

Zwischenbericht¹

Evaluation des Eignungstests für das Medizinstudium in der Schweiz – Zuverlässigkeit der Vorhersage von Studienerfolg

Klaus-D. Hänsgen

Zusammenfassung

Untersucht wurde, inwieweit der Testwert im Eignungstest für das Medizinstudium in der Schweiz (EMS) eine Vorhersage von Studienerfolg ermöglicht. Ein Nachweis dieser Prognosegüte rechtfertigt die Anwendung des EMS als Zulassungskriterium für das Medizinstudium, indem durch die Testbesten (1) das Studium schneller beendet wird und (2) bessere Studienleistungen erzielt werden. Zur Verfügung standen Prüfungsdaten der Personen, die 1998 ein Studium der Humanmedizin in Basel, Bern, Freiburg und Zürich begonnen haben. Für diesen Personenkreis kam erstmals ein Numerus clausus zur Anwendung. Im Sommer oder im Herbst 1999 konnten diese frühestens die 1. Vorprüfung ablegen.

Es kann gezeigt werden, dass sich Personen, welche die Prüfung im ersten Anlauf bestehen, bezüglich des Testwertes signifikant von den anderen Gruppen (Nicht angetreten, Nicht bestanden) unterscheiden. Zwischen der Bestehenswahrscheinlichkeit der Prüfung und dem Testwert besteht ein signifikanter linearer Zusammenhang. Auch die Prüfungs-Notendurchschnitte korrelieren um 0,50 mit den Testwerten, was die aus Deutschland bekannten Korrelationswerte bestätigt. Die Korrelationsmuster mit den einzelnen Fächern zeigen im Übrigen, dass nicht nur dann hohe Korrelationen auftreten, wenn wie im EMS auch zur Prüfung Multiple-Choice-Aufgaben verwendet werden.

Personen, die sich trotz Zulassung nicht immatrikulieren, unterscheiden sich bezüglich des Testwertes signifikant von den Gruppen, welche die Prüfung im ersten Anlauf bestehen. Dies kann als Hinweis gewertet werden, dass die Studieneignung auch beim freiwilligen Rückzug vom Studienplatz nicht ohne Einfluss bleibt .

Es wird bestätigt, dass Personen mit besseren Testleistungen die 1. Vorprüfung früher und mit grösserer Wahrscheinlichkeit bestehen als andere Personen und dass bessere Testleistungen auch mit besseren Prüfungsergebnissen einhergehen. Die Verwendung des EMS als Zulassungskriterium ist aufgrund der Daten für diesen ersten Studienabschnitt gerechtfertigt.

¹ Eine erweiterte Fassung erscheint in: Hänsgen, K.D., Spicher, B. (2000): Eignungstest für das Medizinstudium in der Schweiz (EMS). Bericht über die Testdurchführung 2000 und Ergebnisse zur Evaluation. Bericht 6 des ZTD: Freiburg/Schweiz (in Vorbereitung).

Einleitung

Seit 1998 muss die Zulassung für das Medizinstudium an den Universitäten Basel, Bern, Freiburg und Zürich durch einen Numerus clausus beschränkt werden, weil die Anmeldungen die Kapazitäten deutlich überschreiten. Während 1998 ein Numerus clausus nur für Humanmedizin galt, ist seit 1999 auch Veterinärmedizin betroffen.

Der Numerus clausus basiert auf dem Eignungstest für das Medizinstudium in der Schweiz (EMS), welcher aus Deutschland übernommen worden ist und dort bereits sehr gute Evaluationsergebnisse erzielt hat.²

Neben den juristisch geforderten Kriterien Rechtsgleichheit und Willkürfreiheit bei der Zulassung muss gelten, dass die Zugelassenen das Studium in angemessen kurzer Zeit beenden, um die Ressource der vorhandenen Studienplätze optimal zu nutzen. Die Zulassung ist dann fair, wenn gute Leistungen im EMS mit guten Studienleistungen in Beziehung stehen – wenn die „testbesten“ Zugelassenen das Studium mit besseren Leistungen in kürzerer Zeit beenden als die Personen mit weniger guten Testleistungen.

Bisher wurde anhand der Testdaten der Jahre 1998 und 1999 auch in der Schweiz nachgewiesen, dass der EMS wichtige Evaluationskriterien qualitativ hochstehend erfüllt.³

- Durch eine strenge Standardisierung der Durchführung, die Schulung der Testleiter und vergleichbare Bedingungen an allen Testorten ist die Rechtsgleichheit gegeben.
- Dies gilt insbesondere für die Sprachgruppen, wo durch ein vergleichsweise aufwendiges Analyseverfahren mögliche testbedingte Unterschiede ausgeglichen werden, die tatsächlich vorhandenen Unterschiede jedoch nicht nivelliert werden.
- Ein objektives Auswerteverfahren mit mehreren Kontrollschritten bei der Ermittlung des Testwertes und ein transparentes, gut definiertes Zulassungsverfahren garantieren auch die geforderte Willkürfreiheit.
- Der Test ist fair bezüglich der Geschlechter, der Altersgruppen, der Regionen mit unterschiedlicher Maturitätsquote innerhalb der Deutschschweiz und auch der einzelnen Kantone: der Testwert zeigt keine signifikanten Unterschiede (z.B. Geschlechter, Maturitätsquoten) bzw. bildet vorhandene Realitäten objektiv ab (z.B. Alter).
- Die Testgütekriterien wie Zuverlässigkeit und faktorielle Struktur entsprechen in der Schweiz bei beiden bisherigen Testdurchführungen dem von Deutschland her bekannten hohen Niveau.

Bei der Bewertung der Auswirkungen des Numerus clausus in der Schweiz muss man allerdings beachten, dass es sich um einen vergleichsweise „milden“ NC handelt:

- Nur etwa 1/5 der Personen mit einem gültigen Testergebnis kann kein Studienplatz in der ersten Zuteilungsrunde angeboten werden. Da notwendige Umleitungen an andere Universitäten manchmal nicht die Zustimmung der betroffenen Personen finden, können freige-wordene Studienplätze weiteren Personen angeboten werden. Berücksichtigt man dies, erhielten 1998 insgesamt nur 12% der Testabsolventen gar kein Studienplatzangebot. Es handelt sich also um keine extreme Auslese bei der Zulassung nach der Testleistung.
- Entgegen dem allgemeinen Trend verstärkter Nachfrage nach Hochschulausbildung zeigt sich für das Medizinstudium auch eine dissuasive Wirkung des Tests: die Anmeldezahlen

² Trost, G. (Hrsg.) (1996). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (20. Arbeitsbericht). ITB: Bonn. – Auszug auch im Bericht 4 (1998) und Bericht 5 (1999) des ZTD und im Internet unter http://www.unifr.ch/ztd/ems/berichte/b5/vorhersage_des_studienerfolges.htm. Erwähnenswert sind die hohen Zuverlässigkeitswerte, die mehrfach bestätigte Faktorenstruktur und vor allem die Prognosegüte für schriftliche und auch mündliche Prüfungsleistungen.

