

**swissuniversities**

**swissuniversities**

Effingerstrasse 15, Postfach

3001 Bern

[www.swissuniversities.ch](http://www.swissuniversities.ch)

**Anmeldung und Zulassung  
zum Medizinstudium  
Durchführung des Selektions-  
verfahrens 2022**

**Mantelbericht swissuniversities zuhanden der  
Schweizerischen Hochschulkonferenz**

## **Inhaltsverzeichnis**

1.	Auftrag	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 2022	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

## **Anhang**

Anhang 1: Wissenschaftlicher Bericht des Zentrums für Testentwicklung und Diagnostik der Universität Freiburg über die Durchführung des EMS 2022

## **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 ZTD über die Durchführung des EMS.

Im Vergleich zu den Vorjahren wurden die Zuständigkeiten zwischen swissuniversities und Zentrum für Testentwicklung und Diagnostik der Universität Freiburg (ZTD) optimiert und geklärt. Dies spiegelt sich auch in der Struktur der Berichterstattung: 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) sind insbesondere aufgrund der hohen Kosten des Medizinstudiums und der beschränkten Anzahl praxisbezogener 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 mittels und Numerus Clausus (NC) und Eignungstest für das Medizinstudium (EMS).

### 2.1. Studienplatzkapazitäten

Im Herbstsemester 2022 gab es in der Humanmedizin gegenüber dem Vorjahr eine Erhöhung der Masterstudienplätze um 15 an der Universität Zürich, in der Zahnmedizin blieben die Kapazitäten für 2022/2023 unverändert und in der Veterinärmedizin war an der Universität Zürich auf Masterstufe eine Erhöhung um 10 Plätze zu verzeichnen.

In der Humanmedizin standen somit im 1. Jahr des Bachelorstudiums schweizweit 2'172 Studienplätze zur Verfügung, im 1. Jahr des Masterstudiums 1'460. In der Veterinärmedizin waren es 166 bzw. 136 und in der Zahnmedizin 172 bzw. 140 Studienplätze.

### 2.2. Anmeldungen per 15. Februar

Bis zur Frist vom 15. Februar 2022 gingen insgesamt 7'161 Anmeldungen für alle Hochschulen und Disziplinen ein. Es gab somit etwas weniger Voranmeldungen als im Vorjahr. 2021 waren die Anmeldezahlen höher als erwartet, 2022 lagen sie wieder im langjährigen Trend. Die Zahl der Anmeldung war in allen Disziplinen deutlich höher als die Kapazität.

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

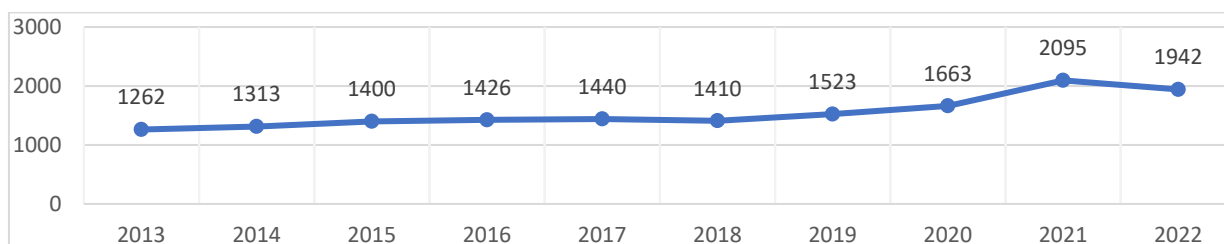


Abbildung 1: Übersicht der Anmeldungen an den Hochschulen ohne NC 2013-2022

Die Anmeldungen an die Hochschulen mit Numerus clausus (Universitäten Basel, Bern, Freiburg, Zürich, Università della Svizzera italiana und ETH Zürich) betragen 2022 für alle Disziplinen zusammengekommen 370 Prozent der vorhandenen Kapazitäten (5'219 Anmeldungen gegenüber 1'410 Studienplätzen).

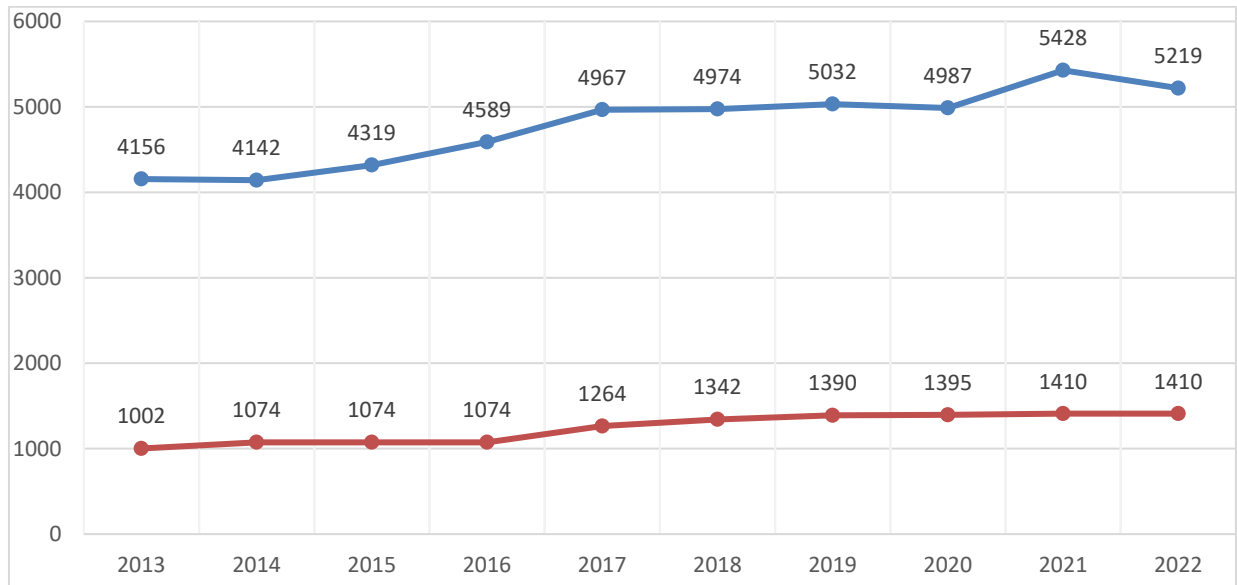


Abbildung 2: Anmeldungen (blau) und Kapazitäten (rot) an den Hochschulen mit NC 2013-2022

### 3. Testteilnahmen Eignungstest für das Medizinstudium 2022

Nach der Frist für die Anmeldung vom 15. Februar kommt es jedes Jahr zu Rückzügen. Einige der angemeldeten Personen immatrikulieren sich nicht (Hochschulen ohne NC) oder melden sich nicht zum EMS an (Hochschulen mit NC). Auch nach der Anmeldung zum EMS kommt es erfahrungsgemäss zu weiteren Rückzügen. 2022 meldeten sich bis zum 15. Februar 5'219 Personen an einer Hochschule mit NC an, 4'027 davon meldeten sich bis zur Frist Ende Mai zum EMS an, 3'840 nahmen am Test teil, 3'838 hatten schliesslich ein gültiges Testergebnis (zwei Abbrüche).

Der EMS kann auf Deutsch, Italienisch und Französisch absolviert werden. Weil die Universitäten Genf, Lausanne und Neuenburg den Zugang zum ersten Studienjahr nicht mittels NC beschränken, nehmen im Verhältnis zur Bevölkerung weniger Personen mit Muttersprache Französisch (347 Personen, 9 Prozent der Testteilnehmenden) oder Italienisch (148 Personen, 4 Prozent) teil als mit Muttersprache Deutsch (3'343 Personen, 87 Prozent). Veterinärmedizin wird in der Schweiz nur in Zürich und Bern angeboten, deshalb liegt der Anteil Frankophoner in dieser Disziplin höher (133 Personen, 30 Prozent).

#### 3.1. Testteilnahmen gesamt und nach Disziplin

2022 nahmen insgesamt (alle Disziplinen) weniger Personen am EMS teil als im Vorjahr.

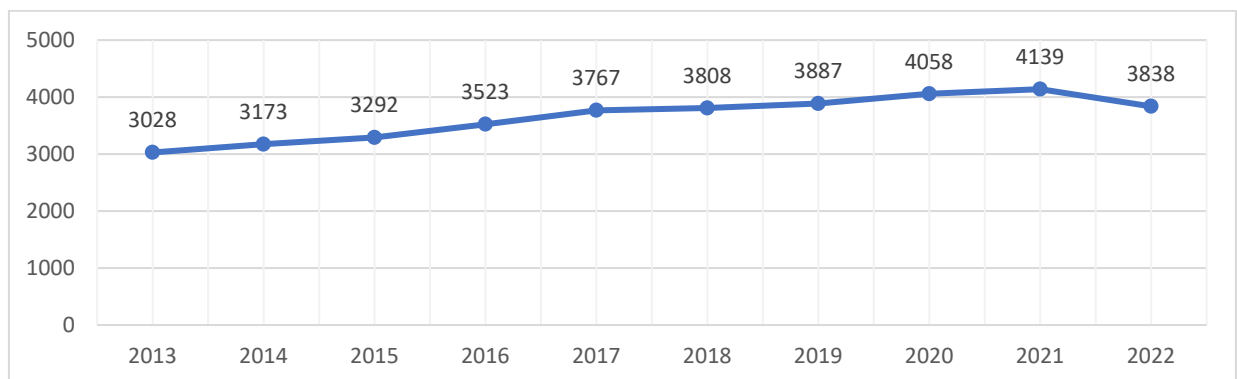


Abbildung 3: Testteilnahmen gesamt 2013-2022

In der Humanmedizin waren 2022 gegenüber 2021 erstmals deutlich weniger Testteilnahmen zu verzeichnen. In der Veterinärmedizin war die Anzahl Testteilnahmen leicht, in der Zahnmedizin deutlich tiefer.

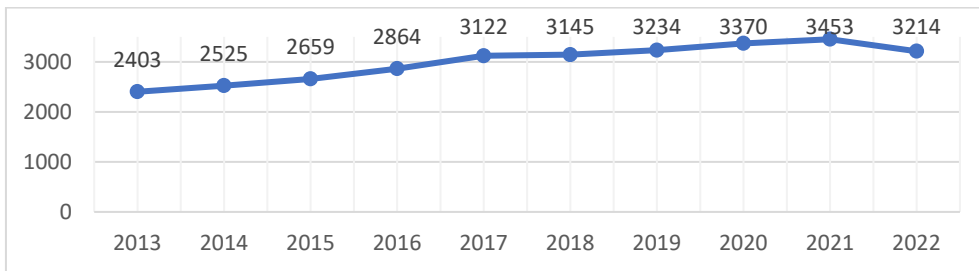


Abbildung 4: Testteilnahmen Humanmedizin inkl. Chiropraktik 2013-2022

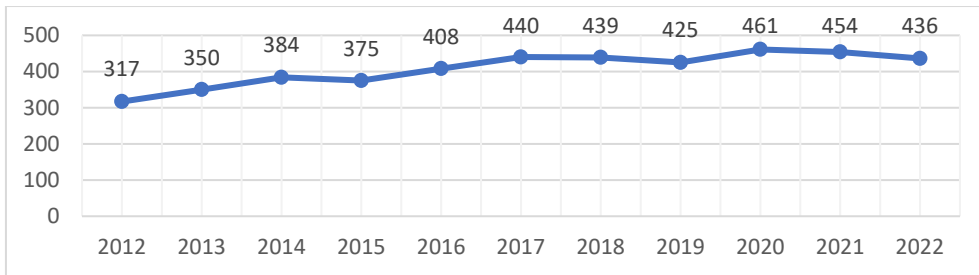


Abbildung 5: Testteilnahmen Veterinärmedizin 2013-2022

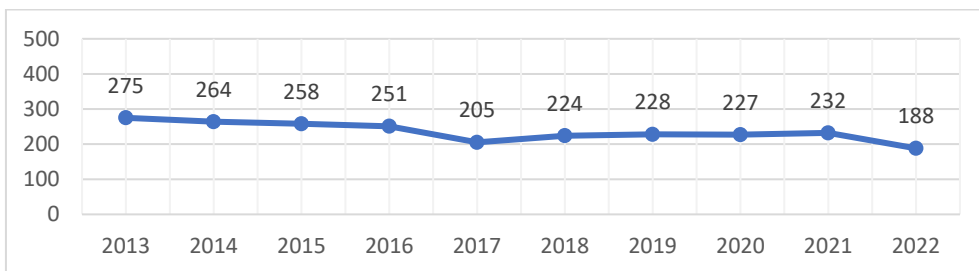


Abbildung 6: Testteilnahmen Zahnmedizin 2013-2022

### 3.2. Testteilnahmen nach Geschlecht und Disziplin

Auch 2022 nahmen mehr Frauen als Männer am EMS teil. In der Humanmedizin (inkl. Chiropraktik) waren es 2'150 Frauen und 1'064 Männer, in der Veterinärmedizin 388 Frauen und 48 Männer und in der Zahnmedizin 135 Frauen und 53 Männer.

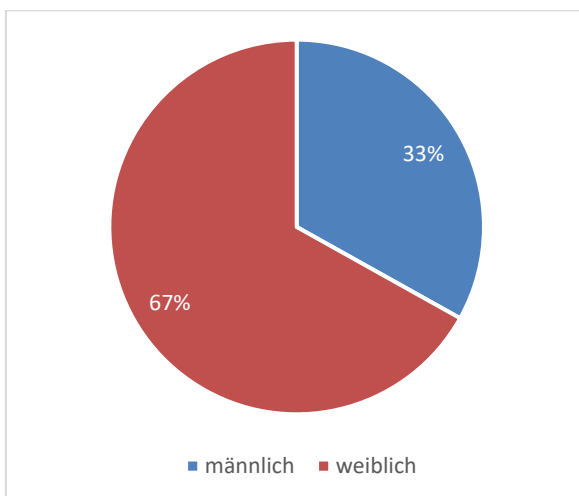


Abbildung 7: Testteilnahmen nach Geschlecht HM inkl. CP

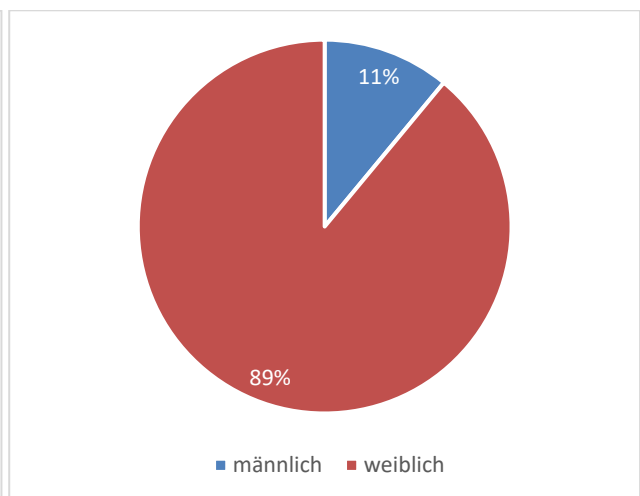


Abbildung 8: Testteilnahmen nach Geschlecht VM

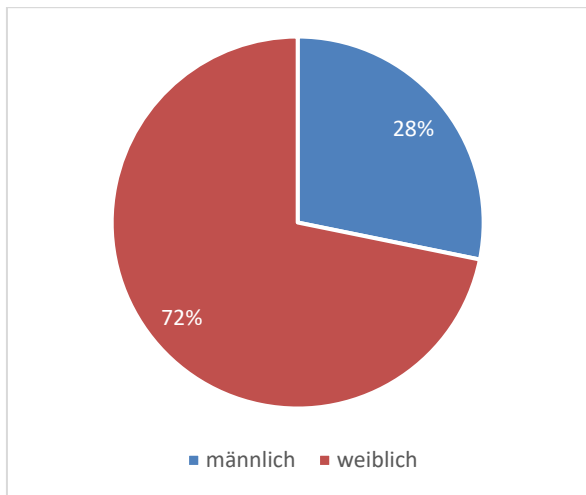


Abbildung 9: Testteilnahmen nach Geschlecht ZM

### 3.3. Testteilnahmen nach Kanton

Von den 3'838 Personen, die 2022 am EMS teilnahmen, hatten 61 ihren Wohnsitz im Ausland und 3'704 in der Schweiz. In der nachfolgenden Tabelle ist die Aufteilung auf die Kantone (Wohnsitz) dargestellt.

