

swissuniversities

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

Anmeldung und Zulassung zum Medizinstudium Durchführung des Selektions- verfahrens 2025

**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. Voranmeldungen per 15. Februar	3
3. Testteilnahmen Eignungstest für das Medizinstudium 2025	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 2025

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, Veterinär- und Zahnmedizin) ist insbesondere aufgrund der hohen Kosten des Medizinstudiums und der maximal verfügbaren praxisbezogenen Ausbildungsplätze in den Spitätern 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 Universität 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 2025/2026 blieben die Studienplatzkapazitäten in der Human- und Veterinärmedizin gegenüber dem Vorjahr unverändert. In der Humanmedizin standen somit im ersten Jahr des Bachelorstudiums schweizweit 2'195 Studienplätze zur Verfügung, im ersten Jahr des Masterstudiums 1'472. In der Veterinärmedizin waren es im Bachelor 172 und im Master 140 Studienplätze. In der Zahnmedizin bleiben die Kapazitäten 2025/2026 im Master unverändert. Die Universität Genf erhöhte die Kapazität im ersten Bachelorstudienjahr, wobei der Zugang zum zweiten Studienjahr weiterhin beschränkt bleibt. Schweizweit standen in der Zahnmedizin im ersten Jahr des Bachelorstudiums 202 und im ersten Jahr des Masterstudiums 140 Studienplätze zur Verfügung.

Studienplatzkapazitäten im zweiten Bachelorstudienjahr an den Universitäten ohne Numerus clausus

An den Universitäten ohne Numerus clausus werden im ersten Bachelorstudienjahr alle Personen immatrikuliert, die die Anforderungen erfüllen, beispielsweise bezüglich Reifezeugnis und Wohnsitz in der Schweiz. Hingegen ist der Zugang zum zweiten Bachelorstudienjahr mittels Concours beschränkt. Vorgesehen ist, dass an der Universität Lausanne im kommenden Jahr 270 Studienplätze im zweiten Bachelorstudienjahr in der Humanmedizin angeboten werden. An der Universität Genf sind es 158 Studienplätze in der Humanmedizin und 30 in der Zahnmedizin.

2.2. Voranmeldungen per 15. Februar

Bis zur Frist vom 15. Februar 2025 gingen insgesamt 7'191 Voranmeldungen für alle Hochschulen und Disziplinen ein. 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 2016-2025

Die Anmeldungen an den Hochschulen mit Numerus clausus (Universitäten Basel, Bern, Freiburg, Zürich, Università della Svizzera italiana und ETH Zürich) betragen 2025 für alle Disziplinen zusammengenommen 361% der vorhandenen Kapazitäten (5'191 Anmeldungen gegenüber 1'439 Studienplätzen).

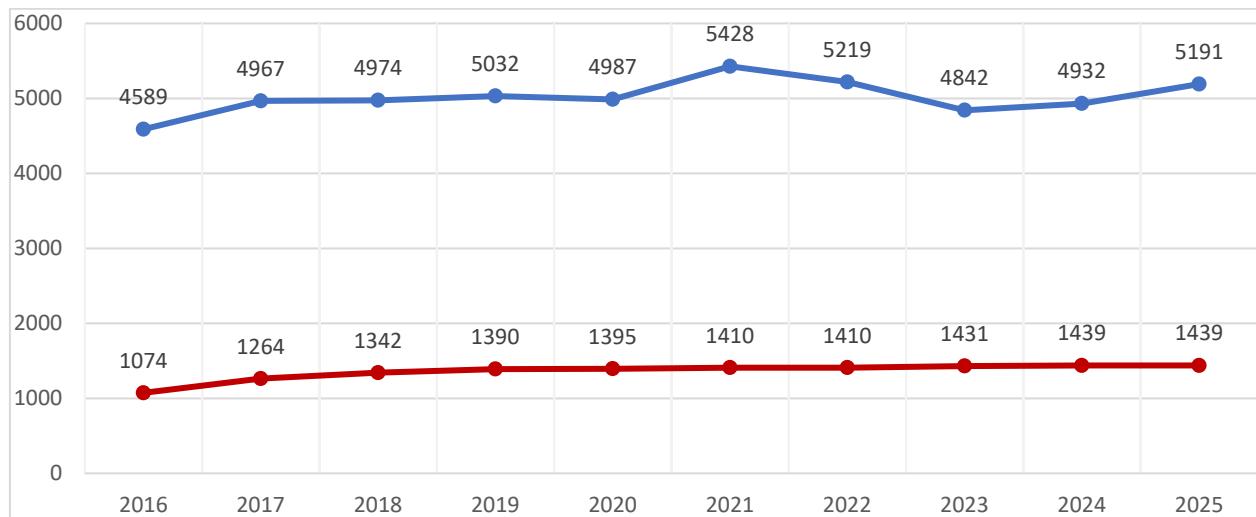


Abbildung 2: Anmeldungen (blau) und Kapazitäten (rot) an den Hochschulen mit Numerus clausus 2016-2025

3. Testteilnahmen Eignungstest für das Medizinstudium 2025

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. 2025 meldeten sich bis zum 15. Februar 5'191 Personen an einer Hochschule mit Numerus clausus an, 3'967 davon meldeten sich bis zur Frist Ende Mai zum EMS an, 3'811 Personen traten zum Test an und 3'809 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 auf Französisch (351 Personen, 9% der Testteilnehmenden) oder Italienisch (156 Personen, 4%) teil als auf Deutsch (3'302 Personen, 87%). Veterinärmedizin wird in der Schweiz nur in Zürich und Bern angeboten, deshalb liegt der Anteil Frankophoner in dieser Disziplin höher (119 von 405 Personen, 29%).

3.1. Testteilnahmen gesamt und nach Disziplin

2025 nahmen insgesamt (alle Disziplinen) etwas mehr Personen am EMS teil als im Vorjahr.

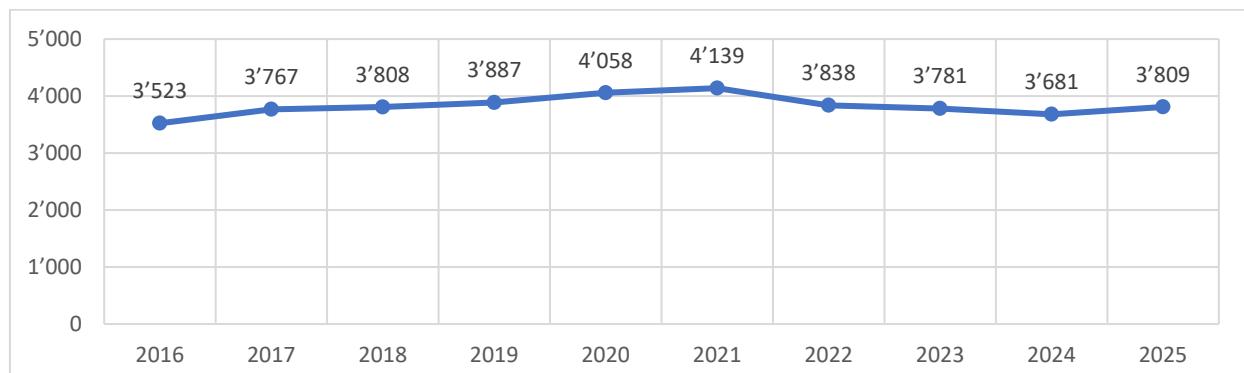


Abbildung 3: Testteilnahmen gesamt 2016-2025

In der Human- und Veterinärmedizin waren 2025 gegenüber 2024 mehr Testteilnahmen zu verzeichnen. In der Zahnmedizin sank die Zahl der Testteilnehmenden ganz leicht.

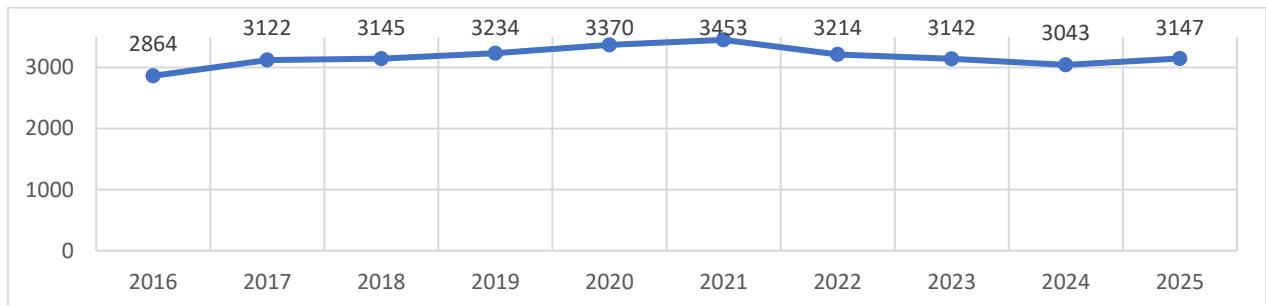


Abbildung 4: Testteilnahmen Humanmedizin inkl. Chiropraktik 2016-2025



Abbildung 5: Testteilnahmen Veterinärmedizin 2016-2025

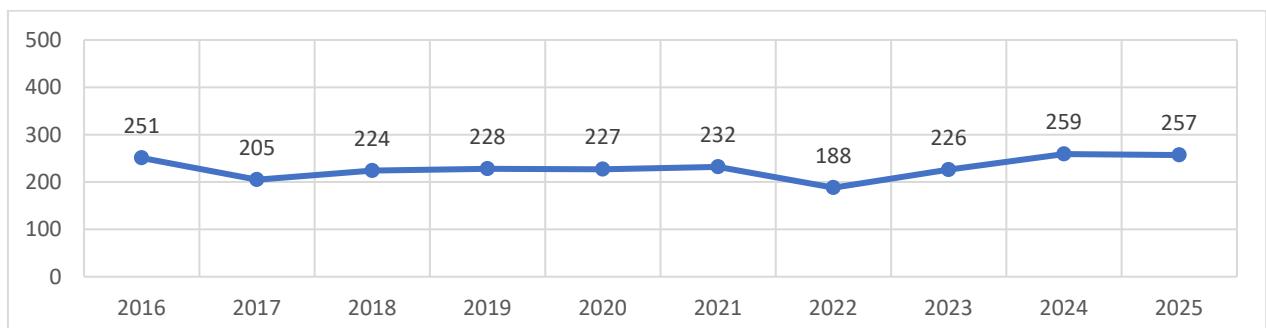


Abbildung 6: Testteilnahmen Zahnmedizin 2016-2025

3.2. Testteilnahmen nach Geschlecht und Disziplin

Auch 2025 nahmen mehr Frauen als Männer am EMS teil. In der Humanmedizin (inkl. Chiropraktik) waren es 2'097 Frauen und 1'050 Männer, in der Veterinärmedizin 360 Frauen und 45 Männer und in der Zahnmedizin 186 Frauen und 71 Männer.

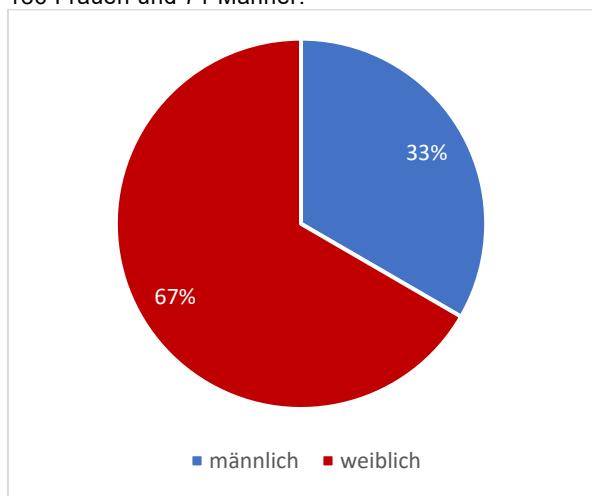


Abbildung 7: Testteilnahmen 2025 nach Geschlecht HM inkl. CP

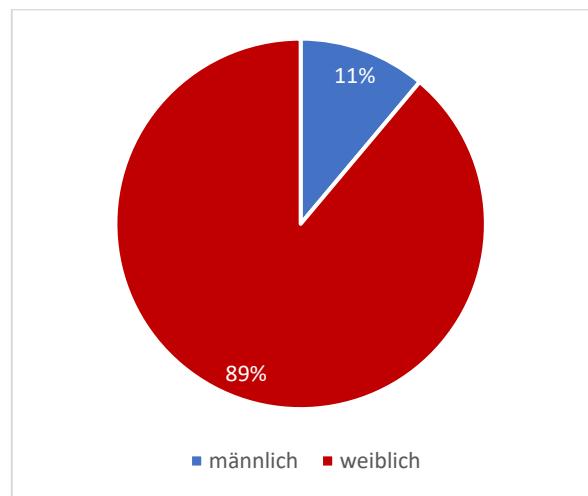


Abbildung 8: Testteilnahmen 2025 nach Geschlecht VM

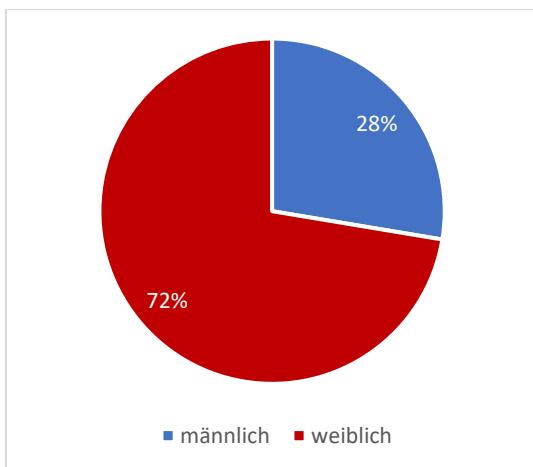


Abbildung 9: Testteilnahmen 2025 nach Geschlecht ZM

3.3. Testteilnahmen nach Kanton

Von den 3'809 Personen, die 2025 am EMS teilnahmen und ein gültiges Ergebnis erzielten, hatten 64 ihren Wohnsitz im Ausland und 3'745 in der Schweiz. Die nachfolgende Tabelle zeigt die Aufteilung auf die Kantone (Wohnsitz).