³ Vgl. Berichte 4 (1998) und 5 (1999) des Zentrums für Testentwicklung.

bleiben stabil bzw. sinken sogar leicht ab. Diese „Selbstregulation“ ist beabsichtigt – wird wahrscheinlich nicht nur nach der Studieneignung erfolgen.

- Schliesslich muss bei der Bewertung von Veränderungen an den Universitäten beachtet werden, dass zunächst noch eine hohe Zahl von Repetenten die Kapazitäten in Anspruch nimmt, bevor sich die Situation entspannt. Die jahrelange Überschreitung der Kapazitäten lässt sich nicht sofort abbauen, die Entlastung wird sich erst allmählich einstellen.

Eine wichtige Frage der Evaluation ist natürlich, inwieweit sich durch den Testwert Studienerfolg tatsächlich vorhersagen lässt. Bisher haben lediglich die 1998 für Humanmedizin Zugelassenen zwei Prüfungsmöglichkeiten gehabt (Sommer 1999 und Herbst 1999). Aufgrund dieser Daten ist eine erste Evaluation erfolgt und soll hier vorgestellt werden.⁴

Untersucht wurden alle Studienanfänger in Humanmedizin, die 1998 aufgrund des EMS an den Universitäten Basel, Bern, Freiburg und Zürich zugelassen wurden. Die Prüfungsnoten wurden aus dem Prüfungsregister des Bundesamts für Gesundheit (BAG)⁵ ermittelt und von den Prüfungsprotokollen individuell neu erfasst. Die Information über das Bestehen der Prüfung wurde direkt aus der Datenbank des BAG extrahiert. Berücksichtigt wurden die Prüfungssessionen Sommer 1999 und Herbst 1999.

Zahl der Personen, die immatrikuliert worden sind		597		100%
Noch nie angetreten (inkl. Studienabbrecher)		54		9,0 %
Im Sommer 1999 erster Versuch bestanden		297		49,7 %
Im Herbst 1999 erster Versuch bestanden		64		10,7 %
Im Sommer 1999 erster Versuch nicht bestanden		138		23,1 %
	<i>Davon im Herbst bestanden</i>		32	23 %
	<i>Davon im Herbst erneut nicht bestanden</i>		6	4 %
	<i>Davon nicht erneut angetreten</i>		100	73 %
Im Herbst 1999 erster Versuch nicht bestanden		44		7,4 %
Zahl der Personen, die sich nicht immatrikuliert haben		153		
Nicht zugelassen (aufgrund niedrigem Testwert)		88		
Zugelassen, aber nicht immatrikuliert (verschiedenste Gründe)		65		

Tabelle 1: Personengruppen nach der Teilnahme an der 1. Vorprüfung (Humanmedizin, Studienbeginn 1998) und Personen, die nicht zugelassen worden sind oder sich trotz Zulassung nicht immatrikuliert haben.

Neun Prozent der 1998 in der Humanmedizin Immatrikulierten haben noch keinen Prüfungsversuch unternommen. Darunter sind auch Personen, die das Studium bereits abgebrochen haben. Eine Bestimmung dieses Anteiles würde Studienverlaufsdaten des Bundesamtes für Statistik erfordern. Wegen der geringen Gruppengrösse wurden diese Daten in diesem Jahr noch nicht beigezogen.

Rund 60% der Immatrikulierten bestehen die Prüfungen beim ersten Antritt. Von den rund 25% Personen, welche die Prüfung im Sommer nicht bestanden haben, bestand nur ¼ im Herbst den zweiten Prüfungsversuch. Summa summarum wird es bei 1/3 der ursprünglich Immatrikulierten kein Bestehen der 1. Vorprüfung nach zwei Semestern geben, was eine wahrscheinliche Studienverlängerung und damit zusätzliche Blockierung von Ressourcen bedeutet.

⁴ Die Evaluation erfolgt im Auftrag des Beirates „Eignungstest“ der Schweizerischen Hochschulkonferenz. Der Evaluationskommission gehören an: Prof. R. Bloch (i.A. des Beirates „Eignungstest“ der SHK), K. Wechsler (i.A. der SHK), H. Neuhaus (i.A. des BAG), K. Holenstein (i.A. des BfS), Prof. K.-D. Hänsgen (i.A. des ZTD).

⁵ Besonders sei Herrn D. Megert vom BAG für die Aufbereitung der Daten für diese Studie gedankt.

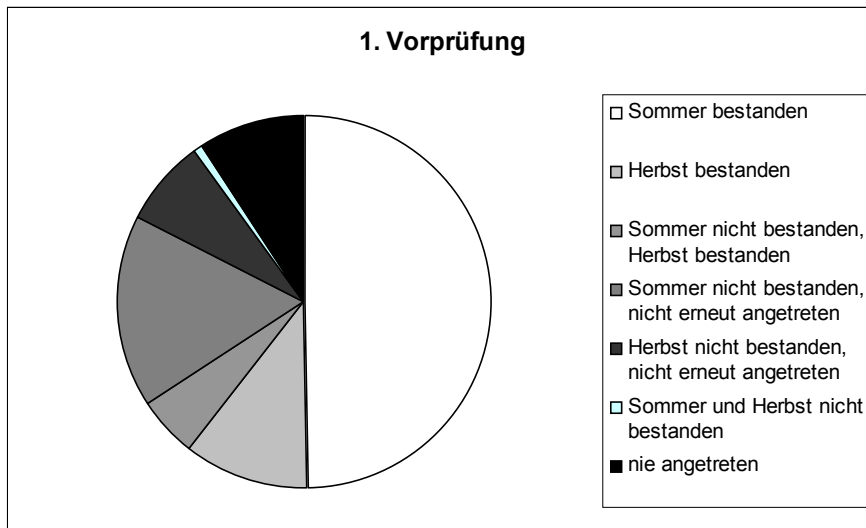


Abbildung 1: Anteile der einzelnen Gruppen nach dem Abschneiden in den Prüfungen

153 Personen haben am Test teilgenommen, sind aber nicht immatrikuliert: 88 Personen wurden nicht zugelassen (Testwert unter dem NC-Kriterium). 65 Personen wurden zugelassen, haben aber den Studienplatz nicht in Anspruch genommen (z.B. wegen einer Umleitung, der Aufnahme eines anderen Studiums, wegen persönlicher Gründe etc.).

Testwert und Prüfungserfolg

Besteht ein Zusammenhang zwischen Testleistung im Eignungstest und Prüfungserfolg? In der Tabelle 2 sind die Mittelwerte und Standardabweichungen für den Testwert der in Tabelle 1 genannten Gruppen dargestellt und in Abbildung 2 die Verteilungen als Boxplots gezeichnet.

