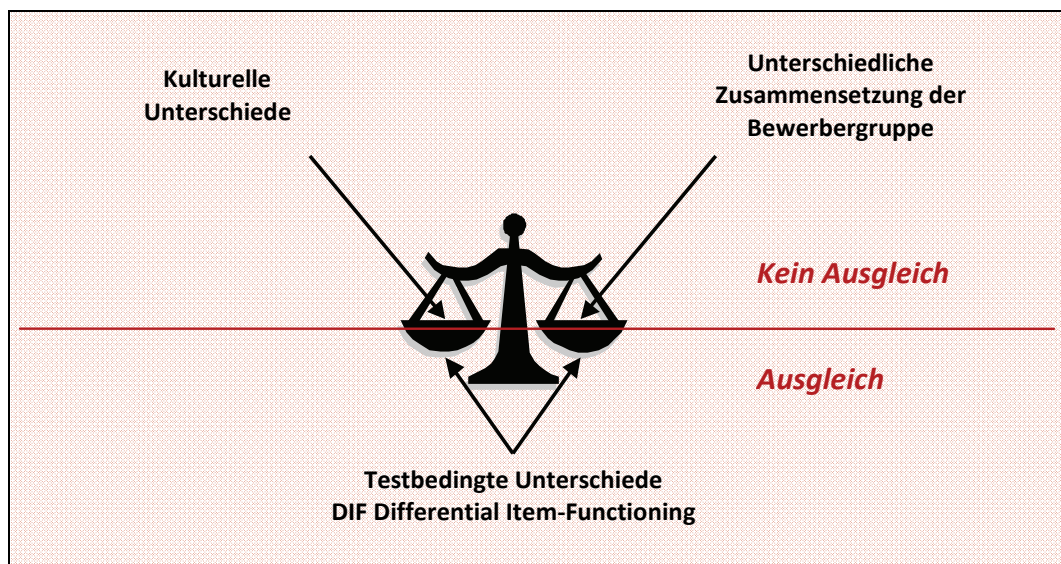


Abbildung 28: Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen.

Zwei Ursachen (Kulturelle Unterschiede und die unterschiedliche Zusammensetzung der Bewerbergruppe) 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 hingegen werden durch die Adaptation verursacht und sollen möglichst vollständig ausgeglichen werden. Mögliche Ursache wäre die 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 Leistungsvoraussetzungen (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 einer Aufgabe 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 Aufgabengruppen. Der Sprachausgleich beschränkt sich demnach auf die stärker sprachabhängigen Aufgabengruppen „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 in der entsprechenden Aufgabengruppe vorhanden sind. Geprüft werden dazu die Abweichungen der jeweiligen Zielsprache von der deutschen Sprachgruppe.

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

Spricht man von DIF, so muss dieses von systematischen Mittelwertsdifferenzen, welche alle Aufgaben einer Gruppe gleichermassen betreffen, abgegrenzt werden. Die Ergebnisse der letzten Jahre haben gezeigt, dass sich die Sprachgruppen in manchen Aufgabengruppen signifikant unterscheiden. Differenzen kommen nicht nur durch einzelne Aufgaben zustande, sondern auch durch eine Verschiebung der Schwierigkeiten aller Aufgaben einer Gruppe. Es ist unwahrscheinlich, dass diese systematischen Unterschiede testbedingt sind. Testbedingte Unterschiede als sprachliche Besonderheiten zeigen sich in spezifischen Abweichungen einzelner Aufgaben. 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.

7.2.1 Sprachvergleich für die Aufgabengruppen

Auch 2014 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Aufgabengruppen auf. Die Punktzahlen der deutschsprachigen Gruppe liegen in der Regel über jenen der anderen Sprachgruppen. 2014 erzielt allerdings die italienische Sprachgruppe in einigen Aufgabengruppen gleichwertige (Textverständnis, Figuren lernen, Konzentriertes und sorgfältiges Arbeiten, Muster zuordnen) oder höhere (Medizinisch-naturwissenschaftliches Grundverständnis) Ergebnisse als die deutsche.

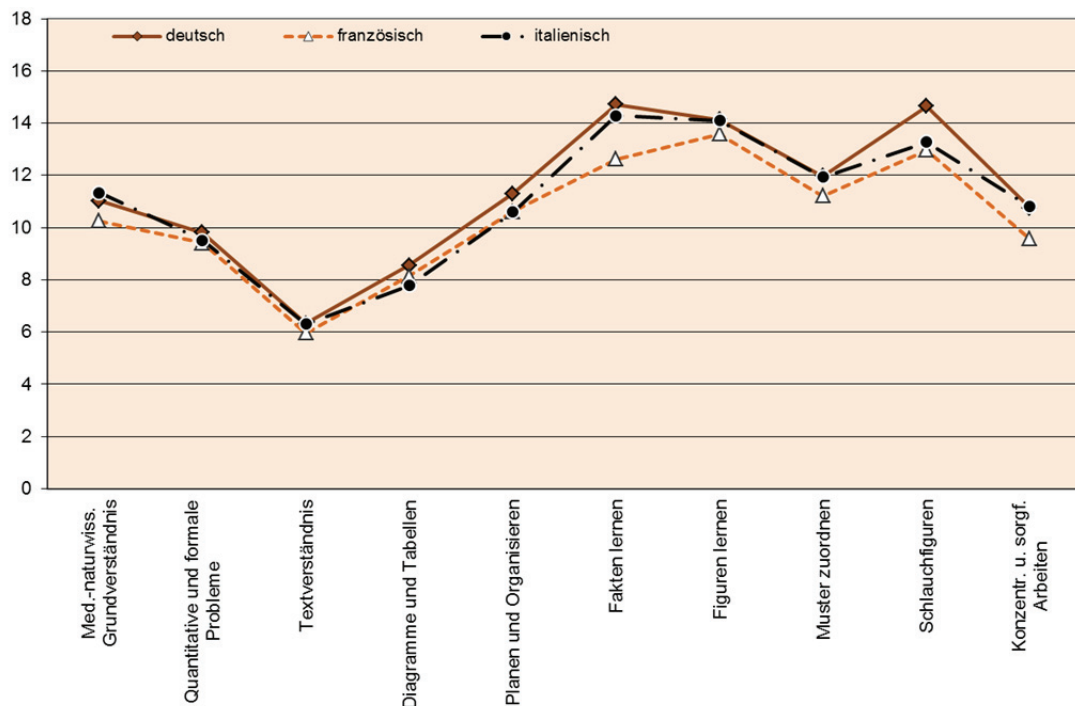


Abbildung 29: Mittelwerte der Punktwerte für die Aufgabengruppen nach Sprachen (unkorrigiert).

Abbildung 30 zeigt den Vergleich zwischen deutsch- und französischsprachigen Testteilnehmern für das Jahr 2014, in Abbildung 31 ist der Vergleich zwischen deutsch- und italienischsprachigen Personen dargestellt. Die senkrechten Linien bezeichnen Mittelwert und Standardabweichung der Differenzen über alle Jahrgänge seit 1998. Höhere Differenzwerte entsprechen besseren Ergebnissen der deutschen Sprachgruppe.

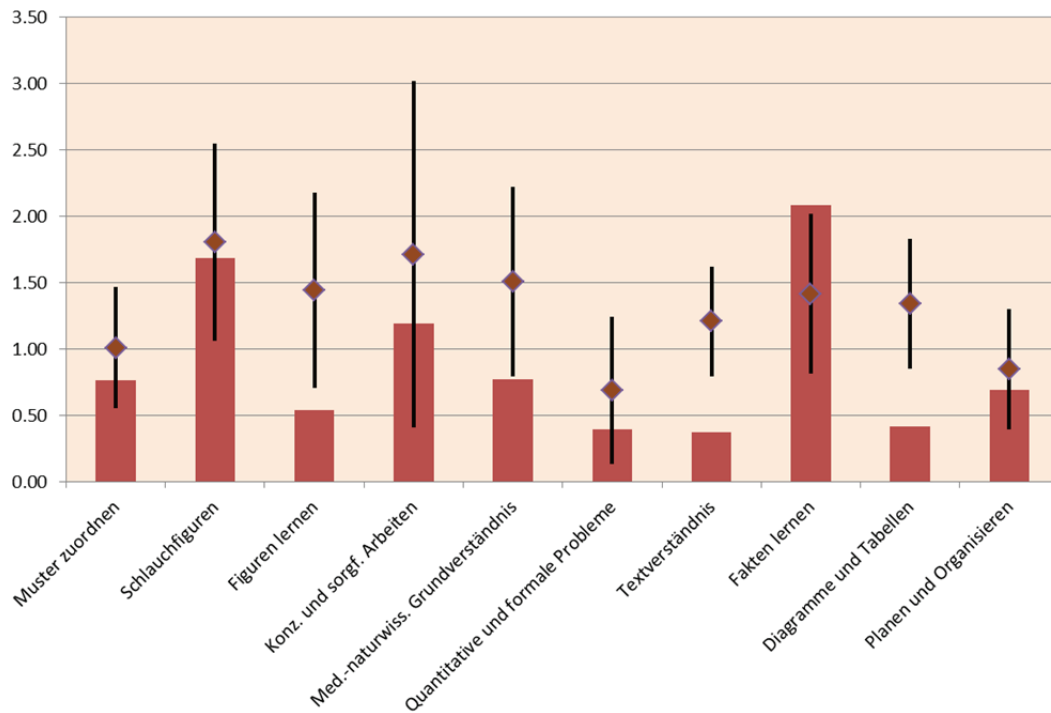


Abbildung 30: Differenz für Aufgabengruppen deutschsprachig-französischsprachig 2014 (unkorrigiert) und Mittelwert mit Standardabweichung aller vorhergehenden Testdurchführungen 1998 – 2013.

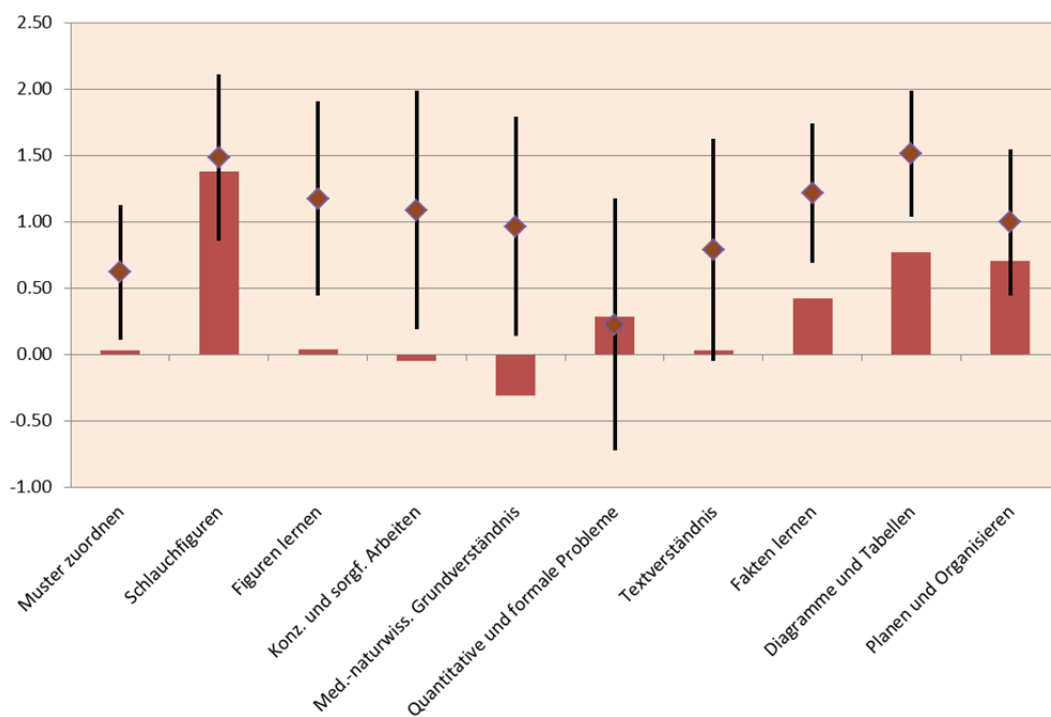


Abbildung 31: Differenz für Aufgabengruppen deutschsprachig-italienischsprachig 2014 (unkorrigiert) und Mittelwert mit Standardabweichung aller vorhergehenden Testdurchführungen 1998 – 2013.

7.2.2 Darstellung des Korrekturverfahrens

7.2.2.1 Identifikation und Ausgleich testbedingter Unterschiede geschehen wie folgt:

- Sprachausgleiche werden nur in den sprachabhängigen Aufgabengruppen „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.
- Damit ein Sprachausgleich in einer bestimmten sprachabhängigen Aufgabengruppe 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 die Aufgabengruppe auch nicht zu Mittelwertunterschieden beim Test-Prozentrang bei. Eine Korrektur allfälliger DIF-Aufgaben (die vorhanden sein können) würde dann eventuell zusätzliche Unterschiede produzieren.
- Es werden jene Aufgaben ausgeglichen, welche über eine systematische Differenz hinaus DIF aufweisen. Für diese Aufgaben 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.
- Als Sprachausgleich wird den französisch- und italienischsprachigen Probanden die Differenz zur entsprechenden Aufgabenschwierigkeit in der deutschsprachigen Version gutgeschrieben, falls sie die Aufgabe 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-Aufgaben mitbestimmt. Man kann das so interpretieren, dass für DIF-Aufgaben bei falscher Antwort zumindest der Schwierigkeitsunterschied zur Referenzgruppe ausgeglichen wird.

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

7.2.2.2 Welche Aufgabengruppen ausgleichen?

Die folgenden Tabellen zeigen die Sprachvergleiche für die sprachabhängigen Aufgabengruppen. Referenz ist die deutschsprachige Testversion. Die französischsprachigen Teilnehmer (Tabelle 20) erzielten in fünf der sechs sprachabhängigen Aufgabengruppen signifikant geringere Punktwerte als die deutschsprachigen Teilnehmer. Bei den italienischsprachigen Kandidaten (Tabelle 22) sind zwei Aufgabengruppen auszugleichen.

Tabelle 20: Signifikanzprüfung deutsche vs. französische Sprachform (sprachabhängige Aufgabengruppen).

Aufgabengruppe	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	2830	11.0	3.54	.001	Ja
	f	216	10.3	3.12		
Quantitative und form. Probleme	d	2830	9.8	3.12	.072	Nein
	f	216	9.4	3.05		
Textverständnis	d	2830	6.3	2.60	.042	Ja
	f	216	6.0	2.45		
Fakten lernen	d	2830	14.7	4.09	.000	Ja
	f	216	12.6	3.84		
Diagramme und Tabellen	d	2830	8.6	2.97	.036	Ja
	f	216	8.1	2.73		
Planen und Organisieren	d	2830	11.3	3.49	.005	Ja
	f	216	10.6	3.32		

Tabelle 21: Signifikanzprüfung deutsche vs. italienische Sprachform (sprachabhängige Aufgabengruppen).

Aufgabengruppe	Sprache	n	m	s	Sig.	Ausgleich
Med.-naturwiss. Grundverständnis	d	2830	11.0	3.54	.332	Nein
	i	127	11.3	3.19		
Quantitative und form. Probleme	d	2830	9.8	3.12	.316	Nein
	i	127	9.5	3.03		
Textverständnis	d	2830	6.3	2.60	.881	Nein
	i	127	6.3	2.72		
Fakten lernen	d	2830	14.7	4.09	.263	Nein
	i	127	14.3	3.81		
Diagramme und Tabellen	d	2830	8.6	2.97	.004	Ja
	i	127	7.8	2.67		
Planen und Organisieren	d	2830	11.3	3.49	.026	Ja
	i	127	10.6	3.39		

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

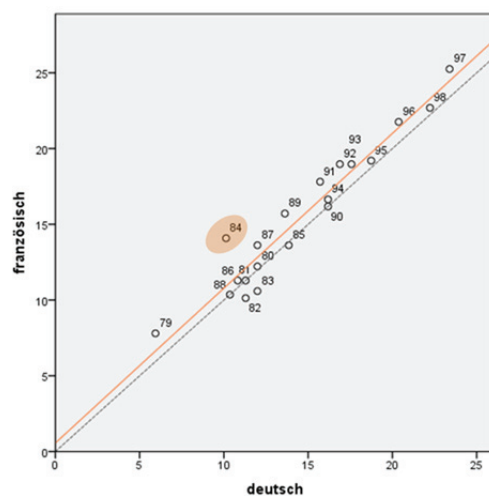
7.2.3 Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes

Eine Möglichkeit zur Identifikation von DIF, auch bei kleinen Stichprobengrößen, ist die Methode „Delta-Plot“. Bei diesem Verfahren werden die zu vergleichenden Aufgaben-Schwierigkeiten z-standardisiert und anschliessend in „Delta-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 wenigen Probanden gelöste) Aufgaben.

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 Aufgabenschwierigkeiten wie auch deren Reihenfolge in beiden Sprach Versionen vergleichbar sind.



Die Diagonale wird in den folgenden Abbildungen durch eine unterbrochene schwarze 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 Aufgabengruppen, deren Ursachen eher nicht testbedingt (Übersetzung), sondern in realen Gegebenheiten zu vermuten sind. Um die Regressionsgerade gruppierte Aufgaben 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 Aufgaben 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 Aufgaben stimmen hingegen in der untersuchten Sprachgruppe bezüglich der „Schwierigkeitshierarchie“ nicht mit der Referenzgruppe überein (sie sind im Vergleich zu den anderen Aufgaben zu schwer oder zu leicht ausgefallen). Die Ursache solcher Differenzen kann eher testbedingt, also beispielsweise in der Übersetzung vermutet werden. Betroffene Aufgaben 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 Aufgabengruppen die kritischen Aufgaben 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 Aufgaben 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 Aufgabenschwierigkeit zur deutschen Aufgabenschwierigkeit gutgeschrieben (also die Differenz der Lösungswahrscheinlichkeit). Dies betrifft nur jene Personen, welche die fragliche Aufgabe nicht korrekt beantwortet haben.

7.2.3.1 Medizinisch-naturwissenschaftliches Grundverständnis

Ein Ausgleich ist nur für die französische Sprachgruppe zu prüfen. Einzig Item 84 liegt über der kritischen Differenz, es werden den Personen, welche diese Aufgabe nicht korrekt gelöst haben 0.17 Punkte gut geschrieben.

Tabelle 22: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“ (Vergleich deutsch-französisch).

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
79	0.92	0.84		5.88	7.77			
80	0.66	0.65		12.03	12.27			
81	0.69	0.69		11.32	11.32			
82	0.69	0.74		11.32	10.14			
83	0.66	0.72		12.03	10.61			
84	0.74	0.57		10.14	14.16		0.17	
85	0.58	0.59		13.92	13.69			
86	0.71	0.69		10.85	11.32			
87	0.66	0.59		12.03	13.69			
88	0.73	0.73		10.37	10.37			
89	0.59	0.50		13.69	15.82			
90	0.48	0.48		16.29	16.29			
91	0.50	0.41		15.82	17.95			
92	0.45	0.36		17.00	19.13			
93	0.42	0.36		17.71	19.13			
94	0.48	0.46		16.29	16.76			
95	0.37	0.35		18.89	19.37			
96	0.30	0.24		20.55	21.97			
97	0.17	0.09		23.62	25.52			
98	0.22	0.20		22.44	22.91			

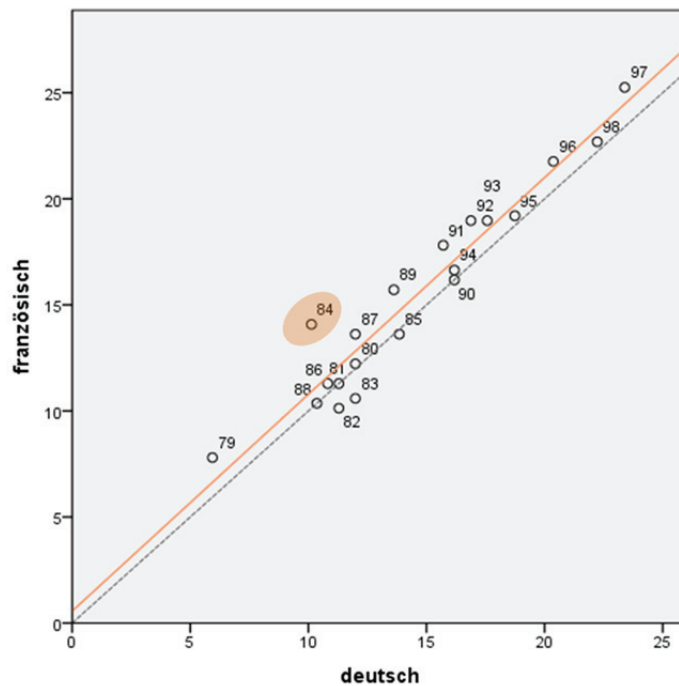


Abbildung 33: Delta-Plot für „Medizinisch-naturwissenschaftliches Grundverständnis“ (deutsch-französisch).

7.2.3.2 Textverständnis

Auch in der Aufgabengruppe „Textverständnis“ wird für die französische Sprachgruppe nahezu eine lineare Verschiebung der Regressionsgeraden beobachtet. Keine der Aufgaben übertrifft die kritische Differenz zu diesem Erwartungswert – es wird kein Ausgleich vorgenommen.

Tabelle 23: DIF-Analyse „Textverständnis“ (Vergleich deutsch-französisch).

