

MEDIZINISCHE AUSBILDUNG

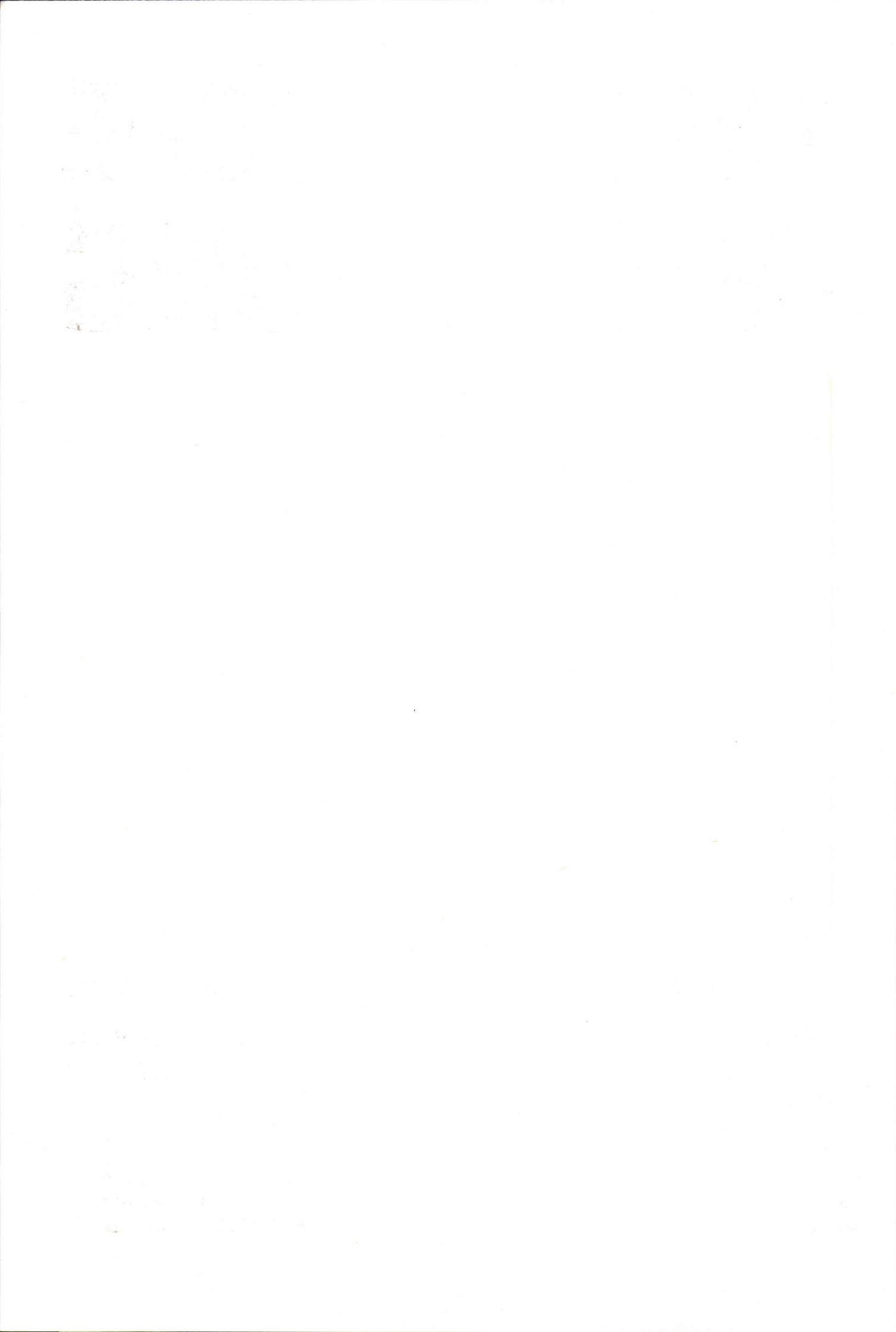


Forum zur Erforschung der ärztlichen Aus-, Weiter- und Fortbildung

Mitteilungsblatt der Gesellschaft für Medizinische Ausbildung
(Deutsche Sektion der Association for Medical Education in Europe)

2/1

Mai 85







MEDIZINISCHE AUSBILDUNG

INHALT

Herausgeber und Schriftleiter

Prof. Dr. D. Habeck
Albert-Schweitzer-Straße 11
D-4400 Münster i. W.

Redaktion

Prof. Dr. K. Hinrichsen, Bochum
Prof. Dr. H. E. Renschler, Bonn
Prof. Dr. K. Schimmelpfennig, Berlin
Prof. Dr. R. Toellner, Münster
Dipl.-Kaufm. P. Tschuschke, Lippstadt

Wissenschaftlicher Beirat

Prof. Dr. F. Anschutz, Darmstadt
Dr. F. C. Bleys, Utrecht
Prof. Dr. W. Hardegg, Heidelberg
Prof. Dr. H. Heimpel, Ulm
Dr. J. D. Hoppe, Düren
Prof. Dr. J. F. d'Ivernois, Paris
Prof. Dr. F. Kemper, Münster
Dr. M. Lischka, Wien
Prof. Dr. J. Moll, Rotterdam
Prof. Dr. H. Pauli, Bern
Prof. Dr. G. Schmidt, Göttingen
Dr. J. F. Steiger, Bern
Prof. Dr. H. J. Walton, Edinburgh

Vorwort

Grußadresse des Präsidenten der Association for Medical Education in Europe und der World Federation for Medical Education

Beiträge

F. C. Bleys und J. G. M. Gerritsma: Eine kurze Einleitung in die ärztliche Entscheidungsfindung

F. C. Bleys: Die ärztliche Entscheidungsfindung und die medizinische Ausbildung; Gespräche mit Art Elstein

G. Gebert: Die schriftliche Prüfung nach der ÄAppO - Probleme der Leistungsmessung und Leistungsbewertung

R. K. Fuchs: Die Absurdität der Bestehensregel für die Prüfungen in der ärztlichen Ausbildung

R. Lohölter: Das Prüfungssystem hat erneut versagt - Anmerkungen zur ärztlichen Vorprüfung vom März 1985

G. P. Burkhard und H. E. Renschler: Teilnehmerbewertung von Vorträgen

M. van der Ven und G. Buchkremer: Einstellungsänderungen bei Medizinstudenten im Verlauf des psychiatrischen Untersuchungskurses

K. Schimmelpfennig: Die Triage sollte aus dem Entwurf der ärztlichen Approbationsordnung herausgenommen werden

H. E. Renschler und D. Habeck: Die amerikanischen Prüfungen für ausländische Ärzte

D. Habeck und H. E. Renschler: Ergebnisse der ECFMG-Prüfung im internationalen Vergleich

Besprechungen

K. Schimmelpfennig: Besprechungen von Lehrfilmen

Resolution

zu Studienaufenthalten deutscher Medizin-Studenten im Ausland

Mitteilungen

Jahrestagung der AMEE in Jerusalem

Bild(ung) und Medizin

C. P. Lange und J. Geyer-Kordes: Zum Titelbild (Frau Dr. med. Marie Heim-Vögtlin mit dem Pflegekind Hanneli)

1

3

4

9

12

19

27

32

39

46

50

59

69

73

73

74

Medizinische Ausbildung erscheint zunächst in zwangloser Reihenfolge. Manuskripte sind in zweifacher Ausfertigung abdruckfertig in 1 1/2 zeiligem Schriftsatz mit einer Zeilenlänge von 10 cm und Gesamthöhe von 28 cm je Seite an die Schriftleitung einzusenden. Für den Verlust eingesandter Manuskripte wird keine Haftung übernommen. - Alle Beiträge sind urheberrechtlich geschützt. Nachdruck und jegliche andere Art von Vervielfältigung (auch auszugsweise) ist nur mit schriftlicher Genehmigung und ausführlicher Quellenangabe zulässig. Mit Verfasseramen veröffentlichte Beiträge decken sich nicht unbedingt mit der Meinung der Redaktion. Preis pro Heft 6 DM, für Studenten 2,50 DM. Mitglieder der Gesellschaft erhalten Medizinische Ausbildung kostenfrei. Postgirokonto: Prof. Dr. Habeck - Sonderkonto Medizinausbildung - Münster, Konto-Nr. 256825-464, Postgironamt Dortmund.



VORWORT

Das vorliegende dritte Heft der Zeitschrift hat länger auf sich warten lassen, als ursprünglich vorgesehen war. Der Grund dafür ist, daß die Finanzierung der Zeitschrift noch nicht sichergestellt ist. So erfolgte die Anfertigung auch dieses Heftes in unprofessioneller Weise.

Inhaltlich steht die bereits für das erste Heft gedachte Grußadresse von Herrn Professor Walton, dem Präsidenten der Association for Medical Education in Europe sowie gleichzeitigem Präsidenten der World Federation for Medical Education an erster Stelle. Zwei Darstellungen sind der bei uns noch wenig diskutierten ärztlichen Entscheidungsfindung gewidmet. Es folgen drei Artikel zum Thema der schriftlichen Prüfungen, von denen die ersten beiden vor der letzten Ärztlichen Vorprüfung im März 1985 der Schriftleitung zugeschickt wurden. Ihre Aufnahme erschien uns aus aktuellem Anlaß besonders sinnvoll, da sie die erneut in Gang gekommene Diskussion unseres derzeitigen Prüfungswesens bereichern dürften.

Weitere Arbeiten gelten der Bewertung von Vorträgen durch Ärzte, der Einstellungsänderung von Medizinstudenten im Laufe des psychiatrischen Untersuchungskurses, dem Thema Triage als Prüfungstoff bei der Novelisierung der ÄAppO und den amerikanischen Prüfungen für ausländische Ärzte sowie Ergebnissen der ECFMG-Prüfungen im internationalen Vergleich. Zwei Beispiele der Besprechung von Lehrfilmen sollen zu entsprechenden Bewertungen audiovisueller Programme anregen. Das Titelbild schließlich ist Frau Dr. Marie Heim-Vögtlein, der ersten Medizinstudentin und Ärztin in der Schweiz gewidmet.

Diese Vielfalt unterschiedlicher Beiträge verdeutlicht die Zielsetzung unserer Gesellschaft, das gesamte Gebiet der medizinischen Ausbildung methodisch zu erforschen und weiter zu entwickeln. Sie steht damit in Übereinstimmung mit den Aktivitäten dieser in verschiedenen ausländischen Staaten seit mehr als 25 Jahren etablierten medizinischen Disziplin, wie es Professor Walton in seiner Grußadresse zum Ausdruck gebracht hat.

Erfreulicherweise war das Echo auf die beiden ersten Hefte so groß, daß genügend Beiträge zur Veröffentlichung eingereicht wurden. Redaktion und Schriftleitung sind bereit, Artikel in die Zeitschrift aufzunehmen, deren Inhalt allein von den jeweiligen Autoren verantwortet wird. Sie wird sich bemühen, daß die Autoren die traditionelle Form wissenschaftlicher Abhandlungen einhalten.

Im Namen der Redaktion

Dietrich Habeck



MESSAGE FROM THE PRESIDENT,
Association for Medical Education in Europe

Professor H J Walton
M.D., Ph.D., F.R.C.P. F.R.C.Psych., D.P.M., M.D. (Hon.) Uppsala

The birth of a new Journal for Medical Education is a most important development, calling for serious recognition. On behalf of the European Association, and of the World Federation for Medical Education, I write to convey the warmest congratulations of the international brotherhood of medical teachers on the founding of Medizinische Ausbildung, and wish the Mitteilungsblatt der Gesellschaft für medizinische Ausbildung great success with this necessary development. As the German Section of AMEE, the Gesellschaft has the responsibility to study, describe and promote medical education - in its undergraduate, postgraduate and continuing phases - for the Federal Republic of Germany. The Gesellschaft also has responsibility to the other countries of Europe : to convey accurately and in detail the condition of medical education in the Federal Republic - its present state, its current problems, its particular strengths, and the proposals for necessary change.

These important tasks will be met more adequately through the existence of a new Journal devoted to medical education. The Journal can also serve to make known throughout West Germany and neighbouring countries the developments taking place in medical education in other nations.

At present the 450 medical schools of Europe, almost all of them long established and often described somewhat disparagingly as "traditional", are under criticism for being resistant to the changes called for by the current upheavals in our society. These shifts which present great new challenges in the training of doctors follow from altered economic circumstances; demographic shifts in populations (greater longevity, for example); the new diseases of affluence (such as the complications of drinking, wrong nutrition and smoking associated with contemporary lifestyles); technological discoveries; and many other social changes which have altered the practice of medicine. The new Journal will be an important instrument in assisting medical schools to evolve new patterns of education and training which will be relevant to their societies and which are needed to create doctors competent to meet the emerging demands for health delivery systems.

If medical schools can respond to these challenges, a promising future lies before them. For that to happen, a national Association such as the Gesellschaft will require authority and national recognition as a foremost body with primary concern for all stages in the training of doctors. No legislation should be passed, for example, on any health education matter without the advice of the Gesellschaft. All medical schools should look to the Gesellschaft as their counterpart on the national scene. Official bodies should liaise with the Gesellschaft about all matter concerning medical education.

Such influences will be fostered by the birth of the new Journal.

In that spirit, I express satisfaction and pleasure on behalf of European medical teachers, that the Gesellschaft has taken the wise and necessary step of launching the Journal. Medical teachers in all our countries will wish it a vigorous, effective and long life.

Henry Walton .