	N	Mittelwert	Standardabweichung	95% Vertrauensintervall	
				Untergrenze	Obergrenze
<i>nicht zugelassen</i>	88	82.8	4.4	81.9	83.8
Sommer - Herbst -	6	96.2	4.6	91.3	101.0
Herbst -	44	96.7	5.1	95.2	98.3
<i>nicht immatrikuliert (aber zugelassen)</i>	65	96.8	7.8	94.9	98.8
Sommer -	100	98.5	6.1	97.3	99.8
nie angetreten	54	99.4	6.8	97.6	101.3
Sommer - Herbst +	32	100.9	5.2	99.1	102.8
Herbst +	64	105.3	7.5	103.5	107.2
Sommer +	297	105.7	7.8	104.8	106.6
Total	750	100.0	10.0	99.3	100.7

Tabelle 2: Kennwerte des Testwertes für die Gruppen der Tabelle 1

Der Gesamt-Mittelwert des Testwertes für alle Immatrikulierten liegt mit 102,9 ca. 1/3 Standardabweichung über dem Mittelwert aller Testteilnehmer von 100 (die Standardabweichung beträgt 10). Die nicht zugelassenen Personen unterscheiden sich erwartungsgemäss deutlich – erfolgte die Zulassung doch nach dem Testwert.

Interessant ist allerdings, dass die Personen, die sich trotz Zulassung nicht immatrikuliert haben, ebenfalls 1/3 Standardabweichung unter dem Mittelwert von 100 liegen. Wie sich beim multiplen Mittelwertvergleich (Tabelle 3) zeigt, unterscheidet sich diese Gruppe signifikant von den Erfolgreichen, welche die Prüfung im ersten Anlauf bestehen. Die „Dissuasion“ *nach* dem Zeitpunkt der

Testteilnahme scheint also nicht völlig unabhängig von der Studieneignung zu sein, obwohl sehr verschiedene Gründe für die Nichtaufnahme des Medizinstudiums denkbar sind.

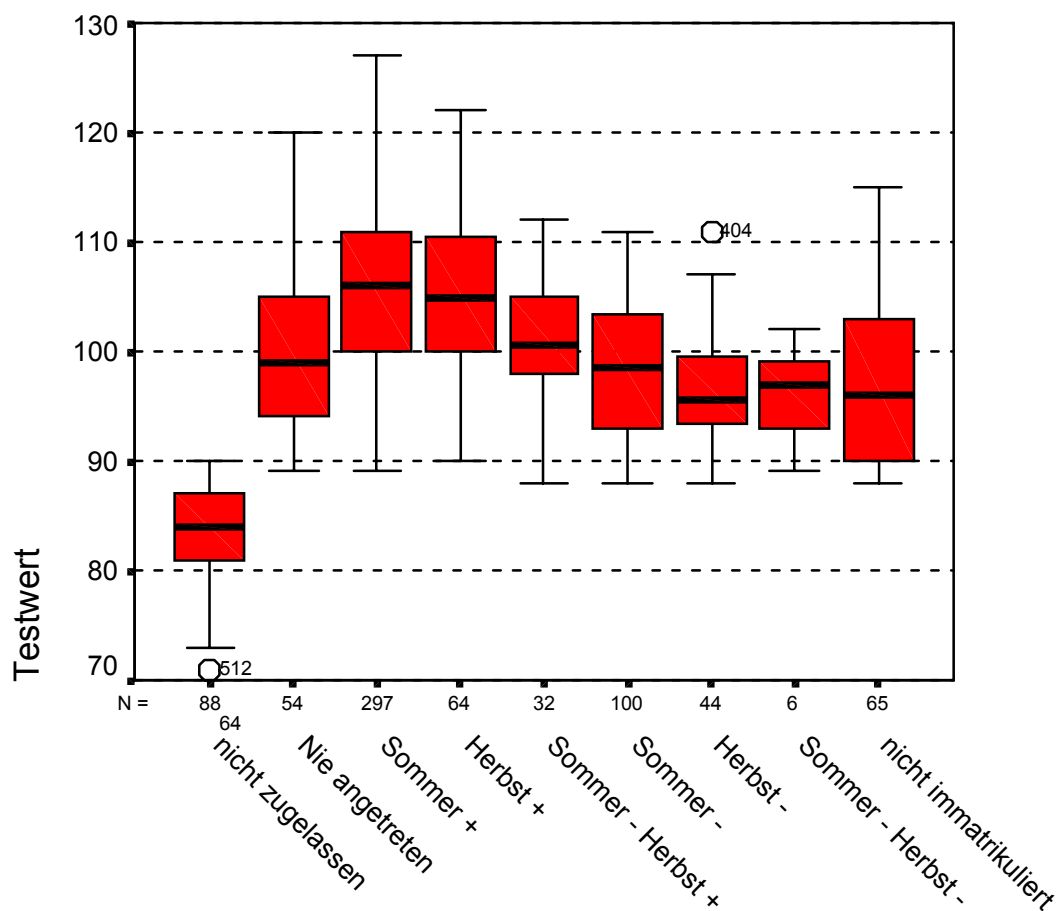


Abbildung 2: Boxplots des Testwertes für die Prüfungsgruppen. Dicke schwarze Linie: Median. Dargestellt sind die Quartile: 2. und 3. Quartil als Balken, 1. und 4. Quartil als Linie. Ausreißer sind markiert (z.B. bei Herbst Erstantritt nicht bestanden; eine Person mit Testwert 110 als nicht typisch für diese Gruppe); + bestanden, - nicht bestanden.

		Subset for alpha = .05		
		1	2	3
nicht zugelassen	88	82.85		
Sommer - Herbst -	6		96.17	
Herbst -	44		96.75	
nicht immatrikuliert (aber zugelassen)	65		96.85	
Sommer -	100		98.56	
nie angetreten	54		99.44	
Sommer - Herbst +	32		100.94	100.94
Herbst +	64			105.38
Sommer +	297			105.71
Sig.		1.000	.151	.151

Tabelle 3: Multipler Mittelwertvergleich (Tukey) für Testwert: Drei homogene Gruppen nach Nicht zugelassen, Bestehen oder Nichtbestehen/Nicht-Angetretensein getrennt. + bestanden, - nicht bestanden.

Im multiplen Mittelwertvergleich werden drei homogene Gruppen unterschieden: Personen, welche die Prüfung bestehen, bilden die eine Gruppe. Klar getrennt davon werden Personen, welche die Prüfungen nicht bestanden haben oder noch nicht zur Prüfung angetreten sind. Die dritte Gruppe sind erwartungsgemäss die Nicht-Zugelassenen, was ja aufgrund des Testwertes erfolgte.

Personen, welche zweimal die Prüfung nicht bestanden haben, unterscheiden sich dabei um fast 10 Punkte (eine Standardabweichung) von denjenigen, welche die Prüfung im ersten Anlauf bestehen. Personen, die erstmals im Herbst antreten und dort nicht bestehen, haben fast genauso niedrige Testwerte wie die Personen, die bereits zweimal nicht bestanden haben.