Wohnkanton	Anzahl
AG	360
AI	6
AR	17
BE	551
BL	210
BS	160
FR	162
GE	20
GL	10
GR	76
JU	19
LU	220
NE	31
NW	15
OW	12
SG	226
SH	34
SO	124
SZ	71
TG	103
TI	143
UR	12
VD	88
VS	92
ZG	87
ZH	896
Ausland	64
Total	3'809

Tabelle 1: Testteilnahmen 2025 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 2025, 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 Testteilnehmenden ein Studienplatz zugeteilt, dass die gesamte Aufnahmekapazität aller Hochschulen im ersten Studienjahr ausgeschöpft wird. Zusätzlich zu den Testteilnehmenden des aktuellen Jahres 2025 müssen dabei jeweils auch die Kandidatinnen und Kandidaten berücksichtigt werden, die den Testwert des Vorjahres anrechnen lassen (2025: 93 Bewerbungen).

5.1. Zulassungen pro Disziplin

In der Humanmedizin (inkl. Chiropraktik) haben 39% der Studienanwärterinnen und -anwärter einen Studienplatz erhalten (1'242 Personen), 61% mussten abgewiesen werden (1'961 Personen). In der Veterinärmedizin konnten 45% der Bewerbungen (186 Personen) zugelassen werden, 55% der Bewerbungen (226 Personen) mussten abgewiesen werden. In der Zahnmedizin konnten 48% (135 Personen) zugelassen werden, 52% (147 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).

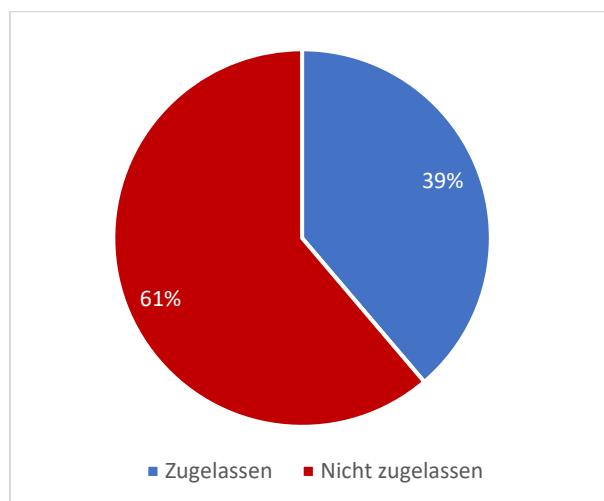


Abbildung 10: Zulassungen 2025 Humanmedizin inkl. Chiropraktik

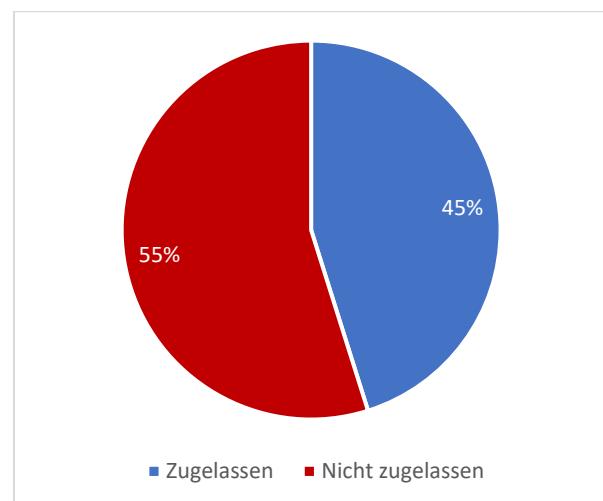


Abbildung 11: Zulassungen 2025 Veterinärmedizin

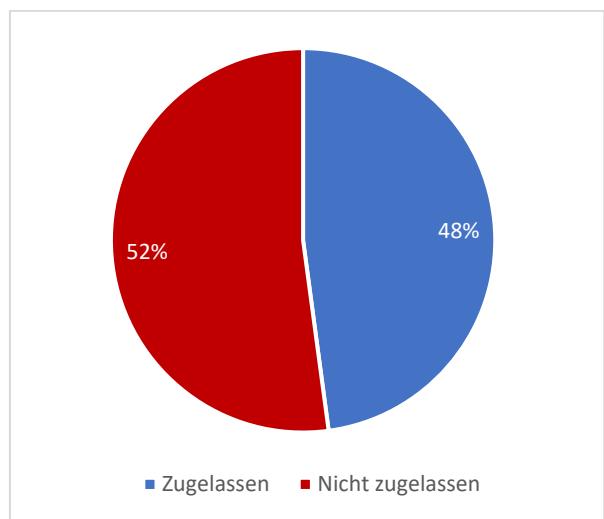


Abbildung 12: Zulassungen 2025 Zahnmedizin

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 2025 zu vielen Umleitungen (198 Personen) zwischen den Hochschulen. 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.

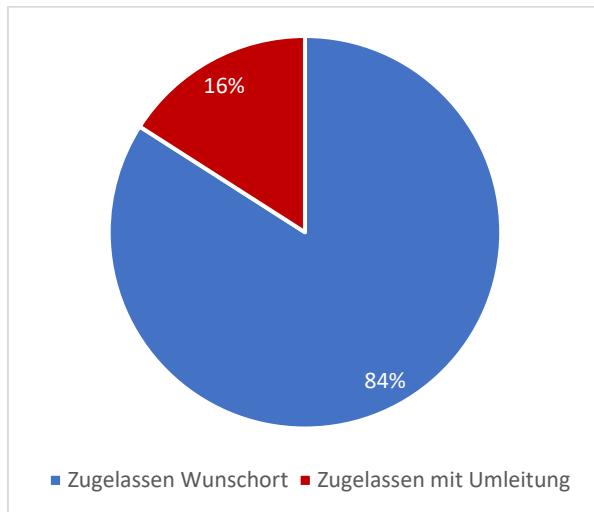


Abbildung 13: Umleitungen 2025 Humanmedizin

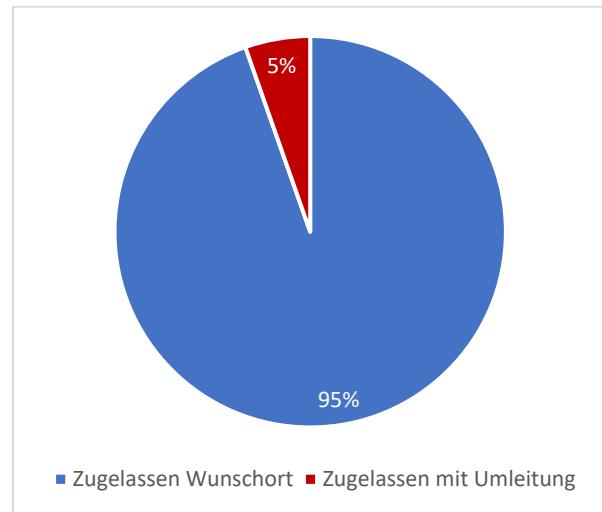


Abbildung 14: Umleitungen 2025 Veterinärmedizin

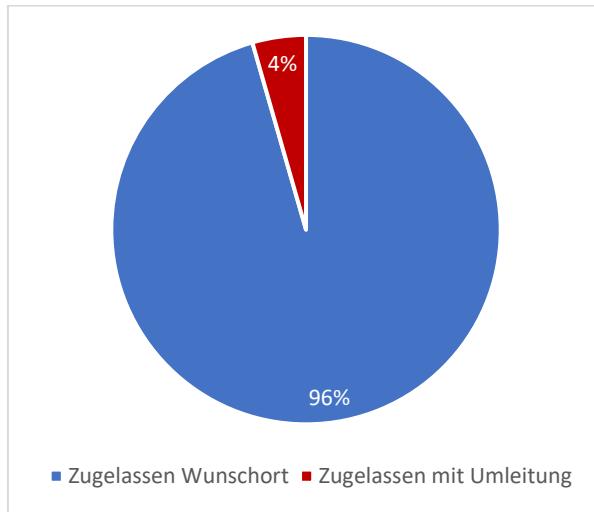


Abbildung 15: Umleitungen 2025 Zahnmedizin

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

5.3. Verzicht auf einen Studienplatz

2025 verzichteten bis zehn Tage vor Vorlesungsbeginn 115 Bewerberinnen und Bewerber nach dem Zulassungentscheid auf den ihnen zugeteilten Studienplatz, davon 93 in der Humanmedizin (inkl. Chiropraktik), 13 in der Veterinär- und neun in der Zahnmedizin. Dank den Überbuchungen (siehe 5.1, Zulassungen pro Disziplin) wurden insgesamt nur wenige Nachnominierungen benötigt. An einer Hochschule wurden jedoch die Überbuchungen in Humanmedizin reduziert, was zu mehr Nachnominierungen führte. Insgesamt waren es 2025 in der Humanmedizin 22 Nachnominierungen. In der Veterinärmedizin und in der Zahnmedizin waren keine Nachnominierungen nötig. Zu Beginn des Studiums blieben 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 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 über die Anzahl über 40-jährige immatrikulierte Personen Bericht erstattet.

Im Herbst 2025 wurden neun Personen im Alter von über 40 Jahren in einem medizinischen Studiengang immatrikuliert. Acht Personen wurden an Hochschulen ohne Numerus clausus und Eignungstest ins erste Studienjahr zugelassen. Eine Person (zu Studienbeginn 43 Jahre alt) erhielt aufgrund ihres Resultats am Eignungstest einen Studienplatz in der Humanmedizin an einer Hochschule mit Numerus clausus.

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

Der Hochschulrat der SHK beauftragte swissuniversities und das ZTD an seiner Sitzung vom 22. Februar 2024, ihn über ihre Zusammenarbeit und die Umsetzung der Empfehlungen A bis E aus der Follow-up Evaluation (durchgeführt im Zeitraum März bis Oktober 2023) auf dem Laufenden zu halten. Mit dem Bericht 2024 kamen swissuniversities und das ZTD dieser Verpflichtung erstmals gemeinsam nach.

Seither konnte bei einigen grundlegenden Fragen eine Klärung erreicht werden. swissuniversities vereinbarte mit dem neuen Direktor des ZTD, der dieses Amt am 1. Mai 2025 übernommen hat, in gewissen Punkten eine Neuorganisation innerhalb des Auftragsverhältnisses zwischen swissuniversities und dem ZTD. So soll das ZTD vermehrt den Auftrag EMS gewissermassen ‘unternehmerisch’ in eigener Verantwortung durchführen. Dies betrifft etwa die Weiterbildung der Mitarbeitenden. Damit wird die Empfehlung *D: Finanzierung von Weiterbildungen klären* hinfällig. Dies liegt neu in der alleinigen Verantwortung des ZTD unter der Leitung des Direktors.

Bei den Empfehlungen *A: Prozesse in einem ZTD-internen Qualitätsmanagementsystem abbilden* und *B: Jährliche Zeiterfassung beibehalten* ist swissuniversities neu ebenfalls nicht mehr involviert. 2025 gab es diesbezüglich keine Änderungen. swissuniversities wird in Absprache mit dem Direktor des ZTD ab 2026 in diesem Bericht über allfällige neue Entwicklungen informieren.

In der Folge dieser Neuorganisation wird das ZTD swissuniversities ab 2027 jeweils ein Globalbudget vorlegen. Die Kosten für den Auftrag EMS werden auf dem Stand des von der SHK bereits genehmigten Budgets 2026 eingefroren. Gleichzeitig erhält das ZTD so auch mehr unternehmerische Freiheiten und kann sich flexibler selbst organisieren.

Zur Empfehlung *C: Möglichkeiten der präventiven Ausrichtung der Risikoszenarien prüfen* gab es 2025 keine Änderungen. Die bestehenden Risikoszenarien werden seit 2024 breiter kommuniziert, insbesondere den Koordinator:innen der acht Testorte sowie den Testleiterinnen und Testleitern. Dies trägt zur präventiven Wirkung bei.

