swissuniversities

swissuniversities
Effingerstrasse 15, Postfach
3001 Bern
www.swissuniversities.ch

Anmeldung und Zulassung zum Medizinstudium

Durchführung des Selektionsverfahrens 2024

Mantelbericht swissuniversities zuhanden der Schweizerischen Hochschulkonferenz

Inhaltsverzeichnis

1. Auftrag und Dank 3
2. Studienplatzkapazitäten und Anmeldung zum Medizinstudium 3
   - 2.1. Studienplatzkapazitäten 3
   - 2.2. Anmeldungen per 15. Februar 3
3. Testteilnahmen Eignungstest für das Medizinstudium 2024 4
   - 3.1. Testteilnahmen gesamt und nach Disziplin 4
   - 3.2. Testteilnahmen nach Geschlecht und Disziplin 5
   - 3.3. Testteilnahmen nach Kanton 6
4. Testdurchführung / Testergebnisse 7
5. Abschluss des Zuteilungsverfahrens 7
   - 5.1. Zulassungen pro Disziplin 7
   - 5.2. Umleitungen 8
   - 5.3. Verzicht auf einen Studienplatz 8
   - 5.4. Altersstruktur 9
6. Follow-up Evaluation Zentrum für Testentwicklung und Diagnostik (ZTD) 9

Anhang

Anhang 1: Wissenschaftlicher Bericht des Zentrums für Testentwicklung und Diagnostik der Universität Freiburg über die Durchführung des EMS 2024

Abkürzungen

CP	Chiropraktik
EMS	Eignungstest für das Medizinstudium
HM	Humanmedizin
NC	Numerus clausus
SHK	Schweizerische Hochschulkonferenz
VM	Veterinärmedizin
ZM	Zahnmedizin
ZTD	Zentrum für Testentwicklung und Diagnostik der Universität Freiburg

1. Auftrag und Dank

Der Hochschulrat der Schweizerischen Hochschulkonferenz (SHK) genehmigte am 25. November 2021 die Grundsätze und die Eckwerte für das Zulassungsverfahren zum Medizinstudium. Er beauftragte swissuniversities, ihm jährlich über die Durchführung des Selektionsverfahrens Bericht zu erstatten. Zentraler Bestandteil des Berichts von swissuniversities ist der wissenschaftliche Bericht des Zentrums für Testentwicklung und Diagnostik (ZTD) der Universität Freiburg über die Durchführung des Eignungstests für das Medizinstudium (EMS).

Anmeldezahlen, verfügbare Studienplätze, Verteilung nach Geschlechtern, Sprachen und Disziplinen werden im Mantelbericht von swissuniversities dargestellt. Der Bericht des ZTD hingegen fokussiert auf die Erstellung, Anwendung und Auswertung des EMS.

swissuniversities dankt allen an der Testdurchführung beteiligten Personen, insbesondere den Mitarbeitenden des ZTD und der beteiligten Hochschulen, für ihren Einsatz und die gute Zusammenarbeit.

2. Studienplatzkapazitäten und Anmeldung zum Medizinstudium

Die Anzahl Studienplätze in der Medizin (Humanmedizin inkl. Chiropraktik, Zahn- und Veterinärmedizin) ist insbesondere aufgrund der hohen Kosten des Medizinstudiums und der maximal verfügbaren praxisbezogenen Ausbildungsplätze in den Spitälern und Kliniken beschränkt. Die Zahl der Anmeldungen für ein Medizinstudium an Schweizer Hochschulen übersteigt die Studienplatzkapazitäten seit Jahren deutlich.

Die Selektion der Studierenden wird unterschiedlich durchgeführt: An den Universitäten Genf, Lausanne und Neuenburg findet die Selektion nach dem ersten Studienjahr mittels «Concours» statt. Die Universitäten Basel, Bern, Freiburg, Zürich (inkl. der Tracks Zürich-Luzern und Zürich-St. Gallen), die Università della Svizzera italiana und die ETH Zürich beschränken den Zugang zum Medizinstudium (Numerus clausus) mittels Eignungstest für das Medizinstudium (EMS).

2.1. Studienplatzkapazitäten

Auf das Studienjahr 2024/2025 erhöhte in der Humanmedizin die Universität Basel die Studienplatzkapazität im ersten Jahr des Masterstudiums gegenüber dem Vorjahr um fünf Plätze. Die Universität Zürich schaffte im Bachelor acht und im Master fünf zusätzliche Plätze. In der Zahnmedizin bleiben die Kapazitäten 2024/2025 unverändert. In der Veterinärmedizin gab es eine Erhöhung um vier Studienplätze auf Masterstufe an der Universität Bern.

In der Humanmedizin standen somit im 1. Jahr des Bachelorstudiums schweizweit 2’195 Studienplätze zur Verfügung. im 1. Jahr des Masterstudiums 1’470. In der Veterinärmedizin und in der Zahnmedizin waren es im Bachelor je 172 und im Master je 140 Studienplätze.

2.2. Anmeldungen per 15. Februar

Bis zur Frist vom 15. Februar 2024 gingen insgesamt 6’678 Anmeldungen für alle Hochschulen und Disziplinen ein. Es gab somit etwas weniger Anmeldungen als im Vorjahr. Die Zahl der Anmeldungen war in allen Disziplinen deutlich höher als die Kapazität.

Abbildung 1 zeigt die Entwicklung der Anmeldungen, Stand 15. Februar, an den Hochschulen ohne Numerus clausus (Universitäten Genf, Lausanne und Neuenburg) über 10 Jahre.

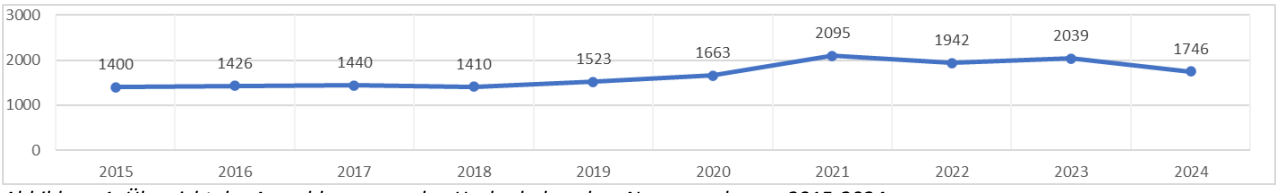


Abbildung 1: Übersicht der Anmeldungen an den Hochschulen ohne Numerus clausus 2015-2024

Die Anmeldungen an den Hochschulen mit Numerus clausus (Universitäten Basel, Bern, Freiburg, Zürich, Università della Svizzera italiana und ETH Zürich) betrugen 2024 für alle Disziplinen zusammengenommen 343% der vorhandenen Kapazitäten (4'932 Anmeldungen gegenüber 1'439 Studienplätzen).

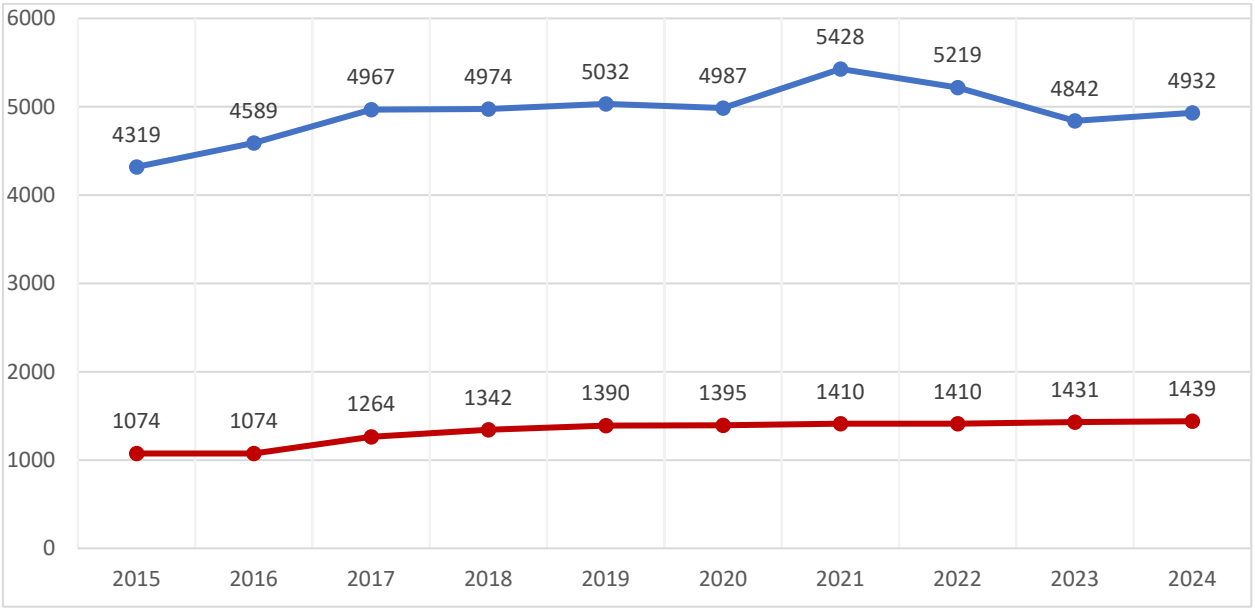


Abbildung 2: Anmeldungen (blau) und Kapazitäten (rot) an den Hochschulen mit Numerus clausus 2015-2024

3. Testteilnahmen Eignungstest für das Medizinstudium 2024

Nach der Frist für die Anmeldung vom 15. Februar kommt es jedes Jahr zu Rückzügen. Einige der angemeldeten Personen immatrikulieren sich nicht (Hochschulen ohne Numerus clausus) oder melden sich nicht zum EMS an (Hochschulen mit Numerus clausus). Auch nach der Anmeldung zum EMS kommt es jedes Jahr zu weiteren Rückzügen. 2024 meldeten sich bis zum 15. Februar 4'932 Personen an einer Hochschule mit Numerus clausus an, 3'813 davon meldeten sich bis zur Frist Ende Mai zum EMS an, 3'681 Personen traten zum Test an und erzielten schliesslich ein gültiges Testergebnis.

Der EMS kann auf Deutsch, Italienisch und Französisch absolviert werden. Weil die Universitäten Genf, Lausanne und Neuenburg den Zugang zum ersten Studienjahr nicht beschränken, nehmen im Verhältnis zur Bevölkerung weniger Personen mit Muttersprache Französisch (320 Personen, 9% der Testteilnehmenden) oder Italienisch (137 Personen, 4%) teil als mit Muttersprache Deutsch (3'224 Personen, 87%). Veterinärmedizin wird in der Schweiz nur in Zürich und Bern angeboten, deshalb liegt der Anteil Frankophoner in dieser Disziplin höher (97 Personen, 26%).

3.1. Testteilnahmen gesamt und nach Disziplin

2024 nahmen insgesamt (alle Disziplinen) etwas weniger Personen am EMS teil als im Vorjahr.

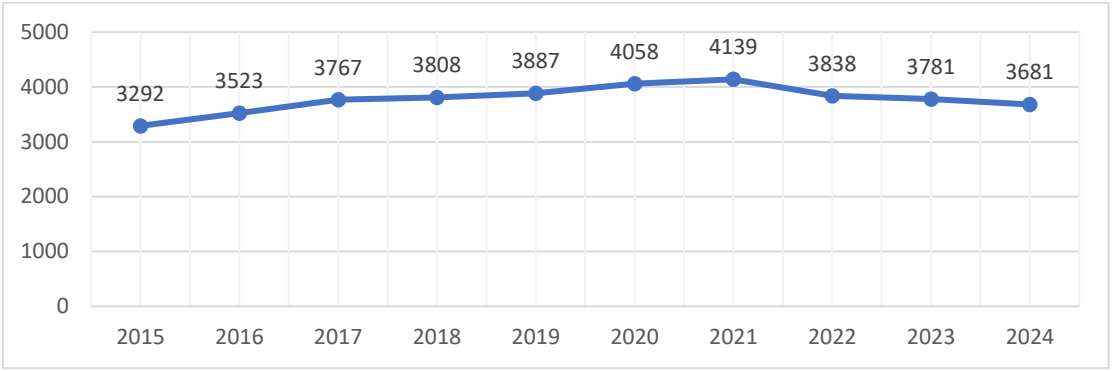


Abbildung 3: Testteilnahmen gesamt 2015-2024

In der Human- und Veterinärmedizin waren 2024 gegenüber 2023 weniger Testteilnahmen zu verzeichnen. In der Zahnmedizin stieg die Zahl der Testteilnehmenden erneut an.

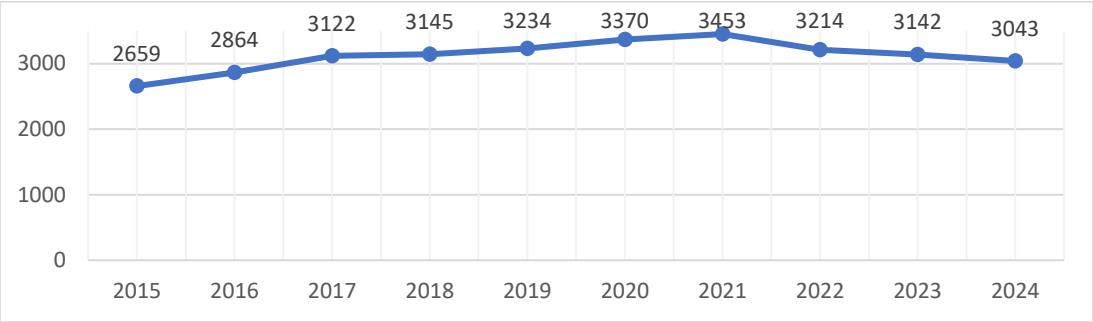


Abbildung 4: Testteilnahmen Humanmedizin inkl. Chiropraktik 2015-2024

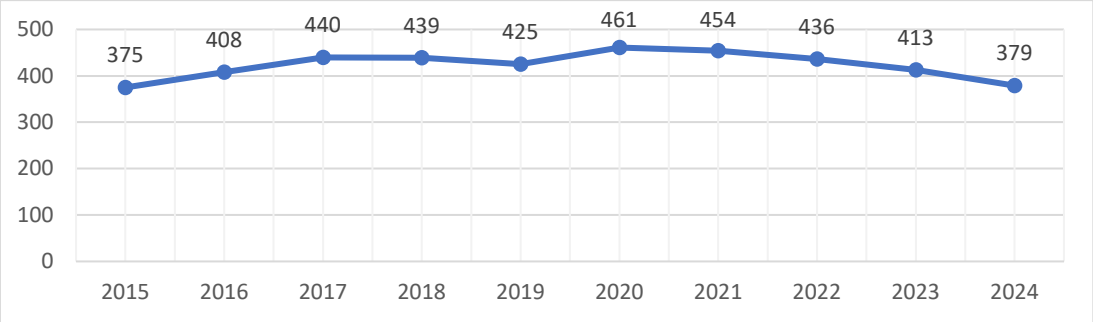


Abbildung 5:Testteilnahmen Veterinärmedizin 2015-2024

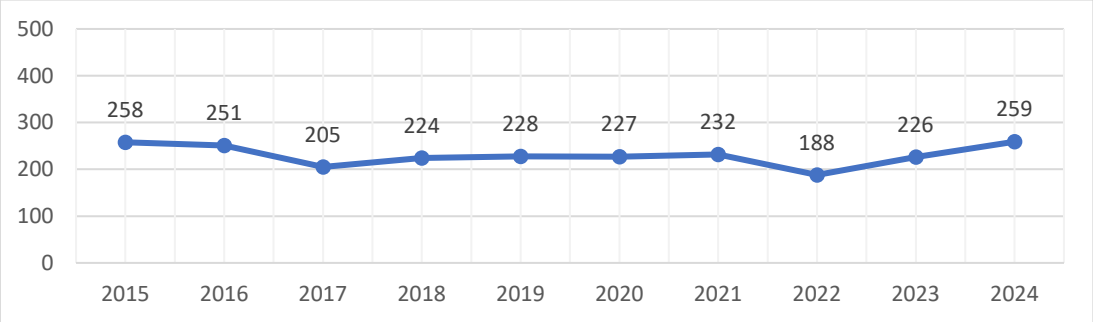


Abbildung 6: Testteilnahmen Zahnmedizin 2015-2024

3.2. Testteilnahmen nach Geschlecht und Disziplin

Auch 2024 nahmen mehr Frauen als Männer am EMS teil. In der Humanmedizin (inkl. Chiropraktik) waren es 2'062 Frauen und 981 Männer, in der Veterinärmedizin 335 Frauen und 44 Männer und in der Zahnmedizin 191 Frauen und 68 Männer.