Tabelle 1: Testteilnahmen 2022 nach Wohnkanton

Wohnkanton	Testteilnehmende
AG	395
AI	9
AR	18
BE	505
BL	240
BS	148
FR	146
GE	31
GL	11
GR	108
JU	20
LU	245
NE	24
NW	10
OW	12
SG	237
SH	37
SO	124
SZ	60
TG	110
TI	137
UR	16
VD	88
VS	83
ZG	112
ZH	851
Ausland	61
<b>Total</b>	<b>3'838</b>

#### 4. Testdurchführung / Testergebnisse

Der EMS 2022 erhielt eine neue Struktur mit jeweils 18 Aufgaben pro Aufgabengruppe. Die neue Struktur ist länger als die Version unter Corona-Bedingungen, aber kürzer als die Originalversion des Tests. Die Kürzung erlaubt auch weiterhin eine Durchführung ohne Mittagspause.

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 2022, 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 Testergebnissen (Test-Prozentrang und Mittlerer Rangplatz) 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 2022 müssen dabei jeweils auch die Kandidatinnen und Kandidaten berücksichtigt werden, die den Testwert des Vorjahres anrechnen lassen (2022: 110 Bewerbungen).

##### 5.1. Zulassungen pro Disziplin

In der Humanmedizin (inkl. Chiropraktik) haben 63% der Studienanwärterinnen und -anwärter keinen Studienplatz erhalten (2065 Personen). In der Veterinärmedizin konnten 60% der Bewerbungen (267 Personen) nicht zugelassen werden, in der Zahnmedizin 36% (77 Personen).

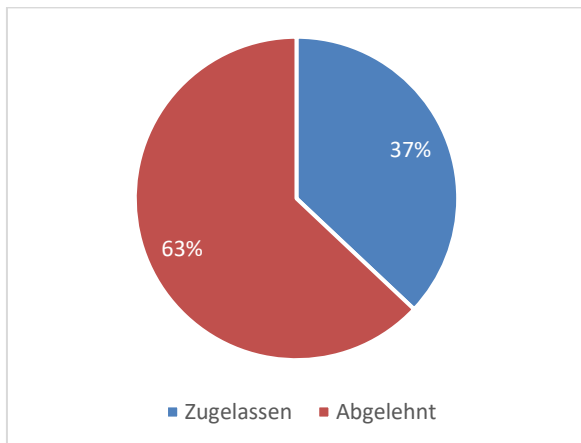


Abbildung 10: Zulassungen Humanmedizin inkl. Chiropraktik

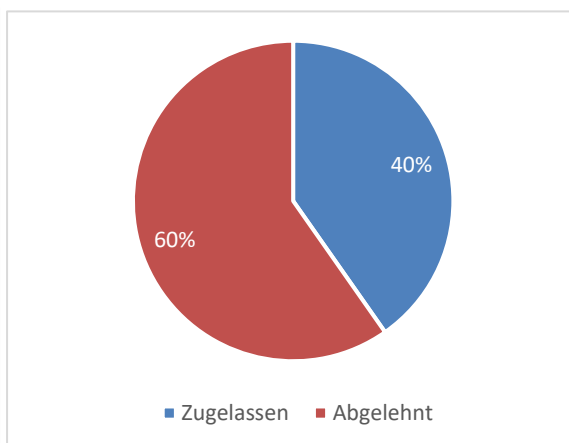


Abbildung 11: Zulassungen Veterinärmedizin

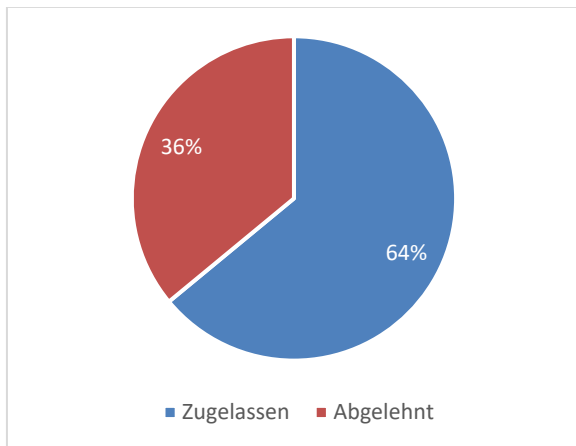


Abbildung 12: Zulassungen Zahnmedizin

## 5.2. Umleitungen

In einem zweiten Schritt – ebenfalls im Anmeldetool MEDON – erfolgt die Verteilung auf die Studienorte. 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 2022 zu vielen Umleitungen zwischen den Hochschulen. Mehrere Hochschulen sind gleichzeitig Absenderin und Empfängerin von umgeleiteten Personen. Auch in der Zahnmedizin ist der Anteil Umleitungen relativ hoch. Veterinärmedizin kann nur an zwei Studienorten studiert werden, deshalb ist der Anteil Umleitungen deutlich geringer.

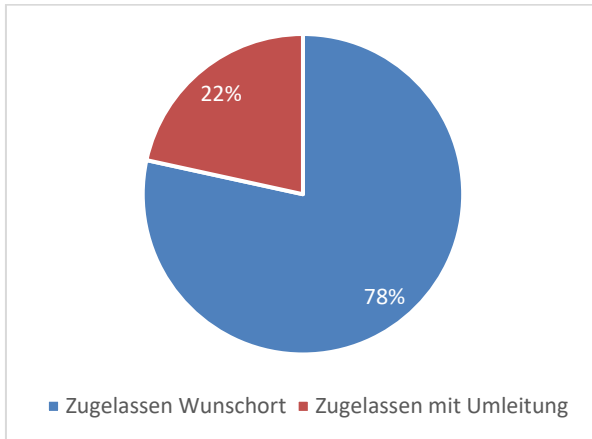


Abbildung 13: Umleitungen 2022 Humanmedizin

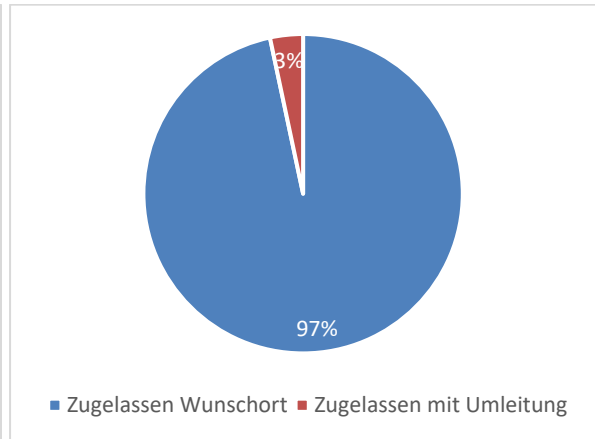


Abbildung 14: Umleitungen 2022 Veterinärmedizin

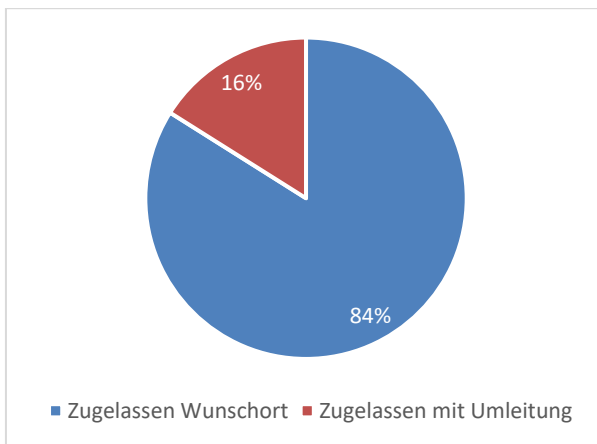


Abbildung 15: Umleitungen 2022 Zahnmedizin

Der Anteil der Bewerberinnen und Bewerber, die einen Studienplatz erhalten haben, jedoch nicht an Ihrer Wunschhochschule, liegt seit 2017 relativ stabil bei insgesamt (alle Disziplinen) zwischen 17 und 20 Prozent.

## 5.3. Verzicht auf einen Studienplatz

2022 verzichteten bis zehn Tage vor Vorlesungsbeginn 120 Bewerberinnen und Bewerber nach dem Zulassungsentcheid auf den ihnen zugeteilten Studienplatz, davon 70 in der Humanmedizin, acht in der Veterinär- und zehn in der Zahnmedizin. Der Anteil war bei den Umgeleiteten doppelt so hoch wie bei den am Wunschort Zugelassenen: Insgesamt knapp 13% der Zugelassenen mit Umleitung (37 Personen) haben auf den Studienplatz verzichtet, gegenüber insgesamt rund 7% der Zugelassenen am Wunschort (83 Personen).



2022 / BERICHT 29

EIGNUNGSTEST FÜR DAS MEDIZINSTUDIUM (EMS)

# BERICHT ÜBER DURCHFÜHRUNG UND ERGEBNISSE

Dr. B. Spicher

Dr. Spicher, Benjamin (2022)

EMS Eignungstest für das Medizinstudium 2022  
Berichte des Zentrums für Testentwicklung, Nr. 29,  
im Auftrag der Schweizerischen Hochschulkonferenz  
(SHK)

Unter Mitarbeit von M. Bernasconi, T. Cruchoad,  
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:  
M. Klaus, K. Balmer

Redaktion: T. Cruchoad





## 1 Inhalt

2	Vorwort .....	5
3	Zusammenfassung .....	6
4	Résumé.....	7
5	Der EMS mit neuer Struktur .....	8
5.1	Vorgaben .....	8
5.2	Umsetzung.....	9
5.2.1	<i>Sonderfall Textverständnis .....</i>	9
5.2.2	<i>Sonderfall Konzentriertes und sorgfältiges Arbeiten .....</i>	9
5.2.3	<i>Sonderfall Figuren lernen / Fakten lernen.....</i>	9
5.3	Erwartete Auswirkungen der neuen Teststruktur .....	9
6	Beschreibung des verwendeten Eignungstests .....	11
6.1	Berechnung der Werte 2022 .....	13
6.2	Mittlerer Rangplatz der Aufgabengruppen .....	14
7	Testanwendung 2022 .....	16
7.1	Verteilungsprüfung.....	16
7.2	Äquivalenz der Sprachversionen.....	17
7.2.1	<i>Sprachvergleich für die Aufgabengruppen.....</i>	19
7.2.2	<i>Darstellung des Korrekturverfahrens .....</i>	23
7.2.3	<i>Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes .....</i>	24
7.2.4	<i>Effekte der Korrektur.....</i>	29
7.3	Vergleichbarkeit der Testlokale .....	32
7.3.1	<i>Versetzter Testbeginn in Zürich.....</i>	34
7.4	Vergleich für die Geschlechter .....	34
7.5	Vergleiche nach Wunschhochschulen .....	36
8	Ergebnisse zur Testgüte .....	37
8.1	Zuverlässigkeit.....	37
8.2	Binnenstruktur .....	38
9	Literatur.....	40
9.1	Originaltest zur Information und Vorbereitung .....	41
9.2	Frühere Berichte des ZTD .....	41

## 2 Vorwort

Nach zwei Jahren in denen der Test, bedingt durch die Corona-Pandemie, unter aussergewöhnlichen Umständen durchgeführt werden musste, war das Jahr 2022 gekennzeichnet durch die Rückkehr zur Normalität – nicht nur bezüglich des Eignungstests. Während einerseits Beschränkungen fielen, sollten andererseits positive Erkenntnisse und Erfahrungen aus den Pandemie Jahren genutzt werden, um wo immer möglich Optimierungen anzubringen und Abläufe zu verschlanken.

Eine Durchführung des Eignungstests ohne Mittagspause stand aufgrund der zunehmenden Teilnehmendenzahlen und des dadurch immer gedrängteren Zeitablaufs bereits vor Corona zur Diskussion. Die Durchführungen 2020 und 2021 haben gezeigt, dass eine verkürzte Test-Durchführung möglich ist und zu annähernd gleich guten diagnostischen Ergebnissen führt, wie der Originaltest. Für das Jahr 2022 konnte die Teststruktur weiter optimiert werden: Die neue Struktur umfasst nun wieder alle Aufgabengruppen, mit jeweils 18 statt 20 Aufgaben. Die Ergebnisse dazu werden in diesem Bericht ausführlich dargestellt.

Das ZTD dankt swissuniversities sowie allen an der Testdurchführung beteiligten Personen, wie Koordinatorinnen und Koordinatoren, externen Übersetzerinnen und Übersetzern, Testleitungen, Aufsichtsteams, Sicherheitspersonal und Transportpersonal, für die wiederum reibungslose, zuverlässige und angenehme Zusammenarbeit.

### 3 Zusammenfassung

Die in diesem Bericht vorgestellten Ergebnisse betreffen den Eignungstest 2022 für das Medizinstudium in der Schweiz (EMS) und die Zulassung zum Studium der Medizin ab Herbstsemester 2022. Für bestimmte Fragestellungen werden Vergleichsdaten früherer Testanwendungen herangezogen.

1. Ein Numerus clausus (NC) ist 2022 für Humanmedizin (seit 1998), Veterinärmedizin (seit 1999) und Zahnmedizin (seit 2004) notwendig für Personen, die sich in Basel, Bern, Freiburg oder Zürich angemeldet haben – in Humanmedizin seit 2017 auch für den Studiengang an der ETH Zürich, für die Luzerner und St. Galler Tracks der Universität Zürich sowie für den Studiengang an der USI im Tessin, der in Basel begonnen wird. An der Universität Zürich gilt der NC weiter auch für Chiropraktik (seit 2008).
2. Die Nachfrage nach Studienplätzen für **Humanmedizin** nahm, gemessen an den Anmeldungen im Februar, nach einem deutlichen Anstieg im Vorjahr erstmals ab. Mit Ausnahme dieses Ausreissers ist der Trend jedoch weiter zunehmend. Die Studienkapazitäten werden 2022 in Humanmedizin zu **389%** (Vorjahr 403%) ausgelastet, in **Veterinärmedizin** zu **340%** (Vorjahr 357%), in **Zahnmedizin** zu **234%** (Vorjahr 261%).
3. **4027 Personen** (Vorjahr 4308) haben sich für die betreffenden Hochschulen bis zur gesetzten Frist im **Mai 2022** zum EMS 2022 angemeldet. Davon bewarben sich **9 Personen** um einen Disziplinenwechsel. Weitere **111 Personen** bewarben sich mit ihrem Testergebnis aus dem Vorjahr und nahmen nicht am diesjährigen Test teil.
4. **3838 Personen** haben den EMS mit gültigem Ergebnis beendet (Vorjahr 4139). 2 Personen haben den Test vorzeitig abgebrochen (Vorjahr 3). 183 Personen sind nicht zum Test erschienen oder haben ihre Testanmeldung zurückgezogen (Vorjahr 161). Die Rückzugsquoten zwischen Februar und Test betragen für Humanmedizin **24.8%**, für Veterinärmedizin **20.7%** und für Zahnmedizin **24.8%**.
5. Der Test fand am 8. Juli 2022 gleichzeitig an **8 Testorten** und in 32 Testlokalen in drei Sprachen (Deutsch: Basel, Bern, Chur, Luzern, St. Gallen, Zürich; Französisch: Freiburg; Italienisch: Lugano) statt. Die Wahl des Testortes war unabhängig vom zukünftigen Studienort möglich.
6. Der EMS 2022 erhielt eine neue Struktur mit jeweils 18 Aufgaben pro Aufgabengruppe. Die neue Struktur ist länger als die Version unter Corona-bedingungen, aber kürzer als die Originalversion des Tests. Die Kürzung erlaubt auch weiterhin eine Durchführung ohne Mittagspause.

## 4 Résumé

Le présent rapport informe sur les résultats concernant le test d'aptitudes aux études de médecine en Suisse (AMS) et les admissions aux études en question au semestre d'automne 2022. Pour certaines problématiques, des données comparatives des éditions antérieures du test (1998-2021) y figurent également.