Auch bei Empfehlung *E: Den Stellenwert der Durchführung von Probeläufen bekannt machen* handelt es sich um eine Daueraufgabe. swissuniversities bittet die Mitglieder der Konferenz der Schweizerischen Gymnasialrektorinnen und Gymnasialrektoren (KSGR) jedes Jahr, kostenlose Probeläufe durchzuführen. Seit 2024 weist sie in der Kommunikation noch stärker als vorher auf die Bedeutung und Wirksamkeit der Probeläufe hin. Es liegt jedoch in der Verantwortung der Gymnasien, Probeläufe durchzuführen.

2025 / BERICHT 32

EIGNUNGSTEST FÜR DAS MEDIZINSTUDIUM (EMS)

BERICHT ÜBER DIE DURCHFÜHRUNG UND ERGEBNISSE

Dr. B. Spicher

Dr. Spicher, Benjamin (2025)
EMS Eignungstest für das Medizinstudium
Berichte des Zentrums für Testentwicklung, Nr. 32,
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



Inhalt

1	Beschreibung des verwendeten Eignungstests.....	4
1.1	Berechnung der Werte.....	6
1.2	Mittlerer Rangplatz der Aufgabengruppen.....	7
2	Testanwendung 2025	9
2.1	Verteilungsprüfung.....	9
2.2	Objekte im Raum	9
2.3	Äquivalenz der Sprachversionen	12
2.4	Effekte der Korrektur	24
2.5	Vergleichbarkeit der Testlokale	27
2.6	Vergleich für die Geschlechter.....	30
2.7	Vergleiche nach Wunschhochschulen.....	32
3	Ergebnisse zur Testgüte	33
3.1	Zuverlässigkeit	33
3.2	Binnenstruktur.....	34
4	Literatur.....	36
4.1	Frühere Berichte des ZTD	37

1 Beschreibung des verwendeten Eignungstests

Der **Eignungstest für das Medizinstudium in der Schweiz (EMS)** wird seit 1998 erfolgreich eingesetzt. Im Laufe seiner Geschichte wurde das Verfahren mehrfach angepasst – von Kürzungen im Testablauf über die Modifikation „innerer Mechanismen“ einzelner Aufgabentypen bis hin zur Einführung neuer Aufgabengruppen.

Seit 2022 umfasst der EMS neun Aufgabengruppen, in denen jeweils maximal 18 Punkte erzielt werden können. Die Gruppen *Muster zuordnen*, *Fakten lernen*, *Figuren lernen* sowie *Konzentriertes und sorgfältiges Arbeiten* werden jährlich durch das Zentrum für Testentwicklung (ZTD) neu erstellt. Die textintensiveren Aufgaben werden in Deutschland einmalig als sogenannte Einstreuaufgaben vorerprobt und anschliessend für den ersten zulassungsrelevanten Einsatz im EMS verwendet.

In den vergangenen Jahren zeigte sich, dass in der Aufgabengruppe *Schlauchfiguren* regelmässig sehr hohe Punktwerte erzielt wurden. Es ist anzunehmen, dass dies auf die Anwendung spezifischer Lösungsstrategien zurückzuführen war, die dazu führten, dass das intendierte zu erfassende Merkmal – das räumliche Denken – nicht mehr zuverlässig abgebildet wurde.

Um diesem Merkmal wieder das gebührende Gewicht zukommen zu lassen, wurde durch das ZTD eine neue Aufgabengruppe entwickelt, die erstmals im Jahr 2025 eingesetzt wurde. Der Analyse dieser Aufgabengruppe *Objekte im Raum* wird im vorliegenden Bericht entsprechend Platz eingeräumt.

Ablauf und Struktur des aktuellen Tests sind in Tabelle 1 dargestellt.

Tabelle 1: Struktur und Ablauf des EMS 2025

Aufgabengruppe	Max. Punkte	Zeit in Min	Zeit pro Aufgabe (min)
Muster zuordnen	18	16	0,9
Medizinisch-naturwissenschaftliches Grundverständnis	18	45	2,5
Objekte im Raum	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	

Die erzielten Punkte werden zu einem Gesamtwert addiert, welcher wiederum in einen Testprozentsrang umgerechnet wird, der die Leistungen der Teilnehmenden mit jenen der Gesamtkohorte vergleicht.

Die Anforderungen an die jährlich neu entwickelten 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.

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. Eine Aufgabe mit ungenügenden Kennwerten in der Aufgabengruppe *Muster zuordnen* wurde 2025 für alle Teilnehmenden als korrekt gelöst gewertet.

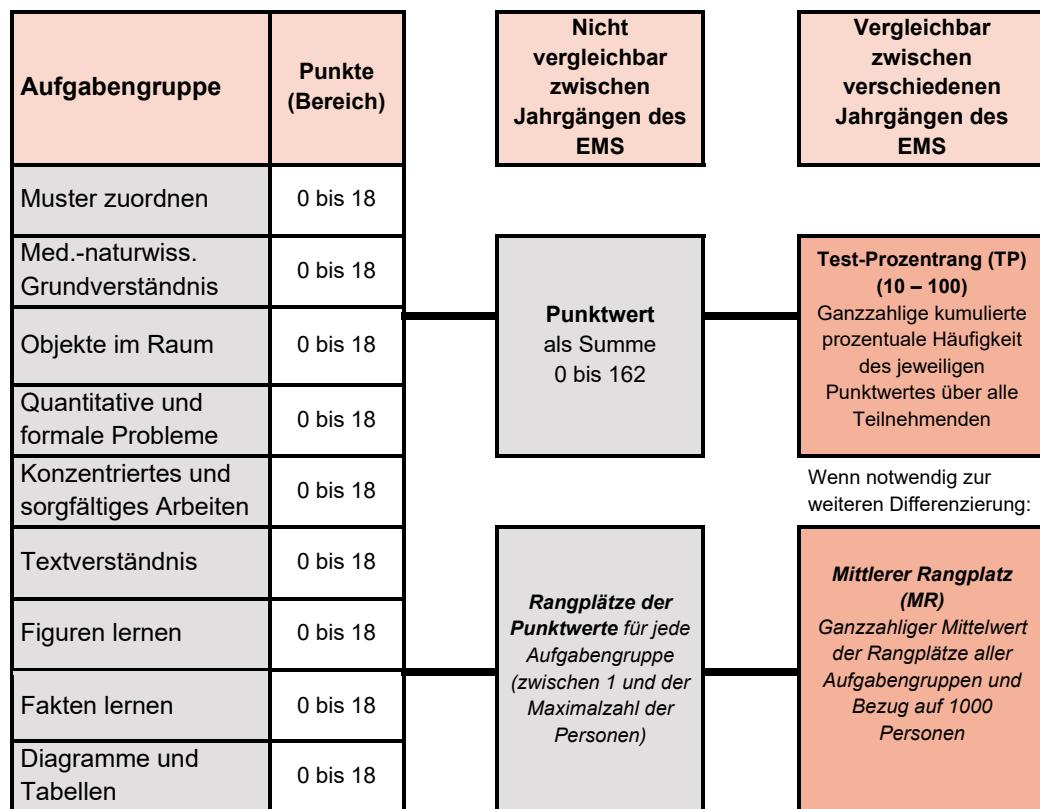


Abbildung 1: Punktwerte der einzelnen Aufgabengruppen 2025 und ihre Zusammenführung über den Punktwert zum Test-Prozentrang und mittleren Rangplatz.

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 übersehnenen 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.

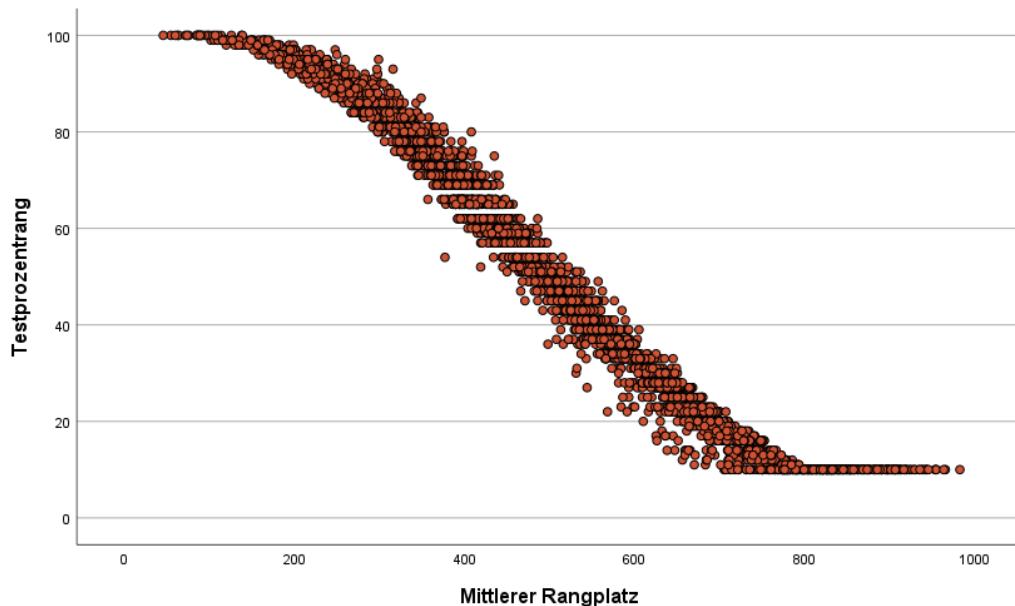


Abbildung 2: Beziehung zwischen Test-Prozentrang (TP) und mittlerem Rangplatz (MR) für 2025 (Werte kleiner als 10 auf werden auf TP 10 angehoben).

Der mittlere Rangplatz MR bewegt sich 2025 zwischen 46 und 982. Niedrige Werte stehen für bessere Leistungen. Mittlerer Rangplatz und Test-Prozentrang korrelieren mit -0.982.

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

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 übertrffen hat als Kandidat B.

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

2 Testanwendung 2025

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	Objekte im Raum	Textverständnis	Med.-naturwiss. Grundv.	Figuren lernen	Fakten lernen	Muster zuordnen	Diagramme und Tabellen	Konzentr. und sorgf. Arbeiten
Mittelwert	88.0	8.1	11.5	7.5	7.9	10.4	10.5	10.8	8.9	12.4
Median	89	8	12	7	8	10	11	10	9	13
Modalwert	98	7	12	7	7	11	11	9	9	13
Stand.-abweichung	18.9	3.1	3.0	2.5	2.8	3.2	3.6	3.8	2.8	3.7
Spannweite	122	18	17	18	17	18	18	17	18	18
25. Perzentil	75	6	9	5	6	8	8	8	7	11
75. Perzentil	102	10	14	10	10	13	13	13	11	14
Schiefe	-0.2	0.3	-0.3	0.3	0.1	-0.1	0.0	0.2	0.0	-1.6
Exzess	-0.3	-0.3	-0.4	-0.1	-0.4	-0.4	-0.6	-0.7	-0.3	4.1

Für die Aufgabengruppe *Textverständnis* wurden 2025 die tiefsten mittleren Punktwerte registriert, gefolgt von *Medizinisch-naturwissenschaftlichem Grundverständnis*. Die höchsten mittleren Punktwerte lieferte aufgrund einer eher schiefen Verteilung die Aufgabengruppe *Konzentriertes und sorgfältiges Arbeiten*. Alle Aufgabengruppen differenzieren sehr gut im relevanten Wertebereich.

2.2 Objekte im Raum

Die neue Aufgabengruppe behält die Idee des Würfels analog zu den *Schlauchfiguren* bei. Allerdings verändert sich nun nicht mehr die Position, und damit die Betrachtungsrichtung des Beobachters, sondern der Betrachter bleibt statisch vor dem Würfel, der nun zweimal nach links/rechts gedreht, und/oder nach oben/unten gekippt wird. Die Teilnehmenden haben zu entscheiden, welche der vorgeschlagenen Kombinationen von Dreh- und Kippbewegungen zu der zweiten Ansicht geführt hat (Abbildung 3).

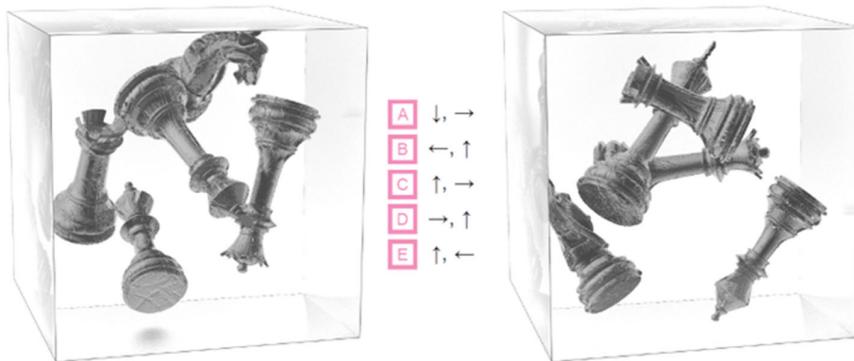


Abbildung 3: Beispiel Objekte im Raum

Der Umstand, dass zwei Veränderungen an der Lage des Würfels vorgenommen werden, hat zur Folge, dass die Konfiguration der Objekte nach der ersten Dreh-/Kippbewegung mental repräsentiert werden muss – der Lösungsweg ergibt sich also nicht direkt aus dem Vergleich der beiden Darstellungen, wodurch analytische Lösungs-Strategien aufgrund der zweidimensionalen Abbildungen, also weitgehend ohne Berücksichtigung der räumlichen Konfiguration, mindestens erschwert werden sollen.