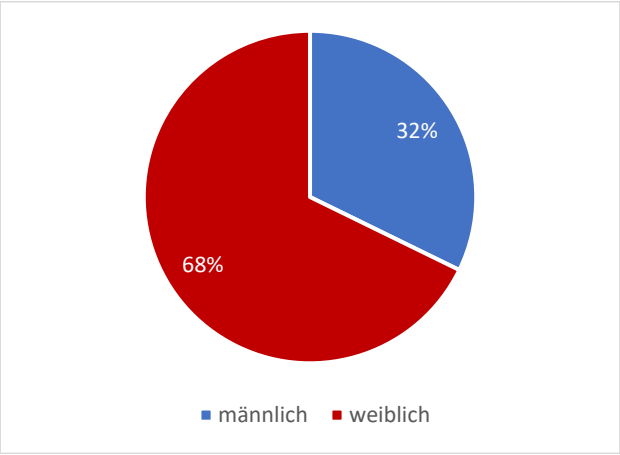


Abbildung 7: Testteilnahmen 2024 nach Geschlecht HM inkl. CP

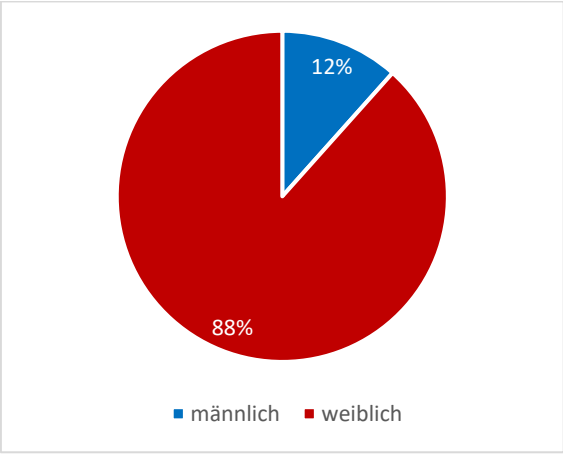


Abbildung 8: Testteilnahmen 2024 nach Geschlecht VM

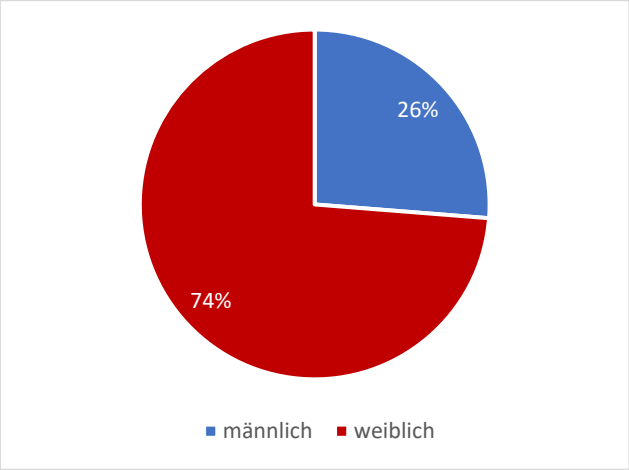


Abbildung 9: Testteilnahmen 2024 nach Geschlecht ZM

3.3. Testteilnahmen nach Kanton

Von den 3'681 Personen, die 2024 am EMS teilnahmen und ein gültiges Ergebnis erzielten, hatten 63 ihren Wohnsitz im Ausland und 3'618 in der Schweiz. Die nachfolgende Tabelle zeigt die Aufteilung auf die Kantone (Wohnsitz).

Wohnkanton	Testteilnehmende
AG	358
AI	7
AR	14
BE	548
BL	170
BS	170
FR	150
GE	19
GL	14
GR	72
JU	15
LU	209
NE	29
NW	22
OW	15
SG	231
SH	29
SO	125
SZ	62
TG	90
TI	125
UR	6
VD	71
VS	92
ZG	61
ZH	914
Ausland	63
Total	3'681

Tabelle 1: Testteilnahmen 2024 nach Wohnkanton

4. Testdurchführung / Testergebnisse

Die Berichterstattung zur Testdurchführung und den Testergebnissen erfolgt im Wissenschaftlichen Bericht des Zentrums für Testentwicklung und Diagnostik (ZTD) der Universität Freiburg über die Durchführung des EMS 2024, der sich im Anhang zum vorliegenden Mantelbericht von swissuniversities befindet.

5. Abschluss des Zuteilungsverfahrens

Nach dem EMS führt swissuniversities basierend auf den vom ZTD ermittelten Testresultaten (Test-Prozentrang und Mittlerer Rangplatz aller Aufgabengruppen) die Zuteilung der Studienplätze durch. Die Zuteilung erfolgt automatisch im Anmeldetool MEDON. In einem ersten Schritt wird für jede Disziplin so vielen Studienanwärterinnen und -anwärtern ein Studienplatz zugeteilt, dass die gesamte Aufnahmekapazität aller Hochschulen im ersten Studienjahr ausgeschöpft wird. Zusätzlich zu den Testteilnehmenden des aktuellen Jahres 2024 müssen dabei jeweils auch die Kandidatinnen und Kandidaten berücksichtigt werden, die den Testwert des Vorjahres anrechnen lassen (2024: 85 Bewerbungen).

5.1. Zulassungen pro Disziplin

In der Humanmedizin (inkl. Chiropraktik) haben 40% der Studienanwärterinnen und -anwärter einen Studienplatz erhalten (1’252 Personen), 60% mussten abgewiesen werden (1'848 Personen). In der Veterinärmedizin konnten 48% der Bewerbungen (187 Personen) zugelassen werden, 52% der Bewerbungen (200 Personen) mussten abgewiesen werden. In der Zahnmedizin konnten 49% (137 Personen) zugelassen werden, 51% (141 Personen) mussten abgewiesen werden. Es werden etwas mehr Personen zugelassen als Kapazitäten vorhanden sind, weil es erfahrungsgemäss auch nach der Zuteilung noch zu Rückzügen kommt (siehe 5.3, Verzicht auf einen Studienplatz).

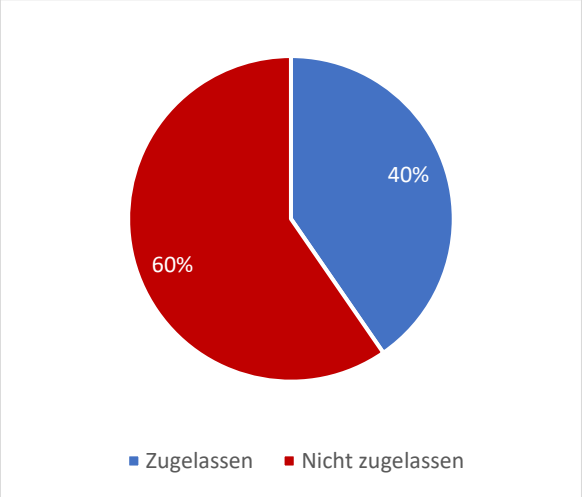


Abbildung 10: Zulassungen 2024 Humanmedizin inkl. Chiropraktik

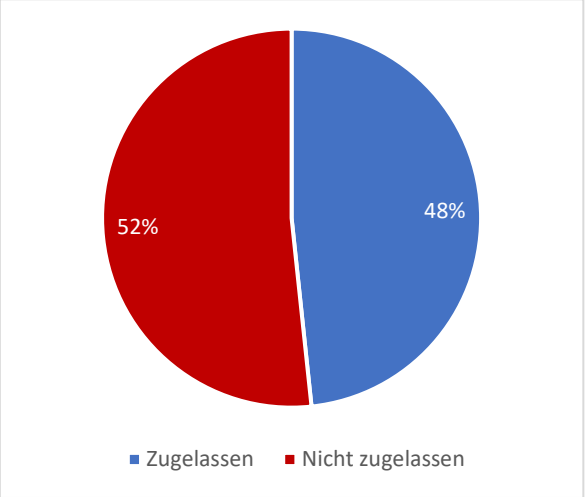


Abbildung 11: Zulassungen 2024 Veterinärmedizin

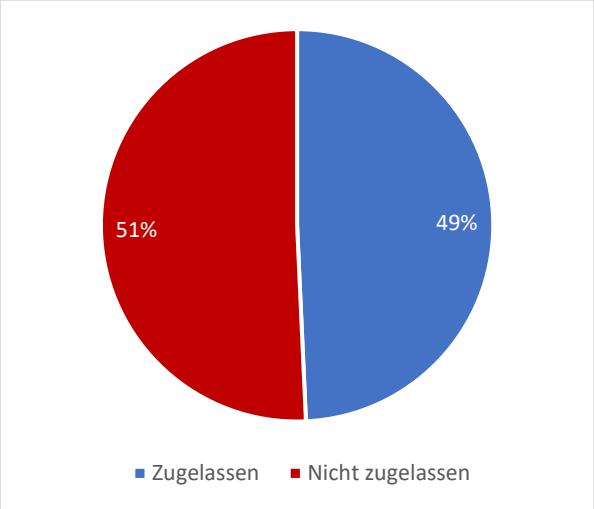


Abbildung 12: Zulassungen 2024 Zahnmedizin

5.2. Umleitungen

In einem zweiten Schritt erfolgt die Verteilung auf die Studienorte, ebenfalls im Anmeldetool MEDON. swissuniversities berücksichtigt dabei so weit wie möglich die Wünsche der Studienanwärterinnen und -anwärter.

In der Humanmedizin führten die Diversifizierung des Studienangebots und die unterschiedliche Nachfrage auch 2024 zu vielen Umleitungen (194 Personen) zwischen den Hochschulen, wenn auch zu etwas weniger als im Vorjahr (214 Personen). Mehrere Hochschulen sind gleichzeitig Absenderin und Empfängerin von umgeleiteten Personen. In der Veterinärmedizin und in der Zahnmedizin kann nur an zwei bzw. drei Studienorten (mit NC) studiert werden, deshalb ist der Anteil Umleitungen geringer.

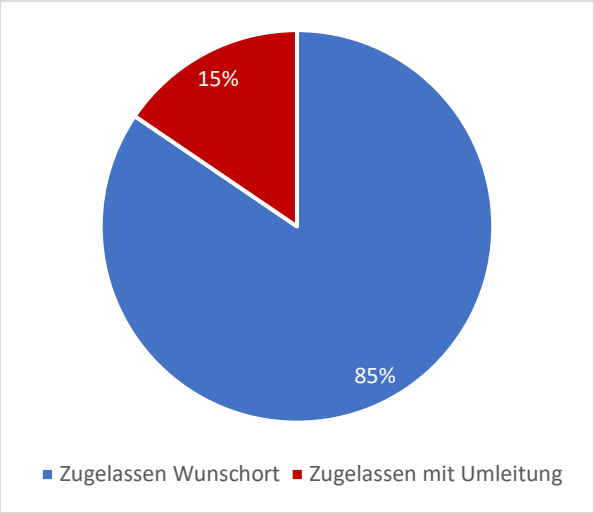


Abbildung 13: Umleitungen 2024 Humanmedizin

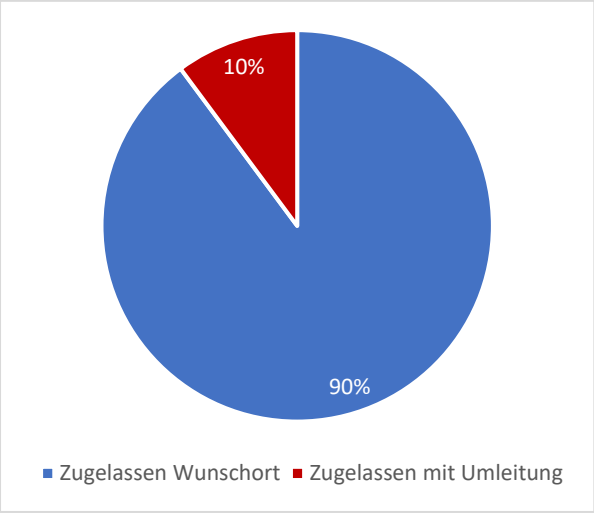


Abbildung 14: Umleitungen 2024 Veterinärmedizin

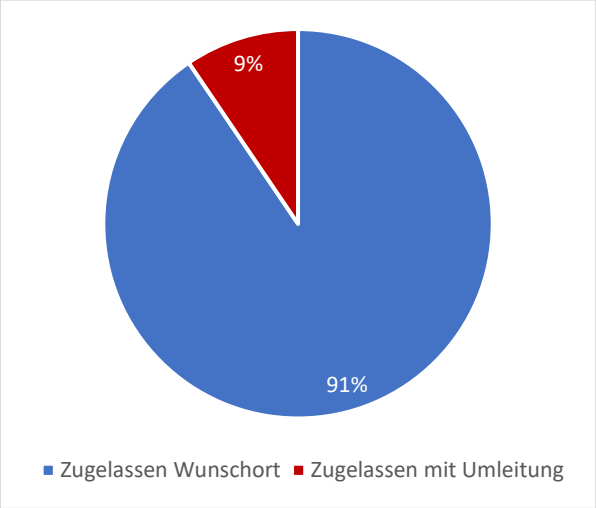


Abbildung 15: Umleitungen 2024 Zahnmedizin

Der Anteil der Bewerberinnen und Bewerber, die einen Studienplatz erhalten haben, jedoch nicht an Ihrer Wunschhochschule, liegt erstmals seit 2017 unter insgesamt (alle Disziplinen) 15%.

5.3. Verzicht auf einen Studienplatz

2024 verzichteten bis zehn Tage vor Vorlesungsbeginn 91 Bewerberinnen und Bewerber nach dem Zulassungsentscheid auf den ihnen zugeteilten Studienplatz, davon 74 in der Humanmedizin (inkl. Chiropraktik), zehn in der Veterinär- und sieben in der Zahnmedizin. Dank den Überbuchungen (siehe 5.1, Zulassungen pro Disziplin) wurden nur wenige Nachnominierungen benötigt: Drei in der Humanmedizin, zwei in der Veterinärmedizin und eine in der Zahnmedizin. Zu Beginn des Studiums bleiben somit keine Plätze vakant.

5.4. Altersstruktur

Der Hochschulrat der Schweizerischen Hochschulkonferenz SHK unterstrich an seiner Sitzung vom 8. November 2024 die öffentlichen Interessen (effiziente und volkswirtschaftlich nachhaltige Zuteilung der beschränkt zur Verfügung stehenden Studienplätze sowie Sicherstellung der öffentlichen Gesundheitsversorgung) und die Verhältnismässigkeit einer möglichen Altersbeschränkung für die Zulassung zum Medizinstudium ab 40 Jahren. Er beschloss aufgrund der aktuellen sehr tiefen Fallzahlen noch keine Empfehlung für eine schweizweite Altersbeschränkung auszusprechen. Er will die Entwicklung der Fallzahlen weiter beobachten und behält sich vor, bei einer Änderung der Situation sämtlichen Trägern der universitären Hochschulen mit medizinischen Ausbildungen die Einführung einer entsprechenden Altersgrenze zu empfehlen. Aus diesen Grund wird im jährlichen Mantelbericht in Zukunft über die Anzahl über 40-jährige immatrikulierte Personen Bericht erstatten.