1. En 2022, un numerus clausus (NC) a été nécessaire en médecine humaine (depuis 1998), en médecine vétérinaire (depuis 1999) et en médecine dentaire (depuis 2004) pour les personnes qui se sont inscrites à Bâle, Berne, Fribourg ou Zurich. En médecine humaine, cela concerne également les nouvelles filières avec NC à l'EPF Zurich, aux Universités de Lucerne et de St-Gall en coopération avec l'Université de Zurich ainsi qu'à l'Università della Svizzera italiana dont le cursus commence à Bâle. À l'Université de Zurich, le NC se poursuit pour la chiropractie (depuis 2008).
2. Après une augmentation significative l'année passée, la demande de places d'études en **médecine humaine**, mesurée sur la base du nombre d'inscriptions reçues en février, a diminué pour la première fois. Mise à part cette exception, la tendance demeure néanmoins toujours à la hausse. En 2022, les capacités d'études en médecine humaine sont dépassées à hauteur de **389 %** (403 % en 2021), en **médecine vétérinaire** à hauteur de **340 %** (357 % en 2021) et en **médecine dentaire** à hauteur de **234 %** (261 % en 2021).
3. **4027 personnes** (4308 en 2021) se sont annoncées pour ces hautes écoles auprès de swiss-universities jusqu'au délai imparti de **mai 2022**, en vue de la passation du test AMS 2022. Parmi elles, **9 personnes** ont sollicité un changement de discipline. Par ailleurs, **111 personnes** ont fait valoir leur résultat de l'année précédente et n'ont ainsi pas pris part au test de cette année.
4. **3838 personnes** ont terminé l'AMS avec des résultats valables (4139 en 2021). 2 personnes ont interrompu prématurément le test (3 en 2021). 183 personnes ont retiré leur inscription au test ou ne se sont pas présentées à celui-ci (161 en 2021). Les taux de retrait entre février et la date de passation du test se sont élevés à **24,8 %** pour la médecine humaine, **20,7 %** pour la médecine vétérinaire et **24,8 %** pour la médecine dentaire.
5. Le test s'est déroulé le 8 juillet 2022 **sur 8 sites de test** simultanément comprenant 32 locaux de test au total et ce, dans trois langues (allemand : Bâle, Berne, Coire, Lucerne, St-Gall, Zurich / français : Fribourg / italien : Lugano). Le choix du lieu du test est indépendant du futur lieu d'études.
6. L'AMS 2022 s'est doté d'une nouvelle structure, avec 18 exercices par groupe d'exercices. La nouvelle structure est plus longue que la version durant le COVID, mais plus courte que la version originale du test. La réduction opérée permet un déroulement du test sans pause de midi.

## 5 Der EMS mit neuer Struktur

Am EMS wurden seit seiner Einführung 1998 diverse Anpassungen und Modernisierungen vorgenommen. Manche betrafen die Struktur des Tests selber (Einführung und später wiederum Verzicht auf die Aufgabengruppe „Planen und Organisieren“), andere betrafen die „Feinjustierung“ von Aufgabenschwierigkeiten um weiterhin eine optimale Differenzierungsfähigkeit zu gewährleisten (Konzentriertes und sorgfältiges Arbeiten, Figuren lernen, Fakten lernen, Muster zuordnen). Allen Versionen gemeinsam war eine hohe Zuverlässigkeit (Reliabilität) der Testergebnisse (um 0.9) – Grundlage für eine faire Selektion der zukünftigen MedizinstudentInnen. Die Mischung der Aufgabengruppen bot ausserdem, innerhalb der Möglichkeiten einer standardisierten und objektiven Beurteilung, einen Ausgleich zwischen verschiedenen Stärken und Schwächen der Teilnehmenden. Eine über die Jahre stabile Faktorenstruktur belegt die Vergleichbarkeit der Testergebnisse auch über die Jahrgänge hinweg.

In den Corona-Jahren 2020 und 2021 wurde der EMS, im Interesse der Teilnehmenden und des Testpersonals, verkürzt. Ziel war dabei in erster Linie, eine Durchführung ohne Mittagspause zu ermöglichen, da beim Verlassen der und Wiedereinlass in die Testlokale die Abstandsregeln nicht einzuhalten gewesen wären. Da es sich 2020 um eine Notmassnahme gehandelt hatte, wurden primär jene Aufgabengruppen eliminiert, die ein zusätzliches Austeilen von Materialien erforderlich gemacht und damit dem Grundsatz der Minimierung von Kontakten widersprochen hätten.

Die erste Durchführung unter diesen Bedingungen im Jahr 2020 durfte, trotz der knappen Vorbereitungszeit für die Materialerstellung, als Erfolg gewertet werden. Vor diesem Hintergrund, und um keine Risiken einzugehen, wurde 2021 das verkürzte Testkonzept noch einmal übernommen – im Bewusstsein, dass dies weiterhin eine Notmassnahme und keine dauerhafte Lösung sein sollte.

Bereits vor den Corona-Jahren war, insbesondere in den grossen Testorten Zürich und Bern, eine mögliche Verkürzung des EMS thematisiert worden. Die Herausforderungen an die Logistik sind dort auch aufgrund der Mittagspause besonders hoch – eine Einhaltung der vorgegebenen Zeiten aufgrund der Menge der KandidatInnen, die innerhalb von 60 Minuten das Lokal verlassen und wieder betreten mussten (inklusive erneuter Zugangskontrollen) nur schwer realisierbar.

Die neue Struktur soll die Durchführung des Tests ab 2022 ohne Mittagspause erlauben. Dabei soll die diagnostische Qualität möglichst optimal erhalten bleiben und gegenüber der Originalversion lediglich minime Einbussen (die bei jeder Verkürzung zu erwarten sind) zur Folge haben.

### 5.1 Vorgaben

Für eine Testung ohne Pause ist eine Obergrenze für die Testdauer von 4 Stunden (netto) anzunehmen. Dies entspricht beispielsweise der Dauer einer schriftlichen Maturitätsprüfung für das Fach Mathematik im Kanton Bern – die Belastung ist den KandidatInnen also bekannt und scheint, angesichts der Wichtigkeit der Entscheidung für die individuelle Zukunft, zumutbar.

Die Zusammensetzung des EMS in seiner originalen Form sollte möglichst erhalten bleiben, um weiterhin einen optimalen Ausgleich zwischen unterschiedlichen Stärken- und Schwächenprofilen zu ermöglichen.



## 5.2 Umsetzung

Der EMS 2022 wurde gegenüber der ursprünglichen Form von 255 auf 235 Minuten netto Durchführungsdauer verkürzt (gegenüber der Corona-Version um 14 Minuten verlängert), liegt damit also knapp unter der kritischen Grenze von vier Stunden. Zusätzliche Zeit ist für Vorbereitung (Instruktionen) und Nachbereitung (Nachzählen von Materialien vor der Entlassung der KandidatInnen) einzurechnen. Erreicht wurde die Einsparung von 20 Minuten Testzeit durch eine Reduktion der (meisten) Aufgabengruppen um jeweils zwei Aufgaben, so dass die maximale Punktzahl pro Aufgabengruppe neu 18 statt 20 Punkte beträgt.

### 5.2.1 Sonderfall *Textverständnis*

Die Aufgabengruppe „Textverständnis“ bestand bereits vorher aus 18 Aufgaben. Dies, weil dort drei längere, komplexe Texte verstanden und zu jedem Text sechs Fragen beantwortet werden müssen. Eine Reduktion auf 5 Fragen pro Text schien möglich, diese stärkere Reduktion (um 3 statt 2 Aufgaben) stand jedoch der hohen Prognosekraft dieser Aufgabengruppe für die Studieneignung gegenüber. Die Aufgabengruppe besteht weiterhin aus 18 Aufgaben – ihr Gewicht wird also am Gesamtergebnis geringfügig erhöht.

### 5.2.2 Sonderfall *Konzentriertes und sorgfältiges Arbeiten*

Um eine zu hohe Gewichtung dieser Aufgabengruppe zu vermeiden, erfolgt die Transformation des Rohwerts hier neu in einen Wert von 0 bis 18 Punkten. Gleichzeitig wurde die Verteilung gegenüber den Vorjahren angepasst: Statt 2.5% erhalten noch 1.5% jeweils den höchsten und den tiefsten Wert, wobei negative Rohwerte in jedem Fall den Wert 0 erhalten.

### 5.2.3 Sonderfall *Figuren lernen / Fakten lernen*

Die Lernphase dieser beiden Aufgabengruppen fand ursprünglich direkt nach der Mittagspause statt. Dazu wurde zunächst ein spezielles Lernheft ausgeteilt, nach Abschluss der Einprägephase wurde dieses wieder eingesammelt und das Testheft B für den Nachmittagsteil ausgeteilt. Dadurch wurde sichergestellt, dass das Lernmaterial bei der nach der Aufgabengruppe „Textverständnis“ folgenden Reproduktionsphase für die Teilnehmenden nicht mehr verfügbar war.

Dieser Ablauf sollte auch in der neuen Version erhalten bleiben, allerdings war kein separates Lernheft mehr nötig. Stattdessen wurde das Lernmaterial in das Testheft A integriert, dieses nach der Lernphase eingesammelt (zusammen mit dem Notizpapier) und Testheft B ausgeteilt. Das bisherige Zeitintervall zwischen Einpräge- und Abrufphase bleibt durch die in ihrer Originallänge dargebotene Aufgabengruppe „Textverständnis“ erhalten. Allerdings wurde die Aufgabe „Konzentriertes und sorgfältiges Arbeiten“ neu am Ende des Tests (statt am Ende des Vormittagsteils) implementiert.

## 5.3 Erwartete Auswirkungen der neuen Teststruktur

Im Vorfeld des EMS 2022 wurden Modellrechnungen zu den Auswirkungen der Testverkürzung angestellt. Berechnet für die Jahre 2016 bis 2019 (letzte Durchführung der Vollversion des EMS vor den Corona-Massnahmen) korrelierten die für die Zulassung relevanten Prozentränge der Vollversion und der modellhaft verkürzten Version mit  $r = .995$ . Veränderungen in der Rangliste waren also punktuell, die verkürzte Version gelangte zu den praktisch gleichen Ergebnissen, wie die jeweiligen Vollversionen.

Hinsichtlich der Testkennwerte (Tabelle 1) ergeben sich wie erwartet gegenüber der Langversion geringe Einbussen, auch diese waren jedoch minimal. Insbesondere waren die beobachteten Differenzen zwischen den Jahren grösser, als die durch die Kürzung bedingten Reduktionen der Kennwerte.

**Tabelle 1: Modellrechnung: Testkennwerte für die lange (EMS) und gekürzte (EMS ab 2022) Version**

<b>Kennwerte</b>	<b>2019</b>	<b>2018</b>	<b>2017</b>	<b>2016</b>
<b>EMS</b>				
Split-half	0.906	0.887	0.899	0.917
Profil	0.826	0.795	0.807	0.761
<b>EMS ab 2022 (Modell)</b>				
Split-half	0.894	0.871	0.89	0.91
Profil	0.811	0.775	0.791	0.735
<b>Differenz</b>				
Split-half	-0.012	-0.016	-0.009	-0.007
Profil	-0.015	-0.020	-0.016	-0.026

Modell: Jeweils die Aufgaben 10 und 11 (mittlere Schwierigkeit) pro Aufgabengruppe wurden aus der Wertung entfernt (ausser bei Textverständnis und konzentriertes und sorgfältiges Arbeiten)

Die tatsächlichen Kennwerte des EMS 2022 werden in einem späteren Kapitel dargestellt.

## 6 Beschreibung des verwendeten Eignungstests

Der EMS besteht ab 2022 wiederum aus neun Aufgabengruppen, die zu einem Gesamtwert verrechnet werden.

Die nachfolgend dargestellte Verkürzung hat es erlaubt, den Test ohne Mittagspause durchzuführen, sowie die Arbeitsschritte des Austeilens und Einsammelns von Materialien zu verringern. Tabelle 2 gibt einen Überblick über den originalen und den verkürzten Testaufbau und die Dauer der einzelnen Aufgabengruppen.

**Tabelle 2: Struktur und Ablauf des EMS „vor Corona“ und ab 2022**

EMS original				EMS ab 2022			
Aufgabengruppe	Aufgaben	Zeit in Min	Zeit pro Aufgabe (min)	Aufgabengruppe	Aufgaben	Zeit in Min	Zeit pro Aufgabe (min)
Muster zuordnen	20	18	0,9	Muster zuordnen	18	16	0,9
MNGV	20	50	2,5	MNGV	18	45	2,5
Schlauch	20	12	0,6	Schlauch	18	10	0,6
QfP	20	50	2,5	QfP	18	45	2,5
Konz	20	8		Figuren einprägen		4	
Mittagspause				Fakten einprägen			
Figuren einprägen		4		Testheft A einsammeln, Testheft B austeilen			
Fakten einprägen		6		TV	18	45	2,5
TV	18	45	2,5	Figuren reproduzieren	18	5	0,3
Figuren reproduzieren	20	5	0,25	Fakten reproduzieren	18	6	0,3
Fakten reproduzieren	20	7	0,35	DuT	18	45	2,5
DuT	20	50	2,5	Konz	18	8	
Total	178	255		Total	162	235	

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 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.

In jeder Aufgabengruppe können neu 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. Jedes Jahr werden neue Aufgaben für die Aufgabengruppen entwickelt und in mehreren Schritten überarbeitet. Die Aufgaben müssen sehr hohe Qualitätsstandards erfüllen, u.a. (1) müssen sie jedes Jahr die Studieneignung gleich zuverlässig messen, (2) muss das Schwierigkeitsspektrum aller Aufgaben annähernd vergleichbar sein, (3) darf kein spezielles Fachwissen vorausgesetzt werden, um die Trainierbarkeit des Tests gering zu halten und (4) muss eine eindeutige, richtige Lösung existieren.

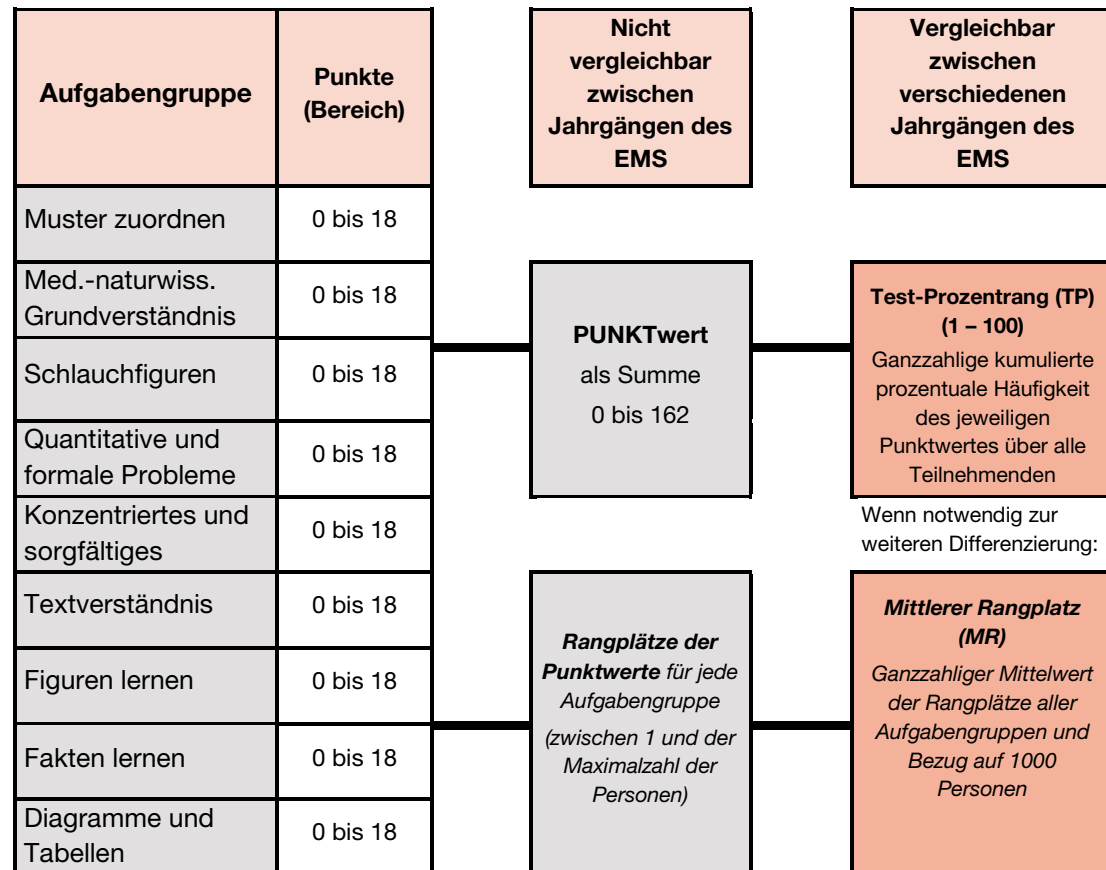
Die Stärken des EMS lassen sich folgendermassen zusammenfassen:

- 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.
- Es wird beachtet, dass für die Beantwortung der Aufgaben kein spezielles fachliches Vorwissen notwendig ist, sondern tatsächlich die „Studierfähigkeit“ als aktuelle Fähigkeit zur Wissensaneignung und Problemlösung gemessen wird. Dadurch ist der Test auch wenig trainierbar, was sich positiv auf die Sozialverträglichkeit auswirkt (kein zusätzlicher Gewinn durch zusätzliche Trainingskurse nachgewiesen, wenn die empfohlene Vorbereitung erfolgt).
- Es ist ein ökonomisches und genau kapazitätsentsprechendes Zulassungsverfahren möglich, die Rechtsfähigkeit des Verfahrens hat sich bei mehreren Überprüfungen (auch gerichtlich) bestätigt.

