

DIE ABSURDITÄT DER BESTEHENSREGEL FÜR DIE PRÜFUNGEN IN DER ÄRZTLICHEN AUSBILDUNG

Roland K. Fuchs, Wachtberg/Bonn

Kritisiert wird hier die Bestehensregel für die Prüfungen, die nach der Approbationsordnung für Ärzte abzulegen sind (§ 14 Abs.5 nach Artikel 1 der Dritten Verordnung zur Änderung der Approbationsordnung für Ärzte vom 15.07.1981. die am 1.August 1981 in Kraft getreten ist):

"(5) Die schriftliche Prüfung ist bestanden, wenn der Prüfling mindestens 60 vom Hundert der gestellten Prüfungsfragen zutreffend beantwortet hat oder wenn die Zahl der vom Prüfling zutreffend beantworteten Fragen die durchschnittliche Prüfungsleistung des jeweiligen Prüfungstermin im gesamtem Bundesgebiet um nicht mehr als 18 vom Hundert dieser durchschnittlichen Prüfungsleistung unterschreitet und nicht unter 50 vom Hundert der gestellten Fragen liegt."

Der Kern der Kritik richtet sich gegen die sogenannte "Gleitklausel" in dieser Bestehensregel, nach der die Bestehens-

grenze bei weniger als 60% der Anzahl der gestellten Aufgaben liegt, solange die Differenz "Durchschnittsleistung minus 18%" weniger als 60%, aber noch mindestens 50% der Aufgabenzahl ergibt.

Nehmen wir als Beispiel die Ärztliche Vorprüfung. Die Gesamtzahl der Aufgaben beträgt 320, 50% entsprechen 160 und 60% 192 Aufgaben. Wann wird diese Gleitklausel wirksam? Dann, wenn "Durchschnittsleistung minus 18%" (mathematisch gleichwertig, aber einfacher zu rechnen ist "Durchschnittsleistung mal 0,82") mindestens gleich 50% oder 160 und nicht größer als 191 ist. Wir bestimmen die Grenzen dieses Intervalls. Dabei muß man berücksichtigen, daß dann, wenn die Bestimmung solcher Grenzen anhand einer Rechenprozedur keine ganzzahligen Punktwerte ergibt, auf die nächsthöhere ganze Punktzahl aufgerundet wird. Denn der Medizinstudent kann nur mit dieser ganzen Punktzahl die rechnerisch ermittelte und um Bruchteile kleinere Grenzleistung erfüllen.

Untere Grenze der Durchschnittsleistung :	davon 82 %	ergibt :	aufgerundet zu :
193,91	mal 0,82	159,01 Punkte	160 Punkten
⋮			
Obere Grenze der Durchschnittsleistung :			
232,92	mal 0,82	190,99 Punkte	191 Punkten

Solange die Durchschnittsleistung in der Ärztlichen Vorprüfung also zwischen 193,91 und 232,92 Punkten liegt (einschließlich; Berechnung hier nur auf 2 Dezimalstellen genau), beläuft sich die Bestehensgrenze nicht auf 192 Punkte, sondern auf 191 oder weniger Punkte, minimal auf 160 Punkte. Dieser Bereich,

in dem diese "18% - Gleitklausel" wirksam wird, ist von hoher praktischer Bedeutung, wie die Ergebnisse der Ärztlichen Vorprüfung seit der Gültigkeit dieser Bestehensregel ab Herbst 1981 und auch die (hier nicht im einzelnen aufgeführten) Ergebnisse der ersten beiden Abschnitte der Ärztlichen Prüfung zeigen.

	Durchschnittsleistung :	Bestehensgrenze nach der "18 % - Gleitklausel" :
Herbst 1981	216,8	178
Frühjahr 1982	210,4	173
Herbst 1982	190,4	(160)*
Frühjahr 1983	186,3	(160)*
Herbst 1983	202,2	166
Frühjahr 1984	196,8	162
Herbst 1984	198,3	163

* Da in diesen beiden Fällen die Durchschnittsleistung unter dem o. a. unteren Grenzwert liegt, tritt hier die absolute Minimalzahl von 50 % richtigen Lösungen als Bestehensgrenze in Kraft.