Personen, die noch nie zur Prüfung angetreten sind, unterscheiden sich bezüglich dem Testwert auch deutlich von denjenigen, welche die Prüfung im ersten Anlauf bestehen. Die Personengruppe, welche die Prüfung im zweiten Anlauf besteht, lässt sich beiden Mittelwertsgruppen zuordnen.

Es zeigt sich insgesamt, dass das Bestehen der Prüfung sehr eng mit den Testwerten zusammenhängt. Fasst man die Prüfungsteilnehmer nach „Bestanden“, und „Nicht bestanden“ zusammen, wird dies am deutlichsten (Tabelle 4). Die Erfolgreichen in der ersten Vorprüfung unterscheiden sich signifikant von den Personen, welche noch nicht angetreten oder nicht bestanden haben (Tabelle 5). Im Boxplot (vgl. Abbildung 3) zeigt sich die sehr geringe Überlappung der Quartile 2 und 3 für „Bestanden“ und „Nicht bestanden“.

	N	Mittelwert	Standardabweichung
nicht bestanden	150	97.9	5.8
nicht angetreten	54	99.4	6.8
bestanden	393	105.2	7.6
Total	597	102.9	7.9

Tabelle 4: Zusammenfassung für Testwert nach Bestehen und Nichtbestehen der 1. Vorprüfung

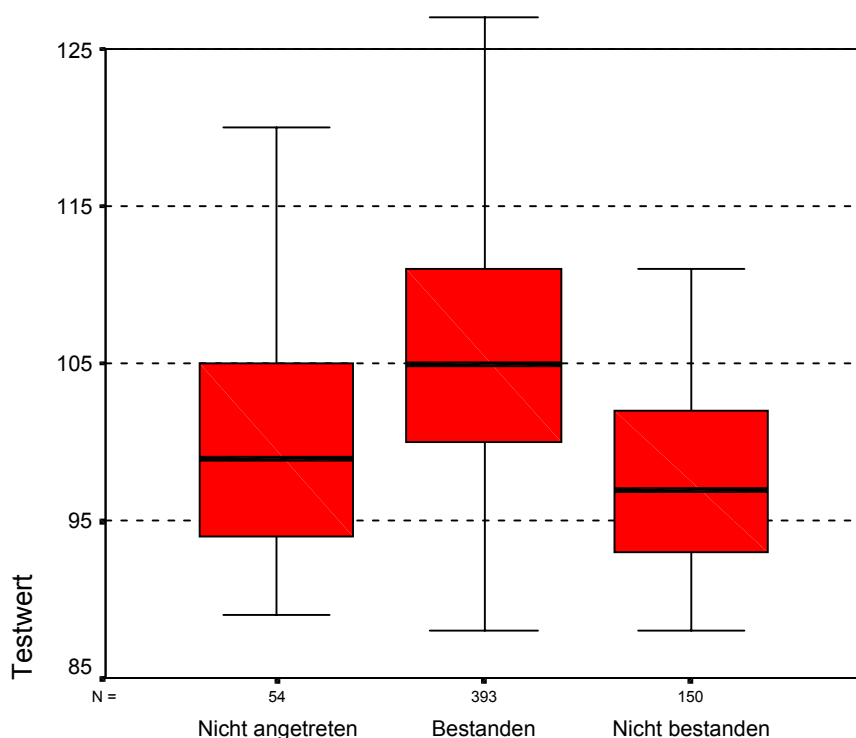


Abbildung 3: Boxplots für Zusammenfassung nach "bestanden", "nicht bestanden" und "nicht angetreten"

	N	Subset for alpha = .05	
		1	2
nicht bestanden	150	97.9	
nicht angetreten	54	99.4	
bestanden	393		105.3
Sig.		.270	1.000

Tabelle 5: Multipler Mittelwertvergleich: Prüfung bestanden wird von "nicht bestanden" und "nicht angetreten" signifikant unterschieden.

Die nachfolgende Abbildung 4 zeigt das Prozentverhältnis von "Bestanden" und "Nicht bestanden" (ohne immatrikulierte Personen, die nicht zur Prüfung angetreten sind) pro Testwert.⁶ Alle Personen mit einem Testwert > 112 haben die Prüfung bestanden. Bei niedrigeren Testwerten zeigt sich eine sehr deutliche monotone Zunahme der Bestehenswahrscheinlichkeit in Abhängigkeit vom Testwert.

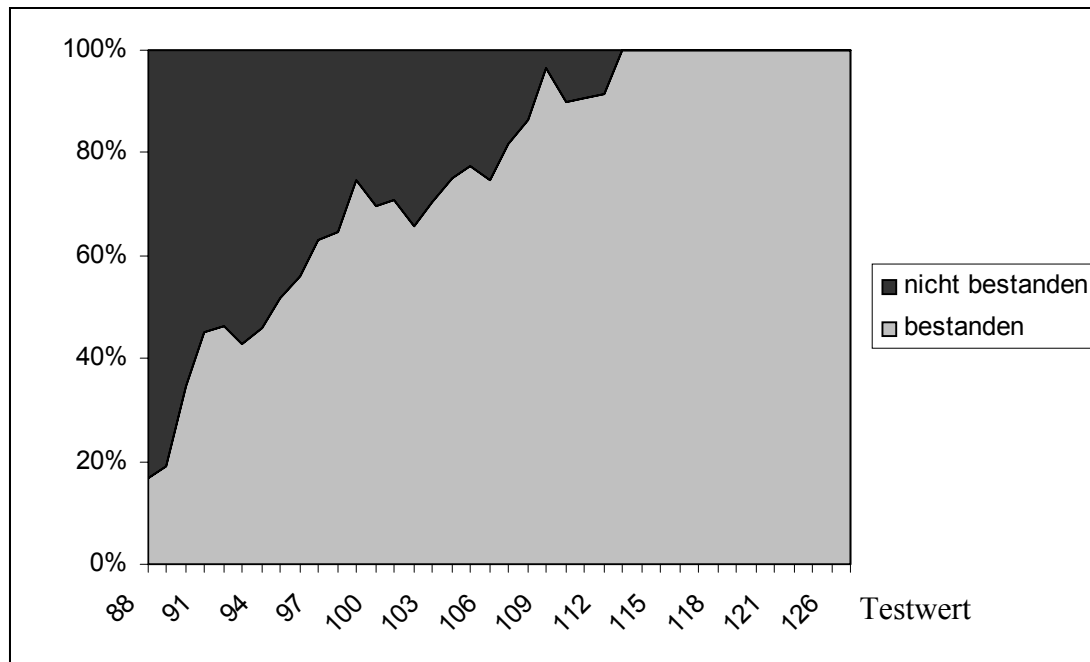


Abbildung 4: Prozentverhältnis "bestanden" und "nicht bestanden" pro individuellem Testwert (x-Achse)

Berechnet man die Bestehenswahrscheinlichkeit der Prüfung für jeden einzelnen Testwert lässt sich für den „linearen“ Teil der Kurve (Wahrscheinlichkeit kleiner als 1, d.h. Testwerte sind kleiner als 112) ein signifikanter linearer Zusammenhang zwischen Testwert und der Bestehenswahrscheinlichkeit nachweisen. Bei einem Bestimmtheitsmass r^2 von 0,76 ($F = 81,8$; $p < 0,001$) fällt dieser ausserordentlich hoch aus (Abbildung 5).