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
41	0.74	0.76		10.14	9.67			
42	0.75	0.65		9.90	12.27			
43	0.57	0.60		14.16	13.45			
44	0.59	0.62		13.69	12.98			
45	0.64	0.53		12.50	15.11			
46	0.22	0.26		22.44	21.49			
53	0.73	0.74		10.37	10.14			
54	0.49	0.44		16.05	17.24			
55	0.59	0.46		13.69	16.76			
56	0.41	0.38		17.95	18.66			
57	0.33	0.32		19.84	20.07			
58	0.28	0.20		21.02	22.91			

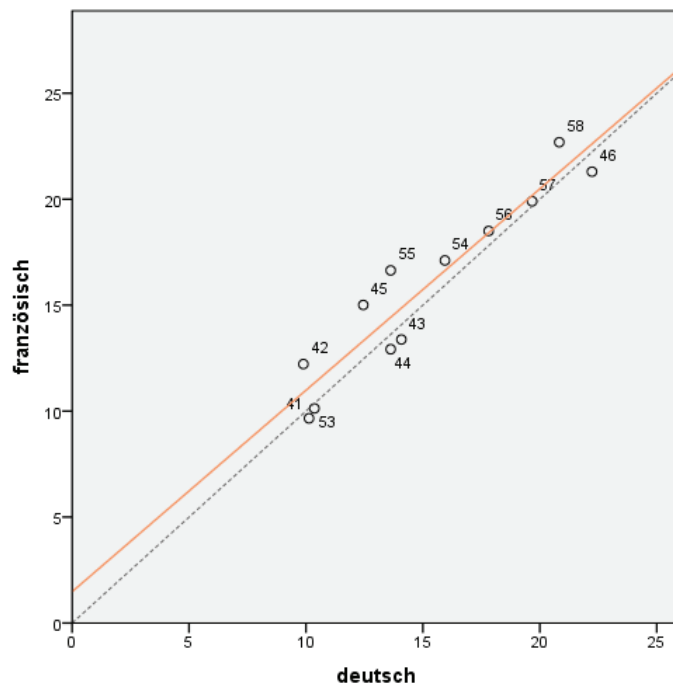


Abbildung 34: Delta-Plot für „Textverständnis“ (deutsch-französisch).

7.2.3.3 Fakten lernen

Auch bei dieser Aufgabengruppe liegen die erreichten Werte der französischen Sprachgruppe nahe an der (wiederum praktisch parallel verschobenen) Regressionsgeraden. Keine Aufgabe erreicht einen auffälligen Wert, es wird kein Ausgleich vorgenommen.

Tabelle 24: DIF-Analyse „Fakten lernen“ (Vergleich deutsch-französisch).

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
119	0.81	0.73		8.48	10.37			
120	0.79	0.74		8.96	10.14			
121	0.80	0.71		8.72	10.85			
122	0.83	0.64		8.01	12.50			
123	0.80	0.74		8.72	10.14			
124	0.60	0.43		13.45	17.47			
125	0.84	0.72		7.77	10.61			
126	0.79	0.64		8.96	12.50			
127	0.70	0.56		11.08	14.40			
128	0.80	0.83		8.72	8.01			
129	0.77	0.59		9.43	13.69			
130	0.85	0.75		7.54	9.90			
131	0.73	0.59		10.37	13.69			
132	0.59	0.41		13.69	17.95			
133	0.68	0.56		11.56	14.40			
134	0.71	0.65		10.85	12.27			
135	0.67	0.58		11.79	13.92			
136	0.75	0.64		9.90	12.50			
137	0.56	0.53		14.40	15.11			
138	0.65	0.61		12.27	13.21			

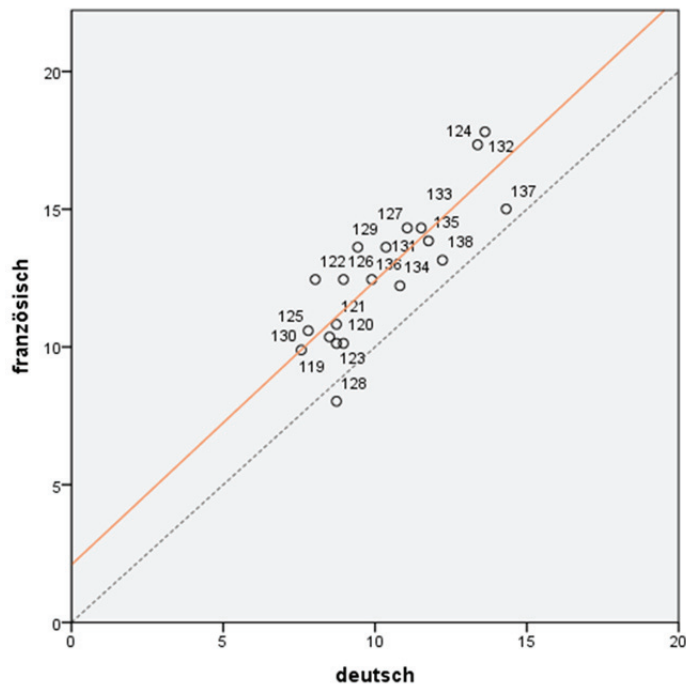


Abbildung 35: Delta-Plot für „Fakten lernen“ (deutsch-französisch).

7.2.3.4 Diagramme und Tabellen

Ein Ausgleich ist hier für beide Sprachgruppen zu prüfen. In der französischsprachigen Version liegen zwei Aufgaben über dem kritischen Wert (maximal 0.29 Bonuspunkte), in der italienischsprachigen Version ist eine Aufgabe (0.21 mögliche Bonuspunkte) auszugleichen.

Tabelle 25: DIF-Analyse „Diagramme und Tabellen“ (Vergleich deutsch-französisch und deutsch-italienisch).

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
159	0.82	0.83	0.86	8.25	8.01	7.30		
160	0.83	0.81	0.83	8.01	8.48	8.01		
161	0.79	0.75	0.69	8.96	9.90	11.32		
162	0.71	0.67	0.64	10.85	11.79	12.50		
164	0.60	0.62	0.65	13.45	12.98	12.27		
166	0.52	0.53	0.50	15.34	15.11	15.82		
167	0.64	0.80	0.63	12.50	8.72	12.74		
168	0.51	0.49	0.44	15.58	16.05	17.24		
169	0.60	0.56	0.61	13.45	14.40	13.21		
171	0.53	0.50	0.42	15.11	15.82	17.71		
172	0.54	0.40	0.33	14.87	18.18	19.84	0.14	0.21
174	0.30	0.31	0.34	20.55	20.31	19.60		
175	0.36	0.31	0.33	19.13	20.31	19.84		
176	0.28	0.13	0.18	21.02	24.57	23.39	0.15	
177	0.30	0.23	0.20	20.55	22.20	22.91		
178	0.23	0.19	0.15	22.20	23.15	24.10		

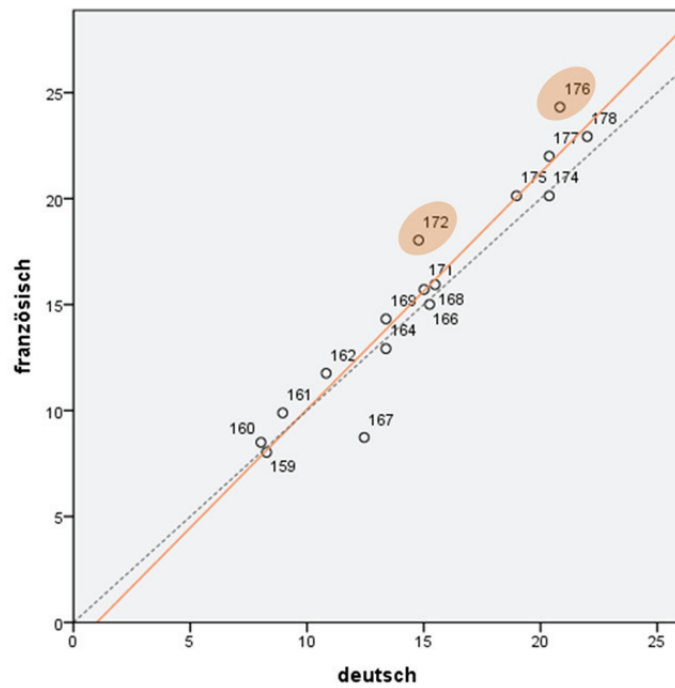


Abbildung 36: Delta-Plot für „Diagramme und Tabellen“ (deutsch-französisch).

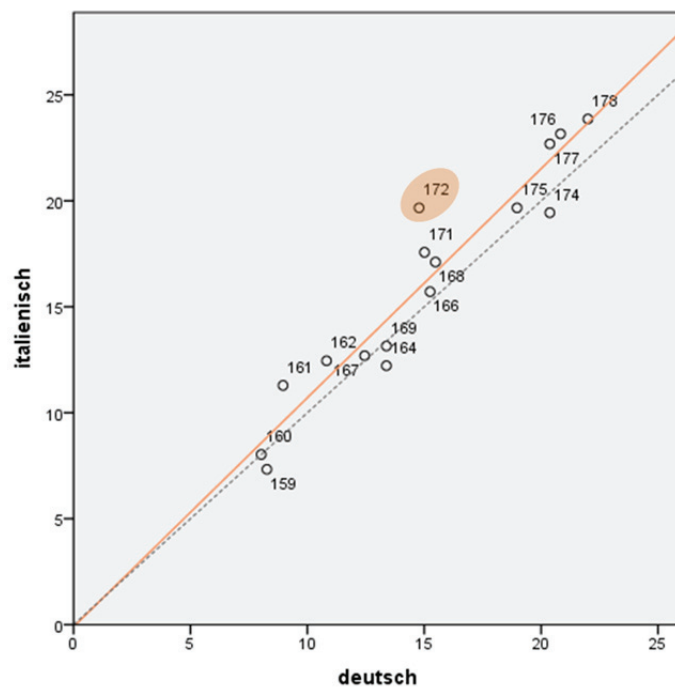


Abbildung 37: Delta-Plot für „Diagramme und Tabellen“ (deutsch-italienisch).

7.2.3.5 Planen und Organisieren

Ein Ausgleich ist in dieser Aufgabengruppe für beide Sprachgruppen zu prüfen. Eine Aufgabe liegt für die französische Sprachgruppe über dem kritischen Wert (0.21 mögliche Bonuspunkte), für die italienische Sprachgruppe ist keine auffällige Aufgabe festzustellen.

Tabelle 26: DIF-Analyse „Planen und Organisieren“ (Vergleich deutsch-französisch und deutsch-italienisch).

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
59	0.79	0.79	0.87	8.96	8.96	7.10		
60	0.17	0.20	0.18	23.39	22.69	23.16		
61	0.71	0.65	0.72	10.82	12.22	10.59		
62	0.71	0.72	0.69	10.82	10.59	11.29		
63	0.58	0.58	0.57	13.85	13.85	14.08		
64	0.79	0.75	0.74	8.96	9.89	10.13		
65	0.52	0.55	0.46	15.25	14.55	16.64		
66	0.52	0.55	0.44	15.25	14.55	17.11		
67	0.54	0.53	0.42	14.78	15.01	17.57		
68	0.44	0.45	0.45	17.11	16.88	16.88		
69	0.60	0.62	0.56	13.38	12.92	14.32		
70	0.66	0.57	0.64	11.99	14.08	12.45		
71	0.52	0.46	0.46	15.25	16.64	16.64		
72	0.44	0.23	0.42	17.11	22.00	17.57	0.21	
73	0.45	0.41	0.35	16.88	17.81	19.20		
74	0.80	0.79	0.74	8.73	8.96	10.13		
75	0.43	0.31	0.35	17.34	20.13	19.20		
76	0.53	0.45	0.49	15.01	16.88	15.94		
77	0.57	0.44	0.59	14.08	17.11	13.62		
78	0.51	0.57	0.48	15.48	14.08	16.18		

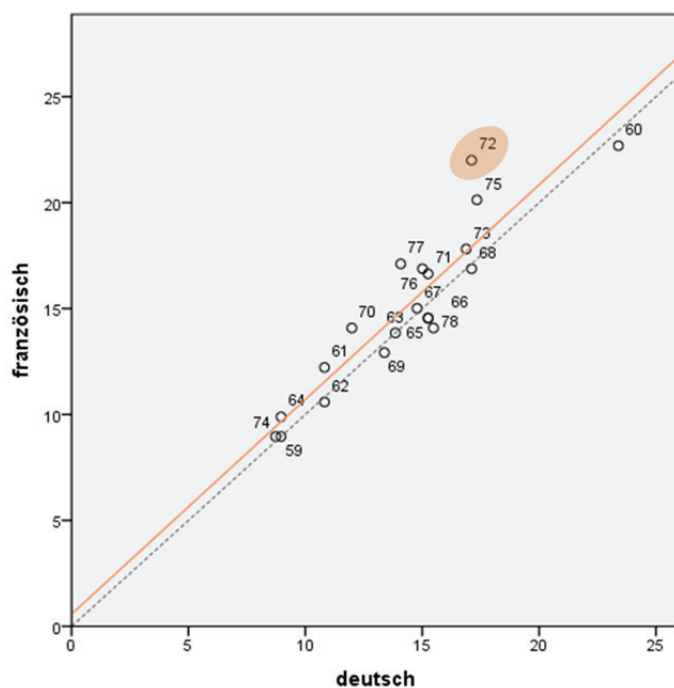


Abbildung 38: Delta-Plot für „Planen und Organisieren“ (deutsch-französisch).

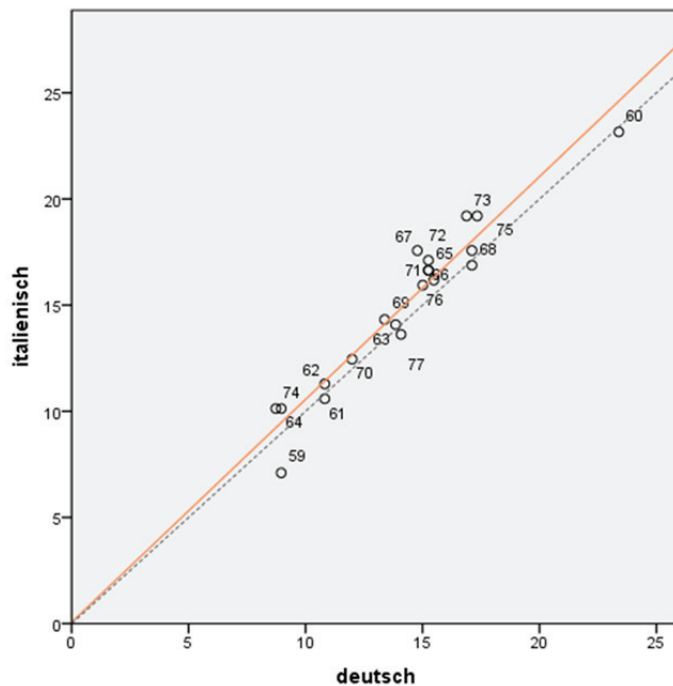


Abbildung 39: Delta-Plot für „Planen und Organisieren“ (deutsch-italienisch).

7.2.4 Effekte der Korrektur

Die durch den Ausgleich resultierenden Bonuspunkte können der Tabelle 27 entnommen werden. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Der maximale Bonus beträgt bei der französischen Sprachgruppe 1 Punkt, die italienische Sprachgruppe erhält keine Bonuspunkte (für die Berechnung des mittleren Rangplatzes wird die Korrektur um 0.21 Punkte bei den betreffenden Personen allerdings vorgenommen).

Die Auswirkungen der Korrektur bleiben damit insgesamt gering.

Tabelle 27: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen.

		n	Minimum	Maximum	m	s
Franz.	Planen und Organisieren	216	0.00	0.21	0.16	0.09
	Med.-naturw. Grundverständnis	216	0.00	0.17	0.07	0.08
	Diagramme und Tabellen	216	0.00	0.29	0.21	0.09
	GESAMT (Gerundet)	216	0	1	0.52	0.50
Ital.	Diagramme und Tabellen	127	0.00	0.21	0.14	0.10
	GESAMT (Gerundet)	127	0	0	0.00	0.00

Die Mittelwertsdifferenzen für jede Aufgabengruppe nach dem Sprachausgleich sind in Tabelle 28 dargestellt. Die Ergebnisse multipler Mittelwertsvergleiche 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 (auf dem 1%-Niveau) höhere Werte erzielt haben als die beiden anderen Sprachgruppen.

Auch nach dem Sprachausgleich bleiben Differenzen zwischen den Sprachgruppen erhalten – diese können jedoch nicht mehr in der Übersetzung des Tests begründet sein. In Abbildung 40 sind die Ausmasse der (ungerundeten) Bonuspunkte für die bisherigen Testdurchführungen dargestellt. Die senkrechten Linien bezeichnen Mittelwert und Standardabweichung der Jahrgänge 1999-2013.

Tabelle 28: Test-Prozentrang und Punktwerte der Aufgabengruppen für die Sprachen.

		n	Mittelwert	Stand.abw.	Minimum	Maximum
Test-Prozentrang (d**,i*)>f	deutsch	2830	52.05	28.18	10	100
	französisch	216	40.68	25.33	10	97
	italienisch	127	47.63	28.69	10	98
Quantitative und formale Probleme	deutsch	2830	9.80	3.12	1	16
	französisch	216	9.41	3.05	1	16
	italienisch	127	9.52	3.03	1	16
Schlauchfiguren d**>(i,f)	deutsch	2830	14.65	3.50	2	20
	französisch	216	12.97	3.56	5	20
	italienisch	127	13.27	3.60	5	20
Textverständnis	deutsch	2830	6.33	2.60	0	12
	französisch	216	5.96	2.45	0	12
	italienisch	127	6.30	2.72	0	12
Planen und Organisieren	deutsch	2830	11.30	3.49	1	20
	französisch	216	10.77	3.30	2	17
	italienisch	127	10.60	3.39	3	18
Med.-naturwiss. Grundverständnis d**>j*>f	deutsch	2830	11.03	3.54	1	20
	französisch	216	10.34	3.10	3	19
	italienisch	127	11.34	3.19	4	19
Figuren lernen	deutsch	2830	14.13	3.63	1	20
	französisch	216	13.59	3.48	3	20
	italienisch	127	14.09	3.67	4	20
Fakten lernen (d,i)**>f	deutsch	2830	14.71	4.09	1	20
	französisch	216	12.63	3.84	2	20
	italienisch	127	14.29	3.81	5	20
Muster zuordnen d**>f	deutsch	2830	11.97	3.16	0	20
	französisch	216	11.21	2.88	5	19
	italienisch	127	11.94	2.66	5	19
Diagramme und Tabellen d*>i	deutsch	2830	8.55	2.97	0	16
	französisch	216	8.36	2.69	1	16
	italienisch	127	7.92	2.63	1	15
Konzentriertes und sorgf. Arbeiten d**>f, i*>f	deutsch	2830	10.75	4.52	0	20
	französisch	216	9.56	4.10	0	20
	italienisch	127	10.80	4.25	0	20

/: signifikanter Einfluss des Faktors „Sprache“ bei varianzanalytischer Prüfung (1%- bzw. 5%-Niveau); darunter signifikante Unterschiede bei multiplem Mittelwertvergleich.

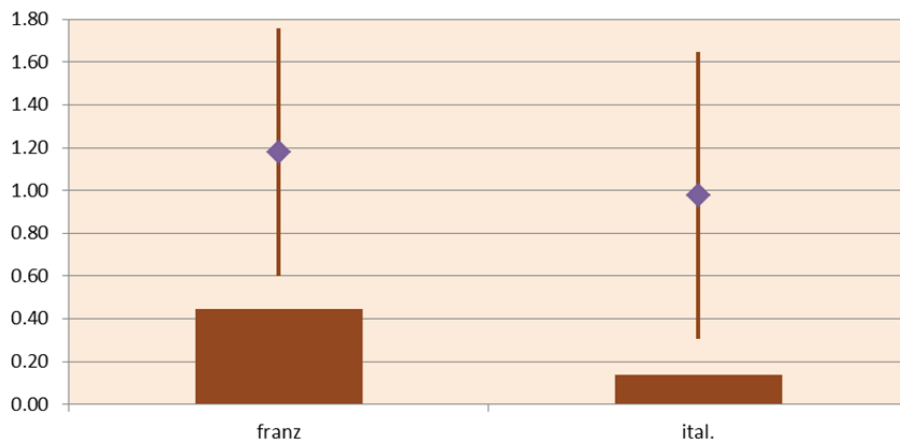


Abbildung 40: Sprachausgleich 2014 (ungerundet) und Mittelwert mit Standardabweichung aller vorhergehenden Testdurchführungen 1998 – 2013.

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.