In diesen Terminen mit Ausnahme von Herbst 1982 und Frühjahr 1983 wurde die Bestehensgrenze mit Hilfe der beschriebenen Gleitklausel festgelegt. Im Ersten Abschnitt der Ärztlichen Prüfung wurde die Bestehensgrenze in allen sieben Prüfungsterminen von Herbst 1981 bis Herbst 1984 nach der Gleitklausel bestimmt, im Zweiten Abschnitt nur in vier Terminen: im Frühjahr 1982, im Herbst 1982 und im Frühjahr 1984 lag die Durchschnittsleistung höher, so daß die absolute Anzahl von 60% richtiger Lösungen die Bestehensgrenze bildete, ebenso wie in allen schriftlichen Prüfungen des Dritten Abschnitts in diesem Zeitraum.

Warum, in welcher Hinsicht erweist sich nun diese variable Festlegung der Bestehensgrenze als absurd, die augenscheinlich doch den Prüflingen nur zum Wohle gereichen müßte? Warum und unter welchen Bedingungen können Medizinstudenten für vermehrte Anstrengungen bestraft werden, die ihnen zusätzliche Punkte bringen?

Das läßt sich nur anhand von konstruierten Beispieldaten erläutern, deren Zahl klein gewählt werden muß, damit das Ganze überschaubar wird. Wir nehmen wieder die Ärztliche Vorprüfung, die Anzahl der Prüflinge sei 20. Diese Prüflinge sind in den Beispielen mit den Buchstaben A bis T bezeichnet. Um

unterschiedliche Auswirkungen von Faktoren zeigen zu können, die die Leistung beeinflussen, müssen wir folgende gedankliche Hilfskonstruktion heranziehen:

Wir haben eine Gruppe (Stichprobe, Population) von Prüflingen, von der wir uns nur vorstellen, daß sie sich mit einer bestimmten Verteilung an Kenntnissen der Ärztlichen Vorprüfung unterziehen würde (findet also tatsächlich nicht statt). Diese mit "Gruppe 1" bezeichneten Prüflinge würden eine bestimmte Durchschnittsleistung M_1 erreichen. Nun stellen wir uns weiter vor, daß diese Gruppe vermehrte intensive Prüfungsvorbereitungen unternimmt und dadurch in der Lage ist, mehr Punkte zu erreichen als ihr ohne diese zusätzlichen Anstrengungen in der nun stattfindenden Prüfung zu erreichen möglich gewesen wäre. Ihre Ergebnisstruktur mit einer höheren Durchschnittsleistung M_2 wird hier als "Gruppe 2" bezeichnet. Um also die Auswirkungen zusätzlicher Prüfungsvorbereitungen sichtbar zu machen, vergleichen wir die bloß gedachten Ergebnisse einer Prüfungsgruppe ohne zusätzliche Lernanstrengungen -hier als Gruppe 1 bezeichnet- mit ihren tatsächlichen Leistungen in der Ärztlichen Vorprüfung, nachdem sie eine Anzahl zusätzlicher Prüfungsvorbereitungen unternommen hat -hier als Gruppe 2 bezeichnet-

net-. Wem dieser Gedankengang zu künstlich oder unverständlich erscheint, der kann sich die beiden Gruppen auch als reale Prüfungspopulationen mit gleicher Fähigkeitsverteilung vorstellen, wobei

Gruppe 2 über zusätzliche Lernerfahrungen gegenüber Gruppe 1 verfügt. Für die Logik der Argumentation und die Schlußfolgerungen macht dies keinen Unterschied.

Beispielfall 1 :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
<u>Gruppe 1</u>	92	132	150	161	163	164	165	167	169	170	171	205	208	223	241	256	269	275	276	279
	+12	+12	+12	+12	+12	+12	+12	+12	+12	+12	+12	+12	+12	+12	+11	+11	+11	+11	+11	+11
<u>Gruppe 2</u>	104	144	162	173	175	176	177	179	181	182	183	217	220	235	252	267	280	286	287	290