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 Studienerfolgs, 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.

## 6.1 Berechnung der Werte 2022

Die Aufgabengruppen des EMS liefern Punkte als Summe richtig gelöster Aufgaben zwischen 0 und 18. 2022 konnten somit maximal 162 Punkte erreicht werden. Aufgrund der Kennwertanalysen können alle Aufgaben gewertet werden.



**Abbildung 1: Punktwerte der einzelnen Aufgabengruppen 2022 und ihre Zusammenführung über den Punktwert zum Test-Prozentrang und mittleren Rangplatz.**

Für die Aufgabengruppe „Konzentriertes und sorgfältiges Arbeiten“ müssen insgesamt 1600 Zeichen bearbeitet werden – davon sind 400 Zielzeichen zu markieren. Es können in der zur Verfügung stehenden Zeit in der Regel nicht alle Zeichen bearbeitet werden. Die Position des letzten markierten Zeichens bestimmt, wie viele Zeichen als bearbeitet gewertet werden. Alle übersehenen und fälschlich markierten Zeichen vor diesem letzten bearbeiteten Zeichen zählen als Fehler und werden von der Menge der korrekt markierten Zeichen abgezogen. Der resultierende Wert ist der „Rohwert“, der in eine Skala zwischen 0 und 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 nicht zwischen Tests verschiedener Jahre vergleichbar ist.

Aus der Gesamtpunktzahl (Punktwert) werden zwei Werte berechnet:

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

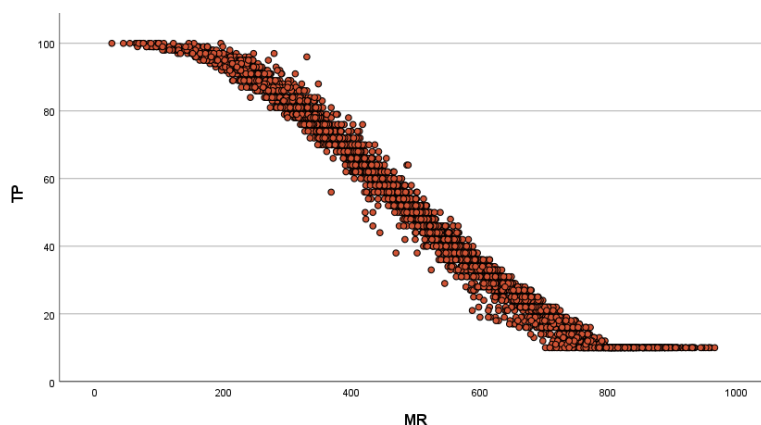
Der Test-Prozentrang (TP) zeigt an, wie viele Prozent aller teilnehmenden Personen ein schlechteres oder gleich gutes Ergebnis erreicht haben. Die Differenz zu 100 zeigt an, wie viele Prozent der Personen ein besseres Testergebnis erreicht haben. Er wird direkt aus der kumulierten Häufigkeits-Verteilung 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.

## 6.2 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 zugelassen werden können. Diese Rangplätze werden dann 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.

Werden in 2 Aufgabengruppen maximal 20 Punkte erzielt, gehen immer 40 Punkte in den Punktwert ein. Haben in einer Aufgabengruppe 100 Personen diesen Wert erreicht, erhalten diese Personen den mittleren Rangplatz 50.5. Haben in einer anderen Aufgabengruppe nur 10 Personen diesen Wert erreicht, beträgt der entsprechende Rangplatz 5.5. Die zweite Aufgabengruppe wird besser bewertet, weil die 20 Punkte dort schwerer zu erzielen waren.

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



**Abbildung 2: Beziehung zwischen Test-Prozentrang (TP) und mittlerem Rangplatz (MR) für 2022 (Werte kleiner 10 auf TP 10 angehoben).**

Der mittlere Rangplatz MR bewegt sich 2022 zwischen 26 und 966. Niedrige Werte stehen für bessere Leistungen. Mittlerer Rangplatz und Test-Prozentrang korrelieren mit 0.985 (trotz der Anhebung aller TP-Werte unter 10 auf 10).

**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 übertroffen hat als Kandidat B.

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

## 7 Testanwendung 2022

### 7.1 Verteilungsprüfung

Die Verteilung des Punktwerts entspricht annähernd einer Normalverteilung. Metrische Prüfverfahren sind bedingt aussagefähig. Die Punktwerte der einzelnen Aufgabengruppen sind, wie in allen Vorjahren, statistisch nicht als normalverteilt anzusehen (Tabelle 3).

**Tabelle 3: 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	91.5	7.5	13.6	8.0	8.0	11.7	9.5	13.8	7.0	12.3
Median	93	7	14	8	8	12	9	14	7	13
Modalwert	94	7	17	9	7	12	9	16	6	13
Stand.-abweichung	18.7	2.9	3.5	3.0	3.0	3.2	3.4	3.0	2.7	2.9
Spannweite	124	18	18	18	17	16	17	18	18	18
25. Perzentil	79	5	11	6	6	10	7	12	5	11
75. Perzentil	105	9	16	10	10	14	12	16	9	14
Schiefe	-0.2	0.3	-0.8	0.3	0.1	-0.2	0.3	-0.8	0.3	-1.3
Kurtosis	-0.3	0.1	0.0	0.0	-0.4	-0.4	-0.4	0.5	0.0	3.8

Bei der Aufgabengruppe „Diagramme und Tabellen“ wurden 2022 die tiefsten mittleren Punktwerte erreicht, gefolgt von „Quantitative und formale Probleme“ – „Schlauchfiguren“ und *Muster zuordnen* waren die leichtesten. Alle Aufgabengruppen differenzieren ausreichend bis sehr gut im relevanten Wertebereich.

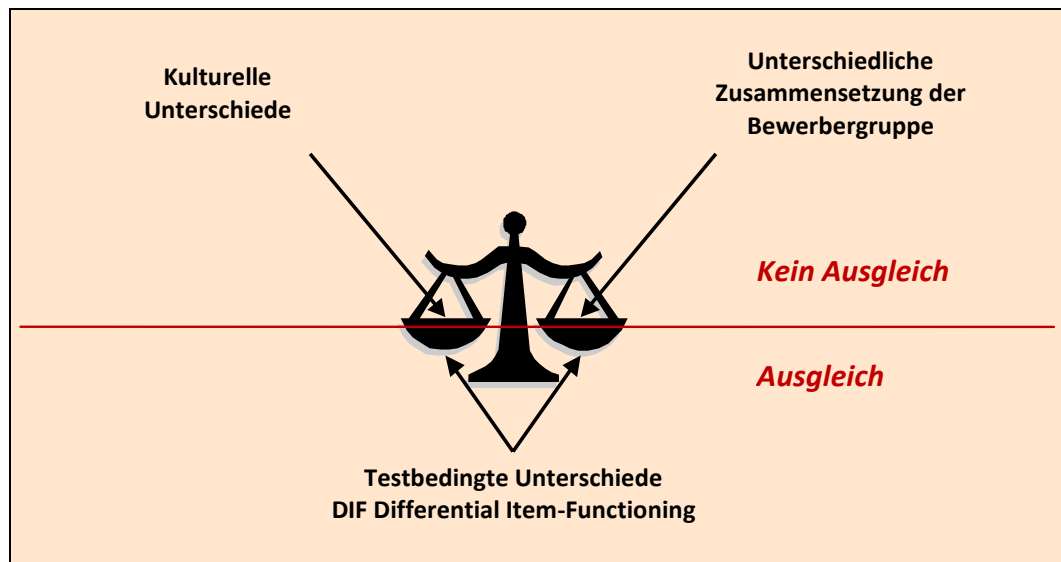


## 7.2 Äquivalenz der Sprachversionen

Sprachrelevante Aufgaben wurden nach den gleichen Kriterien wie in den Vorjahren von der deutschen Vorlage adaptiert. Eine ausführliche Beschreibung des Verfahrens kann dem Bericht 5 (1999) entnommen werden. Zielsetzung bleibt, möglichst übereinstimmende Aufgabenschwierigkeiten (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, 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 Mittelwertdifferenzen, welche alle Aufgaben einer Gruppe gleichermaßen betreffen, abgegrenzt werden. Die Ergebnisse der letzten Jahre haben gezeigt, dass sich die Sprachgruppen in manchen Aufgabengruppen signifikant unterscheiden. Differenzen kommen nicht nur durch einzelne Aufgaben zustande, sondern auch durch eine Verschiebung der Schwierigkeiten aller Aufgaben einer Gruppe. Es ist unwahrscheinlich, dass diese systematischen Unterschiede testbedingt sind. Testbedingte Unterschiede als sprachliche Besonderheiten zeigen sich in spezifischen Abweichungen einzelner Aufgaben. Systematische Unterschiede sind dagegen mit grösserer

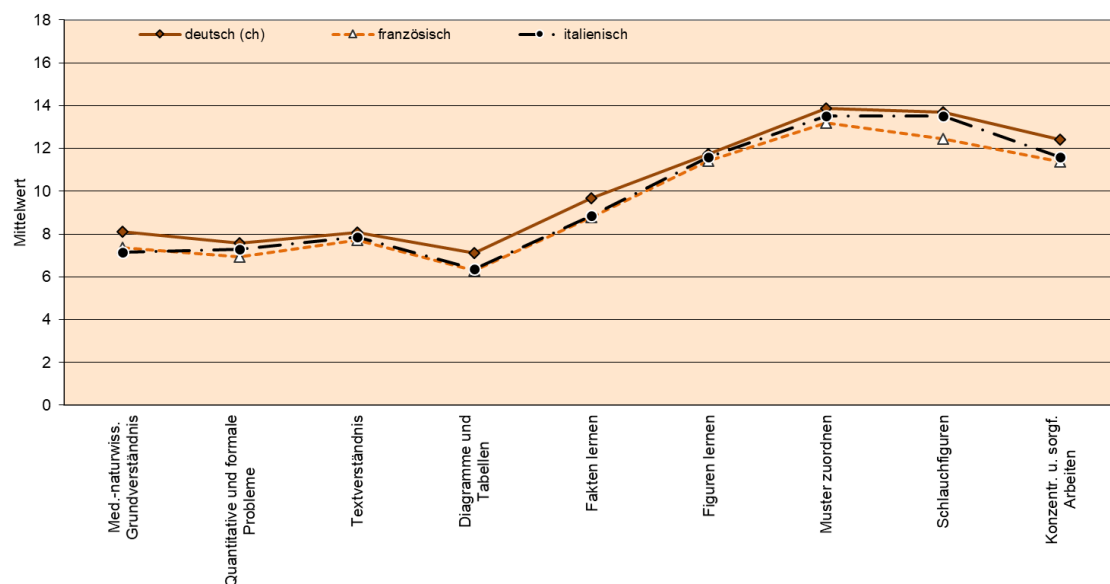
Wahrscheinlichkeit den Faktoren „Kulturunterschiede“ und „unterschiedliche Repräsentativität“ zuzurechnen.

Es ergibt sich für das Ausgleichsverfahren die nachfolgende Zielstellung:

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

### 7.2.1 Sprachvergleich für die Aufgabengruppen

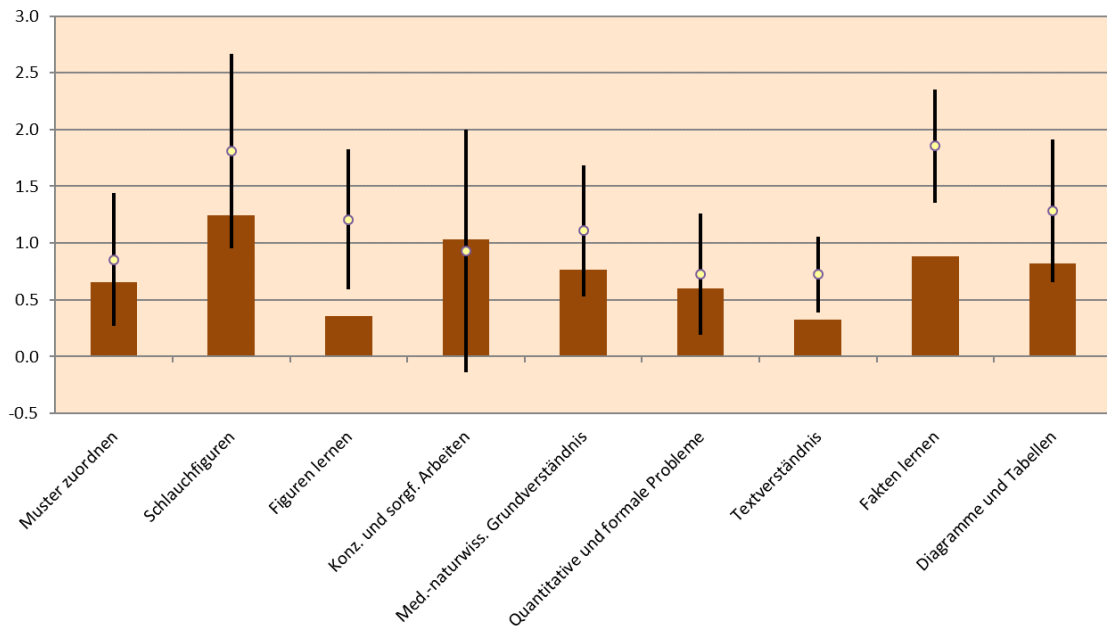
Auch 2022 treten Unterschiede zwischen den Sprachgruppen in sprachabhängigen und wenig sprachabhängigen Aufgabengruppen auf – der grösste Unterschied ist bei der sprachfreien Aufgabe „Schlauchfiguren“ zwischen der deutschen und französischen Sprachgruppe zu beobachten. Die unkorrigierten Punktzahlen nach Sprachen sind nachfolgend (Abbildung 4) abgebildet.



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

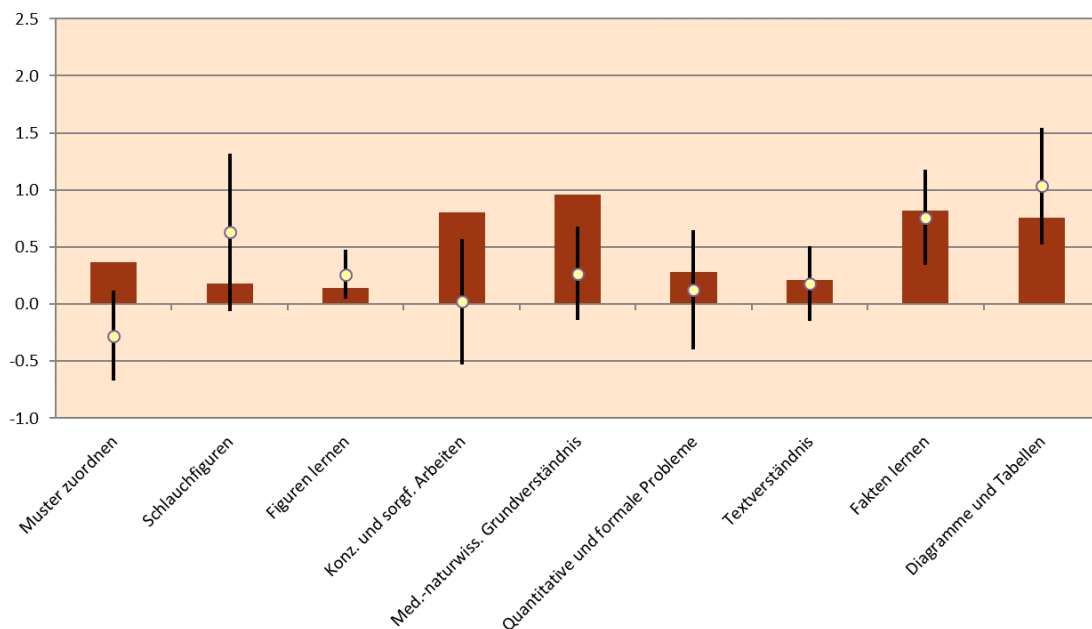
Abbildung 5 zeigt den Vergleich zwischen deutsch- und französischsprachigen Testteilnehmenden für das Jahr 2022, in Abbildung 6 ist der Vergleich zwischen deutsch- und italienischsprachigen Personen dargestellt. Die senkrechten Linien bezeichnen Mittelwert und Standardabweichung der Differenzen über die Jahrgänge für die letzten 10 Jahre. Positive Differenzwerte deuten auf bessere Ergebnisse der deutschen Sprachgruppe hin, negative auf solche der jeweils anderen Sprachgruppe.