In der Abbildung 6 wurden die Personen einbezogen, die noch zu keiner Prüfung angetreten sind. Auch dieser Zusammenhang ist monoton: für Nichtantreten sind ebenfalls niedrigere Testwerte charakteristischer. Er ist allerdings wie erwartet nicht so deutlich wie bei den Personen, welche die Prüfung nicht bestehen. Sehr verschiedene Möglichkeiten sind denkbar, warum noch keine Prüfung absolviert worden ist, nicht alle müssen dabei von der Leistung abhängen.

⁶ Eine Glättung der Kurve gleicht Einzelschwankungen aufgrund der Schwankung der Personenzahl pro Testwert aus. Dazu wurden für jeden Testwert die Häufigkeiten der beiden Nachbarwerte zu seiner Häufigkeit addiert.

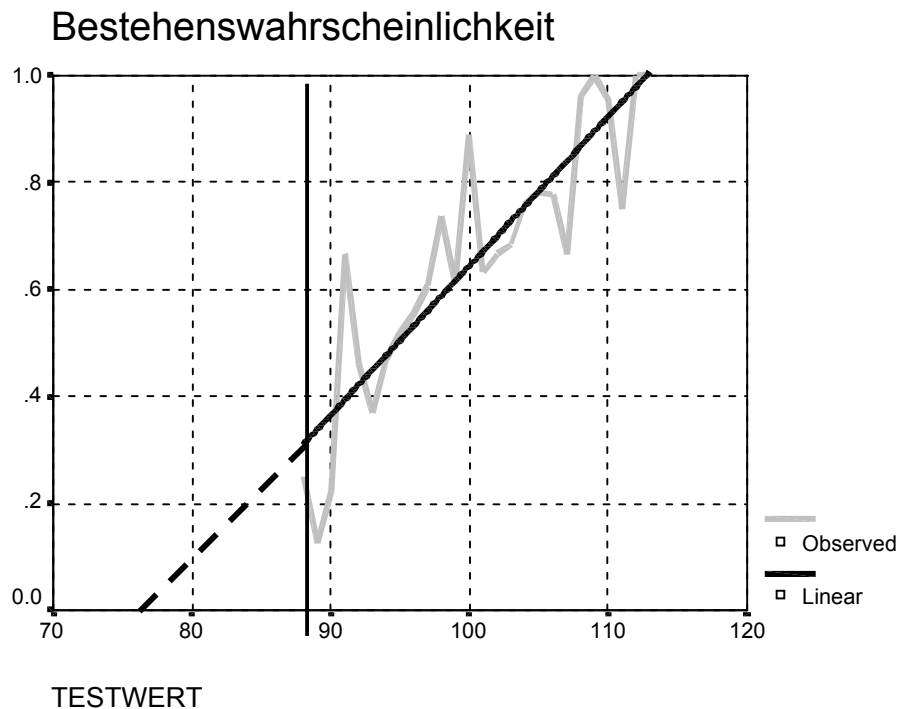


Abbildung 5: Lineare Regression der Bestehenswahrscheinlichkeit für jeden Testwert. Grau: beobachtete Werte, schwarz: Regressionsgerade. Gestrichelt: Schätzung für die Testwerte von Personen, die nicht an Prüfungen teilnahmen (nicht zugelassen worden sind).

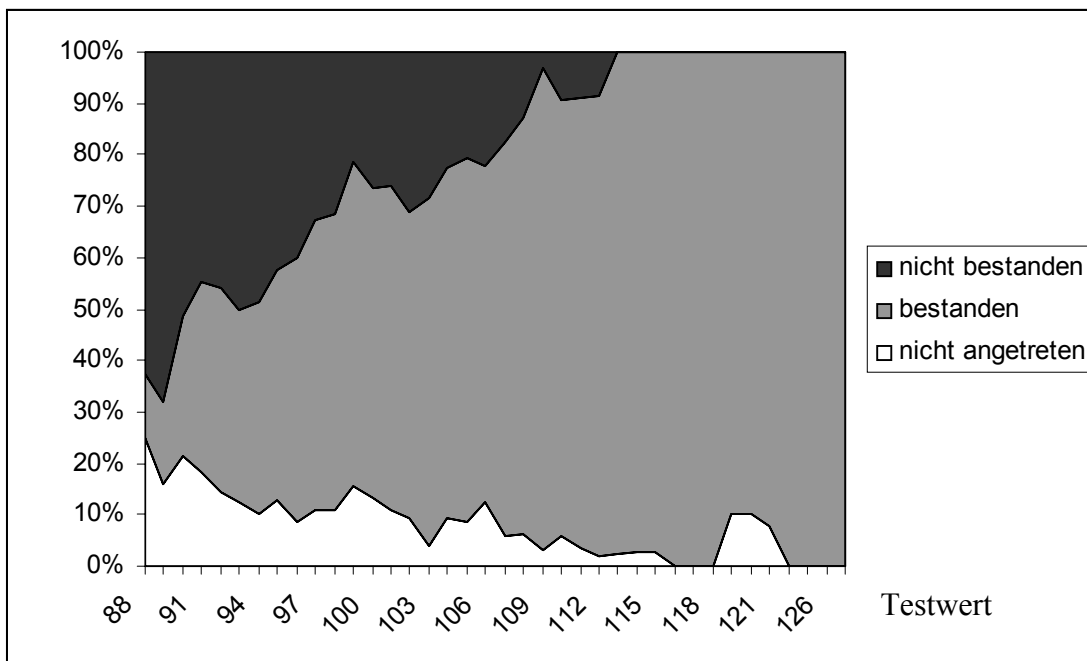


Abbildung 6: Prozentverhältnis "bestanden", "nicht bestanden" und "nicht angetreten" pro einzeltem Testwert (x-Achse)

Die recht deutliche Abhängigkeit des Prüfungserfolges vom Testwert und damit der Studieneignung in der ersten Vorprüfung belegt, dass der Test tatsächlich so etwas wie ein „Probestudium“ ist, dass die Komplexität und die Art der Aufgaben richtig gewählt sind. Wie im Studium erfasst der Test auch die Fähigkeit, sich in einer neuen Situation zurechtzufinden und in kurzer Zeit effektive Lösungsstrategien zu entwickeln. Die zum Teil extreme Höhe dieses Zusammenhanges

zeigt, dass sich zumindest in der Anfangsphase des Studiums die Test- und Studienanforderungen sehr ähneln.

In dieser ersten Studienphase ist auch die Abbrecherrate bekanntlich am höchsten. Deshalb ist es gerechtfertigt, die Studienzulassung vor allem auf den Erfolg in dieser kritischen Phase auszurichten und die begrenzte Ressource der Studienplätze fair an Personen mit den besten Erfolgsaussichten zu verteilen.