Betrachten wir zunächst den Beispielfall 1, Gruppe 1 (also die gedachte Leistung einer Prüfungspopulation ohne vermehrte Prüfungsanstrengungen, wenn sie sich zu diesem Zeitpunkt der Prüfung unterziehen würde): Ihre Durchschnittsleistung wird hier mit $M_1 = 196,8$ richtigen Lösungen (= Punkten) angenommen und liegt damit in dem Intervall, für das die Bestehensgrenze nach der "Gleitklausel" berechnet wird. Diese beläuft sich in diesem Fall auf 162 Punkte. Damit hätten in Gruppe 1 die Prüflinge A - D nicht bestanden - genau 20%. Diese Zahlenverhältnisse sind denen bei der Frühjahrsprüfung 1984 nachempfunden: $M_{V-F 84} = 196,8$, die Bestehensgrenze 162 Punkte, keinen Prüfungserfolg hatten 19,7% der Prüflinge. Nach zusätzlichen Lernanstrengungen und umfangreichen weiteren Prüfungsvorbereitungen legt diese Gruppe nun die Ärztliche Vorprüfung ab, und erzielt insgesamt ein besseres Ergebnis: Ihr Durchschnittsergebnis beträgt $M_2 = 208,5$ Punkte, so daß die Bestehensgrenze wieder mit Hilfe der Gleitklausel bestimmt worden ist. Auch diese angenommenen Zahlen entsprechen in etwa einer realen Prüfung, nämlich der Frühjahrsprüfung 1982, die eine Durchschnittsleistung von $M_{V-F 82} = 210,4$, eine Bestehensgrenze von 173 Punkten und eine Mißerfolgsquote von 17,2% aufwies.

Der Leistungszuwachs der Gruppe 2 im Beispielfall 1 von 234 Punkten ist möglichst gleich auf alle 20 Prüflinge verteilt worden: die ersten 14 haben je 12 Punkte, die letzten 6 Prüflinge je 11 Punkte zusätzlich erhalten. Daß dieser Leistungszuwachs von durchschnittlich 11,7 Punkten beträchtlich ist, kann m.E. gar nicht bezweifelt werden: er beträgt immerhin das 11,7 - Fache (!) der Einheit, die über Bestehen und Nicht-Bestehen der gesamten Prüfung entscheidet!

Durch den gleichmäßigen Punktezuwachs bei allen Prüflingen - eine für die Prüfungspraxis nicht gerade sehr realistische Konstruktion - hat diesmal Prüfling D Glück gehabt: er liegt mit seiner Leistung gerade noch über der (neuen) Bestehensgrenze von 171 Punkten. Damit haben nur die Kandidaten A bis C in Gruppe 2 nicht bestanden: die Mißerfolgsquote ist nun auf 15% gesunken. Das alles entspricht sicher landläufigen Vorstellungen wie es sich mit diesen Prüfungen verhält bzw. verhalten sollte: wenn die erreichten Punktzahlen ansteigen, liegt eine leistungsstärkere Gruppe vor, logischerweise fallen auch weniger durch! Damit scheint die Welt doch in Ordnung zu sein - oder etwa nicht? Wo liegt der Hund begraben?

Wer nur ein wenig Ahnung von Wesen und Aufbau und der Funktionsweise solcher Leistungsprüfungen hat, oder wer sich gar schon einmal selbst damit beschäftigt hat, der wird sofort bemerken, daß wir stillschweigend und unreflektiert immer von "Leistungszuwachs" gesprochen haben. Welche fundamentale Annahme steckt da aber als notwendige Voraussetzung drin? - Sehr richtig: die Prüfungen müssen vergleichbar sein, sie müssen gleichwertig sein, und zwar bezüglich ihrer Schwierigkeit. Denn ohne weitere Informationen ist im Beispielfall 1 nicht entscheidbar, ob die Gruppe 2 tatsächlich leistungsfähiger als die Gruppe 1 ist, oder ob die Prüfung der Gruppe 2 leichter war als die der Gruppe 1. Und genau das ist die Situation bei den Prüfungen nach der Approbationsordnung für Ärzte, die vom Mainzer Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) ausgearbeitet werden. Trotz umfangreicher Recherchen ist es mir nicht gelungen, eine einzige (!) wissenschaftliche Arbeit des IMPP aufzutreiben, in der diese fundamentale Frage der Vergleichbarkeit der Prüfungsfragen unter dem Gesichtspunkt ihrer Schwierigkeit untersucht wird. Wird es da nicht höchste Zeit, daß das IMPP solche Arbeiten durchführt und auch veröffentlicht, wie das im "Abkommen über die Errichtung und Finanzierung des Instituts für medizinische und pharmazeutische Prüfungsfragen" verankert ist? Darin heißt es in Artikel 2:

"(2) Das Institut leistet im Rahmen dieses Abkommens entsprechend seinen Möglichkeiten einen Beitrag zur angewandten Forschung auf dem Gebiete der Methodik des Prüfungswesens. Es unterrichtet die obersten Gesundheitsbehörden der Länder und die für das Hochschulwesen zuständigen Minister (Senatoren) der Länder laufend über die für Reformen des Prüfungswesens relevanten Ergebnisse der Arbeit."