EINE KURZE EINLEITUNG IN DIE ÄRZTLICHE ENTSCHEIDUNGSFINDUNG

F. C. Bleys und J. G. M. Gerritsma, Utrecht

Für das Studium des ärztlichen Denkprozesses sind neuerdings zwei Ausdrücke besonders aktuell geworden: die Problemlösung (Medical Problem Solving) und die Entscheidungsfindung (Medical Decision Making oder Clinical Decision Analysis). Die Definition dieser Begriffe ist nicht ohne weiteres von herkömmlichen Theorien abzuleiten und die genaue Abgrenzung zwischen den beiden Ausdrücken ist auch noch nicht ganz eindeutig. Die Problemlösung und die Entscheidungsfindung schliessen sich gegenseitig nicht aus, sondern analysieren unterschiedliche Phasen des ärztlichen Denkprozesses und betrachten diese aus grundverschiedenen Blickpunkten.

Ein Beispiel

Eine Diskussion zwischen zwei kanadischen praktischen Ärzten soll das als Beispiel verdeutlichen (6):

Dr. McWhinney: "(...) Jeder praktisch tätige Arzt fängt gleich im ersten Kontakt mit Patienten an, sich Hypothesen zu bilden. Bereits während der Anamnese formulieren wir ständig Hypothesen, wir ändern sie oder ersetzen sie durch andere (...). Während des weiteren Untersuchungsganges, der sowohl die gezielte Anamnese und die körperliche Untersuchung, als die Laboruntersuchungen umfasst, versuche ich Tests zu finden, die meine Hypothesen entweder unterstützen oder widerlegen können(...) Ich wähle vorzugsweise solche Tests, die einen grossen "Nettoerlös" abwerfen, nämlich Tests, mit denen ich gut zwischen zwei Hypothesen unterscheiden kann."

Dr. Morissy: "Ich werde Ihnen das Problem einer 42-jährigen Patientin vorlegen, die seit einem Monat kaum gegessen und mindestens 15 Kilo abgenommen hat. Sie sagt, sie habe jede Tatkraft verloren und klagt über Abgeschlagenheit. Nach dem Essen wird es ihr immer übel, aber sie erbricht sich nie."

Dr. McWhinney: "Bei der Abmagerung denke ich zunächst an drei Krankheitsbilder: 1. Hyperthyreose, 2. Diabetes mellitus, 3. eine maligne Geschwulst. Eine weitere Hypothese aufgrund der Appetitlosigkeit lautet, dass sie an einer Hepatitis leidet (...). Um die letztgenannte Hypothese zu prüfen, möchte ich noch die folgenden Fragen stellen: Hat sich die Farbe des Stuhles oder des Harns geändert? Hat sich die Hautfarbe geändert? Ist ein Kontakt mit einem Gelbsüchtigen nachweisbar? Hat sie eine Injektion bekommen oder eine Bluttransfusion?"

Dr. Morissy: "Die Antwort auf all ihre Fragen lautet: nein. Ich selbst hätte übrigens anders verfahren. Obwohl Sie einige Fragen mit hohem Nettoerlös gestellt haben, haben andere wenig erbracht. Für mich ist es im Moment unwichtig, ob die Patientin eine Injektion oder eine Bluttransfusion bekommen hat. Die Antwort

liefert einen Hinweis für die Art der Hepatitis, vorausgesetzt, es sei wirklich eine Hepatitis. Ich muss aber zuallererst wissen, ob sie überhaupt eine Hepatitis hat. Die Tests mit einem hohen Nettoerlös sind deshalb die Fragen nach der Farbe des Stuhles und des Harns."

Problemlösung versus Entscheidungsfindung

In unserem Beispiel personifiziert Dr. McWhinney die Problemlösung, Dr. Morissy die Entscheidungsfindung. Die Problemlösungstheorie geht davon aus, dass die in Frage kommenden Lösungen zuerst erdacht werden müssen: die Arbeitshypothesen, die vorläufigen und definitiven Diagnosen und zum Schluss eine eventuell einzuleitende Therapie. Die ärztliche Entscheidungsfindung dagegen führt immer zur Wahl zwischen zwei oder mehreren im voraus feststehenden Alternativen.

Die Problemlösungstheorie beschreibt und erklärt das beobachtete ärztliche Verfahren als einen mentalen Vorgang, der für jeden Arzt verschieden sein darf und auch tatsächlich ist (2 , 3). Die Analyse der ärztlichen Entscheidungsfindung versucht die optimale Entscheidung oder Strategie zu finden, wenn zwei oder mehrere Entscheidungen oder Strategien möglich sind; die Entscheidungsfindung ist also ein vorschriftmässiger, stochastischer Prozess.

Ein Kartenspiel

Damit Sie den Unterschied zwischen dem beschreibend-erklärenden und dem präskriptiv-stochastischen Prozess selbst erfahren können, laden wir Sie zu folgendem Kartenspiel mit ihren Kindern ein: Aus zwei Stöcke machen Sie ein Spiel mit jeweils drei roten Karten und einer schwarzen. Die Karten werden gemischt und das Spiel beginnt. Es geht darum, zu erraten, ob beim Abnehmen eine rote oder eine schwarze Karte obenauf liegt; jeder darf reihum zehnmal hintereinander ziehen (7).

Man kann voraussagen, dass die Kinder mit einem intuitiven Wahrscheinlichkeitsgefühl von drei rot zu eins schwarz raten werden. Nachdem jedes Kind zehnmal geraten hat, sind Sie an der Reihe, aber Sie müssen eine andere Strategie nehmen: Sie wählen stets rot. Dadurch machen Sie die Wahrscheinlichkeit, eine falsche Voraussage zu machen, statistisch so klein wie möglich und damit werden Sie das Spiel wahrscheinlich gewinnen (und als Spielverderber angeprangert werden!). Sie haben nach den Regeln der Entscheidungsfindung gespielt.

Am Beispiel mit den Karten werden die Voraussetzungen klar, die die Entscheidungsfindung überhaupt möglich machen. Zuallererst muss das Problem genau definiert

sein und weiterhin müssen alle alternativen Lösungen und deren Konsequenzen bekannt sein und zwar im quantitativen Sinn (8). Sie hatten nicht nur eine Ahnung, auf welche Weise Sie gewinnen könnten, sondern Sie wussten ganz genau, Sie hatten bei jeder Wahl einer Karte eine Chance von drei zu eins, den Zug zu gewinnen. Die Entscheidungsfindung ist also nicht imstande, dem Arzt aus heiterem Himmel eine Diagnose zu überreichen, aber sie kann ihm helfen, auf die richtige Spur zu kommen bei der Bewertung seiner Hypothesen(4)

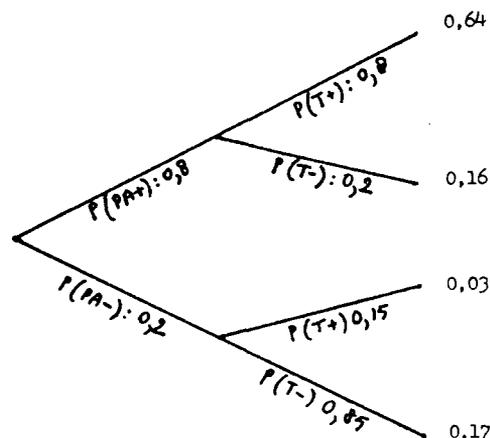


Fig. 1

P: Wahrscheinlichkeit
 PA: perniziöse Anämie
 T+: positives Testergebnis
 T-: negatives Testergebnis

Da es sich hier um zwei von einander unabhängige Wahrscheinlichkeiten handelt ($P(PA+)/P(PA-)$, bzw. $P(T+/T-)$), gilt hier die statische Produktregel (Product Rule).

Wahrscheinlichkeitsbäume liegen immer auf der Seite und werden von links nach rechts gezeichnet.

Ein öfters angewandetes Verfahren in der Entscheidungsfindung ist die Baumumstellung (Tree Inversion) (Fig.2)

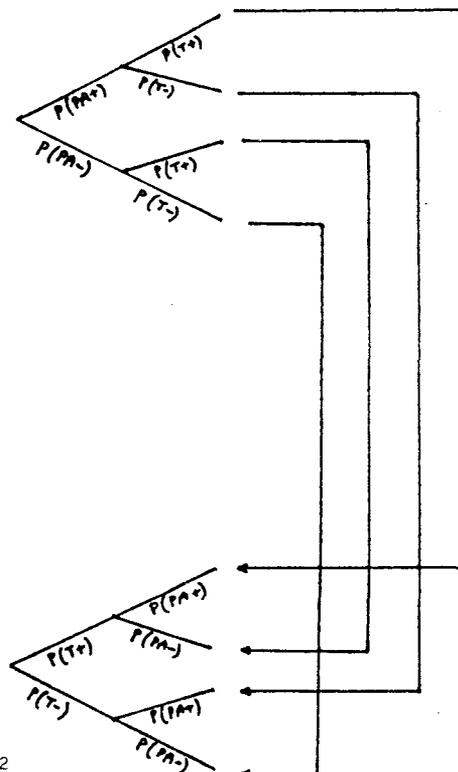


Fig. 2

Ein medizinisches Beispiel

Ein medizinisches Beispiel(1):Albert Bäcker, 61 Jahre alt, leidet seit vier bis fünf Monaten an einem brennenden Gefühl in seinen Füßen und stolpert nachts öfters. Seine Zunge ist glatt und gerötet. Die Milchsäuredehydrogenase im Blut ist erhöht. Das Hämoglobin beträgt 6,6 mmol/l (10,6 g%).

Sein Arzt denkt sofort an eine perniziöse Anämie. Die Frage erhebt sich nun, in welchem Masse die Bestimmung des Vitamin B₁₂ der Diagnose eine grössere Wahrscheinlichkeit geben kann. Aus dem Gesichtspunkt der Analyse des ärztlichen Verfahrens ist der Arzt damit an der Grenze zwischen der Problemlösung und der Entscheidungsfindung angelangt, was deutlich illustriert, dass die Problemlösung und die Entscheidungsfindung einander ergänzen können.

Wie bereits erwähnt, muss der Arzt jetzt, um die Entscheidungsfindung erfolgreich durchführen zu können, das Problem quantitativ definieren. Der erste Schritt ist, dass er die Wahrscheinlichkeit bestimmt oder schätzt, aufgrund derer, dieser Patient, ohne weitere Tests, perniziöse Anämie hat (die a priori Chance). Dieser erste Schritt, der zum Beispiel auch bei einem präventiven Bevölkerungsscreening die Bestimmung der Prävalenz der gesuchten Krankheit sein kann, wird manchmal unterlassen, was eine falsche Entscheidungsfindung zur Folge hat (5).

Der Arzt taxiert die Chance, dass Albert Bäcker aufgrund dessen, was er bereits von ihm weiss, perniziöse Anämie hat, auf 80% (0,8).

Bäume zeichnen

Ein Test gibt niemals hundertprozentige Sicherheit, da er falsch-positive (false positive) und falsch-negative (false negative) Befunde ergeben kann (positive Resultate ohne Krankheit, beziehungsweise negative Resultate bei Krankheit). Deshalb kann der Arzt die Testergebnisse nur dann richtig beurteilen, wenn er über den Prozentsatz der richtig-positiven (true positive) und richtig-negativen (true negative) Tests informiert ist (positive Resultate bei Krankheit, beziehungsweise negative Resultate ohne Krankheit).

Für die Bestimmung des Vitamin B₁₂ nehmen wir eine richtig-positive Rate von 80% (0,8) an im Falle einer perniziösen Anämie und eine richtig-negative Rate von 85% (0,85). Der Arzt kann jetzt einen sogenannten Wahrscheinlichkeitsbaum (Probability Tree) zeichnen, der so heisst, da er einem verzweigten Winterbaum ähnlich sieht (Fig. 1).

Der nächste Schritt ist das sogenannte Umfallen (Folding Back), das heisst, wir gehen von den Ergebnissen aus und berechnen, was ein positiver, beziehungsweise negativer Test impliziert. Hierbei soll die statistische Summenregel (Summation Rule) angewendet werden, da es sich um von einander abhängige Wahrscheinlichkeiten ($P(T^+, PA^+)$, $P(T^+, PA^-)$, $P(T^-, PA^-)$, $P(T^-, PA^+)$) handelt, dessen Summe eins ergeben muss (Fig. 3).

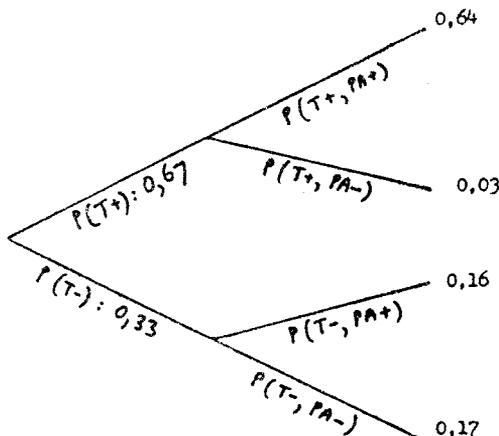


Fig. 3

Dadurch wird die Wahrscheinlichkeit sichtbar gemacht, dass Albert Bäcker eine perniziöse Anämie hat, wenn bei ihm ein erniedrigter Vitamin B₁₂-Gehalt nachgewiesen werden kann. Denn diese Wahrscheinlichkeit lässt sich jetzt kalkulieren:

$$P(T^+, PA^+) = \frac{0,64}{0,67} = 0,955$$

$P(T^+, PA^+)$ wird auch der positive Voraussagewert (P_v^+) (Positive Predictive Value) des Testes für diesen individuellen Patient genannt. Denn wir haben nicht vergessen die a priori Chance mitzuberechnen!