Im Herbst 2024 wurden neun Personen im Alter von über 40 Jahren in einem medizinischen Studiengang immatrikuliert. Sieben Personen wurden an Hochschulen ohne Numerus clausus und Eignungstest ins erste Studienjahr zugelassen. Zwei Personen erhielten einen Studienplatz an einer Hochschule mit Numerus clausus aufgrund ihres Ergebnisses am Eignungstest, beide in der Veterinärmedizin.

6. Follow-up Evaluation Zentrum für Testentwicklung und Diagnostik (ZTD)

Der Hochschulrat der SHK beauftragte swissuniversities und das ZTD, ihn über ihre Zusammenarbeit und die Umsetzung der Empfehlungen A bis E auf dem Laufenden zu halten. Im neuen Abschnitt 6 werden die wichtigsten Entwicklungen zusammengefasst.

Empfehlung A: Prozesse in einem ZTD-internen Qualitätsmanagementsystem abbilden

Alle Arbeitsschritte sind ZTD-intern dokumentiert und werden fortlaufend an die sich ändernden Bedingungen (neue Aufgabengruppen, angepasste Testabläufe, neu zu erstellende Vorbereitungsmaterialien) angepasst. Die relevanten Schnittstellen zwischen swissuniversities und ZTD sind schriftlich geregelt.

Empfehlung B: Jährliche Zeiterfassung beibehalten

Die Zeiterfassung wird beim ZTD weitergeführt. Für den «EMS-Zyklus» 2024 hat das ZTD swissuniversities die Zeiterfassung Ende Oktober zur Verfügung gestellt. Um die Aussagekraft zu verbessern, wird eine Optimierung (zum Beispiel passendere Kategorien von Aufgaben) geprüft.

Empfehlung C: Möglichkeiten der präventiven Ausrichtung der Risikoszenarien prüfen

Die bestehenden Risikoszenarien wurden 2024 breiter kommuniziert, insbesondere den Koordinator:innen der acht Testorte sowie den Testleiter:innen. Dies trägt zur präventiven Wirkung bei.

Durch die Erhöhung von einem auf zwei Termine des TMS in Deutschland ergeben sich keine zusätzlichen Risiken für den EMS. Die Lieferung der Aufgaben für das Folgejahr ist in jedem Fall gewährleistet.

Empfehlung D: Finanzierung von Weiterbildungen klären

Das ZTD und swissuniversities haben erste Abklärungen getroffen, um die Finanzierung der Weiterbildungen des Personals des ZTD (Sach- und Personalkosten) zu klären.

Empfehlung E: Den Stellenwert der Durchführung von Probeläufen bekannt machen

Die Mitglieder der Konferenz der Schweizerischen Gymnasialrektorinnen und Gymnasialrektoren (KSGR) werden jedes Jahr gebeten, kostenlose Probeläufe durchzuführen. swissuniversities hat 2024 in der Kommunikation noch stärker als bisher auf die Bedeutung und Wirksamkeit der Probeläufe hingewiesen. Zusätzlich hat die Generalsekretärin 2024 auch die Mitglieder der Fachkonferenz der Schweizerischen Hochschulkonferenz (SHK), also die zuständigen Amtschefinnen und -chefs der Kantone, für dieses Anliegen sensibilisiert.

Erfahrungen aus Deutschland zeigen, dass eine direkte Involvierung von ZTD und swissuniversities in die Vorbereitung von Teilnehmenden, z.B. in Form von Online-Kursen, nicht empfehlenswert ist. Wenn Kurse interaktiv gestaltet werden sollen, können nicht alle Interessierten teilnehmen. Dies verbessert die Chancengleichheit nicht, sondern wirkt sich negativ aus. Das ITB hat entsprechende Kurse online angeboten und nach Kritik von Testteilnehmenden wieder darauf verzichtet.

Das ZTD hat sinnvolle Bearbeitungshinweise von kommerziellen Anbietern in die kostenlosen Vorbereitungsmaterialien wie die Broschüre «TestInfo» und die Beispielaufgaben aufgenommen.

ZTD CTD

UNI FR

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

2024 / BERICHT 31

EIGNUNGSTEST FÜR DAS MEDIZINSTUDIUM (EMS)

BERICHT ÜBER DIE DURCHFÜHRUNG UND ERGEBNISSE

Dr. B. Spicher

Dr. Spicher, Benjamin (2024)
EMS Eignungstest für das Medizinstudium 2024
Berichte des Zentrums für Testentwicklung, Nr. 31,
im Auftrag der Schweizerischen Hochschulkonferenz (SHK)

Unter Mitarbeit von M. Bernasconi, T. Cruchaud, L. Equey, M. Strazzeri und Y. de Zordo

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

Redaktion: T. Cruchaud

© 2024 ZTD Zentrum für Testentwicklung und Diagnostik
am Departement für Psychologie der Universität Freiburg
Direktorin: Prof. Dr. P. Klumb
Route d'Englisberg 9, CH-1763 Granges-Paccot
E-Mail: ztd@unifr.ch
Internet: http://www.unifr.ch/ztd

1 Inhalt

1 Beschreibung des verwendeten Eignungstests 4
1.1 Berechnung der Werte 6
1.2 Mittlerer Rangplatz der Aufgabengruppen 7
2 Testanwendung 2024 9
2.1 Verteilungsprüfung 9
2.2 Äquivalenz der Sprachversionen 10
2.2.1 Sprachvergleich für die Aufgabengruppen 12
2.2.2 Darstellung des Korrekturverfahrens 14
2.2.3 Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes 15
2.3 Effekte der Korrektur 22
2.4 Vergleichbarkeit der Testlokale 25
2.5 Vergleich für die Geschlechter 28
2.6 Vergleiche nach Wunschhochschulen 30
3 Ergebnisse **zur Testgüte** 31
3.1 Zuverlässigkeit 31
3.2 Binnenstruktur 32
4 Literatur 34
4.1 Frühere Berichte des ZTD 35

1 Beschreibung des verwendeten Eignungstests

Beim EMS handelt es sich um eine Adaptation des deutschen **Tests für Medizinische Studiengänge (TMS)** mit Weiterentwicklungen (durch das ZTD) für die Schweiz. Der TMS mit 9 Aufgabengruppen hat sich in Deutschland bewährt und wird dort seit 1986 eingesetzt (1986 bis 1996 deutschlandweit, seit 2007 erneut für ausgewählte Universitäten). In der Schweiz wird der Test seit 1998 erfolgreich angewendet – bis 2003 mit identischer Struktur wie beim deutschen TMS, seit 2004 durchlief der EMS verschiedene Anpassungen.

Seit 2022, nach zwei Durchführungen als spezielle, „coronabedingte" Versionen, besteht der EMS aus neun Aufgabengruppen, die zu einem Gesamtwert verrechnet werden. Ablauf und Struktur des aktuellen Tests sind in Tabelle 1 dargestellt.

Tabelle 1: Struktur und Ablauf des EMS 2024

EMS Struktur (seit 2022) Aufgabengruppe	Max. Punkte	Zeit in Min	Zeit pro Aufgabe (min)
Muster zuordnen	18	16	0,9
Medizinisch-naturwissenschaftliches Grundverständnis	18	45	2,5
Schlauchfiguren	18	10	0,6
Quantitative und formale Probleme	18	45	2,5
Figuren lernen		4	
Fakten lernen		6	
Wechsel von Testheft A zu Testheft B			
Textverständnis	18	45	2,5
Figuren reproduzieren	18	5	0,3
Fakten reproduzieren	18	6	0,3
Diagramme und Tabellen	18	45	2,5
Konzentriertes und sorgfältiges Arbeiten	18	8	
Total	162	235	

In jeder Aufgabengruppe können maximal 18 Punkte erzielt werden. Die Aufgabengruppen *Muster zuordnen*, *Fakten lernen*, *Figuren lernen* sowie *Konzentriertes und sorgfältiges Arbeiten* werden jährlich durch das Zentrum für Testentwicklung neu erstellt, die anderen, textintensiveren Aufgaben werden in Deutschland einmalig als Einstreuaufgaben vorerprobt und für den ersten zulassungsrelevanten Einsatz im EMS verwendet.

Jedes Jahr werden neue Aufgaben für alle Aufgabengruppen entwickelt. Die Anforderungen an die Aufgaben sind hoch, 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.

Ausgangspunkt der Aufgabenentwicklung des TMS/EMS war eine differenzierte Anforderungsanalyse des Medizinstudiums, an der Lehrbeauftragte sowie Experten und Expertinnen 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. Von diesen haben vor allem aufgrund der Ergebnisse zur Validität dann 9 Aufgabengruppen Eingang in den TMS und den EMS gefunden.

Die Stärken des EMS lassen sich folgendermassen zusammenfassen:

- Die Auswahl der Testanforderungen erfolgt aus einer grösseren Menge möglicher Studienanforderungen aufgrund einer Anforderungsanalyse. Dabei werden ständige Anpassungen an sich verändernde Bedingungen geprüft und wo nötig vorgenommen.
- Der wissenschaftliche Nachweis der Vorhersagbarkeit des Studienerfolgs mittels EMS ist gewährleistet. Dies erlaubt die gesetzlich geforderte Studienzulassung nach Eignung.
- Die Konstruktion der Aufgaben erfolgt durch Experten und es findet eine anschliessende empirische Überprüfung statt, damit die Aufgaben alle Gütekriterien erfüllen, die gewünschte Fähigkeit tatsächlich messen und optimal „schwierig“ sind.
- Für die Beantwortung der Aufgaben ist kein spezielles fachliches Vorwissen notwendig, es wird tatsächlich die „Studierfähigkeit“ als aktuelle Fähigkeit zur Wissensaneignung und Problemlösung gemessen. 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.

1.1 Berechnung der Werte

Alle Aufgabengruppen des EMS liefern Punkte als Summe richtig gelöster Aufgaben zwischen 0 und 18. Es können somit maximal 162 Punkte erreicht werden. Basierend auf den Kennwertanalysen wurden 2024 alle Aufgaben gewertet.

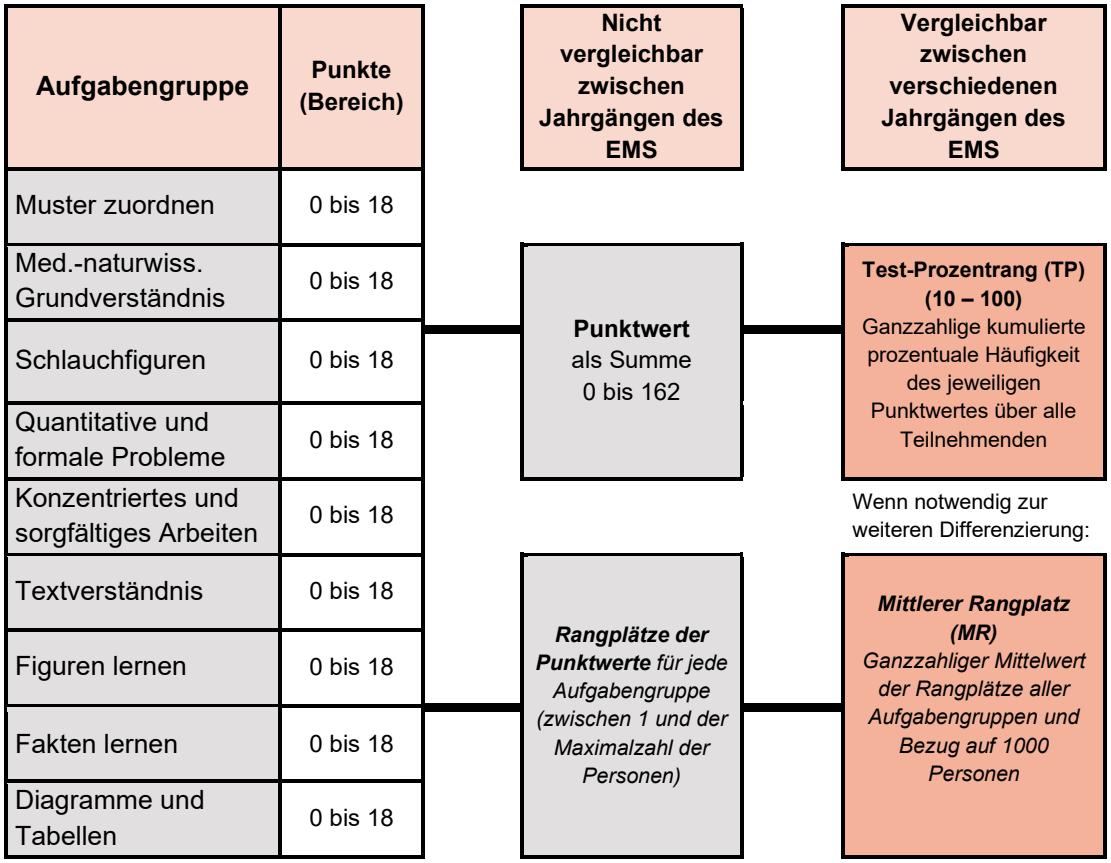


Abbildung 1: Punktwerte der einzelnen Aufgabengruppen 2024 und ihre Zusammenführung über den Punktwert zum Test-Prozentrang und mittleren Rangplatz.

Die Aufgabengruppe *Konzentriertes und sorgfältiges Arbeiten* besteht aus 1600 Zeichen, wovon 400 nach alljährlich wechselnden Regeln als Zielzeichen zu markieren sind. Es können allerdings in der zur Verfügung stehenden Zeit kaum 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 18 transformiert wird, um mit den anderen Aufgabengruppen gleichgewichtig zum Punktwert addiert zu werden. Die schlechtesten 1.5% (aber bis zu mindestens Rohwert 0) erhalten den Punktwert 0, die besten 1.5% erhalten den Punktwert 18. Der Wertebereich des Rohwertes für die verbleibenden 97% wird in 17 äquidistante Abschnitte eingeteilt und den Punktwerten 1 bis 17 zugeordnet. Obwohl sich die Testanforderung zwischen den Jahren ändert, resultieren so vergleichbare Punktwerteverteilungen.

Alle Punkte der Aufgabengruppen werden zu einer Summe addiert (Punktwert). Dieser Wert hat den Nachteil, dass er aufgrund schwankender Schwierigkeiten nicht zwischen Tests verschiedener Jahre vergleichbar ist.

Aus der Gesamtpunktzahl (Punktwert) werden deshalb 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 erreicht haben. Die Differenz zu 100 zeigt an, wie viele Prozent der Personen ein besseres Testergebnis 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. Test-Prozentrang und mittlerer Rangplatz sind über die Jahre vergleichbar – dies ist Voraussetzung dafür, dass ein Testergebnis ins Folgejahr übernommen werden kann.

1.2 Mittlerer Rangplatz der Aufgabengruppen

Nur bei gleichem Test-Prozentrang werden die Personen in der Reihenfolge des mittleren Rangplatzes aller Aufgabengruppen (MR) berücksichtigt, wenn nicht mehr alle mit entsprechendem Test-Prozentrang zugelassen werden können. Diese Rangplätze werden über alle Aufgabengruppen gemittelt. 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. Zur Herstellung der Vergleichbarkeit der Rangplätze zwischen den Jahren wird der mittlere Rangplatz auf die Basis von 1000 Personen umgerechnet.

Eine Beispielrechnung zur Ermittlung des mittleren Rangplatzes wird nachfolgend dargestellt.