Es bleibt abzuwarten, ob die nachfolgenden Studienabschnitte ebenso genau vorhergesagt werden können. Dies wäre in dieser Höhe wie bei der ersten Vorprüfung sogar etwas überraschend.

Testwert und Prüfungsnote

Erfasst wurden die Prüfungsnoten, um innerhalb der Gruppen weiter differenzieren zu können. Es muss zunächst festgehalten werden, dass zwei Typen von Examen angewendet worden sind.⁷ Die Mehrzahl der Personen hat ein Examen nach Typ 1 abgelegt – in Bern kam neu ein Examen vom Typ 2 zur Anwendung. Gleiche Benennung der Examen und gleiche Noten bedeuten allerdings nicht, dass die Anforderungen und die Noten-Massstäbe der Universitäten vergleichbar sein müssen (siehe Abbildung 7). Einige Prüfungen wurden schriftlich, einige mündlich abgenommen. Bei einigen schriftlichen Prüfungen kam ein Multiple-Choice-Verfahren zur Anwendung, einige wurden frei beantwortet. Dies ist in der Tabelle 8 entsprechend gekennzeichnet.

Die Analysen werden vor allem durchgeführt, um die aus Deutschland bekannten Ergebnisse zu überprüfen. Die Benotungen verschiedener Universitäten wurden zusammengefasst, obwohl sie sich jeweils systematisch unterscheiden können. Erhoben wurde pro Person die jeweils letzte Note pro Teilprüfung – bei Wiederholungsprüfungen also die Note der Wiederholung.

Note	Examen Typ 1 BE: 64, BS: 132, FR: 75, ZH: 215				Examen Typ 2 BE: 49			
	Physik und Physiologie	Chemie und Biochemie	Allgemeine u. Humanbiologie I	Allgemeine u. Humanbiologie II	Examen 1 alle Lehrinhalte	Examen 2 alle Lehrinhalte	Examen 3 alle Lehrinhalte	Examen 4 mündlich
1	3	2	5	2				
2	35	40	23	15	2	2	3	
3	105	89	60	61	5	4	3	4
4	180	187	119	109	23	14	12	16
5	134	147	170	185	18	29	26	21
6	29	20	108	113	1		5	8
N	486	485	485	485	49	49	49	49

Tabelle 6: Notenverteilung für die Examen (Quelle: Prüfungsprotokolle im BAG), jeweils letzte Note pro Fach für jede Person. Note 6 ist die beste Note, ab Note 4 gilt eine Teilprüfung als bestanden.

Für die Examen vom Typ 1 können Korrelationsanalysen durchgeführt werden, weil die Fallzahl ausreichend hoch ist und auch die Notengebung zwischen den Stufen ausreichend variiert. Für Examen vom Typ 2 muss wegen der geringen Variation der Noten eine grössere Fallzahl abgewartet werden, wobei sich allerdings bereits jetzt deutliche Unterschiede zum Examenstyp 1 zeigen (s.u.).

⁷ Die Medizinalprüfungen für Ärzte an allen medizinischen Fakultäten sind durch eidg. Prüfungsverordnungen geregelt. Während für den ersten Examenstyp die Verordnung über die Prüfung von Ärzten vom 19.11.80 gilt, regelt beim zweiten Examenstyp z.Zt. eine eidg. Verordnung vom 1.11.99 das besondere Prüfungsverfahren im Rahmen des im Herbst 96 eingeführten Berner Reformcurriculums (dies löst die erste Derogationsverordnung vom 24.11.96 ab).

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Prüfungsdurchschnitt	Vergleich schriftliche Vorprüfung Deutschland
Testwert	.46	.44	.48	.50	.54	.50⁸
N	486	485	485	485	485	

Tabelle 7: Korrelation der Prüfungsnoten (alle vorliegenden Fälle mit Daten) mit dem EMS; Rangkorrelation nach Spearman – der Masskorrelationskoeffizient nach Pearson weicht um maximal 0.02 ab.

Die Korrelationen liegen für Prognose-Koeffizienten im Niveau sehr hoch (Werte zwischen 0.3 und 0.4 werden üblicherweise erwartet). Mit 0,54 liegt die prognostische Validität des Testwertes für den Notendurchschnitt ausserordentlich hoch und bestätigt die aus Deutschland bekannten Ergebnisse (dort Korrelation mit der Ärztlichen Vorprüfung, die in der Regel nach 4 Semestern abgelegt wird).

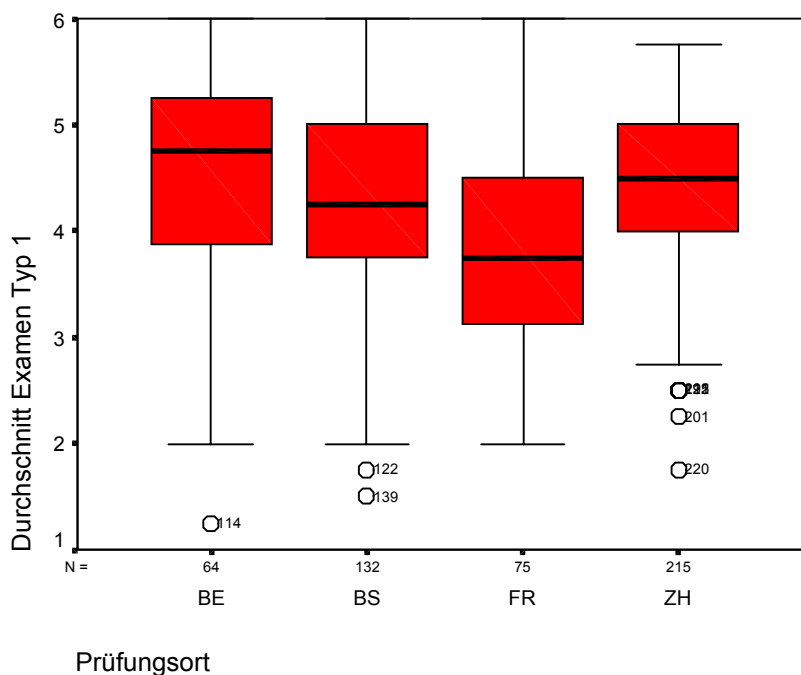


Abbildung 7: Boxplot der Durchschnittsnote für Examen Typ 1 nach Prüfungsorten

In Tabelle 8 werden die Korrelationen der Benotungen für die einzelnen Universitäten gegenübergestellt. Gemessen an der Erwartung für Koeffizienten prognostischer Validität sind alle im Bereich einer ausreichenden Höhe. Die „niedrigste“ Korrelation in Basel entspricht von Niveau her immer noch einer guten prognostischen Validität und lässt die Annahme eines gesicherten Zusammenhanges zwischen Noten und Testwert uneingeschränkt zu.

