

gentlich waren es unter 16 Prozent, doch weil man durch die fiktive Leistungsgutschrift die Gleitklausel mit dem 50-Prozent-Anker auch noch benutzte, wurden es 18,7 Prozent).

Damit war der Anker gelichtet: zum Bestehen der Prüfung mit 320 Fragen genügten jetzt konkret real nur noch 135 richtige Antworten oder 42 Prozent anstatt der rechtlich verankerten Mindestleistung von 50 Prozent (= 160 richtige Antworten), fiktiv real sind es 164 = 51 Prozent.

Ein weiterer Effekt des Bonus von 29 Punkten liegt im Bereich des Möglichen: ein neuer Rekord an „Einserkandidaten“. Wie dem auch sei, im System der staatlichen Prüfungen besteht ein Circulus vitiosus für Ungereimtheiten.

Wie war die Frühjahrsprüfung vor dem Bonus?

Am schlechtesten schnitten die 86 Prüfungsteilnehmer in Hannover ab mit einer Durchfallquote von rund 80 Prozent, am besten die 184 in Würzburg oder die 343 in Heidelberg mit jeweils 25 Prozent. In Hannover waren etwa 60 Prozent Wiederholer, in Würzburg rund 10 Prozent, in Heidelberg 13 Prozent. Das gibt natürlich Anlaß, nachzudenken. Die Wiederholer waren nicht nur in Hannover am stärksten vom erneuten Mißerfolg betroffen. Wie immer der Schwierigkeitsgrad der Prüfung nach ihrer Standardisierung im Vergleich zu den früheren zu bewerten gewesen wäre, was die Selektion als solche anbelangt, dürfte die nicht korrigierte Prüfung kaum aus dem Rahmen gefallen sein.

Außenstehende können verlässlich nichts belegen, was die Frage der Prüfungsqualität und der Prüfungsqualität des Frühjahrstermins betrifft. Der Schlüssel zur Erkenntnis liegt im Institut für medizinische und pharmazeutische Prüfungsfragen.

Wie geht es wohl weiter?

Wenn man nur den politisierenden Effekt der Durchfallquote im Blick hat, dann war der Bonus von 29 richtigen Antworten zu hoch, d. h. die 18,7 Prozent Mißerfolgsrate entschieden zu niedrig. In der Frühjahrsprüfung 1983 lag die Quote bei 23,6 Prozent. Es spricht nichts dagegen, daß sie schon in einer der nächsten Prüfungen zwischen 25 und 30 Prozent liegen kann. Besteht dann oder schon bei 20 Prozent ein Anspruch auf einen Bonus? Ist dies dann der Zeitpunkt, daß die IMPP-Leitung das erhält, was sie, bisher noch vergebens, gefordert hat: die Gleitklausel ohne Anker?

Oder wird eine ständige Kommission von Professoren gebildet, mit der Aufgabe, das Prüfungsergebnis aufzubessern, wenn die Durchfallquote zu hoch erscheint. Beide Möglichkeiten würden gesetzlich nicht vorgesehene Gnadenakte entbehrlich machen. Aber wem ist dann geholfen? Effizienz der Ausbildung und der Prüfung wären auf eine Leerformel reduziert.

Literatur

(1) Harth, O.: Ausbildungskapazität – ein destruierender Begriff für das Medizinstudium. Deutsches Ärzteblatt 81 (1984) Heft 39, S. 1–3; (2) Harth, O.: I. Viele Details, kein Überblick. Kapazitätsproblematik und Prüfungssystem bedingen die Misere. DUZ 40, Nr. 19 (1984) 21–24; (3) Harth, O.: II. Triumph des Details. DUZ 40, Nr. 20 (1984) 15–17; (4) Golenhofen, K.: Bocksprünge auf Mainzer Art. Das IMPP hat aus den Fehlern der Vergangenheit nichts gelernt. DUZ Nr. 14 (1985) 15–18; (5) Golenhofen, K.: Arbeitsbericht der Sachverständigenkommission zur Prüfung der Gründe für das Ergebnis der Ärztlichen Vorprüfung im März 1981. DUZ 38, Nr. 3 (1982) 25–29.