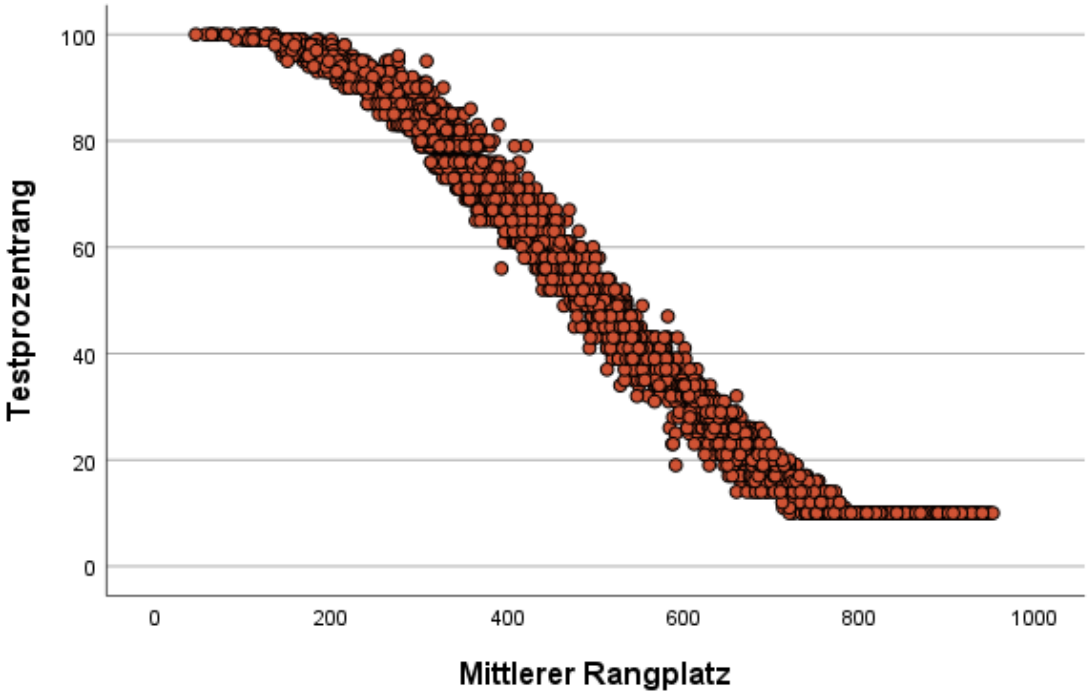


Abbildung 2: Beziehung zwischen Test-Prozentrang (TP) und mittlerem Rangplatz (MR) für 2024 (Werte kleiner als 10 auf werden auf TP 10 angehoben).

Der mittlere Rangplatz MR bewegt sich 2024 zwischen 90 und 969. Niedrige Werte stehen für bessere Leistungen. Mittlerer Rangplatz und Test-Prozentrang korrelieren mit -0.987.

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 und Kandidatinnen die maximale Punktzahl 20. Bei gleicher Punktzahl erhält jede Person den durchschnittlichen Rang dieser Personen (die 1 bis 10 belegen).
- Kandidatin A hat diese maximale Punktzahl erreicht und belegt damit in dieser Aufgabengruppe zusammen mit den anderen 9 Kandidaten und Kandidatinnen Rang 5.5 (Durchschnitt der Ränge 1 bis 10, also (1+10)/2=5.5)).
- Kandidat B hat, zusammen mit 14 weiteren Kandidaten und Kandidatinnen, 19 Punkte erzielt. Diese Kandidaten und Kandidatinnen 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.
- Kandidatin A erreicht hier 19 Punkte, zusammen mit nur 3 weiteren Kandidaten und Kandidatinnen. Sie alle erhalten in dieser Aufgabengruppe den durchschnittlichen Rang (2+5)/2=3.5.

Kandidatin A und Kandidat B erreichten beide damit die gleiche Punktzahl von 39 Punkten (und den gleichen Test-Prozentrang).

Der mittlere Rangplatz für Kandidatin A und Kandidat B berechnet sich wie folgt:

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

Und auf 1000 Personen:

- Kandidatin A: 4.5/100 × 1000=45
- Kandidat B: 9.5/100 × 1000=95

Hätten 1000 Personen am Test teilgenommen, resultiert für Kandidatin A im Mittel über beide Aufgabengruppen der 45. Rang, für Kandidat B der 95. Rang. Kandidatin A wird dafür belohnt, in der ersten Aufgabengruppe die maximale, aber vor allem auch in der zweiten Aufgabengruppe zusammen mit nur sehr wenigen Kandidaten und Kandidatinnen 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 Kandidatin A und Kandidat B ist dadurch gerechtfertigt, dass Kandidatin A trotz gleicher Punktzahl über beide Aufgabengruppen mehr Kandidaten und Kandidatinnen übertroffen hat als Kandidat B.

Müsste man zwischen beiden Personen unterscheiden, wäre Kandidatin A die bessere.

2 Testanwendung 2024

2.1 Verteilungsprüfung

Die Verteilung des Punktwerts entspricht annähernd einer Normalverteilung. Metrische Prüfverfahren sind bedingt aussagefähig. Die Punktwerte der einzelnen Aufgabengruppen sind statistisch nicht als normalverteilt anzusehen (Tabelle 2).

Tabelle 2: Kennwerte der Punktwertskalen Gesamt und Aufgabengruppen

	Punktwert	Quantitative und formale Probleme	Schlauchfiguren	Textverständnis	Med.-naturwiss. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme und Tabellen	Konzentr. und sorgf. Arbeiten
Mittelwert	82.7	8.2	13.5	6.6	7.0	9.4	10.1	10.2	8.0	9.8
Median	83	8	14	6	7	9	10	10	8	10
Modalwert	87	7	16	7	6	9	9	10	8	10
Stand.-abweichung	18.9	3.1	3.0	2.5	2.8	3.2	3.6	3.8	2.8	3.7
Spannweite	109	18	17	16	17	18	17	18	18	18
25. Perzentil	69	6	12	5	5	7	7	7	6	7
75. Perzentil	97	10	16	8	9	12	13	13	10	12
Schiefe	-0.1	0.3	-0.8	0.3	0.2	0.1	0.2	0.2	0.1	-0.2
Exzess	-0.5	-0.2	0.2	0.0	-0.3	-0.4	-0.7	-0.4	-0.3	-0.1

Für die Aufgabengruppe *Textverständnis* wurden 2024 die tiefsten mittleren Punktwerte registriert, gefolgt von *Medizinisch-naturwissenschaftlichem Grundverständnis*. Die höchsten mittleren Punktwerte lieferte die Aufgabengruppe *Schlauchfiguren*. Alle Aufgabengruppen differenzieren gut bis sehr gut im relevanten Wertebereich.

2.2 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) des ZTD entnommen werden. Zielsetzung bleibt, möglichst übereinstimmende Aufgabenschwierigkeiten (vergleichbare Ergebnisse bei vergleichbaren Fähigkeiten) 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. Die Schlussredaktionen der Testversionen in französischer und italienischer Sprache werden von mehreren nativ französisch- bzw. italienischsprachigen Personen durchgeführt. Diese wurden dahingehend geschult, dass es vor allem auf eine Übertragung der Schwierigkeit in die jeweilige Sprache ankommt.

Die Gleichheit ist nicht allein durch Übersetzung zu beeinflussen, sondern hängt auch von kulturellen und weiteren Faktoren ab. Bei allen Adaptationen muss mit Unterschieden hinsichtlich der Aufgabenschwierigkeiten zwischen den Sprachversionen gerechnet werden.

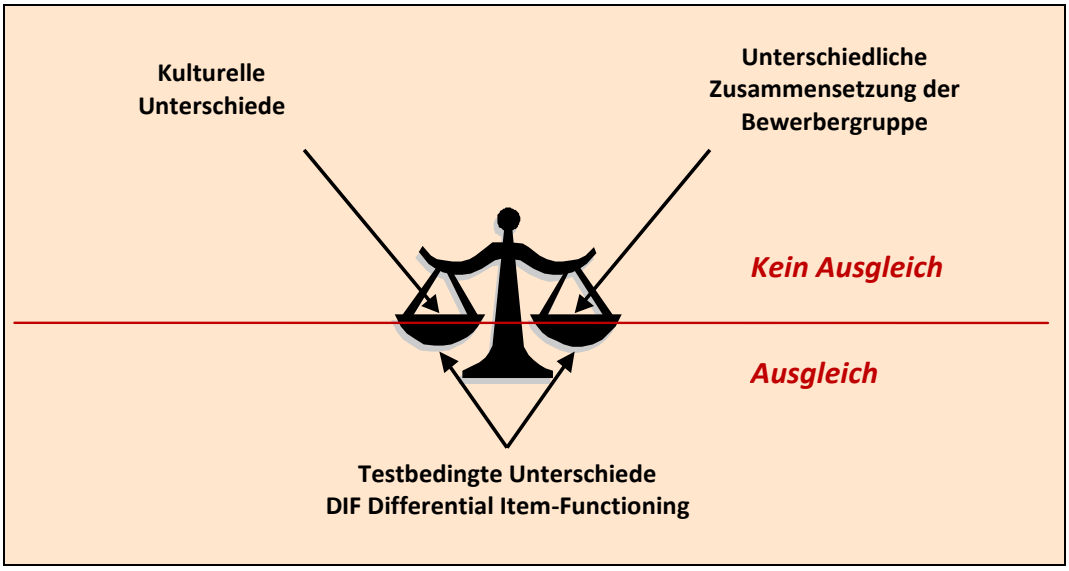


Abbildung 3: Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen.

Zwei Ursachen (kulturelle Unterschiede und die unterschiedliche Zusammensetzung der Bewerbergruppe) führen potenziell zu „echten“ Personen-Unterschieden, die nicht auf den Test oder seine Adaptation zurückgeführt werden können. Sie müssen durch die Ergebnisse entsprechend abgebildet werden – hier erfolgt kein Ausgleich.

Testbedingte Unterschiede hingegen können durch die Adaptation verursacht sein 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, oder die Art und Intensität der Vorbereitung 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 repräsentieren. 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 Teilnehmenden macht dies wahrscheinlich: Die Auswahl kann die „Spitze“ oder das „Ende“ der Leistungsrangreihe aller Maturanden und Maturandinnen 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 fünf von neun Aufgabengruppen. Der Sprachausgleich beschränkt sich demnach auf die stärker sprachabhängigen Aufgabengruppen *Medizinisch-naturwissenschaftliches Grundverständnis*, *Quantitative und formale Probleme*, *Textverständnis*, *Fakten lernen* sowie *Diagramme und Tabellen*.

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 sprachabhängigen Aufgabengruppen zu erwarten, während die 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.

2.2.1 Sprachvergleich für die Aufgabengruppen

Auch 2024 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Aufgabengruppen auf. Die unkorrigierten Punktzahlen nach Sprachen sind nachfolgend (Abbildung 4) abgebildet.

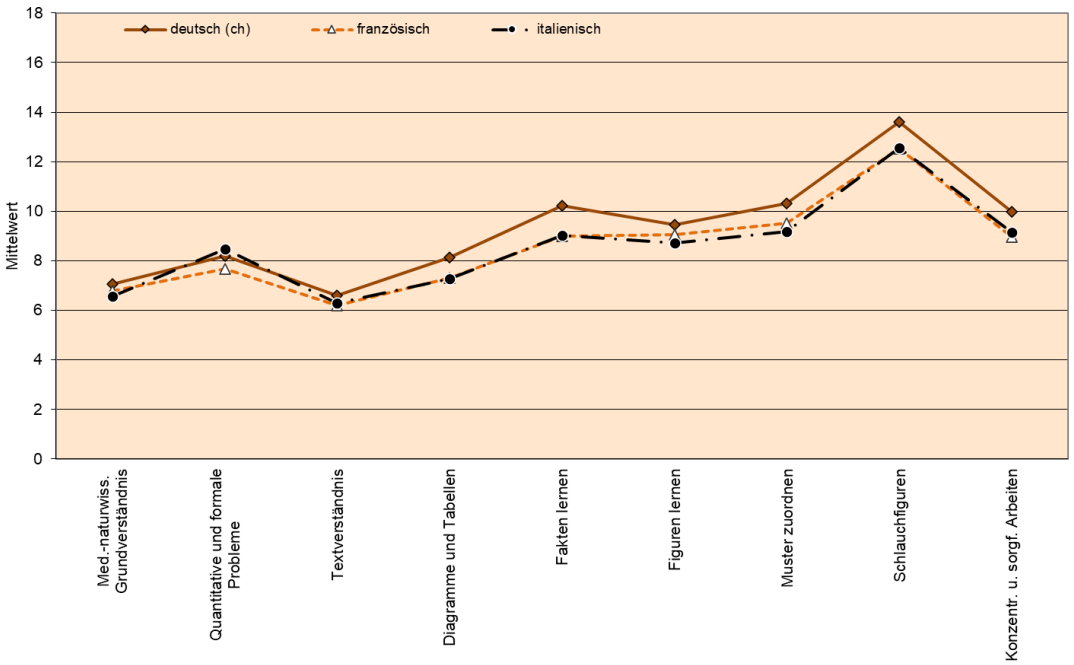


Abbildung 4: Mittelwerte der Punktwerte für die Aufgabengruppen nach Sprachen (unkorrigiert).

Abbildung 5 zeigt den Vergleich zwischen deutsch- und französischsprachigen Testteilnehmenden für das Jahr 2024, in Abbildung 6 ist der Vergleich zwischen deutsch- und italienischsprachigen Personen dargestellt. Die senkrechten Linien bezeichnen Mittelwert und Standardabweichung der Differenzen über die Jahrgänge für die vorangegangenen 10 Jahre. Positive Differenzwerte deuten auf bessere Ergebnisse der deutschen Sprachgruppe hin, negative auf solche der jeweils anderen Sprachgruppe.

Für die französische Sprachgruppe treten mehrheitlich geringere Unterschiede auf, als dies in den letzten 10 Jahren der Fall war, für *Muster zuordnen* und *Konzentriertes und sorgfältiges Arbeiten* sind die Ergebnisse vergleichbar zu den Vorjahren.

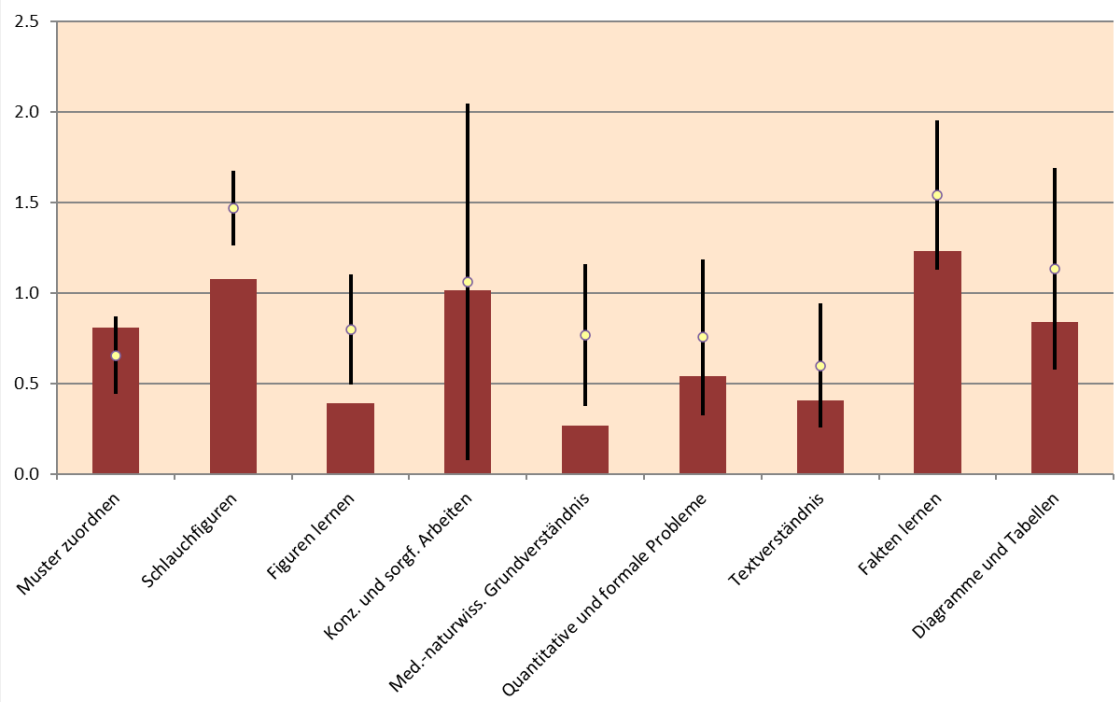


Abbildung 5: Differenz für Aufgabengruppen deutschsprachig-französischsprachig 2024 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2014-2023.