Diese Notwendigkeit der angewandten Forschung auf dem Gebiet der Methodik des Prüfungswesens wird hinsichtlich der Schwierigkeit des Prüfungswesens und damit folgerichtig bezüglich ihrer Vergleichbarkeit im Sinne des Nachweises ihrer Gleichwertigkeit umso dringlicher, wenn man bedenkt:

- Die Angabe eines globalen, gemittelten Schwierigkeitswertes für eine Prüfung kann nichts anderes als Augenwischerei sein, wenn sonst keine Informationen geliefert werden. Beispiel: Mitgeteilt wird der mittlere Prozentwert für eine Prüfung, der gleich dem Mittelwert der Schwierigkeitswerte der einzelnen Prüfungsaufgaben ist, mit beispielweise 0,50. Dieser Wert von 0,50 kann zustande gekommen sein durch eine Mittelung von 0,97 und 0,03. Dann ist die Bezeichnung dieser Prüfung als von mittlerer Schwierigkeit ebenso absurd wie es die Prüfung selbst ist: sie besteht nämlich aus einer Aufgabe, die fast alle (97%) lösen - eine solche Aufgabe prüft nicht mehr irgendeine Form der Leistungsfähigkeit - und aus einer Aufgabe die von fast niemandem (gerade nur 3%) gelöst werden kann - eine solche Aufgabe prüft auch keine Form der Leistungsfähigkeit mehr, da sie bei fast allen nicht vorhandenes Spezialwissen voraussetzt (es sei denn, es ist nachgewiesen, daß diese 3% Lösungen nicht durch Zufall, sondern tatsächlich von den 3% Allerbesten erreicht wurden). Gefordert werden muß dagegen, daß der mittlere Wert von 0,50 aus Schwierigkeitswerten der Einzelaufgaben um 0,49 und 0,51 oder um 0,48 und 0,52 und ähnlich zustande gekommen ist. Wann setzt das IMPP dieses um, und belegt, daß der für eine Prüfung ausgewiesene Schwierigkeitsgrad tatsächlich weitgehend den Einzelschwierigkeitswerten der Prüfungsaufgaben entspricht? Denn nur dann ist eine Voraussetzung geschaffen, um ihn als echten Schwierigkeitsgrad dieser

Prüfung ansehen zu können und auszuschließen, daß er nur "künstlich" als Ergebnis einer Mittelwertsbildung aus wild gemischten Einzelwerten zustande gekommen ist, die man mit Fug und Recht gar nicht so zusammenfassen dürfte!

- Auch zahlenmäßig gleiche Schwierigkeitswerte für Prüfungsaufgaben aus verschiedenen Prüfungen bedeuten nicht selbstverständlich schon das Gleiche hinsichtlich der Leistungsfähigkeit, wenn nicht weitere Informationen über die Populationen der Prüflinge vorliegen. Dieser Punkt soll hier jedoch wegen seiner Wissensvoraussetzungen im Prüfungswesen nicht weiter ausgeführt

Beispielfall 2 :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
<u>Gruppe 1</u>	92	132	150	161	163	164	165	167	169	170	171	205	208	223	241	256	269	275	276	279
	+ 55	+ 36	+ 7	+ 6	+ 2	+ 5	+ 3	+ 2	+ 1	+ 0	+ 46	+ 23	+ 27	+ 13	+ 3	+ 2	+ 1	+ 1	+ 1	+ 0
<u>Gruppe 2</u>	147	168	157	167	165	169	168	169	170	170	217	228	235	236	244	258	270	276	277	279

Die Gruppe 1 ist die gleiche wie im Beispielfall 1, sie stellt wieder die gedachte, bloß vorgestellte Ausgangssituation für unsere Überlegungen dar. Wir erinnern uns: Wir nehmen an, diese Gruppe macht nach zusätzlichen Vorbereitungen, die aus vermehrtem Lernen von Altfragensammlungen, intensiviertem Lehrbuchstudium und Einübung m.c. spezifischer Lösungsstrategien bestehen, die Ärztliche Vorprüfung und erreicht Ergebnisse, deren Struktur als Gruppe 2 dargestellt ist. Der Gesamtwachsbetrag beläuft sich wieder auf 234 Punkte, so daß auch hier wieder die Durchschnittsleistung $M_2 = 208,5$ Punkte ausmacht und die anhand der Gleitklausel bestimmte Bestehensgrenze 171 Punkte. Aber man sehe und staune: wider alles Erwarten beträgt auf einmal die Mißerfolgsquote exakt 50%! Wie ist das nur möglich? Kann das noch mit rechten Dingen zugehen?