Die ursprüngliche Frage, nämlich inwieweit die Bestimmung des Vitamins B₁₂ die Diagnose einer perniziösen Anämie wahrscheinlicher macht, ist quantitativ gelöst: von 0,8 auf 0,955.

Eine Wette

Oberflächlich betrachtet, mag die Zunahme der Wahrscheinlichkeit von 0,8 auf 0,955 nicht gross erscheinen; diese ist jedoch deutlicher sichtbar, wenn man diese Entscheidungsfindung auffasst wie eine Wette, wie beim Kartenspiel mit Ihren Kindern. Dabei wetteten Sie drei zu eins auf rot. Die geeignete Formel dafür lautet:

$$\frac{p}{1 - p}$$

Für Albert Bäcker war die Wette, bevor das Vitamin B₁₂ bestimmt worden war:

$$\frac{0,8}{1 - 0,8} = \text{vier zu eins.}$$

Nachdem ein erniedrigtes Vitamin B₁₂ gefunden worden war, aber:

$$\frac{0,955}{1 - 0,955} = 21 \text{ zu eins!}$$

Das Zweimalzwei

Es ist nicht unbedingt nötig Bäume zeichnen zu können! Die Zahlen können auch in eine Vierfelder-(2x2-, Entscheidungs-) Matrix eingeführt werden.

Der Grundmodell einer 2x2-Matrix sieht wie folgt aus (Tafel 1).

	Krankheit +	Krankheit -	
T+	TP	FP	TP + FP
T-	FN	TN	FN + TN
	TP + FN	FP + TN	TP + FP + TN + FN

Tafel 1

- T+: positives Testergebnis
- T-: negatives Testergebnis
- TP: richtig positiv (positives Testergebnis bei Krankheit)
- FP: falsch positiv (positives Testergebnis ohne Krankheit)
- TN: richtig negativ (negatives Testergebnis ohne Krankheit)
- FN: falsch negativ (negatives Testergebnis bei Krankheit)

Wenn die Zahlen des Falles Albert Bäckers in eine solche 2x2-Matrix, mit der Kalkulation der a priori Chance, eingetragen werden, ist das Resultat (Tafel 2).

	PA +	PA -	
T+(Vitamin B ₁₂ erniedrigt)	0,64	0,03	0,67
T-(Vitamin B ₁₂ normal)	0,16	0,17	0,33
a priori Chance	0,8	0,2	1

Tafel 2

Hieraus kann die Wahrscheinlichkeit eines positiven Testresultats, wenn Albert Bäcker eine perniziöse Anämie hat, abgeleitet werden, sowie die Wahrscheinlichkeit eines negativen Resultats, wenn er nicht an dieser Krankheit leidet:

$$P(T^+, PA^+) = \frac{0,64}{0,67} = 0,955$$

$$P(T^-, PA^-) = \frac{0,17}{0,33} = 0,515$$

$P(T^+, PA^+)$ ist, wie bereits erwähnt, der positive Voraussagewert (P_v^+), das heisst, der Bruchteil der positiven Test-

resultate (meistens in einem Prozentsatz ausgedrückt), der mit der Wirklichkeit (der Anwesenheit der Krankheit) übereinstimmt. $P(T, PA-)$ ist der negative Voraussagewert (P_{V-}), das heisst der Bruchteil der negativen Testresultate, der mit der Wirklichkeit (der Abwesenheit der Krankheit) übereinstimmt.

Die Voraussagewerte sind nicht nur eine Funktion der Eigenschaften des Tests, sondern auch der a priori Chance. Die Voraussagewerte für die Bestimmung des Vitamins B_{12} sind Albert Bäcker infolgedessen nach Mass geschnitten worden. Würde man mittels der Bestimmung des Vitamins B_{12} die ganze Bevölkerung auf das Bestehen einer perniziöse Anämie untersuchen wollen, so gilt eine andere a priori Chance und erhält man andere Voraussagewerte.

Bayes

Im Grunde haben wir bei der Anfertigung einer 2x2-Matrix für den Fall Albert Bäckers implizit bereits das Theorem von Bayes angewandt. Das Theorem von Bayes kann aber auch algebraisch ausgedrückt werden. Um das deutlich zu machen, kommen wir zuerst auf einige Eigenschaften des Testes selbst zurück.

Die Sensitivität (Sensitivity) eines Testes ist die Chance, dass bei Anwesenheit der Krankheit der Test ein positives Resultat ergibt; auf eine Formel gebracht:

$$S_e = \frac{TP}{TP + FN}$$

Die Spezifizität (Specificity) eines Testes ist die Chance, dass bei Abwesenheit der Krankheit der Test ein negatives Resultat ergibt; auf eine Formel gebracht:

$$S_p = \frac{TN}{TN + FP}$$

Wenn die Bestimmung des Vitamins B_{12} eingetragen wird:

$$S_e = \frac{0,64}{0,64 + 0,16} = 0,8 \quad \text{und}$$

$$S_p = \frac{0,17}{0,17 + 0,03} = 0,85$$

Infolge des Theorems von Bayes ist:

$$P_{V+} = \frac{S_e \times \text{die a priori Chance}}{(S_e \times \text{die a priori Chance}) + (1 - S_e) \times (1 - \text{die a priori Chance})}$$

$$P_{V-} = \frac{S_p \times (1 - \text{die a priori Chance})}{S_p \times (1 - \text{die a priori Chance}) + (1 - S_p) \times \text{die a priori Chance}}$$

Für Albert Bäcker also:

$$P_{V+} = \frac{0,8 \times 0,8}{(0,8 \times 0,8) + (1 - 0,85) \times (1 - 0,8)} = 0,955$$

$$P_{V-} = \frac{0,85 \times (1 - 0,8)}{(0,85 \times (1 - 0,8)) + ((1 - 0,8) \times 0,8)} = 0,515$$

Drei verschiedene Methoden um dasselbe zu kalkulieren, die Voraussagewerte der Bestimmung des Vitamins B_{12} im Fall Albert Bäckers, sind jetzt demonstriert worden: Der Wahrscheinlichkeitsbaum, die 2x2-Matrix und die Bayes-Formel. Die zuletzt erwähnte Methode macht dem praktisch tätigen Arzt ohne Zweifel die Entscheidungsfindung am wenigsten schmackhaft, da die Formel wenig Einsicht bringt und sehr kompliziert ist.

Vom Wahrscheinlichkeitsbaum zum Entscheidungsfindungsbaum

Zurück nach den Bäumen! Bis jetzt haben wir von Wahrscheinlichkeitsbäumen gesprochen. Wenn wir den zuletzt gezeichneten Wahrscheinlichkeitsbaum anlässlich Albert Bäcker noch einmal zeigen, wird sofort klar, dass dieser auch

ein Entscheidungsfindungsbaum genannt werden kann, denn er zeigt genau die Chance mit der wir die Diagnose perniziöse Anämie richtig, beziehungsweise nicht stellen, wenn wir diese Krankheitsbestimmung von dem Vitamin B_{12} -Test abhängig machen (Fig. 4).

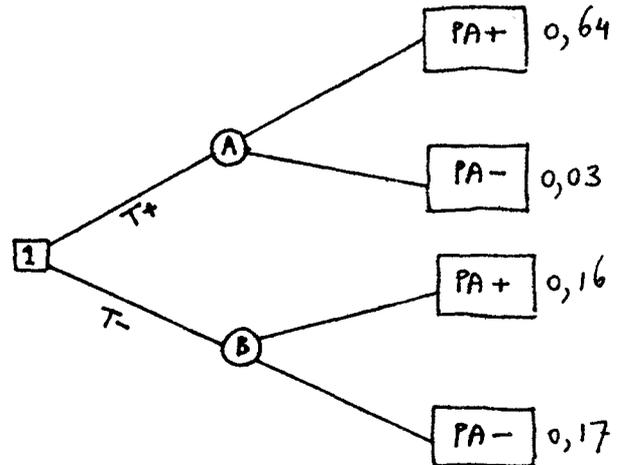


Fig. 4

Er zeigt ebenso gut die Chance, mit der wir die Diagnose perniziöse Anämie richtig, beziehungsweise nicht stellen, wenn wir diese Krankheitsbestimmung nicht von dem Vitamin B_{12} -Test abhängig machen. Mit andern Worten: Diese Entscheidungsfindungsbaum zeigt alle Konsequenzen einer Bestimmung oder aber Unterlassung des Vitamin B_{12} -Tests im Fall Albert Bäckers im voraus. In der alltäglichen Sprache des Arztes auch: Er zeigt alle Gefahren von Unter- und Überdiagnostik.

Beim erneuten Zeichnen des Baumes haben wir einige Elemente hinzugefügt: Kreise und ein Viereck. In der Nomenklatur der Entscheidungsfindung ist ein Viereck ein Entscheidungsknoten, ein Kreis ein Chanceknoten. An einem Entscheidungsknoten muss oder kann der Arzt eine Entscheidung treffen, zum Beispiel das Vitamin B_{12} bestimmen lassen oder nicht, die Ereignisse am Chanceknoten dagegen kann er nicht beeinflussen. Die Entscheidungsknoten werden beziffert, die Chanceknoten erhalten Buchstaben. Die grossen Rahmen enthalten die Endergebnisse der verschiedenen Entscheidungen und Ereignisse.

Der Entscheidungsbaum anlässlich Albert Bäcker war natürlich ganz einfach. Um einen besseren Eindruck zu vermitteln, zeigen wir darum einen verzweigten Baum, den Fall einer 25-jährigen Frau mit einem Knoten in der Brust (5) (Fig. 5).

DIE ÄRZTLICHE ENTSCHEIDUNGSFINDUNG UND DIE MEDIZINISCHE AUSBILDUNG; GESPRÄCHE MIT ART ELSTEIN

F. C. Bleys, Utrecht

Noch immer glauben viele praktisch tätigen Ärzte, dass ihr Denken und Handeln ebenso "wissenschaftlich" unterbaut sei, wie das ihrer Kollegen, die ausschliesslich in der Forschung tätig sind. Damit meinen sie, ihr Vorgehen im Patientenkontakt sei induktiv: nämlich zuerst Daten sammeln, diese dann einordnen und analysieren und schliesslich eine vorläufige Diagnose stellen, die zu einer weiteren Datensammlung Anlass geben kann. So ist ungefähr der empirische Zyklus De Groot's. Auch der problemorientierte Status von Weed ist von dieser Überzeugung geprägt, was einer der Gründe war, dass er viel kritisiert wurde.

Abgesehen davon, dass es sehr fraglich ist, ob die grossen wissenschaftlichen Entdeckungen auf induktive Weise zustandekommen (zum Beispiel die Erzählung Watson's von der Entdeckung der Struktur der DNS), wurde am Ende der 70er Jahre bewiesen, dass die praktisch tätigen Ärzte nicht induktiv denken, sondern hypothetisch-deduktiv, dass heisst, dass sie gleich im ersten Stadium des Patientenkontakts Hypothesen (provisorische Diagnosen) haben, diese nachprüfen und falls nötig, durch andere, bessere ersetzen.

Der grosse Initiator der Idee des hypothetisch-deduktiven Denkens der Ärzte war Professor Art Elstein von der Staatlichen Universität Michigans in den Vereinigten Staaten, der im Jahre 1978 zusammen mit seinem Mitarbeiter darüber das Buch "Medical Problem Solving" veröffentlichte. Vielen war es, als ob die Anarchie ihren Einzug in die medizinische Ausbildung gehalten hätte, da jegliche herkömmliche, normative Systematik in der Didaktik der Medizin überflüssig geworden schien. Viele ältere Dozenten reagierten, als ob Elstein eine Entheiligung der medizinischen Tradition verübt hätte.

Mittlerweile haben sie gerade mit Elstein nicht nur einen Anhänger, aber auch einen Erneuerer bekommen. Er ist ein erklärter Vorkämpfer der Analyse der ärztlichen Entscheidungsfindung (Medical Decision Making oder Clinical Decision Analysis) geworden. Während eines längeren Aufenthaltes an seiner Universität in East Lansing, Michigan, hatte ich Gelegenheit ausführlich mit ihm über die ärztliche Entscheidungsfindung und deren Anwendung in der medizinischen Ausbildung zu sprechen.

Elstein

Art Elstein ist ein typischer jüdischer Ostküsten-Intellektueller, der im provinziellen East Lansing ziemlich ent wurzelt scheint. Als er während einer Käse-und-Wein-Party feststellte, dass er früher längere Zeit in Boston einige Häuser von meiner Frau entfernt gewohnt hatte, übermannte ihn die Rührung und das Eis zwischen ihnen war sofort gebrochen.

Er ist ein beflügelter Redner, der geschickt und treffend formuliert, unterstrichen von heftigen Gesten, wobei er mitunter wirkt wie Woody Allen. Er lässt sich ungerne unterbrechen und es fällt ihm schwer Einwänden unmittelbar Gehör zu schenken. Aber er verarbeitet sie trotzdem und kommt später darauf zurück: "Vongestern sagtest du, dass; ich habe darüber nachgedacht und nun meine ich"

Deskriptiv und präskriptiv

Ich erwähnte Elstein von meiner Auseinandersetzung mit einem Professor für innere Krankheiten, der sagte, die Medizinstudenten können das Lösen von ärztlichen Problemen nur dadurch richtig lernen, dass sie während des diagnostischen Prozesses die Probleme der Patienten suallereerst in grosse physiologische Gruppen einordnen, also induktiv. Auf meinen Einwand, kein erfahrener Arzt tue es auf diese Weise, erwiderte der Professor, dass dann alle erfahrenen Ärzte dringend Fortbildung brauchen. Ich meinte, ebenso wie auch Ayer ("Sprache, Wahrheit, Logik") schreibt, dass es uns mit Hilfe von Arbeits-hypothesen möglich ist, unsere Erfahrung vorausszusagen. Das ist gerade das, was ein praktisch tätiger Arzt wünscht, nämlich eine ziemlich zuverlässige Prognose einer Krankheit und des Resultats einer Therapie zu machen.