Die Unterschiede zwischen deutschsprachigen und italienischsprachigen Teilnehmenden fallen insbesondere bei den vier sprachunabhängigen Aufgabengruppen höher aus als üblich.

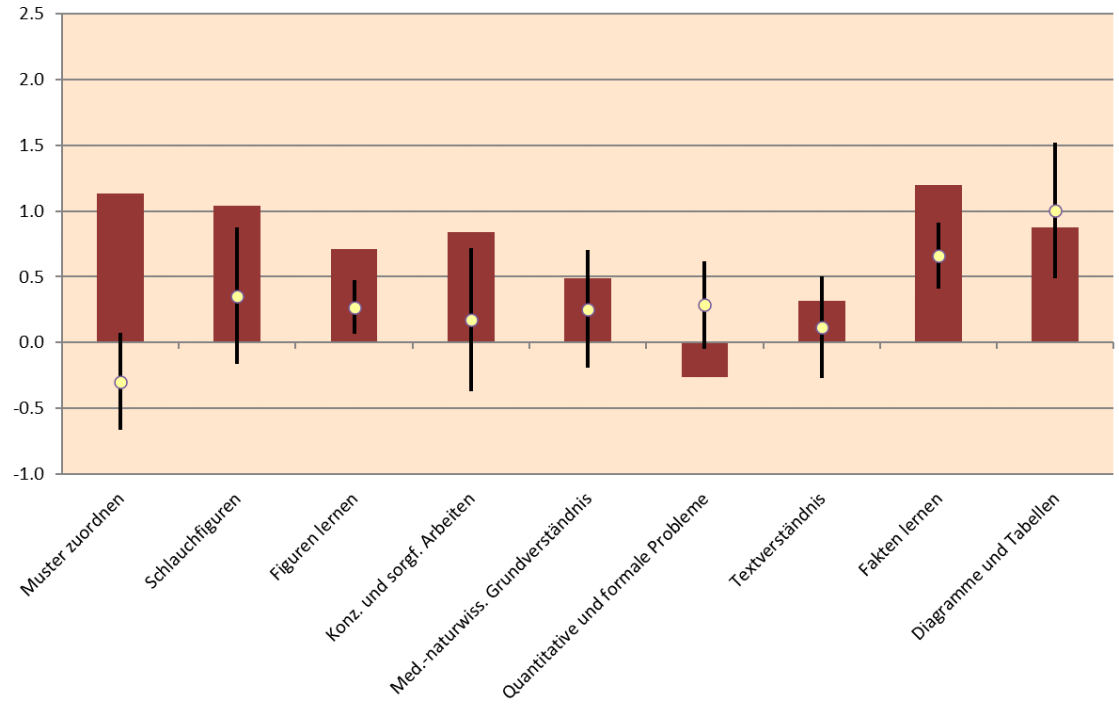


Abbildung 6: Differenz für Aufgabengruppen deutschsprachig-italienischsprachig 2024 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2014-2023.

2.2.2 Darstellung des Korrekturverfahrens

Das nachfolgend dargestellte Korrekturverfahren bleibt weiterhin für alle nicht-deutschsprachigen Testabsolvierungen relevant, da sichergestellt werden muss, dass durch die sprachliche Adaptation des Tests keinerlei Benachteiligungen auftreten.

2.2.2.1 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* sowie *Diagramme und Tabellen* vorgenommen. Dass Unterschiede in den nichtsprachlichen Tests testbedingt sind (und ausgeglichen werden müssen), ist unwahrscheinlich. Die Aufgaben selber erfordern praktisch keine Sprachkompetenzen und 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 hinsichtlich des Test-Prozentrangs 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, die also eine andere Position in der Rangreihe der Itemschwierigkeiten einnehmen. Für diese Aufgaben ist der Verdacht am ehesten begründet, dass es sich um testbedingte Differenzen handelt.

- 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 theoretisch maximal vergeben werden, und sich auf die Gewichtung der Aufgabengruppen negativ auswirken. Aus diesem Grund wird lediglich die Differenz der Aufgabenschwierigkeit gutgeschrieben. Nur Personen, welche die betreffende Aufgabe nicht korrekt gelöst haben, erhalten diesen Bonus.

2.2.2.2 Welche Aufgabengruppen ausgleichen?

Die folgenden Tabellen zeigen die Sprachvergleiche für die sprachabhängigen Aufgabengruppen. Referenz ist die deutschsprachige Testversion. Die Gesamtheit der französischsprachigen Teilnehmenden erzielte in vier der sprachabhängigen Aufgabengruppen signifikant geringere Punktwerte als die deutschsprachigen Teilnehmenden. Für die italienische Sprachgruppe sind drei Aufgabengruppen betroffen.

Tabelle 3: Signifikanzprüfung deutsche vs. französische Sprachform (sprachabhängige Aufgabengruppen).

Aufgabengruppe	Sprache	n	m	s	Sig.	Cohen's d	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3224	7.05	2.801	.053	0.095	Nein
	f	320	6.79	2.656			
Quantitative und form. Probleme	d	3224	8.21	3.110	**.046**	**0.174**	**Ja**
	f	320	7.67	3.015			
Textverständnis	d	3224	6.61	2.541	**.003**	**0.159**	**Ja**
	f	320	6.20	2.521			
Fakten lernen	d	3224	10.24	3.556	**.000**	**0.348**	**Ja**
	f	320	9.01	3.257			
Diagramme und Tabellen	d	3224	8.14	2.837	**.000**	**0.296**	**Ja**
	f	320	7.31	2.658			

Die Effektstärken (Cohen's d) für *Quantitative und formale Probleme* und *Textverständnis* sind in ihrem Ausmass als vernachlässigbar, jene für *Fakten lernen* und *Diagramme und Tabellen* sind als gering zu bezeichnen.

Tabelle 4: Signifikanzprüfung deutsche vs. italienische Sprachform (sprachabhängige Aufgabengruppen).

Aufgabengruppe	Sprache	n	m	s	Sig.	Cohen's d	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3224	7.05	2.801	**.023**	**0.174**	**Ja**
	i	137	6.57	2.620			
Quantitative und form. Probleme	d	3224	8.21	3.110	I > d	-0.086	Nein
	i	137	8.47	2.883			
Textverständnis	d	3224	6.61	2.541	.078	0.124	Nein
	i	137	6.29	2.573			
Fakten lernen	d	3224	10.24	3.556	**.000**	**0.337**	**Ja**
	i	137	9.04	3.541			
Diagramme und Tabellen	d	3224	8.14	2.837	**.000**	**0.308**	**Ja**
	i	137	7.27	2.775			

Die Effektstärke für *medizinisch-naturwissenschaftliches Grundverständnis* ist in ihrem Ausmass als vernachlässigbar, jene für *Fakten lernen* und *Diagramme und Tabellen* sind als gering zu bezeichnen.

Mit der Bestimmung der zu prüfenden Aufgabengruppen ist noch nicht eruiert, ob es sich bei den Differenzen zwischen den Stichproben um mögliche adaptationsbedingte Unterschiede handelt. Zu diesem Zweck ist die nachfolgend dargestellte Analyse der Aufgaben der betreffenden Aufgabengruppen notwendig.

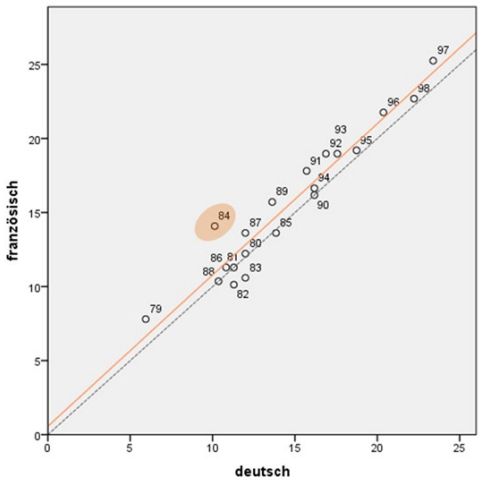
2.2.3 Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes

Eine Möglichkeit zur Identifikation von DIF, auch bei relativ 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 und Probandinnen gelöste) Aufgaben.

Die aus den Delta-Werten abgeleitete Regressionsgerade Y = AX + B beschreibt die Beziehung zwischen den analysierten 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 Sprachversionen 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 7: Beispiel für einen Delta-Plot.

Auffällige Abweichungen einzelner Punkte von dieser Regressionsgeraden hingegen deuten auf Aufgaben hin, welche zusätzlich zu systematischen Verschiebungen 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 Richtung tieferer Punktwerte für die interessierende Sprachgruppe abweichen. Als Sprachausgleich wird den Personen, die die entsprechende Aufgabe nicht gelöst haben, die Differenz der Aufgabenschwierigkeit zur deutschen Aufgabenschwierigkeit gutgeschrieben (also die Differenz der Lösungswahrscheinlichkeit).

2.2.3.1 Medizinisch-naturwissenschaftliches Grundverständnis

Die meisten Aufgaben der italienischsprachigen Version liegen eng an der Regressionsgeraden. Einzig Aufgabe 25 zeigt einen bedeutsamen Unterschied. Teilnehmenden aus der italienischen Sprachgruppe werden bei nicht korrekter Beantwortung dort 0.2 Punkte gutgeschrieben.

Tabelle 5: DIF-Analyse *Medizinisch-naturwissenschaftliches Grundverständnis*

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
19	0.71	-	0.56	8.76	-	11.89		
20	0.52	-	0.55	12.72	-	12.09		
21	0.51	-	0.42	12.93	-	14.80		
22	0.51	-	0.55	12.93	-	12.09		
23	0.44	-	0.42	14.38	-	14.80		
24	0.47	-	0.56	13.76	-	11.89		
25	0.45	-	0.25	14.18	-	18.34		**0.20**
26	0.45	-	0.40	14.18	-	15.22		
27	0.42	-	0.48	14.80	-	13.55		
28	0.43	-	0.34	14.59	-	16.47		
29	0.39	-	0.39	15.43	-	15.43		
30	0.40	-	0.35	15.22	-	16.26		
31	0.24	-	0.19	18.55	-	19.59		
32	0.24	-	0.23	18.55	-	18.76		
33	0.15	-	0.15	20.43	-	20.43		
34	0.21	-	0.21	19.18	-	19.18		
35	0.25	-	0.22	18.34	-	18.97		
36	0.28	-	0.29	17.72	-	17.51		

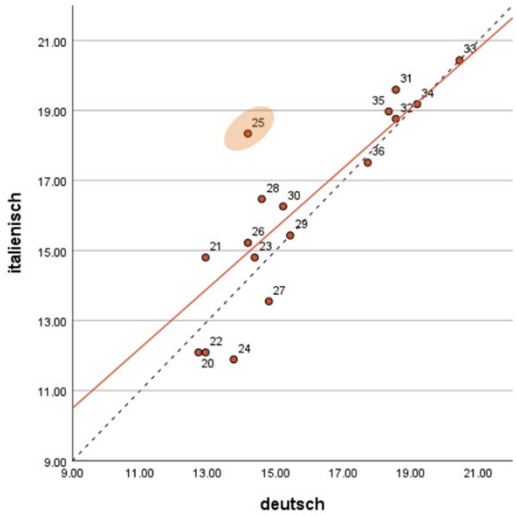


Abbildung 8: Delta-Plot für *Medizinisch-naturwissenschaftliches Grundverständnis* (deutsch-italienisch).

2.2.3.2 *Quantitative und formale Probleme*

Nur die französische Sprachgruppe weist hier bedeutsame Differenzen auf, allerdings liegen alle Aufgaben eng an der Regressionsgeraden. Trotz signifikanter Differenz hinsichtlich der Punktwerte dieser Aufgabengruppe gibt es auf Aufgabenebene keine Hinweise auf übersetzungsbedingte Einflüsse für die französische Sprachgruppe.

Tabelle 6: DIF-Analyse *Quantitative und formale Probleme*

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
55	0.82	0.84	-	6.47	6.05	-		-
56	0.68	0.71	-	9.39	8.76	-		-
57	0.68	0.68	-	9.39	9.39	-		-
58	0.56	0.47	-	11.89	13.76	-		-
59	0.55	0.48	-	12.09	13.55	-		-
60	0.53	0.44	-	12.51	14.38	-		-
61	0.38	0.33	-	15.63	16.68	-		-
62	0.56	0.46	-	11.89	13.97	-		-
63	0.49	0.44	-	13.34	14.38	-		-
64	0.43	0.38	-	14.59	15.63	-		-
65	0.48	0.53	-	13.55	12.51	-		-
66	0.30	0.29	-	17.30	17.51	-		-
67	0.40	0.40	-	15.22	15.22	-		-
68	0.41	0.42	-	15.01	14.80	-		-
69	0.22	0.16	-	18.97	20.22	-		-
70	0.28	0.22	-	17.72	18.97	-		-
71	0.21	0.17	-	19.18	20.01	-		-
72	0.24	0.25	-	18.55	18.34	-		-

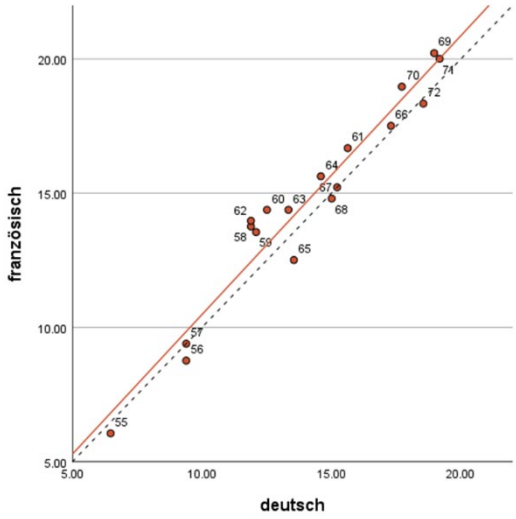


Abbildung 9: Delta-Plot für *Quantitative und formale Probleme* (deutsch-französisch).

2.2.3.3 *Textverständnis*

Auch bei dieser Aufgabengruppe liegen alle Aufgaben eng an der Regressionsgeraden. Trotz signifikanter Differenz hinsichtlich der Punktwerte dieser Aufgabengruppe gibt es auf Aufgabenebene keine Hinweise auf übersetzungsbedingte Einflüsse für die französische Sprachgruppe.

Tabelle 7: DIF-Analyse *Textverständnis*

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
73	0.67	0.67		9.59	9.59			-
74	0.61	0.57		10.84	11.68			-
75	0.42	0.37		14.80	15.84			-
76	0.38	0.28		15.63	17.72			-
77	0.42	0.45		14.80	14.18			-
78	0.25	0.20		18.34	19.38			-
79	0.65	0.59		10.01	11.26			-
80	0.57	0.62		11.68	10.64			-
81	0.30	0.29		17.30	17.51			-
82	0.33	0.31		16.68	17.09			-
83	0.29	0.30		17.51	17.30			-
84	0.26	0.23		18.13	18.76			-
85	0.37	0.32		15.84	16.88			-
86	0.33	0.27		16.68	17.93			-
87	0.17	0.17		20.01	20.01			-
88	0.21	0.23		19.18	18.76			-
89	0.11	0.08		21.26	21.88			-
90	0.28	0.25		17.72	18.34			-

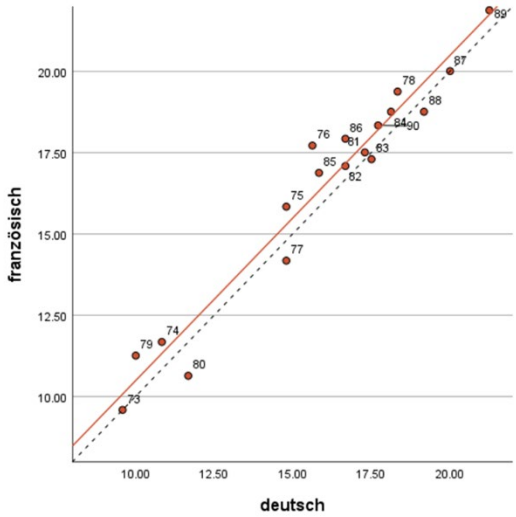


Abbildung 10: Delta-Plot für *Textverständnis* (deutsch-französisch).