Ebenfalls angepasst wurden die Würfeldarstellungen. Die leicht gedrehte und nach unten gekippte Ansicht wurde anlässlich einer Pilot-Befragung gegenüber der perspektivfreien Ansicht bevorzugt und als hilfreich bei der Einschätzung der Lage der Objekte beurteilt.

2.2.1 Vorbereitungsmaterialien

Bei der Einführung einer neuen Aufgabengruppe am EMS war es für die Wahrung der Testfairness zentral, die zukünftigen Teilnehmenden frühzeitig über die Neuerungen zu informieren. Während die offiziellen Informationen und Übungsmaterialien üblicherweise im Februar des Testjahres veröffentlicht werden, erschienen die entsprechenden Informationen für den EMS 2025 bereits im November 2024. Sie beinhalteten folgende Elemente (in allen drei Sprachen):

- Information über die Änderung auf den Webseiten von swissuniversities und ZTD
- TestInfo: Beschreibung der Änderungen hinsichtlich der Testabläufe
- Beispielaufgaben: Die 8 Beispielaufgaben für die Aufgabengruppe *Schlauchfiguren* wurden durch 8 Beispielaufgaben für *Objekte im Raum* ersetzt, inklusive ausführlicher Beschreibung der Lösungswege.
- Probelauf *Objekte im Raum*: Für die Durchführung von Probelaufen wurden weitere 18 Aufgaben zur Verfügung gestellt. Personen, die beabsichtigten, an Probelaufen teilzunehmen, wurden darauf hingewiesen, diese Aufgaben erst nach dem Probelauf eingehender zu analysieren. Personen, die nicht an einem Probelauf teilnehmen wollten, konnten die Aufgaben ebenfalls für die individuelle Vorbereitung nutzen.
- Probelauf für Testleiter: Das Dokument für Durchführende eines (schulischen) Probelaufs beinhaltete die aktualisierten Hinweise für die Durchführung mit der neuen Aufgabengruppe.

2.2.2 Itemkennwerte

Wie bei allen Aufgabengruppen wurde eine Verteilung von leichten, mittleren und schwierigen Aufgaben angestrebt. Aus Vorstudien mit ähnlichen Aufgaben konnten entsprechende Erfahrungen gesammelt werden, aber wie bei anderen durch das ZTD erstellten Aufgabengruppen gilt auch hier, dass die eigentlichen Testaufgaben nicht vorerprobt werden konnten.

Die 18 verwendeten Items weisen die in Tabelle 3 dargestellten Kennwerte auf. Insgesamt erreicht die Aufgabengruppe eine Skalenreliabilität (Cronbachs Alpha) von .697. Damit erreicht sie einen Wert im oberen Bereich der anderen Aufgabengruppen. Für den Zulassungsentscheid ist lediglich die Gesamtreliabilität des Tests massgeblich, nicht jene einzelner Aufgabengruppen.

Tabelle 3: Itemkennwerte für Objekte im Raum

	Mittelwert	Std.-Abweichung	Korrigierte Item-Skala Korrelation
Item 1	.98	.156	.135
Item 2	.86	.350	.195
Item 3	.76	.429	.223
Item 4	.87	.333	.277
Item 5	.73	.442	.377
Item 6	.77	.422	.411
Item 7	.76	.429	.326
Item 8	.37	.483	.195
Item 9	.58	.494	.360
Item 10	.64	.479	.345
Item 11	.64	.480	.343
Item 12	.81	.395	.268
Item 13	.69	.461	.377
Item 14	.50	.500	.326
Item 15	.39	.489	.203
Item 16	.22	.417	.066
Item 17	.35	.478	.244
Item 18	.54	.498	.325

N=3811

Im Mittel wurden in dieser Aufgabengruppe 11.5 Punkte erreicht, die Standardabweichung beträgt 3.0. Die Differenzierungsfähigkeit ist damit sehr gut gewährleistet. In der Tendenz kann der Anteil an schwierigen Aufgaben zukünftig noch etwas erhöht werden.

2.3 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.

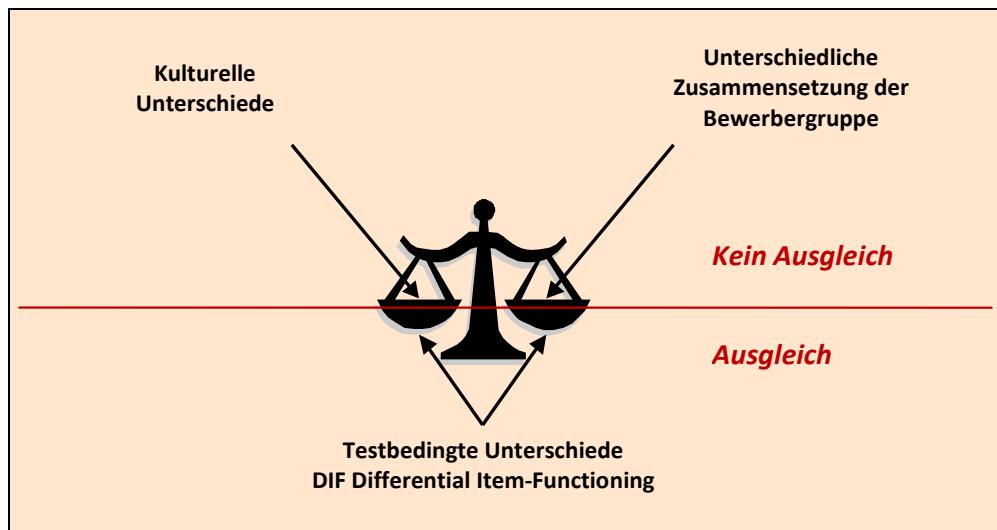


Abbildung 4: Faktoren mit Wirkung auf Unterschiede zwischen den Sprachversionen.

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 gröserer

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.

2.3.1 Sprachvergleich für die Aufgabengruppen

Auch 2025 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Aufgabengruppen auf. Die unkorrigierten Punktzahlen nach Sprachen sind nachfolgend (Abbildung 5) abgebildet.

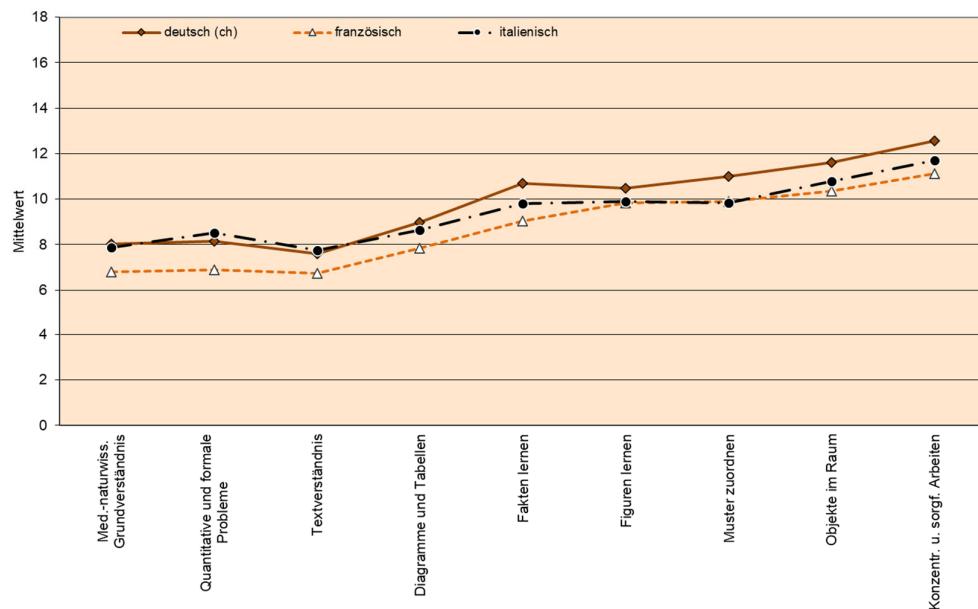


Abbildung 5: Mittelwerte der Punktwerte für die Aufgabengruppen nach Sprachen (unkorrigiert).

Abbildung 6 zeigt den Vergleich zwischen deutsch- und französischsprachigen Testteilnehmenden für das Jahr 2025, in Abbildung 7 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 (ausgenommen *Objekte im Raum*). Positive Differenzwerte deuten auf bessere Ergebnisse der deutschen Sprachgruppe hin, negative auf bessere Ergebnisse der jeweils anderen Sprachgruppe.

Für die französische Sprachgruppe treten mehrheitlich Unterschiede im Rahmen oder leicht über den vorangegangenen 10 Jahren auf. Die Differenz für *Objekte im Raum* liegt in der Größenordnung der anderen Aufgabengruppen.

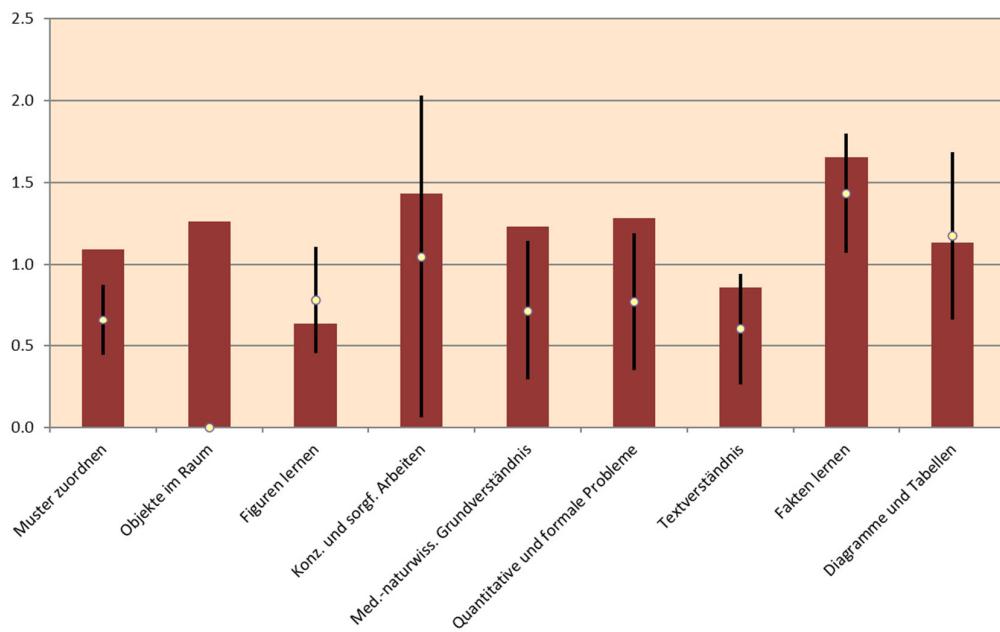


Abbildung 6: Differenz für Aufgabengruppen deutschsprachig-französischsprachig 2025 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2015-2024.

Die Unterschiede zwischen deutschsprachigen und italienischsprachigen Teilnehmenden fallen bei drei sprachunabhängigen Aufgabengruppen höher aus als üblich. Ansonsten liegen sie im üblichen Rahmen oder darunter. In zwei Aufgabengruppen erzielen italienischsprachige Teilnehmende bessere Ergebnisse als die deutsche Sprachgruppe. *Objekte im Raum* liegt im oberen Bereich der anderen Aufgabengruppen.