Anschrift des Verfassers:
Prof. Dr. med. Otto Harth
Physiologisches Institut
der Universität
Saarstraße 21
6500 Mainz

Die wirklich ungeeigneten Fragen wurden gar nicht eliminiert

Zur Nachbesserung der ärztlichen Vorprüfung, Frühjahr 1985

Klaus Hinrichsen

*„Ist dies schon Tollheit, hat es doch Methode“
Shakespeare, Hamlet, 2.2*

Schon bald nach Bekanntwerden der Ergebnisse der ärztlichen Vorprüfung vom Frühjahr 1985 („Katastrophen-Physikum“) habe ich mich öffentlich dafür eingesetzt, die Prüfungsfragen durch eine Gruppe unabhängiger Sachverständiger kritisch zu überprüfen (Deutschlandfunk, 10. Mai 1985, Erstes Deutsches Fernsehen, 23. Mai 1985). Ich habe die dem Deutschen Ärztetag vorgelegte Resolution (Travemünde, 18. Mai 1985; dazu Heft 22/1985) und die des Medizinischen Fakultätentages (Ulm, 8. Juni 1985) mit vorformuliert und nachhaltig unterstützt. Dabei schwebte mir etwas vor, wie es die Schweiz seit Oktober 1983 praktiziert: Die Examinatoren legen die Bewertungskriterien fest, nach denen die Noten erteilt werden; dabei können die Fragen verschieden gewichtet werden. Lassen die Fragen oder Antworten einen offensichtlichen inhaltlichen oder formalen Mangel erkennen, so werden sie bei der Bewertung nicht berücksichtigt (Verordnung über die Einzelheiten des Verfahrens bei den eidgenössischen Medizinalprüfungen vom 30. Juni 1983). Die Ex-

aminatoren also, nicht das dienstleistende Institut (im Falle der Schweiz das „Institut für Ausbildungs- und Examensforschung der Medizinischen Fakultät der Universität Bern, IAE“), entscheiden über die Bestehensgrenze und über die Notengrenzen.

Was bei uns jetzt geschehen ist, hat mit dem ursprünglichen Vorschlag und mit der eidgenössischen Regelung nur wenig zu tun. Entschiedene Kritik ist an dem Arbeitsergebnis der vom Bundesrat beauftragten Kommission ebenso zu üben wie an der Umsetzung dieses Ergebnisses in der „Nachbesserung“.

Das Kommissionsergebnis

Bei der seit 1973 angewandten Prüfungsmethode nach dem Antwort-Wahl-Verfahren (Multiple choice) werden in der in der Bundesrepublik praktizierten Form fünf Wahlantworten vorgegeben, von denen jeweils nur eine als zutreffend auszuwählen ist. Würde ein Laie ohne jede Vorkenntnis die Antworten zufällig verteilen, käme er auf möglicherweise 20 Prozent richtige Antworten (Rate-Wahrscheinlichkeit). Medizinstudenten mit Vorkenntnissen können nicht mehr „zufällig“ antworten; daran hindern sie ihre Kenntnisse. Aber auch ihnen bleibt eine Rate-Wahrscheinlichkeit – vermutlich etwas niedriger als 20 Prozent –, wenn sie den Empfehlungen im Vorspann des Prüfungsheftes folgen, keine Frage auszulassen, sondern als Ultima optio zu raten. – Die Fragen aller Prüfungen des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) in Mainz werden nach der Prüfung teststatistisch ausgewertet. Die Item-Analyse weist für jede gestellte Frage aus, wie viele Teilnehmer sich für welche der fünf Antworten entschieden haben, sie gibt Schwierigkeitsgrad und Trennschärfe an. Post festum ist man also klüger, wenn es auch keine fe-

17 Welche Zyklusphase des Endometriums ist unter anderem durch gestreckte tubuläre Drüsen charakterisiert?

(A) Desquamations-Regenerationsphase
(B) frühe Proliferationsphase
(C) frühe Sekretionsphase
(D) späte Sekretionsphase
(E) Ischämiephase

121 Die Gedächtnisforschung hat unter anderem untersucht, ob ein Zusammenhang zwischen dem Behalten von Wörtern und der emotionalen Bedeutung dieser Wörter besteht.

Hierbei ergab sich:

(1) Neutrales wird besser behalten als Unangenehmes.
(2) Angenehmes wird besser behalten als Unangenehmes.
(3) Sowohl Angenehmes wie Unangenehmes werden besser behalten als Neutrales.
(4) Unangenehmes wird besser behalten als Angenehmes.

(A) nur 1 ist richtig
(B) nur 3 ist richtig
(C) nur 1 und 4 sind richtig
(D) nur 2 und 3 sind richtig
(E) nur 3 und 4 sind richtig

Geeignet oder nicht geeignet? Hier zwei der insgesamt 29 Fragen, die im nachhinein ausgeschieden wurden (eine Frage an den Leser: Hätten Sie's gewußt?)

sten Werte gibt, wann eine Frage nachträglich als geeignet und wann als ungeeignet eingestuft werden müßte.