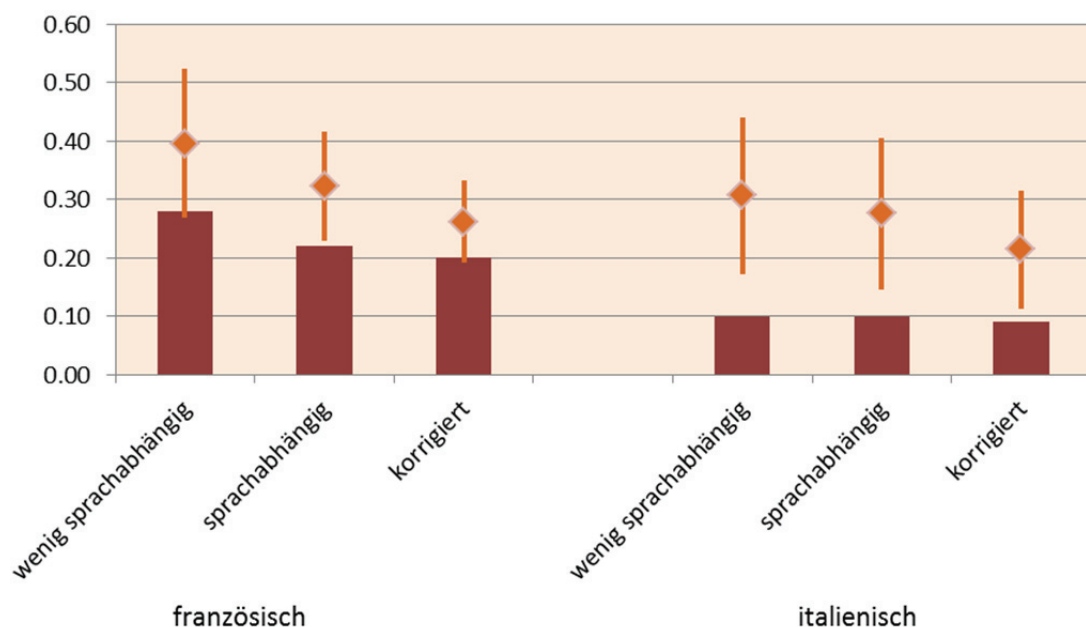


Abbildung 41: Mittlere Abweichung (Delta z) für wenig sprachabhängige und sprachabhängige Aufgabengruppen (korrigiert und unkorrigiert), 2014 und Vergleich zu 1998-2013 (Mittelwert und Streuung).

Beide Sprachgruppen weisen nach der Bonusbepunktung in den sprachabhängigen Aufgabengruppen eine geringere Differenz („korrigiert“) zur deutschen Sprachgruppe auf als in den wenig-sprachabhängigen Aufgabengruppen. Eine spezifische Benachteiligung durch die Sprachabhängigkeit von Aufgabengruppen darf weiterhin ausgeschlossen werden.

7.3 Vergleichbarkeit der Testlokale

Die Durchführungsbedingungen an den einzelnen Testorten sind standardisiert. Hinsichtlich des Test-Prozentrangs treten keine Unterschiede zwischen den (deutschsprachigen) Testlokalen auf. Die Werte von drei Aufgabengruppen unterscheiden sich signifikant zwischen den Testlokalen. Diese Unterschiede sind allerdings mit regionalen Unterschieden konfundiert.

Tabelle 29: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den (deutschsprachigen) Testlokalen.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testprozentrang	<i>Zwischen Gruppen</i>	24271.259	26	933.510	1.178	.244
	<i>Innerhalb Gruppen</i>	2221913.616	2803	792.691		
	<i>Total</i>	2246184.875	2829			
Muster zuordnen	<i>Zwischen Gruppen</i>	302.151	26	11.621	1.168	.254
	<i>Innerhalb Gruppen</i>	27893.068	2803	9.951		
	<i>Total</i>	28195.219	2829			
Med.-naturwiss. Grundverständnis	<i>Zwischen Gruppen</i>	320.896	26	12.342	.984	.486
	<i>Innerhalb Gruppen</i>	35149.899	2803	12.540		
	<i>Total</i>	35470.795	2829			
Schlauchfiguren	<i>Zwischen Gruppen</i>	291.514	26	11.212	.913	.592
	<i>Innerhalb Gruppen</i>	34439.129	2803	12.287		
	<i>Total</i>	34730.643	2829			
Quant. und formale Probleme	<i>Zwischen Gruppen</i>	584.910	26	22.497	2.333	.000
	<i>Innerhalb Gruppen</i>	27031.855	2803	9.644		
	<i>Total</i>	27616.765	2829			
Textverständnis	<i>Zwischen Gruppen</i>	282.556	26	10.868	1.611	.026
	<i>Innerhalb Gruppen</i>	18913.550	2803	6.748		
	<i>Total</i>	19196.106	2829			
Figuren lernen	<i>Zwischen Gruppen</i>	326.378	26	12.553	.950	.537
	<i>Innerhalb Gruppen</i>	37047.666	2803	13.217		
	<i>Total</i>	37374.044	2829			
Fakten lernen	<i>Zwischen Gruppen</i>	446.405	26	17.169	1.024	.430
	<i>Innerhalb Gruppen</i>	46990.993	2803	16.765		
	<i>Total</i>	47437.398	2829			
Diagramme und Tabellen	<i>Zwischen Gruppen</i>	212.515	26	8.174	.924	.575
	<i>Innerhalb Gruppen</i>	24797.034	2803	8.847		
	<i>Total</i>	25009.549	2829			
Konzentr. u. sorgf. Arbeiten	<i>Zwischen Gruppen</i>	723.513	26	27.827	1.365	.103
	<i>Innerhalb Gruppen</i>	57153.307	2803	20.390		
	<i>Total</i>	57876.820	2829			
Planen und Organisieren	<i>Zwischen Gruppen</i>	474.796	26	18.261	1.506	.048
	<i>Innerhalb Gruppen</i>	33981.098	2803	12.123		
	<i>Total</i>	34455.894	2829			

Aus den Protokollen der Testabnahme und der Analyse der erzielten Ergebnisse nach Aufgabengruppen ergeben sich keine Hinweise, die als Einschränkungen der Chancengleichheit zu bewerten wären. Insgesamt gibt es keinen Anlass, Korrekturen aufgrund der Testlokalzuordnung vorzunehmen.

Tabelle 30: Darstellung der detaillierten Ergebnisse nach Testlokalen (deutschsprachig).

		Testprozentrang	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	51.01	12.21	10.43	14.89	9.34	5.90	14.77	14.49	7.96	10.73	11.14	70
	2	49.26	11.89	10.43	14.86	8.99	5.92	14.40	14.79	8.19	10.48	10.93	136
	3	53.37	11.90	11.00	15.07	8.94	6.13	14.54	14.99	8.69	11.06	11.21	70
	4	55.09	12.10	11.31	15.50	9.56	6.87	14.49	14.68	8.53	10.91	11.21	68
	5	49.99	12.23	10.63	14.62	9.01	6.20	14.48	14.52	7.82	11.70	10.93	71
	6	52.44	11.90	11.09	14.89	9.73	6.46	14.00	14.69	8.69	10.28	11.44	131
	7	50.87	11.74	10.70	14.94	9.79	6.62	14.27	15.24	8.56	9.76	10.78	94
	8	44.27	11.29	10.86	14.12	9.20	6.20	13.25	13.81	8.19	9.81	11.12	59
	9	51.09	11.64	10.69	14.71	9.93	6.00	14.18	15.03	8.53	10.74	10.97	118
	10	45.80	11.20	11.15	14.34	8.84	6.49	14.05	13.67	8.08	10.38	10.98	61
	11	49.05	11.76	10.87	14.60	9.51	6.22	13.64	14.49	8.45	10.60	10.76	144
	12	53.43	12.39	10.77	14.79	10.07	6.08	13.92	15.11	8.56	11.41	11.00	195
	13	56.07	11.71	11.57	14.79	10.68	6.69	14.32	15.13	8.76	10.93	11.43	68
	14	54.52	12.30	10.85	14.99	10.01	6.34	14.49	15.19	8.60	11.33	11.69	67
	15	54.88	11.97	11.76	14.88	10.59	6.87	14.79	14.28	8.94	10.21	11.54	68
	16	58.07	12.17	11.58	14.48	10.70	7.20	14.20	15.47	9.14	11.28	11.81	115
	17	53.87	12.06	11.31	14.78	9.96	6.15	14.08	14.72	8.56	10.17	11.48	99
	18	53.67	12.34	10.80	14.96	9.99	5.87	14.48	14.82	8.76	10.55	11.06	89
	19	55.95	12.88	11.46	14.80	10.21	6.38	14.02	14.69	8.55	11.79	11.42	107
	20	50.97	11.75	10.92	14.13	9.94	6.18	13.88	14.67	8.48	10.47	10.84	89
	21	51.69	12.51	11.01	14.67	9.74	6.50	13.59	15.19	8.69	10.19	11.57	70
	22	57.05	12.02	11.85	14.74	10.03	6.77	14.86	15.09	8.89	11.11	11.40	65
	23	52.99	11.68	11.13	14.30	9.95	6.57	13.65	14.73	8.78	11.24	12.09	135
	24	54.81	11.69	11.37	15.04	9.88	6.58	14.39	14.79	8.47	10.96	12.33	109
	25	49.04	11.79	10.92	14.39	9.56	6.04	14.21	14.38	8.30	10.86	11.12	145
	26	49.11	11.93	10.97	14.18	9.50	6.23	14.09	14.53	8.41	10.74	10.99	151
	27	50.61	11.99	11.07	14.19	10.15	6.36	13.93	14.09	8.83	10.23	11.69	236
	Gesamt	52.05	11.97	11.03	14.65	9.80	6.33	14.13	14.71	8.55	10.75	11.30	

Markiert: Mittelwerte für Test-Prozentrang und die Punktwerte der Aufgabengruppen im multiplen Mittelwertsvergleich nach Tukey, die signifikant tiefer als in mindestens einem anderen Testlokal liegen.

7.4 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. Seit 2006 hat sich der Männeranteil bei 40% stabilisiert, das gleiche Verhältnis findet sich seit 2010 in der Zahnmedizin. In der Veterinärmedizin liegt der Männeranteil unter 20%. Die mittleren Test-Prozentränge nach Geschlecht, Sprache und Disziplin können Tabelle 31 entnommen werden.

Tabelle 31: Statistiken für den Test-Prozentrang nach Geschlechtern für Sprache und Disziplin.

		Testsprache			Wunschdisziplin			Gesamt
		deutsch	französisch	italienisch	HM	VM	ZM	
männlich	m	52.2	46.7	48.4	53.7	43.5	37.7	51.8
	s	28.7	29.1	28.9	28.5	25.8	28.0	28.8
	n	n=1048	n=61	n=41	n=989	n=63	n=98	n=1150
weiblich	m	51.2	37.3	46.5	51.9	44.1	42.2	50.0
	s	29.0	24.6	29.8	29.13	26.83	28.73	28.9
	n	n=1782	n=155	n=86	n=1536	n=321	n=166	n=2023

m: Mittelwert, s: Standardabweichung, n: Personenzahl

Für die zahlenmässig bedeutendste deutschsprachige Teilnehmergruppe wird ein Unterschied von 1.0 Test-Prozentrangpunkten registriert. Dieser bewegt sich im Rahmen der bis 2013 verwendeten Testwertunterschiede. Es gibt weiterhin keinen Anlass, in der Schweiz von deutlichen oder testbedingten Genderdifferenzen auszugehen.

Wenn man die Unterschiede disziplin- und sprachspezifisch betrachtet, zeigen sich die aus den Vorjahren bekannten Differenzierungen. Der Unterschied in der deutschen Sprachgruppe für Humanmedizin beträgt 1.5 Test-Prozentrangpunkte.

Tabelle 32: Kombinierte Statistiken für den Test-Prozentrang nach Geschlechtern, Sprache und Disziplin.

		Humanmedizin			Veterinärmedizin			Zahnmedizin		
		d	f	i	d	f	i	d	f	i
männlich	m	54.1	48.6	50.1	43.6	45.6	36.2	38.0	13.0	50.0
	s	28.5	28.1	29.4	24.6	32.3	19.1	27.8	5.7	50.9
	n	910	45	34	44	14	5	94	2	2
weiblich	m	52.6	44.3	47.2	49.1	31.2	46.7	42.5	35.7	39.2
	s	29.2	26.6	29.9	27.1	20.7	29.4	28.5	37.2	35.1
	n	1403	72	61	222	80	19	157	3	6

Tabelle 33: Prüfung zur Varianzhomogenität und der Mittelwertunterschiede (t-Test) für Test-Prozentrang zwischen Geschlechtern (nur für ausreichende Zellengrößen).

Sprache	Disziplin	Levene-Test Gleichheit der Varianzen			t-Test Gleichheit des Mittelwerts		
			F	Sig.	t	df	Sig.
Deutsch	HM	homogen	2.401	0.121	1.254	2311.0	0.210
	VM	homogen	1.733	0.189	-1.248	264.0	0.213
	ZM	homogen	0.539	0.463	-1.215	249.0	0.226
Französisch	HM	homogen	0.391	0.533	0.825	115.0	0.411
Italienisch	HM	homogen	0.299	0.543	0.465	93.0	0.643

Abbildung 42 verdeutlicht die Vergleichbarkeit der Geschlechterdifferenzen mit den Vorjahren. Die diesjährigen Punktwert-Unterschiede auf Ebene der Aufgabengruppen (Balken) sind überwiegend geringer als jene aus den Vorjahren (Mittelwert und Standardabweichung), weisen aber die bekannte Ausprägungsrichtung auf. Auffallend sind die Aufgabengruppen „Schlauchfiguren“ und „Quantitative und formale Probleme“, in denen die Differenz zu den männlichen Kandidaten etwas geringer als gewohnt ausfällt. Im Konzentrierten und sorgfältigen Arbeiten erzielten die Kandidatinnen deutlich bessere Werte als ihre männlichen Mitbewerber.

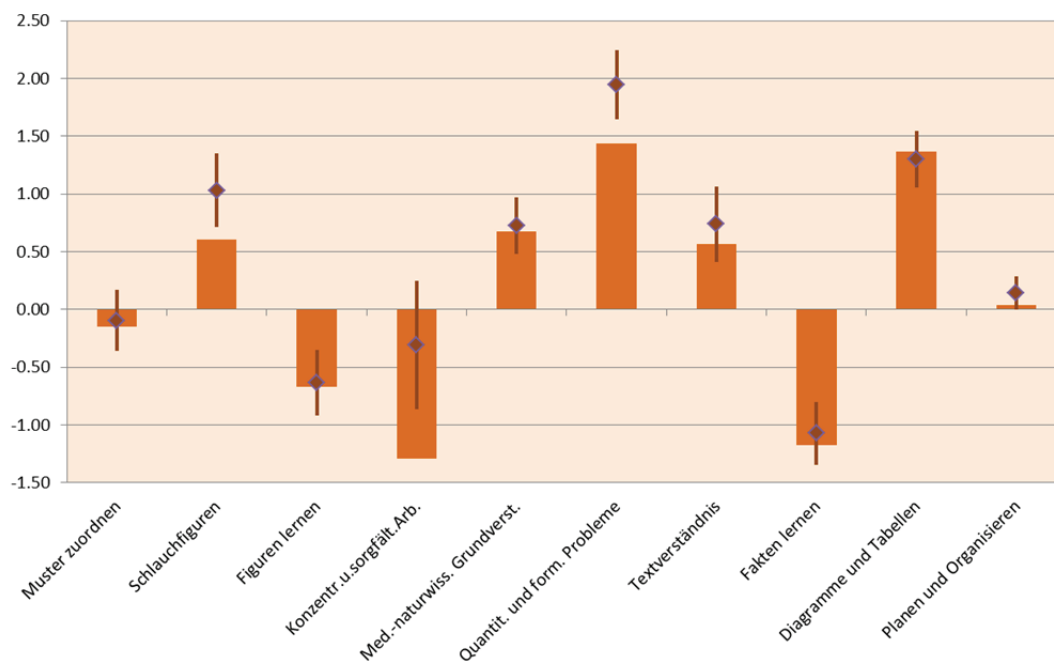


Abbildung 42: Geschlechtsspezifische Differenzen nach Aufgabengruppen (m-w) mit Vergleich (Mittelwert und Standardabweichung) für die Jahre 1998-2013.

7.5 Vergleiche für Altersgruppen

Für die Überprüfung der Alterseffekte wurden die Gruppen wie nachstehend dargestellt gebildet.

Tabelle 34: Gruppenbildung für das Jahr der Maturitätsprüfung bezogen auf die Geburtsjahre (Humanmedizin, deutschsprachig).

Maturitätsjahr	Geburtsjahr			Total
	bis 1989	1990-1992	ab 1993	
1981	1	0	0	1
1987	1	0	0	1
1989	1	0	0	1
1991	2	0	0	2
1993	0	1	0	1
1994	1	0	0	1
1995	0	0	1	1
1996	1	0	0	1
1997	2	0	0	2
1998	1	0	0	1
1999	2	1	0	3
2000	2	0	0	2
2001	2	0	0	2
2002	4	0	0	4
2003	2	0	0	2
2004	7	0	0	7
2005	6	0	1	7
2006	10	0	0	10
2007	10	0	1	11
2008	14	1	0	15
2009	21	8	1	30
2010	21	39	4	64
2011	8	91	1	100
2012	18	178	49	245
2013	21	218	540	779
2014	33	127	860	1020
Total	169	664	1458	2313

Hinsichtlich des Test-Prozentrangs unterscheidet sich 2014 die jüngste Altersgruppe von der ältesten Gruppe mit später Maturität. Die (zahlenmässig geringe) Gruppe der Ältesten mit früher Maturität erreicht die höchsten Test-Prozentränge. Die Punktwerte der deutschsprachigen Altersgruppen (Humanmedizin) für die einzelnen Aufgabengruppen (mit Kennzeichnung der signifikanten Unterschiede) sind in Tabelle 35 aufgeführt.

Tabelle 35: Test-Prozentrang und Punktwerte für die drei Altersgruppen nach Geburtsjahren (Humanmedizin, deutschsprachig)

	Geburtsjahr und Maturität	Mittelwert	Standard-abw.	Quadr.-summe	df	MQ	F	Sig.	Homogene Gruppen
Test-Prozentrang	ab 1994	54.50	28.66	18955.4	3	6318.5	7.615	.000	
	1991 - 1993	51.83	28.95						
	bis 1990, frühe Maturität	58.42	28.79						
	bis 1990, späte Maturität	43.14	29.67						
Muster zuordnen	ab 1994	11.95	3.20	75.1	3	25.0	2.5	.058	
	1991 - 1993	12.16	3.17						
	bis 1990, frühe Maturität	12.16	2.54						
	bis 1990, späte Maturität	11.37	3.04						
Med.-naturwiss. Grundverständnis	ab 1994	11.41	3.51	283.0	3	94.3	7.6	.000	
	1991 - 1993	10.89	3.51						
	bis 1990, frühe Maturität	12.22	3.56						
	bis 1990, späte Maturität	10.35	3.59						
Schlauchfiguren	ab 1994	14.62	3.52	65.2	3	21.7	1.7	.155	
	1991 - 1993	14.91	3.47						
	bis 1990, frühe Maturität	14.69	3.78						
	bis 1990, späte Maturität	14.26	3.76						
Quantitative und formale Probleme	ab 1994	10.24	3.10	323.5	3	107.8	11.4	.000	
	1991 - 1993	9.81	2.94						
	bis 1990, frühe Maturität	9.98	3.74						
	bis 1990, späte Maturität	8.74	3.24						
Textverständnis	ab 1994	6.61	2.59	135.7	3	45.2	6.7	.000	
	1991 - 1993	6.23	2.55						
	bis 1990, frühe Maturität	7.00	2.65						
	bis 1990, späte Maturität	5.83	2.81						
Figuren lernen	ab 1994	14.19	3.55	106.4	3	35.5	2.7	.044	
	1991 - 1993	14.23	3.72						
	bis 1990, frühe Maturität	13.31	3.59						
	bis 1990, späte Maturität	13.48	3.87						
Fakten lernen	ab 1994	14.81	4.02	68.9	3	23.0	1.4	.241	
	1991 - 1993	15.02	4.11						
	bis 1990, frühe Maturität	15.24	3.45						
	bis 1990, späte Maturität	14.31	4.33						
Diagramme und Tabellen	ab 1994	8.97	2.90	399.8	3	133.3	15.5	.000	
	1991 - 1993	8.36	2.94						
	bis 1990, frühe Maturität	9.75	3.15						
	bis 1990, späte Maturität	7.63	3.09						
Konzentriertes und sorgfältiges Arbeiten	ab 1994	10.94	4.59	162.4	3	54.1	2.6	.048	
	1991 - 1993	10.67	4.43						
	bis 1990, frühe Maturität	11.27	4.24						
	bis 1990, späte Maturität	9.90	4.52						
Planen und Organisieren	ab 1994	11.69	3.34	364.7	3	121.6	10.4	.000	
	1991 - 1993	11.06	3.52						
	bis 1990, frühe Maturität	12.02	3.62						
	bis 1990, späte Maturität	10.37	3.67						

Für die älteste Gruppe aufgeteilt nach früher und später Maturitätsprüfung; Varianz-analytische Prüfung der Unterschiede. In den letzten Spalten sind die bezüglich des Mittelwertes homogenen Gruppen dargestellt. Schattierungen in der gleichen Spalte bedeuten, dass sich die markierten Gruppen NICHT voneinander unterscheiden.