Die Tabelle enthält auch Angaben über die Art der Prüfung. Zwei Prüfungen werden mündlich, der Rest wird schriftlich abgenommen. Bei den schriftlichen Prüfungen gibt es Multiple-Choice-Aufgaben ebenso wie „Essay“-Aufgaben (Fragen sind durch kurze Texte zu beantworten), in einigen schriftlichen Prüfungen wird dies kombiniert.

Es deutet sich keine Regelmäßigkeit an, dass die Korrelationen auf die methodische Variante „Multiple-Choice“ zurückgeführt werden können (die bekanntlich auch im EMS eingesetzt wird). Vor allem in Zürich besteht die höchste Korrelation zu einer Prüfung vom „Essay“-Aufgabentyp, auch die kombinierten Prüfungen korrelieren nicht geringer mit dem Testwert als diejenigen, die nur Multiple-Choice-Aufgaben verwenden. In Basel schliesslich werden die drei „geringsten“

⁸ Trost, G. (Hrsg.) (1994). Test für Medizinische Studiengänge (TMS): Studien zur Evaluation (18. Arbeitsbericht). ITB: Bonn, Seite 144.

Korrelationen des Testwertes auch mit Multiple-Choice-Prüfungen gefunden. Der Frage wird in späteren Evaluationen natürlich weiter nachzugehen sein – es ergeben sich aber bisher keine Hinweise, dass es sich bei den Korrelationen von Studieneignung (Testwert) mit Wissen (Prüfung) um „artifizielle“ methodenbedingte Korrelationen handelt.

	Physik und Physiologie	Chemie und Biochemie	Allgemeine und Humanbiologie I	Allgemeine und Humanbiologie II	Prüfungsdurchschnitt
Bern (Ex. 1) (n=64)	SCH-EM	SCH-M	SCH-M	SCH-M	
	.64	.51	.48	.58	.59
Basel (n=132)	SCH-M	SCH-EM	SCH-M	SCH-M	
	.33	.35	.30	.35	.38
Freiburg (n=75)	MÜ	SCH-M	MÜ	SCH-M	
	.42	.52	.44	.46	.56
Zürich (n=215)	SCH-EM	SCH-EM	SCH-E	SCH-M	
	.51	.47	.57	.51	.60

Tabelle 8: Korrelationen des Testwertes mit den Prüfungsnoten (Examen Typ 1) für die Universitäten (n.s: nicht signifikant). SCH: schriftlich, MÜ: mündlich, E: Essay, M: Multiple Choice, EM: Kombination E und M

Höhere Korrelationen zwischen Testwert und Prüfungsnoten deuten auf einen grösseren Einfluss der individuellen Studieneignung auf den Prüfungserfolg hin – sehr hohe bedeuten aber auch, dass es weniger an Kompensationsmöglichkeiten gibt, eventuelle Defizite auf dem Gebiet der Fähigkeiten auszugleichen – etwa durch erhöhte Anstrengungen, Aufwendung von mehr Zeit (was beim Test bekanntlich nicht möglich ist) oder die Nutzung von Hilfe beim Wissenserwerb. Bei den Examen handelt es sich im Unterschied zum Test in der Regel um Wissensprüfungen, der Wissenserwerb sollte idealerweise so erfolgen, dass durch solche Kompensationsmechanismen trotz unterschiedlicher Ausgangsniveaus Wissen auf vergleichbaren Niveaustufen erworben wird. Insofern sollte eine „zu hohe“ Korrelation auch Denkanstoss sein, über optimale und individuell abgestimmte Fördermöglichkeiten nachzudenken.

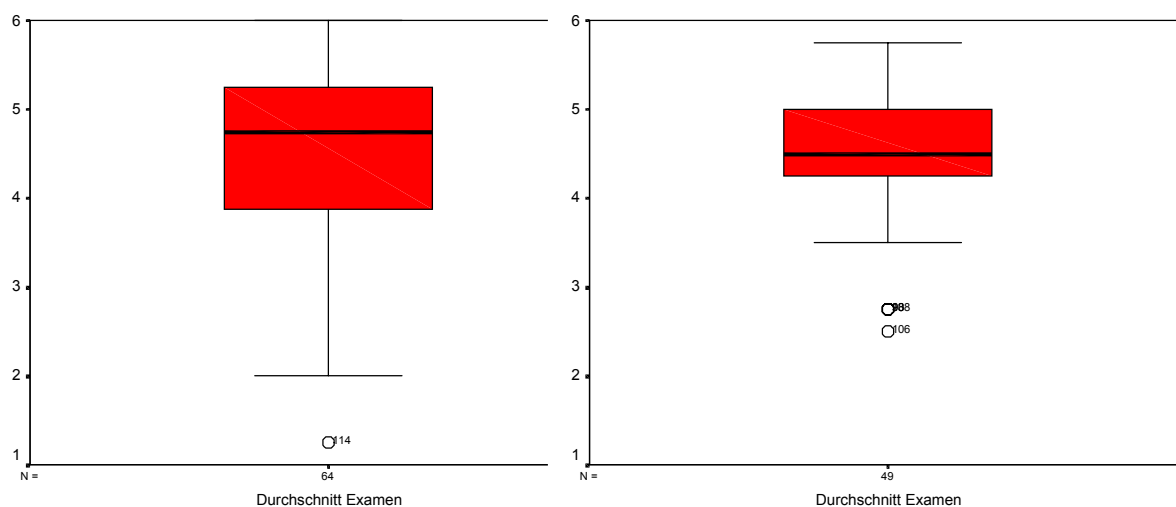


Abbildung 8: Boxplot für Examensdurchschnitte Typ 1 (links) und Typ 2 (rechts) für Universität Bern