2.2.3.4 *Fakten lernen*

In der Aufgabengruppe *Fakten lernen* werden der italienischen Sprachgruppe für die Aufgabe 109 bei Falschantwort 0.18 Punkte gutgeschrieben. Die französischsprachige Version zeigt keine Auffälligkeiten.

Tabelle 8: DIF-Analyse *Fakten lernen*

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
109	0.80	0.68	0.62	6.89	9.39	10.64		**0.18**
110	0.84	0.79	0.80	6.05	7.09	6.89		
111	0.83	0.82	0.88	6.26	6.47	5.22		
112	0.65	0.58	0.67	10.01	11.47	9.59		
113	0.56	0.51	0.47	11.89	12.93	13.76		
114	0.68	0.53	0.55	9.39	12.51	12.09		
115	0.47	0.48	0.44	13.76	13.55	14.38		
116	0.31	0.31	0.31	17.09	17.09	17.09		
117	0.61	0.51	0.51	10.84	12.93	12.93		
118	0.60	0.48	0.55	11.05	13.55	12.09		
119	0.38	0.31	0.35	15.63	17.09	16.26		
120	0.41	0.36	0.33	15.01	16.05	16.68		
121	0.42	0.43	0.28	14.80	14.59	17.72		
122	0.60	0.50	0.49	11.05	13.14	13.34		
123	0.65	0.53	0.52	10.01	12.51	12.72		
124	0.37	0.29	0.35	15.84	17.51	16.26		
125	0.64	0.62	0.63	10.22	10.64	10.43		
126	0.41	0.30	0.30	15.01	17.30	17.30		

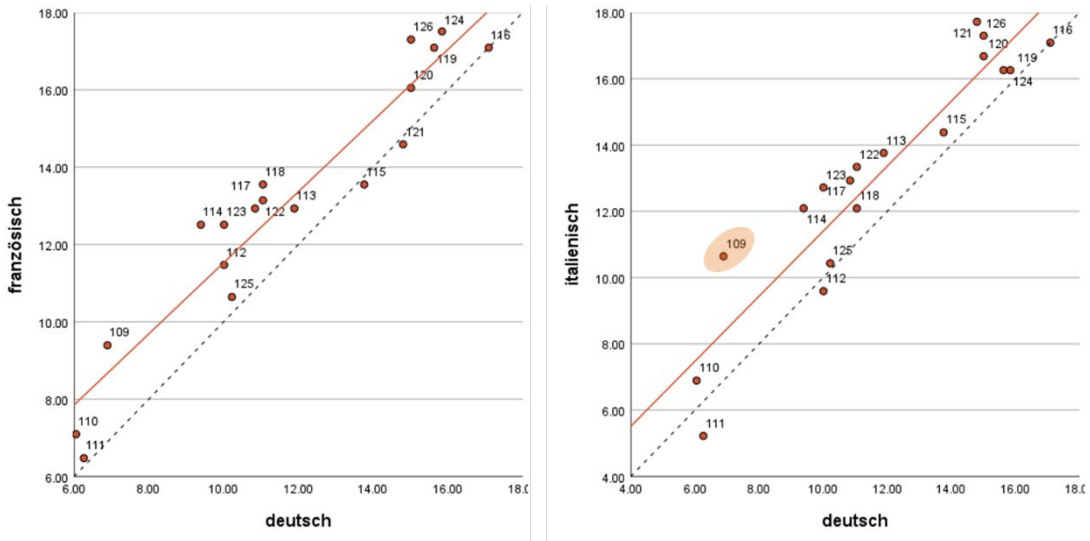


Abbildung 11: Delta-Plot für *Fakten lernen* (deutsch-französisch und deutsch-italienisch).

2.2.3.5 Diagramme und Tabellen

In der Aufgabengruppe *Diagramme und Tabellen* werden der französischen Sprachgruppe bei Falschantwort für die Aufgabe 138 0.16 Punkte gutgeschrieben. In der italienischsprachigen Version sind drei Aufgaben auszugleichen.

Tabelle 9: DIF-Analyse *Diagramme und Tabellen*

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
127	0.82	0.67	0.77	6.47	9.59	7.51		
128	0.75	0.75	0.63	7.93	7.93	10.43		
129	0.64	0.58	0.66	10.22	11.47	9.80		
130	0.70	0.61	0.47	8.97	10.84	13.76		**0.23**
131	0.63	0.66	0.42	10.43	9.80	14.80		**0.21**
132	0.46	0.39	0.34	13.97	15.43	16.47		
133	0.46	0.48	0.45	13.97	13.55	14.18		
134	0.52	0.47	0.64	12.72	13.76	10.22		
135	0.50	0.51	0.47	13.14	12.93	13.76		
136	0.39	0.38	0.42	15.43	15.63	14.80		
137	0.41	0.29	0.27	15.01	17.51	17.93		**0.14**
138	0.40	0.24	0.30	15.22	18.55	17.30	**0.16**	
139	0.38	0.28	0.36	15.63	17.72	16.05		
140	0.24	0.23	0.25	18.55	18.76	18.34		
141	0.21	0.13	0.13	19.18	20.84	20.84		
142	0.13	0.12	0.17	20.84	21.05	20.01		
143	0.28	0.29	0.28	17.72	17.51	17.72		
144	0.23	0.21	0.24	18.76	19.18	18.55		

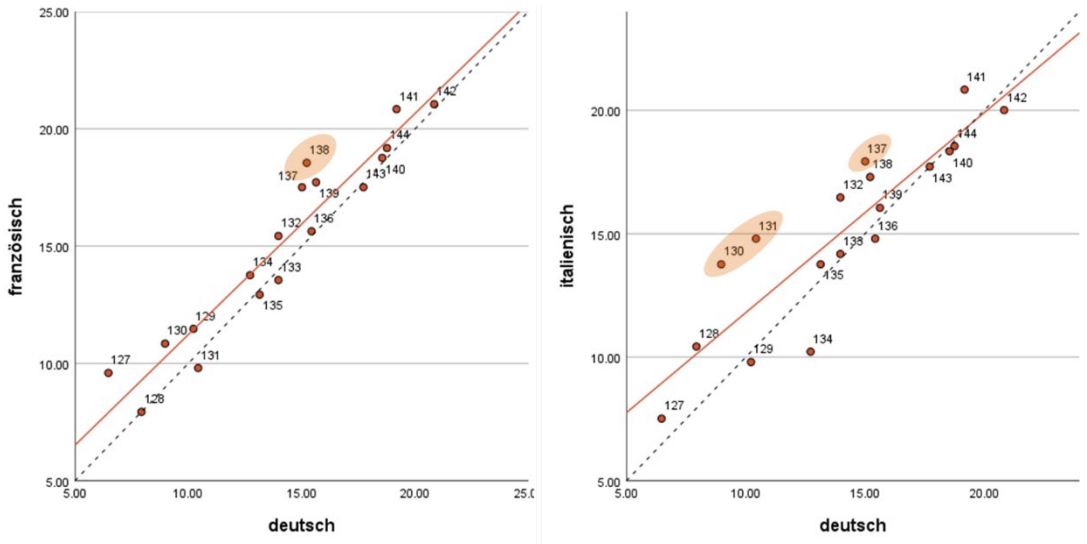


Abbildung 12: Delta-Plot für *Diagramme und Tabellen* (deutsch-französisch und deutsch-italienisch).

2.3 Effekte der Korrektur

Die durch den Ausgleich resultierenden Bonuspunkte können Tabelle 10 entnommen werden. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Teilnehmende aus der italienischen Sprachgruppe erhalten maximal einen Bonuspunkt. Sowohl in der französischen wie in der italienischen Sprachgruppe kommen aber die Bonuspunkte bei der Ermittlung der Rangplätze zum Tragen.

Tabelle 10: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen.

		n	Minimum	Maximum	m	s
Franz.	Diagramme und Tabellen	320	0.00	16.00	0.12	0.07
	GESAMT (Gerundet)	320	0.00	0.00	-	-
Italienisch	Med.-naturwiss. Grundverständnis	137	0.00	0.20	0.15	0.09
	Fakten lernen	137	0.00	0.18	0.07	0.09
	Diagramme und Tabellen	137	0.00	0.58	0.35	0.17
	GESAMT (Gerundet)	137	0.00	1.00	0.69	0.47

Die Mittelwertsdifferenzen für jede Aufgabengruppe nach dem Sprachausgleich sind in Tabelle 11 dargestellt. Die Ergebnisse multipler Mittelwertsvergleiche können der ersten Spalte entnommen werden. Differenzen zwischen den Teilnehmenden sind dort in der Form d** > f, i angegeben. Das Beispiel würde bedeuten, dass die deutschsprachigen Kandidaten und Kandidatinnen signifikant höhere Werte erzielt haben als die beiden anderen Sprachgruppen, die sich untereinander nicht signifikant unterscheiden.

Der Bonus der französischen Sprachgruppe war 2024 deutlich geringer als in den Vorjahren, die italienische Sprachgruppe hat mit im Mittel 0.69 Punkten einen vergleichsweise hohen Ausgleich erhalten (Abbildung 14).

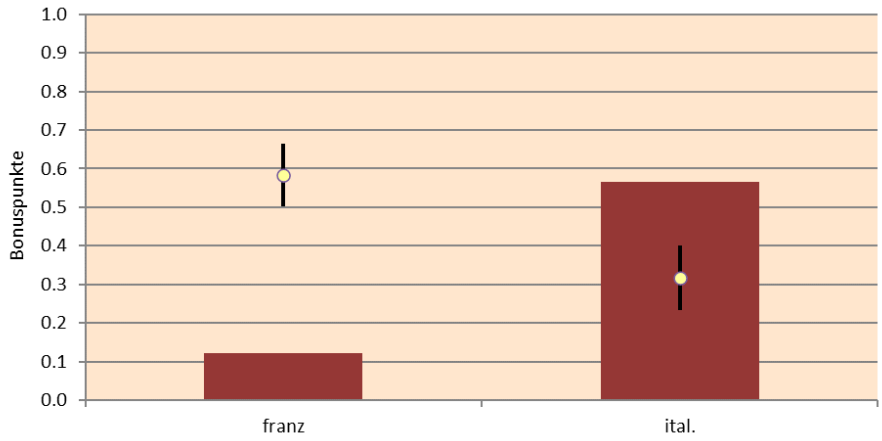


Abbildung 13: Sprachausgleich 2024 (ungerundet) mit Mittelwert und Standardabweichung der Testdurchführungen 2014-2023

Tabelle 11: Test-Prozentrang und Punktwerte inklusive Sprachausgleich.

		Mittelwert	Stand.-abw.	Minimum	Maximum
Test-Prozentrang d** > f, i	deutsch	52.45	28.09	10	100
	französisch	42.41	27.04	10	99
	italienisch	43.82	28.05	10	100
Quantitative und form. Probleme d, i** > f	deutsch	8.21	3.11	0	18
	französisch	7.67	3.01	1	17
	italienisch	8.47	2.88	1	17
Schlauchfiguren d**> f, i	deutsch	13.59	2.95	1	18
	französisch	12.51	3.12	3	18
	italienisch	12.55	3.40	1	18
Textverständnis d* > f	deutsch	6.61	2.54	0	16
	französisch	6.20	2.52	0	14
	italienisch	6.29	2.57	1	12
Med.-naturwiss. Grundverständnis	deutsch	7.05	2.80	0	17
	französisch	6.79	2.66	1	14
	italienisch	6.72	2.59	2	13
Figuren lernen	deutsch	9.44	3.22	0	18
	französisch	9.05	2.93	2	17
	italienisch	8.73	3.46	1	18
Fakten lernen d** > f, i	deutsch	10.24	3.56	1	18
	französisch	9.01	3.26	3	18
	italienisch	9.10	3.50	3	18
Muster zuordnen d** > f, i	deutsch	10.32	3.80	0	18
	französisch	9.51	3.38	1	18
	italienisch	9.18	3.16	2	18
Diagramme und Tabellen d** > f	deutsch	8.14	2.84	0	18
	französisch	7.43	2.63	0	15
	italienisch	7.62	2.69	2	16
Konz. und sorgf. Arbeiten d**, i* > f	deutsch	9.90	3.71	0	18
	französisch	8.87	3.62	0	18
	italienisch	9.06	3.67	0	18

n: deutsch 3224 / französisch 320 / italienisch 137

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

Die Differenzen zwischen der deutschsprachigen Referenzgruppe und den anderssprachigen Teilnehmenden werden nach folgender Formel berechnet:

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

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.

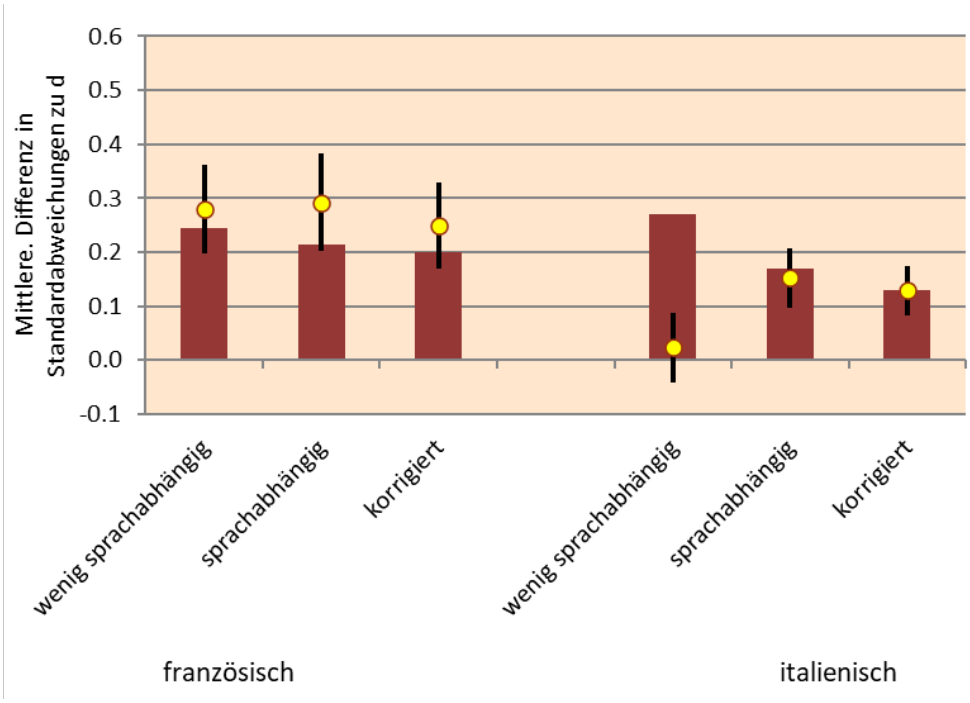


Abbildung 14: Mittlere Abweichung (Delta z) für wenig sprachabhängige und sprachabhängige Aufgabengruppen (korrigiert und unkorrigiert), 2024 und Vergleich zu 2014-2023 (Mittelwert und Streuung).

Beide Sprachgruppen weisen bei den sprachunabhängigen Aufgabengruppen höhere Differenzen aus als bei den stärker sprachabhängigen, was im direkten Vergleich zu den letzten Jahren darauf hindeutet, dass Ergebnisdifferenzen zwischen den Sprachgruppen weitgehend auf jährliche Leistungsschwankungen zwischen den Gruppen zurückzuführen sein dürften. Nach der Korrektur wird die Differenz in den sprachabhängigen Aufgabengruppen weiter vermindert.