Die im Auftrage des Bundesrates tätig gewordene Kommission hat 29 Fragen aus der Ärztlichen Vorprüfung vom Frühjahr 1985 für ungeeignet erklärt. Kriterien waren nicht die teststatistischen Werte, sondern die für den Arzt allgemein erforderlichen Kenntnisse (§ 14 Approbationsordnung), also die Relevanz der Fragestellung, die Zumutbarkeit des Frageninhaltes in bezug auf den Ausbildungsstand, die Deutlichkeit oder Eindeutigkeit der Formulierung

der vorgegebenen Antworten. Bei der Beurteilung der verwendeten Fragen ging es dabei meist um die Frage, ob außer der von den Sachverständigen vorgeschlagenen, letztlich vom IMPP festgelegten einen richtigen Antwort auch eine der Distraktoren genannten Falschantworten (auch/oder doch ein bißchen) zutreffend sein könnte. – Die Ergebnisse der Kommission sind in der Tat überraschend. Während z. B. 77 Prozent der Prüfungsteilnehmer wußten, daß Monozyten Makrophagen sind, und nur 2 Prozent der Prüfungsteilnehmer die falsche Aussage, kleine Alveolarepithelzellen (nicht die Alveolarmakrophagen!)

seien Makrophagen, für richtig hielten, schloß sich die Kommission diesen 2 Prozent an und erklärte die Frage für ungeeignet. Wie in diesem Beispiel wurden andere Fragen für ungeeignet gehalten, bei denen 93 Prozent, 84 Prozent oder 75 Prozent der Prüfungsteilnehmer die richtige Antwort wählten, während die Kommission glaubte, den 2 Prozent, 4 Prozent oder 9 Prozent der Prüflinge folgen zu sollen, die eine bestimmte, von der Kommission auch für möglich gehaltene andere Antwort für richtig hielten.

Richtig oder falsch ist in Sachfragen keine Mehrheitsfrage. Soll eine Prüfung auch in der Leistungsspitzengruppe hinreichend differenzieren, muß sie auch schwierige Fragen enthalten, die nur von wenigen Teilnehmern richtig beantwortet werden. Soll aber eine Prüfung das Ziel erreichen, für die überwiegende Zahl der Teilnehmer eine möglichst weitgehende Übereinstimmung von Lehrstoff und Prüfungswissen zu dokumentieren, sind solche Zahlen zumindest ein Indiz. Oder umgekehrt: Finden nur sehr wenige Prüfungsteilnehmer die richtige Antwort, kann das am hohen Niveau der Frage, am Nichtwissen der Teilnehmer liegen, es kann aber seine Ursache auch im Frageninhalt oder der Fragenform oder in mangelnder Eindeutigkeit der Fragestellung haben.

Die Erwartung, daß die Kommission jene Fragen herausfinden würde, bei denen solche Probleme bestehen, hat sich nicht erfüllt. Im Prüfungsgebiet Biologie–Anatomie, aus dem die schon genannten Beispiele stammen, waren zwei Fragen mit besonders schlechten Kenndaten vorhanden, bei denen nur 12 Prozent beziehungsweise 16 Prozent der Prüfungsteilnehmer die richtige Antwort wählten. Diese Fragen (Nr. 95 und Nr. 141) sind nicht für ungeeignet erklärt worden, sondern wurden „nur“ als verbesserungsbedürftig eingestuft und blieben in der Wertung.

Schlimmer sieht es in anderen Stoffgebieten aus. Von den 10 für ungeeignet erklärten Fragen im Stoffgebiet Physik–Physiologie hatten nur drei Fragen schlechte Antwortquoten (Frage 13: 25 Prozent, Frage 62: 14 Prozent, Frage 128: 21 Prozent). Drei andere, als geeignet belassene Fragen wiesen nur 10 Prozent oder weniger richtige Antworten auf (Frage 8: 9 Prozent, Frage 17: 10 Prozent, Frage 19: 5 Prozent).

Unter den als geeignet belassenen Fragen sind immerhin noch achtzehn weitere, deren positive Antwortquoten nur fünfundsiebzig Prozent oder weniger betragen (Fragen: 1: 24 Prozent, 10: 22 Prozent, 21: 13 Prozent, 22: 22 Prozent, 29: 21 Prozent, 35: 23 Prozent, 39: 25 Prozent, 49: 12 Prozent, 50: 20 Prozent, 51: 24 Prozent, 53: 19 Prozent, 56: 15 Prozent, 57: 20 Prozent, 60: 18 Prozent, 101: 25 Prozent, 104: 23 Prozent, 106: 25 Prozent, 129: 20 Prozent. Im Stoffgebiet Psychologie–Soziologie verbleiben unter den als geeignet belassenen Fragen drei dieser Kategorie (Fragen: 9: 23 Prozent, 88: 21 Prozent, 131: 17 Prozent), die von der Kommission nicht erfaßt wurden.