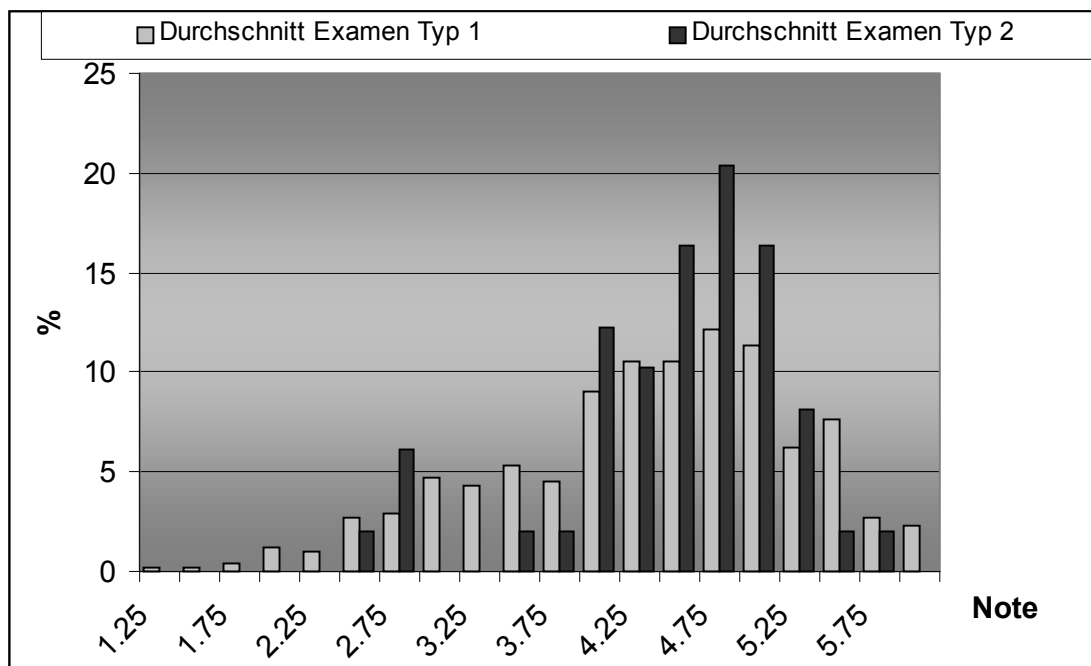


Abbildung 9: Verteilung der Prüfungsdurchschnittsnoten für die Examen Typ 1 (alle) und Typ 2 (Bern)

Zum Examen Typ 2 sei eine Anmerkung gestattet, weil sich hier ein Problem anzudeuten scheint. In Abbildung 8 (nur Bern) und 9 (Gesamt Typ 1 und Typ 2) werden die Verteilungen der Prüfungsdurchschnitte beider Examenstypen verglichen. Die Prüfungsdurchschnittsnoten des Examens Typ 2 treten gehäuft innerhalb von 1,5 Notenstufen auf und differenzieren deutlich weniger als die vom Examen Typ 1. Der EMS-Testwert der beider Gruppen in Bern unterscheidet sich dabei nicht (Examen 1 Mittelwert=103,3, Examen 2 Mittelwert=103,7), man muss also von vergleichbaren Fähigkeiten der Geprüften ausgehen.

Im Mittelwert unterscheiden sich Notendurchschnitte des Examenstyps 1 (4,3) nicht vom Examenstyp 2 (4,5). Dennoch ist die Bestehensquote für Examen Typ 2 mit 87,8% deutlich höher als Examen Typ 1 mit 71,8%. Hierfür können natürlich bessere Leistungen ebenso verantwortlich sein wie Probleme der Leistungsbewertung. Letztendlich wäre es eine erwartete Auswirkung der Zulassung nach der Studieneignung, dass alle Personen gute Leistungen erreichen und die Prüfung bestehen.

Da die Extremmeidung aber beide Extreme betrifft, scheint es sich um ein Problem der Notengebung zu handeln: Die Benotung im Examen Typ 2 könnte u.a. auf einen „Extremscheue-Effekt“ zurückgeführt werden, der eine Meidung schlechterer wie besserer Noten und eine bevorzugte Vergabe mittlerer Noten beinhaltet (siehe auch Tabelle 6). Bei neu eingeführten Notensystemen kann dies in der Anfangsphase auftreten, solange nur wenige Vergleichsmöglichkeiten existieren und/oder Misserfolge (für das Individuum wie das Ausbildungssystem) vermieden werden sollen. Wenn dies so bliebe, würde sich mangels einer Variation der Leistungsbewertung durch die Prüfung die Korrelationsanalysen mit dem EMS-Testwert erübrigen. Die Evaluation des EMS liesse sich unter diesen Umständen auf die Analyse des Bestehens der Prüfung beschränken.

Schlussfolgerungen:

Für die Evaluation des EMS bleibt festzuhalten:

1. Der Testwert **prognostiziert den Prüfungserfolg** mit sehr hoher Güte. Personen mit hohen Testwerten treten früher zur 1. Vorprüfung an und bestehen diese auch mit höherer Wahrscheinlichkeit im ersten Anlauf. Die Verwendung des EMS als Zulassungskriterium für den NC bietet die Gewähr, dass Personen bevorzugt einen Studienplatz erhalten, die das Studium wahrscheinlich schneller beenden (zunächst nur für die 1. Vorprüfung nachgewiesen). Damit können die vorhandenen Studienplatz-Ressourcen besser genutzt werden und letztendlich mehr Personen studieren.
2. Personen mit hohen Testleistungen erreichen mit hoher Wahrscheinlichkeit auch **bessere Prüfungs-Einzelnoten und Notendurchschnitte**. Geht man davon aus, dass die Prüfungsleistungen nebst weiteren Fähigkeiten und Fertigkeiten künftiger Ärzte für die Qualität der Berufsausbildung nicht unerheblich sind, so werden diese Leistungsniveaus angehoben.
3. Es finden sich Hinweise, dass Personen, die sich trotz zugeteiltem Studienplatz **nicht immatrikulierten, geringere Testwerte** erzielt haben als die erfolgreichen Personen (Prüfung im ersten Anlauf bestanden). Unter Berücksichtigung der Punkte 1 und 2 kann man dies – mit der gebotenen Vorsicht – als Hinweis interpretieren, dass die Studienmotivation und somit Dissuasionseffekte zumindest teilweise von der Studieneignung in gewünschter Weise beeinflusst sind und nicht unbedingt die Geeigneten „abgeschreckt“ werden.
4. Die insgesamt milde Wirkung des Numerus clausus innerhalb der Schweiz verhindert, dass bei der Zulassung eine Überbewertung des Eignungsaspektes erfolgt. Für den tatsächlichen Studienverlauf ist er jedoch in einer recht starken Art und Weise wirksam.

Die Eignung des EMS als Selektionsinstrument für die Studienzulassung kann aufgrund dieser Studie bestätigt werden.

Die Analyse der nachfolgenden Prüfungen und Jahrgänge ist geplant und wird diese Ergebnisse weiter differenzierbar machen.

Technische Anmerkungen:

Der Datenschutz bei dieser Studie erfolgte unter Beachtung der gesetzlichen Vorschriften: in den Auswertedateien wurden alle Personen-Identifikationsmerkmale gelöscht. Ergebnisse werden so dargestellt, dass ein konkreter Personenbezug nicht ersichtlich ist.

Die Erfassung der Prüfungsnoten und der Prüfungsergebnisse im BAG erfolgte ohne Kenntnis der Testleistungen im EMS. Es wurden durch die Schweizerische Hochschulkonferenz (SHK) lediglich die Immatrikulationsnummern der Studienanfänger mitgeteilt und daraufhin im BAG die zugehörigen Prüfungsdaten ermittelt. Die Datei mit den Prüfungsergebnissen wurde der SHK übergeben und erst dort mit den Testergebnissen im EMS verbunden. Diese Datei wurde anschliessend anonymisiert und sie liegt der SHK als Masterdatei auch zu Kontrollzwecken der Ergebnisse vor, um eine ausreichende Objektivität des Evaluationsverfahrens zu gewährleisten. Alle Berechnungen im ZTD erfolgten dann mit einer Kopie dieser Masterdatei.