Elstein lachte. "Da werden doch einige Dinge durcheinander gebracht! Deine Meinung und die des Professors schliessen einander gar nicht aus."

Die ärztliche Problemlösung kann man von drei verschiedenen Gesichtspunkten, entsprechend den drei grösseren kognitiven Theorien, betrachten: Die Psychologie der Informationsverarbeitung (Information-Processing Psychology), die Gruppenschlussfolgerungstheorie (Social Judgment Theory) und die Entscheidungsfindungsanalyse (Decision Analysis). Das frühere Werk Elsteins aus dem Ende der 70er Jahre, über die ärztliche Problemlösung, ist eine Psychologie der Informationsverarbeitung und hat als Ziel, den Denkvorgang des Arztes zu charakterisieren, dadurch, dass er Schritt für Schritt in psychologischer Hinsicht festgelegt und erklärt wird. Bei der Gruppenschlussfolgerungstheorie steht nicht der ärztliche Denkvorgang selbst zentral, sondern es werden bei einzelnen Gruppen von Ärzten die Anfangsdaten und die Schlussfolgerungen miteinander statistisch verglichen.

Die beiden ersten Theorien sind deskriptiv; jedoch können Normen für die medizinische Ausbildung davon abgeleitet werden, wenn die Handlungweise von Experten als massgebend angesehen wird. Die Analyse der Entscheidungsfindung zielt dagegen auf eine Korrektur intuitiver Entscheidungen durch die Anwendung formaler Modelle und ist infolgedessen vorschreibend, präskriptiv. Sie schliesst die hypothetisch-deduktive Denkart nicht aus, sondern

ergänzt sie. Bei jedem Schritt im hypothetisch-deduktiven Denkprozess kann die Analyse der Entscheidungsfindung quantitativ angeben, was die jeweiligen Konsequenzen der Hypothesen sind, sowie die Risiken und den Voraussagewert (Predictive Value) weiterer Schritte im voraus kalkulieren, so wie beim in-voraus-Denken im Schachspiel.

Entscheidungen in Ungewissheit

Die Analyse der Entscheidungsfindung sieht die ärztliche Tätigkeit an als eine Reihe von riskanten Wahlen in Ungewissheit. Sie macht den Grad von Ungewissheit explizit und quantitativ.

Elstein: "Die Ärzte und insbesondere die bedürftigen Patienten unter dem Druck einer lähmenden Unsicherheit, Sie werden auf falsche Weise Sicherheit an, durch möglichst viel Information zu sammeln, besonders in Form von Laborbefunden. Sie sehen nicht ein, dass gerade eine Häufung von Befunden das Urteil trüben kann, da sie gelernt haben, das Labor gebe perfekte Information. Sie haben keine Ahnung von falsch-positiven und falsch-negativen Resultaten und noch weniger von der Beeinflussung des Wertes eines Tests durch die Prävalenz der gesuchten Krankheit in der Bevölkerung. Daran hat der herkömmliche Unterricht in der Statistik Mitschuld, da er die Berechnung und die Interpretation von Signifikanzniveaus, t-Tests, usw. betont und nicht die Statistik von Bayes, die für den praktisch tätigen Arzt viel relevanter ist. Das Theorem von Bayes nimmt eine hervorragende Stelle ein bei der ärztlichen Entscheidungsfindung."

Wie kommt es mit den Statistikkenntnissen der Ärzte und Medizinstudenten aus? Ein kürzlich erschienener Artikel von Casscells und Mitarbeitern aus dem "New England Journal of Medicine" deutet, aus dem hervorgeht, dass nur 10% der befragten Ärzte und Medizinstudenten den Voraussagewert eines sehr einfachen Tests richtig berechnen können.

Elstein: "Mit der Analyse der ärztlichen Entscheidungsfindung schenken wir unseren Studenten nicht Sicherheit, dass ist ja unmöglich, sondern ihnen die Werkzeuge überreichen, so dass sie die Ungewissheiten rational verarbeiten und handhaben können."

Allmählich ist Elstein im Gespräch an dem speziell von ihm entwickelten Kurs in Entscheidungsfindung gelangt, wovüber er mit grossem Enthusiasmus erzählt.

Alt und neu

Elstein und seine Mitarbeiter entwickelten aus einer herkömmlichen Einführung in die Klinik einen Kurs, der die psychologischen Prinzipien des Diagnostizierens, sowie ethische und ökonomische Fragen und selbstverständlich auch die Analyse der ärztlichen Entscheidungsfindung betont. "Eigentlich sind spezielle Kurse für die Entscheidungsfindung unnatürlich, da das ganze Studium davon geprägt werden sollte. Ich sehe diese Kurse als einen Anfang zum weiteren Ausbau."

"Es ist mir übrigens aufgefallen, dass es schwieriger ist, innovative Curricula zu erneuern als herkömmliche. Wie du weisst, haben wir an der Staatlichen Universität Michigans zwei Curricula nebeneinander, ein innovatives und ein traditionelles. Im traditionellen Curriculum muss man nur einzelne Elemente abändern, im innovativen Curriculum, worin alles miteinander zusammenhängt, ist

ein viel grösserer Umbau erforderlich."

Er findet die physiologischen Gruppen untauglich für die Entscheidungsfindung. "Heutzutage, wo Frühdiagnosen gestellt werden sollen und diagnostische Entscheidungen getroffen werden müssen, bevor klar ist, in welche Gruppe die Krankheit einzuordnen ist, und ausserdem die Krankheiten im Gegensatz zu früher nicht mehr einfach in eine einzige Gruppe eingeteilt werden können - Diabetes mellitus gehört zum Beispiel nicht nur zum endokrinen System, sondern auch zu Herz und Kreislauf, sowie zum Nervensystem -, sollte man von den diagnostischen Möglichkeiten selbst ausgehen." Die physiologischen Gruppen haben sich überlebt

"Vergiss nicht," fügt Elstein hinzu, "dass man vor hundert Jahren der Physiologie ebenso fremd gegenüberstand, wie jetzt der Analyse der Entscheidungsfindung. Ich bin davon überzeugt, dass die Analyse der Entscheidungsfindung durchsetzen wird, nicht, weil die Ärzte vernünftiger werden - das massgebende Buch über die Analyse der ärztlichen Entscheidungsfindung, dessen Mitautor ich bin, ist mehr in den Literaturverzeichnissen genannt, als wirklich gelesen worden -, sondern als Folge allgemeiner wirtschaftlicher Forderungen, durch finanzielle Einschränkungen, die überall in der westlichen Welt in der Gesundheitsorge erforderlich werden. Die Analyse der ärztlichen Entscheidungsfindung ist das einzige Mittel, womit die Obrigkeiten die Entscheidungen der Ärzte unmittelbar beeinflussen können und damit ist sie das einzige wirklich effektive Mittel zur Kosteneinschränkung. Das kann man bereits in den Vereinigten Staaten wahrnehmen, wo ausgerechnet der Vorkämpfer einer freien Wirtschaft, Reagan, die medizinische Betreuung der Pensionäre, die grossenteils vom Staat bezahlt werden muss, dadurch einschränkt, dass nur ein im voraus bestimmter Betrag für jede Krankheit, die sogenannten "Diagnosis Related Groups", überwiesen wird. Damit zwingt er die Ärzte indirekt, die von ihnen angewandten diagnostischen Mittel wohlüberlegt zu wählen."

Nur eine Ahnung davon

"Ich glaube, wir schweifen vom Thema ab," unterbrach ich ihn. "Apropos ihr Buch: Ich habe es wirklich gelesen und ich frage mich, ob die Materie nicht etwas zu schwierig sei für deine Studenten; es betrifft schliesslich Studenten im 1. bis 4. Semester."

"Aber nein, es ist nicht meine Absicht versierte Entscheidungsfindungsanalytiker aus ihnen zu machen! Ich halte es nicht für nötig, dass jeder Arzt bei jedem klinischen Problem einen prächtigen Entscheidungsfindungsbaum (Decision Tree) aus seinem Ärmel schüttelt. Das ist mitunter viel zu kompliziert und zeitraubend für einen vielbeschäftigten praktizierenden Arzt. In der Zukunft kann übrigens der Computer ihm dabei behilflich sein; es gibt bereits Software für die ärztliche Entscheidungsfindung. Was ich anstrebe, ist, dass die Medizinstudenten eine Ahnung bekommen von der rationalen Entscheidungsfindung. Die meisten Ärzte wissen doch auch nicht alle Einzelheiten der Physiologie. Aber sie haben eine Ahnung davon, was es ihnen ermöglicht auf wissenschaftliche Weise zu funktionieren."

Der Kurs

"Zurück zum Thema!" mahnte ich wieder.

"Einverstanden! Betrachten wir einen Kurs; hier habe ich zum Beispiel einen papieren Fall einer Anämie, ein sogenanntes "Focal Problem".

Der Kurs fängt an mit einer von mir gegebenen allgemeinen Einleitung in die Grundlagen der logischen, rationellen klinischen Entscheidungsfindung. Zuvor haben die Studenten Literaturlaufgaben studiert; amerikanische Studenten sind in dieser Hinsicht äusserst diszipliniert.

Während der drei folgenden Unterrichtsstunden behandelt ein Internist zuerst die klinischen Manifestationen und die üblichen Laborbefunde bei Anämie, ehe er auf die Entscheidungsfindung eingeht. Dieses Verfahren kommt den Studenten noch nicht spanisch vor und macht sie nicht kopfscheu. Ich halte nichts vom aufdringlichen Dozieren von Neuigkeiten, das erzeugt nur Widerstand.

Aber dann werden die Studenten konfrontiert mit den Eigenschaften von Labortests: Wahrscheinlichkeit, falsch-positiven und falsch-negativen Befunden, illustriert am Beispiel der Anämie. Aus ihren Köpfen wird nachdrücklich herausgeputzt, dass die Information eines Laborbefundes hundertprozentig sei.

Die vierte Unterrichtsstunde ist dem Theorem von Bayes gewidmet, das an Hand eines klinischen Problems, eines Kindes mit Anämie erläutert wird. Wir gehen nicht gleich auf die statistischen Berechnungen ein, sondern zuerst müssen die Studenten intuitiv und mit Hilfe von Lehrbüchern Wahrscheinlichkeiten einschätzen. Die Studenten erhalten durch die herkömmliche medizinische Ausbildung ein sehr verzerrtes Bild der Prävalenz der verschiedenen Krankheiten. Auch werden die Kosten der Labortests und der Behandlung unter die Lupe genommen.

Dies alles findet nicht nur während der Unterrichtsstunden statt, sondern auch nach vorbereitendem Selbststudium, anhand von papieren Fallsimulationen und Fragebogen mit Literaturanweisungen."

Ich bemerkte, dass diese Papiere nicht sehr kompliziert aussehen.

"Der grosse Fehler des beginnenden Dozenten in einem innovativen Kurs ist immer, dass er den Aufbau, sowie die Anleitungen allzu kompliziert macht, was ihm viel zu viel Zeit nimmt," fügte Elstein hinzu.

Es dauert mehrere Unterrichtsstunden, bis die Studenten mit dem Zeichnen von Entscheidungsfindungsbäumen anfangen können. In der gezeigten Anleitung betrifft es einen Fall möglicher perniziöser Anämie. Die Studenten müssen in jedem Stadium des diagnostischen Prozesses dieser Fallsimulation quantitativ angeben, wievielmehr Gewissheit jeder Schritt verschafft. Auf diese Weise lernen sie, dass in diesem speziellen Fall die Anamnese und einfache Laborbefunde bereits grosse Wahrscheinlichkeit erbringen, dass die Bestimmung des Vitamin B₁₂ viel zur richtigen Diagnose beiträgt, aber ein Schillingtest ganz überflüssig ist.

Für jedermann!

Nachdem ich die zum Kurs gehörenden Papiere durchgeblättert hatte, musste ich Elstein sagen: "Du hast gesagt, dass du deine Studenten nicht zu schwer belasten wolltest, aber ich bin froh, zu sehen, dass du

auch die Papiere für den Dozenten beigelegt hast. Denn diese enthalten alle Antworten fix und fertig!"

Um ihn zu necken, fragte ich ihn noch, ob die Analyse der ärztlichen Entscheidungsfindung doch nicht nur eine Modelalaune sei, die nicht anhalten werde.

"Aber ganz gewiss nicht!" erwiderte er heftig, "die Entscheidungsfindung soll und muss der Schwerpunkt des zukünftigen ärztlichen Denkens und Handelns werden. Die Analyse der Entscheidungsfindung ist nicht nur für die Ärzte äusserst wichtig, sondern für uns alle. Das Leben ist für jedermann eine ununterbrochene Aneinanderreihung von Entscheidungen. Darf die nicht etwas wissenschaftlicher werden?"