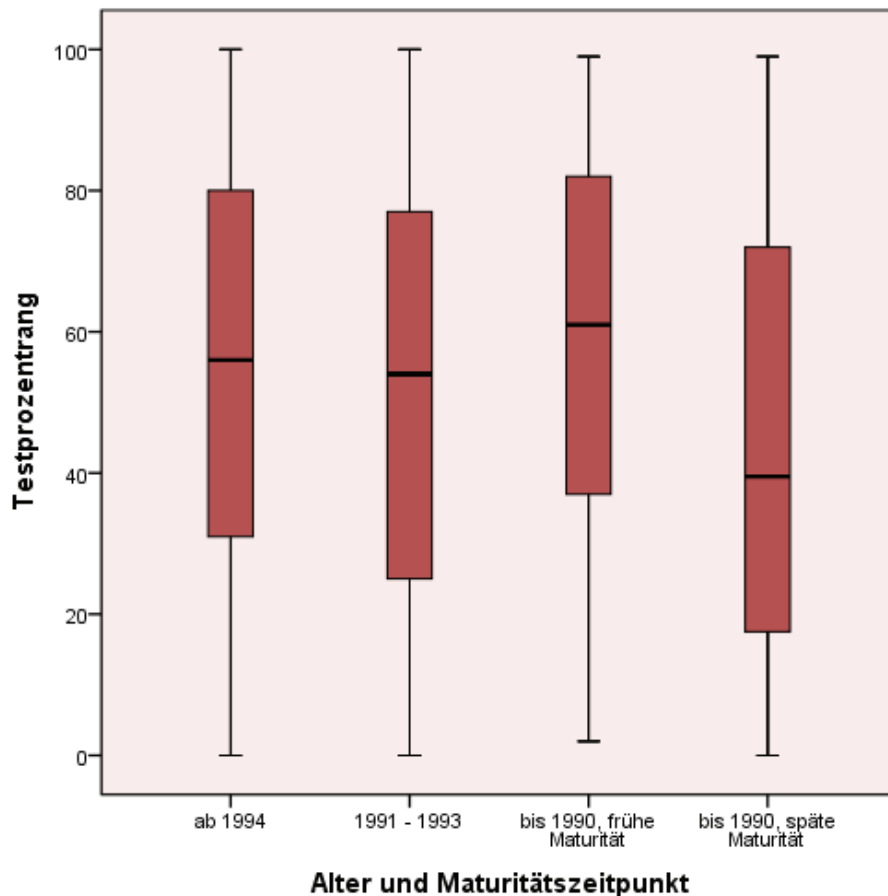


Abbildung 43: Boxplot für Test-Prozentrang der Altersgruppen (Geburtsjahre) und Maturitätsjahr (Humanmedizin, deutsche Sprachgruppe).

7.6 Vergleiche nach Wunschuniversitäten

Die Tabelle 36 führt die Test-Prozenträge getrennt nach Disziplin und gewünschter Universität (erste Wahl) aller deutschsprachigen Kandidaten auf. Für die Bewerber in Zahnmedizin sind keine Differenzen feststellbar, in den übrigen Disziplinen unterscheiden sich nur die Kandidaten mit den Wunschuniversitäten Bern und Zürich signifikant.

Obwohl die Nachfrage natürlich auch nach regionalen Gesichtspunkten erfolgt (z.B. nach Wohnortnähe), gibt es damit kaum bedeutsame Fähigkeitsunterschiede zwischen den Kandidaten verschiedener Universitäten. Vergleicht man nur die Zugelassenen (Tabelle 37) treten keine signifikanten Differenzen zwischen den Universitäten auf. Dies ist auch ein Hinweis darauf, dass die Attraktivität der vier Universitäten bei der Studienwahl nicht von Leistungsunterschieden determiniert wird und es „anspruchsvollere“ und „weniger anspruchsvollere“ Universitäten gäbe. Insgesamt nivellieren sich hier offenbar auch die früher festgestellten Unterschiede zwischen den einzelnen Kantonen.

Tabelle 36: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschuniversitäten, (d).

	Wunsch-universität	n	Mittel-wert	Stand-ardabw.	Quadrat-summe	Df	MQ	F	Sig.
HM	Bern	584	50.6*	28.5	8494.9	3	2831.6	3.4	.017
	Basel	539	52.1	28.8					
	Freiburg	76	52.3	31.4					
	Zürich	1114	55.1*	29.0					
ZM	Bern	69	40.6	29.8	127.3	3	42.4	0.1	.984
	Basel	54	39.7	28.5					
	Freiburg	6	39.8	30.9					
	Zürich	122	41.4	27.6					
VM	Bern	117	44.4*	26.5	3048.1	1	3048.1	4.3	.038
	Zürich	149	51.2*	26.6					
Gesamt	Bern	770	48.8*	28.5	11186.8	3	3728.9	4.5	.004
	Basel	593	51.0	28.9					
	Freiburg	82	51.3	31.4					
	Zürich	1385	53.4*	28.9					

*: signifikante Differenzen

Tabelle 37: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschuniversitäten, zugelassene, deutschsprachige Personen.

	Wunsch-universität	n	Mittel-wert	Stand-ardabw.	Quadrat-summe	Df	MQ	F	Sig.
HM	Bern	178	84.7	8.6	123.4	3	41.1	0.5	.659
	Basel	173	85.8	8.7					
	Freiburg	30	85.5	9.0					
	Zürich	413	85.5	8.9					
ZM	Bern	32	68.9	15.9	240.0	3	80.0	0.3	.826
	Basel	26	66.0	13.7					
	Freiburg	3	64.0	19.3					
	Zürich	59	65.7	17.5					
VM	Bern	48	71.1	15.5	113.1	1	113.1	0.5	.481
	Zürich	77	73.0	14.8					
Gesamt	Bern	258	80.2	13.1	1163.0	3	387.7	2.4	.068
	Basel	199	83.2	11.6					
	Freiburg	33	83.5	11.6					
	Zürich	549	81.6	13.1					

8 Ergebnisse zur Testgüte

8.1 Zuverlässigkeit

Die zulassungsrelevante Reliabilität (Zuverlässigkeit) kann anhand zweier Koeffizienten berechnet werden. Die Reliabilitätsschätzung 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 der entsprechenden Aufgabengruppen.

Entscheidend sind die Kennwerte des Punktwertes, welcher nach der Standardisierung als Test-Prozentrang für die Zulassung verwendet wird.

Der Zuverlässigkeitswert des Punktwertes liegt 2014 mit 0.92 wiederum im gewohnt hohen Bereich der Vorjahre. Auch die Konsistenzen des Testprofils liegt mit 0.82 im üblichen Bereich bisheriger Werte. Hierbei ist zu beachten, dass zu hohe Werte für sehr gleichartige Aufgabengruppen (mit der Frage, ob man einzelne weglassen kann) sprechen würden, sehr niedrige Werte für eine heterogene Testbatterie, die nicht ohne weiteres zu einem Test-Prozentrang 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. Test-Prozentrang für die Eignungsmessung zu verwenden.

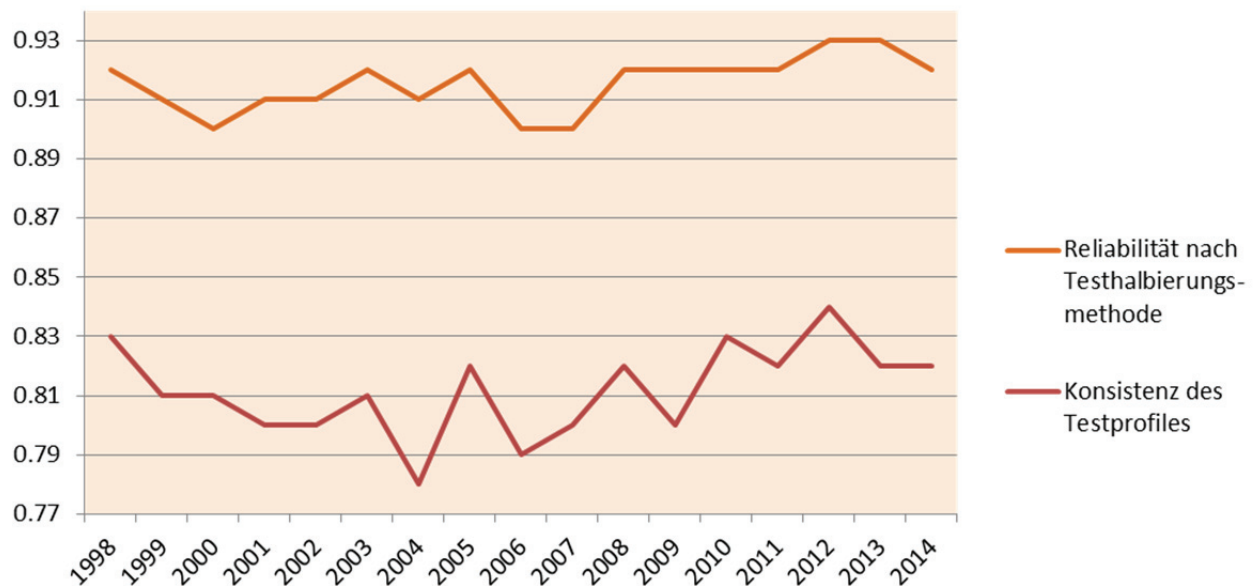


Abbildung 44: Reliabilität des Punktwertes (Split Half) und Konsistenz des Testprofils 1998 bis 2014.

Die Zuverlässigkeitswerte der einzelnen Aufgabengruppen sind für die Testverwendung nicht bedeutsam, da keine Entscheidungen auf der Basis einzelner Aufgabengruppen getroffen und diese nicht einzeln interpretiert werden. Hier müssten lediglich Eingriffe erfolgen, wenn der Test-Prozentrang selbst eine zu geringe Zuverlässigkeit aufweisen würde (was aber ausgeschlossen werden kann).

Insgesamt werden auch auf Ebene der Aufgabengruppen mit den Vorjahren vergleichbare Kennwerte erreicht (Abbildung 45 und Abbildung 46).

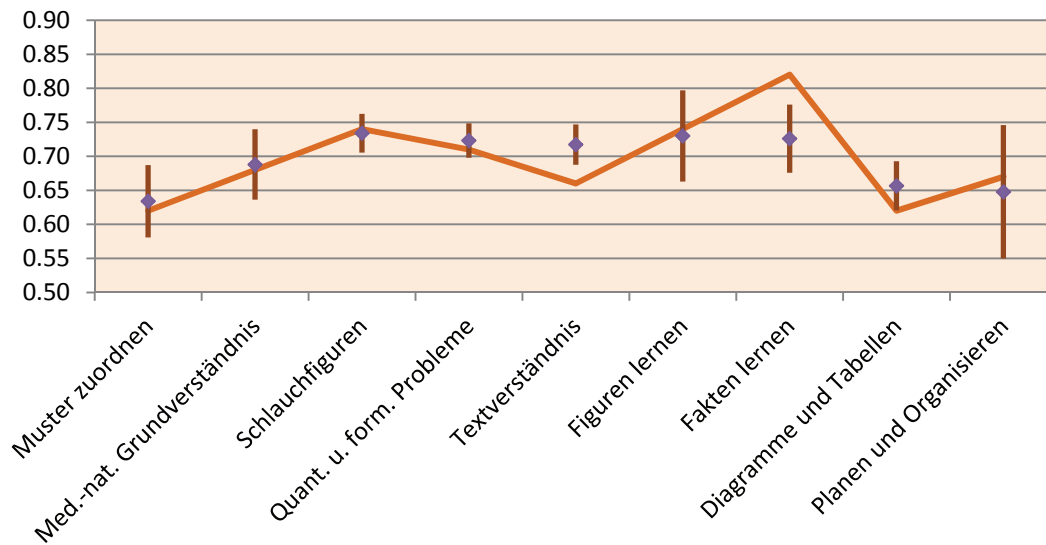


Abbildung 45: Zuverlässigkeit nach der Testhalbierungsmethode 2014 und Vergleich der mit den Jahren 1998 bis 2013 (Mittelwert und Standardabweichung).

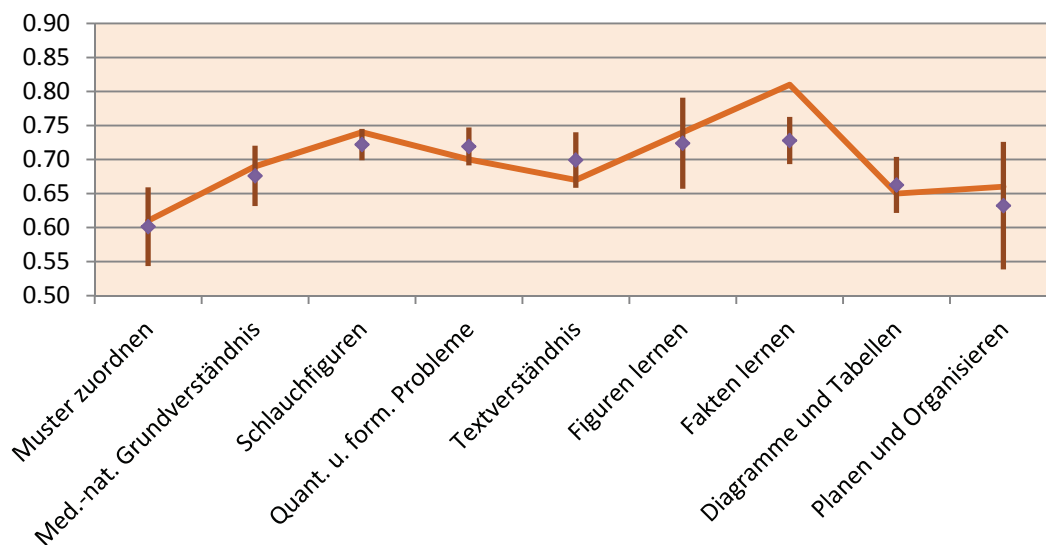


Abbildung 46: Skalenkonsistenzen 2014 und Vergleich der mit den Jahren 1998 bis 2013 (Mittelwert und Standardabweichung).

8.2 Binnenstruktur

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

Tabelle 38: Korrelationen zwischen Punktwerten der Aufgabengruppen sowie mit dem Test-Prozentrang.

	Med.-naturw. Grundv.	Schlauchfiguren	Quant. u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Planen und Organisieren	Test-Prozentrang
Muster zuordnen	0.22	0.47	0.26	0.20	0.35	0.38	0.21	0.34	0.20	0.57
Med.-naturw. Grundverst.		0.29	0.54	0.61	0.25	0.25	0.58	0.23	0.49	0.68
Schlauchfiguren			0.36	0.28	0.42	0.44	0.32	0.30	0.26	0.65
Quant. u. form. Probleme				0.46	0.24	0.27	0.62	0.16	0.49	0.66
Textverständnis					0.23	0.22	0.49	0.17	0.41	0.61
Figuren lernen						0.46	0.20	0.35	0.23	0.59
Fakten lernen							0.23	0.27	0.23	0.62
Diagramme und Tabellen								0.16	0.51	0.65
Konzent. u. sorgf. Arbeiten									0.20	0.53
Planen und Organisieren										0.61

Die ermittelten Werte entsprechen mit wenigen Abweichungen den aus den Vorjahren bekannten Korrelationen (Abbildung 47). Die Struktur des EMS bleibt auch 2014 stabil.

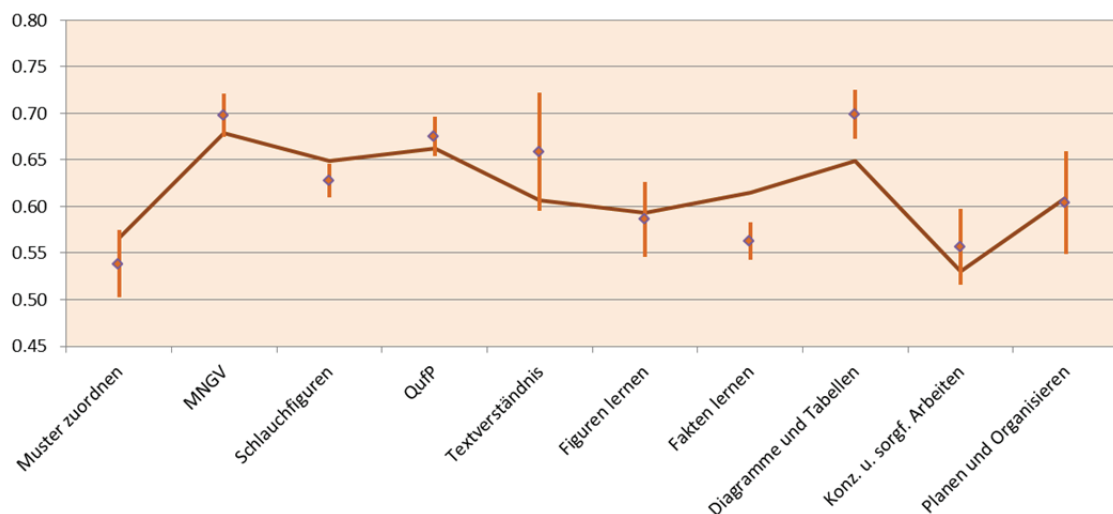


Abbildung 47: Korrelationen zwischen Punktwerten und Aufgabengruppen 2014 und 1998 bis 2013 (Mittelwert und Standardabweichung).

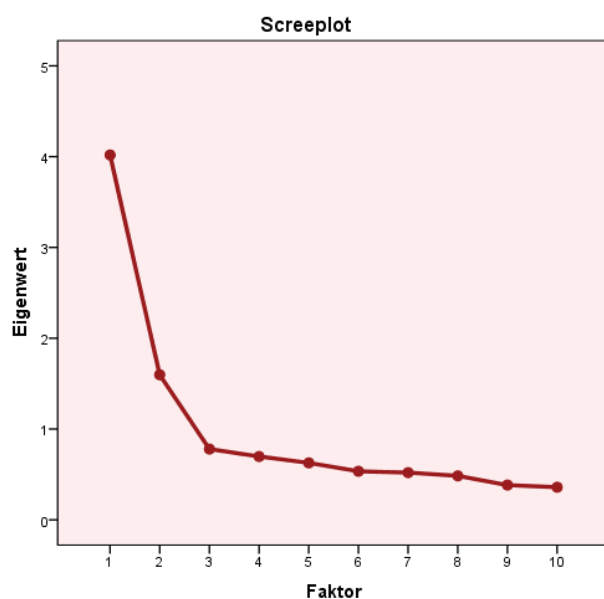


Abbildung 48: Screeplot (Eigenwerteverlauf) der Faktoren 1 bis 10 für die Aufgabengruppen des EMS 2014.

Der Screeplot stellt den Eigenwerteverlauf der Faktoren grafisch dar. Die Ein-Faktorenlösung ist deutlich hervorgehoben, die Zweifaktorenlösung erreicht die bekannte Aufteilung in die Faktoren „Schlussfolgerndes Denken“ und „Werkzeugfunktionen“.

Tabelle 39: Faktorenanalyse: Verlauf der unrotierten Lösung (2010-2014).

Unrotierte Lösung															
Faktor	Eigenwert					% Varianz					Kumuliert %				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
1	4.04	3.98	4.14	4.02	4.02	40.43	39.75	41.43	40.16	40.18	40.4	39.75	41.43	40.16	40.18
2	1.48	1.44	1.64	1.58	1.60	14.76	14.37	16.39	15.80	15.97	55.2	54.12	57.82	55.95	56.16
3	0.88	0.86	0.80	0.89	0.78	8.82	8.59	8.04	8.89	7.79	64.0	62.71	65.86	64.84	63.94
4	0.73	0.74	0.67	0.68	0.70	7.26	7.41	6.74	6.75	6.98	71.3	70.12	72.61	71.60	70.92
5	0.58	0.60	0.57	0.58	0.63	5.77	5.98	5.73	5.81	6.27	77.0	76.10	78.34	77.40	77.19
6	0.56	0.58	0.50	0.51	0.53	5.63	5.76	5.04	5.06	5.34	82.7	81.86	83.38	82.46	82.53
7	0.54	0.54	0.49	0.50	0.52	5.39	5.36	4.95	5.01	5.20	88.1	87.22	88.33	87.47	87.74
8	0.51	0.50	0.46	0.49	0.48	5.08	5.01	4.64	4.85	4.84	93.1	92.23	92.97	92.32	92.58
9	0.37	0.42	0.38	0.40	0.38	3.74	4.20	3.80	4.04	3.83	96.9	96.43	96.77	96.35	96.41
10	0.31	0.36	0.32	0.37	0.36	3.13	3.57	3.23	3.65	3.59	100	100.0	100.0	100.00	100.00

Tabelle 40: Varianzanteile der 2-Faktorenlösungen 2010-2014 (rotiert).

Varimax-rotierte Lösungen														
Eigenwert					% Varianz					Kumuliert %				
2010	2011	2012	2013	2014	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
3.3	3.0	3.2	3.2	3.1	33.3	30.5	31.8	32.2	30.9	33.3	30.5	31.8	32.2	30.9
2.2	2.5	2.6	2.4	2.5	21.9	24.9	26.0	23.7	25.3	55.2	55.5	57.8	55.9	56.2

Ein allgemeiner Fähigkeitsfaktor klärt weiterhin 40% der Varianz auf – es gibt quasi einen grossen „gemeinsamen Nenner“ in allen Aufgabengruppen. Mit der Zwei-Faktorenlösung können nach wie vor weit über 50% der Varianz aufgeklärt werden. 2014 stellt die Aufgabe „Konzentriertes und sorgfältiges

Arbeiten“ aber einen eigenen starken dritten Faktor dar, der in der drei-Faktorenlösung dazu führen würde, dass alle übrigen den Werkzeugfunktionen zuzuordnenden Aufgabengruppen zusammenfallen. Aufgrund des Eigenwerteverlaufs scheint diese Lösung nicht angezeigt (Abbildung 49).