Für die französische Sprachgruppe treten, mit Ausnahme von „Konzentriertes und sorgfältiges Arbeiten“, deutlich geringere Unterschiede auf, als dies in den letzten 10 Jahren der Fall war.



**Abbildung 5: Differenz für Aufgabengruppen deutschsprachig-französischsprachig 2022 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2012-2021.**

Die Ergebnisse der italienischsprachigen Teilnehmenden liegen mehrheitlich im Rahmen der bekannten Abweichungen. Das auffälligste Ergebnis betrifft die Aufgabengruppe *Muster zuordnen*. Dort erzielte die italienischsprachige Gruppe mehrheitlich bessere Ergebnisse als die deutschsprachige Referenzgruppe – in diesem Jahr sind die italienischsprachigen Ergebnisse allerdings etwas schlechter ausgefallen.



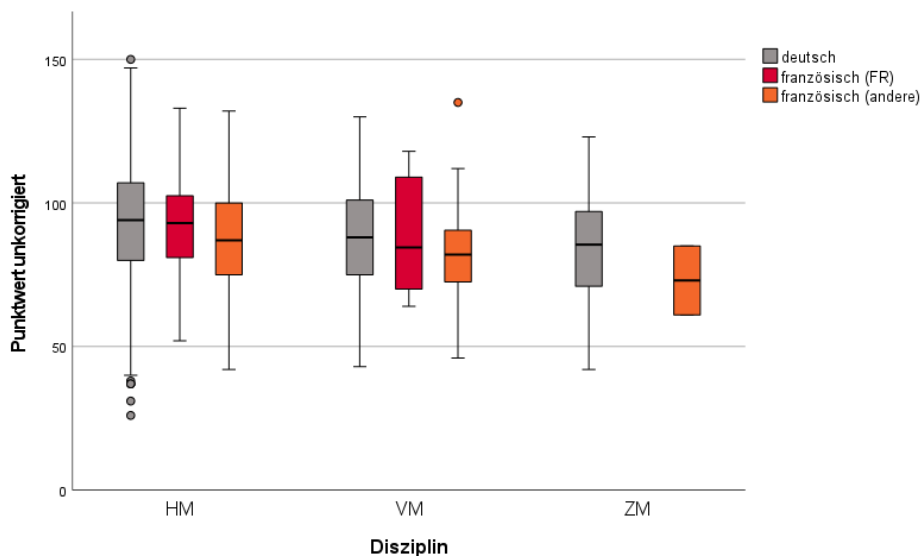
**Abbildung 6: Differenz für Aufgabengruppen deutschsprachig-italienischsprachig 2022 (unkorrigiert) und Mittelwerte mit Standardabweichungen der Testdurchführungen 2012-2021.**

### 7.2.1.1 Vergleich von französischsprachigen KandidatInnen aus dem Kanton Freiburg und ausserhalb

Im Vorfeld des EMS 2022 kam es zu politischen Diskussionen, die sich um die Zukunft der (französischsprachigen) ärztlichen Versorgung im Kanton Freiburg drehten. Vorgebracht wurde, dass französischsprachige Personen durch die verfügbaren Vorbereitungsangebote in deutscher Sprache beim EMS benachteiligt wären. Swissuniversities und das ZTD propagieren andererseits seit Jahren die Durchführung von Probeläufen in allen drei Testsprachen und stellen entsprechende Unterlagen allen Schulen zur Verfügung. Frühere Evaluationen der Vorbereitungsstrategien der Teilnehmenden haben gezeigt, dass das Angebot solcher Probeläufe dennoch insbesondere in französischsprachigen Regionen weniger bekannt ist und genutzt wird. Diesem Umstand wurde, neben den bereits bestehenden kostenfreien Angeboten, 2022 zusätzlich entgegengewirkt, indem das bereits bestehende Angebot an Probeläufen im Kanton Freiburg stärker propagiert wurde. Zusätzlich wurde an der Universität Freiburg ein weiterer Probelauf durch „NC-Wiki“ angeboten. Dabei handelt es sich um eine studentische, nicht gewinnorientierte Initiative.

Sollten die durch swissuniversities und ZTD propagierten Massnahmen Wirkung zeigen, müsste sich dies insbesondere auf die im Kanton Freiburg wohnhaften Personen in einem grösseren Ausmass auswirken, da zu erwarten wäre, dass der Anteil teilnehmender Personen aus diesem Kanton gegenüber anderen Kantonen erhöht war.

In den Vorjahren unterschieden sich diese Personen nicht signifikant von „ausserkantonale“ wohnhaften, französischsprachigen KandidatInnen. Nachfolgend (Abbildung 7) sind die entsprechenden Werteverteilungen der unkorrigierten Punktwerte für die drei Disziplinen und die Sprachgruppen deutsch (n=2877/286/180), französisch „Freiburg“ (n=48/10/0) und französisch „andere“ (n=164/123/2) dargestellt. Zwischen der deutschen Kohorte und den französischsprachigen KandidatInnen aus dem Kanton Freiburg treten 2022 in Humanmedizin und Veterinärmedizin (keine Teilnehmenden für Zahnmedizin) keine signifikanten Mittelwertsdifferenzen auf, wohl aber zwischen den deutschsprachigen Personen und den französischsprachigen Testteilnehmenden, die nicht im Kanton Freiburg wohnhaft sind.



**Abbildung 7: Vergleich nach Disziplinen zwischen französischsprachigen KandidatInnen mit Wohnkanton Freiburg und anderen französischsprachigen Personen, sowie Vergleichsgruppe „deutsch“ (Punktwert ohne Sprachausgleich).**

Die Differenzen zwischen den beiden französischsprachigen Gruppen erreichen aufgrund der geringen Gruppengrösse der im Kanton Freiburg wohnhaften Personen nur auf Ebene einiger Aufgabengruppen signifikantes Niveau, die erreichten Punktwerte liegen jedoch praktisch

durchwegs (Ausnahme siehe unten) über denjenigen der französischsprachigen Personen aus anderen Kantonen. Für die mindestens tendenziell unterschiedlichen Leistungen der beiden Gruppen können natürlich verschiedene Ursachen relevant sein. Der Umstand, dass zwischen den Ergebnissen deutschsprachiger und im Kanton Freiburg wohnhafter französischsprachigen Personen in diesem Jahr keine signifikanten Unterschiede bezüglich der erreichten Punktwerte (mehr) auftreten, spricht hingegen zumindest nicht gegen die Wirksamkeit von Probelaufen – und relativiert gleichzeitig die oft propagierte höhere Effizienz kostenpflichtiger Vorbereitungsangebote.

Tabelle 4 weist die detaillierten Werte und Mittelwertsvergleiche für die einzelnen Aufgabengruppen aus. Zwischen deutschsprachigen und im Kanton Freiburg wohnhaften französischsprachigen Personen treten auch auf Ebene der Aufgabengruppen kaum (mit Ausnahme von „Konzentriertes und sorgfältiges Arbeiten“) signifikante Mittelwertsdifferenzen auf. In drei Aufgabengruppen erreichen die französischsprachigen Personen aus dem Kanton Freiburg signifikant (trotz geringer Gruppengrösse) bessere Ergebnisse, als französischsprachige Personen aus anderen Kantonen.

**Tabelle 4: Signifikanzprüfungen für Prozenträge und Aufgabengruppen: d, f(FR), f(andere)**

Sprachgruppe		Mittelwert	Stand.abw.	Minimum	Maximum
Test-Prozentrang $d^{**} > f(a)$	d	52.4	28.24	10	100
	f(FR)	49.6	27.97	10	99
	f(andere)	40.6	25.42	10	100
Muster zuordnen $d^{**} > f(a)$	d	13.9	2.98	0	18
	f(FR)	13.4	2.96	6	18
	f(andere)	13.2	3.12	4	18
MNGV $d^{**} > f(a)$	d	8.1	3.04	0	17
	f(FR)	7.7	2.84	0	15
	f(andere)	7.3	2.86	1	15
Schlauchfiguren $d^{**}, f(FR)^* > f(a)$	d	13.7	3.43	0	18
	f(FR)	13.7	3.28	4	18
	f(andere)	12.2	3.58	3	18
QfP $(d, f(FR))^{**} > f(a)$	d	7.6	2.86	0	18
	f(FR)	7.9	3.00	2	14
	f(andere)	6.8	2.82	1	16
Textverständnis $d^{**}, f(FR)^* > f(a)$	d	8.1	2.97	0	18
	f(FR)	8.5	2.85	3	17
	f(andere)	7.5	2.77	2	16
Figuren lernen	d	11.7	3.22	2	18
	f(FR)	11.8	3.36	5	18
	f(andere)	11.3	3.15	2	18
Fakten lernen $d^{**} > f(a)$	d	9.7	3.41	1	18
	f(FR)	9.2	3.12	3	17
	f(andere)	8.7	3.10	2	18
DuT $d^{**} > f(a)$	d	7.1	2.66	0	18
	f(FR)	6.6	2.62	2	15
	f(andere)	6.2	2.47	1	15
Konz. und sorgf. Arbeiten $d > f(FR)^*, f(a)^{**}$	d	12.4	2.95	0	18
	f(FR)	11.4	2.95	0	17
	f(andere)	11.4	2.83	0	18

n: deutsch 3343 / französisch (FR) 58 / französisch (andere) 289

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

Die Effizienz des Probelauf-Angebots dürfte nicht alleine auf den eigentlichen Anlass zurückzuführen sein. Vielmehr ist anzunehmen, dass bereits das Vorhandensein entsprechender Angebote in der geografischen Nähe die Bereitschaft zur intensiveren Vorbereitung erhöht. Auch können Personen, die selber nicht teilnehmen, beispielsweise über

Freunde und Bekannte, die vom Probelauf erzählen, zu grösseren Vorbereitungsanstrengungen motiviert werden. Die genauen Wirkmechanismen dürften vielschichtig sein – in jedem Fall scheinen sie zu funktionieren.

## 7.2.2 Darstellung des Korrekturverfahrens

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

### 7.2.2.1 Identifikation und Ausgleich testbedingter Unterschiede geschehen wie folgt:

- Sprachausgleiche werden nur in den sprachabhängigen Aufgabengruppen „Medizinisch-naturwissenschaftliches Grundverständnis“, „Quantitative und formale Probleme“, „Textverständnis“, „Fakten lernen“ 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. Der Ausgleich wird nicht symmetrisch vorgenommen, da es darum geht, eine Benachteiligung der Vergleichsgruppe gegenüber der Referenzgruppe zu vermeiden.
- Als Sprachausgleich wird den französisch- und italienischsprachigen Probanden die Differenz zur entsprechenden Aufgabenschwierigkeit in der deutschsprachigen Version gutgeschrieben, falls sie die Aufgabe nicht korrekt gelöst haben. Um jede Benachteiligung auszuschliessen, wird nicht nur die Differenz zur Regressionsgeraden ausgeglichen. Die Lage dieser Regressionsgeraden wird auch von allen DIF-Aufgaben mitbestimmt. Man kann das so interpretieren, dass für DIF-Aufgaben bei falscher Antwort zumindest der Schwierigkeitsunterschied zur Referenzgruppe ausgeglichen wird.

Eine generelle Gutschrift des Mittelwertunterschiedes pro Aufgabengruppe für die französisch- und italienischsprachige Gruppe könnte für einzelne Personen zu mehr Punkten pro Aufgabengruppe führen, als 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.

### 7.2.2.2 Welche Aufgabengruppen ausgleichen?

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

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

Aufgabengruppe	Sprache	n	m	s	Sig.	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3343	8.12	3.037	<b>.000</b>	<b>Ja</b>
	f	347	7.36	2.858		
Quantitative und form. Probleme	d	3343	7.56	2.863	<b>.000</b>	<b>Ja</b>
	f	347	6.94	2.876		
Textverständnis	d	3343	8.07	2.968	<b>.015</b>	<b>Ja</b>
	f	347	7.71	2.806		
Fakten lernen	d	3343	9.66	3.415	<b>.000</b>	<b>Ja</b>
	f	347	8.77	3.106		
Diagramme und Tabellen	d	3343	7.12	2.661	<b>.000</b>	<b>Ja</b>
	f	347	6.29	2.495		

**Tabelle 6: Signifikanzprüfung deutsche vs. italienische Sprachform (sprachabhängige Aufgabengruppen).**

Aufgabengruppe	Sprache	n	m	s	Sig.	DIF-Prüfung
Med.-naturwiss. Grundverständnis	d	3343	8.12	3.037	<b>.000</b>	<b>Ja</b>
	i	148	7.16	2.847		
Quantitative und form. Probleme	d	3343	7.56	2.863	.117	Nein
	i	148	7.28	2.733		
Textverständnis	d	3343	8.07	2.968	.203	Nein
	i	148	7.86	3.162		
Fakten lernen	d	3343	9.66	3.415	<b>.002</b>	<b>Ja</b>
	i	148	8.84	3.091		
Diagramme und Tabellen	d	3343	7.12	2.661	<b>.000</b>	<b>Ja</b>
	i	148	6.36	2.408		

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

### 7.2.3 Identifikation von DIF-Aufgaben und Bestimmung des Korrekturwertes

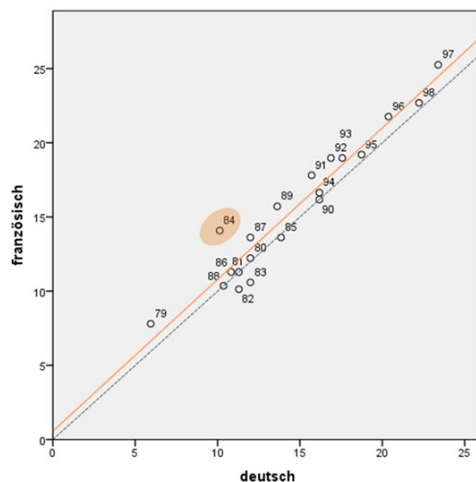
Eine Möglichkeit zur Identifikation von DIF, auch bei relativ kleinen Stichprobengrößen, ist die Methode „Delta-Plot“. Bei diesem Verfahren werden die zu vergleichenden Aufgabenschwierigkeiten z-standardisiert und anschliessend in „Delta-Werte“ transformiert.

Die Transformation erfolgt über die Formel:  $\Delta = 13 - 4z$



Dies bedeutet, dass  $\Delta$  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 8: 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

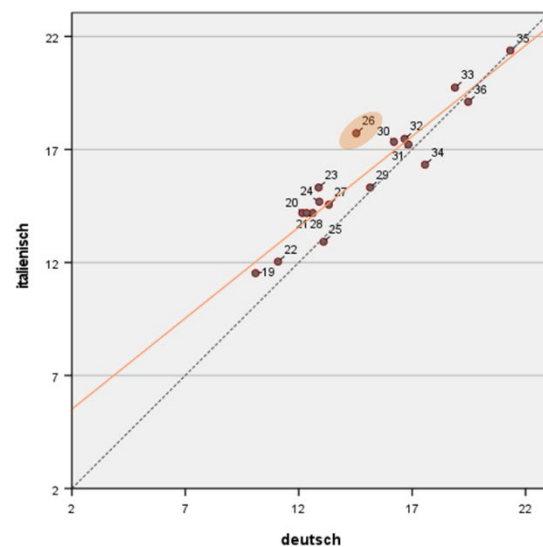
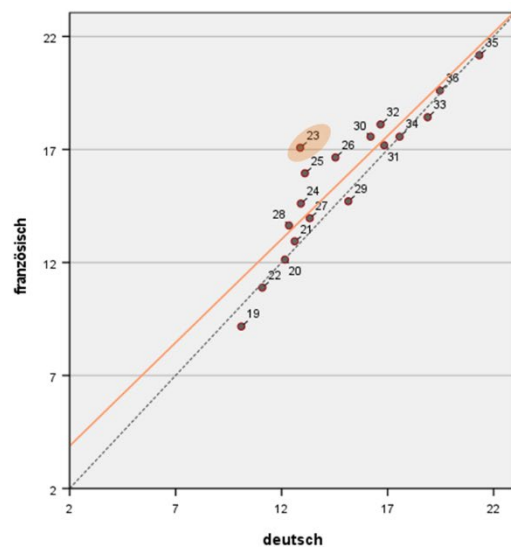
massgebend. Es werden also jene Aufgaben ausgeglichen, **welche mindestens 1.5 Punkte** (Longford, Holland & Thayer, 1993) von der Regressionsgeraden in entgegengesetzter Richtung zur Diagonalen abweichen. Als Sprachausgleich wird der benachteiligten Sprachgruppe die Differenz der betreffenden Aufgabenschwierigkeit zur deutschen Aufgabenschwierigkeit gutgeschrieben (also die Differenz der Lösungs-wahrscheinlichkeit). Dies betrifft ausschliesslich jene Personen, welche die fragliche Aufgabe nicht korrekt beantwortet haben.