Literatur

1. Elstein, A.S., M.M. Holmes, M.M. Ravitch, u.A.: Medical Decisions in Perspective: Applied Research in Cognitive Psychology. Persp. Biol. Med. 26 (1983) 486-501
2. Elstein, A.S., D.R. Rovner, M.L. Rothert: A Preclinical Course in Decision Making. Med. Decis. Making 2 (1982) 209-216
3. Griner, P.F., R.J. Mayewski, A.I. Mushlin, u.A.: Selection and Interpretation of Diagnostic Tests and Procedures; Principles and Applications. Ann. intern. Med. 94 (1981) 553-600
4. Weinstein, M.C., H.V. Fineberg, A.S. Elstein, u.A.: Clinical Decision Analysis. W.B. Saunders Cy. Philadelphia, London, Toronto. 1980
5. Wright, H.J., D.B. Macadam: Clinical Thinking and Practice. Churchill Livingstone, Edinburgh, London, New York. 1979

Ich danke Frau Dr. G.L. Kuiper-Hindemith für die Korrektur meines Deutsch."

Dr. F. C. Bleys, Simulations- und Instruktionslabor der Reichsuniversität Utrecht, Postfach 16250, NL 3500 CG Utrecht

DIE SCHRIFTLICHE PRÜFUNG NACH DER ÄAPPO PROBLEME DER LEISTUNGSMESSUNG UND LEISTUNGSBEWERTUNG

Gerfried Gebert, Mainz

Das Bestehen der Ärztlichen Prüfung gilt nach der Bundesärzteordnung als Nachweis dafür, daß der Kandidat die fachlichen Voraussetzungen für die Erteilung der Approbation als Arzt erfüllt. Die Ärztliche Vorprüfung dient analog der Beurteilung der Frage, ob der Kandidat über die für die Zulassung zum klinischen Ausbildungsabschnitt für erforderlich gehaltenen Kenntnisse verfügt.

Die Prüfungen sind Leistungsprüfungen, die verhindern sollen, daß unzureichend fähige Kandidaten den Beruf des Arztes ausüben. Ihr Ziel ist nicht, einen bestimmten Anteil der Kandidaten bestehen oder nicht bestehen zu lassen.

Prüfungsstoff der schriftlichen Examina

Der globale Rahmen für den Inhalt des in den einzelnen Stoffgebieten zu prüfenden Wissens ist die ÄAppO mit ihren Prüfungsstoffkatalogen. Inhaltliche Leitlinie für deren Umsetzung in zu messende Kenntnisse ist die Vorschrift des § 14(2) ÄAppO: Die Prüfungsfragen sind auf die für den Arzt allgemein erforderlichen Kenntnisse abzustellen.

Es wäre nicht sachgerecht, diese Vorgabe im Sinne von "für den Allgemeinarzt erforderlichen" Kenntnissen auszulegen. Dies ergibt sich bereits daraus, daß in der Bundesrepublik Deutschland nur etwa ein Viertel der berufstätigen Ärzte als niedergelassene Allgemeinärzte bzw. Praktische Ärzte

arbeiten. Es geht vielmehr um die Kenntnisse, die die allgemein als Grundlage der ärztlichen Tätigkeit angesehen werden und die die Basis für die durch Weiterbildung erfolgende Qualifikation zur selbständigen Tätigkeit in einem der zahlreichen Sektoren des Berufsfeldes "Arzt" liefern.

Dieser Ansatz entspricht auch der Berufswirklichkeit, denn selbst die Tätigkeit als Praktischer Arzt, bei der die selbständige Berufsausübung nicht an eine vorhergehende Weiterbildung gebunden ist, wird nach der Statistik der Kassenärztlichen Bundesvereinigung überwiegend erst nach vierjähriger und längerer Assistenzarztzeit aufgenommen, d.h. nach nur formell nicht dokumentierter Weiterbildung.

Problematischer ist die Anwendung des in § 14(2) formulierten Prinzips auf die Ärztliche Vorprüfung. Dies gilt insbesondere für die Fächer Physik, Chemie und Biologie, in denen weniger "für den Arzt allgemein erforderliche Kenntnisse" als naturwissenschaftliche Grundlagen für den Erwerb dieser Kenntnisse zu prüfen sind.

Messung von Prüfungsstoffkenntnis mit m.c.-Fragen

Bei Antwortauswahlaufgaben hat der Kandidat keine Möglichkeit, durch Rückfragen Interpretationshilfen in bezug auf die Fragestellung zu erlangen oder problematische Sachverhalte zu diskutieren. Als Prüfungsstoff kommen daher nur eindeutig formulierbare und ebenso eindeutig als richtig oder

falsch einstuftbare Fakten oder Zusammenhänge in Frage. Hypothesen sollten nur geprüft werden, wenn sie allgemein anerkannt sind, und auch dann nur unter Kennzeichnung ihres Charakters (z.B. Symptome ersten Ranges der Schizophrenie nach K. SCHNEIDER).

Durch Antwortauswahlfragen wird der Prüfungsstoff in Form von Prüfungsgegenständen parzelliert. Bei Fragen, in denen eine richtige Aussage (Lösung) zusammen mit mehreren falschen (Distraktoren) angeboten wird, ist als Prüfungsgegenstand der Prüfungsstoff anzusehen, dessen Kenntnis zur Bewertung der Lösung erforderlich ist.

Die Anforderung durch eine Frage kann vom einfachen Wiedererkennen eines ausformulierten Sachverhalts bis zur Verwertung des Wissens in Form von Analyse oder Synthese reichen (taxonomische Fragenschwierigkeit).

Die Zahl der Prüfungsgegenstände, die den Prüfungsstoff eines Faches bilden, läßt sich grob abschätzen, wenn man die gegenüber den Prüfungsstoffkatalogen der AAppO differenziertere Auflistung der Prüfungsthemen in den Gegenstandskatalogen des IMPP zu Hilfe nimmt. Im Fach Anatomie beispielsweise enthält der GK 1 in 21 Kapiteln insgesamt 362 Themen vom Abstraktionsniveau "Rektum", "Schädelbasis" usw. Zu jedem dieser Themen lassen sich im Durchschnitt mindestens 15 Prüfungsgegenstände formulieren, wenn man die Spannweite der in der Ausbildung vermittelten Kenntnisse berücksichtigt (Form, Struktur, Funktion, Nachbarschaftsbeziehungen, Gefäße, Nerven usw.). Danach umfaßt der Prüfungsstoff im Fach Anatomie mindestens 5000 Prüfungsgegenstände. Bei einer anderen Schätzmethode kann man von den in den bisherigen Prüfungsterminen eingesetzten Fragen ausgehen. Im Fach Physiologische Chemie ergab eine Analyse der Prüfungen bis einschließlich Herbst 1984 (eigene, noch unveröffentlichte Ergebnisse) über 1500 lösungsrelevante Prüfungsgegenstände. Die Zahl der pro Termin neu geprüften Sachverhalte zeigte noch keine abnehmende Tendenz, so daß der Umfang des gesamten Prüfungsstoffs des Faches erheblich größer sein muß als der des bisher geprüften Stoffes.

Prüfungsleistung in der Antwortauswahlprüfung in Relation zur Prüfungsstoffkenntnis des Kandidaten

Als Maß der schriftlichen Prüfungsleistung dient die Zahl der richtig beantworteten Fragen im Verhältnis zur Zahl der gestellten Fragen. Die von der Beantwortung der einzelnen Frage gegebene dichotome (ja/nein) Information über die Kenntnis des geprüften Gegenstands wird durch Mittelung über die Fragen eines Termins zu einer quantitativen Aussage umgeformt. Ein solches Verfahren erscheint nur gerechtfertigt, wenn die Beantwortung der einzelnen Fragen eine zumindest formal vergleichbare Information über die Prüfungsstoffkenntnis liefert.

Bei Fragen vom Typ positive Einfachauswahl (eine richtige von fünf angebotenen Alternativen) ist zur korrekten Beantwortung in der Regel die Kenntnis eines Prüfungsgegenstands notwendig. Fragen vom sogenannten Verknüpfungstyp, bei denen die Richtigkeit von zwei Aussagen und (im Fall, daß beide zutreffend sind) das Vorhandensein einer kausalen Beziehung beurteilt werden muß, verlangen die Kenntnis von mindestens zwei Prüfungsgegenständen. Bei Fragen vom Kombinationstyp (nur Aussagen 1,2 und 5 sind richtig) können bis zu vier Prüfungsgegenstände lösungsrelevant geprüft werden. Der gegenüber Einfachauswahlaufgaben erhöhte Anspruch von Verknüpfungs- und Kombinationsfragen an die Prüfungsstoffkenntnis kommt in der geringeren Häufigkeit richtiger Antworten bei diesen Fragentypen zum Ausdruck. Von den den Studenten noch nicht bekannten Fragen der Vorprüfungen von Herbst 1979 bis Frühjahr 1984 wurden im Fach Anatomie 53,4 % der Fragen vom Typ positive Einfachauswahl, aber nur 44,6 % der Kombinationsfragen richtig beantwortet. Im Fach Physiologische Chemie betrug die entsprechende Zahl 58,6 % bzw. 46,8 % (eigene, noch unveröffentlichte Untersuchungen).

Bei Verwendung von mehreren Fragentypen mit unterschiedlichem Anspruch an die Kenntnis des Prüfungsstoffs hat die in einem Termin erbrachte Prüfungsleistung eines Kandidaten den Charakter einer höchstens im Rahmen einer Rangordnung in-

terpretierbaren Zahl ohne quantifizierbaren Aussagewert für die Prüfungsstoffkenntnis.

Auch wenn nur Fragen mit vergleichbarem Anspruch an die Prüfungsstoffkenntnis eingesetzt werden, muß bedacht werden, daß mit den Fragen eines Termins nur ein kleiner Bruchteil des Prüfungsstoffs erfaßt werden kann. Im Fach Anatomie z.B. kann an einem Termin (ca. 80 Fragen) nur die Kenntnis von 1-2 % der Prüfungsgegenstände ermittelt werden, wenn Einfachauswahlaufgaben gestellt werden. Die Prüfungen haben somit den Charakter einer Stichprobe, und die Genauigkeit von Hochrechnungen aus dem Prüfungsergebnis auf die Prüfungsstoffkenntnis ist an die Repräsentativität und die Validität der Stichprobenzusammenstellung gebunden.

Festlegung der zu fordernden Prüfungsleistung

Es ist nicht möglich, die Qualität ärztlicher Leistungen allgemeinverbindlich zu quantifizieren und aus der späteren Berufstätigkeit von erfolgreichen Examenkandidaten retrospektiv einen Maßstab für die Bewertung der Examenleistung abzuleiten. Die Entscheidung über die Bestehensgrenze ist daher eine Ermessensentscheidung, die sich am Prinzip der Verantwortung gegenüber dem potentiellen Patienten und der Fairneß gegenüber dem Kandidaten zu orientieren hat.

Dabei ist möglich :

1. eine apodiktische Festlegung eines geforderten Mindestanteils richtiger Antworten unter Außerachtlassung des Aussagewerts der Aufgaben über die Prüfungsstoffkenntnis des Kandidaten

2. die Fixierung einer unteren Grenze für die Prüfungsstoffkenntnis, wobei der Mindestanteil richtiger Antworten als Folgegröße unter Berücksichtigung der Aussagefähigkeit und -genauigkeit des Prüfungsverfahrens bestimmt wird.

Die normorientierte Bestehensregel (Gleitklausel), bei der die Leistung eines Kandidaten an der aller Kandidaten eines Prüfungstermins gemessen wird, ist sachlich und möglicherweise auch rechtlich unbefriedigend. Mit einer solchen Bewertung wird von vornherein festgelegt, daß eine (kleine) Zahl von Kandidaten ungeachtet ihrer wahren Prüfungsleistung nicht bestehen darf und daß der (weit überwiegende) Rest der Kandidaten bestehen muß. Derzeit "beweist" nach der Gleitklauselregelung ein Kandidat seine Qualifikation für den Übergang ins Praktische Jahr, wenn er mit seiner Prüfungsleistung hinter nicht mehr als 94 % seiner Kommilitonen zurückbleibt. Für die Bewältigung dieser Aufgabe hat er zudem drei Anläufe zur Verfügung.

Wenn man den oben unter 2. beschriebenen Ansatz einer Festlegung der geforderten Mindestkenntnis des Prüfungsstoffs wählt, könnte die Bewertung "bestanden" z.B. an den Nachweis gebunden werden, daß der Kandidat mindestens die Hälfte des Stoffes beherrscht. Eine derartige Anforderung darf nicht als niedrig eingeschätzt oder gar mit dem Begriff "Halbwissen" abqualifiziert werden. Bei dem enormen Umfang des in der Ausbildung zum Arzt zu erwerbenden Wissens (allein in der Ärztlichen Vorprüfung umfaßt der Prüfungsstoff mindestens 20 000 Gegenstände) stellt ihre Erfüllung eine respektable Leistung dar.

Die in der Prüfung mit Antwortauswahlfragen (im Gegensatz zur mündlichen Prüfung) wegfallende Gewichtung der Prüfungsgegenstände nach ihrer Bedeutung sollte nicht zum Anlaß genommen werden, den Prüfungsstoff in sogenanntes Kern- oder Basiswissen und in sonstiges Wissen einzuteilen und der Beantwortung von Fragen zum "Basiswissen" größere Bedeutung zuzumessen. Abgesehen davon, daß es kaum möglich ist, einen Kon-

sens über die Zuordnung der Prüfungsgegenstände zum Basiswissen zu erreichen, wird der unterschiedlichen Bedeutung der einzelnen Prüfungsthemen bereits dadurch Rechnung getragen, daß zu wichtigeren Bereichen mehr Prüfungsgegenstände formuliert und entsprechend auch mehr Fragen gestellt werden.

