

**EMS • Eignungstest für das
Medizinstudium in der Schweiz**
Vorhersage des Prüfungserfolges
K.-D. Hänsgen und B. Spicher

Inhaltsverzeichnis

1	EINLEITUNG.....	3
2	EVALUATION: EIGNUNGSTEST UND PRÜFUNGSERFOLG	7
2.1	Prüfungsverlauf für Testkohorten 1998 und 1999	9
2.1.1	Erste Vorprüfung und Prüfungsverlauf	9
2.1.2	Zweite Vorprüfung und Prüfungsverlauf	13
2.2	Prüfungserfolg und Testwert.....	14
2.2.1	Erste Vorprüfung und Prüfungserfolg gesamt.....	15
2.2.2	Effekt der Zulassung von der Warteliste	22
2.2.3	Erfolgsvorhersage für verschiedene Universitäten.....	23
2.2.4	Erfolgsvorhersage für Geschlechter	25
2.2.5	Erfolgsvorhersage für Alter und Maturitätsjahr	26
2.2.6	Erfolgsvorhersage für Sprachgruppen.....	27
2.2.7	Unterschiede zwischen Human- und Veterinärmedizin	28
2.2.8	Zweite Vorprüfung	31
2.3	Testwert und Prüfungsnoten (Erste Vorprüfung).....	33
2.4	Schlussfolgerungen zur Eignungsfeststellung.....	40

aus: Hänsgen, Klaus-Dieter; Spicher, Benjamin (2001).

EMS Eignungstest für das Medizinstudium in der Schweiz 2001

Bericht – Vorhersage Prüfungserfolg – Schlussfolgerungen

Berichte des Zentrums für Testentwicklung, Nummer 7, im Auftrag der Schweizerischen Universitätskonferenz (SUK)

Freiburg/Schweiz: Zentrum für Testentwicklung und Diagnostik, 2001

Unter Mitarbeit von Katharina Stress (Lektorat),

Mauro Bernasconi, Steve Fürst, Sébastien Simonet, Maurizio Strazzeri

Redaktion: Tanja Läng

1 Einleitung

Ein Numerus Clausus (NC) wird für die Zulassung zum Studium der Medizin seit 1998 an den Universitäten Basel, Bern, Freiburg und Zürich angewendet. Jede Zulassungsbeschränkung erfordert Kriterien, nach denen die Auswahl der Personen erfolgt. Als Zulassungskriterium wird in der Schweiz ein Eignungstest (EMS) verwendet. Die recht intensiven politischen Diskussionen zum Zeitpunkt seiner Einführung werden nicht mehr geführt – NC und der Eignungstest sind in der öffentlichen Meinung faktisch eine „feste Einrichtung“ geworden. Volksentscheide in Zürich und Bern sind ohnehin deutlich pro NC ausgefallen – an der breiten öffentlichen Akzeptanz hat es nie gefehlt.

Die Entscheidung für einen NC ist eine politische Entscheidung, die unter Abwägung der verschiedenen Interessen getroffen wird. Die Schweiz gehört weiterhin zu den Ländern mit der grössten Ärztedichte. Die Gesundheitskosten wachsen und lassen sich nur schwer eindämmen – erst jüngst sind die Krankenkassen-Beiträge wieder drastisch gestiegen. Von dieser Seite gerät der NC nun unter Druck, auch berufspolitisch zu wirken und für eine Eindämmung der absehbaren und befürchteten „Ärztenschwemme“ zu sorgen. Es scheint daher wieder an der Zeit, darauf hinzuweisen, was seine Grundlagen sind – und was nicht.

Die Interkantonale Universitätsvereinbarung (IUV) vom 20.2.1997 fordert im Artikel 10 eine Gleichbehandlung bei Zulassungsbeschränkungen. Die gesetzlichen Grundlagen dafür sind auf der Ebene der Kantone geregelt und sagen aus, dass die Zulassung beschränkt werden kann, wenn Ausbildungs-Kapazitäten nicht ausreichen und keine andere Lösung gefunden werden kann. Es handelt sich also um eine bildungspolitische und keine berufspolitische Massnahme. Rückwirkungen auf den Arbeitsmarkt erfolgen allenfalls indirekt über eine langfristige Anpassung der Ausbildungskapazitäten an den Bedarf. Auch das Interesse für den Beruf spielt eine Rolle. Dieses wird beeinflusst durch die erlebten Aussichten, auf dem Gebiet der Medizin eine zukunftsichere und befriedigende Tätigkeit zu finden. Eine zu grosse „Dissuasion“ durch schlechtere Aussichten und Perspektiven würde ebenso zu Fehlregulationen führen wie eine zu grosse Idealisierung des Medizinerberufes. Ausserdem könnte eine Dissuasion die Geeigneten stärker abschrecken – etwa, wenn diese auch offener für andere Lebensperspektiven sind.

Der NC ist durch Entscheide des Bundesgerichts gestützt¹ und wird grundsätzlich gutgeheissen – aber gleichzeitig werden zwei Forderungen an das Zulassungsverfahren gestellt, die ausserordentlich bedeutsam scheinen: Es müssen Rechtsgleichheit und Willkürfreiheit gewährleistet sein². Die Gleichbehandlung wird durch diese beiden Begriffe also weiter differenziert und in der Sache bestätigt.

Die Kantone sind zuständig für die gesetzlichen Grundlagen des Numerus Clausus. In diesen gesetzlichen Grundlagen für den NC der Kantone ist geregelt, dass bei der Zulassung von der Eignung auszugehen ist. Der Eignungsbegriff wird nicht weiter differenziert, um offen für das Einbeziehen aktueller Forschungsergebnisse oder internationaler Erfahrungen zu sein – schliesslich ist die Schweiz nicht das einzige Land, in welchem

¹ Urteile des Schweizerischen Bundesgerichtes 2P.171/1998/mks vom 3. März 1999 sowie 2P.228/1998 – 2P.332/1998/luc vom 20. August 1999.

² "Wie alle staatlichen Leistungen sind auch Studienplätze zwangsläufig ein beschränkt vorhandenes Gut. Verfassungsrechtlich besteht ein Anspruch auf eine willkürfreie und rechtsgleiche Regelung bei der Zulassung zu den vorhandenen Studienplätzen". Urteil vom 3. März 1999 S. 7.

Zulassungsbegrenzungen notwendig sind. Sie findet sich diesbezüglich in einer Reihe mit den meisten Industriestaaten wieder.

Auch in den bisher bekannten Entwürfen zum neuen Medizinalberufe-Gesetz ist der Eignungsbegriff weiterhin in der Diskussion als Zulassungskriterium. Obgleich dort der Versuch unternommen wird, den Eignungsbegriff weiter zu differenzieren (etwa als „psychosoziale Eignung“¹), bleiben die gewählten Modelle grundsätzlich gültig.

Dass Eignung nicht weiter definiert ist, birgt allerdings auch Risiken für die Diskussion. Es werden alle denkbaren Wunschvorstellungen über „den guten Arzt/die gute Ärztin“ als Eignungskriterien genannt und damit als Zulassungskriterien zum Studium vorgeschlagen. Dies geschieht ohne Rücksicht darauf,

- (1) ob diese Kriterien bereits vor Beginn des Studiums vergleichbar für die verschiedenen Bewerbergruppen ausgeprägt sind und eingeschätzt werden können (oder beispielsweise Personen mit Berufserfahrung Vorteile hätten - was gegen die Gleichbehandlung sprechen würde);
- (2) ob die Vorhersagbarkeit der Entwicklung dieser Kriterien über einen so langen Zeitraum wie das Studium überhaupt gegeben ist (oder diese Entwicklung vor allem während des Studiums erst in Gang gesetzt wird);
- (3) ob überhaupt Methoden existieren, welche diese Kriterien mit ausreichender Genauigkeit messen können, was für die Rechtsfähigkeit der Zulassung eine unbedingte Voraussetzung ist (oder dass durch Training jede/r sehr leicht einen „guten Eindruck“ hinterlassen kann, ohne dass der sachlich gerechtfertigt wäre);
- (4) ob das Zulassungsverfahren auf der Basis dieser Kriterien in zeitlicher und finanzieller Hinsicht noch praktikabel ist.

Betrachten wir eine psychologische Definition von Eignung, so ist die „*hinreichende Übereinstimmung von individuellen Leistungsvoraussetzungen und aufgabenabhängigen Anforderungen bei konkreten Tätigkeiten*“² der gemeinsame Nenner aller Definitionen. Will man rechtsgleich und willkürfrei Personen nach unterschiedlicher Eignung auswählen, müssen die Anforderungen der Tätigkeiten und vor allem Erfolgskriterien für die Erfüllung der Anforderungen ausreichend definiert sein. Je globaler und undifferenzierter die Anforderungen sind (was bei der Definition „guter Ärzte“ nicht ausbliebe), umso ungenauer wird man Personen nach der erfolgreichen Erfüllung dieser Anforderung (was Eignung bestimmt) differenzieren können. Umso weniger begründet würde die Zulassung sein.

Eignung sollte auch nicht mit Neigung verwechselt werden. Der Begriff der Neigung wird eher mit dem Interesse, der Motiviertheit für den Medizinalberuf in Verbindung gebracht. Es wäre ideal, wenn die Informationen vor der Studienbewerbung so weit ausreichen würden, dass realistische Erwartungen in Bezug auf das Studium und die spätere berufliche Tätigkeit bestehen. Hier haben Universitäten und auch die Praxis gute Möglichkeiten, entsprechende Informationen zu liefern – bis hin zu „Schnupperpraktika“, wo sich die Bewerber und Bewerberinnen direkt vor Ort informieren und ihren Berufswunsch festigen, ihre Neigung zum Medizinalberuf erkennen und prüfen können. Es wäre allerdings vermessen, in solchen Praktika die Eignung zum Bestehen der Studienanforderungen oder

¹ „Psychosoziale Eignung“ meint alle psychischen und sozialen Prozesse und ihre Wechselwirkungen und ist daher als Begriff so global, wie wenn man nur von „Eignung“ spricht.

² Schorr, A. (Hrsg.) (1993). Handwörterbuch der Angewandten Psychologie. Stichwort „Eignungsdiagnostik“ (H.J. Liebel). Bonn: Deutscher Psychologen-Verlag.

generell für einen Medizinalberuf abzuschätzen. Interesse ist – wie in vielen anderen Bereichen, nicht mit Fähigkeit bzw. Eignung gleichzusetzen.

In den jetzt vorliegenden und nachfolgend dargestellten Ergebnissen zur Prognose der Studienergebnisse durch den Eignungstest kann schlüssig nachgewiesen werden, dass der EMS sowohl Studiendauer als auch Studienqualität sehr gut vorhersagen kann – beides sind wichtige Erfolgskriterien und damit Determinanten der Eignung. Es werden diejenigen bevorzugt zugelassen, die ihr Studium in optimaler Zeit mit besseren Leistungen abschliessen (bisher nachgewiesen bis zur 2. Vorprüfung).

Die Eignungsabklärung muss langfristig nicht auf den Leistungsaspekt wie im vorliegenden EMS begrenzt sein. Diese Begrenzung ist entstanden, weil die Leistungserfassung bisher die einzige Methode ist, welche alle Kriterien der Wissenschaftlichkeit und Ökonomie erfüllt. In diesem Zusammenhang wird häufig bemängelt, dass Merkmale der sozialen Kompetenz nicht als Eignungskriterien berücksichtigt werden.

Der wohl wichtigste Grund ist das Fehlen geeigneter Methoden. Lässt man soziales Verhalten in Situationen beobachten (etwa in Eignungsgesprächen), müssen die Beurteiler auf einen einheitlichen Masstab hin geschult werden. Ausserdem sind für jede Beurteilung mindestens 2-3 Urteiler notwendig, damit die Entscheidungen rechtsfähig sind. Das ist für alle Bewerberinnen und Bewerber um ein Medizinstudium nicht bezahlbar. Ein weiterer Nachteil ergibt sich aus der hochgradigen Trainierbarkeit von Sozialverhalten. Es gibt auf dem Gebiet der sogenannten Assessment-Center (einer verbreiteten Methode der Eignungsdiagnostik) einen „boomenden“ Trainingsmarkt, wo die Bewerberinnen und Bewerber entsprechend vorbereitet werden – zum Teil zu beträchtlichen Kosten. Das kann für die Bewerbung zum Medizinstudium nicht anstrebenswert sein.

Objektivere, ökonomischere und weniger trainierbare Methoden zur Erfassung von Aspekten sozialer Kompetenz als Bestandteile zukünftiger EMS-Versionen sind allerdings nicht gänzlich ausgeschlossen. Hier muss die Grundlagenforschung – beispielsweise zur sozialen Wahrnehmung – noch wichtige Vorleistungen erbringen. Dies sollte in Verbindung mit der Analyse der neuen Anforderungen an das Medizinstudium, die sich aus den Reformmodellen der einzelnen Universitäten ergeben, geschehen. Wichtig wäre, entsprechende Forschungen rechtzeitig anzuregen. Auch diese neuen Methoden müssen genauso intensiv entwickelt und überprüft werden wie die jetzt verwendeten Testteile.

Es ist legitim, auch andere Zulassungskriterien daraufhin zu überprüfen, ob sie die notwendigen Anforderungen erfüllen und dabei bezahlbar bleiben. Ein Teil der Wunschvorstellungen, Eignung auch auf Aspekte der beruflichen Eignung auszudehnen, wird an zu hohen Kosten oder fehlender Erfassbarkeit durch Methoden sowie mangelnder Rechtsgleichheit bzw. Willkürfreiheit scheitern. In der Tabelle 1 wurde der Versuch unternommen, die einzelnen Kriterien, die immer wieder in die Diskussion eingebracht werden, vergleichend zu bewerten. Letztlich wird deutlich, dass der jetzige EMS gegenüber allen anderen Zulassungskriterien in jedem einzelnen Bewertungsaspekt mit zu den Besten gehört. Die übrigen Kriterien haben Nachteile, welche die Praktikabilität einschränken.

	Eignungstest	Maturitätsnote	Aufnahmegespräche	Sozialpraktika	Intrauniversitäre Selektion
Erfasst Eignung?	Ja, wissenschaftlich nachgewiesene Vorhersage von Studiendauer und Studienleistungen	Möglich, nicht nachgewiesen (Benotungsunterschiede werden sich hier begrenzend auswirken)	Möglich, wenn standardisierte Beurteilung, mehrere Beurteiler und Training dieser Beurteiler	Nein, dient mehr dem Erkennen der Neigung. Betreuer würden durch Notwendigkeit der Eignungseinschätzung überfordert – Anforderung wie Aufnahmegespräche!	Ist nicht dem Test überlegen!
Kapazitätsentsprechende Zulassung?	Ja	Ja	Ja	Nein alle erfolgreichen Praktikanten wollen dann auch studieren	Weiter Überlastung der Kapazität im 1. Jahr, weiter grössere Abbrecherzahl nach 1 Jahr; Prüfung ist versteckter NC.
Rechtsgleichheit	Ja	Nein kantonale, schul- und lehrerbedingte Benotungsunterschiede	Nur bei Training der Beurteiler und Standardisierung der Bedingungen (aufwändig)	Nur wenn wirklich gleiche Praktikums- und Beurteilungbedingungen bestehen würden	Ja
Willkürfreiheit	Ja				Wenn Prüfungen vergleichbar sind
Ökonomie	Ja	„anfallend“	Nicht für alle realisierbar (dann zu teuer)	Nein (in Zürich erwies es sich als nicht durchführbar wg. fehlender Plätze)	Hohe Kosten im 1. Jahr
Negative Folgen?	Nein	Rückwirkung auf Benotungssystem der Maturität	„Trainingsmarkt“ würde wachsen	Geschultes Betreuungspersonal nötig (Kosten)	Versteckter NC

Tabelle 1: Mögliche Zulassungskriterien und deren Bewertung

Zusammenfassend kann man festhalten:

- Wenn eine Zulassung nach Eignung erfolgen soll, muss das „WOFÜR geeignet“ klar sein (also die Anforderungen müssen bekannt sein).
- Die Berufseignung ist bezüglich ihrer Anforderungen wenig eindeutig – es gibt auch verschiedene Disziplinen und Tätigkeiten innerhalb der Medizin, die sich stark unterscheiden. Die Beschränkung auf Studieneignung ist sinnvoll, denn...
- ... der Erfolg bei der Erfüllung der Anforderungen muss klar messbar sein (bei Studieneignung etwa die Studiendauer und die Studienleistungen). Berufserfolg ist kaum ausreichend definierbar.
- Es müssen Zulassungskriterien vorhanden sein, welche diesen Erfolg vorhersagbar machen und damit eine Unterscheidung der Personen nach der Eignung zulassen.
- Das Zulassungskriterium ist nur dann gerechtfertigt, wenn wissenschaftlich nachweisbar ist, dass es dafür sorgt, dass die Geeigneteren eine höhere Chance erhalten, zum Studium zugelassen zu werden.