werden. Doch auch hier muß mit Nachdruck gefragt werden: Wann veröffentlicht das IMPP Belege bzw. Nachweise anhand der entsprechenden Verfahren darüber, daß die Prüfungsaufgaben auch tatsächlich gleichwertig sind und nicht nur den Augenschein gleicher Zahlenwerte aufweisen? Schließlich hat jeder Prüfling das verbrieftete Recht darauf, nicht Prüfer- oder Prüfungswillkür ausgesetzt zu sein, sondern unabhängig vom Prüfungstermin eine echte Gleichbehandlung zu erfahren!

Nun aber endlich zum Kernstück der Kritik: Betrachten wir den Beispielfall 2.

Die Ursache liegt in der andersartigen Aufteilung des Gesamtwachsbetrages auf die einzelnen Prüflinge, die jedoch im Gegensatz zu der gleichmäßigen Aufteilung im Beispielfall 1 bedeutend mehr der realistischen Situation in der Prüfungspraxis entsprechen dürfte und dazu noch am ehesten als Zeichen der Validität der Prüfung angesehen werden könnte. Denn die relativ schwachen Prüflinge, hier nur durch A und eventuell noch B vertreten, machen wohl grundsätzlich etwas falsch bzw. haben fundamentale Lücken. Sie profitieren aus den Zusatzanstrengungen am meisten. Entsprechend weniger können sich die Kandidaten C bis J steigern, da sie bereits dicht unterhalb ihrer "Kapazitätsgrenze" angelangt sind (insofern würde sich die Prüfung als valide erweisen). In der Gruppe der Erfolgreichen wiederholt sich das Muster des Zuwachses im Prinzip wieder: die es gerade

eben oder auch mit sicherem Punktepolster geschafft haben, haben es mit Sicherheit an Fleiß, Anstrengungsbereitschaft oder Strategieeinsatz - gemessen an ihrer "eigentlichen" Kapazität - fehlen lassen, so daß bei ihnen die größten individuellen Steigerungsbeträge möglich sind. Je weiter man jedoch nach oben auf dem Leistungskontinuum kommt (etwa ab Prüfung 0), desto schwieriger werden naturgemäß bei dieser bereits vorhandenen Leistungshöhe zusätzliche Steigerungen.

Nochmals einige zusätzliche Anmerkungen zu den hier angestellten Modellüberlegungen, damit keine Mißverständnisse entstehen oder unausgeräumt bleiben:

Die Situation die wir in den beiden Beispielsfällen unter "Gruppe 1" verstehen, wird in der Realität natürlich gar nicht vorhanden sein. Sie stellt für uns nur eine Denkhilfe dar, um eine bestimmte Leistungssituation zu veranschaulichen und um uns klar zu machen, wie sich dem gegenüber zusätzliche Lernanstrengungen auswirken würden. Und wichtiger: wie diese zusätzlichen Lernanstrengungen im Rahmen dieser Argumentation sichtbar und damit faßbar i.e.S. gemacht werden können. Ebenso willkürlich ist natürlich auch die Höhe des hier verwendeten Gesamtsteigerungsbetrages von 234 Punkten, der genau zur Hälfte auf die 10 erfolgreichen und auf die 10 nicht erfolgreichen Prüflinge aufgeteilt worden ist. Es dürfte jedem einsichtig sein, daß es neben dem gewählten noch eine große Menge an Verteilungsmustern gibt, die zum gleichen Ergebnis führen (z. B. Steigerungen nur in der Gruppe der nicht Erfolgreichen bei geringerem Gesamtbetrag, oder Steigerungen nur in der Gruppe der Erfolgreichen). Und selbstverständlich müssen die Unterschiede zwischen den individuellen Zuwachsbeträgen nicht so kraß ausfallen, wie dies hier zur Verdeutlichung konstruiert worden ist.