### 7.2.3.1 Medizinisch-naturwissenschaftliches Grundverständnis

Die Aufgabe 23 liegt für die französische Sprachgruppe über der kritischen Differenz, es werden den Personen, welche diese Aufgabe nicht korrekt gelöst haben, 0.22 Punkte gutgeschrieben. Für die italienische Sprachgruppe ist Item 26 betroffen, dort werden 0.17 Punkte gutgeschrieben.

**Tabelle 7: DIF-Analyse „Medizinisch-naturwissenschaftliches Grundverständnis“**

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
19	0.71	0.76	0.63	10.10	9.17	11.53		
20	0.59	0.60	0.49	12.16	12.13	14.19		
21	0.57	0.55	0.49	12.62	12.94	14.19		
22	0.65	0.66	0.60	11.09	10.89	12.04		
23	0.56	0.33	0.43	12.88	17.08	15.32	0.22	
24	0.55	0.46	0.46	12.91	14.61	14.69		
25	0.54	0.39	0.55	13.10	15.95	12.92		
26	0.47	0.35	0.30	14.54	16.65	17.72		0.17
27	0.53	0.50	0.47	13.33	13.96	14.56		
28	0.58	0.52	0.49	12.35	13.64	14.19		
29	0.43	0.46	0.43	15.15	14.71	15.32		
30	0.38	0.31	0.32	16.20	17.57	17.34		
31	0.34	0.33	0.32	16.84	17.19	17.22		
32	0.35	0.28	0.31	16.67	18.11	17.47		
33	0.23	0.26	0.19	18.89	18.43	19.74		
34	0.31	0.31	0.37	17.57	17.57	16.33		
35	0.10	0.11	0.10	21.33	21.17	21.38		
36	0.20	0.20	0.22	19.47	19.61	19.11		



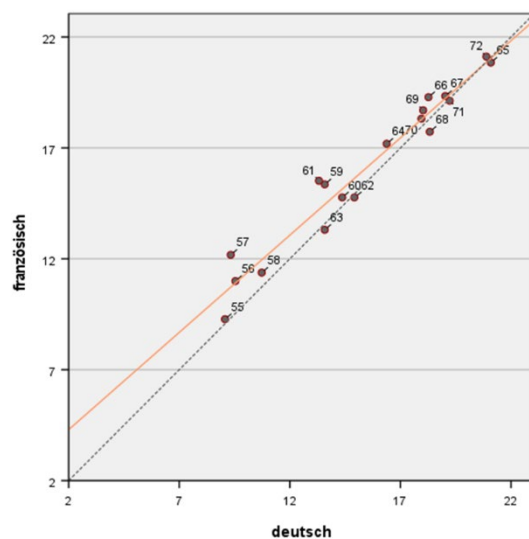
**Abbildung 9: Delta-Plot für „Medizinisch-naturwissenschaftliches Grundverständnis“ (deutsch-französisch und deutsch-italienisch) .**

### 7.2.3.2 Quantitative und formale Probleme

Alle Aufgaben liegen eng an der Regressionsgeraden. Trotz signifikanter Differenz hinsichtlich des Punktwerts dieser Aufgabengruppe gibt es auf Aufgabenebene keine Hinweise auf übersetzungsbedingte Einflüsse für die französische Sprachgruppe.

**Tabelle 8: DIF-Analyse „Quantitative und formale Probleme“**

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
55	0.76	0.75		9.07	9.28		-	
56	0.73	0.66		9.54	11.00		-	
57	0.75	0.59		9.33	12.18		-	
58	0.67	0.64		10.73	11.38		-	
59	0.52	0.42		13.58	15.36		-	
60	0.48	0.46		14.37	14.77		-	
61	0.53	0.41		13.32	15.52		-	
62	0.45	0.46		14.92	14.77		-	
63	0.52	0.53		13.58	13.31		-	
64	0.37	0.33		16.38	17.19		-	
65	0.12	0.13		21.09	20.85		-	
66	0.27	0.21		18.27	19.29		-	
67	0.23	0.21		19.03	19.34		-	
68	0.26	0.30		18.33	17.73		-	
69	0.28	0.24		18.03	18.70		-	
70	0.29	0.27		17.95	18.32		-	
71	0.22	0.22		19.23	19.13		-	
72	0.13	0.12		20.89	21.12		-	



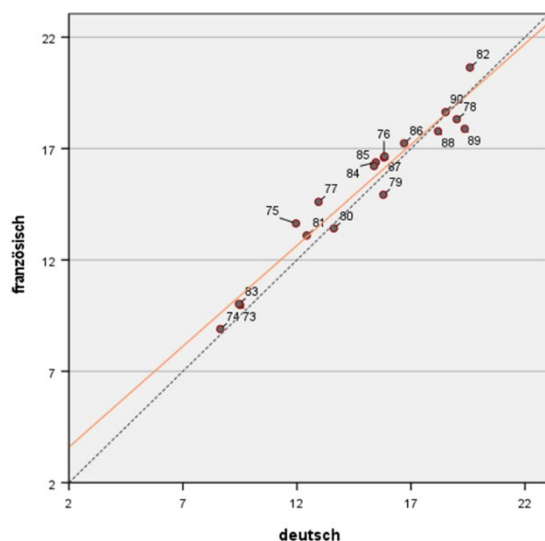
**Abbildung 10: Delta-Plot für „Quantitative und formale Probleme“ (deutsch-französisch).**

### 7.2.3.3 Textverständnis

Alle Aufgaben liegen eng an der Regressionsgeraden. Trotz signifikanter Differenz hinsichtlich des Punktwerts dieser Aufgabengruppe gibt es auf Aufgabenebene keine Hinweise auf übersetzungsbedingte Einflüsse für die französische Sprachgruppe.

**Tabelle 9: DIF-Analyse „Textverständnis“**

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
73	0.74	0.71		9.51	9.98		-	
74	0.78	0.77		8.64	8.90		-	
75	0.61	0.52		11.96	13.64		-	
76	0.40	0.36		15.82	16.60		-	
77	0.55	0.46		12.95	14.61		-	
78	0.23	0.27		19.01	18.32		-	
79	0.40	0.45		15.79	14.93		-	
80	0.52	0.53		13.62	13.42		-	
81	0.58	0.54		12.43	13.10		-	
82	0.20	0.14		19.59	20.64		-	
83	0.74	0.71		9.46	10.03		-	
84	0.42	0.37		15.46	16.38		-	
85	0.42	0.38		15.38	16.22		-	
86	0.35	0.32		16.70	17.24		-	
87	0.40	0.35		15.85	16.65		-	
88	0.27	0.29		18.19	17.78		-	
89	0.21	0.29		19.36	17.89		-	
90	0.25	0.25		18.52	18.64		-	



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

### 7.2.3.4 Diagramme und Tabellen

In der Aufgabengruppe „Diagramme und Tabellen“ werden der französischen Sprachgruppe bei Falschantwort für die Aufgabe 134 0.19, für die Aufgabe 136 0.15 Punkte gutgeschrieben. Für die italienischsprachige Gruppe fällt keine Aufgabe auf.

Tabelle 10: DIF-Analyse „Diagramme und Tabellen“

Aufgabe	Schwierigkeit			Delta-Wert			Bonus f	Bonus i
	d	f	i	d	f	i		
127	0.72	0.55	0.64	9.87	12.94	11.41		-
128	0.78	0.77	0.76	8.75	8.84	9.01		-
129	0.73	0.67	0.81	9.72	10.73	8.13		-
130	0.57	0.53	0.45	12.56	13.42	14.82		-
131	0.50	0.36	0.36	13.89	16.49	16.46		-
132	0.46	0.57	0.37	14.62	12.56	16.33		-
133	0.46	0.47	0.45	14.73	14.55	14.94		-
134	0.42	0.22	0.39	15.48	19.07	16.08	0.19	-
135	0.34	0.33	0.30	16.87	17.03	17.72		-
136	0.29	0.14	0.20	17.78	20.64	19.61	0.15	-
137	0.30	0.22	0.28	17.60	19.24	18.10		-
138	0.22	0.17	0.23	19.07	20.15	18.98		-
139	0.28	0.25	0.17	18.05	18.59	20.12		-
140	0.20	0.15	0.11	19.46	20.47	21.25		-
141	0.36	0.38	0.38	16.57	16.11	16.21		-
142	0.16	0.18	0.15	20.35	19.94	20.50		-
143	0.22	0.22	0.22	19.21	19.24	19.11		-
144	0.10	0.10	0.11	21.34	21.44	21.25		-

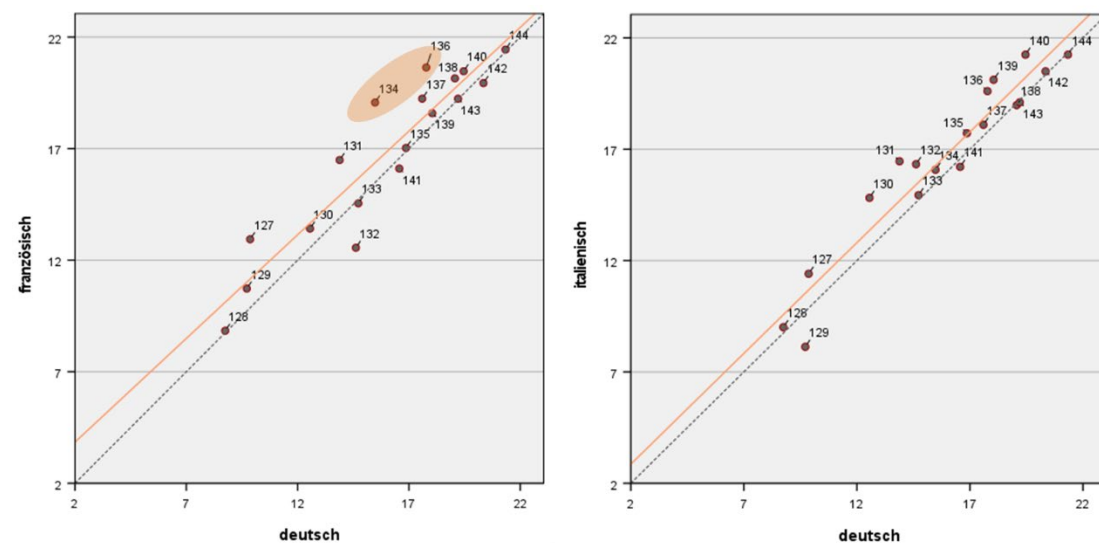


Abbildung 12: Delta-Plot für „Diagramme und Tabellen“ (deutsch-französisch und deutsch-italienisch).

### 7.2.4 Effekte der Korrektur

Die durch den Ausgleich resultierenden Bonuspunkte können Tabelle 22 entnommen werden. Für die Bestimmung des Punktwertes wird auf ganze Punkte gerundet. Der maximale Bonus in der französischen Sprachgruppe beträgt 1 Punkt, in der italienischen Sprachgruppe kommen die (ungerundeten) Bonuspunkte auf Ebene der betroffenen Aufgabengruppen zur Ermittlung des mittleren Rangplatzes zum Tragen.

**Tabelle 11: Mittelwerte und Standardabweichungen der Korrekturwerte für die Personen der französisch- und italienischsprachigen Gruppen.**

		n	Minimum	Maximum	m	s
französisch	Med.-naturw. Grundverständnis	347	0.00	0.22	0.15	0.10
	Diagramme und Tabellen	347	0.00	0.34	0.27	0.10
	GESAMT (Gerundet)	347	0.00	1.00	0.48	0.50
italienisch	Med.-naturw. Grundverständnis	148	0.00	0.17	0.12	0.08
	Diagramme und Tabellen	148	0.00	0.15	0.11	0.07
	GESAMT (Gerundet)	148	0.00	0.00	-	-

Die Mittelwertsdifferenzen für jede Aufgabengruppe nach dem Sprachausgleich sind in Tabelle 12 dargestellt. Die Ergebnisse multipler Mittelwertvergleiche 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.

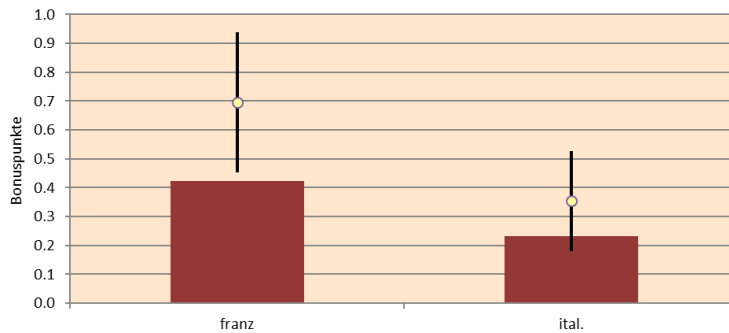
**Tabelle 12: Test-Prozentrang und Punktwerte inklusive Sprachausgleich.**

		Mittelwert	Stand.- abw.	Minimum	Maximum
Test-Prozentrang $d^{**} > f, i$	deutsch	52.42	28.24	10	100
	französisch	42.13	26.04	10	100
	italienisch	45.05	27.42	10	99
Quantitative und formale Probleme $d^{**} > f$	deutsch	7.56	2.86	0	18
	französisch	6.94	2.88	1	16
	italienisch	7.28	2.73	1	14
Schlauchfiguren (d, i) $^{**} > f$	deutsch	13.71	3.43	0	18
	französisch	12.44	3.57	3	18
	italienisch	13.53	3.59	2	18
Textverständnis	deutsch	8.07	2.97	0	18
	französisch	7.71	2.81	2	17
	italienisch	7.86	3.16	0	18
Med.-naturwiss. Grundverständnis (d,i) $^{**} > f$	deutsch	8.12	3.04	0	17
	französisch	7.51	2.84	0	15
	italienisch	7.27	2.82	0	14
Figuren lernen	deutsch	11.73	3.22	2	18
	französisch	11.40	3.19	2	18
	italienisch	11.59	3.08	3	17
Fakten lernen $d^{**} > (f,i)$	deutsch	9.66	3.41	1	18
	französisch	8.77	3.11	2	18
	italienisch	8.95	3.07	2	18
Muster zuordnen $d^{**} > f$	deutsch	13.88	2.98	0	18
	französisch	13.21	3.09	4	18
	italienisch	13.51	3.24	5	18
Diagramme und Tabellen $d^{**} > (f,i)$	deutsch	7.12	2.66	0	18
	französisch	6.56	2.44	1	15
	italienisch	6.36	2.41	2	13
Konz. und sorgf. Arbeiten $d^{**} > (f,i)$	deutsch	12.41	2.95	0	18
	französisch	11.39	2.84	0	18
	italienisch	11.61	2.72	0	18

n: deutsch 3343 / französisch 347 / italienisch 148

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

Der Sprachbonus sowohl der französischen wie der italienischen Sprachgruppe war 2022 geringer als in den Vorjahren (Abbildung 13).

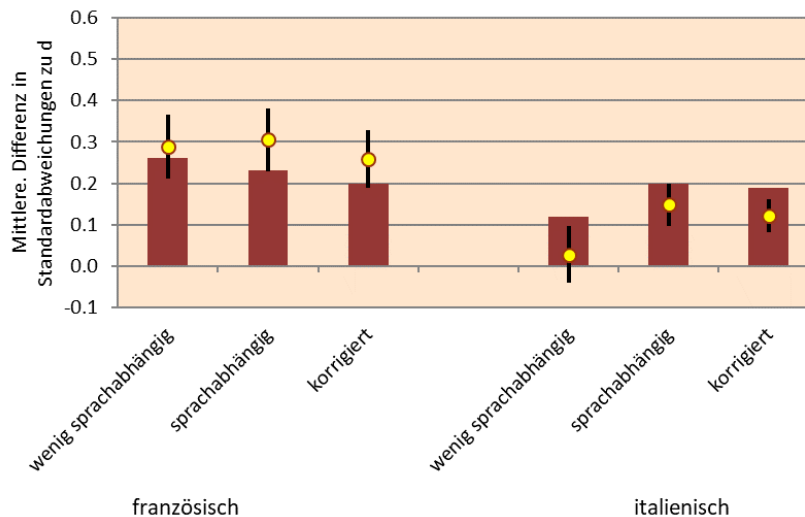


**Abbildung 13: Sprachausgleich 2022 (ungerundet) mit Mittelwert und Standardabweichung der Testdurchführungen 2012-2021**

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 14: Mittlere Abweichung (Delta z) für wenig sprachabhängige und sprachabhängige Aufgabengruppen (korrigiert und unkorrigiert), 2022 und Vergleich zu 2012-2021 (Mittelwert und Streuung).**

Die französische Sprachgruppe weist bei den sprachunabhängigen Aufgabengruppen eine höhere Differenz aus als bei den stärker sprachabhängigen. Nach der Korrektur wird letztere weiter vermindert. Die Abweichungen der italienischsprachigen Personen von der deutschen Sprachgruppe sind insgesamt etwas höher als in den Vorjahren, sowohl bei sprachabhängigen wie wenig sprachabhängigen Aufgabengruppen.