Fehlerquellen der Kenntnisschätzung über die Prüfungsleistung

Systematische Fehler bei dem Rückschluß von dem Anteil richtiger Antworten bei den Fragen eines Termins auf die Prüfungsstoffkenntnis eines Kandidaten sind zu erwarten, wenn

1. Fragen zum Einsatz kommen, deren Beantwortung auch ohne das in der Ausbildung zum Arzt vermittelte Wissen möglich ist
2. verschiedene Fragentypen mit konstruktionsbedingt unterschiedlichen Anforderungen an die Prüfungsstoffkenntnis verwendet werden
3. die geprüften Gegenstände nicht ausreichend repräsentativ für den gesamten Prüfungsstoff sind
4. den Kandidaten bereits bekannte Fragestellungen wiederholt eingesetzt werden.

Bei Häufung wiederholter, aus Fragensammlungen bekannter Fragen bezieht sich die Prüfung statt auf den gesamten Prüfungsstoff hauptsächlich auf den bereits geprüften Stoff, der bisher nur einen geringen Teil des gesamten ausmacht. Studenten, die sich schwerpunktmäßig nach Fragensammlungen vorbereiten, erzielen bei hohem Anteil wiederholter Fragen ein besseres Ergebnis als es ihrer eigentlichen Fachkenntnis entspricht.

Einfluß der Fragenschwierigkeit auf die Prüfungsleistung

Als eigentliche Fragenschwierigkeit ist die Anforderung zu definieren, die die Fragestellung

an die Umsetzung der Kenntnis des Prüfungsgegenstands in die richtige Beantwortung stellt (taxonomisches Niveau). Für die Lösung entscheidend wird ein höherer Fragenschwierigkeitsgrad überhaupt erst, wenn der Kandidat über eine gewisse Kenntnis des geprüften Sachverhalts verfügt.

In den Antwortauswahlprüfungen in der Ausbildung zum Arzt werden (nicht nur in der Bundesrepublik Deutschland) überwiegend Fragen auf dem unteren Niveau des Wiedererkennens lehrbuchgemäß formulierter Sachverhalte gestellt. Der Anteil richtiger Antworten, den ein Kandidat erzielt, hängt deshalb hauptsächlich von seiner bloßen Kenntnis der geprüften Gegenstände ab. Gleiches gilt für den Anteil der Kandidaten, der eine Frage richtig beantwortet hat. Ungeachtet dessen wird der Prozentsatz der Kandidaten mit richtiger Lösung einer Frage als Fragenschwierigkeitsindex (FSI) bezeichnet, obwohl es sich eher um einen Index der Kenntnis vom geprüften Gegenstand handelt.

Der FSI stellt eine von der Leistungsfähigkeit der geprüften Population abhängige Größe dar.

Die Häufigkeitsverteilung des FSI aller denkbaren Fragen kann somit nur hypothetisch auf der Basis einer Bezugspopulation definiert werden. Im Idealfall wäre von einer Normalverteilung auszugehen.

Für die Fragen der einzelnen Prüfungstermine ist auch wegen der zu verlangenden Chancengleichheit für die Kandidaten eine vergleichbare (an dem FSI einer Bezugspopulation orientierte) FSI-Verteilung zu fordern. Dies kann als erreicht angesehen werden, wenn (nach Aus-

schluß von Populationsunterschieden) kein statistisch signifikanter Unterschied zwischen den FSI-Verteilungen des einzelnen Termins und der Gesamtzahl der denkbaren Fragen (Fragengrundgesamtheit) besteht.

Die FSI-Häufigkeitsverteilung in der Fragenrundgesamtheit ist einer direkten Bestimmung nicht zugänglich, weil die für die Zusammenstellung der Prüfungen nach einem bestimmten Verfahren erarbeiteten Fragen nur einen kleinen Teil der zum Prüfungsstoff erarbeitbaren Fragen darstellen. Außerdem steht keine konstante Bezugspopulation, sondern nur die von Prüfungstermin zu Prüfungstermin verschiedene jeweilige Kandidatenpopulation zur Ermittlung der FSI-Verteilung zur Verfügung. Wenn die Schwankung der Populationsleistung jedoch gering gegenüber der Streubreite des FSI ist, kann durch die Zusammenfassung der Ergebnisse einer Reihe von Prüfungsterminen geeignete Information über die wahrscheinliche FSI-Verteilung in der Fragenrundgesamtheit erhalten werden.

Die Sicherstellung einer von Termin zu Termin vergleichbaren FSI-Verteilung ist bei den Prüfungen nach der ÄApp0 nicht über eine gezielte Zusammenstellung der Prüfungsfragenstichproben realisierbar. Die Verwendung von durch Einsatz in früheren Terminen vorgetesteten Fragen verbietet sich, weil dadurch das Prüfungsergebnis systematisch verfälscht wird (s.o.). Eine prospektive Schätzung des FSI durch Sachverständige ist weder zuverlässig noch genau genug möglich. Der einzig sinnvolle Weg für die Zusammenstellung von nach der FSI-Häufigkeitsverteilung vergleichbaren Stichproben geht daher über eine Zufallsauswahl.

Sind die bei den bisherigen Prüfungen beobachteten Unterschiede in der FSI-Verteilung der nicht systematisch verschiedenen Prüfungsfragen mit der Annahme einer Zufallsschwankung vereinbar?

Die Untersuchung dieser Frage ist auch von prospektivem Interesse, denn sie ermöglicht Aussagen über die künftig zu erwartende Schwankungsbreite der Prüfungsschwierigkeit auf der Basis der bisherigen Erfahrungen.

Als Material dienten die Fragen der Fächer Anatomie, Physiologie und Physiologische Chemie der Ärztlichen Vorprüfungen vom Herbst 1979 bis zum Frühjahr 1984 (10 Termine). Zur Gewährleistung der Vergleichbarkeit wurden nur Fragen des Typs positive Einfachauswahl herangezogen und von diesen auch nur solche, deren lösungsentscheidender Inhalt den Kandidaten noch nicht aus früheren Fragen bekannt sein konnte.

Als Beispiel wird die FSI-Häufigkeitsverteilung der Fragen des Faches Anatomie dargestellt:

FSI (%)	n	Fragenanzahl graphisch
6-15	3	ooo
16-25	2	oo
26-35	18	oooooooooooooooooooo
36-45	13	oooooooooooooooooooo
46-55	20	oooooooooooooooooooo
56-65	23	oooooooooooooooooooo
66-75	19	oooooooooooooooooooo
76-85	10	oooooooooooo
86-95	2	oo

Die Verteilungsparameter (Mittelwert und Standardabweichung) für die zusammengefaßten Fragen aller Termine waren für das Fach

Anatomie	53,4 ± 18,0 %
Physiologie	54,6 ± 20,3 %
Physiol. Chemie	58,6 ± 18,5 %

In allen drei Fächern fielen Mittelwert und Median der FSI-Verteilung praktisch zusammen, d.h. die Verteilungen waren weitgehend symmetrisch. Zur Prüfung der Zufallsannahme wurde untersucht, ob die FSI-Mittelwerte der einzelnen Prüfungen innerhalb des 95 % -Vertrauensbereichs blieben, der sich aus der FSI-Häufigkeitsverteilung der einzelnen Fragen des Faches und dem Umfang der Stichprobe des Termins errechnen läßt, wenn man von dem Prinzip normalverteilter FSI-Werte ausgeht.

Die von Herbst 1979 bis Frühjahr 1984 aufgetretenen Schwankungen im FSI-Mittel der Termine waren eher geringer als es bei einer Zufallsauswahl aus einer Fragengrundgesamtheit unter Annahme einer konstanten Populationsleistung zu erwarten war. Der 95 %- Vertrauensbereich für den FSI - Mittelwert wurde in keinem Fall überschritten. Innerhalb des 68 % -Vertrauensbereichs, in dem nach Fallsgesichtspunkten nur 6 - 7 der 10 FSI-Mittelwerte pro Fach zu erwarten waren, blieben im Fach Anatomie 9, im Fach Physiologie 8 und im Fach Physiologische Chemie 6 der 10 Terminmittelwerte.

In der Diskussion der Zufallsbedingtheit der Schwankungen läßt sich das obige Ergebnis allerdings nur im Sinne des Fehlens eines Gegenbeweises interpretieren.

Ein positiver Hinweis für den Zufallscharakter der Unterschiede in den FSI-Mittelwerten (bezogen auf eine konstante Populationsleistung) läßt sich auf einem anderen Weg gewinnen. Die Leistungsfähigkeit der Kandidaten ist in den Herbstterminen im Mittel besser als in den Frühjahrs-terminen, denn im Herbst ist der Anteil von Erstteilnehmern nach Mindeststudiendauer höher, und diese Kandidaten sind nach der IMPP-Statistik leistungstärker als die übrige Population. Wenn

sich die FSI-Verteilung der Frühjahrs- und der Herbsttermine nur zufällig unterscheidet, muß der FSI-Mittelwert im Herbst durchschnittlich höher liegen als im Frühjahr.

Bei Zusammenfassung der Fragen aus den jeweils 5 Terminen erhält man als FSI - Mittel für

	Herbst	Frühjahr
Anatomie	55,7 %	50,4 %
Physiologie	54,3 %	54,8 %
Physiol. Chemie	61,7 %	54,2 %

Die nach dem Unterschied in den Kandidatenpopulationen zu erwartende Differenz tritt deutlich zutage, wenn man vom Fach Physiologie absieht. In diesem Fach wird der Unterschied durch das mehr als 14 Prozentpunkte über dem sonstigen Mittel liegende Ergebnis der Prüfung vom Frühjahr 1984 verwischt. Die Fragen dieses Termins waren (bedingt durch das Ausscheiden der Sachverständigen der Physiologie) nicht mehr von einer regulären Fachkommission erarbeitet worden.

Wie hoch ist die Fallsschwankung der Prüfungsschwierigkeit bei Prüfungen mit noch unbekanntem Fragen vom Typ positive Einfachauswahl?

Als Maß für die zu erwartende Fallsschwankung der Prüfungsschwierigkeit kann der Vertrauensbereich für den FSI-Mittelwert des Prüfungstermins dienen, der sich unter Annahme einer konstanten Leistungsfähigkeit der geprüften Population abschätzen läßt. Unter Zugrundelegung der Normalverteilungsannahme für den FSI ergibt sich die Standardabweichung des FSI-Terminmittelwerts vom wahren Mittelwert der Fragengrundgesamtheit aus der Standardabweichung der Einzelfragen-FSI geteilt durch die Wurzel aus der Zahl der Fragen pro Termin.

Bei den drei untersuchten Fächern, die etwa zwei

Drittel der Fragen der Ärztlichen Vorprüfung stellen, liegt die (auf eine konstant leistungsfähige Population bezogene) FSI-Standardabweichung der einzelnen Fragen höchstens bei 18 Prozentpunkten, denn in dem beobachteten Wert von 18 - 20 Prozentpunkten ist zusätzlich u.a. der Frühjahrs/Herbst - Unterschied in der Populationsleistung enthalten.

Wenn pro Termin 80 Fragen eines Faches gestellt würden, wäre von einer Standardabweichung des FSI-Mittelwerts von $18/\sqrt{80} = \text{ca. } 2$ Prozentpunkten auszugehen. Bei Hochrechnung auf die 320 Fragen der Ärztlichen Vorprüfung wäre (unter Annahme einer vergleichbaren Einzelfragenstreuung in den anderen Fächern) die Standardabweichung des FSI-Mittelwerts auf nur einen Prozentpunkt zu schätzen.

Modellrechnung für eine Bestehensgrenze bei einer Prüfung aus unveröffentlichten Einfachauswahlaufgaben bei einer geforderten Prüfungsstoffkenntnis von 50 %

Bei Angebot einer richtigen zusammen mit vier falschen Alternativen wird die Lösung auch ohne Kenntnis des geprüften Gegenstands in 20 % der Fälle zufällig angekreuzt. Wenn eine zur statistischen Realisierung der Zufallschance α reichende Fragenzahl angeboten wird, ist von einem Kandidaten, der die Hälfte des Prüfungsstoffs beherrscht, eine Prüfungsleistung von 60 % richtig beantworteter Fragen zu erwarten.

Bei einer derartigen Ermittlung der Bestehensgrenze sollte zugunsten des Kandidaten berücksichtigt werden, daß die Prüfung an seinem Ter-

min zufällig schwieriger sein kann als an den sonstigen Terminen. Für die Ärztliche Vorprüfung mit 320 Fragen ist der 99 % - Vertrauensbereich des FSI-Mittelwerts als dreifache Standardabweichung, also etwa 3 Prozentpunkte, zu schätzen. Da die Bestehensgrenze in der Nähe des FSI-Mittelwerts liegt, könnten diese 3 Prozent auf die eigentlich zu fordernden 60 % richtiger Antworten angerechnet werden. Zusätzlich könnte als Vorhalt für nachträglich als ungeeignet erkennbare (z.B. nicht eindeutig lösbare) Fragen ein Vorwegabzug von zusätzlich zwei Prozentpunkten vorgesehen werden. Bei einer so ermittelten Bestehensgrenze von 55 % richtiger Antworten würde (unter den Bedingungen der Modellrechnung) sichergestellt, daß praktisch alle Kandidaten, die die Lösung von 50 % der gestellten Fragen aufgrund ihrer Prüfungsstoffkenntnis finden, auch bestehen. Zugunsten der Zweifelsfälle würde in Kauf genommen, daß die Prüfung auch von einigen bestanden wird, die nicht über die verlangte Prüfungsstoffkenntnis verfügen.