2 Evaluation: Eignungstest und Prüfungserfolg

Wenn auf politischer Ebene der Entscheid eines Numerus Clausus getroffen wird, muss ein **Zulassungskriterium** den Zugang zum Studium regeln. Ablehnungen müssen, wenn sie unvermeidbar sind, ausreichend begründet sein. Ein solches Zulassungskriterium muss Chancengleichheit gewährleisten, objektiv, fair und wissenschaftlich fundiert sein.

„**Eignung**“, welche entsprechend der gesetzlichen Grundlagen der Kantone als Zulassungskriterium dienen soll, muss in objektive, messbare und überprüfbare Erfolgskriterien umgesetzt werden. Eine Person ist dann geeigneter, wenn sie diese Erfolgskriterien besser erfüllt. Wenn sich der Numerus Clausus als **bildungspolitische** Massnahme begreift (und nicht als berufspolitisches Planungsinstrument den Arbeitsmarkt der Medizin regeln soll), muss die **Studieneignung** die Zulassung leiten.

Es bieten sich zwei **Erfolgskriterien** für Studieneignung an: Studienplätze bevorzugt an diejenigen zu vergeben, die

- (1) in möglichst kurzer bzw. **optimaler Zeit** das Studium absolvieren, sowie
- (2) das Studium in **guter Qualität** bzw. mit auf hohem Niveau erworbenen Fähigkeiten und Fertigkeiten absolvieren.

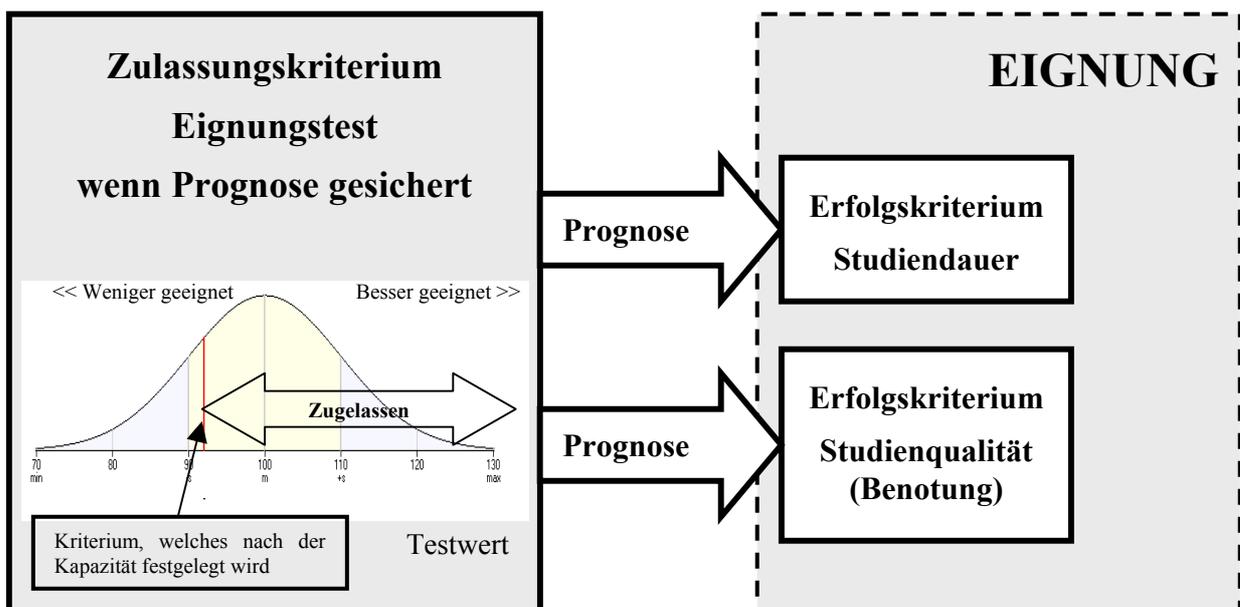


Abbildung 1: Zusammenhang Eignung – Erfolgskriterium – Zulassungskriterium

In der Abbildung 1 wird der Zusammenhang noch einmal verdeutlicht: Eignung wird anhand zweier Erfolgskriterien definiert (Studiendauer und -qualität). Indem nachgewiesen wird, dass das Zulassungskriterium diese beiden Erfolgskriterien vorhersagt (also Eignung tatsächlich misst), können Geeignetere tatsächlich bevorzugt zugelassen werden.

Durch Vermeidung einer verlängerten Studiendauer oder gar von Studienabbrüchen werden die begrenzt vorhandenen Ausbildungsressourcen nicht mehr als notwendig beansprucht und somit früher wieder für Nachfolgende frei. Im Medizinstudium ergeben sich Begrenzungen beispielsweise durch eine nicht beliebig erweiterbare Patientenzahl an den Universitätsspitalern und vergleichsweise teuer ausgestattete Übungs- und Praktikumsplätze. Man muss aber nicht nur diese „volkswirtschaftliche“ Perspektive sehen,

denn auch für die Studierenden besteht der Nutzen darin, dass weniger verlorene Lebenszeit auf dem Weg zum Beruf entsteht, wenn weniger Abbrüche nach einem oder sogar zwei Jahren auftreten (vor dem Numerus Clausus brechen etwa 1/3 der Personen ab!).

Gesucht ist daher ein Zulassungskriterium, welches Studiendauer und Studienqualität möglichst vor Antritt des Studiums optimal vorhersagbar macht.

Intrauniversitäre Selektionen sind in der Regel teuer und es bleiben in der Anfangszeit weiterhin Ausbildungs-Engpässe erhalten. Nachteilige Wirkungen auf die Persönlichkeitsentwicklung können erwartet werden, weil man mit den Mitstudierenden in dieser Zeit konkurrenziert, statt das für den späteren Beruf notwendige Kooperationsverhalten zu erlernen.

Der Eignungstest EMS soll diese Studieneignung vorhersagen. Wenn er im Vergleich zu anderen Kriterien ausserdem konkurrenzlos geringe Kosten, hohe Chancengleichheit und Fairness gewährleistet, ist seine Wahl gut begründet.

Eine wichtige Frage der Evaluation des EMS als Zulassungskriterium ist daher die Analyse der Beziehungen

- (1) zwischen Testleistung und dem Bestehen der Prüfungen mit direktem Einfluss auf die Studiendauer und
- (2) zwischen Testleistung und den Prüfungsnoten als Ausdruck der Studienqualität.

Bereits im Jahre 2000 wurde eine erste Evaluation der Testkohorte 1998 vorgelegt (Hänsgen, Spicher 2000), wo nach einem Jahr Studiendauer gute Vorhersagewerte ermittelt werden konnten.

Für die Analyse des Zusammenhanges von Testergebnis und Prüfungserfolg stehen in diesem Jahr die Kohorten der Testteilnahme 1998 und 1999 zur Verfügung. 1998 war der Numerus Clausus für Humanmedizin, 1999 für Human- und Veterinärmedizin an den Universitäten Basel, Bern, Freiburg und Zürich notwendig. In der Tabelle 2 wird dargestellt, wie viele Testteilnehmer jeweils ein Studium begonnen haben. Die Analyse von Prüfungsnoten wird allerdings dadurch erschwert, dass zwischen den Universitäten und auch innerhalb der Jahrgänge verschiedene Prüfungsarten und Notensysteme zur Anwendung kommen. Im Rahmen der Studienreform wird teilweise noch experimentiert und Noten werden durch Kreditpunkte abgelöst, die nur noch zwischen Bestehen und Nichtbestehen differenzieren.

		Studienbeginn Medizin			Bisher kein Studienbeginn Medizin*	Testteilnahme gesamt
		1998	1999	Gesamt		
Test-jahr	1998	597	3	600	150	750
	1999		666	666	189	855
Total		597	669	1266	339	1605

Tabelle 2: Test 1998 und 1999 und Studienbeginn; *: Ablehnungen und Rückzüge

Die 1998 Immatrikulierten haben bisher vier mögliche Prüfungstermine wahrnehmen können (Sommer und Herbst 1999, Sommer und Herbst 2000). Wurde die erste Vorprüfung 1999 bestanden, kann im Jahre 2000 bereits die zweite Vorprüfung abgelegt worden sein.

Eine zweijährige Studiendauer bis zur zweiten Vorprüfung wäre damit das mögliche Optimum. Die 1999 Immatrikulierten konnten bisher nur im Sommer oder Herbst 2000 die erste Vorprüfung ablegen.

Die Prüfungsdaten des Jahres 2000 wurden wiederum im Bundesamt für Gesundheit (BAG) erfasst¹. Zur Quantifizierung der Studiendauer nummerieren wir die möglichen Prüfungstermine für jeden Studierenden vom Studienbeginn an fortlaufend. Erste und zweite Prüfungssession liegen damit im Sommer bzw. Herbst nach dem ersten Studienjahr, dritte und vierte Prüfungssession im Sommer bzw. Herbst nach dem zweiten Studienjahr.

2.1 Prüfungsverlauf für Testkohorten 1998 und 1999

Die Studienzulassung erfolgt praktisch in mehreren Schritten. Nicht ausgenutzte Plätze werden in zwei Runden an Personen vergeben, die zunächst aufgrund des Testwertes nicht zugelassen werden können. Testteilnehmende, die sich nicht immatrikuliert haben, werden danach unterschieden, ob der Testwert unter- oder oberhalb des Testwertes liegt, welcher der genauen Ausnutzung der Kapazität entspricht. Von Interesse sind dabei nur die Ergebnisse der Personen, die sich nicht immatrikuliert haben, vom Testwert her aber klar über dem Kapazitäts-Kriterium liegen. Sind dies gehäuft Geeignete, weniger Geeignete oder ist dies bezüglich der Eignung gleichmässig verteilt? Nachfolgend werden einige Statistiken über den Studienverlauf dargestellt, die jetzt und vor allem zukünftig Aufschluss über veränderte Studienverläufe unter den Bedingungen eines NC geben.

2.1.1 Erste Vorprüfung und Prüfungsverlauf

Abbildung 2 verdeutlicht, dass die Verhältnisse der beiden Testkohorten nach jeweils einem Studienjahr vergleichbar sind. Innerhalb des NC-Zeitraumes treten bisher keine Veränderungen auf. Für die Zeit vor dem NC fehlen entsprechende Vergleichsdaten.

Es wäre damit zu rechnen, dass sich die Zahl der bestandenen Prüfungen in der ersten und zweiten Session langfristig erhöht. Dabei ist auch der vergleichsweise milde Numerus Clausus in der Schweiz zu beachten. Eine anfängliche Ablehnungsrate von 20% ist im internationalen Vergleich bereits recht niedrig. Rückzüge und die nachfolgende erneute Verteilung der Plätze an Personen mit niedrigeren Testwerten führen dazu, dass auch weniger Geeignete zum Studium zugelassen werden – dann zumeist nicht an ihrer Wunschuniversität. Mit anderen Worten: der NC führt nicht unbedingt und überall zu einer veränderten Zusammensetzung der Studierenden – seine Wirkung besteht vor allem in der kapazitätsentsprechenden Auslastung der Universitäten und dem Abbau der Überlastung.

Diese Überlastungen können im übrigen auch nicht sofort abgebaut werden – Repetenten aus den Vorjahren beanspruchen Kapazitäten noch 1 bis 2 Jahre länger, wenn die hier festgestellten Prüfungsverläufe auch für frühere Zeiten gelten. Hoffnung auf Erhöhung der

¹ Wir danken Herrn Megert, Frau Hirvi sowie Herrn Neuhaus vom BAG herzlich für die Zusammenarbeit. Das BAG erhielt die Immatrikulationsnummern der Studienanfänger von der Rektorenkonferenz der Schweizer Universitäten (CRUS). Prüfungserfolgsdaten wurden in elektronischer Form zur Verfügung gestellt, Noten von den Prüfungsbogen erfasst. Dabei waren EMS-Ergebnisse nicht bekannt. Erst durch die CRUS wurden die Prüfungsergebnisse mit den Testergebnissen verbunden – ein Original befindet sich zu Nachweiszwecken dort. Die Evaluationsberechnungen wurden dann am ZTD mit einer Kopie dieser Datenbank vorgenommen, was eine ausreichende Objektivität der Evaluation gewährleisten sollte.

Bestehensquoten besteht allerdings auch, weil die kapazitätsentsprechende Auslastung der Universitäten nunmehr auch Reformstudiengänge erlaubt, die zu einer neuen Studienqualität beitragen können.

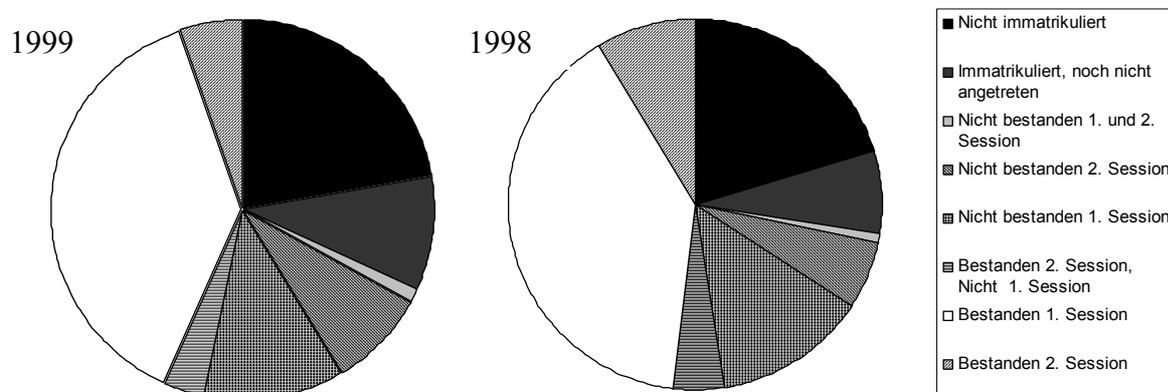


Abbildung 2: Vergleich der Testkohorten 1998 und 1999 nach jeweils einem Studienjahr

Die Tabelle 3 stellt das Prüfungsverhalten beider Testkohorten für die erste Vorprüfung differenziert zusammen. Für die Kohorte 1998 wurde der Stand 1999 noch einmal mit dargestellt, um einen Vergleich zur Kohorte 1999 nach ebenfalls einem Jahr Studium zu ermöglichen. Die Ergebnisse rechtfertigen eine gemeinsame Analyse beider Kohorten für das erste Jahr.

Eine Zusammenfassung des Verlaufes der ersten Vorprüfung ist nach verschiedenen Kriterien möglich. Die Tabelle 4 zeigt, dass die Zahl der Prüfungsanläufe nach jeweils einem Jahr in etwa vergleichbar ist – in der Testkohorte 1998 nach 2 Jahren mehr Personen im zweiten Anlauf bestanden oder zweimal nicht bestanden haben – also ein Jahr länger für ihr Studium benötigt haben. Noch deutlicher wird dies in Tabelle 5, wo die immatrikulierten Personen danach zusammengefasst werden, wie lange das Studium minimal bis zur ersten Vorprüfung dauert. Bei rund 14 Prozent der Kohorte 1998 ist sicher, dass mehr als ein Jahr Studienverlängerung bis zum Bestehen der ersten Vorprüfung notwendig ist (siehe dazu auch die Abbildung 3).