Die im Stoffgebiet Chemie–Physiologische Chemie herausgenommene Frage Nr. 115 wies immerhin noch 41 Prozent richtige Antworten auf. Nicht erfaßt wurden hingegen 11 weitere Fragen mit nur 25 Prozent oder weniger richtiger Antwortquote, von denen allein sechs (42: 17 Prozent, 72: 13 Prozent, 94: 15 Prozent, 116: 16 Prozent, 139: 15 Prozent, 152: 18 Prozent) deutlich unter der Rate-Wahrscheinlichkeit liegen. – Im Durchschnitt lag der Anteil der richtigen Antworten bei den von der Kommission für ungeeignet eingestuften Fragen im Stoffgebiet Physik–Physiologie immerhin noch oberhalb 30 Prozent, im Stoffgebiet Psychologie–Soziologie oberhalb 45 Prozent und im Stoffgebiet Biologie–Anatomie sogar oberhalb 60 Prozent.

Ohne hier die weiteren teststatistischen Kenndaten der Fragen im einzelnen zu diskutieren, ist schon aufgrund dieser Übersicht festzustellen, daß es der Kommission nicht gelungen ist, jene Fragen als „ungeeignet“ einzustufen, bei denen die Studierenden besondere Schwierigkeiten hatten. Für das Prüfungsgebiet Biologie–Anatomie kann ich fachlich begründen, daß die Kommission zumindest bei sechs der neun als „ungeeignet“ erklärten Fragen in ihrer Begründung zu sachlich falschen Schlußfolgerungen gekommen ist.

Umsetzung des Ergebnisses

Ziel einer kritischen Fragenrevision und gegebenenfalls einer „Nachbesserung“ müßte es sein, die *individuelle* Prüfungsleistung der einzelnen Teilnehmer zu korrigieren. Es müßte also eine erneute Auswertung aller Antwortbögen erfolgen, mit der Feststellung, wie viele der als geeignet verbliebenen Fragen der einzelne Prüfling richtig beantwortet hat, unter Neuberechnung der Bestehensgrenze. Dabei wird offensichtlich: Je „besser“ die Antworten der Teilnehmer bei den „ungeeigneten“ Fragen waren, um so mehr Pluspunkte aus diesen Fragen kommen bei der Revision für die Betroffenen in Wegfall.

Bei einer pauschalen Gutschrift von 29 Punkten für alle Prüfungsteilnehmer wird nicht das Ergebnis bei den „geeigneten“ Fragen „korrigiert“: Wer einen Teil der ungeeigneten Fragen doch richtig beantwortet hatte, erhält nun Pluspunkte für Fehler, die er bei den in der Wertung belassenen Fragen gemacht hat. Die pauschale Gutschrift von Punkten für alle Prüfungsteilnehmer ist keine Korrektur der individuellen Prüfungsleistung, sondern faktisch eine Senkung der Bestehensgrenze von bisher 50 Prozent auf nunmehr 41 Prozent. Selbst wenn man die neue Bestehensgrenze

an den verbliebenen 291 Fragen errechnet (wozu wegen der pauschalen Gutschrift kein Anlaß besteht), liegt sie mit 45 Prozent deutlich unter dem Minimum von 50 Prozent, das die Approbationsordnung fordert. Das Ergebnis ist durchschlagend: Statt 41,4 Prozent haben nach dieser „Nachbesserung“ nur noch 18,7 Prozent die Prüfung nicht bestanden. Wäre eine echte Korrektur der Einzelleistungen erfolgt, würde die Auswirkung auf die Nichtbestehensquote weit geringer sein, dem Vernehmen nach weiterhin knapp oberhalb der 40-Prozent-Grenze gelegen haben. Das ließe Rückschlüsse auf das wahre Leistungsvermögen der nun Gutschrift-Begünstigten zu, wenn die wirklich ungeeigneten Fragen eliminiert worden wären.

Bei dieser Sachlage hätte man sich den Umweg mit der Kommission ersparen können. Nicht die Arbeitsergebnisse der Kommission, so angehtbar sie im einzelnen sind, gingen in die „Korrektur“ ein. Die Kommission hat offensichtlich nur den numerischen Vorschlag für die faktische Änderung der Bestehensgrenze geliefert.