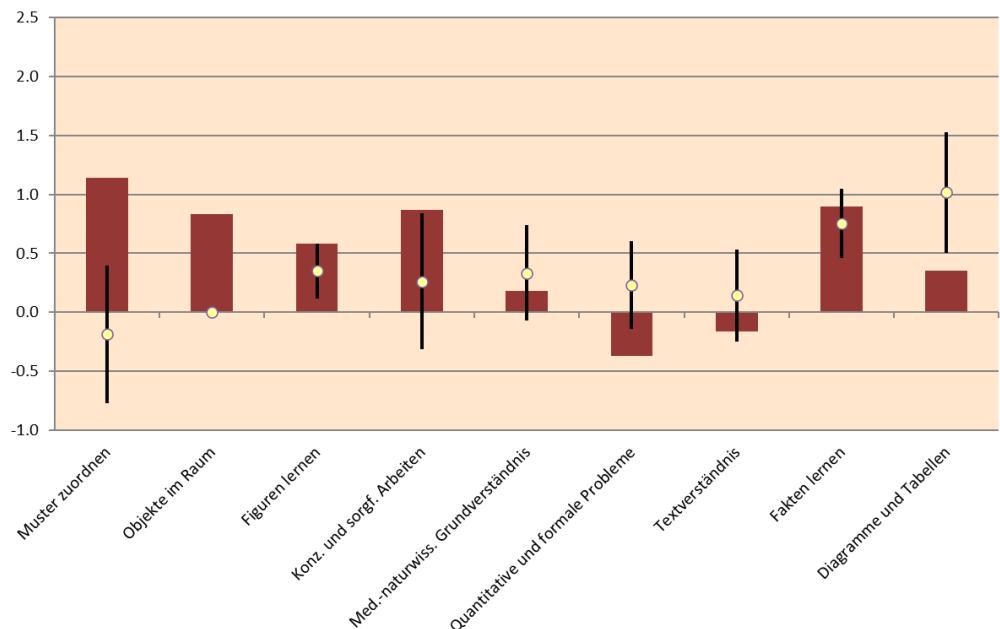


Abbildung 7: Differenz für Aufgabengruppen deutschsprachig-italienischsprachig 2025 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2015-2024.

2.3.2 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.

2.3.2.1 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 haben, erhalten diesen Bonus.

2.3.2.2 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 allen fünf sprachabhängigen Aufgabengruppen signifikant geringere Punktwerte als die deutschsprachigen Teilnehmenden. Für die italienische Sprachgruppe ist eine der fünf Aufgabengruppen betroffen.

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

Aufgabengruppe	Sprache	n	m	s	Sig.	Cohen's d	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3304	8.0	2.9	.000	0.43	Ja
	f	351	6.8	2.9			
Quantitative und form. Probleme	d	3304	8.2	3.3	.000	0.39	Ja
	f	351	6.9	3.3			
Textverständnis	d	3304	7.6	3.0	.000	0.28	Ja
	f	351	6.7	2.8			
Fakten lernen	d	3304	10.7	3.6	.000	0.47	Ja
	f	351	9.0	3.4			
Diagramme und Tabellen	d	3304	9.0	2.8	.000	0.40	Ja
	f	351	7.9	2.8			

Die Effektstärken (Cohen's d) der Differenzen sind als klein (*Textverständnis*) bis mittel (um 0.5) einzuschätzen.

Tabelle 5: Signifikanzprüfung deutsche vs. italienische Sprachform (sprachabhängige Aufgabengruppen).

Aufgabengruppe	Sprache	n	m	s	Sig.	Cohen's d	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3304	8.0	2.9	0.225	0.06	Nein
	i	156	7.9	2.8			
Quantitative und form. Probleme	d	3304	8.2	3.3	i > d	-0.11	Nein
	i	156	8.5	3.3			
Textverständnis	d	3304	7.6	3.0	i > d	-0.05	Nein
	i	156	7.8	3.1			
Fakten lernen	d	3304	10.7	3.6	.001	0.255	Ja
	i	156	9.8	3.5			
Diagramme und Tabellen	d	3304	9.0	2.8	.063	0.125	Nein
	i	156	8.6	2.9			

Die Effektstärke der Differenz bei *Fakten lernen* ist in ihrem Ausmass als klein einzuschätzen.

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.

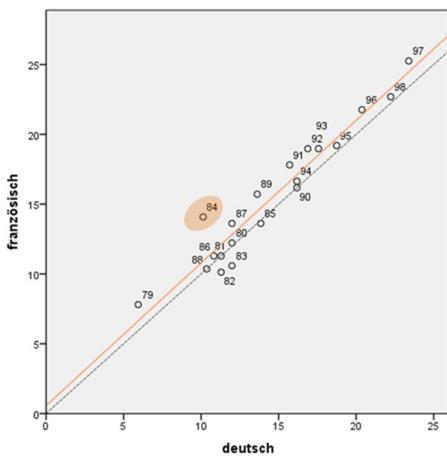
2.3.3 Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes

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.



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).

2.3.3.1 Medizinisch-naturwissenschaftliches Grundverständnis

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

Tabelle 6: DIF-Analyse Medizinisch-naturwissenschaftliches Grundverständnis

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
19	0.67	0.61	-	9.60	11.09	-		-
20	0.63	0.50	-	10.49	13.48	-		-
21	0.65	0.52	-	10.13	13.10	-		-
22	0.60	0.49	-	11.33	13.67	-		-
23	0.60	0.66	-	11.29	9.84	-		-
24	0.56	0.51	-	12.07	13.29	-		-
25	0.57	0.37	-	11.98	16.31	-	0.20	-
26	0.59	0.46	-	11.48	14.30	-		-
27	0.51	0.42	-	13.22	15.11	-		-
28	0.52	0.41	-	13.00	15.49	-		-
29	0.38	0.28	-	15.99	18.26	-		-
30	0.32	0.30	-	17.35	17.75	-		-
31	0.25	0.19	-	18.91	20.33	-		-
32	0.32	0.30	-	17.47	17.82	-		-
33	0.37	0.31	-	16.22	17.63	-		-
34	0.10	0.08	-	22.22	22.78	-		-
35	0.23	0.26	-	19.39	18.76	-		-
36	0.15	0.14	-	21.19	21.40	-		-

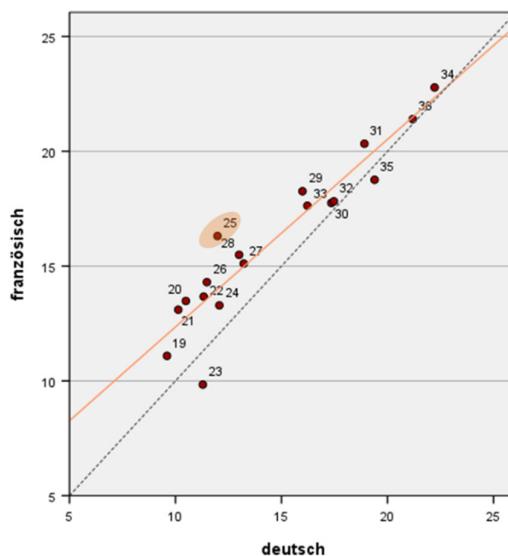


Abbildung 9: Delta-Plot für Medizinisch-naturwissenschaftliches Grundverständnis (deutsch-französisch).

2.3.3.2 Quantitative und formale Probleme

Die Aufgaben streuen hier stärker in beiden Richtungen um die Regressionsgerade. Zwei Aufgaben überschreiten die kritische Distanz in der relevanten Richtung. Für die Aufgaben 57 und 59 werden Teilnehmenden aus der französischen Sprachgruppe bei falscher Beantwortung 0.21, beziehungsweise 0.22 Punkte gutgeschrieben.

Tabelle 7: DIF-Analyse Quantitative und formale Probleme

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
55	0.59	0.54	-	11.41	12.60	-		-
56	0.72	0.72	-	8.68	8.71	-		-
57	0.61	0.39	-	11.11	15.80	-	0.21	-
58	0.58	0.48	-	11.70	13.98	-		-
59	0.56	0.34	-	12.11	16.94	-	0.22	-
60	0.58	0.56	-	11.61	12.10	-		-
61	0.50	0.34	-	13.41	16.87	-		-
62	0.49	0.33	-	13.66	17.12	-		-
63	0.55	0.51	-	12.33	13.23	-		-
64	0.41	0.36	-	15.53	16.56	-		-
65	0.40	0.35	-	15.61	16.68	-		-
66	0.36	0.34	-	16.57	16.94	-		-
67	0.34	0.26	-	17.04	18.63	-		-
68	0.38	0.32	-	16.14	17.31	-		-
69	0.31	0.21	-	17.66	19.95	-		-
70	0.30	0.32	-	17.87	17.31	-		-
71	0.32	0.34	-	17.35	16.87	-		-
72	0.17	0.15	-	20.76	21.15	-		-

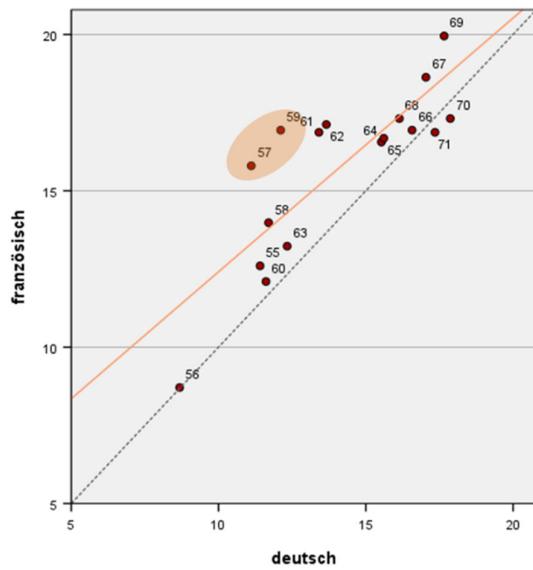


Abbildung 10: Delta-Plot für Quantitative und formale Probleme (deutsch-französisch).

2.3.3.3 Textverständnis

Hier liegen alle Aufgaben eng an der Regressionsgeraden. Trotz signifikanter Differenz hinsichtlich der Punktswerte dieser Aufgabengruppe gibt es auf Aufgabenebene keine Hinweise auf mögliche übersetzungs- oder andere sprachbedingte Einflüsse für die französische Sprachgruppe.

Tabelle 8: DIF-Analyse Textverständnis

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
73	0.65	0.57	-	10.06	11.85	-	-	-
74	0.57	0.45	-	12.01	14.49	-	-	-
75	0.49	0.42	-	13.71	15.11	-	-	-
76	0.32	0.29	-	17.43	18.07	-	-	-
77	0.37	0.27	-	16.23	18.44	-	-	-
78	0.14	0.14	-	21.35	21.46	-	-	-
79	0.68	0.60	-	9.58	11.34	-	-	-
80	0.56	0.56	-	12.03	12.22	-	-	-
81	0.49	0.47	-	13.71	14.17	-	-	-
82	0.44	0.45	-	14.73	14.61	-	-	-
83	0.42	0.40	-	15.23	15.55	-	-	-
84	0.32	0.25	-	17.47	18.88	-	-	-
85	0.51	0.48	-	13.13	13.79	-	-	-
86	0.42	0.29	-	15.21	18.07	-	-	-
87	0.45	0.37	-	14.58	16.37	-	-	-
88	0.26	0.25	-	18.81	18.95	-	-	-
89	0.34	0.33	-	16.93	17.19	-	-	-
90	0.15	0.13	-	21.06	21.52	-	-	-

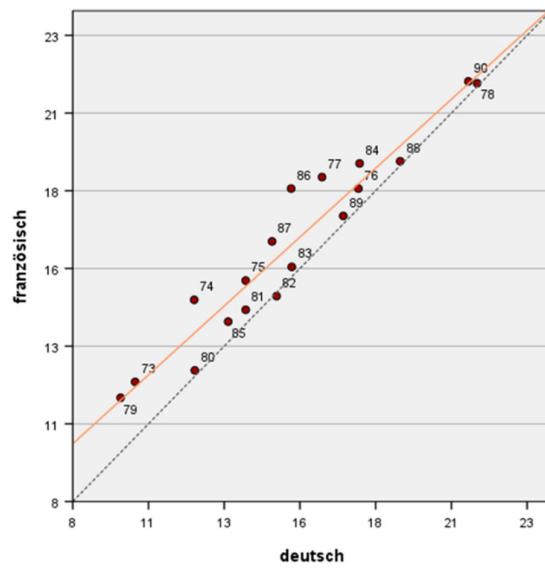


Abbildung 11: Delta-Plot für Textverständnis (deutsch-französisch).

2.3.3.4 Fakten lernen

In der Aufgabengruppe *Fakten lernen* werden der französischen Sprachgruppe für Item 111, der italienischen für die Items 111 und 124 potentiell Punkte gutgeschrieben..