Auch nach 2 Jahren haben „nur“ 63% der Personen die erste Vorprüfung erfolgreich abgelegt – 52% waren es schon nach einem Jahr. Zweimal nicht bestanden haben 6% nach 2 Jahren – etwa 1/3 der Zugelassenen verbleiben nach 2 Jahren als „offen“, was die erste Vorprüfung betrifft. Es ist möglich, dass einige davon nicht mehr immatrikuliert sind, einige aber noch später antreten werden.

Prüfungsverlauf	Test 1998				Test 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht immatrikuliert, TW unter Kriterium	107	14.3%	107	14.3%	88	10.3%
Nicht immatrikuliert, TW über Kriterium	43	5.7%	46	6.1%	101	11.8%
Immatrikuliert, noch nicht angetreten	33	4.4%	54	7.2%	83	9.7%
Nicht bestanden 3. und 4. Session	2	0.3%	Zeitlich nicht möglich			
Nicht bestanden 2. und 4. Session	2	0.3%				
Nicht bestanden 2. und 3. Session	13	1.7%				
Nicht bestanden 1. und 4. Session	5	0.7%				
Nicht bestanden 1. und 3. Session	18	2.4%				
Nicht bestanden 1. und 2. Session	6	0.8%	6	0.8%	11	1.3%
Nicht bestanden 4. Session	2	0.3%	Zeitlich nicht möglich			
Nicht bestanden 3. Session	7	0.9%				
Nicht bestanden 2. Session	14	1.9%	44	5.9%	71	8.3%
Nicht bestanden 1. Session	24	3.2%	100	13.3%	101	11.8%
Bestanden 2. Session, Nicht 1. Session	32	4.3%	32	4.3%	30	3.5%
Bestanden 3. Session, Nicht 1. Session	43	5.7%	Zeitlich nicht möglich			
Bestanden 4. Session, Nicht 1. Session	9	1.2%				
Bestanden 3. Session, Nicht 2. Session	15	2.0%				
Bestanden 4. Session, Nicht 2. Session	1	0.1%				
Bestanden 4. Session, Nicht 3. Session	1	0.1%				
Bestanden 1. Session	302	40.3%	297	39.6%	325	38.0%
Bestanden 2. Session	64	8.5%	64	8.5%	45	5.3%
Bestanden 3. Session	6	0.8%	Zeitlich nicht möglich			
Bestanden 4. Session	1	0.1%				
Total	750		750		855	

Tabelle 3: 1. Vorprüfung für Testkohorte 1998 (Stand 1999 und 2000) sowie Testkohorte 1999 (Stand 2000). Sessionen wurden pro Studierenden vom Studienbeginn an durchnummeriert (siehe Text)

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht immatrikuliert, TW unter Kriterium	107	14.3%	107	14.3%	88	10.3%
Nicht immatrikuliert, TW über Kriterium	43	5.7%	46	6.1%	101	11.8%
Immatrikuliert, noch nicht angetreten	33	4.4%	54	7.2%	83	9.7%
Zweimal Nicht bestanden	46	6.1%	6	0.8%	11	1.3%
Einmal Nicht bestanden	47	6.3%	144	19.2%	172	20.1%
Zweiter Anlauf Bestanden	101	13.5%	32	4.3%	30	3.5%
Sofort Bestanden	373	49.7%	361	48.1%	370	43.3%
Total	750	100.0%	750	100.0%	855	100.0%

Tabelle 4: Zusammenfassung 1. Vorprüfung nach der Zahl der Prüfungsanläufe

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Studienverlängerung mehr als 1 Jahr	82	13.67%				
Studienverlängerung mindestens 1 Jahr	120	20.00%	204	34.17%	266	39.94%
„minimale“ Studienzzeit	398	66.33%	393	65.83%	400	60.06%
Total	600		597		666	

Tabelle 5: Zusammenfassung 1. Vorprüfung nach dem Einfluss auf die Studienzzeit

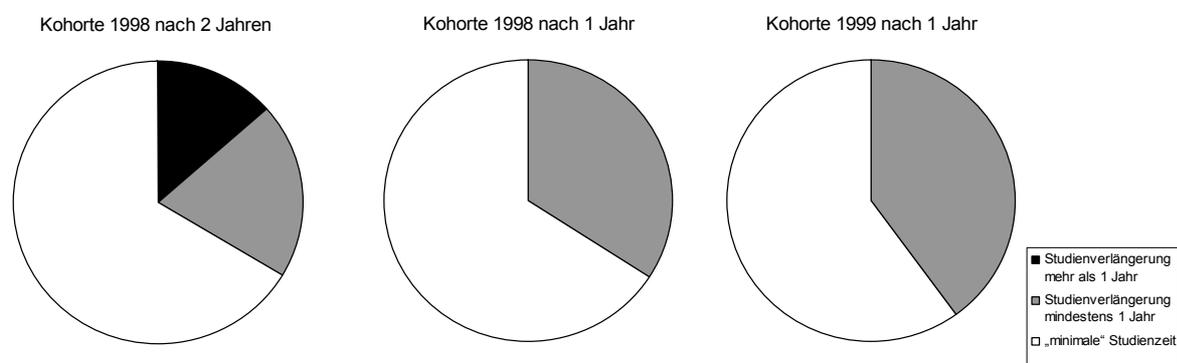


Abbildung 3: Vergleich der Test-Kohorten nach dem Einfluss auf die Studienzzeit

In der Tabelle 6 werden die Personen nach dem Prüfungserfolg zusammengefasst. Wenn zwei Prüfungsanläufe unternommen wurden, zählt das jeweils letzte Resultat. Im Herbst 2000 haben geringfügig weniger Personen die erste Vorprüfung bestanden als im Herbst

1999. Zu beachten ist, dass in der zweiten Test-Kohorte Veterinärmedizin mit enthalten ist – 1998 bezog sich der NC nur auf Humanmedizin. Später wird gezeigt, dass die Bestehensquote in Veterinärmedizin geringer ist als in Humanmedizin (siehe Seite 28). Hier könnte sich allerdings auch andeuten, dass die erste Vorprüfung vom Charakter her immer noch als ein „versteckter“ Numerus Clausus aufgefasst wird und eine auf die Verteilung der Leistungen orientierte Benotungsstrategie vorherrscht. Durch eine Homogenisierung des Leistungsniveaus als Folge des Numerus Clausus – oder auch als Folge der verbesserten Studienbedingungen (weil die Kapazitäten nicht mehr so überschritten sind) – wird die Differenzierungsfähigkeit zwischen Personen geringer und das gewählte Bestehenskriterium für Prüfungen lässt bei strengen Maßstäben mehr Personen die Prüfung nicht bestehen.

Prüfungsverlauf	Testkohorte 1998				Testkohorte 1999	
	Stand Herbst 2000		Stand Herbst 1999		Stand Herbst 2000	
Nicht bestanden	93	16.4%	150	27.6%	183	31.4%
Bestanden	474	83.6%	393	72.4%	400	68.6%
Total	567	100.0%	543	100.0%	583	100.0%

Tabelle 6: Zusammenfassung 1. Vorprüfung nach dem Erfolg

2.1.2 Zweite Vorprüfung und Prüfungsverlauf

Die zweite Vorprüfung kann nur von Personen der Testkohorte 1998 abgelegt worden sein – entweder im Sommer oder Herbst 2000; die Tabelle 7 fasst den Prüfungsverlauf differenziert zusammen.

Prüfungsverlauf	Testkohorte 1998	
Nicht immatrikuliert, unter Zulassungskriterium	107	14.3%
Nicht immatrikuliert, über Zulassungskriterium	43	5.7%
Wegen Nichtbestehen 1. Prüfung nicht möglich	126	16.8%
Zeitlich unmöglich (1. Vorprüfung zu spät bestanden)	79	10.5%
Noch nicht angetreten (aber möglich)	49	6.5%
Einmal Nicht bestanden 4. Session	41	5.5%
Einmal Nicht bestanden 3. Session	11	1.5%
Bestanden 3. Session	110	14.7%
Bestanden 4. Session	182	24.3%
Bestanden 4. Session, Nicht 3. Session	2	0.3%
Total	750	

Tabelle 7: 2. Vorprüfung für Testkohorte 1998, Sessionen vom Beginn des Studiums an fortlaufend durchnummeriert

346 Personen sind zur 2. Vorprüfung angetreten – lediglich 15% haben sie nicht bestanden (Tabelle 8). Weitere 49 Personen hätten zu dieser Prüfung antreten können, haben dies aber nicht getan. Diese Prüfung hat damit nicht mehr den Charakter eines „versteckten“ NC. Bisher konnten nur Personen die 2. Vorprüfung ablegen, welche die erste Vorprüfung in minimal möglicher Zeit bestanden haben (die also ohnehin die geeignetsten wären). Es ist hier also noch kein repräsentativer Leistungsquerschnitt vorhanden.

Prüfungsverlauf	Testkohorte 1998	
Nicht bestanden	52	15.0%
Bestanden	294	85.0%
Total	346	100.0%

Tabelle 8: Antritte 2. Vorprüfung nach Bestehen

Der Zusammenhang zwischen erster und zweiter Vorprüfung wird in der Tabelle 9 dargestellt.

Testkohorte 1998	2. Vorprüfung							Total
	Zeitlich nicht möglich	Noch nicht angetreten	1 x Nicht bestanden 4. Session	1 x Nicht bestanden 3. Session	Bestanden 3. Session	Bestanden 4. Session	Bestanden 4. Session, Nicht bestanden 3. Session	
Bestanden 1. Session	3*	25	17	10	<u>105</u>	<u>40</u>	2	302
Bestanden 2. Session		21	10		<u>1</u>	<u>32</u>		64
Bestanden 3. - 4. Session	7							7
Bestanden 2. Session, Nicht bestanden 1. Session		3	14	1	<u>4</u>	<u>10</u>		32
Bestanden 3. - 4. Session, Nicht bestanden 1. - 3. Session	69							69
	79	49	41	11	<u>110</u>	<u>182</u>	2	474

Tabelle 9: 2. Vorprüfung für Personen der Testkohorte 1998, welche die 1. Vorprüfung bestanden haben. *: Studienbeginn erst 1999. Unterstrichen: minimale Studiendauer

2.2 Prüfungserfolg und Testwert

Die Vorhersage des Prüfungserfolges aufgrund des Testwertes wird als wichtigstes Evaluationskriterium betrachtet. Dies hat vor allem Einfluss auf die Studiendauer. Bekanntlich ist der EMS ein Test, der auf die Vorhersage der Studierfähigkeit hin ausgerichtet wurde. Er ist kein Wissenstest – es wird vielmehr geprüft, inwieweit die Erarbeitung von Wissen (z.B. aus Texten), das Erkennen und die Anwendung logischer Regeln oder die Konzentrations- und Merkfähigkeit gelingt. Für die Lösung der Aufgaben notwendiges Wissen wird jeweils bei den einzelnen Aufgaben mit dargestellt.

Bei den Vorprüfungen handelt es sich dagegen um den tatsächlich im Studium erworbenen Wissensstand (Fähigkeiten und später auch Fertigkeiten). Der Test ist dann ein gutes Zulassungskriterium, wenn dieser Wissenserwerb von den geprüften kognitiven Fähigkeiten so stark abhängt, dass eine Verwendung als Zulassungskriterium begründet ist. Es ist klar, dass auch andere Komponenten den Prüfungserfolg mit bestimmen (Motivation, aufgewendete Zeitressourcen und soziale Situation, Lernbedingungen usw.).

2.2.1 Erste Vorprüfung und Prüfungserfolg gesamt

Die Testwerte der einzelnen Gruppen nach dem Prüfungsverlauf sind in Tabelle 10 und als Boxplots in der Abbildung 4 dargestellt. In beiden Jahrgängen ergeben sich vergleichbare Testleistungen für die Gruppen. Unterschiedliche Zeiträume (ein oder zwei Jahre bis zur Prüfung) haben dabei nur einen geringen Einfluss. Auch in der dritten und vierten Prüfungssession hängt der Prüfungserfolg vom Testwert ab.

Die Personengruppe „Prüfung sofort bestanden“ unterscheidet sich von allen anderen signifikant, deren mittlerer Testwert liegt eine halbe Standardabweichung über dem aller anderen Gruppen (multipler Mittelwertvergleich für alle Personen, Tabelle 11). Personen, die im zweiten Anlauf bestanden haben, noch nicht angetreten sind sowie sich nicht immatrikuliert haben (aber über dem Kapazitäts-Kriterium liegen) bilden die zweite Gruppe mit einem mittleren Testwert um 100. Personen, die nicht bestanden haben, unterscheiden sich wiederum deutlich von dieser Gruppe – der Mittelwert liegt deutlich unter 100.

	Test 1998 (nach 2 Jahren)			Test 1999 (nach 1 Jahr)			Gesamt		
	n	m	s	n	m	s	n	m	s
Nicht immatrikuliert, TW unter Kriterium	107	83.91	4.639	88	83.18	4.372	195	83.58	4.523
Nicht immatrikuliert, TW über Kriterium	43	99.77	6.982	101	100.92	6.431	144	100.58	6.597
Noch nicht angetreten	33	99.61	6.955	83	98.27	7.727	116	98.65	7.510
Zweimal Nicht bestanden	46	97.17	6.441	11	95.91	7.148	57	96.93	6.535
Einmal Nicht bestanden	47	96.34	5.446	172	97.31	6.855	219	97.10	6.578
Zweiter Anlauf Bestanden	101	99.94	5.407	30	101.67	7.862	131	100.34	6.067
Sofort Bestanden	373	105.54	7.721	370	105.43	8.560	743	105.48	8.144
Total	750	100.02	9.965	855	100.02	10.018	1605	100.02	9.990

Tabelle 10: Testwert Mittelwert (m) und Standardabweichung (s) für Gruppen nach der ersten Vorprüfung

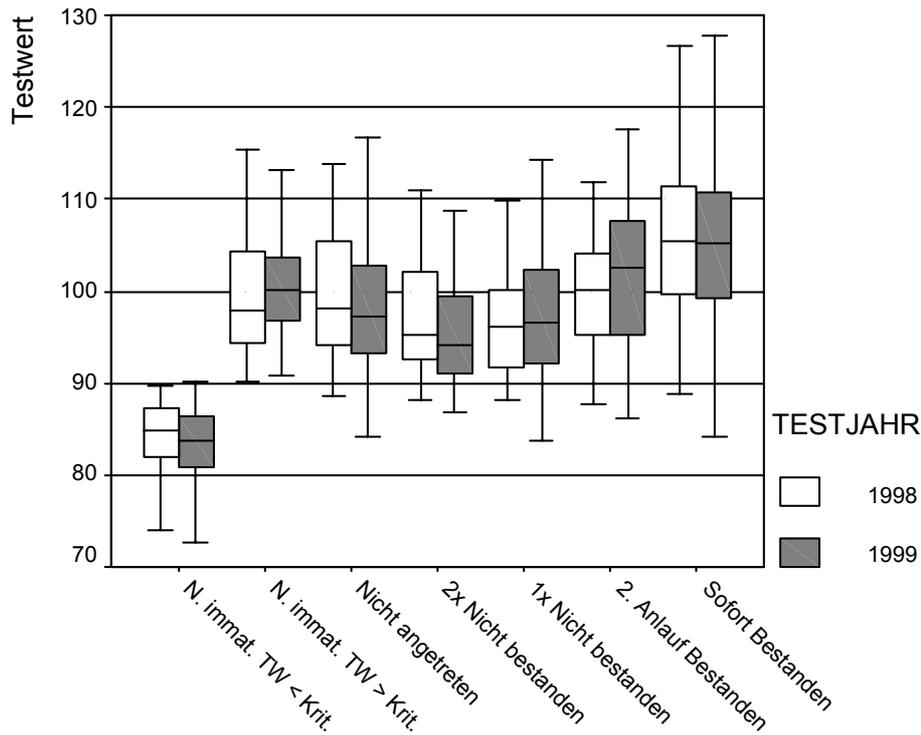


Abbildung 4: Boxplots für Testwerte (Median und 1. bis 4. Quartil) nach dem Prüfungsverhalten für die Kohorten 1998 und 1999

Erfolg 1. Vorprüfung nach Prüfungsanläufen	n	Subset for alpha = .05			
		1	2	3	4
Nicht immatrikuliert, TW unter Kriterium	195	83.607			
Zweimal Nicht bestanden	57		96.875		
Einmal Nicht bestanden	219		97.086		
Noch nicht angetreten	116		98.568	98.568	
Zweiter Anlauf Bestanden	131			100.261	
Nicht immatrikuliert, TW über Kriterium	144			100.547	
Sofort Bestanden	743				105.474
Signifikanz		1.000	.457	.263	1.000

Tabelle 11: Multipler Mittelwertvergleich des Testwertes (Tukey) für das Prüfungsverhalten beider Kohorten zusammengefasst

Gruppiert man die immatrikulierten Personen nach der sich ergebenden Studienverlängerung gegenüber dem optimalen Studienablauf (Tabelle 12 bzw. Abbildung 5), ergibt sich wiederum eine klare Trennung der Personengruppe mit einem Jahr Studiendauer bis zur ersten Vorprüfung von allen anderen (Mittelwertprüfung siehe Tabelle 13). Die beiden Gruppen mit notwendiger Verlängerung unterscheiden sich allerdings nicht voneinander.