2.4 Vergleichbarkeit der Testlokale

2024 erfolgte die Testabnahme in 32 Testlokalen/Sektoren verteilt auf acht Testorte. Eine Besonderheit stellen wiederum die sehr kleinen Testlokale (Lokale 3, 23 und 24) dar, die für Personen mit spezifischen Einschränkungen bereitgestellt wurden. Diese beiden Lokale werden aufgrund der zu geringen Fallzahlen von den weiteren Analysen ausgeschlossen.

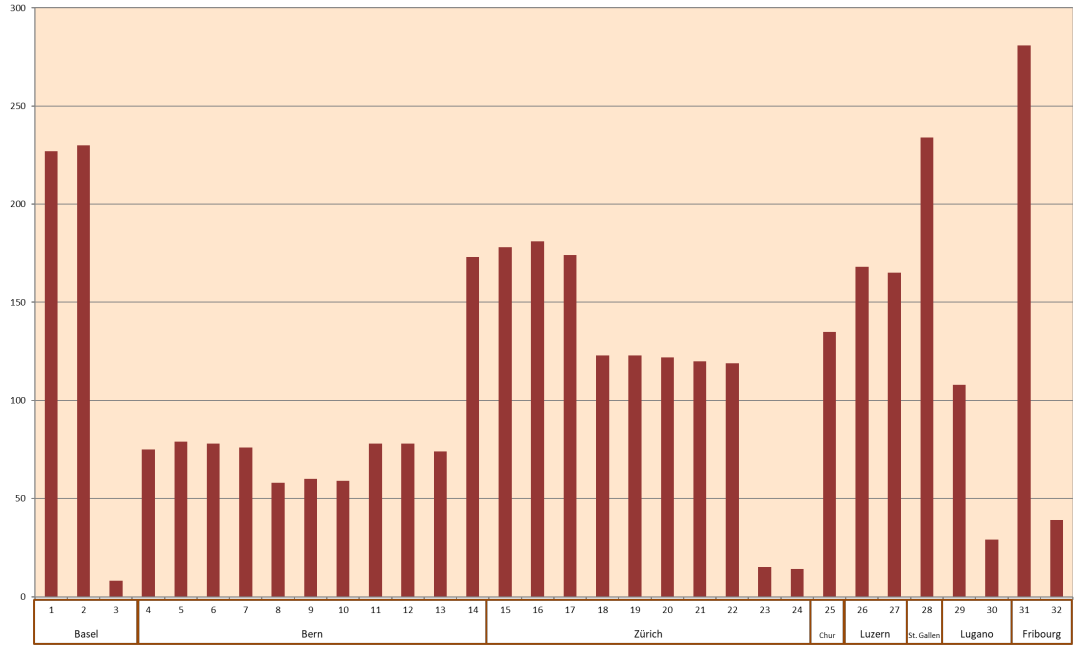


Abbildung 15: Testlokale mit Teilnehmerzahlen.

Tabelle 12: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den deutschsprachigen Testlokalen (ohne Lokale 3, 23 und 24).

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testprozentrang	*Zwischen*	45803.04	24	1908.460	2.446	0.000
	Innerhalb	2467440.48	3162	780.342		
	Total	2513243.53	3186			
Muster zuordnen	*Zwischen*	511.80	24	21.325	1.482	0.062
	Innerhalb	45512.08	3162	14.393		
	Total	46023.88	3186			
Med.-naturwiss. Grundverständnis	*Zwischen*	293.76	24	12.240	1.564	0.040
	Innerhalb	24753.15	3162	7.828		
	Total	25046.91	3186			
Schlauchfiguren	*Zwischen*	382.37	24	15.932	1.838	0.008
	Innerhalb	27414.26	3162	8.670		
	Total	27796.63	3186			
Quant. und formale Probleme	*Zwischen*	711.00	24	29.625	3.108	0.000
	Innerhalb	30137.61	3162	9.531		
	Total	30848.61	3186			
Textverständnis	*Zwischen*	236.93	24	9.872	1.537	0.046
	Innerhalb	20310.36	3162	6.423		
	Total	20547.29	3186			
Figuren lernen	*Zwischen*	308.25	24	12.844	1.241	0.193
	Innerhalb	32720.39	3162	10.348		
	Total	33028.63	3186			
Fakten lernen	*Zwischen*	531.23	24	22.135	1.759	0.013
	Innerhalb	39795.25	3162	12.585		
	Total	40326.48	3186			
Diagramme und Tabellen	*Zwischen*	314.77	24	13.115	1.636	0.026
	Innerhalb	25344.56	3162	8.015		
	Total	25659.33	3186			
Konzentr. und sorgf. Arbeiten	*Zwischen*	515.93	24	21.497	1.568	0.039
	Innerhalb	43339.86	3162	13.706		
	Total	43855.79	3186			

Zwischen den (deutschsprachigen) Testlokalen treten wenige signifikante Differenzen auf. Ein multipler Mittelwertsvergleich nach Tukey identifiziert Testlokal 4 mit geringeren Werten bei mehreren Aufgabengruppen und hinsichtlich des Testprozentranges. Die Durchführungsbedingungen an den einzelnen Testorten sind standardisiert – aus den Testdurchführungsprotokollen ergeben sich keine Hinweise auf mögliche Ursachen der auftretenden Unterschiede. Es gibt keinen Anlass, Korrekturen aufgrund der Testlokalzuordnung vorzunehmen.

Tabelle 13: Darstellung der detaillierten Ergebnisse nach Testlokalen (ohne Lokale 3, 23 und 24)

Lokal	Testprozentrang	Muster zuordnen	Med.-naturwiss. Grundverständnis	Schlauchfiguren	Quant. und formale Probleme	Textverständnis	Figuren lernen	Fakten lernen	Diagramme und Tabellen	Konzentr. u. sorgf. Arbeiten	Lokalgrösse
1	47.8	10.0	6.8	13.3	7.6	6.6	9.2	10.0	7.7	9.5	227
2	50.2	10.8	6.5	13.7	7.8	6.5	9.3	10.0	7.8	9.7	230
4	42.5	9.6	6.0	12.6	7.4	6.1	9.1	9.4	7.1	9.1	75
5	50.6	10.0	7.2	13.2	7.8	7.3	9.0	10.3	7.9	9.6	79
6	51.1	10.1	7.3	13.6	7.8	6.6	9.7	10.1	8.1	9.8	78
7	50.3	9.7	7.4	13.2	7.9	6.3	9.4	10.2	8.1	9.5	76
8	49.0	9.6	7.4	13.6	7.9	6.6	9.2	9.8	8.0	9.6	58
9	51.3	10.3	7.2	13.8	8.5	6.6	9.3	9.9	7.7	9.7	60
10	55.4	10.4	7.3	14.1	8.2	6.7	9.8	11.1	8.4	10.0	59
11	49.9	10.1	6.9	13.4	7.4	6.6	9.6	10.2	7.7	10.0	78
12	50.4	9.6	7.0	12.9	8.6	6.8	9.4	10.2	8.0	9.5	78
13	51.7	10.1	6.8	13.2	8.0	6.9	9.3	10.4	8.0	9.9	74
14	57.6	10.3	7.3	14.0	9.1	6.6	9.9	10.7	8.5	10.3	173
15	53.1	10.6	7.1	13.6	8.3	6.4	9.0	10.7	8.2	10.2	178
16	55.5	10.6	7.1	13.5	8.1	6.7	9.4	10.7	8.2	10.4	181
17	55.2	10.4	7.2	13.7	8.8	6.8	9.7	10.3	8.6	10.1	174
18	57.1	10.8	7.5	13.7	8.5	6.7	9.9	10.5	8.4	10.4	123
19	57.2	10.6	7.1	14.1	8.7	6.9	10.0	10.4	8.1	10.5	123
20	54.5	10.5	7.2	13.6	8.8	6.7	9.5	9.8	8.6	10.1	122
21	59.0	10.9	7.6	14.3	8.6	7.3	10.1	10.6	8.4	10.5	120
22	54.2	10.4	7.1	14.1	8.7	6.4	9.5	10.6	8.3	9.8	119
25	55.5	10.8	7.2	13.8	8.5	6.7	9.6	10.4	8.5	10.2	135
26	53.5	10.6	7.1	13.6	8.1	6.8	9.5	10.6	8.3	10.0	168
27	47.3	9.9	6.8	13.2	7.4	6.2	9.1	10.1	7.9	9.1	165
28	48.1	9.7	7.0	13.3	8.3	6.2	9.1	9.3	8.2	9.5	234
29	42.5	9.0	6.6	12.4	8.3	6.3	8.8	9.0	7.1	8.9	108
30	48.9	10.0	6.4	13.2	9.1	6.3	8.6	9.1	8.1	9.5	29
31	42.8	9.6	6.8	12.5	7.7	6.2	9.2	9.0	7.3	9.0	281
32	39.3	9.0	6.8	12.8	7.6	6.2	8.1	9.2	7.4	8.1	39
Total	51.3	10.2	7.0	13.5	8.2	6.6	9.4	10.1	8.0	9.8	3644

Markiert: Mittelwerte für Aufgabengruppen im multiplen Mittelwertsvergleich (Tukey), die signifikant tiefer als in mindestens einem anderen deutschsprachigen Testlokal liegen. Italienisch- und französischsprachige Durchführungen (Lokale 29 bis 32) sind farblich abgesetzt.

2.5 Vergleich für die Geschlechter

Die Forderung nach Gleichbehandlung bedeutet, dass bei gleicher Eignung die gleichen Chancen auf eine Zulassung bestehen. Identische Zulassungsquoten für die Geschlechter dürfen folglich bereits aufgrund der unterschiedlichen Repräsentativität der beiden Gruppen, die sich in einem deutlichen Ungleichgewicht der Teilnehmendenzahlen äussert, nicht erwartet werden.

Nach 2006 hatte sich der Männeranteil in Humanmedizin vorübergehend bei rund 40 % stabilisiert, seit 2014 ist er allerdings weiter gesunken. 2024 fiel er gegenüber dem Vorjahr weiter leicht ab, 29.7 % der Bewerbenden mit gültigem Testergebnis waren männlich, 70.3 % weiblich (über alle Disziplinen).

Tabelle 14: Statistiken für den Test-Prozentrang nach Geschlechtern für Sprache und Disziplin.

		Testsprache			Wunschdisziplin			Gesamt
		deutsch	französisch	italienisch	HM	VM	ZM	
männlich	m	53.9	40.4	41.6	53.7	43.8	42.0	52.6
	s	28.1	26.8	26.0	28.3	26.5	24.5	28.2
	n	985	63	46	982	44	68	1094
weiblich	m	51.8	42.9	44.9	53.0	42.6	40.0	50.7
	s	28.1	27.1	29.1	28.3	25.9	24.8	28.1
	n	2239	257	91	2061	335	191	2587

m: Mittelwert, s: Standardabweichung, n: Personenzahl

Für die zahlenmässig bedeutendste deutschsprachige Teilnehmergruppe wird 2024 ein Unterschied von 2.1 Test-Prozentrangpunkten registriert. Die Differenz ist damit noch deutlich geringer als in den Vorjahren und ist mit einer Effektstärke von 0.045 (Cohen's d) als vernachlässigbar zu bezeichnen.

Für die gesamte Kohorte beträgt die mittlere Differenz 1.9 Prozentrangpunkte, was auch darauf zurückzuführen ist, dass in den französischen und italienischen Sprachgruppen die weiblichen Teilnehmenden höhere Werte erreichen.

Tabelle 15: Kombinierte Statistiken für den Test-Prozentrang nach Geschlechtern, Sprache und Disziplin.

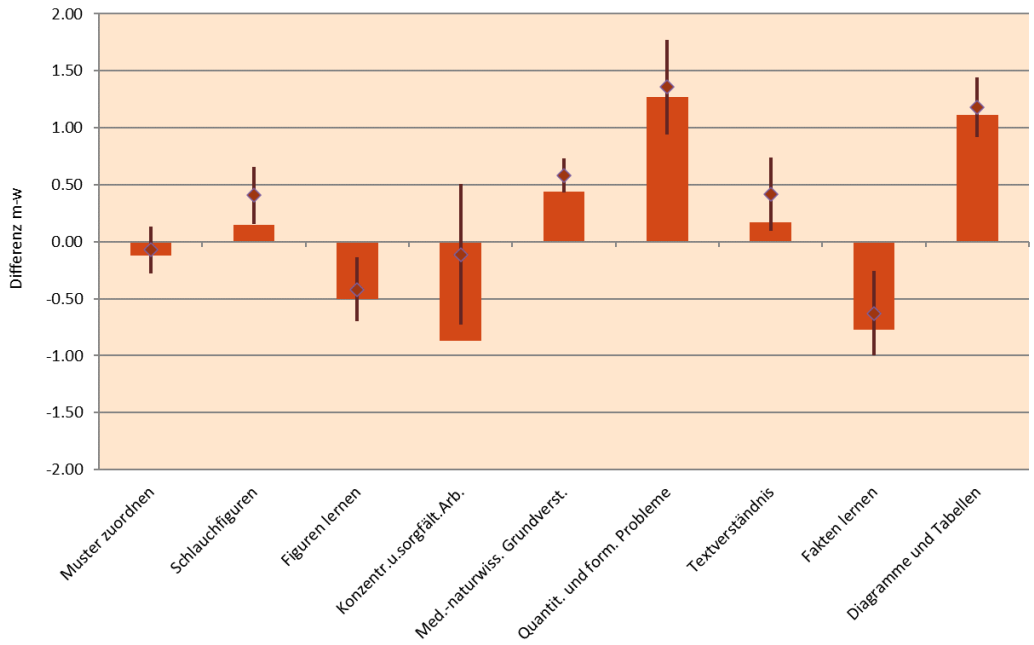
		Humanmedizin			Veterinärmedizin			Zahnmedizin		
		d	f	i	d	f	i	d	f	i
männlich	m	55.0	41.1	41.9	47.9	36.6	*35.7*	41.8	*52.0*	*45.0*
	s	28.2	27.0	26.2	25.7	27.7	*30.2*	24.7	*0.0*	*31.1*
	n	892	49	41	28	13	*3*	65	*1*	*2*
weiblich	m	53.7	48.0	47.5	46.7	32.9	30.2	40.1	*28.0*	*43.0*
	s	28.3	28.0	29.4	26.1	22.4	22.2	24.7	*15.6*	*46.7*
	n	1814	171	76	238	84	13	187	*2*	*2*

Betrachtet man die Unterschiede disziplin- und sprachspezifisch, zeigt sich für die deutschsprachige Gruppe mit Wunschdisziplin Humanmedizin eine geringe aber signifikante Differenz von 1.3 Prozentrangpunkten. In der französischen Sprachgruppe beträgt die Differenz dort 6.9 Prozentrangpunkte zugunsten der weiblichen Teilnehmenden. Die anderen Unterschiede sind nicht signifikant.

Tabelle 16: 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	0.89	0.35	4.09	2805	**0.000**
	VM	homogen	0.02	0.89	0.68	276	0.248
	ZM	homogen	1.11	0.29	1.04	218	0.150
Französisch	HM	homogen	0.32	0.58	2.07	121	*0.020*
Italienisch	HM	homogen	2.94	0.09	0.47	210	0.319

Abbildung 17 verdeutlicht die Vergleichbarkeit der Geschlechterdifferenzen mit den Vorjahren. Die diesjährigen Punktwert-Unterschiede auf Ebene der Aufgabengruppen fallen gegenüber früheren Durchführungen vergleichbar aus, sind einzig bei *Konzentriertes und sorgfältiges Arbeiten* etwas höher als üblich.



(positive Werte bedeuten höhere Ergebnisse für die männliche, negative Werte für die weibliche Stichprobe)

Abbildung 16: Geschlechtsspezifische Differenzen nach Aufgabengruppen (deutschsprachig, Humanmedizin) mit Vergleich (Mittelwert und Standardabweichung) für die Vorjahre 2014-2023.