Und nun zum Ergebnis und seinen Konsequenzen: Ich halte es für absurd und kann mir nicht vorstellen, daß vom Verordnungsgeber die Wirkung der Gleitklausel so intendiert war, daß die Medizinstudenten für vermehrte Anstrengungen geradezu bestraft werden - statt 20% fallen 50% durch! Schauen wir uns nochmals Beispielsfall 2 genau an: gegenüber der (nur vorgestellten) Situation ohne zusätzlichen Anstrengungen, in der nur 20% durchgefallen wären (hätten sie zu diesem Zeitpunkt und bei diesem Wissensstand die Prüfung angelegt), fallen mit diesen zusätzlichen Lernanstrengungen 50% durch- zusätzlich die Prüflinge E - J. Und das, obwohl bis auf den Prüfling J alle anderen Prüflinge in der Gruppe der nicht Erfolgreichen einen zum Teil erheblichen Punktezugewinn zu verzeichnen haben. Und das ist doch absurd- bessere Leistungen und dafür, ja sogar zum Teil dadurch verursacht, erfolglos! Das ist die Auswirkung der Gleitklausel in Verbindung mit dem Pech für die jetzt erfolglosen Prüflinge E - J, dem Pech nämlich, daß sich auch die Prüflinge A - D und die Prüflinge K - S zum Teil wesentlich verbessert haben. Und zwar haben die sich so stark verbessert, daß deren Verbesserungen zusammen mit ihren eigenen Verbesserungen die Durchschnittsleistung und die über die Gleitklausel daran gekoppelte Bestehensgrenze so weit nach oben getrieben haben, daß die Verbesserungen der Prüflinge E - I (J weist im Beispielsfall keine Verbesserung auf) nicht mehr ausreichen, um trotz verbesserter Leistungen über die nun gültige Bestehensgrenze zu kommen! Absurd! Es wird damit zu einer Sache von Glück und Pech, also des Zufalls, ob man eine Prüfung besteht, deren Bestehensgrenze über eine solche Gleitklausel festgelegt wird. Das Glück bzw. das Pech besteht nämlich darin, daß das eigene Bestehen- zumindest wenn man

mit seiner Leistung im Umfeld der Bestehensgrenze liegt- abhängig wird von den anderen, die (zufällig?) mit einem die Prüfung ablegen. Gibt es sehr viele gute Prüflinge und wenig sehr schwache, stehen für einen Prüfling im Bereich um die Bestehensgrenze die Chancen sehr schlecht. Ein solcher Prüfling müßte sich eben wünschen, daß sehr viele andere Prüflinge mit ganz geringen Punktzahlen teilnehmen und damit Durchschnittsleistung und Bestehensgrenze wesentlich drücken. Sollten etwa die ausländischen Studenten mit ihren um Vielfaches geringeren Leistungen in diesen für sie mit verbalen Spitzfindigkeiten gespickten Prüfungen und die Prüfungswiederholer diese Funktion...?- Ich wage nicht, weiterzudenken.

Noch von einer anderen Seite her erhalten die hier skizzierten Überlegungen über die absurden Auswirkungen der Gleitklausel Bestätigung. Und zwar durch die vom IMPP geübte Praxis, bei offensichtlich fehlerhaft konstruierten oder gedruckten Prüfungsaufgaben, wenn sich dies nach der Prüfungsdurchführung herausstellt, diese Prüfungsaufgaben allen Prüfungsteilnehmern als gelöst anzuerkennen und die entsprechende Punktzahl anzurechnen. Gegen diese Praxis haben Medizinstudenten verschiedentlich Klagen erhoben, zu Recht muß man wohl sagen. Und auch die Rechtsprechung scheint diese Praxis in der letzten Zeit als unzulässig anzusehen, wenn auch aus anderen Gründen. Was passiert hier durch die Anwendung der Gleitklausel?