Prüfungsverlauf	1998			1999			Total		
	n	m	s	n	m	s	n	m	s
Studienverlängerung mehr als 1 Jahr	82	98.1	6.8	83*	98.2	7.7	165*	98.2	7.2
Studienverlängerung mindestens 1 Jahr	120	98.4	5.6	183	97.2	6.8	303	97.7	6.4
"minimale" Studienzeit	398	105.2	7.8	400	105.1	8.5	798	105.2	8.2
	600	102.9	8.0	666	102.1	8.8	1266	102.5	8.4

Tabelle 12: Mittelwert (m) und Streuung (s) des Testwertes für die Studiendauer, *: Personen, die nach einem Jahr noch nie zur Prüfung angetreten sind

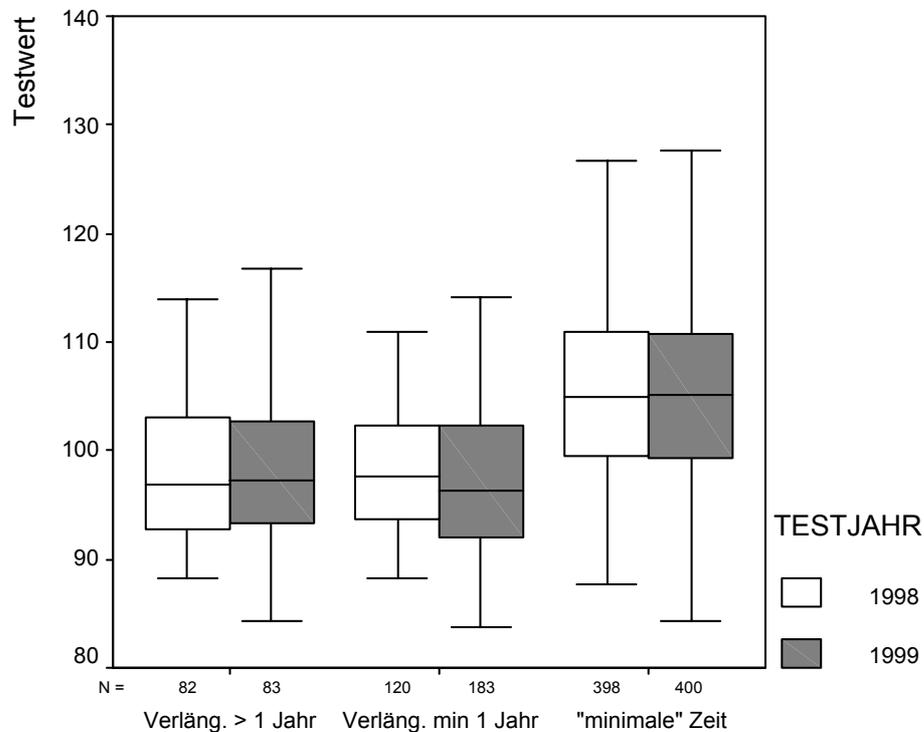


Abbildung 5: Boxplots für die Zusammenfassung der Personen nach der Studiendauer für beide Kohorten

Prüfungsverlauf	n	Subset alpha = .05	
		1	2
Studienverlängerung mindestens 1 Jahr	303	97.6	
Studienverlängerung mehr als 1 Jahr	165	98.1	
"minimale" Studienzzeit	798		105.1
Signifikanz		.728	1.000

Tabelle 13: Multipler Mittelwertvergleich des Testwertes nach Studiendauer für beide Kohorten

Die Testwerte derjenigen Personen, die zur ersten Vorprüfung angetreten sind, wurden nach dem Bestehen dieser Prüfung unterschieden und in der Tabelle 14 dargestellt. In beiden Jahren ergeben sich etwa gleiche Unterschiede (siehe auch Boxplot in Abbildung 6) und die Unterschiede sind statistisch signifikant (siehe Tabelle 15).

Testjahr		n	m	s
1998	Nicht bestanden	93	96.75	5.941
	Bestanden	474	104.35	7.638
1999	Nicht bestanden	183	97.22	6.860
	Bestanden	400	105.15	8.558

Tabelle 14: Mittelwert (m) und Streuung (s) des Testwertes für den Studienerfolg

Test-jahr	Varianzhomogenität	Levene's Test Varianzhomogenität		t-test Mittelwertgleichheit		
		F	Sig.	t	df	Sig. (2s)
1998	Varianzen inhomogen	9.573	.002	-10.714	158.107	.000
1999	Varianzen inhomogen	9.731	.002	-11.937	433.237	.000

Tabelle 15: Mittelwertvergleich (t-Test) des Testwertes nach dem Bestehen der ersten Vorprüfung

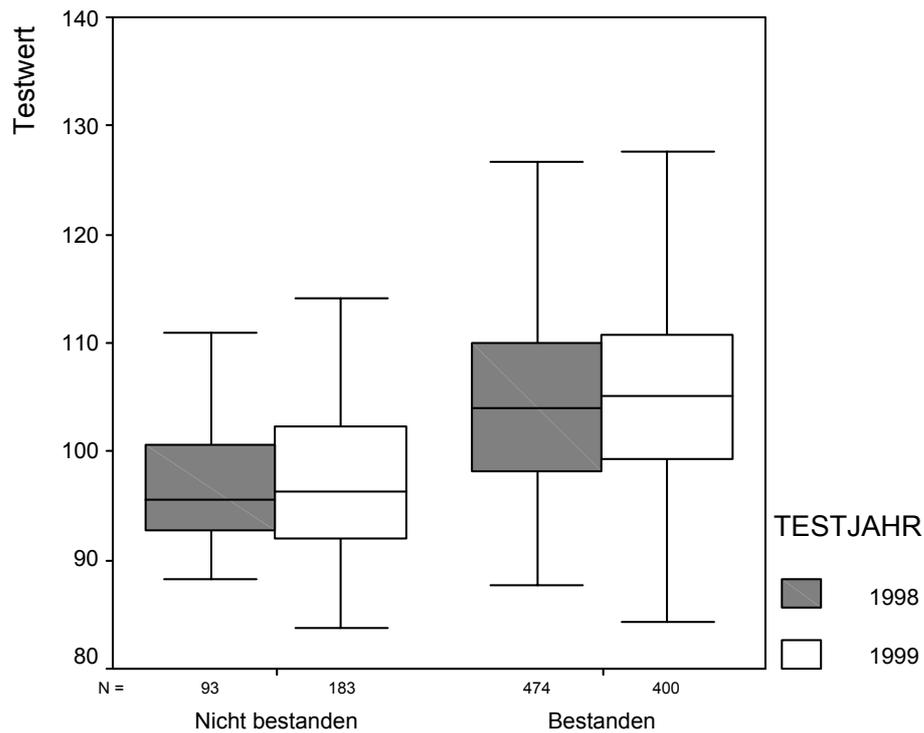


Abbildung 6: Boxplots des Testwertes für beide Kohorten nach dem Bestehen der ersten Vorprüfung

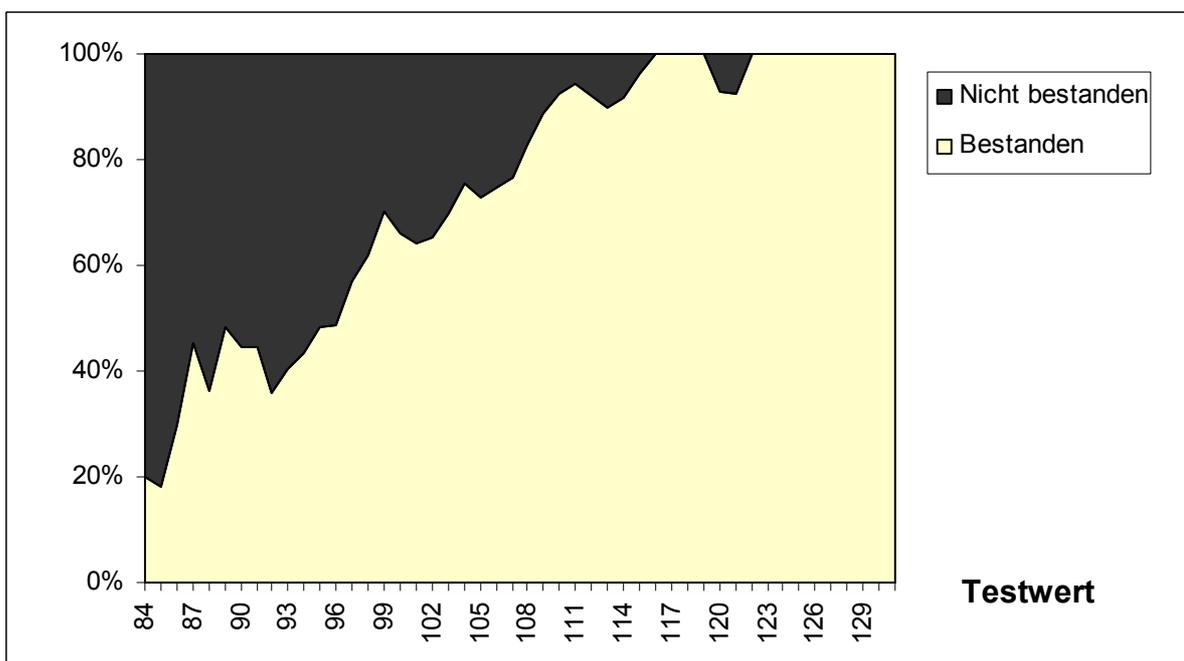


Abbildung 7: Wahrscheinlichkeit des Bestehens der 1. Vorprüfung pro Testwert für Kohorte 1998 nach 2 Jahren

Die Bestehenswahrscheinlichkeiten der Prüfung pro individuellem Testwert wurden berechnet und in Abbildung 7 (Kohorte 1998 nach 2 Jahren) und Abbildung 8 (Kohorte 1999 nach einem Jahr) dargestellt. Wiederum ergeben sich im unteren Bereich der Testwerte lineare Zusammenhänge zwischen Testwert und Bestehenswahrscheinlichkeit. In der Kohorte 1998 bleibt dieser Zusammenhang auch nach zwei Jahren Analysezeitraum deutlich

(Bestimmtheitsmass 0.83). Für die Kohorte 1999 nach einem Jahr Prüfungszeitraum ist das Bestimmtheitsmass mit 0.94 höher. Dies bedeutet, dass tatsächlich mit sinkendem Testwert die Wahrscheinlichkeit monoton abnimmt, eine Prüfung zu bestehen. Dies bedeutet allerdings auch, dass unterhalb eines bestimmten Testwertes die Wahrscheinlichkeit, eine Prüfung zu bestehen, recht gering wird. Durch die erneute Zuweisung von durch Rückzügen freiwerdender Studienplätze an diesen Personenkreis wird die Gesamt-Erfolgsrate des Jahrganges dann nachweislich verringert.

	Quadratsumme	Df	MQ	F	Sig.
Regression	.869	1	.869	122.630	.000
Residual	.170	24	.007		
Total	1.039	25			

Tabelle 16: Varianzanalyse der Regression: linearer Teil zwischen Testwert 88 und 113
Bestimmtheitsmass (adjustiert): $r^2 = 0.83$

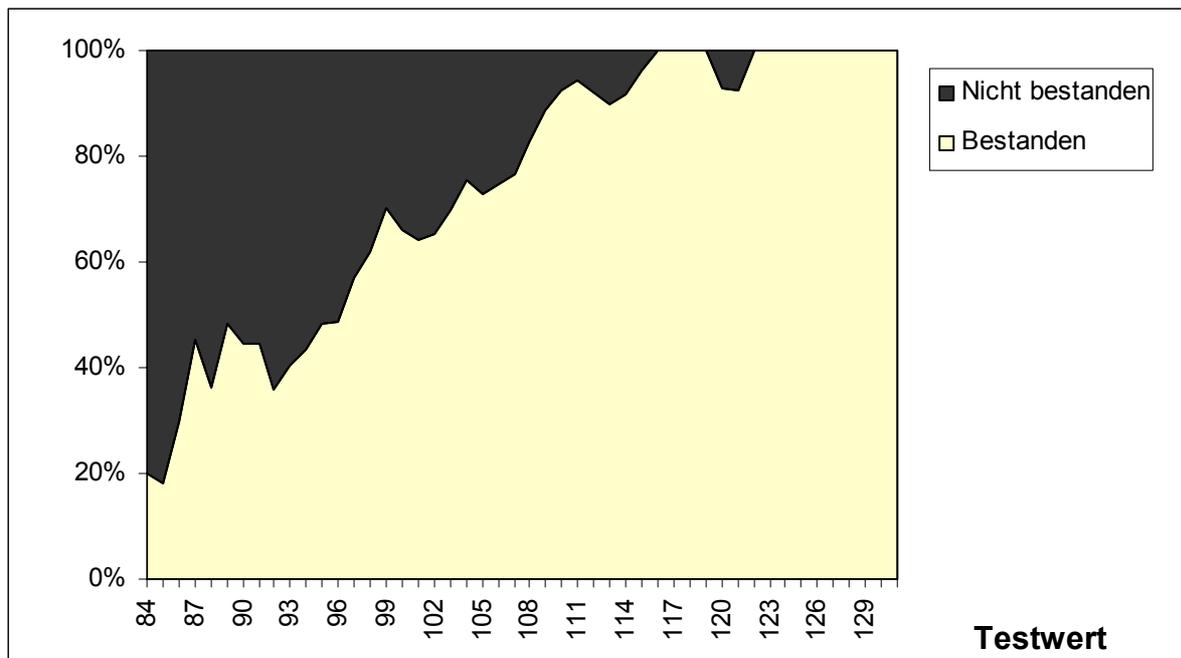


Abbildung 8: Wahrscheinlichkeit des Bestehens der 1. Vorprüfung pro Testwert für Kohorte 1999 nach 1 Jahr

	Quadratsumme	Df	MQ	F	Sig.
Regression	1.650	1	1.650	569.948	.000
Residual	.090	31	.003		
Total	1.740	32			

Tabelle 17: Varianzanalyse der Regression linearer Teil zwischen Testwert 84 und 116:
Bestimmtheitsmass (adjustiert): $r^2 = 0.94$