Unter den gemäss Modell zugelassenen Personen (ohne Nachrücker) die 2024 am Test teilgenommen haben, beträgt der Frauenanteil in Humanmedizin 66.3%, in Veterinärmedizin 88.8%, in Zahnmedizin 72.6%, in Chiropraktik 83.3%.

2.6 Vergleiche nach Wunschhochschulen

Die Tabelle 17 führt die Test-Prozentränge getrennt nach Disziplin und gewünschter Hochschule (erste Wahl) aller deutschsprachigen Kandidaten und Kandidatinnen auf. Nach post-hoc Vergleich nach Tukey sind alle drei Disziplinen homogen, die Unterschiede zwischen den Universitäten sind statistisch nicht relevant.

Tabelle 17: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschhochschulen (d).

	Wunsch-hochschule	n	Mittel-wert	Stand-ardabw.	Quadrat-summe	Df	MQ	F	Sig.
HM	Basel	496	50.7	28.2	10491.0	6	1748.5	2.2	.041
	Bern	796	54.1	27.6					
	ETH	211	57.6	28.6					
	Freiburg	43	60.2	28.6					
	Zürich	1012	54.7	28.5					
	Zürich (LU)	59	54.2	27.7					
	Zürich (SG)	83	54.4	29.6					
VM	Bern	116	47.6	25.7	103.1	1	103.1	0.2	0.697
	Zürich	150	46.3	26.4					
ZM	Basel	50	41.8	26.2	1037.1	2	518.5	0.8	0.429
	Bern	86	37.7	22.4					
	Zürich	116	42.1	25.7					

Unter den zugelassenen Personen treten zwischen den Wunschuniversitäten in der Humanmedizin dementsprechend ebenfalls keine relevanten Unterschiede auf (Tabelle 18).

Tabelle 18: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschhochschulen, nur zugelassene*, deutschsprachige Personen.

	Wunsch-hochschule	n	Mittel-wert	Stand-ardabw.	Quadrat-summe	Df	MQ	F	Sig.
HM	Basel	175	82.1	10.6	1090.5	6	181.7	1.7	.106
	Bern	321	82.3	9.8					
	ETH	96	84.7	10.1					
	Freiburg	22	84.5	11.2					
	Zürich	421	83.3	10.4					
	Zürich (LU)	26	79.7	9.8					
	Zürich (SG)	35	84.9	9.3					
VM	Bern	65	66.1	17.8	262.1	1	262.1	0.9	0.336
	Zürich	75	68.9	15.8					
ZM	Basel	23	66.7	15.2	380.8	2	190.4	0.8	0.451
	Bern	34	61.6	13.7					
	Zürich	56	64.6	16.4					

*Zulassung nach erstem, provisorischem Zulassungsmodell (ohne Nachrücker)

3 Ergebnisse zur Testgüte

3.1 Zuverlässigkeit

Von besonderem Interesse bleiben unter dem Blickwinkel der Neustrukturierung des Tests 2022 die Ergebnisse zur Testgüte. Aufgrund der vorangegangenen Modellierung wären hinsichtlich der Kennwerte keine auffälligen Auswirkungen der Kürzung im Vergleich zur Originalversion des EMS zu erwarten gewesen.

Die zulassungsrelevante Reliabilität (Zuverlässigkeit) über den Gesamttest kann anhand zweier Koeffizienten berechnet werden. Die Reliabilitätsschätzung nach der Testhalbierungsmethode (Teilung nach gerad- und ungeradzahligen Aufgaben) ist eine der gebräuchlichsten Zuverlässigkeitsschätzungen. Wegen der Stichprobengrössen werden nur die Ergebnisse des deutschsprachigen Tests der Schweiz verwendet.

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

Beide Werte sind, neben der Qualität der eigentlichen Aufgaben, auch abhängig von der Aufgabenzahl (im Fall der Konsistenz des Testprofils der Anzahl Aufgabengruppen). Nach zwei Jahren coronabedingter Kürzungen umfasste der Test 2022 wieder mehr Aufgaben, maximal waren statt 118 nun 162 Punkte möglich (die maximale Punktzahl des Originaltests lag bei 178 Punkten).

Der Zuverlässigkeitswert des Punktwertes erreicht 2024 mit .90 einen Wert im üblichen («vor-Corona») Bereich, das Testprofil mit 0.84 weiterhin eine sehr hohe Konsistenz. Beide Werte liegen damit mindestens im Bereich des Originaltests. Dieser Bereich scheint deshalb optimal, weil vergleichbare Werte auch in den Jahren mit einer erfolgreichen Evaluation des Zusammenhanges von Studienerfolg und Eignungstest gefunden worden sind.

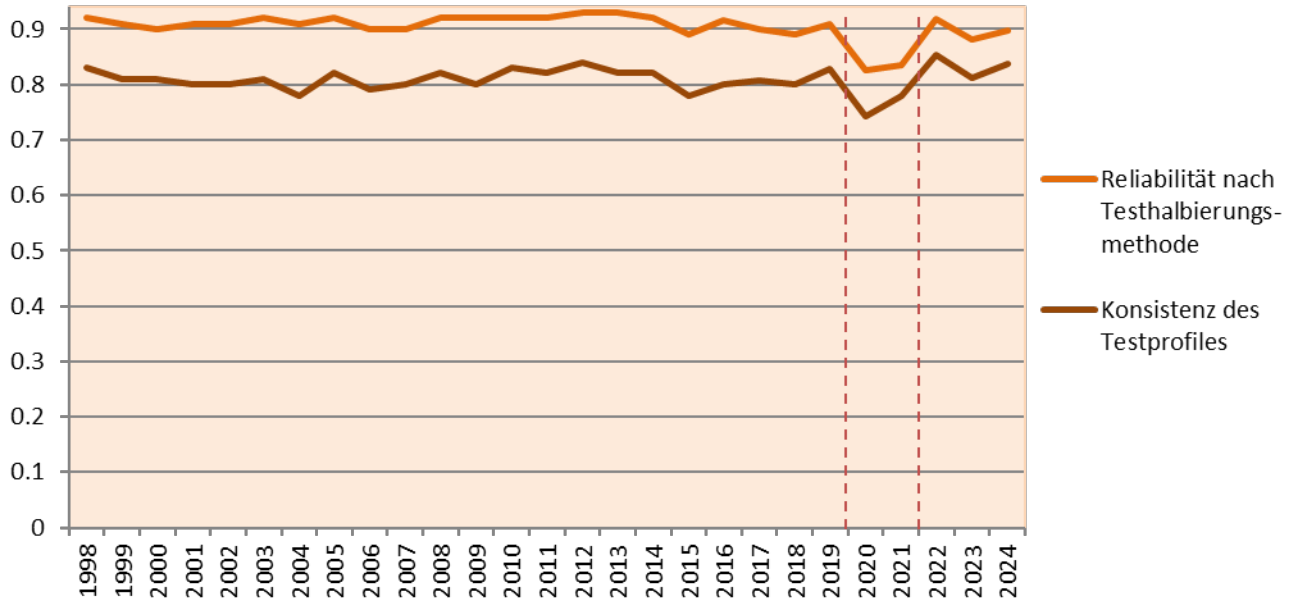


Abbildung 17: Reliabilität des Punktwertes (Split Half) und Konsistenz des Testprofils 1998-2023.

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. Die einzelnen Aufgaben werden im Zug der Testauswertung auf die Eindeutigkeit der Fragestellung hin

geprüft und bezüglich Schwierigkeit und Trennschärfe analysiert. 2024 mussten keine Aufgaben aus Qualitätsgründen aus der Wertung entfernt werden. Die verwendeten Aufgabengruppen weisen die geforderten Varianzen und Schwierigkeiten auf, um zwischen den Kandidierenden optimal zu differenzieren.

3.2 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 19: Korrelationen zwischen Punktwerten der Aufgabengruppen und Gesamt-Punktwert.

	Med.-naturwis. Grundv.	Schlauch-figuren	Quant. u. formale Probl.	Textver-ständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Punkt-wert
Muster zuordnen	.277**	.502**	.274**	.247**	.365**	.452**	.274**	.438**	.675**
Med.-naturw. Grundverst.		.343**	.466**	.467**	.263**	.340**	.499**	.314**	.644**
Schlauchfiguren			.338**	.300**	.397**	.472**	.356**	.464**	.707**
Quant. u. form. Probleme				.410**	.255**	.291**	.540**	.299**	.636**
Textverständnis					.279**	.313**	.437**	.299**	.600**
Figuren lernen						.458**	.255**	.376**	.624**
Fakten lernen							.300**	.418**	.700**
Diagramme und Tabellen								.335**	.650**
Konzent. u. sorgf. Arbeiten									.687**

Die Varianzanteile der unrotierten Lösung für die sechs statt neun Aufgabengruppen entsprechen den bekannten Grössenordnungen (Tabelle 20).

Tabelle 20: Faktorenanalyse: Verlauf der unrotierten Lösung (2020-2024).

Unrotierte Lösung															
Faktor	Eigenwert					% Varianz					% Kumuliert				
	2020	2021	2022	2023	2024	2020	2021	2022	2023	2024	2020	2021	2022	2023	2024
1	2.73	2.95	4.17	3.74	3.97	45.52	49.17	46.37	41.51	44.06	45.52	49.17	46.37	41.51	44.06
2	0.94	0.97	1.14	1.28	1.23	15.63	16.09	12.68	14.23	13.62	61.15	65.26	59.06	55.74	57.68
3	0.71	0.60	0.72	0.82	0.69	11.86	9.94	8.01	9.07	7.67	73.01	75.20	67.06	64.81	65.35
4	0.67	0.57	0.61	0.63	0.60	11.17	9.55	6.74	7.05	6.69	84.17	84.75	73.81	71.86	72.03
5	0.52	0.48	0.57	0.60	0.57	8.60	7.97	6.34	6.62	6.36	92.77	92.72	80.14	78.48	78.40
6	0.43	0.44	0.53	0.52	0.52	7.23	7.28	5.89	5.75	5.78	100	100	86.03	84.22	84.18
7			0.44	0.50	0.50			4.91	5.53	5.50			90.94	89.75	89.68
8			0.42	0.47	0.48			4.63	5.25	5.35			95.57	95.00	95.03
9			0.40	0.45	0.45			4.43	5.00	4.97			100	100	100

Tabelle 21: Varianzanteile der 2-Faktorenlösung 2020-2024 (rotiert).

Varimax-rotierte Lösungen															
Eigenwert					% Varianz					% Kumuliert					
2020	2021	2022	2023	2024	2020	2021	2022	2023	2024	2020	2021	2022	2023	2024
2.3	2.4	2.7	2.7	2.7	37.7	40.7	30.0	29.5	30.2	37.7	40.7	30.0	29.5	30.2
1.4	1.5	2.6	2.4	2.5	23.5	24.6	29.1	26.2	27.5	61.2	65.3	59.1	55.7	57.7

Ein allgemeiner Fähigkeitsfaktor klärt 44 % der Varianz auf – es gibt einen grossen „gemeinsamen Nenner“ in allen Aufgabengruppen. Mit der Zwei-Faktorenlösung können 58 % der Varianz aufgeklärt werden. Die Aufgabe *Konzentriertes und sorgfältiges Arbeiten* ist je nach konkreter Aufgabenstellung eher dem einen oder dem anderen Faktor zuzuordnen – misst aber in jedem Fall (wie intendiert) eine durch die anderen Aufgabengruppen weniger berücksichtigte Fähigkeit. 2024 ist diese eher dem Faktor „schlussfolgerndes Denken“ zuzuordnen.

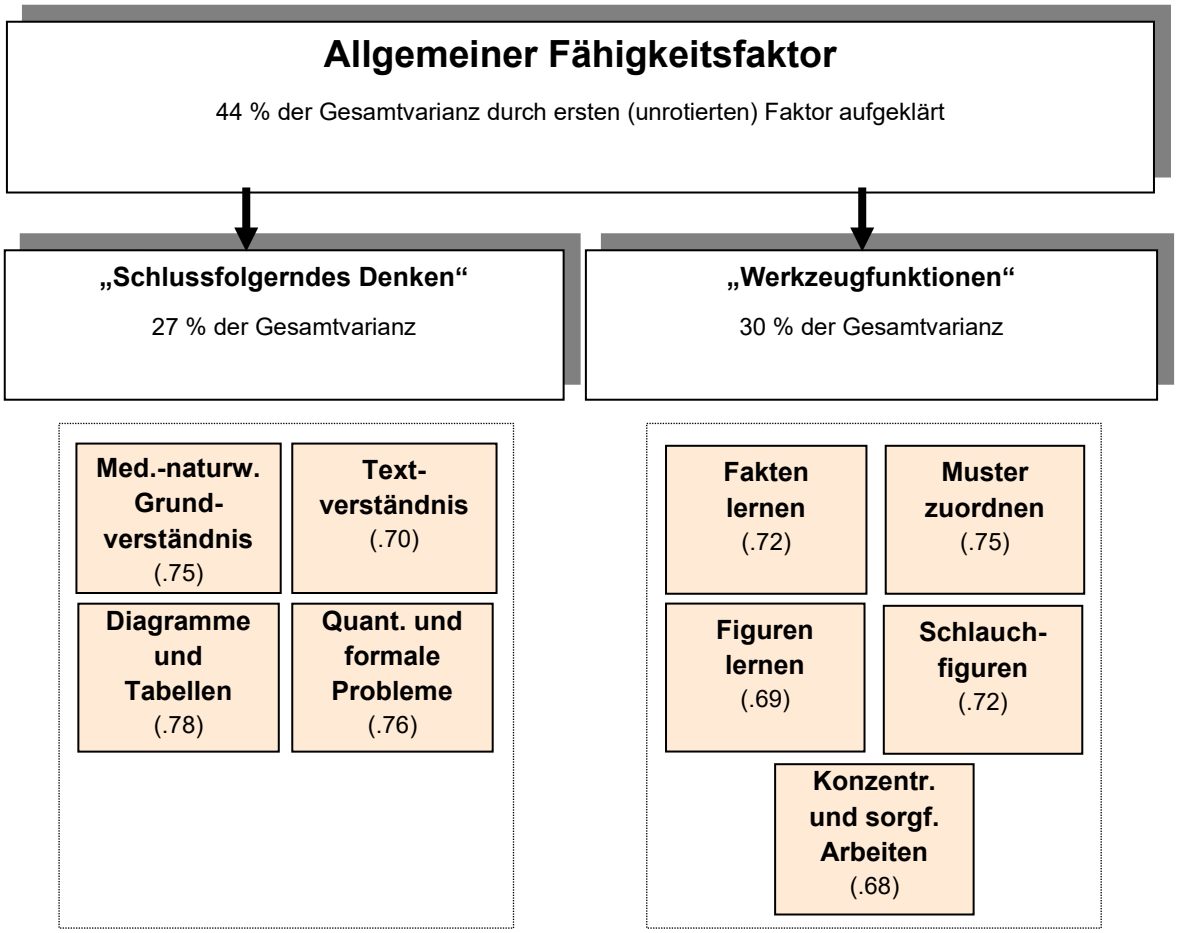


Abbildung 18: Struktur der Aufgabengruppen des EMS, hierarchische Faktorenstruktur 2024 mit Faktorladungen der Aufgabengruppen auf rotierten Faktoren.

Die aktuellen Werte entsprechen durch den wiederhergestellten Faktor „Werkzeugfunktionen“ den Ergebnissen vor den Corona-Jahren bis 2019. Insgesamt zeigt die neue Teststruktur keine nachteiligen Auswirkungen auf die Testgüte.

4 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.

Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995a). Un test d'aptitudes aux études de médecine est-il faisable en Suisse? Bulletin des médecins suisses, 7, S. 267 - 274.

Hänsgen, K.-D., Hofer, R. & Ruefli, D. (1995b). Der Eignungstest für das Medizinstudium in der Schweiz. Schweizerische Ärztezeitung, 37, S. 1476 - 1496.

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.

4.1 Frühere Berichte des ZTD

Über EMS-Seite: http://www.unifr.ch/ztd/de/ems/berichte.html