Resumé

Die hier vorgelegte, unter Verwendung von bisher eingesetzten Fragen der Ärztlichen Vorprüfung durchgeführte Analyse zeigt, daß bei Vermeidung systematischer Fehler eine quantitative, präzise und zuverlässige Messung der Prüfungsstoffkenntnisse mit Antwortauswahlaufgaben möglich ist. Eine nicht standardisierte m.c.-Prüfung gestattet dagegen nur Aussagen über die Rangordnung der Kandidaten nach ihrer Leistung an einem Termin.

Prof. Dr. G. Gebert,
Friedrich-Schneider-Str. 5, 6500 Mainz 1

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 gesamt 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

DAS PRÜFUNGSYSTEM HAT ERNEUT VERSAGT

ANMERKUNGEN ZUR ÄRZTLICHEN VORPRÜFUNG VOM MÄRZ 1985

R. Lohölter, Frankfurt/Main

Einleitung

Die FAZ hat es auch diesmal - wie schon 1981 - zuerst gewußt: "Ergebnis der ärztlichen Vorprüfung alarmiert Länderbehörden" (Ausgabe v. 30.3.1985). Vier Jahre nach dem Märzphysikum von 1981 ist erneut von einem "Katastrophen-Physikum" die Rede.

Die folgenden Ausführungen sollen versachlichen, informieren - aber auch Stellung beziehen. Sie sind eine erste Analyse aufgrund gegenwärtig zugänglicher Daten, die weiterer Vertiefung und Detaillierung durch eine Sachverständigen-Kommission bedarf.

Ergebnisse

Die Prüfung vom März d.J. war das mit deutlichem Abstand schwerste Multiple-Choice-Physikum, das es bisher gegeben hat. Die durchschnittliche Prüfungsleistung lag mit 53,2% richtig beantworteten Fragen noch um 4,6% unter dem bislang niedrigsten Wert vom März 1981; sie lag um beinahe 9% unter

den durchschnittlichen Leistungen der Jahre 1981 - 1984.

Betrachtet man die in den einzelnen Fächern bzw. Stoffgebieten erreichten Ergebnisse, fällt folgendes auf (Tab. 1):

In der Physik, der Biologie und Psychologie/Soziologie gibt es nur geringfügige Abweichungen vom Durchschnitt der vergangenen vier Jahre. Die Leistungen in Anatomie und Chemie liegen schon deutlich unter den bisherigen Mittelwerten; es sind auch die schlechtesten Einzelergebnisse in diesem Zeitraum. Krasse Einbrüche liegen in der Physiologie und Biochemie vor; beide Fächer haben ihre bisher niedrigsten Werte in der MC-Geschichte. Geradezu atemberaubend ist die Negativleistung in der Physiologie: Das jetzige Ergebnis liegt um 21,4% unter dem Mittelwert der letzten vier Jahre, es übersteigt die Ratewahrscheinlichkeit gerade um 20%.

Tab. 1 : Fachergebnisse bei der ärztlichen Vorprüfung 1981 - 1985
(Prozentwerte richtig beantworteter Fragen)

Prüfungsfach	M 81	A 81	M 82	A 82	M 83	A 83	M 84	A 84	M 85	Durchschnitt 1981 - 1984 (10)	Differenz (10) - (9)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)		
Physik	50,1	53,9	54,3	60,0	58,6	64,1	48,9	51,1	53,2	55,1	- 1,9
Physiologie	55,4	72,3	64,7	51,6	50,6	65,0	68,4	69,0	40,7	62,1	-21,4
Chemie	58,1	57,2	57,4	63,0	60,4	60,4	60,2	61,4	52,2	59,8	- 7,6
Biochemie	59,5	63,4	62,6	63,7	58,3	54,6	53,4	57,1	46,7	59,1	-12,4
Biologie	62,7	69,3	72,1	65,2	65,2	64,3	65,8	73,8	68,6	67,3	+ 1,3
Anatomie	57,0	60,3	64,0	65,0	60,1	61,1	60,6	60,5	55,3	61,1	- 5,8
Psy/Soz.	58,9	84,7	76,3	52,8	59,6	72,0	65,6	59,8	63,7	66,2	- 2,5

Da Biochemie und Physiologie zusammen beinahe 40% der Prüfungsfragen stellen, ergibt sich, daß das bisher negativste Physikum wesentlich von diesen beiden Fächern geprägt ist.

Notenverteilung

Von 5675 Prüflingen haben ganze 4 - d.h. 0,7 Promille - die Note "sehr gut" erhalten; dagegen 2388 (42,1%) die Note "mangelhaft". Man muß sich hier vor Augen halten, daß ein erheblicher Teil der Studienanfänger mit hervorragenden Kenntnissen (und Noten) auch in den naturwissenschaftlichen Fächern in die Universitäten kommt. Wenn weniger als ein Promille eine "1" erreicht, dann ist das ein unmißverständlicher Hinweis darauf, daß das Märzphysikum sich in seinem Schwierigkeitsgrad von den Prüfungen der vergangenen Jahre krass unterschieden hat.

Die Notenverteilung (0,07% "sehr gut", 3,6% "gut", 19,9% "befriedigend", 34,4% "ausreichend" und 42,1% "nicht ausreichend") bestätigt auch die Kritik an dem Benotungsmechanismus, der mit der 4. Änderung der Approbationsordnung für Ärzte vom Dezember 1983 eingeführt wurde: "Für jede Verteilung der Noten gilt, daß die beste Note an 10% der Teilnehmer vergeben werden soll. Ein Verfahren, bei dem dieser Anteil wesentlich von 10% abweicht, löst bei Fachleuten Überraschung aus (Bloom, Hastings, Madaus)." (Renschler und Burkhard, 1984, S. 20) Der geltende Benotungsmechanismus ist eine politische Setzung, die wissenschaftlich nicht begründbar ist; er ist nicht geeignet, auch bei schweren Prüfungen eine akzeptable Notenverteilung sicherzustellen.

Auswertungsfehler

Auch bei der simplen Auswertung der Prüfungsfragen ist dem IMPP ein Fehler unterlaufen. Um Täuschungsversuche zu

erschweren, wird die Prüfung bekanntlich in zwei Auflagen gedruckt (Auflagen A und B), so daß in der Prüfung nebeneinander sitzende Studenten zwar insgesamt dieselben Fragen bearbeiten, jedoch in einer unterschiedlichen Reihenfolge. Beim Märzphysikum ist das Kuriosum passiert, daß ein und dieselbe Frage in Auflage B mit Antwort "E" als richtig bewertet wurde (Frage 105), in Auflage A jedoch mit "C" (Frage 135). Die Lösung C wäre für beide korrekt gewesen. Die Folge dieser krassen Fehlleistung war, daß den Studenten, die bei korrekter Auswertung der Auflage B das Physikum gerade bestanden hätten - das Markieren von "C" bei Frage 105 vorausgesetzt -, mitgeteilt wurde, sie seien durch die Prüfung gefallen.

Obendrein ist diese Fehlleistung beim ABC der Prüfungsauswertung nicht den Landesprüfungsämtern, sondern den Betroffenen selber aufgefallen.

Eine gütige Ironie liegt darin, daß der Fehler letztlich die Mißerfolgsquote verringert, da Prüflingen, die "E" angekreuzt haben, im Nachhinein kein Punkt aberkannt wurde, wodurch einige bestanden haben, die bei sachgemäßer Auswertung durchgefallen wären ("begünstigender Verwaltungsakt"). Dieser Sachverhalt soll hier nicht strapaziert werden, aber ich glaube doch, daß er anzeigt, daß entgegen allen bisherigen Annahmen auch einfachste Kontrollroutinen bei der Auswertung der Prüfung im IMPP nicht fehlerfrei funktioniert haben.

Wiederholungsfragen

Die "Sachverständigenkommission zur Prüfung der Gründe für das Ergebnis der ärztlichen Vorprüfung im März 1981" hat in ihrem Bericht auch die Bedeutung sog. Altfragen bzw. Wiederholungsfragen für das Gesamtergebnis der Prüfungen analysiert; sie stellte fest,

daß die Handhabung des Wiederholungsfragenanteils in den Jahren 1979 -1981 durch das IMPP "auf einen unverantwortlich leichtfertigen Umgang mit dem empfindlichen Instrument der Prüfungsschwierigkeit" hindeute (Sachverständigen-Kommission, 1982a, S. 27).

Während im Herbst 1980 40% der Fragen Wiederholungsfragen waren, wurde deren Anteil im März 1981 auf 20% reduziert. Diese Reduktion war für den Ergebniseinbruch vom Frühjahr 1981 wesentlich verantwortlich.

Am Beispiel der Physiologie läßt sich zeigen, wie sehr der Altfragenanteil das Ergebnis determiniert. Die Sachverständigen-Kommission hat in einer zweiten Stellungnahme darauf aufmerksam gemacht, daß im Fach Physiologie "das Prüfungsheft vom Herbst 1981 wieder ähnlich viele Altfragen wie im Herbst 1980" enthielt (Sachverständigen-Kommission, 1982b, S. 16).

Das Ergebnis schnellte denn auch auf 72,3% durchschnittlich richtig beantwortete Fragen hoch (s. Tab. 1).

Im August 1982 wurde der Anteil der Altfragen auf weniger als 20% - davon die meisten modifiziert - gesenkt mit der Folge, daß die durchschnittliche Prüfungsleistung in der Physiologie auf 51,6% absackte (Diel et al., 1983, S. 196). Es darf vermutet werden, daß der Wiederholungsfragenanteil beim letzten Physikum noch weit geringer war.

Um die Prüfungsergebnisse zu stabilisieren, hatten sich die Gesundheitsminister 1981 offenbar auf einen bestimmten Anteil von Altfragen in den Prüfungen geeinigt: "Obwohl nun die Gesundheitsminister im Vorjahre bezüglich der Konstanz der Prüfungsschwierigkeit auch auf einen bestimmten Anteil von Altfragen abgehoben haben, konnte bei der letzten Gesundheitsministerkonferenz am 4. Juni 1982 in

München eine Klärung dahingehend erreicht werden, daß unter Altfragen nicht nur wortidentische Fragen zu verstehen seien, sondern auch modifizierte Altfragen, bei denen entweder der Fragenstamm oder die Distraktoren verändert sind." (Göttsching, 1982, S. 81f.)

Unübersehbar ist diese Einigung bezüglich der Konstanz der Prüfungsschwierigkeit, deren genauer Inhalt der Öffentlichkeit unbekannt ist, im März 1985 wirkungslos geblieben. Weder die Ergebnisberichte über die schriftlichen Prüfungen noch die Item-Analysen, die das IMPP den Universitäten zur Verfügung stellt, enthalten Informationen über Wiederholungsfragen; es liegt hier eine Grauzone der Steuerung der Prüfungsschwierigkeit vor, die dringend der Erhellung bedarf. Es ist zu fordern, daß die Entwicklung des Anteils der Wiederholungsfragen - und zwar getrennt nach unveränderten und modifizierten - für die Zeit von 1981 bis 1985 offengelegt wird.

Staatsvertrag

Als Folge des "März-Debakels" von 1981 waren von der GMK Änderungen des Staatsvertrages zur Errichtung des IMPP sowie bei den Richtlinien des Verwaltungsrats über die Arbeitsweise des Institutes angekündigt worden. Den medizinischen Fakultäten und wissenschaftlichen Fachgesellschaften wurde in Artikel 8 des geänderten Staatsvertrags ein förmliches Vorschlagsrecht für die Mitglieder der Sachverständigen-Kommissionen eingeräumt. Außerdem sollte "durch die Schaffung von Kontroll-Kommissionen eine Kontrolle des Aufgabenhefts (d.h. der in einem Prüfungstermin vorgesehenen Fragen) erfolgen...Die für jeden Prüfungsabschnitt vorgesehenen Kontrollkommissionen sind bereits in Funktion...Ihre

letztlich geäußerte Fachmeinung ist ausschlaggebend für die endgültige Zusammensetzung des Prüfungsheftes."

(Göttsching, 1982, S. 80)

Richtschnur der Kontrollkommissionen für die medizinischen Prüfungen ist § 14 Abs. 2 ÄAppO: "Die Prüfungsfragen müssen auf die für den Arzt allgemein erforderlichen Kenntnisse abgestellt sein und zuverlässige Prüfungsergebnisse ermöglichen."

Es muß leider festgestellt werden, daß die Kontrollkommission bei der Überprüfung des Aufgabenheftes vom März 1985 aus bislang unbekanntem Gründen offenbar versagt hat.

Gleitklausel

Die politisch Verantwortlichen in Bund und Ländern sind in der Pflicht, Abhilfe zu schaffen. Bereits nach dem verunglückten Märzphysikum von 1981 sind Änderungen bei der Arbeitsweise des IMPP angekündigt und realisiert worden. Wie sich nun gezeigt hat, mit wenig Erfolg. 1978 hatte der Gesetzgeber die sog. Gleitklausel bei den Prüfungen abgeschafft, damals auch in der trügerischen Gewißheit, daß das IMPP die Prüfungen im Griff habe. Diese Gleitklausel war ein vernünftiger Mechanismus zum Ausgleich von allzu schweren Prüfungen. 1981, bei der 3. Novellierung der Approbationsordnung, wurde der nächste Fehler gemacht, indem die Gleitklausel zwar richtigerweise wieder eingeführt, aber mit einem starren 50%-Anker versehen wurde. Im März 1985 hat sich nun erwiesen, daß auch die Regelung von 1981 mangelhaft war: Sie reicht nicht aus, um ein Gegengewicht gegen Prüfungen zu bilden, die so extrem ausfallen wie das letzte Physikum.