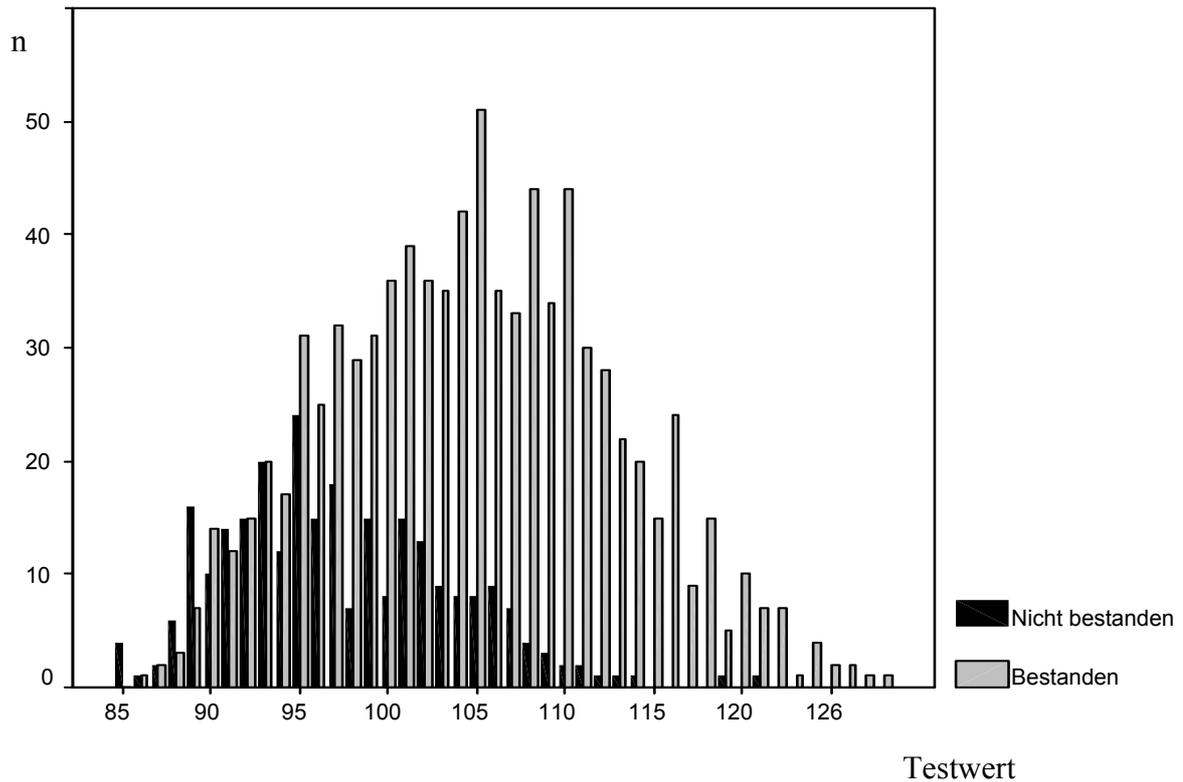


Abbildung 9: Häufigkeitsverteilung für den Testwert nach dem Bestehen der ersten Vorprüfung

In der Abbildung 9 sind die Häufigkeitsverteilungen nach dem Bestehen der Prüfung dargestellt. Beide Verteilungen sind deutlich unterschieden. Mittels Diskriminanzanalyse wurde die Trennbarkeit der beiden Gruppen geprüft. Bei 77.5% der Personen lässt sich allein aufgrund des Testwertes korrekt vorhersagen, ob die Prüfung bestanden wird oder nicht. In der Tabelle 19 werden die Ergebnisse einer Diskriminanzanalyse vorgestellt, die auf gewichteten Koeffizienten für die einzelnen Untertests beruht. Es wäre denkbar, dass eine differenziertere Verrechnung der Untertestwerte eine bessere Trennung erzielt. Es zeigt sich aber, dass dann nur 79.1% richtig klassifiziert werden – der Gewinn gegenüber der Verwendung des Testwertes als Summe der Untertestwerte von 1.6% zu vernachlässigen ist. Der Testwert ist in der aktuell verwendeten Form damit ein für die Erfolgsprognose valider Koeffizient. Die Gewichte der einzelnen Untertests entsprechen dabei in etwa der Reihenfolge der Korrelation mit den Noten, wobei die sehr studienanforderungsnahen Anforderungen ebenfalls besser abschneiden als Konzentration, räumliches Vorstellen oder Merkfähigkeit.

Tatsächlicher Prüfungserfolg	Vorhergesagt aufgrund des Testwertes		Total
	Nicht bestanden	Bestanden	
Nicht bestanden	68	208	276
Bestanden	50	822	872

Tabelle 18: Optimale Klassifikation nach dem Bestehen der ersten Vorprüfung aufgrund des Testwertes: 77.5% der Personen werden korrekt klassifiziert

Untertests	Standardisierte Koeffizienten der Diskriminanzfunktion	Strukturmatrix
Quantitative und formale Probleme	.415	.695
Diagramme und Tabellen	.255	.665
Med.-naturw. Grundverständnis	.187	.644
Textverständnis	.181	.612
Muster zuordnen	.245	.489
Konzentr. und sorgf. Arbeiten	.354	.426
Schlauchfiguren	-.069	.382
Figuren lernen	.066	.361
Fakten lernen	.122	.352

Tabelle 19: Optimale Klassifikation nach dem Bestehen der ersten Vorprüfung aufgrund einer Diskriminanzfunktion der einzelnen Untertestwerte: 79.1% werden korrekt klassifiziert

2.2.2 Effekt der Zulassung von der Warteliste

Die durch Rückzüge freiwerdenden Plätze werden bekanntlich an Personen vergeben, deren Testwert unterhalb des kapazitätsentsprechenden Testwerts (nachfolgend als Kriterium bezeichnet) liegt, die zunächst also eine Warteliste bilden. Dieses Kriterium wird nur nach der Kapazität festgelegt – es trennt keinesfalls Geeignete von Ungeeigneten. Stünden weniger Plätze zur Verfügung, würde das Kriterium höher festgelegt. Bei mehr Plätzen wäre es entsprechend niedriger. Die nachfolgende Statistik bezieht beide Kohorten ein.

Erfolg 1. Vorprüfung				
		Nicht bestanden	Bestanden	
Testwert unter Kriterium	Anzahl	38	25	63
	Prozentverhältnis	60.3%	39.7%	
Testwert über Kriterium	Anzahl	238	849	1087
	Prozentverhältnis	21.9%	78.1%	
Erfolg 2. Vorprüfung				
		Nicht bestanden	Bestanden	
Testwert unter Kriterium	Anzahl	1	1	2 (!)
	Prozentverhältnis	50.0%	50.0%	
Testwert über Kriterium	Anzahl	51	293	344
	Prozentverhältnis	14.8%	85.2%	

Tabelle 20: Erfolg in beiden Vorprüfungen und kapazitätsentsprechender Testwert (Kriterium), Erläuterungen siehe Text

Aus Tabelle 20 wird ersichtlich, dass 63 Personen zur ersten Vorprüfung angetreten sind, die das Kriterium eigentlich nicht erreicht haben und über die Warteliste zugelassen worden sind. Deren Erfolgsaussicht zum Bestehen dieser Prüfung beträgt nur 40% - im Vergleich zu 78% für Personen mit einem Testwert über dem Kriterium. Die zweite Vorprüfung erreichten bisher sogar nur zwei Personen mit einem Testwert unterhalb des Kriteriums. In der Tabelle 21 sind Mittelwerte und Standardabweichungen des Testwertes für die einzelnen Gruppen angegeben. Personengruppen, deren Testwert unter dem Kriterium liegt – die aber dennoch zugelassen wurden, weisen deshalb eine sehr geringe Variation der Testwerte auf. Die Erfolgreichen unterscheidet von den Nicht Erfolgreichen auch hier eine halbe Standardabweichung – der wesentlichere Unterschied ist allerdings die schon erwähnte Halbierung der Erfolgswahrscheinlichkeit.

Diese Befunde zeigen das Risiko auf, welches eine Kapazitätserhöhung bei gleichbleibendem Bewerbungsverhalten birgt: die Studiendauer wird länger und die Erfolgsaussichten werden geringer. Vermutlich wird auch die Abbrecherzahl steigen. Mehr Aufmerksamkeit sollte den Fragen gewidmet werden, wie sich mehr Geeignete für ein Medizinstudium bewerben, ob die Dissuasionsfaktoren möglicherweise die „Falschen“ abschrecken und selektiv wirken. Mit dem Eignungstest steht ein Instrument zur kapazitätsentsprechenden Zulassung zur Verfügung. Die Zugelassenen haben gute Chancen, das Studium mit Erfolg abzuschließen. Unterstellt man, dass steigende Maturandenzahlen eigentlich auch ein steigendes Interesse für das Medizinstudium hervorbringen müssten, bleiben nur Dissuasionsfaktoren als eine Ursache für das hohe, aber gleichbleibende Interesse für das Fach Medizin. Ein Teil davon wird in beruflichen Faktoren liegen. Dissuasion wurde in der Zeit vor Einführung des NC sogar gefördert und auch der Test wirkte anfangs dissuasiv. Vielleicht wäre jetzt die Zeit für eine Überprüfung dieses Herangehens. Wenn mehr Personen den Tests als Chance sehen würden, ihre Eignung festzustellen (und ihre Neigung zu diesem Studium prüfen), würden mehr Geeignete das Studium beginnen.

		n	m	s	
Testwert unter Kriterium	Vergleichswert alle Personen	274	84.9	4.5	
	Erste Vorprüfung	Nicht angetreten	16	88.3	1.7
		Nicht bestanden	38	87.8	2.0
		Bestanden	25	88.8	1.5
Testwert über Kriterium	Erste Vorprüfung	Nicht bestanden	238	98.5	5.8
		Bestanden	849	105.2	7.7

Tabelle 21: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Gruppen nach dem Erfolg in der ersten Vorprüfung und dem Erreichen des Kriteriums

2.2.3 Erfolgsvorhersage für verschiedene Universitäten

Die Vorhersagekraft des Testwertes für das Bestehen der Prüfung ist nicht spezifisch für einzelne Universitäten. Aus der Tabelle 22 wird ersichtlich, dass in allen Universitäten der Unterschied mindestens 5 Testwert-Punkte zwischen Bestehen und Nichtbestehen beträgt. Die Testwert-Niveaus zwischen den Universitäten unterscheiden sich weniger stark als

zwischen den beiden Gruppen nach dem Prüfungsbestehen (siehe Abbildung 10). Unterschiede finden sich allerdings in den Bestehensquoten – die liegen aber offenbar nicht (nur) in unterschiedlicher „Strenge“ bei den Prüfungen begründet, sondern auch in Leistungsunterschieden der Personen (Unterschiedliche „Bewerberprofile“ der Universitäten und Folgen der Umleitungen sind zu beachten).

Universität	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
Bern	46 (16%)	98.0	6.2	235 (84%)	104.2	7.8
Basel	48 (21%)	97.1	5.4	186 (79%)	102.6	8.2
Freiburg	69 (40%)	93.8	6.2	104 (60%)	102.3	9.0
Zürich	113 (25%)	98.6	6.7	347 (75%)	106.8	7.4

Tabelle 22: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für die Universitäten – beide Kohorten zusammengefasst

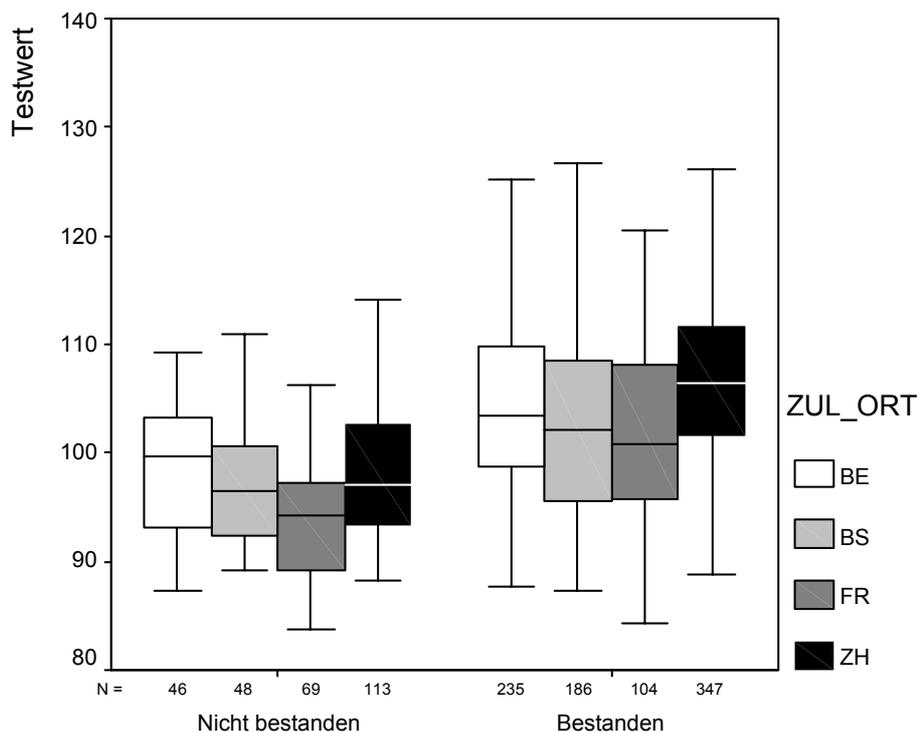


Abbildung 10: Boxplots des Testwertes für die verschiedenen Universitäten nach dem Bestehen der Prüfung

Die besondere Situation von Freiburg wird ebenfalls sichtbar: Bekanntlich ist die Rate der umgeleiteten Studierenden dort am höchsten. Würde man dort die Kapazitäten weiter erhöhen, führte dies vermutlich zur Zulassung von noch mehr Personen, die geringere Aussichten auf das Bestehen der Prüfung haben.

2.2.4 Erfolgsvorhersage für Geschlechter

Auch bezüglich der beiden Geschlechter findet sich der gleiche deutliche Unterschied des Testwertes für die beiden Gruppen nach dem Bestehen der Prüfung (Tabelle 23 und Abbildung 11). Bekanntlich waren 1998 und 1999 keine signifikanten Unterschiede der Testwerte für beide Geschlechter nachweisbar, wobei tendenziell die Testwerte der Männer immer ca. 1.5 Punkte besser waren. Für die beiden Gruppen, welche die Prüfung bestanden haben, sind die Testwerte hier nun nahezu identisch – die Gruppen, welche die Prüfung nicht bestanden haben, unterscheiden sich nur um 0.4 Punkte.

Dies ist ein Beleg, dass mögliche Testwertunterschiede zwischen den Geschlechtern keinesfalls als Chancenunterschiede zu interpretieren wären, sondern sehr wahrscheinlich die unterschiedliche Repräsentativität mit einem Überwiegen weiblicher Studierender dafür verantwortlich sein wird. Die testwertabhängigen „Chancen“, die erste Vorprüfung zu bestehen, sind praktisch bei gleicher Eignung identisch. Dieser Befund entkräftet jegliche Befürchtungen, dass es zu Ungleichbehandlungen bezüglich der Geschlechter kommt.