Tabelle 9: DIF-Analyse Fakten lernen

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
109	0.88	0.83	0.88	5.09	6.19	4.97		
110	0.83	0.86	0.74	6.18	5.50	8.08		
111	0.78	0.55	0.62	7.32	12.41	10.76	0.23	0.16
112	0.70	0.63	0.69	9.12	10.59	9.35		
113	0.83	0.73	0.76	6.27	8.39	7.79		
114	0.52	0.44	0.49	13.07	14.67	13.73		
115	0.44	0.30	0.38	14.87	17.88	15.99		
116	0.44	0.30	0.31	14.77	17.82	17.55		
117	0.60	0.57	0.59	11.15	11.91	11.47		
118	0.53	0.40	0.49	12.77	15.55	13.73		
119	0.33	0.28	0.36	17.17	18.19	16.56		
120	0.59	0.49	0.50	11.54	13.67	13.45		
121	0.65	0.49	0.60	10.05	13.61	11.33		
122	0.40	0.32	0.35	15.55	17.44	16.84		
123	0.47	0.34	0.44	14.10	17.06	14.72		
124	0.60	0.46	0.45	11.35	14.30	14.58		0.15
125	0.40	0.37	0.40	15.65	16.24	15.71		
126	0.71	0.66	0.74	8.82	9.84	8.08		

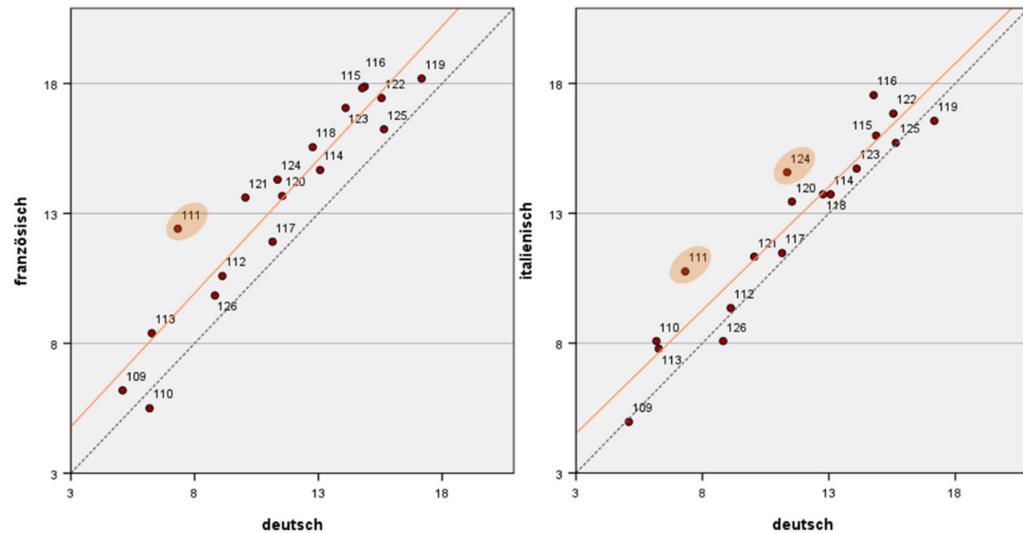


Abbildung 12: Delta-Plot für Fakten lernen (deutsch-französisch und deutsch-italienisch).

2.3.3.5 Diagramme und Tabellen

In der Aufgabengruppe *Diagramme und Tabellen* werden der französischen Sprachgruppe bei Falschantwort für die Aufgabe 134 0.24 Punkte gutgeschrieben.

Tabelle 10: DIF-Analyse Diagramme und Tabellen

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
127	0.82	0.81	-	6.36	6.63	-		-
128	0.77	0.70	-	7.49	9.08	-		-
129	0.79	0.69	-	7.08	9.21	-		-
130	0.73	0.69	-	8.27	9.27	-		-
131	0.66	0.56	-	9.99	12.10	-		-
132	0.59	0.58	-	11.48	11.66	-		-
133	0.64	0.47	-	10.47	14.11	-		-
134	0.65	0.41	-	10.09	15.49	-	0.24	-
135	0.53	0.40	-	12.73	15.74	-		-
136	0.49	0.41	-	13.64	15.49	-		-
137	0.49	0.46	-	13.59	14.36	-		-
138	0.39	0.38	-	15.91	16.18	-		-
139	0.23	0.21	-	19.46	19.83	-		-
140	0.31	0.27	-	17.73	18.44	-		-
141	0.31	0.24	-	17.63	19.26	-		-
142	0.16	0.18	-	21.01	20.58	-		-
143	0.22	0.24	-	19.55	19.13	-		-
144	0.20	0.17	-	20.08	20.77	-		-

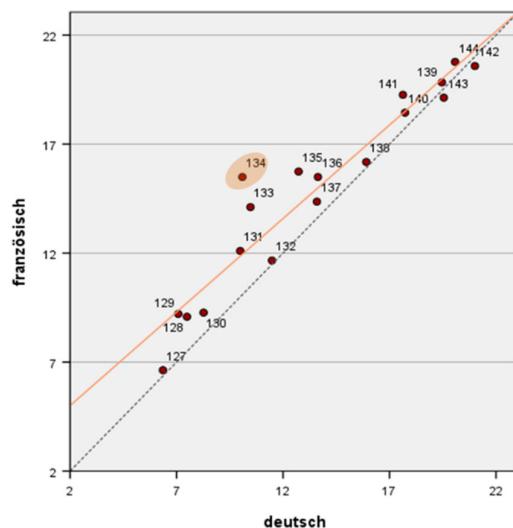


Abbildung 13: Delta-Plot für Diagramme und Tabellen (deutsch-französisch).

2.4 Effekte der Korrektur

Die durch den Ausgleich resultierenden Bonuspunkte können Tabelle 11 entnommen werden. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Teilnehmende aus der französischen 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 11: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen.

		n	Minimum	Maximum	m	s
Franz.	Medizinisch-naturwiss. Grundverständnis	351	0	0.2	0.13	0.10
	Quantitative und formale Probleme	351	0	0.43	0.27	0.15
	Fakten lernen	351	0	0.23	0.10	0.11
	Diagramme und Tabellen	351	0	0.24	0.14	0.12
	GESAMT (Gerundet)	351	0	1.00	0.66	0.47
Italienisch	Fakten lernen	156	0	0.31	0.14	0.10
	GESAMT (Gerundet)	156	0	0	-	-

Die Mittelwertsdifferenzen für jede Aufgabengruppe nach dem Sprachausgleich sind in Tabelle 12 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.

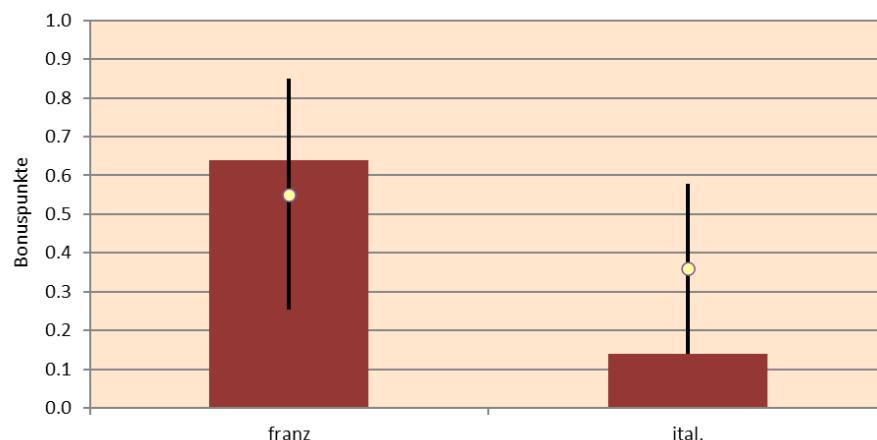


Abbildung 14: Sprachausgleich 2025 (ungerundet) mit Mittelwert und Standardabweichung der Testdurchführungen 2015-2024

Tabelle 12: Test-Prozentrang und Punktwerte inklusive Sprachausgleich.

		Mittelwert	Stand.-abw.	Minimum	Maximum
Test-Prozentrang d** > i* > f	deutsch	52.80	28.03	10	100
	französisch	38.53	26.87	10	100
	italienisch	46.01	26.51	10	99
Quantitative und form. Probleme d, i** > f	deutsch	8.16	3.27	0	18
	französisch	7.16	3.21	1	18
	italienisch	8.53	3.33	1	17
Objekte im Raum d** > f, i	deutsch	11.62	3.11	1	18
	französisch	10.36	3.48	2	18
	italienisch	10.79	3.09	2	17
Textverständnis d, i** > f	deutsch	7.59	3.03	0	18
	französisch	6.73	2.84	0	16
	italienisch	7.75	3.07	1	17
Med.-naturwiss. Grundverständnis d, i** > f	deutsch	8.04	2.88	0	17
	französisch	6.94	2.82	1	15
	italienisch	7.86	2.81	1	15
Figuren lernen d** > f	deutsch	10.48	3.16	0	18
	französisch	9.85	3.51	2	18
	italienisch	9.90	3.35	2	18
Fakten lernen d > i* > f**	deutsch	10.69	3.55	0	18
	französisch	9.14	3.32	2	18
	italienisch	9.93	3.41	1	18
Muster zuordnen d** > f, i	deutsch	10.99	3.83	1	18
	französisch	9.90	3.60	1	18
	italienisch	9.85	2.75	1	17
Diagramme und Tabellen d**, i* > f	deutsch	8.98	2.81	0	18
	französisch	8.00	2.80	2	17
	italienisch	8.63	2.89	2	15
Konz. und sorgf. Arbeiten d** > f, i	deutsch	12.56	3.13	0	18
	französisch	11.13	3.33	0	18
	italienisch	11.69	3.56	0	17

n: deutsch 3304 / französisch 351 / italienisch 156

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

Die Differenzen zwischen der deutschsprachigen Referenzgruppe und den anderssprachigen Teilnehmenden 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.

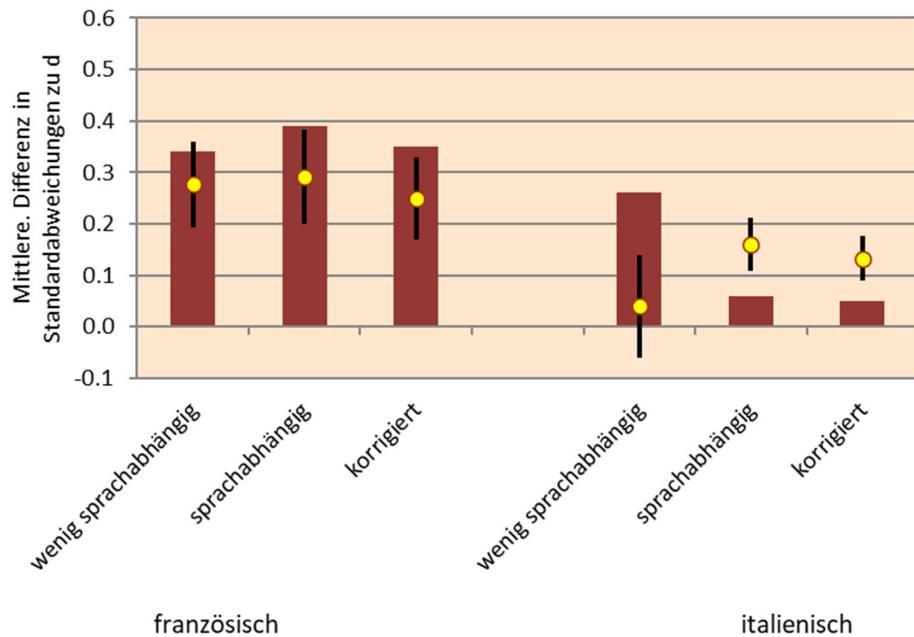


Abbildung 15: Mittlere Abweichung (Delta z) für wenig sprachabhängige und sprachabhängige Aufgabengruppen (korrigiert und unkorrigiert), 2025 und Vergleich zu 2015-2024 (Mittelwert und Streuung).

Beide Sprachgruppen weisen bei den wenig sprachabhängigen Aufgabengruppen höhere Differenzen auf als im Mittel der vergangenen 10 Jahre. Die französische Sprachgruppe weist auch in den sprachabhängigen Aufgabengruppen hohe Unterschiede auf, während diese bei der italienischen Sprachgruppe deutlich geringer ausfallen als in den vergangenen Jahren.

2.5 Vergleichbarkeit der Testlokale

2025 erfolgte die Testabnahme in 33 Testlokalen/Sektoren verteilt auf acht Testorte. Eine Besonderheit stellen wiederum die sehr kleinen Testlokale (Lokale 4, 24 und 25) dar, die für Personen mit Sonderbehandlungen bereitgestellt wurden. Diese Lokale werden aufgrund der zu geringen Fallzahlen von den weiteren Analysen ausgeschlossen.

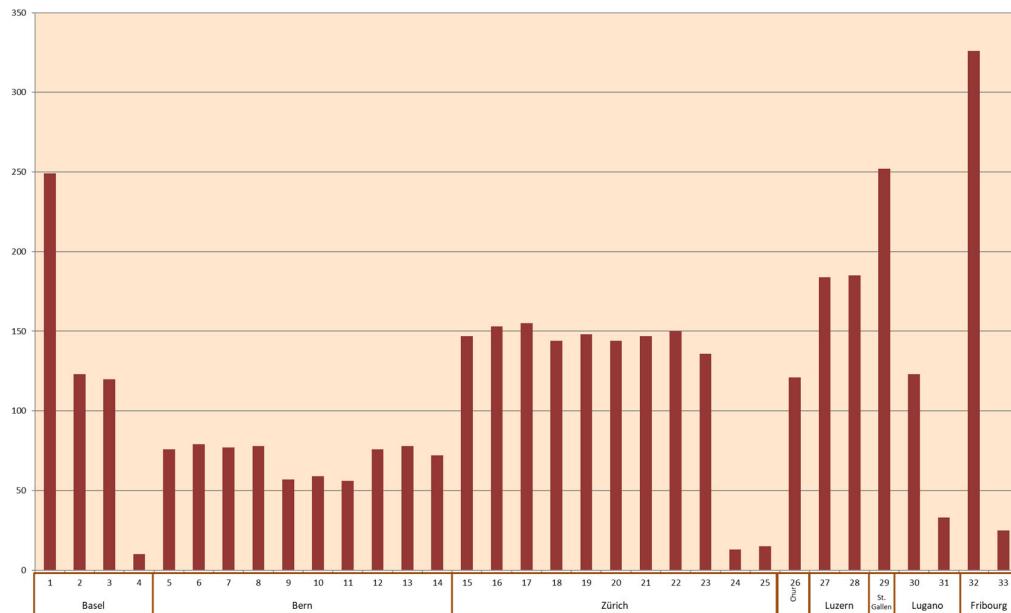


Abbildung 16: Testlokale mit Teilnehmerzahlen.