### 7.3 Vergleichbarkeit der Testlokale

2022 erfolgte die Testabnahme in 32 Testlokalen/Sektoren verteilt auf acht Testorte. Eine Besonderheit stellen wiederum die sehr kleinen Testlokale (Lokale 4 und 24) dar, die für Personen mit besonderem Schutzbedürfnis oder spezifischen Einschränkungen bereitgestellt wurden.

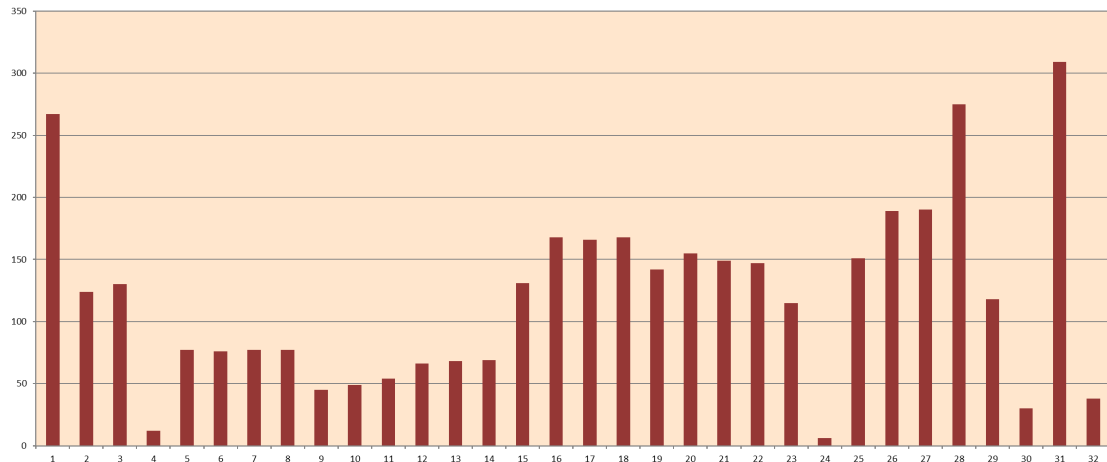


Abbildung 15: Testlokale mit Teilnehmerzahlen.

Tabelle 13: Varianzanalytische Prüfung der Homogenität für Test- und Punktwerte zwischen den deutschsprachigen Testlokalen.

		Quadratsumme	df	Mittl. Quadrat. Abweichung	F	Sig.
Testprozentrang	Zwischen	36802.157	27	1363.043	1.719	0.012
	Innerhalb	2628907.294	3315	793.034		
	Total	2665709.451	3342			
Muster zuordnen	Zwischen	305.211	27	11.304	1.274	0.156
	Innerhalb	29419.259	3315	8.875		
	Total	29724.470	3342			
Med.-naturwiss. Grundverständnis	Zwischen	295.913	27	10.960	1.190	0.228
	Innerhalb	30522.779	3315	9.207		
	Total	30818.692	3342			
Schlauchfiguren	Zwischen	379.752	27	14.065	1.197	0.222
	Innerhalb	38964.567	3315	11.754		
	Total	39344.319	3342			
Quant. und formale Probleme	Zwischen	418.273	27	15.492	1.903	0.003
	Innerhalb	26984.473	3315	8.140		
	Total	27402.746	3342			
Textverständnis	Zwischen	316.481	27	11.722	1.334	0.116
	Innerhalb	29125.909	3315	8.786		
	Total	29442.390	3342			
Figuren lernen	Zwischen	228.875	27	8.477	0.814	0.738
	Innerhalb	34527.903	3315	10.416		
	Total	34756.778	3342			
Fakten lernen	Zwischen	444.552	27	16.465	1.417	0.075
	Innerhalb	38528.455	3315	11.622		
	Total	38973.007	3342			
Diagramme und Tabellen	Zwischen	339.100	27	12.559	1.785	0.008
	Innerhalb	23326.754	3315	7.037		
	Total	23665.854	3342			
Konzentr. und sorgf. Arbeiten	Zwischen	307.195	27	11.378	1.312	0.130
	Innerhalb	28757.543	3315	8.675		
	Total	29064.738	3342			



Zwischen den (deutschsprachigen) Testlokalen treten wenige signifikante Differenzen auf. Lokal 13 (68 Teilnehmende) erreicht signifikant geringere Testprozenträge als die Personen in den anderen Testlokalen – bezüglich der einzelnen Aufgabengruppen treten dort jedoch keine signifikanten Differenzen auf. Die Durchführungsbedingungen an den einzelnen Testorten sind standardisiert – aus den Testdurchführungsprotokollen ergeben sich keine Hinweise auf mögliche Ursachen vereinzelt auftretender Unterschiede. Es gibt keinen Anlass, Korrekturen aufgrund der Testlokalzuordnung vorzunehmen.

**Tabelle 14: Darstellung der detaillierten Ergebnisse nach Testlokalen.**

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	49.8	13.7	8.0	13.3	7.4	7.9	11.8	9.2	7.1	12.4	267
2	50.2	14.0	7.5	13.7	7.2	7.7	11.8	9.0	6.9	12.1	124
3	52.8	14.3	8.2	14.0	7.5	8.3	11.7	9.3	6.9	12.4	130
4	49.8	13.5	7.8	14.8	6.2	7.0	12.8	9.6	6.6	12.6	12
5	43.7	13.3	7.4	13.3	6.8	7.6	11.2	9.2	6.0	11.8	77
6	48.5	13.7	7.9	13.2	7.3	8.2	10.9	9.6	7.2	12.0	76
7	51.1	14.1	8.1	13.4	7.3	7.8	11.6	9.5	6.7	12.5	77
8	51.4	13.8	8.1	13.3	7.1	7.8	12.1	9.7	6.9	12.3	77
9	43.7	13.1	7.4	12.9	7.2	7.6	11.4	9.2	6.4	11.2	45
10	57.8	13.8	8.8	13.7	8.4	8.2	11.8	10.0	7.4	12.7	49
11	46.2	13.3	8.1	13.1	7.5	7.6	11.5	9.2	6.4	11.8	54
12	54.9	13.6	8.3	13.9	7.8	8.5	11.5	10.4	7.4	12.7	66
13	42.5	13.2	7.4	12.9	6.5	7.6	11.0	9.3	6.5	11.5	68
14	53.6	14.1	8.4	13.8	7.4	8.5	11.5	10.0	6.9	12.2	69
15	52.1	13.8	8.1	13.7	7.6	7.6	11.7	9.6	7.1	12.5	131
16	55.5	14.3	8.4	13.6	8.3	8.3	12.2	9.8	7.1	12.4	168
17	56.4	14.2	8.2	14.0	7.8	8.5	11.9	10.0	7.5	12.7	166
18	56.1	14.1	8.5	13.8	7.8	8.2	11.9	10.0	7.4	12.6	168
19	53.0	13.6	8.3	13.8	7.8	8.0	11.5	9.6	7.2	13.1	142
20	55.1	14.1	8.5	14.1	7.8	8.3	11.7	10.0	7.3	12.6	155
21	53.8	14.0	8.2	13.8	7.7	8.5	11.7	9.8	7.3	12.5	149
22	55.2	13.9	8.3	14.3	7.7	8.4	11.7	9.8	7.5	12.5	147
23	53.0	13.9	8.3	13.6	7.5	8.1	11.7	9.8	7.5	12.5	115
24	62.7	14.8	8.7	15.2	7.2	8.7	11.0	11.0	8.3	13.5	6
25	56.5	14.5	8.1	14.1	7.6	7.7	12.0	10.6	7.2	12.4	151
26	51.5	13.7	7.9	13.7	7.7	7.8	11.8	9.5	7.3	12.4	189
27	51.0	13.9	8.0	13.9	7.1	8.2	11.6	9.7	6.8	12.4	190
28	51.7	13.7	8.1	13.7	7.8	8.1	11.9	9.4	7.0	12.4	275
29	47.3	13.7	7.5	13.9	7.4	8.0	11.8	8.7	6.5	11.8	118
30	36.1	12.9	5.7	11.9	6.8	7.4	10.7	9.3	6.0	10.7	30
31	41.8	13.3	7.3	12.5	6.9	7.7	11.3	8.7	6.2	11.3	309
32	45.0	12.8	7.7	12.3	7.4	7.6	12.3	9.5	6.8	12.2	38
Total	51.2	13.8	8.0	13.6	7.5	8.0	11.7	9.5	7.0	12.3	3838

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.

### 7.3.1 Versetzter Testbeginn in Zürich

Auch 2022 wurde der Test in Zürich zeitversetzt in zwei Gruppen durchgeführt.

Für die Testlokale 19 bis 24 in Zürich war der Testbeginn um 9.45 Uhr vorgesehen, für die Lokale 15 bis 18 um 11.15 Uhr. Neben der eigentlichen Einlassprozedur sollten durch die versetzten Einladungszeiten auch die öffentlichen Verkehrsmittel (z. B. betroffene Buslinien) entlastet werden.

Bis zum Beginn der späteren Durchführung war es in allen anderen Testorten in der Schweiz (ausser in Notfällen) untersagt, das Testlokal zu verlassen – beispielsweise bei einem Testabbruch. Dadurch wurde sichergestellt, dass keine aufgabenbezogenen Informationen an die Personen mit späterem Testbeginn weitergegeben werden konnten.

Fasst man die KandidatInnen in den Testlokalen in Zürich nach frühem oder spätem Testbeginn zusammen, müsste sich ein allfälliger Einfluss des Testzeitpunkts entsprechend in den Testleistungen manifestieren. Tatsächlich weisen die beiden Gruppen praktisch identische Werte auf – die Lokale mit frühem Beginn einen mittleren Punktwert von 93.7 ( $s=18.66$ ;  $n=714$ ), jene mit späterem Beginn einen mittleren Punktwert von 94.00 ( $s=19.06$ ;  $n=633$ ).

Es gibt keinerlei Anlass, für eine der beiden Gruppen systematische Vor- oder Nachteile aufgrund der unterschiedlichen Startzeit zu vermuten.

## 7.4 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 nicht erwartet werden.

Nach 2006 hatte sich der Männeranteil in Humanmedizin vorübergehend bei rund 40% stabilisiert, seit 2014 ist er allerdings weiter gesunken. 2022 waren 30.4% der Bewerbenden mit gültigem Testergebnis männlich, 69.6% weiblich (über alle Disziplinen).

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

		Testsprache			Wunschdisziplin			Gesamt
		deutsch	französisch	italienisch	HM	VM	ZM	
männlich	m	55.4	47.9	38.7	55.0	45.2	44.2	54.1
	s	28.4	25.8	24.1	28.4	25.9	25.6	28.3
	n	1042	67	56	1064	48	53	1165
weiblich	m	51.1	40.8	48.9	52.1	42.3	38.1	49.9
	s	28.1	25.9	28.7	28.2	26.0	25.6	28.1
	n	2301	280	92	2150	388	135	2673

m: Mittelwert, s: Standardabweichung, n: Personenzahl

Für die zahlenmässig bedeutendste deutschsprachige Teilnehmergruppe wird 2022 ein Unterschied von 4.3 Test-Prozentrangpunkten registriert. Die Differenz hat gegenüber 2021 deutlich (damals 6.5) abgenommen.

**Tabelle 16: Kombinierte Statistiken für den Test-Prozentrang nach Geschlechtern, Sprache und Disziplin.**

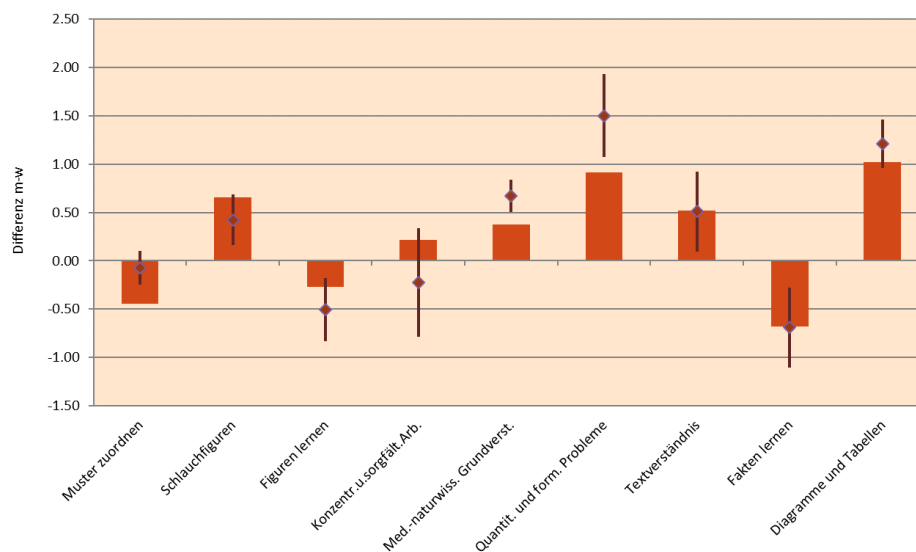
		Humanmedizin			Veterinärmedizin			Zahnmedizin		
		d	f	i	d	f	i	d	f	i
männlich	m	56.1	49.6	40.1	47.6	36.7	42.8	46.3		18.0
	s	28.5	26.3	24.1	26.9	20.6	29.1	25.3		9.5
	n	958	58	48	35	9	4	49		4
weiblich	m	52.7	45.5	48.2	45.0	35.2	59.4	38.7	23.0	10.5
	s	28.2	27.3	28.1	26.3	23.1	29.4	25.6	18.4	0.7
	n	1919	154	77	251	124	13	131	2	2

Betrachtet man die Unterschiede disziplin- und sprachspezifisch, zeigt sich für die deutschsprachige Gruppe mit Wunschdisziplin Humanmedizin eine signifikante Differenz von 3.4 Test-Prozentrangpunkten, und in Zahnmedizin von 7.6 Test-Prozentrangpunkten. Die anderen Unterschiede sind nicht signifikant.

**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	0.14	0.71	3.03	2875	<b>0.001</b>
	VM	homogen	0.02	0.89	0.56	284	0.288
	ZM	homogen	0.32	0.57	1.78	178	<b>0.038</b>
Französisch	HM	homogen	1.44	0.23	0.99	210	0.162
Italienisch	HM	homogen	2.42	0.12	-1.65	123	0.051

Abbildung 16 verdeutlicht die Vergleichbarkeit der Geschlechterdifferenzen mit den Vorjahren. Die diesjährigen Punktwert-Unterschiede auf Ebene der Aufgabengruppen fallen gegenüber früheren Durchführungen vergleichbar aus.



(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 2012-2021.**

Unter den gemäss Modell (ohne Nachrücker) zugelassenen Personen die 2022 am Test teilgenommen haben, beträgt der Frauenanteil in Humanmedizin 64.8%, in Veterinärmedizin 87.6%, in Zahnmedizin 68.5% , in Chiropraktik 60.0%.

## 7.5 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 die Bewerber und Bewerberinnen zu Humanmedizin und Veterinärmedizin sind signifikante Differenzen einzig zwischen Freiburg und Zürich (Luzerner Track) feststellbar.

**Tabelle 18: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschhochschulen (d).**

	Wunschhochschule	n	Mittelwert	Standardabw.	Quadratsumme	Df	MQ	F	Sig.
HM	Basel	577	52.7	28.5	13481.8	6	2247.0	2.8	<b>.010</b>
	Bern	737	52.0	27.3					
	ETH	264	57.5	30.3					
	Freiburg	41	48.0	30.6					
	Zürich	1099	55.1	28.5					
	Zürich (LU)	69	59.2	24.9					
	Zürich (SG)	86	49.3	26.0					
VM	Bern	122	42.8	24.6	1291.2	1	1291.2	1.9	0.174
	Zürich	164	47.1	27.6					
ZM	Basel	44	32.8	24.1	3706.2	2	1853.1	2.9	0.060
	Bern	38	42.5	28.4					
	Zürich	98	43.7	24.8					

Unter den zugelassenen Personen treten zwischen den Wunschuniversitäten keine signifikanten Unterschiede auf (Tabelle 19).