Geschlecht	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
männlich	100 (20%)	96.8	7.0	407 (80%)	104.6	8.3
weiblich	176 (27%)	97.2	6.3	467 (73%)	104.8	7.9

Tabelle 23: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für die Geschlechter – beide Kohorten zusammengefasst

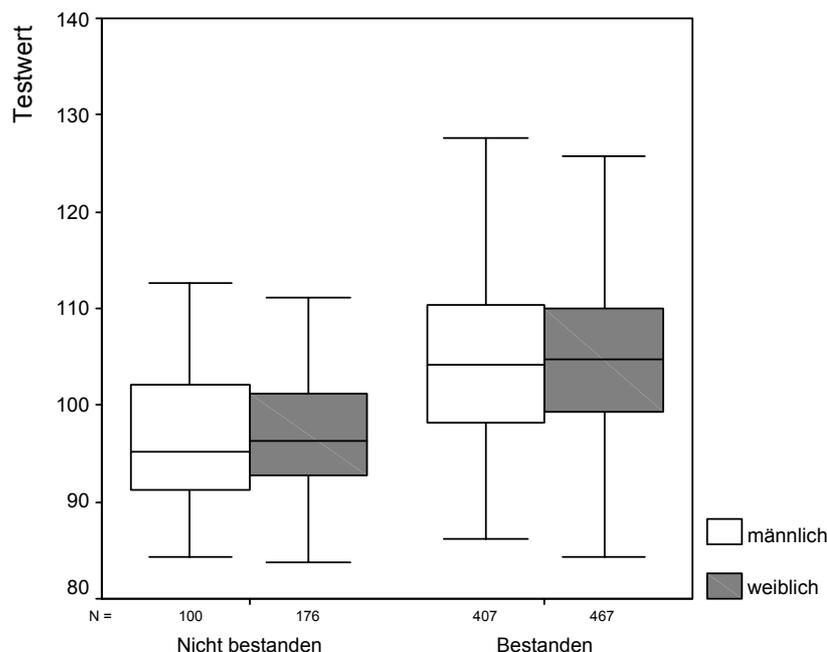


Abbildung 11: Boxplots des Testwertes für die Geschlechter nach dem Bestehen der Prüfung

2.2.5 Erfolgsvorhersage für Alter und Maturitätsjahr

In allen Testjahren ergab sich bisher der Befund, dass die Testwerte von Personen, die älter sind und die Maturitätsprüfung spät abgelegt haben, geringer sind als von Gleichaltrigen, welche die Maturität zum frühestmöglichen Zeitpunkt abgelegt haben. Nachfolgend werden zwei Gruppen „älterer Personen“ (für Geburtsjahre vor 1975) gebildet, die nach dem Ablegen der Maturität nach 1995 („späte Maturität“) oder vor 1996 („frühe Maturität“) unterschieden werden. Diese werden mit den „jüngeren Personen“ verglichen.

In allen drei Gruppen treten wiederum deutliche Testwertunterschiede hinsichtlich des Bestehens der Prüfung auf. Bei der älteren Gruppe mit später Maturitätsprüfung (die tendenziell geringere Testwerte haben) fällt die hohe Bestehensrate von 84% ins Auge – die Testwerte dieser Personen liegen dagegen mit 101.1 im Mittel etwas geringer als für die anderen beiden Gruppen.

Alter und Maturitätsabschluss	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
geb. vor 1975, Matur vor 1996	15 (44%)	96.5	4.7	19 (56%)	105.7	8.1
geb. vor 1975, Matur nach 1995	7 (16%)	94.9	5.1	37 (84%)	101.1	6.0
geb. nach 1974	254 (24%)	97.1	6.7	816 (76%)	104.8	8.1

Tabelle 24: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Gruppen nach Alter und Maturitätsjahr – beide Kohorten zusammengefasst

Die Testwerte unterscheiden in allen drei Gruppen Erfolgreiche von weniger Erfolgreichen deutlich (mindestens $\frac{1}{2}$ Standardabweichung). In der älteren Gruppe mit später Maturität können mit anderen Gruppen vergleichbare Erfolge allerdings auch mit niedrigeren Testwerten erreicht werden. Besonders im Boxplot (Abbildung 12) wird deutlich, dass für die ältere Gruppe mit später Maturität die Verteilungen beider Teilgruppen nach dem Prüfungsbestehen in Richtung niedriger Testwerte verschoben sind. Dies könnte ein Beleg dafür sein, dass die grössere „Lebenserfahrung“ dann für einen Bonus sorgt, wenn die „Lernerfahrung“ durch eine unlängst abgelegte Maturitätsprüfung aufgefrischt worden ist.

Die ältere Gruppe mit früherer Maturität hat bezogen auf die Testwerte identische Chancen wie die jüngere Gruppe, eine Prüfung zu bestehen. Diese Gruppe ist im biografischen Ablauf auch bis zum Ablegen der Maturität mit der „jüngeren Gruppe“ identisch – es wird nach der Maturität jedoch kein oder ein anderes Studium aufgenommen. Die Gruppe liegt bezüglich des Testwertes in allen bisher analysierten Jahren immer niedriger als die Personengruppe, die sofort das Studium aufnimmt. Die Chancengleichheit ist erfüllt: gleiche Testwerte bedeuten nachweislich für beide Gruppen altersunabhängig eine gleiche Eignung im Studium. Daher ist es sehr wahrscheinlich, dass dieser Eignungsunterschied schon nach dem Zeitpunkt der Maturitätsprüfung bestand. Wenn die Personen entsprechende Erfahrungen in der Schule machten (wie leicht oder schwer diese fiel), kann dies auch von einer frühen Studienaufnahme abgehalten haben.

Vorgreifend können wir feststellen, dass für die Gruppe der Älteren mit später Maturitätsprüfung der **einzige Beleg vorliegt**, dass im Vergleich zu anderen Gruppen niedrigere Testwerte eine höhere Bestehenswahrscheinlichkeit der Prüfung bedeuten. Auch hier unterscheidet sich der Testwert innerhalb der Personengruppe zwischen Erfolgreichen

und weniger Erfolgreichen. Es handelt sich um eine lineare Verschiebung um 4 bis 5 Punkte nach unten. Eventuelle Konsequenzen für das Zulassungsverfahren wären zu bedenken. Personen des „zweiten Bildungsweges“ (um die es sich vor allem handelt) unterscheiden sich im „biografischen Verlauf“ des Lebens von denen, welche die Maturitätsprüfung auf dem ersten Bildungsweg abgelegt haben. Es wäre denkbar, dies im Sinne einer Punktegutschrift zu berücksichtigen, die abhängig vom Alter bei Ablegen der Maturitätsprüfung gegeben wird.

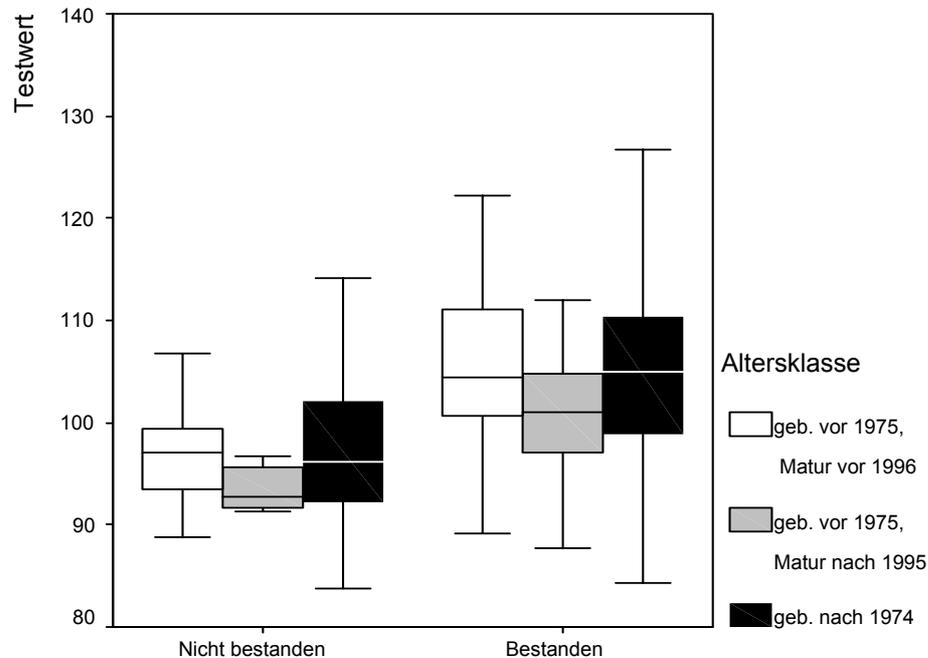


Abbildung 12: Boxplots des Testwertes für die Gruppen nach Alter und Maturitätsjahr nach dem Bestehen der Prüfung

2.2.6 Erfolgsvorhersage für Sprachgruppen

Auch für die drei Sprachgruppen treten vergleichbare Unterschiede nach dem Prüfungserfolg auf, wobei er in der italienischen Sprachgruppe tendenziell etwas weniger deutlich ist (sie ist allerdings auch die zahlenmässig kleinste Gruppe). Dieser Befund ist bedeutsam, weil er auch zeigt, dass es keine unterschiedliche Eignung bei gleichem Testwert zwischen den Sprachgruppen gibt. Die Chancengleichheit zwischen den Sprachgruppen ist gegeben.

Sprache	Nicht bestanden			Bestanden		
	n	m	s	n	m	s
deutsch	216 (21%)	97.1	6.3	806 (79%)	104.8	8.1
französisch	41 (53%)	95.9	7.5	37 (47%)	104.1	7.3
italienisch	19 (38%)	98.3	6.7	31 (62%)	102.7	7.7

Tabelle 25: Mittelwerte (m) und Standardabweichungen (s) des Testwertes für Sprachgruppen – beide Kohorten zusammengefasst

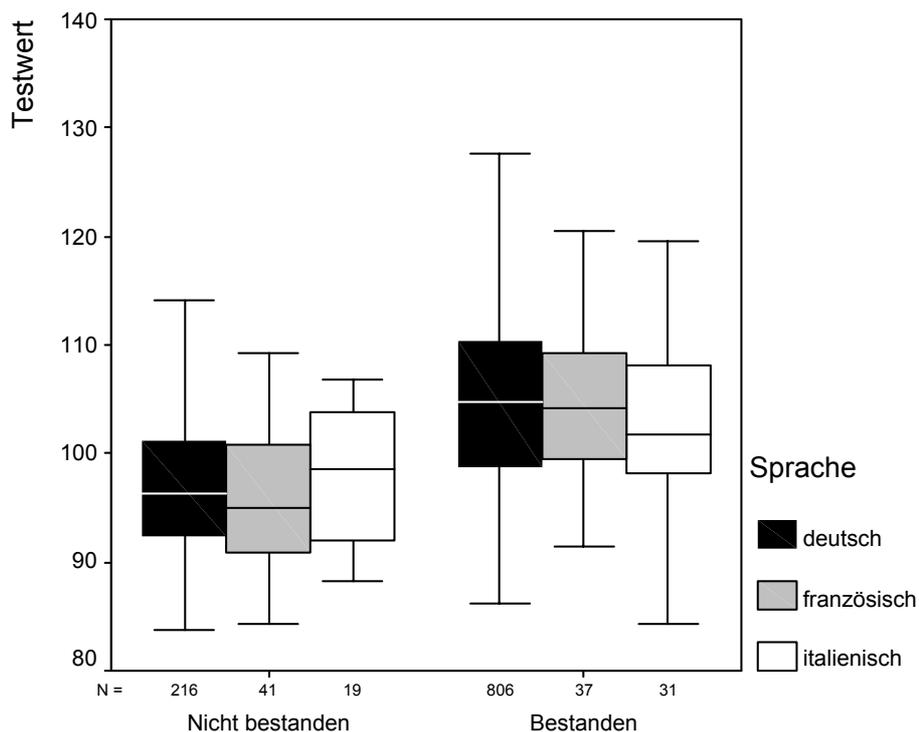


Abbildung 13: Boxplots des Testwertes für Sprachgruppen nach Bestehen der Prüfung

2.2.7 Unterschiede zwischen Human- und Veterinärmedizin

Einbezogen wurde der Testjahrgang 1999, wo erstmals für Human- und Veterinärmedizin ein Numerus Clausus zur Anwendung kam. Berücksichtigt sind in diesen Statistiken diejenigen Personen, die 1999 tatsächlich immatrikuliert worden sind. Tabelle 26 zeigt zunächst, dass in Veterinärmedizin weniger Personen die erste Vorprüfung bestehen als in Humanmedizin. Ist die Prüfung dort strenger – oder unterscheiden sich die Personen hinsichtlich ihrer Eignung?

Prüfungsverlauf	Humanmedizin		Veterinärmedizin	
	N	%	N	%
Noch nicht angetreten	65	12.4%	18	12.8%
2 x Nicht bestanden 1. und 2. Session	10	1.9%	1	0.7%
1 x Nicht bestanden 2. Session	54	10.3%	17	12.1%
1 x Nicht bestanden 1. Session	60	11.4%	41	29.1%
Bestanden 1. Session	272	51.8%	53	37.6%
Bestanden 2. Session	40	7.6%	5	3.5%
Bestanden 2. Session, Nicht bestanden 1. Session	24	4.6%	6	4.3%
Total	525		141	

Tabelle 26: Prüfungsverlauf für 1999 immatrikulierte Personen nach Disziplin

Für Eignungsunterschiede spricht, dass sich die mittleren Testwerte für Personen, welche die Prüfung bestehen, nicht signifikant zwischen Human- und Veterinärmedizin unterscheiden. Gleiches gilt für die mittleren Testwerte der weniger Erfolgreichen¹. Bei gleichem Testwert (sprich gleicher Eignung) ist der gleiche Prüfungserfolg wahrscheinlich. Die Zusammensetzung der Studierenden in Veterinärmedizin unterscheidet sich von derjenigen in Humanmedizin: die Testwerte sind im Mittel geringer.

	Humanmedizin			Veterinärmedizin		
	n	Testwert		n	Testwert	
		Mittelwert	Stand.abw.		Mittelwert	Stand.abw.
Nicht bestanden	124	96.9	7.2	59	97.9	6.0
Bestanden	336	105.3	8.8	64	104.4	6.8

Tabelle 27: Testwert für 1999 immatrikulierte Personen nach Disziplin und Prüfungserfolg

	Humanmedizin			Veterinärmedizin		
	n	Testwert		n	Testwert	
		Mittelwert	Stand.abw.		Mittelwert	Stand.abw.
Noch nicht angetreten	65	98.7	8.1	18	96.5	5.8
Zweimal Nicht bestanden	10	96.0	7.4	1	94.2	-
Einmal Nicht bestanden	114	97.0	7.2	58	97.9	6.1
Zweiter Anlauf Bestanden	24	101.2	8.0	6	103.2	7.4
Sofort Bestanden	312	105.6	8.8	58	104.5	6.8

Tabelle 28: Testwert für 1999 immatrikulierte Personen nach Disziplin und Prüfungsverlauf

¹ t-Test Vergleich Human- zu Veterinärmedizin für Bestanden: $t = 0,37$, $p \leq 0.67$; für Nicht bestanden: $t = 0.98$, $p \leq 0.32$.