Zwischen den (deutschsprachigen) Testlokalen treten kaum signifikante Differenzen auf (Tabelle 13). Ein multipler Mittelwertsvergleich nach Tukey identifiziert lediglich Testlokal 2 mit geringeren Werten in der Aufgabengruppe *Textverständnis* (Tabelle 14). Aus den Testdurchführungsprotokollen ergeben sich keine Hinweise auf mögliche Ursachen. Es gibt keinen Anlass, Korrekturen aufgrund der Testlokalzuordnung vorzunehmen.

Tabelle 13: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den deutschsprachigen Testlokalen (ohne Lokale 4, 24 und 25).

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testprozentrang	<i>Zwischen</i>	25389.600	25	1015.584	1.294	.150
	<i>Innerhalb</i>	2543371.003	3240	784.991		
	<i>Total</i>	2568760.603	3265			
Muster zuordnen	<i>Zwischen</i>	389.774	25	15.591	1.063	.379
	<i>Innerhalb</i>	47539.736	3240	14.673		
	<i>Total</i>	47929.510	3265			
Med.-naturwiss. Grundverständnis	<i>Zwischen</i>	222.905	25	8.916	1.077	.361
	<i>Innerhalb</i>	26828.046	3240	8.280		
	<i>Total</i>	27050.951	3265			
Objekte im Raum	<i>Zwischen</i>	227.659	25	9.106	.938	.551
	<i>Innerhalb</i>	31455.376	3240	9.708		
	<i>Total</i>	31683.035	3265			
Quant. und formale Probleme	<i>Zwischen</i>	538.218	25	21.529	2.023	.002
	<i>Innerhalb</i>	34483.171	3240	10.643		
	<i>Total</i>	35021.389	3265			
Textverständnis	<i>Zwischen</i>	278.192	25	11.128	1.211	.215
	<i>Innerhalb</i>	29765.263	3240	9.187		
	<i>Total</i>	30043.455	3265			
Figuren lernen	<i>Zwischen</i>	318.867	25	12.755	1.273	.164
	<i>Innerhalb</i>	32451.067	3240	10.016		
	<i>Total</i>	32769.934	3265			
Fakten lernen	<i>Zwischen</i>	415.864	25	16.635	1.316	.135
	<i>Innerhalb</i>	40953.223	3240	12.640		
	<i>Total</i>	41369.087	3265			
Diagramme und Tabellen	<i>Zwischen</i>	340.012	25	13.600	1.738	.013
	<i>Innerhalb</i>	25346.886	3240	7.823		
	<i>Total</i>	25686.898	3265			
Konzentr. und sorgf. Arbeiten	<i>Zwischen</i>	354.846	25	14.194	1.443	.072
	<i>Innerhalb</i>	31874.832	3240	9.838		
	<i>Total</i>	32229.678	3265			

Tabelle 14: Darstellung der detaillierten Ergebnisse nach Testlokalen (ohne Lokale 4, 24 und 25)

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	51.3	11.1	8.0	11.7	7.8	7.4	10.6	10.3	8.7	12.3	249
2	48.3	11.5	7.6	11.2	7.6	6.9	10.6	10.7	8.3	11.9	123
3	51.6	11.1	8.0	11.7	7.9	7.2	10.3	10.8	8.7	11.9	120
5	48.1	10.6	7.3	11.6	7.3	7.1	10.0	11.1	8.3	12.1	76
6	54.0	11.1	8.4	11.8	8.1	8.0	10.8	11.2	8.7	12.5	79
7	52.3	11.0	7.8	11.7	7.8	7.7	11.4	10.8	8.8	12.4	77
8	50.4	10.4	7.7	11.5	8.2	7.7	10.4	9.9	8.7	12.4	78
9	52.8	10.6	8.4	11.3	8.1	8.6	10.5	10.5	9.3	12.2	57
10	57.4	11.4	8.6	12.0	7.8	8.4	11.1	10.6	9.7	13.1	59
11	47.9	11.0	7.3	11.2	7.6	7.5	10.0	10.2	8.3	12.5	56
12	52.9	10.9	7.7	11.4	7.7	7.7	10.1	11.1	8.9	12.8	76
13	49.8	10.3	7.9	11.2	7.6	7.5	10.8	10.7	8.6	12.6	78
14	53.8	11.1	7.9	11.3	7.9	7.9	10.7	11.1	9.0	12.5	72
15	55.5	11.1	8.2	12.0	8.8	7.5	10.7	10.8	9.2	12.8	147
16	56.8	11.3	8.5	11.6	8.6	7.9	10.5	11.2	9.1	13.1	153
17	55.2	11.1	8.2	12.1	8.3	7.3	10.3	11.1	9.3	12.9	155
18	56.8	11.0	8.2	11.9	8.8	7.6	10.7	11.2	9.6	12.7	144
19	53.6	10.9	8.1	11.2	8.5	7.6	10.4	10.6	9.3	12.5	148
20	55.1	11.2	8.1	11.9	8.6	7.4	10.6	10.8	9.0	12.7	144
21	55.0	10.9	8.1	11.8	8.4	7.6	11.0	10.9	8.9	12.9	147
22	55.5	11.6	8.3	11.8	8.6	7.8	10.3	10.9	9.3	12.6	150
23	54.4	11.1	8.2	11.8	8.6	7.8	10.7	10.9	8.9	13.1	136
26	55.6	11.4	8.1	11.8	8.6	7.7	10.7	10.7	9.3	12.8	121
27	49.8	10.5	8.0	11.4	7.8	7.6	10.2	10.4	8.8	12.3	184
28	49.6	11.0	7.9	11.4	7.7	7.5	10.2	10.3	8.9	12.3	185
29	49.8	10.4	7.9	11.4	8.2	7.8	10.1	10.2	9.2	12.6	252
30	46.4	9.8	7.9	10.8	8.7	7.8	9.7	9.8	8.5	12.0	123
31	44.6	10.2	7.8	10.7	8.0	7.5	10.7	9.9	9.1	10.5	33
32	38.2	9.8	6.8	10.4	6.9	6.7	9.8	9.0	7.8	11.1	326
33	42.3	10.7	7.4	9.7	6.9	6.5	10.6	9.4	8.8	11.1	25

Markiert: Mittelwerte für Aufgabengruppen im multiplen Mittelwertsvergleich (Tukey), die signifikant tiefer als in mindestens einem anderen deutschsprachigen Testlokal liegen. Italienisch- und französischsprachige Durchführungen (Lokale 30 bis 33) sind farblich abgesetzt.

2.6 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 an Bewerbenden zu Humanmedizin vorübergehend bei rund 40 % stabilisiert, ab 2014 ist er allerdings weiter gesunken. 2025 nahm er gegenüber dem Vorjahr (29.7%) leicht zu, liegt nun bei 33.3%. Über alle Disziplinen betrug der Anteil an männlichen Teilnehmenden 30.6%.

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

		Testsprache			Wunschdisziplin			Gesamt
		deutsch	französisch	italienisch	HM	VM	ZM	
männlich	m	57.1	46.1	56.0	58.2	39.4	40.1	56.3
	s	27.8	29.0	25.7	27.7	25.4	23.0	27.9
	n	1039	78	49	1050	45	71	1166
weiblich	m	50.8	36.4	41.4	50.9	40.7	43.0	49.0
	s	27.9	25.9	25.7	28.2	25.9	26.2	28.0
	n	2265	273	107	2099	360	186	2645

m: Mittelwert, s: Standardabweichung, n: Personenzahl

Für die zahlenmäßig bedeutendste deutschsprachige Teilnehmergruppe wird 2025 ein Unterschied von 6.3 Test-Prozentrangpunkten registriert. Die Differenz ist mit einer Effektstärke von 0.24 (Cohen's d) als gering zu bezeichnen. Für die gesamte Kohorte beträgt die mittlere Differenz 7.3 Prozentrangpunkte. Ein Vergleich der Disziplinen zeigt, dass die Differenz lediglich in Humanmedizin ausgeprägt ist – in Veterinär- und Zahnmedizin erzielten die Bewerberinnen jeweils sogar etwas bessere Ergebnisse als die Bewerber.

Tabelle 16: 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	58.8	48.6	57.6	41.2	32.8	54.5	39.9	52.0	32.7
	s	27.6	29.6	24.2	24.6	24.4	50.2	22.6	27.5	33.4
	n	944	62	44	30	13	2	65	3	3
weiblich	m	52.0	41.1	45.1	47.3	28.3	26.9	43.1	54.6	19.0
	s	28.2	27.5	26.1	26.0	21.1	17.2	26.4	14.2	14.7
	n	1850	162	87	237	106	17	178	5	3

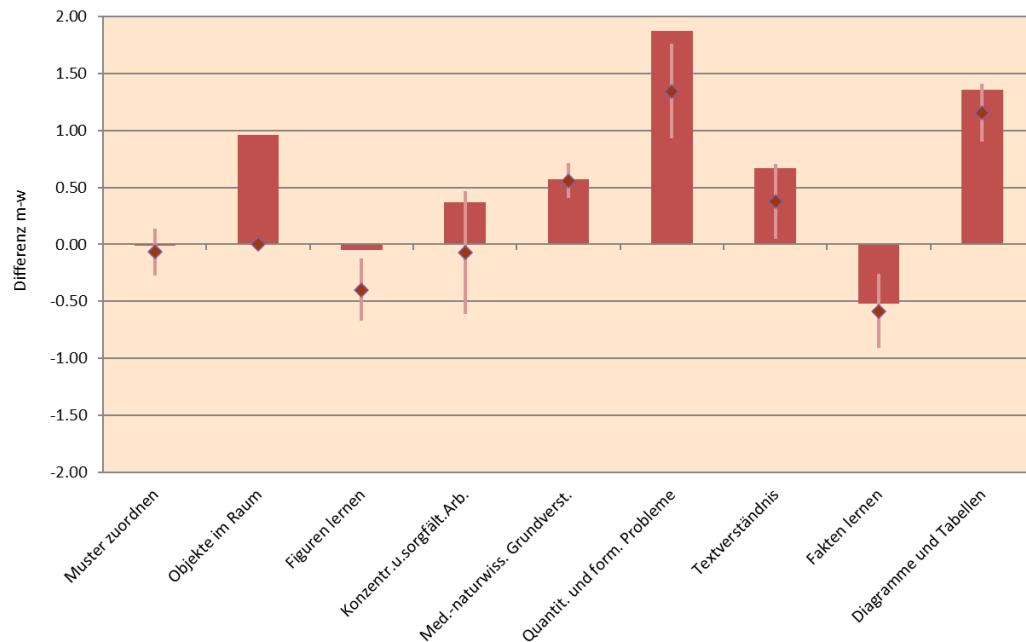
Betrachtet man die Unterschiede disziplin- und sprachspezifisch, zeigt sich für die deutschsprachige Gruppe mit Wunschdisziplin Humanmedizin eine signifikante Differenz von 6.8 Prozentrangpunkten. In der französischen Sprachgruppe beträgt die Differenz 7.5 Prozentrangpunkte zugunsten der männlichen Teilnehmenden, in der italienischen Sprachgruppe sind es für Humanmedizin 12.5 Punkte. Die deutschsprachigen Bewerberinnen

in Veterinärmedizin erzielten im Mittel höhere Ergebnisse als die entsprechenden Bewerber, gleiches gilt für die Zahnmedizin.

Tabelle 17: 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	2.13	0.14	6.07	2792	0.000
	VM	homogen	0.24	0.62	-1.21	265	0.113
	ZM	nicht homogen	4.03	0.05	-0.92	132	0.180
Französisch	HM	homogen	1.37	0.24	1.79	222	0.038
Italienisch	HM	homogen	0.47	0.49	2.66	129	0.004

Abbildung 17 verdeutlicht die Vergleichbarkeit der Geschlechterdifferenzen mit den Vorjahren (ausgenommen *Objekte im Raum*). Die diesjährigen Punktwert-Unterschiede auf Ebene der Aufgabengruppen fallen gegenüber früheren Durchführungen vergleichbar aus, sind einzig bei *quantitative und formale Probleme* etwas höher als üblich. Das bessere Abschneiden männlicher Teilnehmender in der Aufgabengruppe *Objekte im Raum* war aufgrund der zu bearbeitenden Materie zu erwarten (siehe Butler et al. (2006), Bersier et al. (2024))



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

Abbildung 17: Geschlechtsspezifische Differenzen nach Aufgabengruppen (deutschsprachig, Humanmedizin) mit Vergleich (Mittelwert und Standardabweichung) für die Vorjahre 2015-2024.