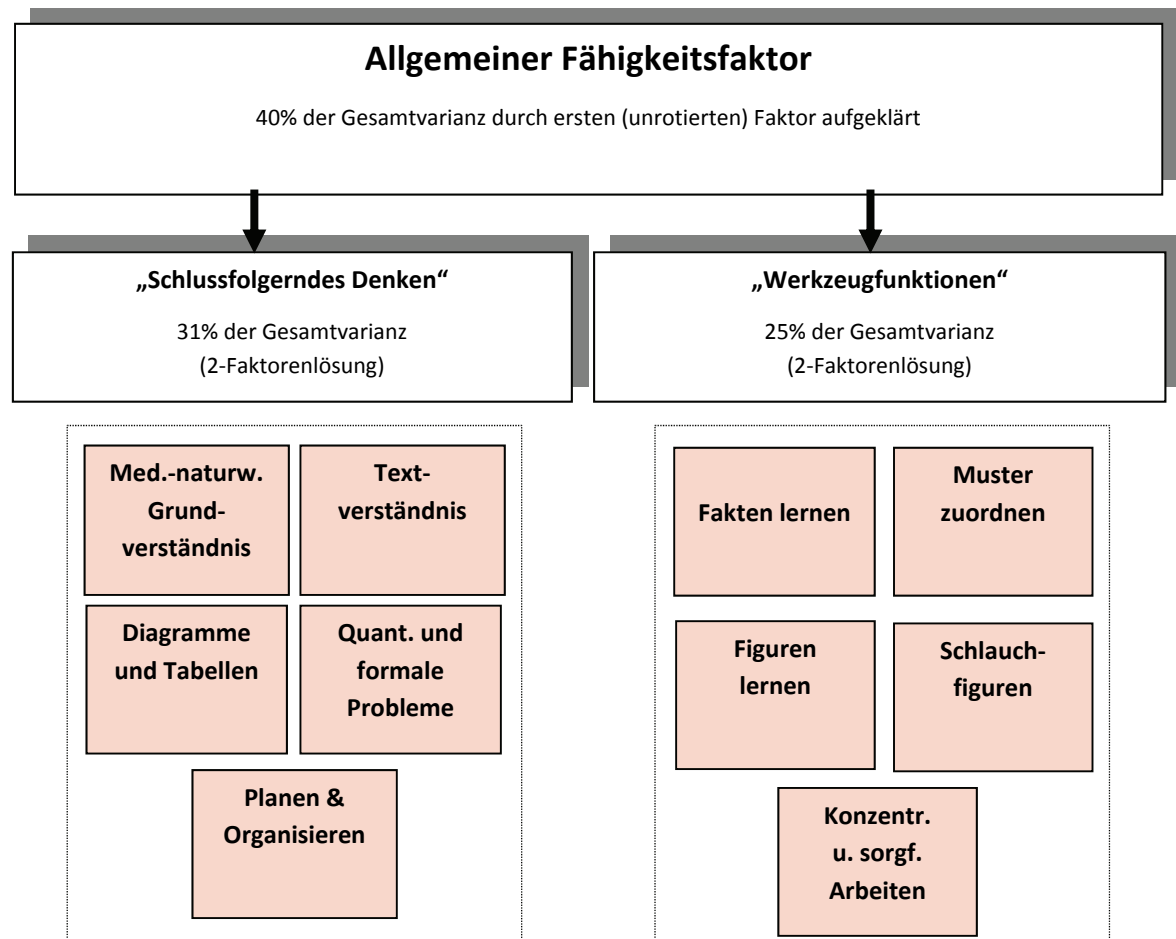


Abbildung 49: Struktur der Aufgabengruppen des EMS, hierarchische Faktorenstruktur 2014.

Aufgrund der wechselnden Anforderungen im „Konzentrierten und sorgfältigen Arbeiten“ sind derartige leichte Fluktuationen plausibel. Bis 2004 wurde die unverändert eingesetzte Version von den Kandidaten teilweise exzessiv geübt (siehe auch im entsprechenden Kapitel) – die Leistung also zunehmend stärker dem Faktor Merkfähigkeit zugeordnet. Dieser „Missstand“ sollte durch die Neugestaltung der Aufgabengruppe behoben werden, was offenbar gelungen ist.

Die Ladungen der einzelnen Aufgabengruppen im ersten unrotierten Faktor sind in Abbildung 50 dargestellt. Sie folgen den bisher bekannten Ausprägungen.

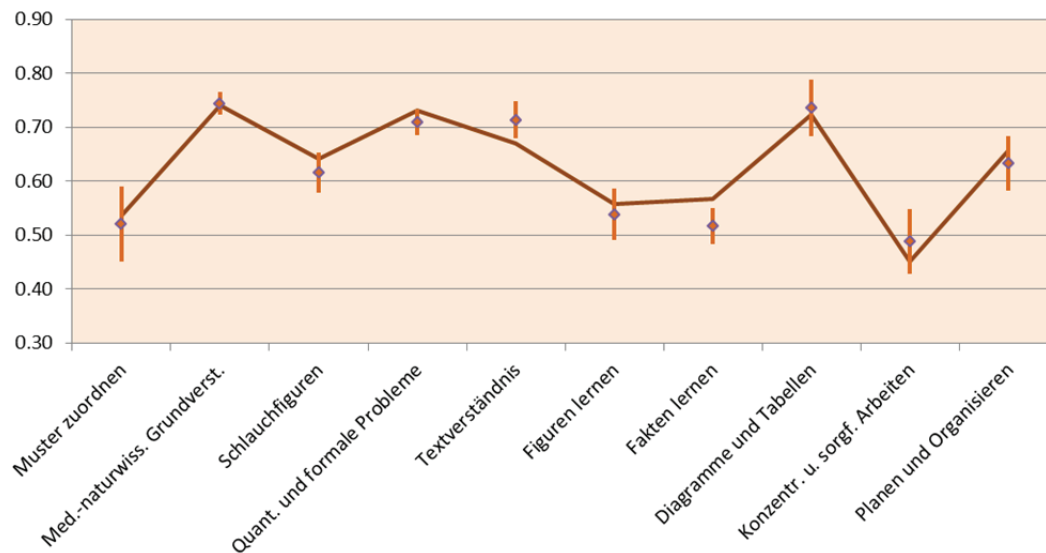


Abbildung 50: Ladungen im ersten Faktor der unrotierten Lösung („Generalfaktor“ bzw. allgemeiner Fähigkeitsfaktor) 2014 und Vergleich (Mittelwert und Standardabweichung) 1998 bis 2013.

Tabelle 41: Ladungen und Kommunalitäten der Zwei-Faktorenlösung 2009 bis 2014, varimax-rotiert.

	Faktor 1						Faktor 2						Kommunalitäten (h^2)					
	09	10	11	12	13	14	09	10	11	12	13	14	09	10	11	12	13	14
Diagramme und Tabellen	.79	.79	.77	.82	.81	.81		(.24)	(.23)				.65	.68	.65	.70	.68	.68
Med.-naturwiss. Grundverständnis	.76	.80	.80	.81	.79	.80							.61	.66	.68	.68	.66	.68
Textverständnis	.79	.77	.76	.82	.76	.74							.65	.63	.61	.68	.59	.57
Quant. und form. Probleme	.80	.74	.76	.76	.74	.77					(.21)	(.21)	.67	.57	.59	.61	.59	.63
Planen und Organisieren	.71	.71	.67	.64	.73	.71				(.25)			.51	.51	.47	.48	.54	.53
Konzentr. u. sorgf. Arbeiten		.55	(.28)	(.33)	.42		.49	.18	.50	.54		.61	.26	.34	.32	.40	.19	.38
Figuren lernen							.72	.77	.74	.75	.78	.73	.53	.60	.57	.56	.63	.55
Muster zuordnen							.69	.69	.69	.74	.73	.71	.48	.50	.49	.56	.57	.52
Fakten lernen							.67	.68	.67	.72	.75	.70	.47	.49	.48	.54	.57	.52
Schlauchfiguren	(.25)	(.25)		(.21)	(.24)	(.27)	.68	.69	.72	.73	.72	.70	.52	.53	.55	.58	.58	.56

8.3 Aufgaben-Trennschärfen

Die Aufgaben-Trennschärfen sind die Korrelationen des Punktwertes für die jeweilige Aufgabengruppe mit den zugeordneten Aufgaben. Positive Korrelationen weisen darauf hin, dass die Leistungsbesten in der jeweiligen Aufgabengruppe auch bei der entsprechenden Aufgabe die richtige Lösung bevorzugt gewählt haben. Zu beachten ist, dass bei sehr leichten und sehr schwierigen Aufgaben wegen der geringeren Antwortvarianz auch die Trennschärfe in der Regel geringer ausfallen wird. Negative Trennschärfen würden auf Aufgaben hinweisen, die entweder 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 Aufgaben von der Auswertung nachträglich ausgeschlossen und nicht gewertet, um Risiken der Fehlbewertung auszuschliessen.

Die nachfolgenden Abbildungen zeigen die Trennschärfen aller Aufgaben im Vergleich mit den Werten aus Deutschland. Unterschiede zu diesen (theoretischen) Kennwerten aus Deutschland können auch durch die Positionierung der Aufgaben innerhalb der Aufgabengruppe mitbegründet sein. So werden etwa Aufgaben am Ende einer Aufgabengruppe weniger häufig bearbeitet (aufgrund fehlender Zeit dann häufig geraten), was neben ihrer Schwierigkeit auch die Trennschärfe negativ beeinflusst.

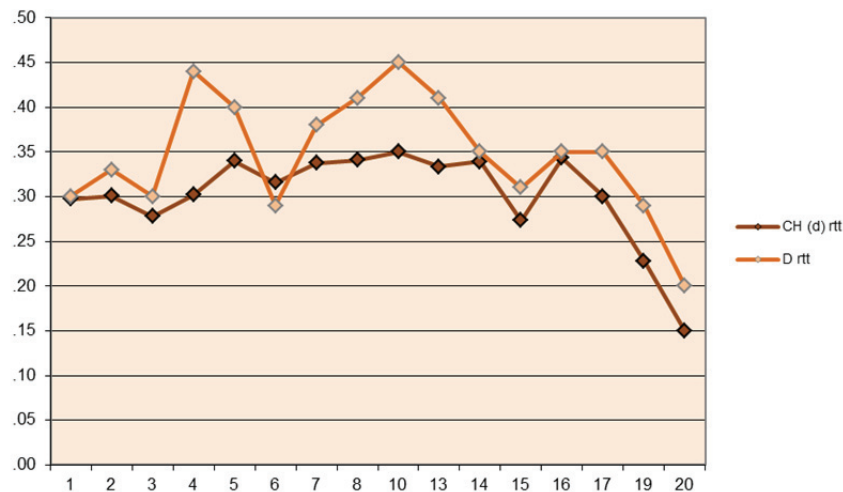


Abbildung 51: Trennschärfen für Aufgabengruppe „Quantitative und formale Probleme“.

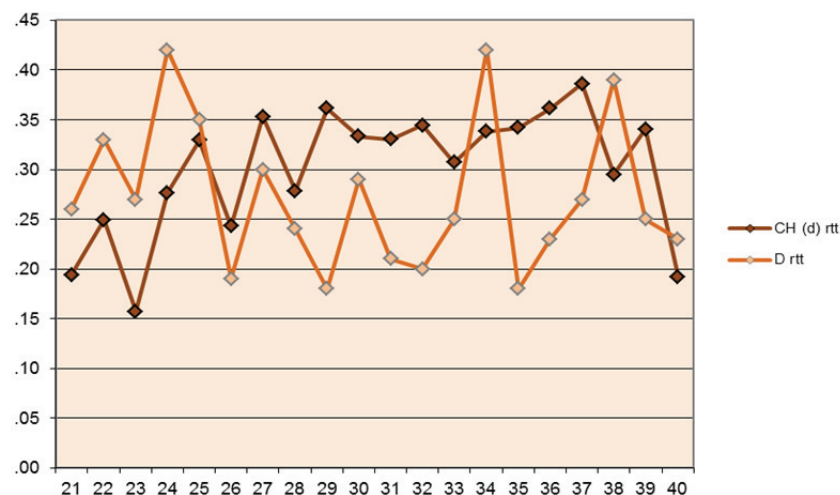


Abbildung 52: Trennschärfen für Aufgabengruppe „Schlauchfiguren“.

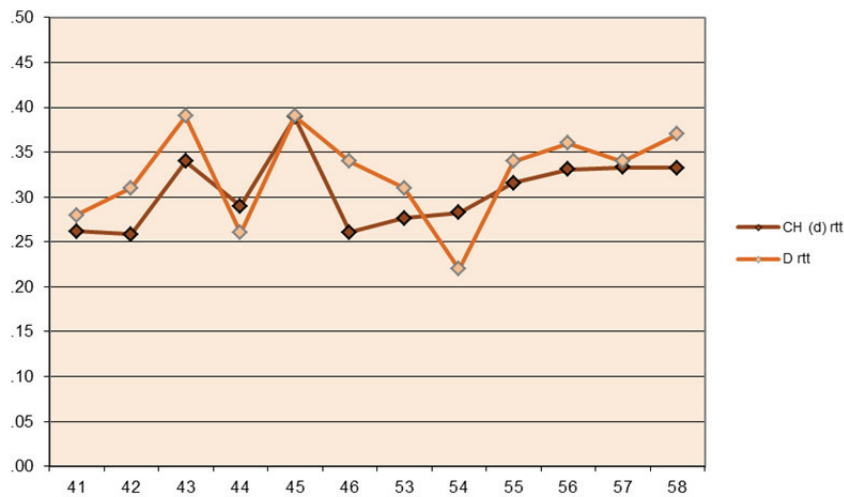


Abbildung 53: Trennschärpen für Aufgabengruppe „Textverständnis“.

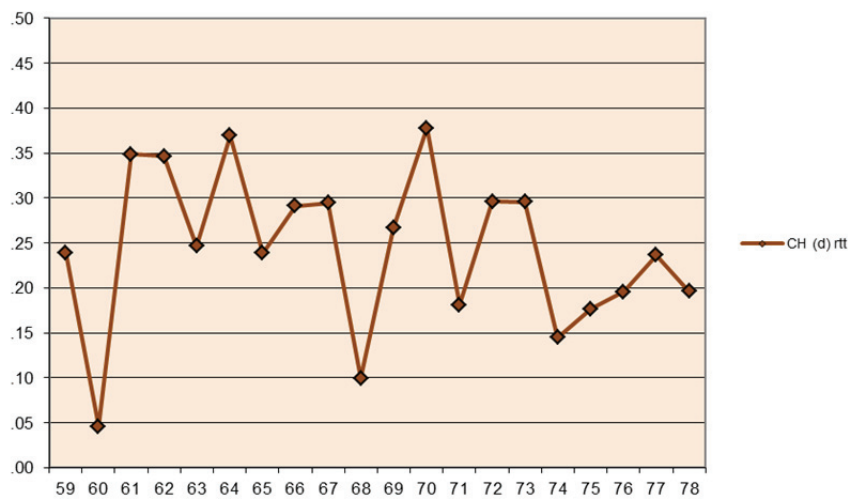


Abbildung 54: Trennschärpen für Aufgabengruppe „Planen und Organisieren“ (keine deutschen Vergleichswerte).

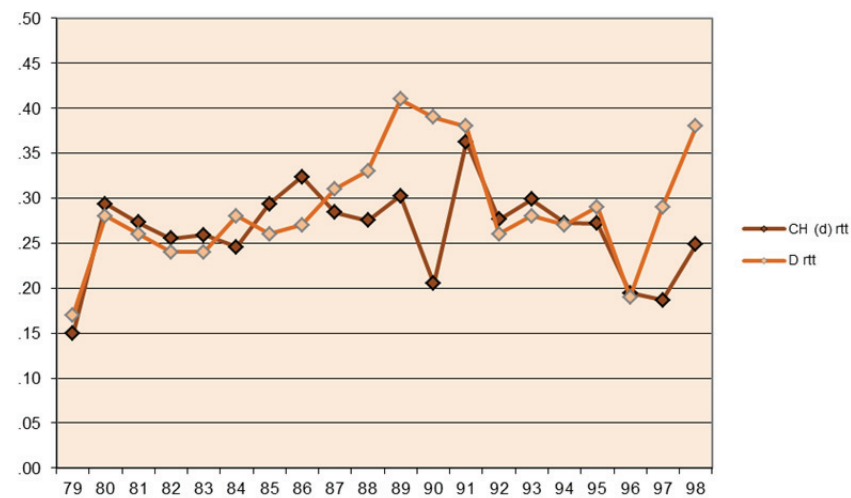


Abbildung 55: Trennschärpen für Aufgabengruppe „Medizinisch-naturwissenschaftliches Grundverständnis“.

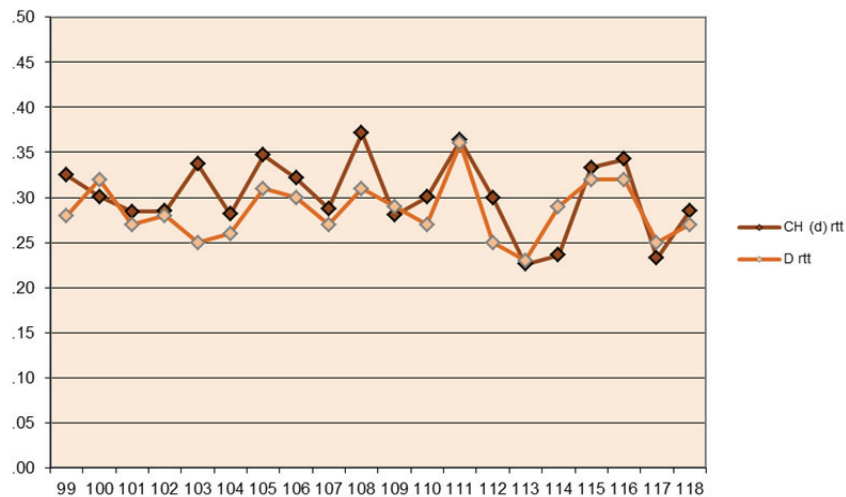


Abbildung 56: Trennschärpen für Aufgabengruppe „Figuren lernen“.

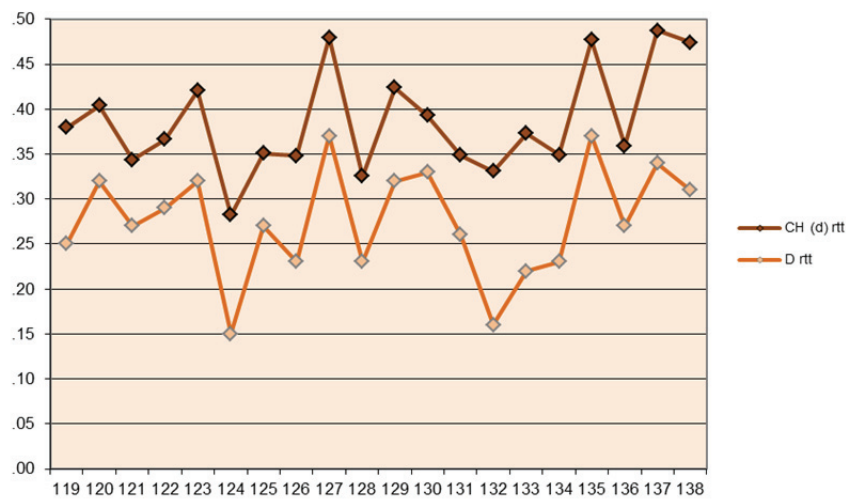


Abbildung 57: Trennschärpen für Aufgabengruppe „Fakten lernen“.

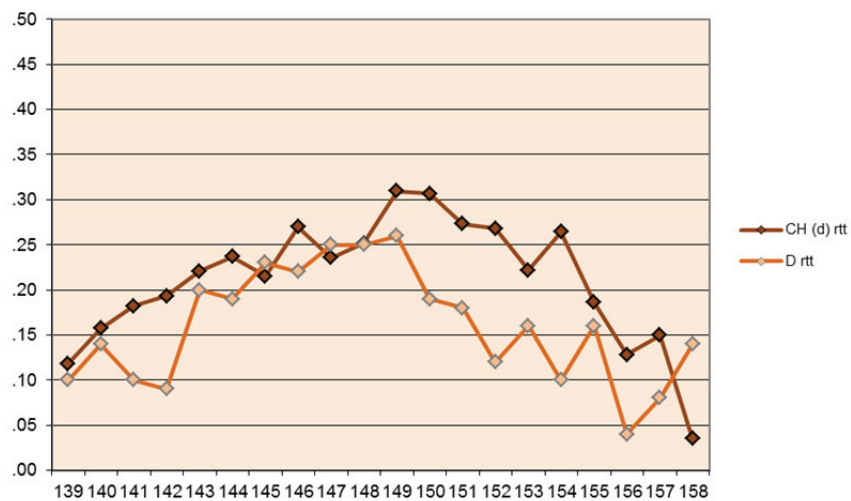


Abbildung 58: Trennschärpen für Aufgabengruppe „Muster zuordnen“.