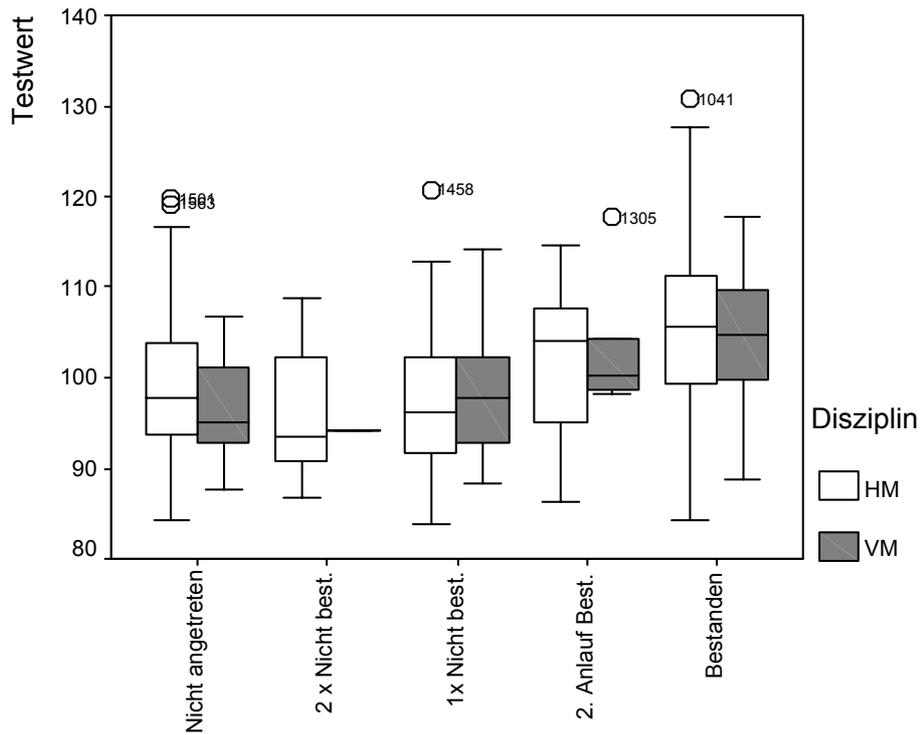


Abbildung 14: Testwertverteilung (Median und Quartile) für 1999 immatrikulierte Personen nach Disziplin und Prüfungsverlauf

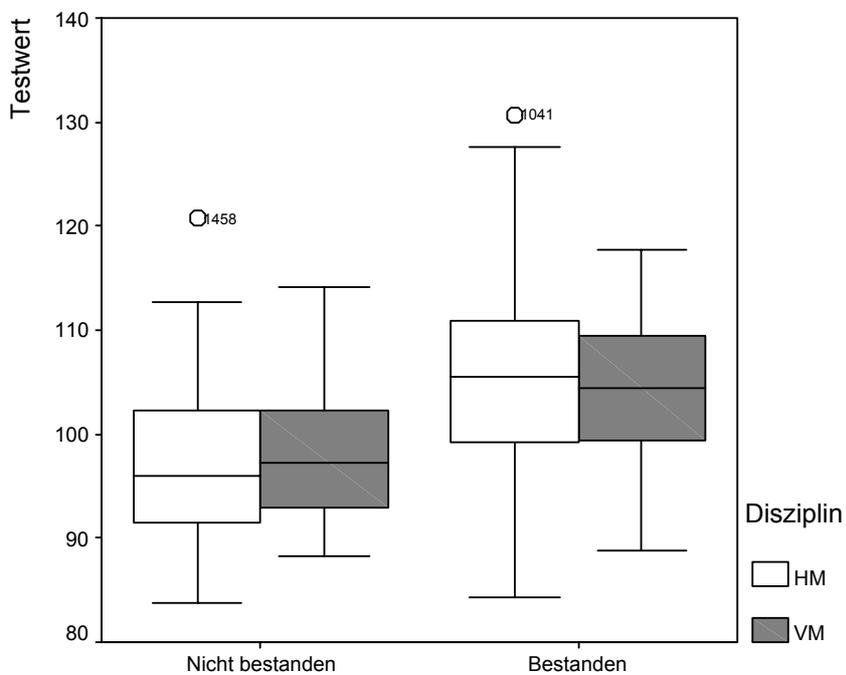


Abbildung 15: Testwertverteilung (Median und Quartile) für 1999 immatrikulierte Personen nach Disziplin und Prüfungserfolg

2.2.8 Zweite Vorprüfung

Die zweite Vorprüfung kann bisher nur durch die Personen der Kohorte 1998 absolviert worden sein. Dabei sind nur Personen angetreten, welche die erste Vorprüfung bestanden haben. Die Zusammensetzung ist daher nicht mit der Situation vor der ersten Vorprüfung zu vergleichen. Personen mit geringeren Testwerten sind in dieser Gruppe noch unterrepräsentiert und durch diese spezielle „Homogenisierung“ ist die Aussagefähigkeit eingeschränkt. Erst in den Folgejahren werden mehr Personen mit geringeren Testwerten über Prüfungsergebnisse für die 2. Vorprüfung verfügen. Deshalb hat auch die Gruppe der Personen, welche die Prüfung nicht bestanden hat, mit im Mittel 102.1 einen vergleichsweise hohen Testwert. Dennoch unterscheiden sich die Testwerte für Erfolgreiche und nicht Erfolgreiche signifikant (Tabelle 29).

Prüfungserfolg	n	m	s		
Nicht bestanden	52	102.15	6.623		
Bestanden	294	105.99	7.605		
Varianzgleichheit angenommen	Levene's Test Varianz		t-test Mittelwert		
	F	Sig.	t	df	Sig. (2-seitig)
	1.374	.242	-3.411	344	.001

Tabelle 29: Mittelwert (m) und Standardabweichungen (s) des Testwertes für 2. Vorprüfung nach Bestehen, sowie Ergebnisse des Mittelwertvergleiches (t-Test)

Prüfungsverlauf	n	m	s
Wegen Nichtbestehen 1. Vorprüfung noch nicht möglich	126	97.5	6.3
Zeitlich noch nicht möglich (1. Vorprüfung in 3. oder 4. Session)	79	100.0	5.4
Noch nicht angetreten, wäre aber möglich gewesen	49	103.9	8.6
1 x Nicht bestanden 4. Session	41	102.5	6.6
1 x Nicht bestanden 3. Session	11	101.0	6.7
Bestanden 4. Session, Nicht bestanden 3. Session	2	100.5	13.4
Bestanden 3. Session	110	105.8	7.8
Bestanden 4. Session	182	106.2	7.4

Tabelle 30: Mittelwert (m) und Standardabweichung (s) für die Gruppen nach dem Erfolg in der zweiten Vorprüfung

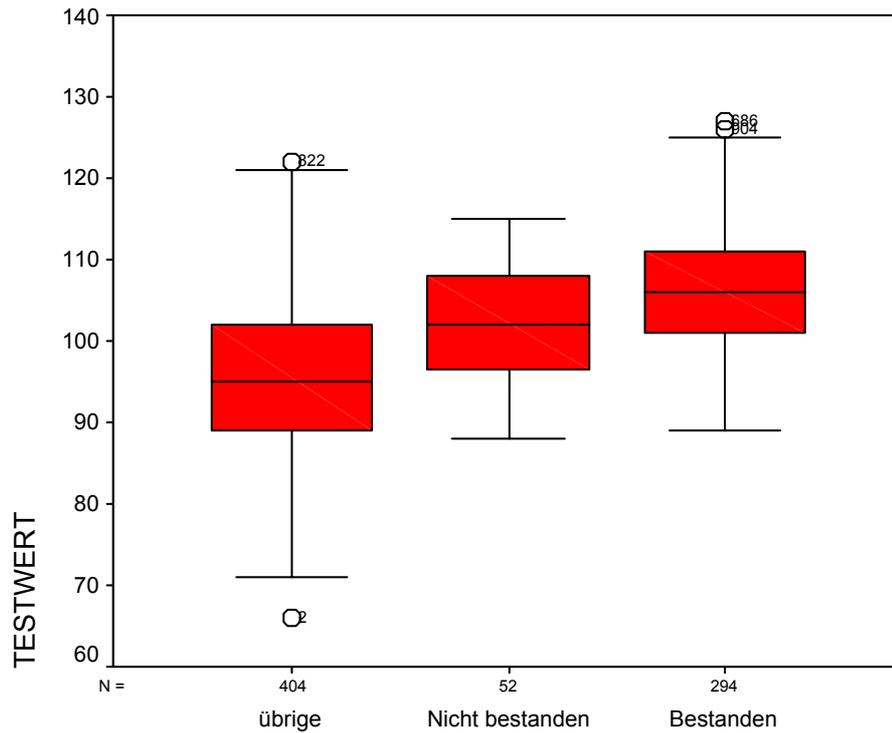


Abbildung 16: Boxplots des Testwertes für die Gruppen nach dem Bestehen der 2. Vorprüfung

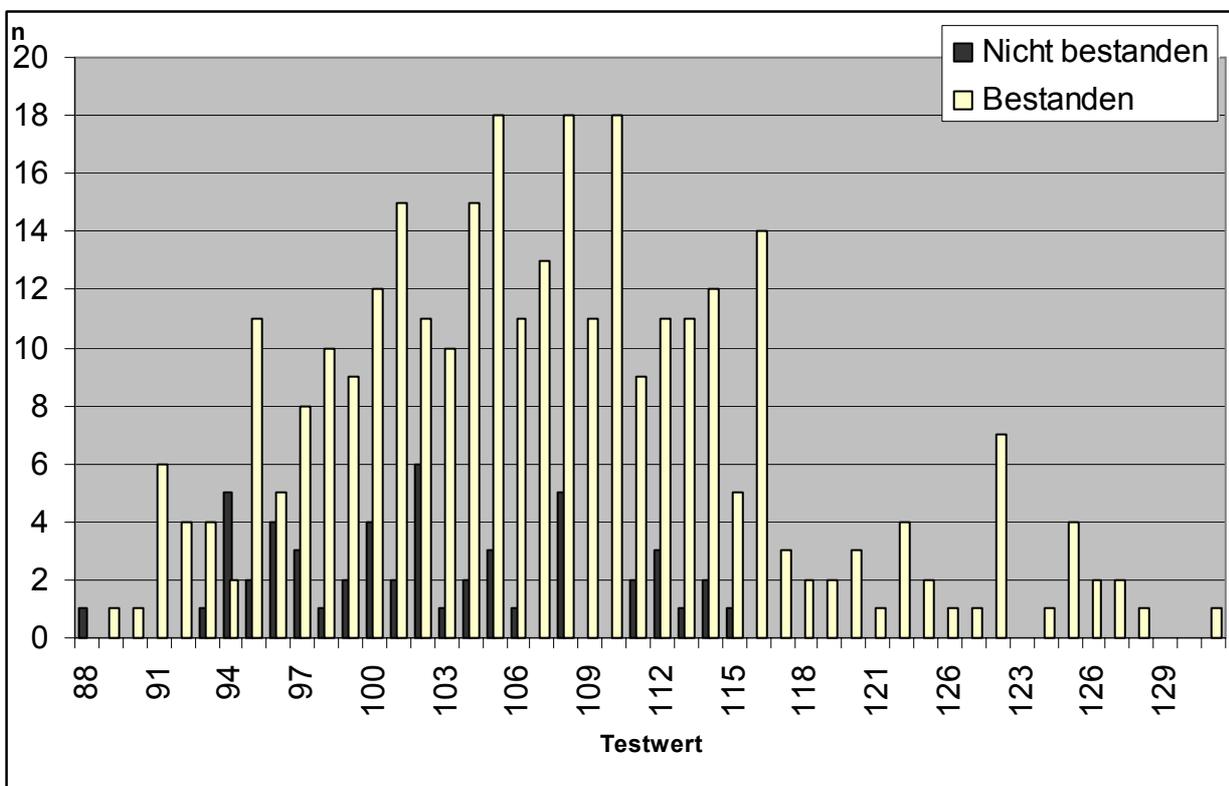


Abbildung 17: Häufigkeitsverteilung des Bestehens der 2. Vorprüfung nach Testwerten

Dass die Prognosekraft des Testwertes für die 2. Vorprüfung keinesfalls geringer sein wird als für die erste Vorprüfung belegt ein Vergleich der punktbiserialen Korrelationen (Assoziationskoeffizient Eta) zwischen Testwert und dem dichotomen Bestehenskriterium. Für die erste Vorprüfung beträgt dieser Wert 0.42 und für die zweite Vorprüfung 0.40. Rechnet man bei der zweiten Vorprüfung alle Personen, die schon zur Prüfung hätten antreten können als „nicht bestanden“ dazu, beträgt Eta sogar 0.44.

2.3 Testwert und Prüfungsnoten (Erste Vorprüfung)

Neben der Vorhersage der Studiendauer ist die Studienleistung das zweite wichtige Evaluationskriterium. Dessen Untersuchung wird dadurch erschwert, weil sich die Bewertungssysteme zwischen den Universitäten und zwischen den Jahren unterscheiden. Durch die Vergabe von (gleichen) Kreditpunkten im Falle des Bestehens der Prüfung wird auf die Notendifferenzierung z.T. bewusst verzichtet (Bern). Nachfolgend werden diejenigen Noten analysiert, die eine leistungsabhängige Streuung aufweisen und an einer ausreichenden Stichprobe gewonnen werden konnten.

Bezüglich der Vorhersage der Noten der zweiten Vorprüfung aufgrund des Testwertes sind noch keine statistisch gesicherten Aussagen möglich. Zum einen ist die Personengruppe mit vorliegenden Noten eine „homogenisierte“ Auswahl der Leistungsbesten der Kohorte 1998 (mit geringerer Stichprobengrösse), wo die Notenstreuung noch zu gering ist. In den Folgejahren werden mehr Personen die zweite Vorprüfung bestehen und es wird wahrscheinlich auch hier zu Notenverteilungen mit mehr Variation kommen.

Die erste Vorprüfung wurde in Bern 1999 (teilweise), sowie in beiden Jahren in Basel, Freiburg und Zürich in vier Fächern abgelegt. Erfasst wurden jeweils die letzten Noten, wenn eine Prüfung zweimal abgelegt wurde. In der Tabelle 31 wird deutlich, dass die Noten ausreichend streuen. Verwendet wurden auch die leistungsabhängig differenzierten Punktwerte in der ersten Vorprüfung nach dem Examenstyp „Reform 2“¹.

Note	Physik und Physiologie (a)	Chemie und Biochemie (b)	Allgemeine und Humanbiologie I (c)	Allgemeine und Humanbiologie II (d)
1	4	7	3	2
2	47	70	32	45
3	192	161	107	118
4	354	390	244	209
5	270	256	353	376
6	56	40	185	174

Tabelle 31: Notenverteilung 1. Vorprüfung gesamt (ohne Bern Reform 1999 und 2000)

¹ Gemäss Verordnung vom 1.11.99 für die medizinische Fakultät Bern. Wir danken Herrn Rainer Hofer vom Institut für Aus-, Weiter- und Fortbildung (IAWF) der Universität Bern für die Bereitstellung dieser Daten.

	Bern				Basel				Freiburg				Zürich			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1		1	1	1	2				2	6		1			2	
2	6	6	9	4	6	11	2	6	17	27	8	14	18	26	13	21
3	5	9	8	6	55	28	17	31	39	31	34	23	93	93	48	57
4	17	11	12	11	87	96	48	44	46	52	73	54	203	230	110	100
5	29	28	18	26	65	81	122	103	46	38	44	57	130	108	169	190
6	7	9	16	16	17	16	43	48	19	15	10	20	13		116	90
Total	64	64	64	64	232	232	232	232	169	169	169	169	457	457	458	458
m	4.9	4.9	5.0	5.2	4.4	4.6	5.1	5.0	4.6	4.4	4.5	4.8	4.3	4.3	5.2	5.3
s	0.6	0.7	0.9	0.7	0.8	0.7	0.7	0.7	1.0	0.9	0.8	0.7	0.7	0.6	0.7	0.7

Tabelle 32: Notenverteilung 1. Vorprüfung nach Universitäten (ohne Bern Reform 1999 und 2000), Fächerbezeichnung siehe Kopf der Tabelle 31

	n	Minimum	Maximum	m	s
Praktische Fertigkeiten (mündliche Prüfung)	111	7	40	27.5	7.3
Multiple-Choice-Prüfung theoretische Fertigkeiten	111	62	173	121.8	25.6

Tabelle 33: Mittelwerte und Standardabweichungen erste Vorprüfung Bern 2000 (Reform 2)

Die Korrelationen des Testwertes beider untersuchten Jahrgänge mit den Noten der ersten Vorprüfung (nach 4 Sessionen) zeigt die Tabelle 34, jene nach 2 Sessionen für den Jahrgang 1999 die Tabelle 35. Die Korrelationen mit dem Testwert sind für alle Fächer und den Durchschnitt hoch. Nach 2 Sessionen sind die Korrelationen deshalb etwas höher, weil hier noch mehr schlechtere Noten der nicht bestandenen Prüfungen eingehen – dieser Anteil verringert sich nach 4 Sessionen und es kommt zu einer Homogenisierung der Notenvariation, die korrelationsmildernd wirkt. Die Ergebnisse nach 2 Sessionen sind mit den Befunden der Kohorte 1998 nach einem Jahr (Hänsgen und Spicher, 2000) voll vergleichbar. Die Untertests korrelieren wie erwartet unterschiedlich mit den Noten. Es sei daran erinnert, dass das Weglassen jedes Untertests die Zuverlässigkeit des Testwertes mindert und seine hohe Güte sich aus allen verwendeten Untertests ableitet.