Unter den gemäss Modell zugelassenen Personen (ohne Nachrücker) die 2025 am Test teilgenommen haben (ohne Übertrag aus dem Vorjahr), beträgt der Frauenanteil in Humanmedizin 61.6%, in Veterinärmedizin 89.9%, in Zahnmedizin 74.8%, in Chiropraktik 33.3% (insgesamt 9 Personen).

2.7 Vergleiche nach Wunschhochschulen

Die Tabelle 18 führt die Test-Prozentränge getrennt nach Disziplin und gewünschter Hochschule (erste Wahl) aller deutschsprachigen Kandidaten und Kandidatinnen auf. Für Humanmedizin treten signifikante Unterschiede auf, Veterinärmedizin und Zahnmedizin unterscheiden sich nicht.

Tabelle 18: 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	544	52.9	28.2	13890.7	6	2315.1	2.9	.007
	Bern	790	53.1	28.4					
	ETH	248	60.8	28.7					
	Freiburg	43	55.7	31.5					
	Zürich	1014	54.4	27.7					
	Zürich (LU)	76	56.6	25.7					
	Zürich (SG)	75	50.8	28.6					
VM	Bern	120	45.9	25.8	109.4	1	109.4	0.2	0.687
	Zürich	147	47.2	26.0					
ZM	Basel	60	42.3	25.7	2.6	2	1.3	0.0	0.998
	Bern	79	42.1	25.1					
	Zürich	104	42.3	25.8					

Unter den zugelassenen Personen weist lediglich der Vergleich zwischen ETH und Zürich (LU) auf signifikante Unterschiede hin, alle anderen Vergleiche sind homogen.

Tabelle 19: 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	204	83.5	9.9	1647.4	6	274.6	2.9	.009
	Bern	302	83.6	9.4					
	ETH	119	86.6	10.6					
	Freiburg	20	85.0	9.1					
	Zürich	401	82.9	9.9					
	Zürich (LU)	32	80.6	8.1					
	Zürich (SG)	29	82.1	10.0					
VM	Bern	65	66.7	13.9	57.0	1	57.0	0.3	0.613
	Zürich	79	68.0	15.7					
ZM	Basel	27	66.4	14.7	72.8	2	36.4	0.1	0.863
	Bern	33	66.1	16.2					
	Zürich	44	67.9	16.0					

*Zulassung nach erstem, provisorischem Zulassungsmodell (ohne Nachrücker)

3 Ergebnisse zur Testgüte

3.1 Zuverlässigkeit

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 (2020 und 2021) umfasste der Test ab 2022 wieder mehr Aufgaben.

Der Zuverlässigkeitswert des Punktwertes erreicht 2025 mit .90 einen Wert im üblichen («vor-Corona») Bereich, das Testprofil mit 0.83 weiterhin eine sehr hohe Konsistenz. Beide Werte liegen damit mindestens im Bereich des Originaltests.

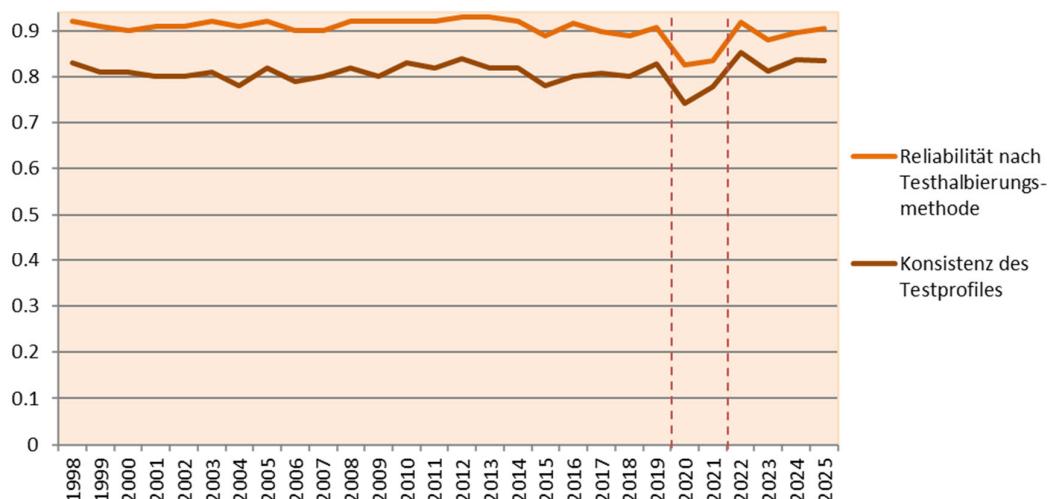


Abbildung 18: Reliabilität des Punktwertes (Split Half) und Konsistenz des Testprofils 1998-2025.

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. 2025 musste eine Aufgabe 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.

Von besonderem Interesse sind hier die Werte für die neue Aufgabengruppe *Objekte im Raum*. Es darf konstatiert werden, dass die Aufgabengruppe im Vergleich mit den anderen Aufgabengruppen den höchsten Zusammenhang mit dem Gesamtpunktwert aufweist.

Tabelle 20: Korrelationen zwischen Punktwerten der Aufgabengruppen und Gesamt-Punktwert.

	Med.-naturw. Grundv.	Objekte im Raum	Quant. u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Punktwert
Muster zuordnen	.271**	.491**	.256**	.232**	.379**	.494**	.218**	.349**	.651**
Med.-naturw. Grundverst.		.354**	.499**	.567**	.285**	.332**	.505**	.334**	.681**
Objekte im Raum			.375**	.308**	.378**	.448**	.341**	.402**	.695**
Quant. u. form. Probleme				.503**	.245**	.279**	.549**	.332**	.671**
Textverständnis					.259**	.303**	.494**	.302**	.653**
Figuren lernen						.451**	.258**	.342**	.611**
Fakten lernen							.271**	.354**	.682**
Diagramme und Tabellen								.360**	.651**
Konzent. u. sorgf. Arbeiten									.635**

Die Varianzanteile der unrotierten Lösung für die neun Aufgabengruppen entsprechen den bekannten Größenordnungen (Tabelle 21).

Tabelle 21: Faktorenanalyse: Verlauf der unrotierten Lösung (2021-2025).

Faktor	Unrotierte Lösung														
	Eigenwert					% Varianz					% Kumuliert				
	2021	2022	2023	2024	2025	2021	2022	2023	2024	2025	2021	2022	2023	2024	2025
1	2.95	4.17	3.74	3.97	3.98	49.17	46.37	41.51	44.06	44.20	49.17	46.37	41.51	44.06	44.20
2	0.97	1.14	1.28	1.23	1.31	16.09	12.68	14.23	13.62	14.57	65.26	59.06	55.74	57.68	58.77
3	0.60	0.72	0.82	0.69	0.69	9.94	8.01	9.07	7.67	7.63	75.20	67.06	64.81	65.35	66.40
4	0.57	0.61	0.63	0.60	0.65	9.55	6.74	7.05	6.69	7.22	84.75	73.81	71.86	72.03	73.62
5	0.48	0.57	0.60	0.57	0.55	7.97	6.34	6.62	6.36	6.13	92.72	80.14	78.48	78.40	79.74
6	0.44	0.53	0.52	0.52	0.49	7.28	5.89	5.75	5.78	5.48	100	86.03	84.22	84.18	85.22
7		0.44	0.50	0.50	0.47		4.91	5.53	5.50	5.19		90.94	89.75	89.68	90.41
8		0.42	0.47	0.48	0.44		4.63	5.25	5.35	4.89		95.57	95.00	95.03	95.30
9		0.40	0.45	0.45	0.42		4.43	5.00	4.97	4.70		100	100	100	100

Tabelle 22: Varianzanteile der 2-Faktorenlösung 2021-2025 (rotiert).

Varimax-rotierte Lösungen														
Eigenwert					% Varianz					% Kumuliert				
2021	2022	2023	2024	2025	2021	2022	2023	2024	2025	2021	2022	2023	2024	2025
2.4	2.7	2.7	2.7	2.7	40.7	30.0	29.5	30.2	29.8	40.7	30.0	29.5	30.2	29.8
1.5	2.6	2.4	2.5	2.6	24.6	29.1	26.2	27.5	28.9	65.3	59.1	55.7	57.7	58.7

Ein allgemeiner Fähigkeitsfaktor klärt 44 % der Varianz auf – es gibt einen grossen „gemeinsamen Nenner“ in allen Aufgabengruppen. Mit der Zwei-Faktorenlösung können fast 59 % 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. Auch 2025 ist sie eher dem Faktor „Werkzeugfunktionen“ zuzuordnen, mit immerhin noch einer Ladung von 0.36 auf dem Faktor „schlussfolgerndes Denken“. Die neue Aufgabengruppe *Objekte im Raum* ist ebenfalls primär den Werkzeugfunktionen zuzuordnen, weist aber mit einer Ladung von 0.30 ebenfalls noch einen substantiellen „analytischen“ Anteil auf.

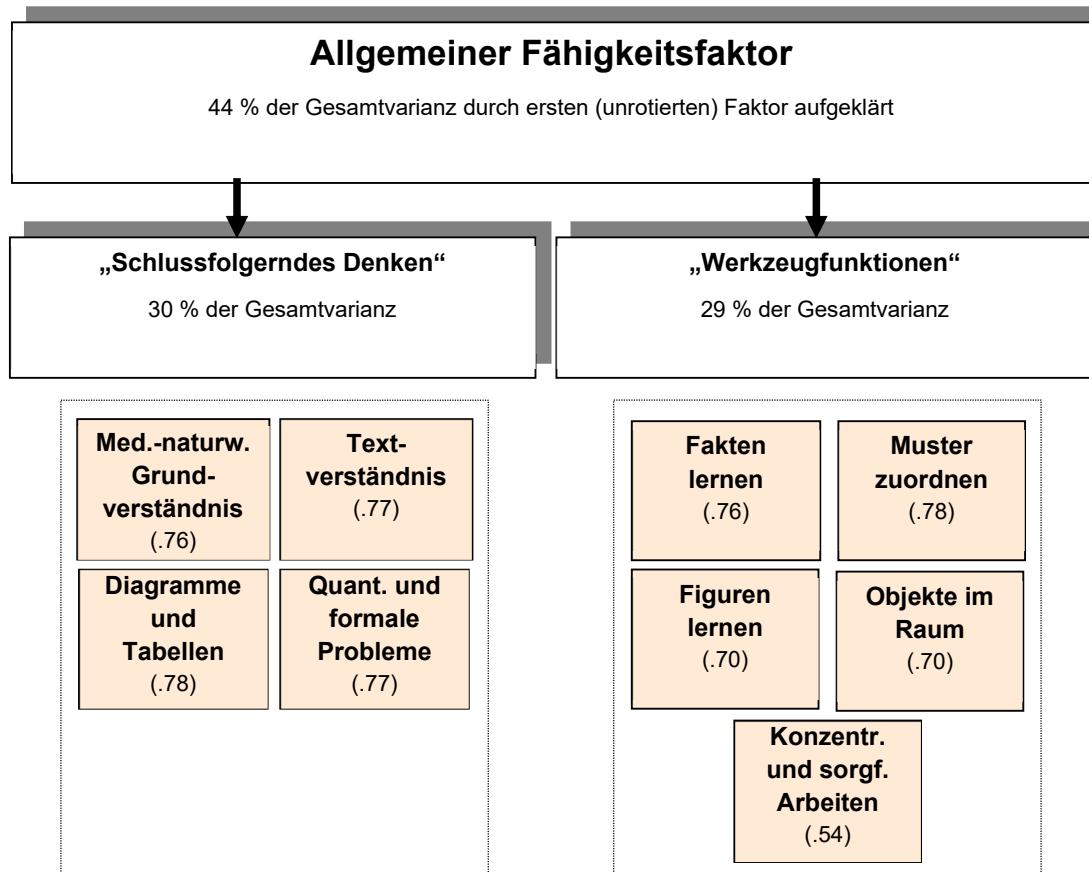


Abbildung 19: Struktur der Aufgabengruppen des EMS, hierarchische Faktorenstruktur 2025 mit Faktorladungen der Aufgabengruppen auf rotierten Faktoren.

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.
- Bersier, N.M., Arbula, AS., Ionta, S. & Rumiati, R.I. (2024). Mental rotation-related neural interactions between gender and cognitive strategy. *Imaging Neuroscience* 2024; 2 1–18. doi: https://doi.org/10.1162/imag_a_00310
- Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y. S. & Silbersweig, D. (2006). Sex differences in mental rotation: top-down versus bottom-up processing. *Neuroimage*, 32(1), 445–456. doi: <http://dx.doi.org/10.1016/j.neuroimage.2006.03.030>
- 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>