**Tabelle 19: Test-Prozentränge und Ergebnisse der Varianzanalyse pro Disziplin nach Wunschhochschulen, nur zugelassene\*, deutschsprachige Personen.**

	Wunschhochschule	n	Mittelwert	Standardabw.	Quadratsumme	Df	MQ	F	Sig.
HM	Basel	213	83.7	9.7	946.2	6	157.7	1.8	.103
	Bern	245	83.4	9.1					
	ETH	118	86.1	9.6					
	Freiburg	15	82.2	9.3					
	Zürich	434	84.9	9.4					
	Zürich (LU)	29	83.2	9.6					
	Zürich (SG)	22	83.5	10.3					
VM	Bern	50	68.2	13.1	657.1	1	657.1	3.3	0.073
	Zürich	75	72.8	14.9					
ZM	Basel	23	51.5	18.9	1406.4	2	703.2	2.1	0.132
	Bern	22	62.6	20.0					
	Zürich	64	58.1	17.8					

\*Zulassung nach erstem, provisorischem Zulassungsmodell (ohne Nachrücker)

## 8 Ergebnisse zur Testgüte

### 8.1 Zuverlässigkeit

Von besonderem Interesse sind 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 Gesamtest kann anhand zweier Koeffizienten berechnet werden. Die Reliabilitätsschätzung nach der Testhalbierungsmethode (Teilung nach gerad- und ungeradzahligem Aufgaben) ist eine der gebräuchlichsten Zuverlässigkeitsschätzungen. Wegen der Stichprobengrößen werden nur die Ergebnisse des deutschsprachigen Tests der Schweiz verwendet.

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 2022 mit .92 wieder einen sehr hohen Wert, das Testprofil mit 0.85 eine sehr gute Konsistenz. Die Zuverlässigkeit des Punktwertes liegt damit im üblichen Bereich des Originaltests, die Konsistenz des Testprofils sogar im oberen Bereich. Der Bereich um 0.80 scheint deshalb optimal, weil vergleichbare Werte auch in den Jahren mit einer erfolgreichen Evaluation des Zusammenhangs von Studienerfolg und Eignungstest gefunden worden sind.

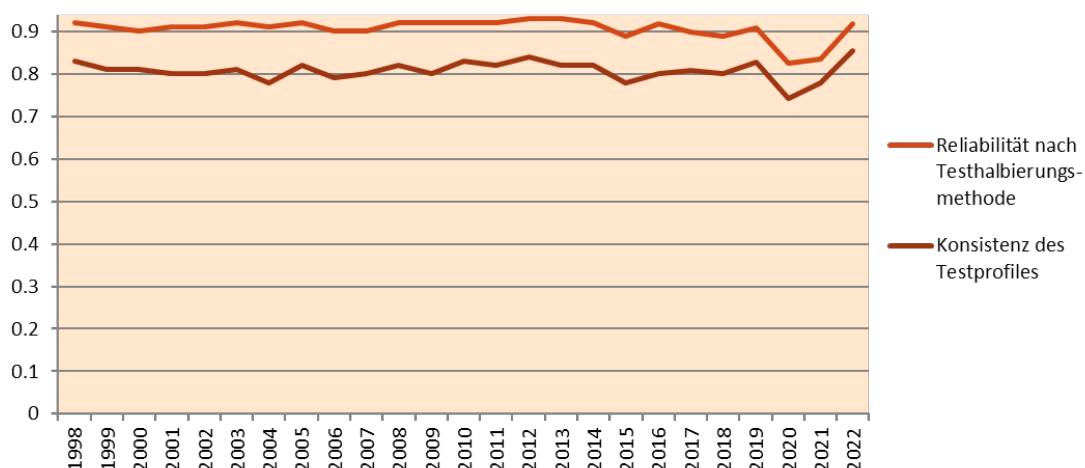


Abbildung 17: Reliabilität des Punktwertes (Split Half) und Konsistenz des Testprofils 1998-2022.

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. 2022 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.

## 8.2 Binnenstruktur

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

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

	Med.-naturwis. Grundv.	Schlauchfiguren	Quant. u. formale Probl.	Textverständnis	Figuren lernen	Fakten lernen	Diagr. und Tabellen	Konzent. u. sorgf. Arbeiten	Punkt-wert
Muster zuordnen	0.37	0.57	0.33	0.32	0.43	0.43	0.32	0.43	0.69
Med.-naturw. Grundverst.		0.42	0.53	0.59	0.33	0.39	0.53	0.31	0.72
Schlauchfiguren			0.40	0.34	0.44	0.47	0.39	0.40	0.73
Quant. u. form. Probleme				0.48	0.29	0.32	0.55	0.33	0.68
Textverständnis					0.29	0.36	0.51	0.30	0.68
Figuren lernen						0.46	0.29	0.35	0.64
Fakten lernen							0.34	0.34	0.68
Diagramme und Tabellen								0.33	0.68
Konzent. u. sorgf. Arbeiten									0.61

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

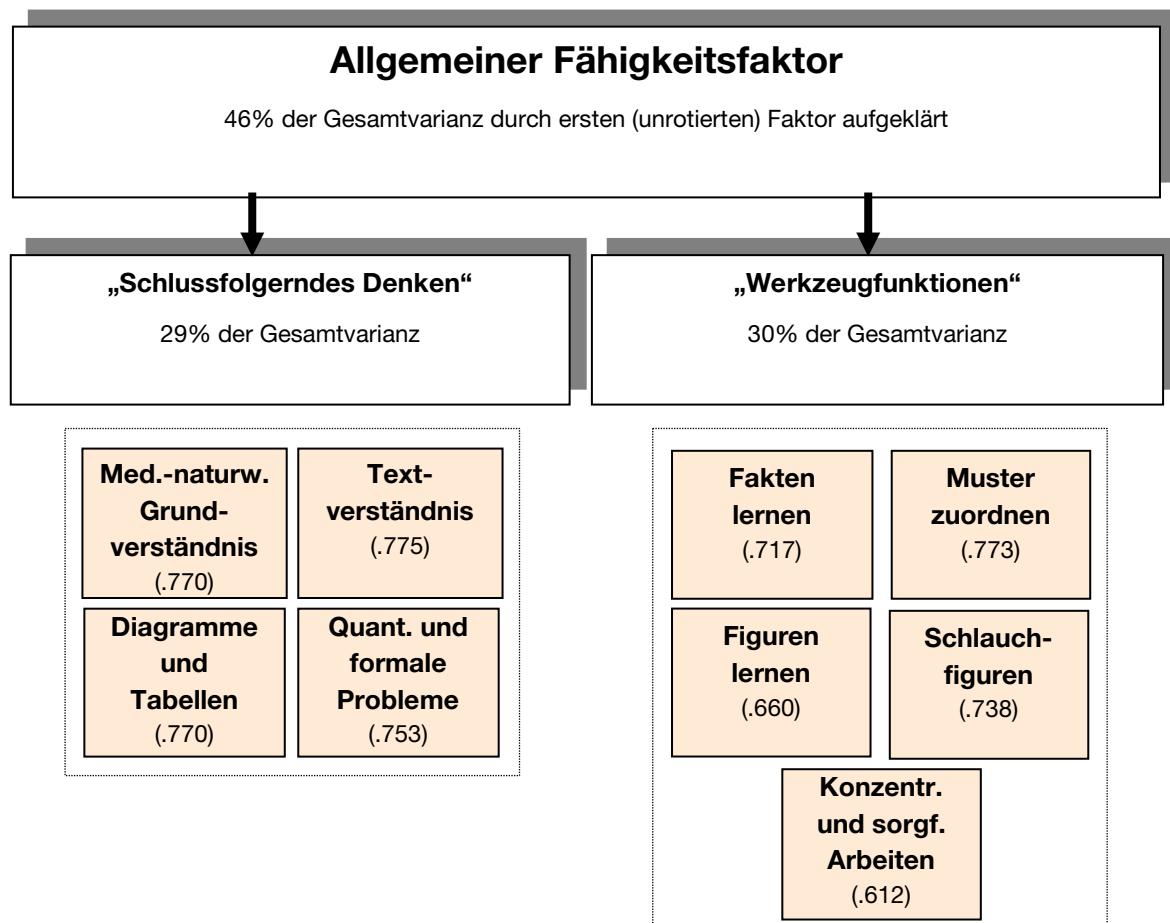
**Tabelle 21: Faktorenanalyse: Verlauf der unrotierten Lösung (2018-2022).**

Faktor	Unrotierte Lösung														
	Eigenwert					% Varianz					% Kumuliert				
	2018	2019	2020	2021	2022	2018	2019	2020	2021	2022	2018	2019	2020	2021	2022
1	3.58	3.84	2.73	2.95	4.17	39.72	42.66	45.52	49.17	46.37	39.72	42.66	45.52	49.17	46.37
2	1.12	1.40	0.94	0.97	1.14	12.43	15.51	15.63	16.09	12.68	52.15	58.17	61.15	65.26	59.06
3	0.82	0.71	0.71	0.60	0.72	9.14	7.93	11.86	9.94	8.01	61.29	66.10	73.01	75.20	67.06
4	0.71	0.63	0.67	0.57	0.61	7.92	7.05	11.17	9.55	6.74	69.20	73.16	84.17	84.75	73.81
5	0.67	0.56	0.52	0.48	0.57	7.41	6.27	8.60	7.97	6.34	76.61	79.43	92.77	92.72	80.14
6	0.58	0.52	0.43	0.44	0.53	6.41	5.75	7.23	7.28	5.89	83.02	85.17	100	100.00	86.03
7	0.56	0.50			0.44	6.23	5.56			4.91	89.25	90.73			90.94
8	0.53	0.43			0.42	5.86	4.82			4.63	95.10	95.55			95.57
9	0.44	0.40			0.40	4.90	4.45			4.43	100	100			100.00

**Tabelle 22: Varianzanteile der 2-Faktorenlösung 2018-2022 (rotiert).**

Varimax-rotierte Lösungen														
Eigenwert					% Varianz					% Kumuliert				
2018	2019	2020	2021	2022	2018	2019	2020	2021	2022	2018	2019	2020	2021	2022
2.4	2.7	2.3	2.4	2.7	27.0	29.8	37.7	40.7	30.0	27.0	29.8	37.7	40.7	30.0
2.3	2.6	1.4	1.5	2.6	25.2	28.3	23.5	24.6	29.1	52.1	58.2	61.2	65.3	59.1

Ein allgemeiner Fähigkeitsfaktor klärt 46% der Varianz auf – es gibt einen grossen „gemeinsamen Nenner“ in allen Aufgabengruppen. Mit der Zwei-Faktorenlösung können 60% der Varianz aufgeklärt werden. Die Aufgabe „Konzentriertes und sorgfältiges Arbeiten“ ist 2022 dem Faktor „Werkzeugfunktionen“ zuzuordnen. Die aktuellen Werte entsprechen durch den wiederhergestellten Faktor „Werkzeugfunktionen“ den Ergebnissen vor den Corona-Jahren bis 2019.



**Abbildung 18: Struktur der Aufgabengruppen des EMS, hierarchische Faktorenstruktur 2022 mit Faktorladungen der Aufgabengruppen auf rotierten Faktoren.**

Insgesamt hatte die neue Teststruktur keine nachteiligen Auswirkungen auf die Testgüte, wohl aber die erwarteten positiven Auswirkungen auf die Testorganisation, indem die Testabnahme ohne Mittagspause ermöglicht wurde. Im Zuge der Neuorganisation einiger Abläufe wurden weitere Optimierungen vorgenommen. Ebenso werden neue Erfahrungen und weitere Optimierungsmöglichkeiten hinsichtlich der Abläufe in die zukünftigen Testabnahmen einfließen.

## 9 Literatur

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In Holland, P.W., Wainer, H. (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Erlbaum.

Beller, M., Gafni, N. (1995). Translated Versions of Israel's interuniversity Psychometric Entrance Test (PET). In T. Oakland & R.K. Hambleton (Eds.), *International Perspectives of Academic Assessment*, S.207-218. Boston: Kluwer.

Beller, M. (1996). Translating, equating and validating Scholastic Aptitude Tests: The Israeli Case. In Hänsgen, K.-D., Hofer, R. & Ruefli, D. (Hrsg). *Eignungsdiagnostik und Medizinstudium*, (S. 14-29), Bericht 2. Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (Second Edition)*. Hillsdale, NJ: Erlbaum.

Cook, L.L. (1998). *Can Scores Obtained on Test Given in Different Languages to Examinees of Different Cultures be Equally Valid?* ICAP: San Francisco.

Deidesheimer Kreis (1997). *Hochschulzulassung und Studieneignungstests: studienfeld-bezogene Verfahren zur Feststellung der Eignung für Numerus clausus und andere Studiengänge*. Göttingen, Zürich: Vandenhoeck und Ruprecht.

Ebach, J., Trost, G. (1997). *Admission to Medical Schools in Europe*. Lengerich: Pabst.

Hänsgen, K.-D. (2014): Wollen wir nicht lieber diplomierte Ärzte? Sind fachliche und soziale Kompetenzen Gegensätze? *Schweizerische Ärztezeitung* 2014; 95: 6 S. 194 f.

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

Hänsgen, K.-D (2007): Numerus clausus in der Medizin – werden die Richtigen ausgewählt für Studium und Beruf? *Schweizerische Ärztezeitung | Bulletin des médecins suisses | Bollettino dei medici svizzeri* | 2007; 88: 46.

Hänsgen, K.-D. (2007): Wie bilden wir die richtigen Studierenden zu Ärzten aus? *Neue Zürcher Zeitung*, 18. Oktober 2007.

Hänsgen, K.-D., Spicher, B. (2002). Numerus clausus: Finden wir mit dem «Eignungstest für das Medizinstudium» die Geeigneten? *Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri* – 2002; 83(31): 1653 – 1660. <http://www.saez.ch/pdf/2002/2002-31/2002-31-842.PDF>

Hänsgen, K.-D., Spicher, B. (2002). Numerus clausus : Le « test d'aptitudes pour les études de médecine » (AMS) permet-il de trouver les personnes les plus aptes? *Schweizerische Ärztezeitung / Bulletin des médecins suisses / Bollettino dei medici svizzeri* – 2002; 83 (47) 2562 – 2569. <http://www.saez.ch/pdf/2002/2002-47/2002-47-1144.PDF>

Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 1: Erfahrungen mit dem EMS als Zulassungskriterium. *Schweizerische Ärztezeitung* Heft 12, S. 666 – 672.

Hänsgen, K.-D., Spicher, B. (2000). Zwei Jahre Numerus clausus und Eignungstest für das Medizinstudium in der Schweiz (EMS). Teil 2: EMS und Chancengleichheit. *Schweizerische Ärztezeitung* Heft 13, S. 723-730.



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

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

Longford, Holland & Thayer, (1993). Stability of the MH D-DIF Statistics Across Populations. In P. W. Holland & H. Wainer (Ed.) Differential Item Functioning, S. 171 – 196. Lawrence Erlbaum Associates: New Jersey.

Oswald, U. (1999). Der Eignungstest 1998 für das Medizinstudium. Schweizerische Ärztezeitung 80, S. 1313 – 1317.

Trost, G. (Hrsg.) (1994). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht). Bonn: ITB.

Trost, G. (Hrsg.) (1995). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (19. Arbeitsbericht). Bonn: ITB.

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

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

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

## 9.1 Originaltest zur Information und Vorbereitung

Es existieren in allen drei Sprachen jeweils drei veröffentlichte Originalversionen.

Deutsch: [www.unifr.ch/ztd/ems/vord.htm](http://www.unifr.ch/ztd/ems/vord.htm)

Französisch: [www.unifr.ch/ztd/ems/vorf.htm](http://www.unifr.ch/ztd/ems/vorf.htm)

Italienisch: [www.unifr.ch/ztd/ems/vori.htm](http://www.unifr.ch/ztd/ems/vori.htm)

## 9.2 Frühere Berichte des ZTD

Über EMS-Seite: [www.unifr.ch/ztd](http://www.unifr.ch/ztd)

EMS-WIKI-Blog [www.ztd.ch](http://www.ztd.ch)  
Kommentare und Fakten