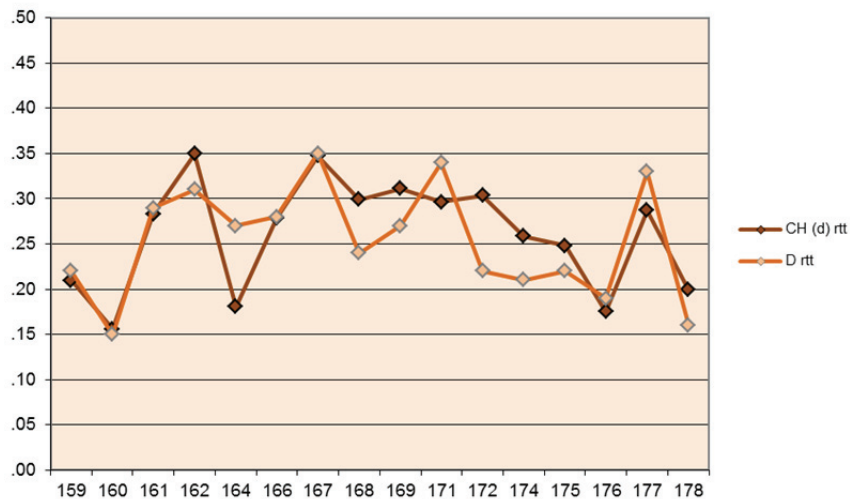


Abbildung 59: Trennschärpen für Aufgabengruppe „Diagramme und Tabellen“.

8.4 Konzentriertes und sorgfältiges Arbeiten

Diese Aufgabengruppe wurde 2004 revidiert und wird nun jährlich mit einer variierenden Anforderung durchgeführt, die vorher nicht bekannt ist. Es hatte sich früher gezeigt, dass durch exzessives Üben (40- bis teilweise 80-mal) 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. Seit 2004 findet sich eine glockenförmige Verteilung des Rohpunktwertes – 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.

Mit der Umstellung ab 2004 wurde gleichzeitig ein Paradigmenwechsel vollzogen. Der ursprünglich verwendete Test bestand aus Zeichen, für die unabhängig von den umgebenden Zeichen entschieden werden musste, ob es sich um ein Zielzeichen handelte. Mit der Einführung der neuen Versionen wurde es auch möglich, Zielzeichen beispielsweise in Abhängigkeit der umgebenden Zeichen zu definieren.

Folgende Parameter bestehen nach wie vor unverändert:

Richtige: Target richtig markiert (Wertebereich 0...400)

Fehler I: Target nicht markiert, „übersehen“ (Wertebereich 0...400)

Fehler II: Distraktor markiert, „falscher Alarm“ (Wertebereich 0...1200)

Der Punktwert (PW) berechnet sich wie folgt:

$$\text{PW} = \text{Richtige} - \text{Fehler I} - \text{Fehler II}$$

Werden mehr Fehler als Richtige Punkte erzielt, 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. Dieses Verfahren gewährleistet, dass auch unterschiedlich schwierige Konzentrationstests zu einer vergleichbaren Bepunktung über die einzelnen Jahre führen und dass der differenzierte Wertebereich nicht durch Ausreisser eingeschränkt wird.

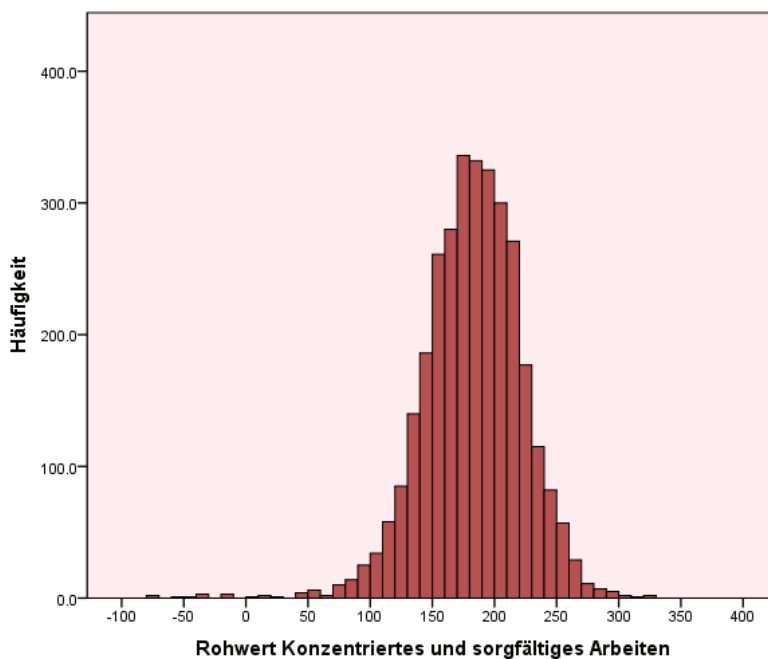


Abbildung 60: Verteilung des Rohwertes für „Konzentriertes und sorgfältiges Arbeiten“ 2014, negative Werte nicht Null gesetzt, Darstellung bis -100. 2 weitere Personen erreichten Werte unter -100.

Die Verteilung des Rohwertes ist in Abbildung 60 dargestellt. Personen mit negativen Werten (mehr Fehler als Richtige) haben die Aufgabe nicht instruktionsgemäss bearbeitet. Insgesamt differenziert die Version 2014 im gesamten Wertebereich, niemand hat das Punktemaximum (400 Rohwertpunkte) erreicht. Auslassungsfehler bleiben deutlich häufiger als Fehlmarkierungen, im Durchschnitt wurde bis Zeile 24 (von 40) gearbeitet. Ein detaillierter Vergleich für die Sprachgruppen mit Differenzierung nach Fehlertyp ist in Tabelle 42 dargestellt.

Tabelle 42: Sprachspezifische Statistiken für Parameter der Aufgabengruppe „Konzentriertes und sorgfältiges Arbeiten“ und gesamt.

		Mittelwert	Standard-abweichung	Minimum	Maximum
Deutsch	Richtige	211.8	38.2	9	346
	Fehler	29.7	28.3	0	572
	... Auslassungen (I)	27.4	25.1	0	235
	... Fehlmarkierungen (II)	2.2	9.0	0	431
	letztes bearbeitetes Zeichen	953.6	200.2	334	1595
Französisch	Richtige	199.5	36.7	110	348
	Fehler	27.4	31.4	0	237
	... Auslassungen (I)	25.9	31.1	0	237
	... Fehlmarkierungen (II)	1.6	2.3	0	22
	letztes bearbeitetes Zeichen	898.4	217.8	440	1594
Italienisch	Richtige	207.1	35.7	123	326
	Fehler	23.2	18.4	2	99
	... Auslassungen (I)	21.7	17.8	1	98
	... Fehlmarkierungen (II)	1.6	1.9	0	15
	letztes bearbeitetes Zeichen	911.9	173.1	524	1441
Gesamt	Richtige	210.8	38.1	9	348
	Fehler	29.3	28.2	0	572
	... Auslassungen (I)	27.1	25.3	0	237
	... Fehlmarkierungen (II)	2.2	8.5	0	431
	letztes bearbeitetes Zeichen	948.2	201.0	334	1595

Abbildung 61 zeigt den Scatterplot für Mengenleistung und die Genauigkeit der Bearbeitung. Es besteht ein signifikanter, aber in der Ausprägung schwacher Zusammenhang ($r=.188$) zwischen Mengenleistung und Fehlerzahl (Tabelle 43), wobei mit zunehmender Bearbeitungsgeschwindigkeit vor allem Auslassungsfehler begangen werden.

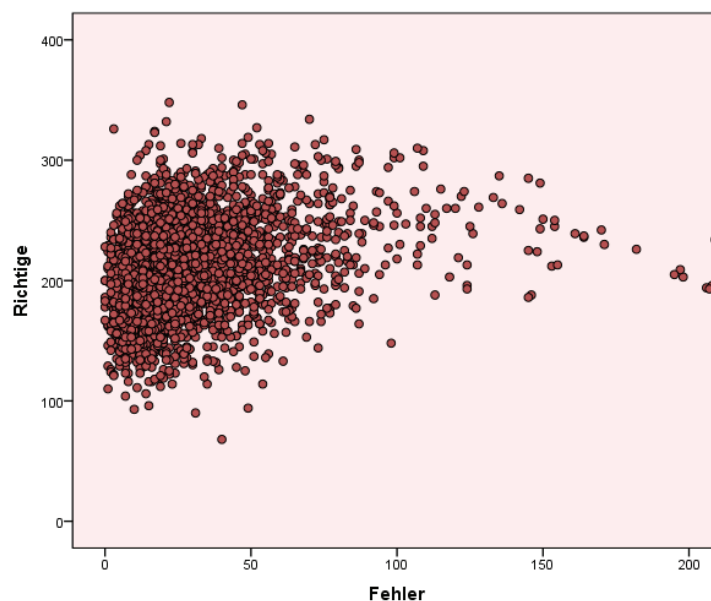


Abbildung 61: Scatterplot für alle Personen – Mengenleistung (Richtige) und Genauigkeit (Fehlerzahl, bis max 200 dargestellt).

Tabelle 43: Korrelationen der Parameter im Konzentrationstest;

	Fehler gesamt	Auslassungen (I)	Fehlmarkierungen (II)
Richtige	.188**	.225**	-.046**
Fehler gesamt		.955**	.475**
Auslassungen (I)	.955**		.194**

**Die Korrelation ist auf dem Niveau von .01 (2-seitig) signifikant.

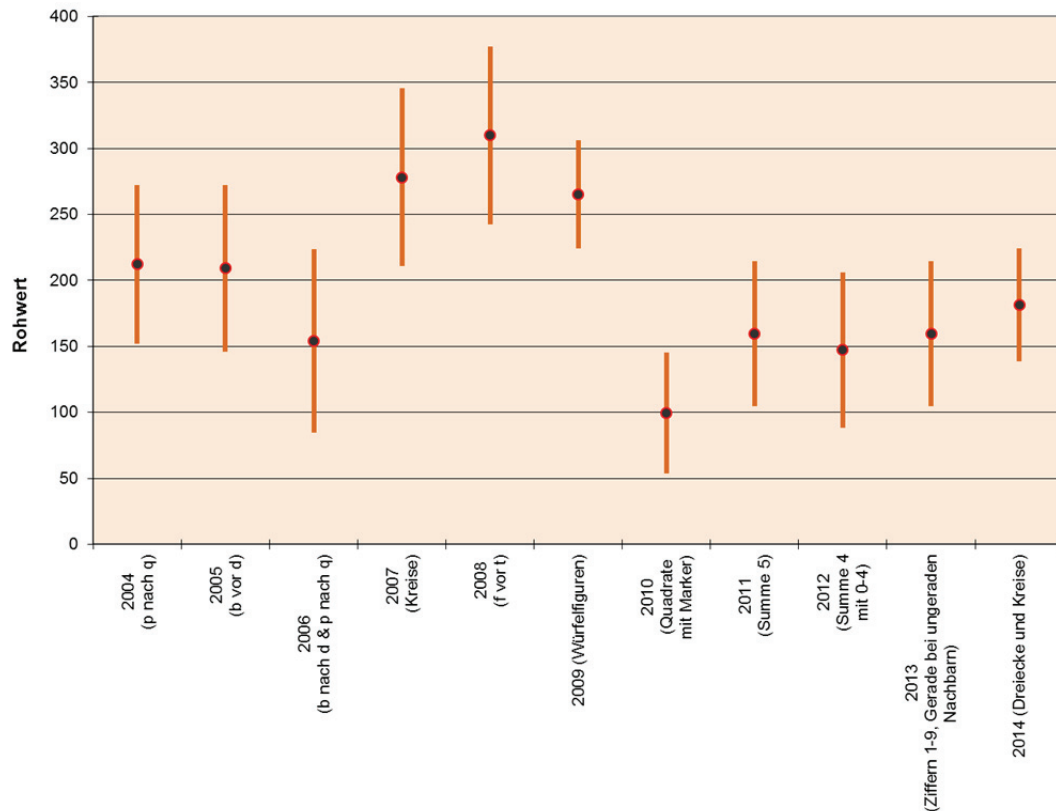


Abbildung 62: Leistungen „Konzentriertes und sorgfältiges Arbeiten“ ($m \pm s$) seit Einführung variierender Instruktionen und Zeichen (2004).

Trotz unterschiedlicher Schwierigkeiten der einzelnen Versionen dieser Aufgabe bleiben die Varianzen stabil, was eine Voraussetzung vergleichbarer Differenzierungsfähigkeit zwischen den Jahren ist. Die Version 2014 ist im Vergleich der bisher 11 Varianten dem mittleren Schwierigkeitsbereich zuzuordnen.

9 Beispielaufgaben für die Aufgabengruppen

Nachfolgend wird pro Aufgabengruppe eine Beispielaufgabe dargestellt. So können lediglich die Prinzipien der Aufgabenstruktur verdeutlicht werden. Die Aufgaben unterscheiden sich innerhalb jeder Aufgabengruppe bezüglich des Schwierigkeitsgrades und der Anforderung.

9.1 Quantitative und formale Probleme

Mit Hilfe dieses Aufgabengruppen 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 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.

1)



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 Textverständnis

Mit Hilfe dieser Aufgabengruppe 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 (T3) und Thyroxin (T4). 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 T3 und T4 als inaktive Speicherformen enthalten sind. Beim Menschen ist in den Follikeln so viel T3 und T4 gespeichert, dass der Organismus damit für etwa 10 Monate versorgt werden kann.

Das für die Hormonbildung erforderliche Jod entstammt der Nahrung und wird von den Epithelzellen als Jodid aus dem Blut aufgenommen. Die Jodidaufnahme erfolgt an der äusseren Zellmembran der Epithelzellen durch eine sogenannte Jodpumpe. Diese wird durch ein Hormon aus der Hirnanhangsdrüse, das TSH, stimuliert und kann pharmakologisch durch die Gabe von Perchlorat gehemmt werden. Ferner gibt es erbliche Schilddrüsenerkrankungen, bei deren Vorliegen die Jodpumpe nicht funktioniert.

Bei Gesunden wird das in die Epithelzellen aufgenommene Jodid im nächsten Schritt unter dem Einfluss eines Enzyms in freies Jod umgewandelt und in die Follikel abgegeben. Die Aktivität dieses Enzyms kann ebenfalls pharmakologisch gehemmt werden.

Die letzten Schritte der Hormonbildung finden in den Follikeln, also ausserhalb der einzelnen Epithelzellen, statt. In dort vorhandene sogenannte Tyrosin-Reste (des Thyreoglobulins) wird zunächst ein Jodatomb eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schliesslich T4, während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T3 hervorgeht. T3 und T4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt.

Diese Freisetzung von T3 und T4 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 T3 und T4; 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 T3 und T4 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 T3 und T4 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 T3 und T4 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 T3 führen?

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

Für die Beantwortung dieser Frage ist das Verständnis der im obigen Text unterstrichenen Stellen wichtig (im Original sind selbstverständlich keine Hervorhebungen). Der Text sagt nichts über eine Umwandlung von Jod in Jodid in den Follikeln aus, und auch der umgekehrte Prozess, die Umwandlung von Jodid in Jod, findet nicht in den Follikeln statt, sondern in den Epithelzellen. Somit gehört der Vorgang (B) nicht zu den vom Text beschriebenen Schritten, die zur Bildung von T3 führen.

9.4 Planen und Organisieren

Die Aufgabengruppe „Planen und Organisieren“ ist seit 1995 Bestandteil des EMS und prüft Fähigkeiten, die für eine effiziente Selbstorganisation im Studium wichtig sind. Er umfasst verschiedene voneinander unabhängige Szenarien, auf die sich jeweils mehrere Aufgaben beziehen. Auch bei dieser Aufgabengruppe wird kein Wissen vorausgesetzt. Es werden Aufgabenstellungen in Form so genannter Szenarien vorgegeben. Zu diesen Szenarien werden dann genau wie bei anderen Aufgabengruppen 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., 08.00 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.
- (A) Nur Aussage I trifft zu.
 (B) Nur Aussage II trifft zu.
 (C) Beide Aussagen treffen zu.
 (D) Keine der beiden Aussagen trifft zu.

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

- Der günstigste Zeitraum, in dem Ihr Freund Ihre Arbeit Korrektur lesen kann, ist die erste Hälfte der letzten Vorlesungswoche.
 - 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 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 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 Aufgabengruppen 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.

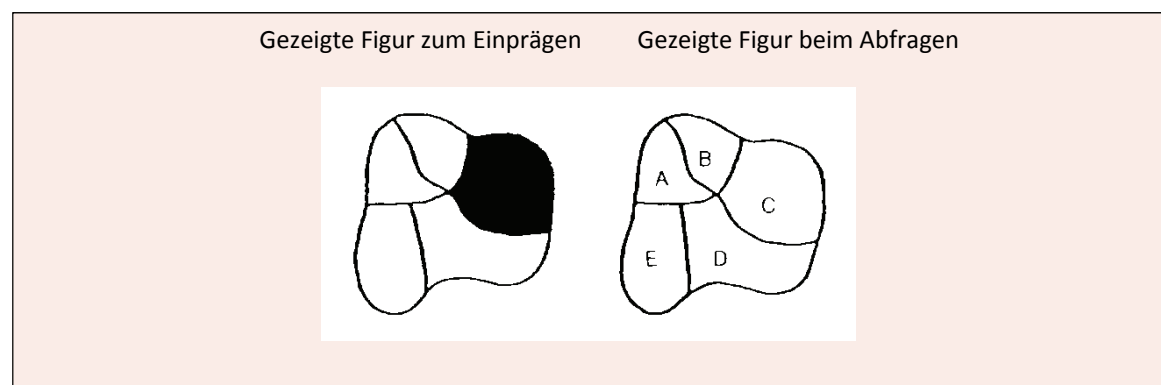
- II. Da es im Kindesalter noch offen ist, in welcher Hälfte des Gehirns die Sprachregion angelegt wird, besteht für ein Vorschulkind immer noch die Möglichkeit, die Muttersprache wieder zu erlernen. Die Sprachregion wird dann in der rechten Hälfte der Hemisphäre angelegt. Somit ist Aussage II ebenfalls 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 Figuren lernen

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

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



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 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:

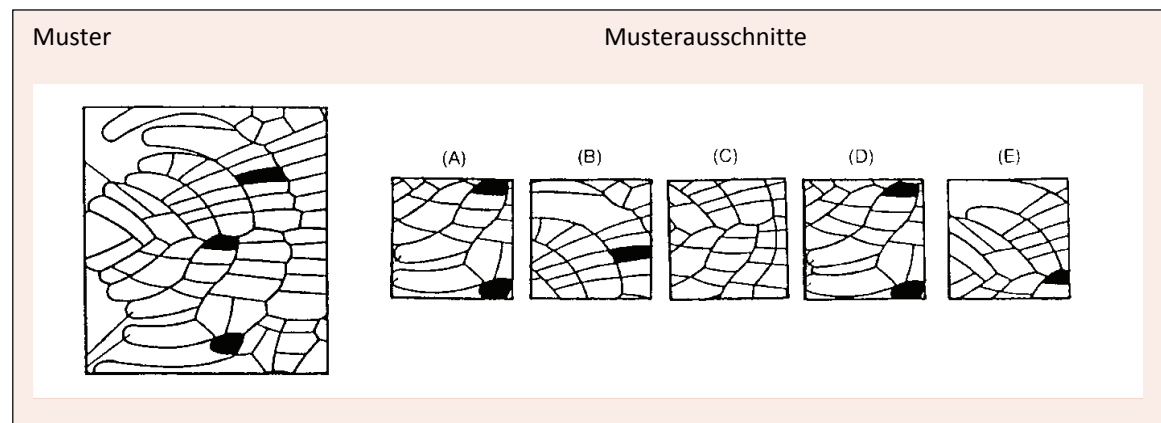
Der Patient mit dem Schädelbasisbruch ist von Beruf...

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

9.9 Muster zuordnen

In dieser Aufgabengruppe 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:



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 diese Aufgabengruppe 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 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 Aufgabengruppen „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 Aufgabengruppen 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: studienfeld-bezogene 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. (2014): Wollen wir nicht lieber diplomierte Ärzte? Sind fachliche und soziale Kompetenzen Gegensätze? *Schweizerische Ärztezeitung* 2014; 95: 6 S. 194 f.

Hänsgen, K.-D. (2013): Ohne Numerus clausus sinkt die Qualität des Studiums. Der externe Standpunkt, *NZZ* am Sonntag, 28.4. 2013, S. 17.

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.

10.1 Originaltest zur Information und Vorbereitung

Es existieren in allen drei Sprachen je zwei veröffentlichte Originalversionen.

Deutsch: www.unifr.ch/ztd/ems/vord.htm

Französisch: www.unifr.ch/ztd/ems/vorf.htm

Italienisch: www.unifr.ch/ztd/ems/vori.htm

10.2 Frühere Berichte des ZTD

Über EMS-Seite: www.unifr.ch/ztd/ems
Im Menü rechts inklusive aller Berichte und FAQ

EMS-WIKI-Blog www.ztd.ch
Kommentare und Fakten

Zusatzbericht EMS 2014: Nichtwertung der 14 Aufgaben

Inhalt

Zusammenfassung	1
Was ist geschehen – welche Massnahmen wurden geprüft und ergriffen?	1
Sind die bekannten Aufgaben auffällig hinsichtlich Schwierigkeit oder Trennschärfe?	2
Gibt es Auffälligkeiten in der Verteilung der Punktwerte?	3
Hat die Bekanntheit Einfluss auf die Bearbeitung unbekannter Aufgaben?	3
Entstanden neue Nachteile dadurch, dass 14 Aufgaben nicht gewertet wurden?	6
Haben sich durch die Nichtwertung die Gewichte im EMS bedeutsam verändert?	7
Gibt es Vorteile durch die Ähnlichkeit beim „Fakten lernen“?	7

Zusammenfassung

Beim diesjährigen Eignungstest für das Medizinstudium (EMS) hat es einen **möglichen Vorteil** für eine unbestimmte Personenzahl durch die vorherige Bekanntheit von 14 Aufgaben gegeben. Deren Ausschluss aus der Wertung war notwendig für die Wiederherstellung der Gleichbehandlung. Die vom Gesetz vorgeschriebene Feststellung der Eignung ist auf der Basis des gewerteten Teils gewährleistet, die Nichtwertung entspricht den Qualitätsstandards für Psychodiagnostik: Psychologische Tests erfordern bestmögliche Bedingungs-gleichheit für alle Personen.