Nehmen wir den Fall eines Medizinstudenten, der mit 166 Punkten genau die mit Hilfe der Gleitklausel festgesetzte Bestehensgrenze von ebenfalls 166 Punkten erreicht um damit die Ärztliche Vorprüfung (zunächst) bestanden hat. Nach dieser ersten Auswertung stellt sich jedoch heraus, daß dieser Student eine dieser 166 Aufgaben zwar so beantwortet

hatte, wie das IMPP dies als richtige Lösung vorgesehen hatte, diese Aufgabe aber offensichtlich eine Doppellösung besitzt, so daß das IMPP sich entschließt diese Aufgabe allen Prüfungsteilnehmern als gelöst anzurechnen. War die Aufgabe vorher von z.B. 40% aller Teilnehmer gelöst worden, führt die Erhöhung ihrer Lösungshäufigkeit auf 100% zwangsläufig zu einer Erhöhung der Durchschnittsleistung (hier um 0,6 Punkt von z.B. 202,12 auf 202,72 Punkte). Damit erhöht sich aufgrund der Gleitklausel aber auch die Bestehensgrenze von 166 auf 176 Punkte (exakt 165,74 und 166,23, die jeweils auf die nächste ganze Punktzahl aufgerundet werden müssen). Und damit ist - wiederum vollkommen absurd - unser Medizinstudent durchgefallen! Absurd deshalb, weil es einem Prüfling zum persönlichen Pech gereicht, daß das IMPP einen Fehler gemacht hat, in Verbindung mit der fatalen Gleitklausel. Und zuweilen erkennt das IMPP in einer Prüfung mehr als eine Aufgabe allen Prüflingen als gelöst an ...!

Ich fasse als Fazit zusammen:

1. Es muß als absurd bezeichnet werden, daß die Medizinstudenten aufgrund der geltenden Bestehensregel - und innerhalb dieser wegen der sog. Gleitklausel - für erhöhte Leistungen bestraft werden, indem sie durchfallen, da sie trotz ihrer Leistungserhöhungen die erhöhte Bestehensgrenze nicht erreichen können. Wie in ausführlichen Modellüberlegungen gezeigt worden ist, passiert das z.B. dann, wenn wenige schwache oder viele leistungsstarke Kandidaten die Prüfungsgruppe bilden. Solche Konsequenzen können nicht im Sinne des Verordnungsgebers sein.
2. Bei einem Medizinstudenten mit einer Punktzahl zwischen 50 und 60% - in diesem Intervall wird die Bestehensgrenze über die sog. Gleitklausel

"Durchschnittsleistung minus 18%" festgelegt - kann es vom Glück oder Pech abhängen, ob er die Prüfung besteht. Sein Glück bzw. Pech besteht nämlich darin, wer mit ihm zusammen die Prüfung macht: sind es viele besonders leistungsstarke Kandidaten, die die Durchschnittsleistung und damit die angekoppelte Bestehensgrenze hochtreiben, hat er Pech und fällt unter diese (erhöhte) Bestehensgrenze und damit durch die Prüfung; sind es besonders viele leistungsschwache Kandidaten, hat er Glück: sie drücken Durchschnittsleistung und Bestehensgrenze nach unten, so daß er die Prüfung besteht! Damit in diesem Sinne die Zufälligkeit in der Zusammensetzung der Leistungspopulation kontrollierbar wird, muß das IMPP nachweisen, daß die Leistungsverteilungen der Medizinstudenten in den von ihm ausgearbeiteten Prüfungen nicht beliebig von unkontrolliert gemischt-schwierigen Aufgaben abhängen, sondern weitgehend nur von den Fähigkeitsverteilungen in den Studentenpopulationen. Damit wird die kritisierte Abhängigkeit des Prüf-

ungserfolges von der jeweiligen Zusammensetzung der Studentenpopulation eindeutig faßbar!

3. Unerläßliche Voraussetzung für diese Kontrollierbarkeit der Leistungsverteilungen ist die Vergleichbarkeit der einzelnen Prüfungen zu den aufeinander folgenden Terminen hinsichtlich ihrer Schwierigkeit. Das IMPP hat die Pflicht, diese Vergleichbarkeit anhand der Schwierigkeiten der einzelnen Aufgaben nachzuweisen. Denn die zu den verschiedenen Zeitpunkten sich der Prüfung stellenden Medizinstudenten haben ein verbrieftes Recht auf Gleichbehandlung hinsichtlich der Schwierigkeit ihrer Prüfung und damit bezüglich der Chance, die Prüfung erfolgreich abzulegen. Keine einzige Publikation des IMPP liegt vor, die auf diese drängenden grundsätzlichen Fragen Antwort gibt. Dabei verfügt das IMPP im Jahre 1984 über 10 (!) Jahre "Prüfungserfahrungen" - alles vertane Zeit im Hinblick auf eine fundierte Weiterentwicklung der Methodik des Prüfungswesens?

Dipl:Psychol. R. K. Fuchs, Compbachweg 35
D 5307 Wachtberg/Pech