Auch die Kreditpunkte der Prüfung in Bern korrelieren sehr hoch mit dem Testwert und den einzelnen Testvariablen. Das ist auch auf die höhere Variation und damit auf die bessere Differenzierung der Bewertung zwischen den Personen zurückzuführen.

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Durchschnitts- note
	N = 509	N = 509	N = 509	N = 509	N = 509
Testwert	.432	.380	.400	.449	.495
	.000	.000	.000	.000	.000
Muster zuordnen	.116	.160	.157	.176	.182
	.009	.000	.000	.000	.000
Med.-naturw. Grundverständnis	.306	.284	.287	.336	.362
	.000	.000	.000	.000	.000
Schlauchfiguren	.195	.152	.157	.207	.212
	.000	.001	.000	.000	.000
Quantitative und formale Probleme	.434	.371	.321	.359	.441
	.000	.000	.000	.000	.000
Textverständnis	.331	.240	.304	.352	.367
	.000	.000	.000	.000	.000
Figuren lernen	.159	.146	.172	.199	.203
	.000	.001	.000	.000	.000
Fakten lernen	.150	.178	.241	.220	.238
	.001	.000	.000	.000	.000
Diagramme und Tabellen	.316	.249	.239	.297	.328
	.000	.000	.000	.000	.000
Konzentr. und sorgf. Arbeiten	.105	.191	.216	.214	.218
	.018	.000	.000	.000	.000

Tabelle 34: Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung Kohorte 1998 nach 4 Prüfungssessionen (1999 und 2000), ohne Bern „Reform 2“

	Mündliche Prüfung gesamt (Bern) N = 111	Multiple Choice Prüfung gesamt (Bern) N = 111	Physik und Physiologie N = 414	Chemie und Biochemie N = 415	Allgemeine und Human- biologie I N = 415	Allgemeine und Human- biologie II N = 415	Durch- schnitts- note N = 414
Testwert	.548	.577	.507	.491	.500	.478	.554
	.000	.000	.000	.000	.000	.000	.000
Muster zuordnen	.335	.322	.198	.197	.230	.212	.233
	.000	.001	.000	.000	.000	.000	.000
Med.-naturw. Grundverständnis	.386	.349	.410	.371	.413	.370	.438
	.000	.000	.000	.000	.000	.000	.000
Schlauchfiguren	.348	.392	.232	.197	.259	.256	.269
	.000	.000	.000	.000	.000	.000	.000
Quantitative und formale Probleme	.454	.494	.507	.464	.403	.357	.484
	.000	.000	.000	.000	.000	.000	.000
Textverständnis	.261	.309	.360	.338	.363	.348	.396
	.006	.001	.000	.000	.000	.000	.000
Figuren lernen	.293	.299	.202	.201	.213	.235	.241
	.002	.001	.000	.000	.000	.000	.000
Fakten lernen	.120	.138	.214	.230	.247	.235	.259
	.209	.147	.000	.000	.000	.000	.000
Diagramme und Tabellen	.484	.517	.457	.442	.375	.362	.454
	.000	.000	.000	.000	.000	.000	.000
Konzentr. und sorgf. Arbeiten	.278	.284	.189	.259	.244	.245	.266
	.003	.003	.000	.000	.000	.000	.000

Tabelle 35: Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung Kohorte 1999 nach 2 Prüfungssessionen 2000 (Bern „Reform 2“ und zusammengefasst Basel, Zürich, Freiburg)

In den beiden nachfolgenden Tabellen werden Human- und Veterinärmedizin verglichen. Festzuhalten ist eine hohe Korrelation der Noten mit den Testergebnissen, wobei sie in der Veterinärmedizin tendenziell sogar etwas höher sind. Die Prognosegüte des Tests für die Studienleistungen ist in beiden Disziplinen gegeben.

	Physik und Physiologie N = 346	Chemie und Biochemie N = 346	Allgemeine und Humanbiologie I N = 346	Allgemeine und Humanbiologie II N = 346	Durchschnitts- note N = 346
Testwert	.495	.474	.482	.444	.533
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Muster zuordnen	.194	.193	.210	.204	.224
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Med.-naturw. Grundverständnis	.407	.361	.395	.340	.422
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Schlauchfiguren	.233	.171	.235	.213	.245
	<i>.000</i>	<i>.001</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Quantitative und formale Probleme	.489	.458	.388	.341	.471
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Textverständnis	.342	.326	.359	.343	.387
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Figuren lernen	.228	.231	.238	.255	.270
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Fakten lernen	.220	.227	.253	.221	.258
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Diagramme und Tabellen	.470	.445	.380	.347	.457
	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Konzentr. und sorgf. Arbeiten	.173	.237	.237	.213	.245
	<i>.001</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>

Tabelle 36: Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung für **Humanmedizin**, Kohorte 1999, ohne Bern „Reform 2“

	Physik und Physiologie N = 68	Chemie und Biochemie N = 68	Allgemeine und Humanbiologie I N = 68	Allgemeine und Humanbiologie II N = 68	Durchschnitts- note N = 68
Testwert	.551	.569	.595	.660	.661
	.000	.000	.000	.000	.000
Muster zuordnen	.210	.213	.314	.250	.279
	.086	.081	.009	.040	.021
Med.-naturw. Grundverständnis	.414	.420	.481	.518	.512
	.000	.000	.000	.000	.000
Schlauchfiguren	.171	.277	.328	.433	.341
	.164	.022	.006	.000	.004
Quantitative und formale Probleme	.573	.460	.441	.397	.515
	.000	.000	.000	.001	.000
Textverständnis	.448	.404	.377	.363	.439
	.000	.001	.002	.002	.000
Figuren lernen	.126	.105	.159	.205	.168
	.305	.394	.195	.093	.171
Fakten lernen	.141	.202	.203	.271	.228
	.252	.099	.097	.025	.061
Diagramme und Tabellen	.327	.357	.327	.388	.387
	.006	.003	.007	.001	.001
Konzentr. und sorgf. Arbeiten	.215	.325	.250	.382	.324
	.078	.007	.040	.001	.007

Tabelle 37: Korrelation des Testwertes und der Punktwerte für Untertests (kursiv: das Signifikanzniveau) mit den Noten 1. Vorprüfung für **Veterinärmedizin**, Kohorte 1999, ohne Bern „Reform 2“

Weitere Teilgruppen werden in der Tabelle 38 verglichen. Die Prognosegüte des Testwertes für die Studiennoten ist nicht spezifisch für Universitäten (wo es auch Variationen der Prüfungsarten Schriftlich Essay, Schriftlich Multiple-Choice und Mündlich gibt, die keinen Einfluss auf die Prognosegüte haben – vergleiche dazu Hänsgen und Spicher 2000). Auch bezüglich der Geschlechter, der Sprachgruppen und der Altersgruppen gibt es keine Hinweise auf Unterschiede hinsichtlich der Prognosegüte des Tests.

	n	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Human- biologie I	Allgemeine und Human- biologie II	Durch- schnittsnote
Studienort						
Bern (Teil 1999)	64	.636	.509	.479	.580	.585
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Basel	233	.394	.350	.304	.377	.410
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Freiburg	169	.499	.527	.381	.389	.533
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Zürich	458	.516	.495	.506	.501	.577
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Alter und Maturität						
geb. vor 1975, Matur vor 1996	30	.501	.579	.520	.426	.590
		<i>.005</i>	<i>.001</i>	<i>.003</i>	<i>.019</i>	<i>.001</i>
geb. vor 1975, Matur nach 1995	34	.407	.341	.450	.312	.465
		<i>.017</i>	<i>.048</i>	<i>.007</i>	<i>.068</i>	<i>.005</i>
geb. nach 1974	859	.472	.436	.447	.462	.521
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Geschlecht						
männlich	411	.458	.403	.427	.407	.489
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
weiblich	512	.497	.473	.466	.501	.557
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
Sprache						
deutsch	829	.472	.433	.446	.460	.524
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>
französisch	52	.571	.549	.464	.501	.596
		<i>.000</i>	<i>.000</i>	<i>.001</i>	<i>.000</i>	<i>.000</i>
italienisch	42	.407	.420	.408	.496	.474
		<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>	<i>.000</i>

Tabelle 38: Korrelation der Noten für die 1. Vorprüfung mit dem Testwert für verschiedene Teilgruppen (kursiv: das Signifikanzniveau)

2.4 Schlussfolgerungen zur Eignungsfeststellung

Die Ergebnisse der vorliegenden Evaluation lassen es zu, für den Zeitraum bis zur zweiten ärztlichen Vorprüfung gesicherte Erkenntnisse zum Zusammenhang zwischen Zulassung aufgrund des Eignungstests und der Eignung zu gewinnen:

- Der Eignungstest für das Medizinstudium in der Schweiz (EMS) erfasst nachweislich Studieneignung, indem (1) Personen mit höheren Testwerten die Prüfungen mit höherer Wahrscheinlichkeit und zeitlich eher bestehen als die mit niedrigeren und (2) Personen mit höheren Testwerten auch eine höhere Studienqualität erreichen, die sich in besseren Prüfungsnoten ausdrückt.
- Der Zusammenhang zwischen Testergebnis und Studieneignung kann auf einem statistisch bedeutsamen Niveau festgestellt werden, welches die Verwendung des Testwertes als Zulassungskriterium rechtfertigt.
- Obwohl die Ergebnisse der zweiten Vorprüfung erst für wenige Personen vorliegen gibt es ausreichend Hinweise, dass die Vorhersagekraft des Testwertes für beide Vorprüfungen gleichermaßen hoch ist.
- Die Universitäten sind jetzt in der Lage, eine kapazitätsentsprechende Zulassung vorzunehmen und dadurch die Studienbedingungen zu verbessern. Das ist der wichtigste Vorteil des EMS, der seine Anwendung rechtfertigt. Ausserdem kann mit seiner Hilfe fair entschieden werden, wer an der „Wunschuniversität“ studieren kann und wer eine Umleitung in Kauf nehmen muss.
- Was sich offenbar nicht ändert, ist die Zahl der Geeigneten, die sich für ein Medizinstudium anmelden. Der Numerus Clausus ist mit rund 20% abgewiesener Bewerbungen sehr mild. Von den 80% Zugelassenen tritt ein Teil das Studium nicht an – Umleitungen von der Wunsch- an eine andere Universität sind eine wichtige Ursache. Dies führt nun dazu, dass die freiwerdenden Studienplätze an weniger Geeignete aus den 20% der Personen mit den geringeren Testwerten vergeben werden und das Niveau der Eignung der Studierenden gegenüber der Situation ohne Numerus Clausus praktisch unverändert bleibt. Die Zulassung von weniger Geeigneten, nur um Kapazitäten auszunutzen, hat nachweislich wenig Erfolgsaussicht: diese Personen haben um die Hälfte verringerte Erfolgchancen bei der ersten Vorprüfung, bei der zweiten Vorprüfung sind bisher sogar nur zwei Personen aus der „Warteliste“ gelangt.
- Das jetzt praktizierte Verfahren könnte ohne jegliche Systemveränderung im schlimmsten Falle dazu führen, dass die Studiendauern sich nur langsam verkürzen, die Abbrecherzahlen und Notenverteilungen sich nur unwesentlich gegenüber dem Zeitraum vor Einführung des NC verändern. Vor allem die bisher immer auch gewollten Dissuasionseffekte durch Test und Information über Studien- und Berufsbedingungen wären hinsichtlich möglicher selektiv-negativer Wirkungen zu hinterfragen. Welche Möglichkeiten zur Systemoptimierung gäbe es, um den relativen Anteil der Geeigneten im Medizinstudium zu erhöhen und dadurch auch den Einsatz der Ausbildungsressourcen zu optimieren?
 - o **Einflüsse auf das Bewerbungsverhalten**, indem vermehrt Geeignete angesprochen werden. Zu analysieren ist dabei, ob durch die „gewollten“ Dissuasionsfaktoren NC, Eignungstest und der geschmälernten Berufsperspektive („Ärztenschwemme“, Kostenregulierung u.a.) selektiv die Geeigneteren stärker abgeschreckt werden, sich zu bewerben. Möglicherweise muss das „Dissuasionsparadigma“ generell revidiert

- werden. Dazu gehört auch die Klärung der Frage, warum der Männeranteil bei den Bewerbungen so viel geringer ist und von Jahr zu Jahr weiter abnimmt¹.
- Personen, die sich trotz Zulassung nicht immatrikulieren, weisen ebenfalls nachweislich geringere Testwerte auf. Da ein grösserer Teil davon die Umgeleiteten betrifft und diejenigen mit niedrigeren Testwerten auch eher umgeleitet werden, ist dies nicht verwunderlich.
 - **Verzicht, Personen zum Studium zuzulassen**, die nachweislich ein „Mindest-Niveau“ der **Eignung nicht erreichen** – wo die Wahrscheinlichkeit, Prüfungen zu bestehen, von vornherein als gering eingeschätzt werden muss. Dabei würde man in Kauf nehmen müssen, die festgelegten Kapazitäten zu unterschreiten (die aktuellen Kapazitäten der ersten beiden Jahre beinhalten ihrerseits noch Schätzungen von Rückzügen, Abbrüchen und Verzögerungen, die auf den Zahlen der Situation vor dem NC beruhen).
 - Breite **Einführung der Reformstudiengänge**, um den Geeigneten dann solche Studienbedingungen zu verschaffen, die eine Optimierung von Studiendauer und Studienqualität erlauben, wodurch ein Studienabbruch dann einmal die Ausnahme würde.

Eine zu klärende Frage für das Zulassungsverfahren selbst wäre die Einführung einer Bonifikation für Personen, die älter sind UND die Maturität später abgelegt haben. Dies ist die einzige Teilgruppe, für die nachgewiesen werden kann, dass sich die Basis für die Vorhersage des Prüfungsbestehens aufgrund der Testwerte linear in Richtung niedrigerer Werte verschiebt. Eine mögliche Lösung wäre, den Mittelwertunterschied der Erfolgreichen als maximale Gutschrift proportional zur Differenz zwischen Lebensalter und Maturitätsalter für die Gruppe zu vergeben, die mehr als 6 Jahre älter sind als die jüngsten normalerweise erwarteten Bewerberinnen und Bewerber.

Beispiel: $104.8 - 101.1 = 3.7$ Punkte Differenz zwischen den erfolgreichen Jüngeren und Älteren mit später Maturität. Maturitätsalter liegt zwischen 18 und 34 Jahre. Gutschrift 3.7 Punkte bei 34 Jahren, 0 Punkte bei 23 Jahren – dazwischen ab 24 Jahren entsprechend einer linearen Funktion.

¹ Diese Analyse ist ohne Auswirkung auf die Chancengleichheit zu verstehen. Ein wichtiger und ermutigender Befund dieser Evaluation ist, dass Männer und Frauen bei gleichem Testwert und damit gleicher Eignung nahezu identische Chancen im Studium haben, Prüfungen in optimalen Zeiträumen zu bestehen und gute Noten zu erreichen. Es geht nicht darum, den absoluten Frauenanteil bei den Bewerbungen zu verringern, sondern im Gegenteil sollte es weiterhin ein Ziel sein, mehr Frauen zum Erwerb der Hochschulreife und der Aufnahme eines Studiums zu motivieren und die entsprechenden Bedingungen dafür zu schaffen.

Es geht um die Frage, warum der Männeranteil absolut und relativ beständig abnimmt. Uns scheint die Vermutung nicht ganz abwegig zu sein, dass – aus einem alten Rollenverständnis heraus – Männer durch den Test, die Möglichkeit einer Ablehnung und die Rückmeldung des Testergebnisses mit einer objektiven Vergleichsmöglichkeit zu anderen Personen stärker abgeschreckt werden als Frauen, wo diese Meidung von Konkurrenzsituationen im Leistungsbereich und „Versagensängste“ nicht so stark aus einem traditionellen Rollenverständnis begründet wären (obgleich sie natürlich auftreten können). Träfe das zu, wäre der stärkere Rückzug der Männer tatsächlich ein Hinweis, dass noch viel für die Überwindung der traditionellen Rollenverständnisse getan werden muss und ist insofern auch bedenklich.