Obwohl die Daten keine Hinweise auf eine bedeutsame, qualitativ oder quantitativ identifizierbare Personengruppe mit Nutzen aus diesem Vorwissen liefern, wurden weitere mögliche Vor-/Nachteile für bestimmte Teilnehmergruppen theoretisch und empirisch geprüft:

- Ein bedeutsamer Vorteil als **Zeitgewinn** bei der Bearbeitung der gewerteten Aufgaben für die Personen mit Vorkenntnis ist aufgrund der empirischen Daten und wissenschaftlicher Kenntnisse zur Diagnostik des „Schlussfolgernden Denkens“ nicht nachweisbar.
- Die Veränderung der **Gewichte** der einzelnen Aufgabengruppen am Gesamtergebnis liegt im Schwankungsbereich der bisher eingesetzten Varianten EMS. Auch der für die Studienerfolgsprognose gewichtige Faktor „Schlussfolgerndes Denken“ ist noch ausreichend repräsentiert.
- Die **Äquivalenz** (Ersetzbarkeit) der Testergebnisse aufgrund der ursprünglichen und der gekürzten Punktschichten ist für die drei betroffenen Aufgabengruppen sowie insbesondere für den Gesamtest ausreichend hoch. Die Reihung der Personen als Basis für die Zulassung bleibt zwischen beiden Versionen sehr gut vergleichbar.

Bei „**Fakten lernen**“ besteht nur eine sehr hohe Ähnlichkeit der Lernliste, die dazu gestellten Fragen sind andere. Die Ergebnisse zeigen, dass Personen, die am wahrscheinlichsten an einem Vorbereitungskurs teilgenommen haben, im Mittel sogar schlechtere Leistungen erzielen als andere. Dies ist psychologisch aufgrund des Konzeptes der Interferenz erklärbar – die Nichtwertung ist bei einer differenzierten Güterabwägung nicht notwendig. Die gewertete Version des EMS 2014 erfüllt alle Anforderungen für eine gültige Testabnahme und Zulassung aufgrund der gemessenen Eignung.

Was ist geschehen – welche Massnahmen wurden geprüft und ergriffen?

Ende Juli uns übersandte Unterlagen eines Trainingsanbieters enthalten 14 mit dem EMS 2014 identische Aufgaben aus 3 Aufgabengruppen: 1 Text der Aufgabengruppe „Textverständnis“ mit 6 Fragen, sowie je 4 Aufgaben „Quantitative und Formale Probleme“ und „Diagramme und Tabellen“. Bei der Aufgabengruppe „Fakten lernen“ gab es Ähnlichkeiten zwischen Trainingsmaterial und Original: Die Lernliste war sehr ähnlich, es wurden aber in der Abrufphase andere Fragen dazu gestellt.

Dieses Trainingsmaterial entspricht einer in Deutschland verwendeten Originalversion des TMS aus dem Jahr 1996, die so nie in der Schweiz eingesetzt wurde. Es handelt sich um ein Original oder eine entsprechende Kopie (nicht abfotografiert, rekonstruiert o.ä.). Es gibt keine Hinweise auf weitere im Umlauf befindliche Originalaufgaben. Sonstige am Markt verfügbare rekonstruierte oder nachempfundene Aufgaben der Trainingsanbieter sind damit qualitativ nicht vergleichbar – aufgrund der Marktbeobachtung in der Schweiz und auch in Deutschland gab es bisher keine Gründe, andere Originalaufgaben zurückzuziehen.

Die Schweiz adaptiert seit 1998 den deutschen TMS (drei Sprachen, Anpassung an Schweizer Gegebenheiten). Im Rahmen der kontinuierlichen Erneuerung werden einzelne Aufgaben ersetzt – so sind die Aufgaben in die Version 2014 gelangt und deshalb sind es auch nur wenige.

Die vorher bekannten identischen Aufgaben, die von einer unbekannten Personenzahl unter irregulären Bedingungen bearbeitet worden sind, wurden aus Sicherheitsgründen nicht gewertet. Die Bedingungen sind dadurch bei den gewerteten Aufgaben **hinsichtlich des Vorwissens** für alle Personen wieder gleich. Die Abbildung 1 stellt die weiteren notwendigen Prüf- und Entscheidungsschritte dar.

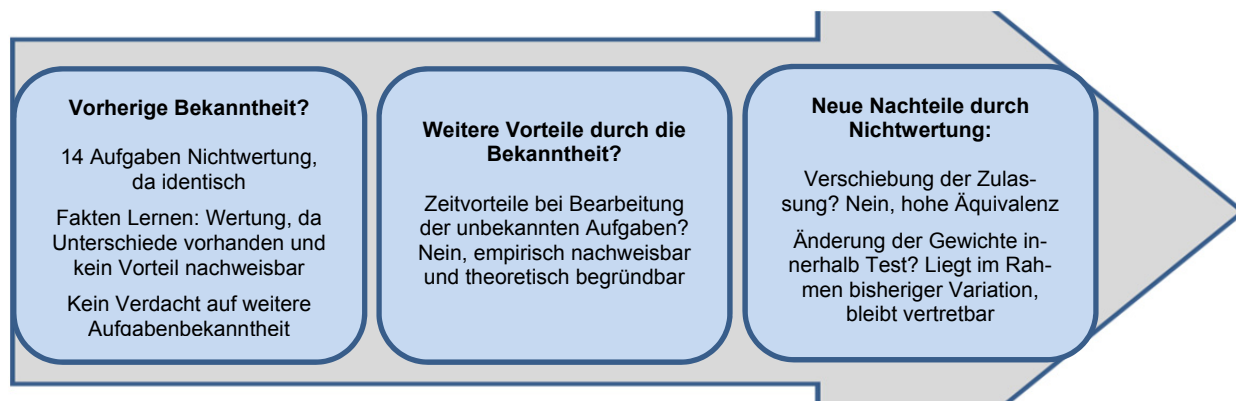


Abbildung 1: Fragestellungen für die Gewährleistung einer gültigen Auswertung 2014

Sind die bekannten Aufgaben auffällig hinsichtlich Schwierigkeit oder Trennschärfe?

Die vorher bekannten Aufgaben reihen sich bezogen auf Schwierigkeit (Lösungswahrscheinlichkeit) und Trennschärfe (Korrelation Lösungswahrscheinlichkeit mit Gesamtleistung der Aufgabengruppe) an den aufgrund der in Deutschland 1996 ermittelten Kennwerten in die jeweilige Aufgabengruppe ein, sind z.B. nicht im Mittel leichter oder mit niedrigerer Trennschärfe verbunden, weil z.B. die Beantwortung anders funktioniert hätte (Tabelle 1). Sie sind über diese vergleichsweise lange Zeit gleich schwer geblieben, was für hohe zeitliche Stabilität der gemessenen Merkmale spricht.

Tabelle 1: Schwierigkeiten (Lösungswahrscheinlichkeiten p) und Trennschärfen (Korrelation Aufgabe mit Gesamtwert r) sowie Standardabweichung (s) für betroffene Aufgabengruppen – ungewertete Aufgaben markiert

Quantitative und Formale Probleme Zeit: 50 Minuten				Diagramme und Tabellen Zeit: 50 Minuten				Textverständnis Zeit: 45 Minuten			
Nr.	p	s	r	Nr.	p	s	r	Nr.	p	s	r
1	.88	.325	.306	159	.83	.379	.207	41	.74	.436	.257
2	.80	.397	.294	160	.83	.377	.158	42	.73	.441	.252
3	.74	.440	.287	161	.78	.414	.303	43	.57	.495	.352
4	.78	.412	.297	162	.71	.456	.374	44	.59	.492	.299
5	.71	.455	.350	163	.69	.463	.375	45	.63	.482	.398
6	.64	.480	.316	164	.61	.489	.174	46	.23	.420	.261
7	.77	.418	.337	165	.67	.470	.253	47	.77	.424	.220
8	.69	.462	.330	166	.52	.500	.298	48	.60	.490	.290
9	.54	.498	.272	167	.65	.478	.352	49	.67	.472	.295
10	.69	.464	.355	168	.50	.500	.314	50	.55	.497	.202
11	.66	.475	.318	169	.59	.491	.308	51	.39	.487	.325
12	.65	.476	.304	170	.47	.499	.304	52	.51	.500	.236
13	.57	.496	.348	171	.53	.499	.312	53	.73	.442	.278
14	.53	.499	.347	172	.53	.499	.330	54	.49	.500	.281
15	.46	.498	.290	173	.42	.493	.285	55	.58	.494	.302
16	.42	.493	.359	174	.30	.459	.270	56	.41	.492	.307
17	.48	.500	.308	175	.36	.479	.264	57	.33	.469	.329
18	.35	.477	.273	176	.26	.440	.180	58	.27	.445	.314
19	.31	.463	.240	177	.29	.453	.293	Schwierigkeiten hier innerhalb der 3 Texte mit je 6 Fragen abgestuft			
20	.29	.454	.163	178	.22	.416	.188				

Gibt es Auffälligkeiten in der Verteilung der Punktwerte?

Wenn nennenswerte Teilgruppen einzelne Aufgaben unter irregulären Bedingungen (z.B. vorherige Bekanntheit) beantwortet haben, müsste sich dies als Auffälligkeit der Antwortverteilung zeigen. Häufungen bei hohen Punktzahlen (bis zu Mehrgipfligkeiten) sprächen für qualitative Unterschiede. Ist der Unterschied eher quantitativ (z.B. nur geringere relative Verbesserungen bei diesen Personen), müsste sich dies als Schiefe zeigen, weil diese Teilgruppe bessere Werte als erwartet erreicht.

Das Maximum von 14 Punkten in den 14 vorher bekannten Aufgaben (sehr gute Personen oder Personen, die optimal von Vorwissen profitierten) erreichen nur 33 Personen. Dies sind 1% der Gesamtgruppe – was die Erwartung auch unter normalen Umständen wäre. Fasst man 13 und 14 Punkte zusammen, sind es 109 Personen, was ebenfalls in der Erwartung liegt. Gäbe es eine nennenswerte Gruppe von Personen, die (fast) alle Aufgaben aufgrund der vorherigen Bekanntheit gelöst hat, müssten hier mehr Personen vertreten sein. Abbildung 2 zeigt, dass sowohl in den gewerteten (unbekannten), als auch in den ungewerteten (möglicherweise bekannten) Aufgaben die Punkteverteilung nahezu ideal einer symmetrischen Glockenverteilung folgt und auch keine Schiefe als Zeichen quantitativer Unterschiede vorhanden ist (Schiefe .00 für gewertete, .08 für ungewertete Aufgaben, Kurtosis bzw. Wölbung: -.50 und -.54). Entweder haben sehr wenige Personen dieses Material tatsächlich gesehen, oder diese haben sich die richtigen Lösungen der vorher bekannten Aufgaben nicht merken können.

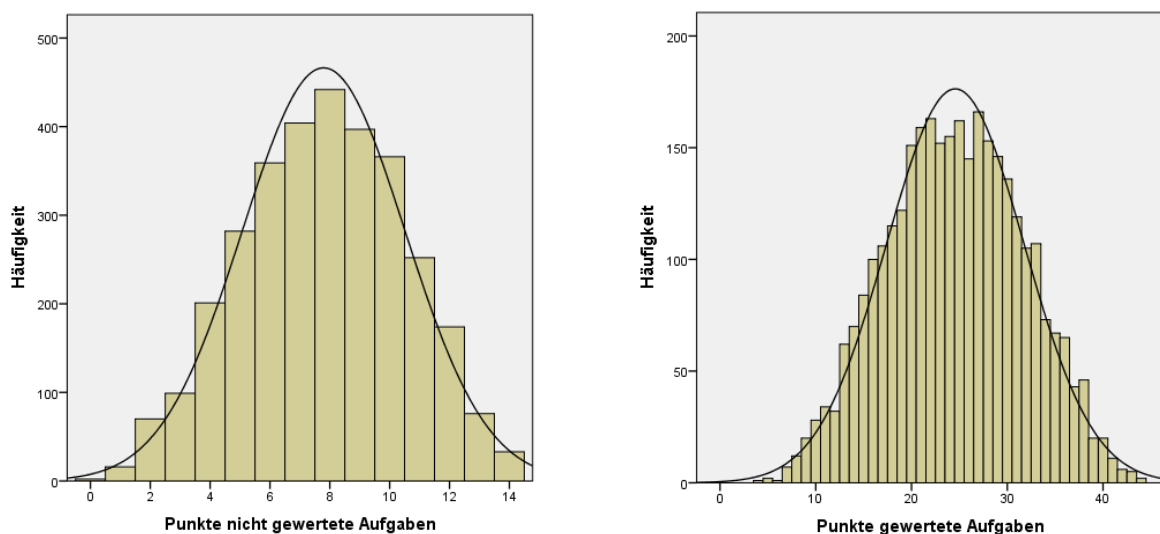


Abbildung 2: Vergleich der Punktwerteverteilung für nicht gewertete und gewertete Aufgaben der drei betroffenen Aufgabengruppe

Hat die Bekanntheit Einfluss auf die Bearbeitung unbekannter Aufgaben?

Es wurden Vorteile von Personen mit Vorkenntnissen der 14 Aufgaben auch bei den gewerteten Aufgaben der betroffenen drei Aufgabengruppen vermutet, etwa durch gewonnene Zeit. Die betroffenen Aufgabengruppen gehören allerdings zu jenen mit vergleichsweise reichlicher Bearbeitungszeit, weil dort kein zu grosser Zeitdruck gewünscht ist (2 x 50 und 1 x 45 Minuten).

Trotz fehlendem generellen Nachweis von Vorteilen bei der Verteilungsanalyse soll dem nachgegangen werden: Abbildung 3 liefert zunächst eine „visuelle“ Übersicht über die Daten. Dort wurden die Personen nach dem EMS-Gesamt-Punktwert in fünf ca. 20% umfassende Leistungs-Gruppen eingeteilt. Analog wurden fünf Klassen nach der Punktzahl nur für die ungewerteten Aufgaben gebildet, in der besten Klasse sind Personen mit 13 und 14 Punkten zusammengefasst. Gewertete und ungewertete Aufgaben sind durch den Bezug auf den gleichen Messbereich hoch korreliert, das Niveau in beiden sollte auch vom Gesamt-Punktniveau abhängen.

Die Boxplots zeigen für die Gruppen nach Punkten in den bekannten Aufgaben ein homogenes Bild: Sie steigen **zwischen** den gleichen Punktwertklassen ungewerteter Aufgaben **innerhalb** jeder Gesamt-Punktwertklasse sowie auch zwischen den Klassen der Gesamt-Punktwerte gleichförmig. Auch die „beste“ Klasse mit 13 und 14 Punkten in den bekannten Aufgaben setzt sich von den übrigen nicht ab, bei einem substantiellen Punktezuwachs durch Zeitgewinn müssten sich die Klassen mit den höchsten Punktzahlen in den ungewerteten Aufgaben „unabhängiger“ von ihrem Gesamtpunktwert darstellen und bessere Ergebnisse erreichen.

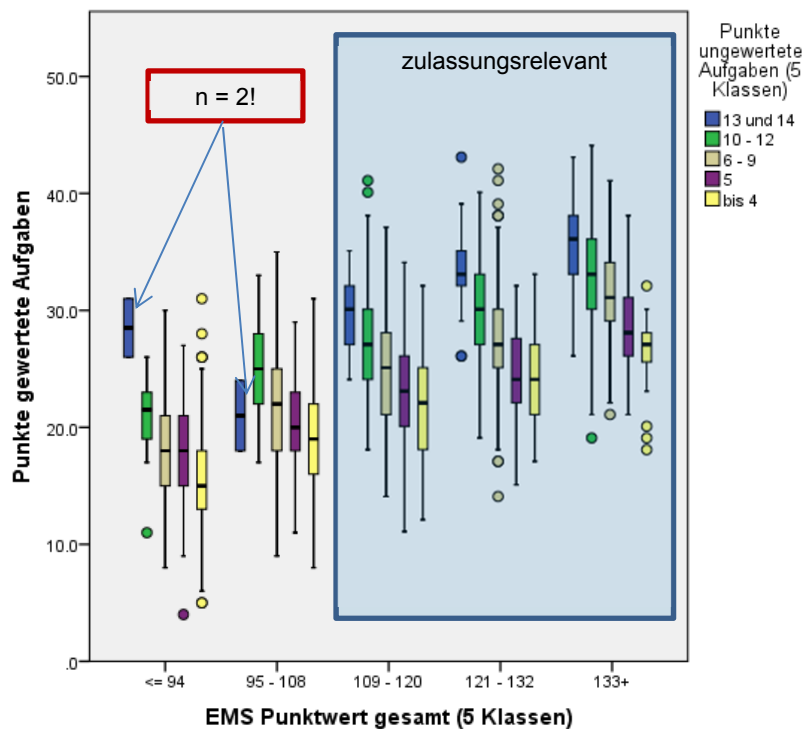


Abbildung 3: Boxplots (Mediane und Quartilabstände) der Punktschritte der bewerteten Aufgaben der drei Aufgabengruppen Diagramme und Tabellen, Textverständnis, Quantitative und formale Probleme für je 5 Punktwertklassen für Gesamt und 5 Klassen Punkte nur in den 14 ungewerteten Aufgaben. 2 Klassen sind nur mit 2 Personen besetzt!
Punkte: Ausreisser

Eine genaue Identifikation jener Personen, die von der Bekanntheit der Aufgaben profitiert haben können, ist nicht realistisch, da Unterschiede ebenso messfehlerbedingt sein können. Für die Anwendung statistischer Prüfmethode sollen stattdessen Personengruppen mit der **höchsten Wahrscheinlichkeit** eines Profitierens von der Aufgabenbekanntheit herausgefiltert und verglichen werden:

- Relative Unterschiede: Ein Weg führte über eine Regressionsanalyse. Über die Punktwerte in allen bewerteten Aufgaben des Faktors „Schlussfolgerndes Denken“ kann abgeschätzt werden, welcher Punktwert in den 14 eliminierten Aufgaben aufgrund des individuellen Leistungsniveaus jedes Kandidaten erwartet wird. 50 Personen, deren Punktschritte diesen Erwartungswert am deutlichsten übertrafen, wurden in die „**Trainingsgruppe**“ eingeteilt – dort ist die höchste Wahrscheinlichkeit gegeben, dass diese Personen von der Bekanntheit der Aufgaben profitiert haben. Nicht alle diese Personen müssen am Trainingskurs teilgenommen haben, der Unterschied kann auch zufällig bzw. messfehlerbedingt sein. Wenn es aber überhaupt einen systematischen Effekt eines Profitierens gibt, der grösser als der Messfehler ist, müssen sich die „trainierten“ Personen in dieser Gruppe gehäuft finden. Andernfalls hätte die Bekanntheit zu überhaupt keinen relevanten Vorteilen geführt. Diese Trainingsgruppe unterscheidet sich nicht von den übrigen Personen bezüglich der mittleren Gesamtleistung (113.2 Vergleichsgruppe, 113.5 Trainingsgruppe als mittlerer Punktwert) – die Differenz zwischen bewerteten und ungewerteten Aufgaben kann auf allen Leistungsstufen auftreten.
- Absolute Unterschiede: Eine zweite „**Extremgruppe 90/50**“ wird aus 49 Personen gebildet, die in den 14 ungewerteten Aufgaben mindestens einen Prozentrang von 90 erreicht haben (zu den besten 10% in diesen Aufgaben gehören, mindestens 11 Aufgaben gelöst haben), sowie im Faktor „Schlussfolgerndes Denken“ über alle bewerteten Aufgaben maximal den Prozentrang 50 erreicht haben (zu den „schlechteren“ 50% in den bewerteten Aufgaben dieses Faktors gehören). Auch hier ist die Diskrepanz zwischen bewerteten und ungewerteten Aufgaben am höchsten – es gilt die gleiche Einschränkung wie oben, dass die Unterschiede auch messfehlerbedingt sein können. Diese Extremgruppe hat bedingt durch die Auswahl durchschnittlich schlechtere Gesamtleistungen als die übrigen Personen – die Leistung in den ungewerteten Aufgaben ist „absolut“ besser.

Vergleicht man die mittleren Lösungswahrscheinlichkeiten für bewertete und ungewertete Aufgaben innerhalb der drei betroffenen Aufgabengruppen, wird deutlich (Abbildung 4): Die durchschnittlichen Lösungswahrscheinlichkeiten beider Personengruppen mit dem wahrscheinlichsten Vorwissen bewegen sich in allen drei Untertests für die bekannten Aufgaben zwischen 0.8 und 0.9 (was durch die Art der Gruppenbildung trivial ist). Bedeutsam ist, dass bei den unbekannten Aufgaben aller drei betroffenen Aufgabengruppen diese Personengruppen jeweils gleiche oder etwas geringere mittlere Punktwerte aufweisen als die Vergleichsgruppe. Das Ausmass dieser Unterschiede entspricht im Niveau genau den Differenzen in beiden nicht betroffenen Aufgabengruppen des gleichen Faktors „Schlussfolgerndes Denken“.