Kritik zu üben ist aber auch an Einzelheiten der Vorgehensweise bei diesem Verfahren. Bekanntlich war die Medizinische Hochschule Hannover am stärksten von allen medizinischen Ausbildungsstätten durch eine Nichtbestehensrate von 79,07 Prozent betroffen. Es ist zumindest eine Frage des politischen Stils, wenn der Kommissionsvorsitzende aus der am stärksten betroffenen Hochschule kommt und die Abwicklung des Bundesratsauftrages durch die niedersächsische Staatskanzlei erfolgt. Das DEUTSCHE ARZTEBLATT sprach von einem „Föderativ-Krimi“ (Heft 30/1985).

Anschrift des Verfassers:
Prof. Dr. med. Klaus Hinrichsen
Lehrstuhl für Anatomie I
Universitätsstraße 150
4630 Bochum 1

Moderne Medizintechnik verspricht höhere Lebensqualität

Kernspintomographen und anderes hochmodernes medizintechnisches Gerät seien nur teure Spielzeuge in den Händen der Ärzte. Oder: Patienten seien einer seelenlosen Technik hilflos ausgeliefert. Diese und ähnliche Meinungen sind vor dem Hintergrund einer verbreiteten Technologiefeindlichkeit zu sehen. Doch die Medizintechnik diene dem Kranken und schließlich auch der Volkswirtschaft. Das betonte Professor Dr. Michael Ungethüm, Vorstandssprecher der Aesculap-Werke in Tuttlingen, bei der Jahrespressekonferenz der deutschen feinmechanischen und optischen Industrie. So können medizintechnische Entwicklungen die Verweildauer der Patienten im Krankenhaus verkürzen, das Wiedereingliedern in den normalen Arbeitsprozeß fördern und Pflegekräfte einsparen. Auch die elektrotechnische Industrie trage mit ihrem technischen Fortschritt nachweisbar zur notwendigen Kostendämpfung im Gesundheitswesen bei, so Dr. Friedrich Kuhr, Vorstandsmittglied bei Siemens, unlängst vor der Deutschen Gesellschaft für Medizinische Physik.

Als moderne bildgebende Verfahren, die erst durch die Fortschritte der Mikroelektronik und Kleinrechner zu der heutigen Perfektion entwickelt werden konnten, kamen in den letzten Jahrzehnten die Nuklearmedizin, die Ultraschalldiagnostik, die Computertomographie und die Kernspintomographie hinzu. Bei den jüngeren Verfahren erhält man Querschnittsbilder ohne die störende Überlagerung der abgebildeten Organe, so daß man Anomalitäten leichter nachweisen und lokalisieren kann.

Ohne den Einsatz modernster Kleinrechner wäre auch die digitale Subtraktionsangiographie nicht

denkbar. Wegen seiner höheren Empfindlichkeit brauchen die Patienten bei diesem Verfahren nicht mehr so stark wie in der konventionellen Angiographie belastet zu werden, so daß man dies neue Verfahren teilweise sogar ambulant anwenden kann.

Von den Fortschritten in der Mikroelektronik und der Optik profitierte auch die ophthalmologische Diagnostik. So kann man heute weitverbreitete und gefährliche Glaukome mittels Bildanalyse wesentlich früher erkennen. Ein anderes Beispiel sind elektronische Geräte, die nachts vorherrschende Lichtverhältnisse simulieren und sich zum Nachweis der Nachtblindheit eignen. Damit trägt die moderne Augenheilkunde mit dazu bei, die Unfallgefahren zu verringern.

Modernste Lichtfasertechniken ermöglichten Erfolge der Endoskopie nicht nur im diagnostischen, sondern insbesondere auch im therapeutischen Bereich. Miniaturisierte Farbfernsehkameras, die auf die Endoskope aufgesetzt werden, dienen der subtilen Diagnose und erleichtern somit auch die endoskopische Mikrochirurgie. Bei der perkutanen Nephroskopie, die die extrakorporale Stoßwellentherapie (ebenfals ein Produkt modernster Medizintechnik) ergänzt, werden die Steine durch eine eingeführte Sonde mittels elektrohydraulischer Schockwellen zerstört.

Viele operative Eingriffe sind erst durch den gemeinsamen Einsatz miniaturisierter chirurgischer Instrumente und von Spezialmikroskopen überhaupt möglich geworden. Ein großer Vorteil der Mikrochirurgie ist, daß kleinere Wunden entstehen, die weitaus schneller heilen und das umgebende Gewebe weniger schädli-