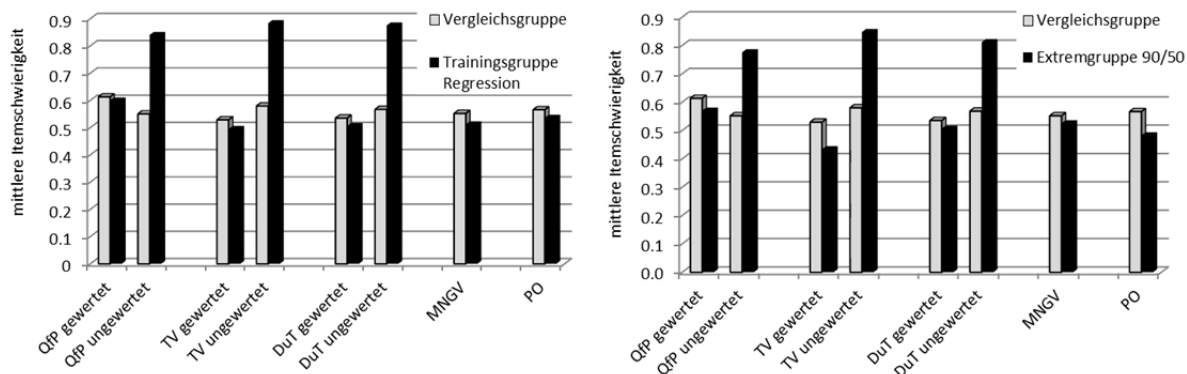


Abbildung 4: Vergleich mittlerer Schwierigkeiten gewertete und nicht gewertete Aufgaben für Gruppen aufgrund Regression und Prozenträngen; QfP: Quantitative und formale Probleme, TV: Textverständnis, DuT: Diagramme und Tabellen, MNGV: Medizinisch-Naturwissenschaftliches Grundverständnis. PO: Planen und Organisieren

Wenn die Bekanntheit einiger Aufgaben der betroffenen Aufgabengruppen einen Einfluss auf die Bearbeitung der übrigen Aufgaben derselben Aufgabengruppe hätte (z.B. durch Zeitgewinn), **müsste die Trainingsgruppe in den gewerteten Aufgaben der betroffenen Aufgabengruppen tendenziell besser abschneiden als die Vergleichsgruppe, und besser als in den beiden nicht betroffenen Aufgabengruppen dieses Faktors sein**. Die detaillierten Ergebnisse entsprechender Mittelwertsvergleiche (t-Test) für die betroffenen Aufgabengruppen sind in Tabelle 2 dargestellt und nur **für die nicht gewerteten Aufgaben wie erwartet signifikant**. (Die Signifikanz bei gewerteten Aufgaben ist wie beschrieben aufgrund der Art der Personenauswahl trivial).

Tabelle 2: Trainingsgruppe und Vergleichsgruppe (n=2780): Punktwerte in betroffenen Aufgabengruppen – signifikante Unterschiede nur bei den 14 ungewerteten Aufgaben, nicht bei den gewerteten. Signifikanzprüfung mittels t-Tests für unabhängige (vertikal) bzw. abhängige (horizontal) Stichproben

Gruppe	Mittlerer Punktwert		Signifikanz der Differenz beobachteter Wert vs. Erwartungswert	Mittlere Differenz zw. Erwartungswert und tatsächlichem Wert	Standardabweichung (beobachtet)
	beobachtet	erwartet aufgrund Regression			
Punktwert in den 14 ungewerteten Aufgaben					
Vergleichsgruppe	7.78	7.86	.023*	-0.08	2.69
Trainingsgruppe	12.16	7.57	.000**	4.59	1.63
Signifikanz Gruppenvergleich	.000**	.282		.000**	
Quantitative und formale Probleme					
Vergleichsgruppe	9.81	9.81	.948	-0.00	3.12
Trainingsgruppe	9.60	9.45	.668	0.15	3.15
Signifikanz Gruppenvergleich	.642	.237		.626	
Diagramme und Tabellen					
Vergleichsgruppe	8.56	8.56	.966	0.00	2.98
Trainingsgruppe	8.10	8.20	.741	-0.90	2.66
Signifikanz Gruppenvergleich	.277	.227		.750	
Textverständnis					
Vergleichsgruppe	6.34	6.34	.972	0.00	2.61
Trainingsgruppe	5.94	6.01	.775	-0.07	2.36
Signifikanz Gruppenvergleich	.280	.086		.792	

*: Signifikant auf dem 5%-Niveau; **: Signifikant auf dem 1%-Niveau

Für alle Vergleiche auf Ebene der Aufgabengruppen gilt: Sowohl der mittlere Punktwert wie auch die mittlere Differenz zum Erwartungswert unterscheiden sich **nicht signifikant**. Auch in der „Trainingsgruppe“ entsprechen die Punktwerte in den gewerteten Aufgaben den **durch das individuelle Leistungsniveau erwarteten Werten**. Für Zeitvorteile durch die Bekanntheit der Aufgaben finden sich hier keine Belege.

Dies wird auch durch theoretische Konzepte der Psychologie gestützt: Bei Aufgaben des Schlussfolgernden Denkens steht Zeit in keiner linearen Beziehung zur Leistung – begrenzend ist vielmehr die zeitlich stabilere Fähigkeit als dispositionelles Merkmal. Einige vergleichbare Tests („Niveau- oder

Powertests“) arbeiten ganz ohne Zeitbegrenzung und die Ergebnisse differenzieren dennoch nach der Fähigkeit. Übersteigt die Schwierigkeit einer Aufgabe die persönliche Fähigkeit, wird entweder geraten oder auch eine falsche Antwort gegeben. Die Personen kommen innerhalb der verfügbaren Zeit an die Grenze, bei welcher die Aufgabenschwierigkeit die eigene Fähigkeit übersteigt. Das Konzept der ansteigenden Schwierigkeit findet sich auch bei den drei betroffenen Aufgabengruppen.

Es wurde auch geprüft, ob **einzelne** weitere Aufgaben der Version 2014 in einem Trainingsmaterial des Anbieters enthalten und damit bekannt waren. Vergleiche einzelner Aufgabenschwierigkeiten der Trainingsgruppe mit den übrigen Personen zeigt, dass sich nur die 14 ungewerteten Aufgaben aus den betroffenen Aufgabengruppen wie erwartet signifikant unterscheiden (Abbildung 5). Würden einzelne weitere Aufgaben ebenfalls bekannt gewesen sein, müssten sie gleichfalls diesem Muster folgen. Diese Analyse wurde auch für alle anderen Aufgabengruppen des EMS 2014 vorgenommen – keine weitere Aufgabe weist derartige signifikante Differenzen auf. Dieser Befund bezieht sich auf diesen einen Trainingsanbieter. Hinweise auf weitere identische Aufgaben anderer Anbieter liegen nicht vor und sind aus verschiedenen Gründen auch weniger wahrscheinlich.

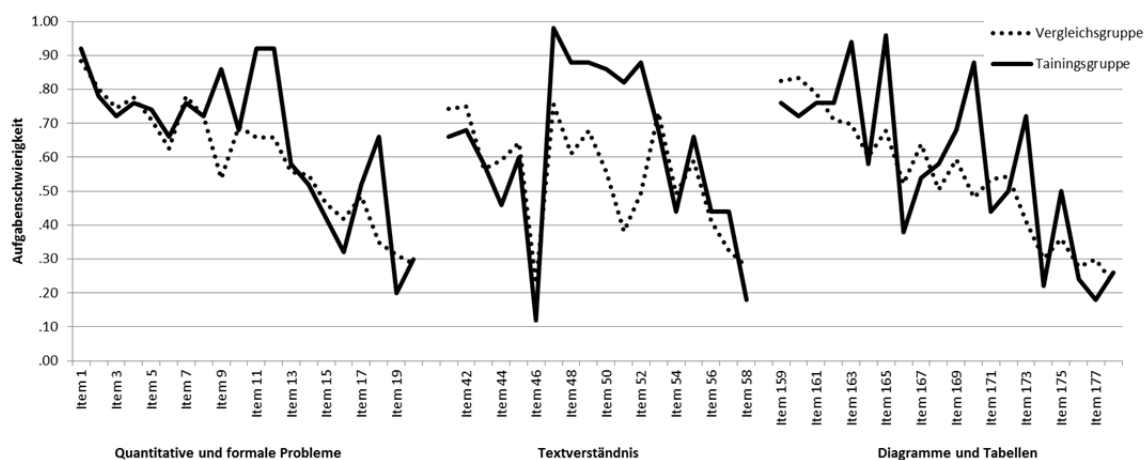


Abbildung 5: Schwierigkeiten der Aufgaben für die Trainingsgruppe im Vergleich zu den anderen Teilnehmenden

Zusammenfassend finden sich keine Hinweise dafür, dass eine Gruppe mit Vorkenntnissen der Aufgaben weitere Vorteile aus der Aufgabenbekanntheit gezogen hätte (etwa durch Zeitvorteile). Die Nichtwertung der Aufgaben hat die möglichen Vorteile korrigiert (selbst wenn auch deren Nachweis nicht wirklich evident erscheint).

Entstanden neue Nachteile dadurch, dass 14 Aufgaben nicht gewertet wurden?

Der Einwand, dass es Personen gibt, die in den 14 Aufgaben besser waren und sich durch die Nichtwertung **systematisch** verschlechtert hätten, gilt nur für jene, die von der Bekanntheit profitiert hätten. Die Leistungen in den gewerteten und ungewerteten Aufgaben nach Aufgabengruppen sind konstruktionsbedingt hoch korreliert. Die Rangreihe nach der Leistung muss sich gleichermassen zeigen, alle nicht systematischen Abweichungen sind aus Sicht der psychologischen Theorie Messfehler. Dies muss so sein, weil alle Einzelaufgaben das gleiche Merkmal messen – nicht wie bei Prüfungen inhaltlich unterschiedliche Facetten eines Themengebietes. Zwischen den Aufgaben gibt es keinen inhaltlichen Bezug, sondern nur formal gleiche Anforderungen (z.B. Textverständnis).

Tabelle 3: Äquivalenzschätzung der Punktwerte für gesamte und verkürzte Aufgabengruppen

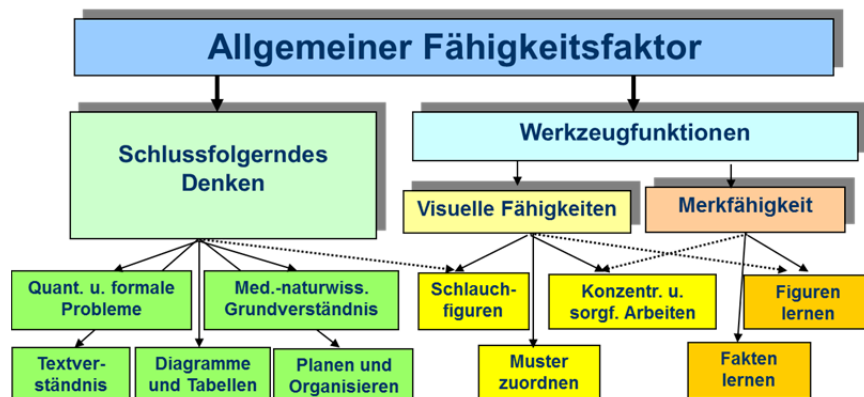
Gültige Version	Ursprüngliche Version	Textverständnis 18 Aufgaben	Quantitative und formale Probleme 20 Aufgaben	Diagramme und Tabellen 20 Aufgaben
Textverständnis (12 Aufgaben)	Korrelation Pearson N	.913** 3173	.474** 3173	.508** 3173
Quantitative und formale Probleme (16 Aufgaben)	Korrelation Pearson N	.499** 3173	.967** 3173	.637** 3173
Diagramme und Tabellen (16 Aufgaben)	Korrelation Pearson N	.530** 3173	.623** 3173	.964** 3173

** : Signifikant auf dem 1%-Niveau

Die **Paralleltest-Reliabilität** prüft die **Äquivalenz von 2 Messungen**. Sind die Messungen mit allen und nur den gewerteten Aufgaben äquivalent, **kann eine Messung die andere ersetzen**, beide Messungen differenzieren die Personen dann auf gleiche Weise. Für die drei betroffenen Aufgabengruppen korrelieren die Summen der ursprünglichen mit den gewerteten Aufgaben mit 0.91 bis 0.97 (Tabelle 3). Reliabilitäten ab 0.85 werden als hoch eingeschätzt – da eine Teilmenge mit einem Ganzen verglichen wird, muss der Wert hier etwas höher sein. In allen drei Fällen liegt eine ausreichende Äquivalenz vor, die gewerteten Aufgaben messen jeweils das gleiche wie ursprünglich alle Aufgaben. Entscheidender ist allerdings die Äquivalenz des zulassungsrelevanten Gesamttests: **Für den gesamten Test erreicht die Äquivalenz für den Punktwert sogar 0.996, für den Test-Prozentrang 0.995 und für den mittleren Rangplatz 0.996. Dies ist mehr als ausreichend für die Annahme einer vergleichbaren Rangreihe nach der Studieneignung und einer vergleichbaren Zulassung.**

Haben sich durch die Nichtwertung die Gewichte im EMS bedeutsam verändert?

Die drei betroffenen Aufgabengruppen wie auch die nicht betroffenen Aufgabengruppen „Medizinisch-Naturwissenschaftliches Grundverständnis“ und „Planen und Organisieren“ gehören zum wichtigen Faktor „**Schlussfolgerndes Denken**“. Dieser varianzstärkste Faktor ist normalerweise mit 98 Aufgaben vertreten, jetzt noch mit 84. Bis 2004 umfasste er nur 78 Aufgaben (wie noch heute im deutschen TMS). Die Gewichtung des Faktors ist also immer noch höher als in den ersten Formen des EMS.



Es wurde erwogen, die Resultate der betroffenen Aufgabengruppen proportional auf die vollen Punktzahlen hochzurechnen. Das wären aber „neue

Spielregeln während des laufenden Spiels“ gewesen. Die Varianz wäre „künstlich“ erhöht worden, es hätten Zwischenwerte gefehlt und durch eine einzige Aufgabe hätte man plötzlich z.B. 1.2 Punkte erhalten oder verloren. Deshalb wurde davon abgesehen.

Gibt es Vorteile durch die Ähnlichkeit beim „Fakten lernen“?

Wirken sich die Unterschiede der Lernliste und die anderen dazu gestellten Fragen auf die Leistungen aus? Analog Abbildung 3 werden dazu in Tabelle 4 die Punktwerte gesamt und in den vorher bekannten Aufgaben hinsichtlich des Punktwerts im „Fakten lernen“ untersucht. In jeder Klasse der nicht gewerteten Aufgabenpunkte steigt der Mittelwert für „Fakten lernen“ mit steigendem Gesamtpunktwert erwartungsgemäss an, da die Leistungen auch korrelieren.

Bei den besten 60% (ab 109 Punkte) liegt das jeweilige Maximum des Mittelwerts (markiert) nie in der Klasse mit der höchsten Punktzahl in den nicht gewerteten Aufgaben (dies müsste der Fall sein, wenn diese Gruppe besonders profitiert hätte).

Tabelle 4: Mittelwerte (m) und Standardabweichungen (s) sowie Personenzahl (n) in der Aufgabengruppe „Fakten lernen“ für je 5 Punktwertklassen und 5 Klassenpunkte in den 14 nicht gewerteten Aufgaben

Klassen Gesamtpunkte	Klassen Punkte in den 14 nicht gewerteten Aufgaben														
	13 und 14			10 bis 12			6 bis 9			5			4 und weniger		
	n	m	s	n	m	s	n	m	s	n	m	s	n	m	s
<= 94	2 (!)	n.b.	n.b.	20	10.5	2.7	179	10.9	3.6	109	10.8	3.5	337	10.5	3.8
95 - 108	2 (!)	n.b.	n.b.	69	12.1	3.2	278	13.2	3.3	111	14.0	3.2	163	13.5	3.8
109 - 120	9	12.1	4.0	166	13.9	3.4	331	14.8	3.5	61	15.8	2.8	93	15.5	3.3
121 - 132	26	14.4	4.3	224	15.8	3.4	290	16.6	2.8	52	18.0	2.2	49	17.7	2.4
133+	70	17.6	2.9	313	17.7	2.6	165	18.1	2.2	26	18.6	1.6	28	18.5	2.1

n.b.: Nicht berechnet, da nur 2 Personen und nicht zulassungsrelevant

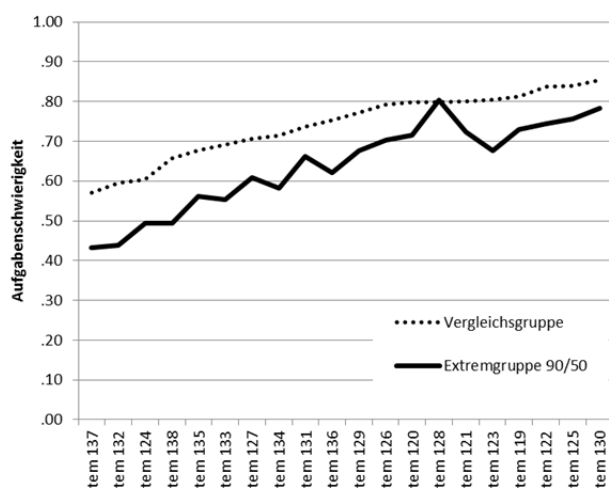


Abbildung 6: Vergleich der Schwierigkeiten der einzelnen Aufgaben bei „Fakten lernen“ für Extrem- und Vergleichsgruppe

Analog des auf Seite 3 beschriebenen Verfahrens der Bildung der „Extremgruppe 90/50“ wird für die Aufgabenanalyse von „Fakten lernen“ eine Extremgruppe aus 148 Personen gebildet, die mindestens 11 der 14 ungewerteten Aufgaben korrekt beantwortet haben und gleichzeitig maximal Prozentrang 50 im Faktor „**Werkzeugfunktionen**“ aufweisen (welchem „Fakten lernen“ zuzuordnen ist). Dies sind Personen, bei denen die Leistung bei „Fakten lernen“ am wahrscheinlichsten über der Erwartung liegt – mit gleichen Einschränkungen wie oben, dass dies auch messfehlerbedingt sein kann. Aus diesem Vergleich ergibt sich **kein Hinweis auf einen Vorteil dieser Extremgruppe** bei der Bearbeitung der Aufgabengruppe „Fakten lernen“. Wie erwartet sind diese Werte schlechter (Abbildung 6).

Weiterhin wird auch für die durch Regressionsanalyse ermittelte Trainingsgruppe (siehe Seite 4) deutlich, dass der Unterschied in der Leistung beim „Fakten lernen“ ebenfalls nicht signifikant ist (Tabelle 5) – für diese Gruppe unterscheiden sich die mittleren Gesamtleistungen bekanntlich ebenfalls nicht. Bei der Bildung dieser Gruppe ist „Fakten lernen“ nicht berücksichtigt worden, der fehlende signifikante Unterschied ist daher besonders aussagefähig.

Tabelle 5: Vergleichsgruppe und Trainingsgruppe aufgrund Regression für „Fakten lernen“. Die Signifikanzprüfung erfolgte mittels t-Tests für unabhängige (vertikal) bzw. abhängige (horizontal) Stichproben

Gruppe	Mittlerer Punktwert		Signifikanz der Differenz beobachteter Wert vs. Erwartungswert	Mittlere Differenz zw. Erwartungswert und tatsächlichem Wert	Standardabweichung (beobachtet)
	beobachtet	erwartet aufgrund Regression			
Fakten lernen (auch hier kein signifikanter Unterschied – für alle 20 Aufgaben)					
Vergleichsgruppe	14.69	14.70	.873	-0.01	4.10
Trainingsgruppe	15.50	14.93	.226	0.6	4.02
Signifikanz Gruppenvergleich	.167	.490		.229	

Die Aufgabengruppe gehörte von Beginn des EMS an zu den leichteren, wobei häufiger hohe bzw. volle Punktzahlen erreicht wurden und die Varianz der Punktwerte niedriger war. Insofern hätten sich Personen mit Vorkenntnissen auch überhaupt weniger von den übrigen absetzen können.

Da sich keine Vorteile für Personengruppen nach möglicher Vorkenntnis nachweisen lassen (sondern in einer Analyse sogar Nachteile) ist die Wertung und Beibehaltung dieser Aufgabengruppe im Rahmen einer Güterabwägung gerechtfertigt.

Eine mögliche psychologische Erklärung, warum Personen mit „Vorkenntnissen“ sogar etwas schlechter sind, sei kurz dargestellt: Das Umlernen einer vorhandenen Lernliste ist in der Regel schwerer als das Neulernen. Beim Abruf der gelernten Dinge kann es zu sogenannten proaktiven Interferenzen zwischen ursprünglicher und veränderter Liste kommen, indem diese nicht ausreichend unterschieden werden können. Die Vorgabe der ähnlichen Liste reduziert die Reproduktionsleistung der ursprünglichen Liste. Underwood (1957) hat dies in dem Satz zusammengefasst „Je mehr {Ergänzung: und auch je intensiver} Listen vorher gelernt wurden, desto geringer ist die Wahrscheinlichkeit, die neue Liste korrekt zu erinnern“. Danach haben Personen, die ähnliche Listen vorher gelernt haben, sogar Nachteile gegenüber Personen, die eine Liste neu lernen, was die Ergebnisse erklären würde.