Es muß betont werden, daß hier ein politischer Fehler vorliegt, für den nicht das IMPP verantwortlich ist. Im Gegenteil hat sich Kraemer 1981 gegen

die starre 50%-Regelung ausgesprochen: "Wir alle sollen bedenken, daß eine Bestehensregel konkret über das Schicksal junger Menschen mitentscheidet. Unter den jetzigen Rahmenbedingungen halten wir daher auch den vorgeschlagenen sog. Anker von 50% in der Bestehensregel für nicht kompatibel." (Kraemer, 1981, S. 58)

Konsequenzen

Angesichts der dargelegten Fakten und Sachverhalte muß es verwundern, wenn die Gesundheitsminister und Senatoren der Länder zur Tagesordnung übergehen wollen; laut FAZ vom 30.3.1985 sagte beispielsweise der rheinland-pfälzische Sozialminister: "Es ist weder ein Anlaß noch ein Ansatz zu Nachbesserungen sichtbar." Der Sprecher des BMJFG versicherte, man habe - anders als 1981 - "die politische Absicht, hart zu bleiben" (FAZ vom 12.4.1985).

Im richtigen Moment - gegenüber falschen Ansprüchen - hart zu bleiben, ist eine Fähigkeit, die gerade auch von Politikern gefordert ist. Nur: Härte gegenüber den Opfern einer unprofessionell arbeitenden Anstalt des öffentlichen Rechts erscheint mir nicht als Ausdruck politischer Souveränität, sondern eher als Zeichen dafür, daß eigene Fehler gar nicht erst zur Kenntnis genommen werden sollen.

Es geht hier nicht darum, der Unzahl emotionsgeladener Ausbrüche gegenüber dem IMPP einen weiteren anzufügen; das Institut erfüllt eine wichtige - und zweifellos undankbare - Aufgabe. Aber nach all den Pannen und Fehlern sind wirksame Reformen im IMPP und auch in der Arbeitsweise seines Verwaltungsrats überfällig; die politisch Verantwortlichen - insbesondere die GMK - hatten sich 1981/82 dieser Notwendigkeit verschlos-

sen (Sachverständigen-Kommission, 1982b, S. 14f.). Vorschläge hierzu sind sowohl vom Medizinischen Fakultätentag als auch von der "Sachverständigen-Kommission" gemacht worden.

Literatur

Diel, R., G. Einhorn, M. Gerspach, R. Lohölter, W. Schoeppe:
Ärztliche Vorprüfung August 1982: Wessen Problem? Eine aktuelle Kurzanalyse.
In: Med. Welt
1983, 34, S. 194-197

Göttsching:
Zur Aufgabenerfüllung des IMPP.
In: Protokoll des Ordentlichen Medizinischen Fakultätentages der Bundesrepublik Deutschland am 11. u. 12. Juni 1982 in Freiburg/Br., S. 76-82

Kraemer, H.-J.:
Zur Entwicklung der schriftlichen Mediziner-Prüfungen von 1974-1981.
In: Protokoll des Ordentlichen Medizinischen Fakultätentages der Bundesrepublik Deutschland am 19. u. 20. Juni 1981 in Mainz, S. 54-59

Renschler, H.E., G.P. Burkhard:
Benotung des Erfolges der ärztlichen Ausbildung.
In: Medizinische Ausbildung
1984, 1, S. 19-24

Sachverständigen-Kommission zur Prüfung der Gründe für das Ergebnis der Ärztlichen Vorprüfung im März 1981: Arbeitsbericht.
In: DUZ
1982a, 30, H. 3, S. 25-29

Sachverständigen-Kommission zur Prüfung der Gründe für das Ergebnis der Ärztlichen Vorprüfung im März 1981: Wachsende Fernsteuerung in der Medizin. Die Sachverständigen-Kommission bezieht erneut Stellung zur Prüfungsgestaltung am IMPP.
In: DUZ
1982b, 30, H. 9, S. 14-16

Verwaltungsrat des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP):
Stellungnahme zum Schlußbericht der unabhängigen Sachverständigen-Arbeitsgruppe zur Prüfung der Gründe für das Ergebnis der Ärztlichen Vorprüfung im März 1981.
In: DUZ
1982, 30, H. 4, S. 22-23

TEILNEHMERBEWERTUNG VON VORTRÄGEN

G. P. Burkhard; H. E. Renschler, Bonn

Neben der Ausbildung kommt der Fortbildung zur Qualitätssicherung der ärztlichen Tätigkeit eine wichtige Bedeutung zu. Hier wie dort sollte es um die Feststellung sowie Steigerung der Effektivität von Lernprozessen, dementsprechend um die Optimierung des Lernverhaltens, nicht zuletzt aber auch um die des Lehrerverhaltens gehen. Gerade diese Frage steht im Gegensatz zu anderen Ländern bei uns stark im Hintergrund. Es wird fast immer darüber diskutiert, was vermittelt werden soll, jedoch nicht wie es vermittelt werden soll, um eine größtmögliche Effektivität zu erreichen.

Hier wird darauf vertraut, daß der Lehrer ankommt und die Fähigkeit, etwas weiterzugeben, beherrscht. Wenn man dem Lehrer klarmachen könnte, daß dieses Gottvertrauen unbegründet ist und in den Augen der Zielgruppe vielfach voll-

kommen fehl am Platz ist, wäre viel erreicht. Erst das Bewußtmachen bzw. -werden einer Situation macht ihre Veränderung möglich.

Ausgehend von dieser Grundsituation entstand die Frage, wie ein Evaluationsansatz zur Feststellung der Lehrerqualität durch die Teilnehmer aussehen könnte, der die Forderung der Reliabilität, Validität, Praktikabilität erfüllt und der zu Entscheidungen hinsichtlich dieses Lehrers bzw. Vortragenden führt. Im folgenden soll berichtet werden über einen entsprechenden Versuch, der seit 1980 im Rahmen eines von der Akademie für Ärztliche Fortbildung Hessen abgehaltenen Notfalldienstseminars durchgeführt wird.

1. Sinn und Möglichkeit der Unterrichtsevaluation

Bei der Planung und Implementierung von Unterrichtsaktivitäten ist die Evaluation eine notwendige Komponente in diesem Prozeß. Trotzdem wird die Evaluation oft bewußt oder nicht bewußt vernachlässigt. Die Gründe für das Unterlassen scheinen dabei vielfältiger Natur zu sein.

a) Zunächst wird der Prozeß der Evaluation oft nicht verstanden. Die umfangreiche Literatur und die Breite dieses Gebietes führen oft zur Verwirrung.

b) Die Wichtigkeit der Evaluation wird vielfach nicht richtig erkannt oder bei richtiger Einschätzung dieses Prozesses werden die Ergebnisse nicht benötigt. Damit erscheint die Evaluation als nutzloser Prozeß.

c) Der Prozeß der Evaluation ist oft recht arbeits- und damit zeitintensiv, wenn er nicht systematisch und regelmäßig durchgeführt wird.

Drei Gründe scheinen jedoch für eine derartige Evaluation zu sprechen:

- a) Der Erfolg eines Unterrichtes kann erst hierdurch sichergestellt werden. Aspekte, die ineffizient sind, müssen bzw. können modifiziert werden.
- b) Der Aufwand für einen Unterricht bzw. Veranstaltung kann nachträglich gerechtfertigt werden.
- c) Es kann nach und nach der Standard erhöht werden.

Nach STUFFELBEAM 1971 bedeutet Evaluation "die Gewinnung von Informationen durch formale Mittel, wie Kriterien,

Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällen von Urteilen in Entscheidungssituationen zu erhalten".

Sie ist demnach nicht Selbstzweck, sondern soll helfen über Entscheidungen zu Verbesserungen zu gelangen. Nach Festlegung der Inhalte, Methoden, Referenten sollte eine Veranstaltung konzipiert werden und die Evaluation anschließend Auskunft geben über die erreichte Qualität. Diese Ergebnisse müssen im Sinne eines Kreisprozesses wieder zu Veränderungen führen (Abb. 1).

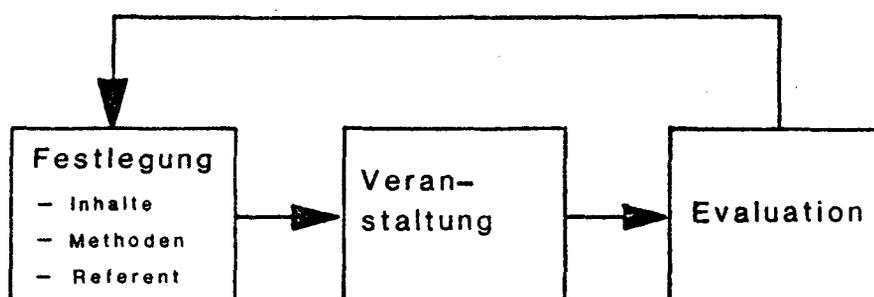


ABB.1: Bedeutung der Evaluation im Rahmen des Unterrichts

2. Operationalisierung

Der vorgegebene Rahmen der Evaluation betraf das eingangs erwähnte Notfalldienstseminar der Landesärztekammer Hessen, das seit 1979 etwa 15 Mal durchgeführt wurde und weiter durchgeführt wird. Es sollte die daran teilnehmenden Ärzte durch 19 Referate zu 5 verschiedenen Notfallthemenbereichen für Notfallbehandlung kompetenter machen. Die Aufgabe bestand in einer Mitfestlegung von Inhalten aufgrund von Analysen von Notfallstatistiken und Befragung der Teilnehmer,

der Mitgestaltung der Veranstaltung, der Schulung der immer wieder eingesetzten Referenten und schließlich in der Evaluation der Veranstaltung. Dieser letzte Aspekt zielte einmal auf den Lerneffekt der Veranstaltung ab, zum anderen auf die Einschätzung der Referenten durch sogenannte Experten und durch Teilnehmer. Diese Einschätzungen wurden für Rückmeldungen an die Referenten herangezogen, damit zur eingangs geforderten Bewußtmachung der eigenen Lehrerkompetenz

und zur langfristigen Verbesserung der Qualität des Vortragsverhaltens. Damit hatte dieser Evaluationsansatz Entscheidungscharakter.

3. Evaluationskriterien der Teilnehmerbewertung

Eine Durchsicht der relevanten deutschen Literatur ergab kaum Hinweise auf Kriterien für eine Bewertung von Vortragenden. Im amerikanischen Bereich fanden sich verschiedene Bewertungssysteme. Beispielhaft seien die 24 Items umfassende Skala 'Rating your lecture' von FOLEY vom Center for Educational Development, University of Illinois, 1976 oder der 'Audience Response Survey' vom Department of Postgraduate Medical Studies mit 12 Items genannt. Oft sind diese Beurteilungsteile jedoch so komplex und differenziert, daß vor der Einschätzung ein sogenanntes Ratertraining erfolgen muß. Dies kann nicht das Ziel sein.

Wie Popham 1975 darstellt, hat es seit 1900 verschiedene Versuche gegeben, befriedigende Kriterien für Effektivität eines Lehrers bzw. Vortragenden zu finden. Bis heute gibt es jedoch keine einheitlichen Kriterien, die sowohl ihre Brauchbarkeit, als auch ihre Genauigkeit und Gültigkeit unter Beweis gestellt haben. Die an derartige Einschätzungs-

4. Stichprobe

Seit 1981 kommen die Einschätzungen nach Relevanz hier 'Notfallrelevanz' und 'Lernerfolg' zum Einsatz, ab 1983 das Kriterium des Gefallens. Die im folgenden dargestellten Ergebnisse basieren auf etwa 1400 teilneh-

Aus dieser gesamten Thematik soll entsprechend der Themenstellung nur die Teilnehmerbewertung weiterverfolgt werden.

kriterien gestellten Anforderungen waren:

- wenige Items
- für Rater gut verständliche Items
- quantitatives Urteil für
 - Feedback an Referenten
 - Rangbildung unter den Referenten.

Als Kriterien, die diese Forderung erfüllen sollten, wurden gewählt:

1. Wie relevant war das Referat, der Vortrag in Bezug auf das Unterrichtsziel?
2. Wie hoch war der Lernerfolg?
3. Wie hat der Vortrag/Vortragende gefallen?

Da eine Aussage bezüglich der Ausprägungsstärke der genannten Kriterien gemacht werden sollte, wurde eine ordinale 5-Punkte-Skala gewählt, die von 1 = sehr hoch bis 5 = überhaupt nicht reichte.

Diese Einschätzungen wurden ergänzt durch qualitative Aussagen, wie freie Angaben und Persönlichkeitsangaben.

menden Ärzten bei 7 Veranstaltungen mit jeweils 19 Referenten. Übergreifend betrug der Rücklauf etwa 76 %. Es kann somit von einer zufriedenstellenden Datenbasis ausgegangen werden.

5. Ergebnisse

Jedes Erfassungsinstrument muß, um die Bewertungsfunktion erfüllen zu können, eine gewisse Meßgenauigkeit und Gültigkeit unter Beweis stellen. Meßgenau ist ein Instrument bzw. Item dann, wenn es bei Wiederholung am gleichen Objekt zum gleichen Ergebnis kommt. Dies ist im vorliegenden Fall nur schwer machbar, da dieselben Teilnehmer denselben Referenten mit identischem Vortragsverhalten in der Regel nicht noch einmal sehen und bewerten. Ein Maß für die Meßgenauigkeit scheint im Grad der Obereinstimmung der Urteile hinsichtlich derselben Referenten zu liegen. Neben den Methoden, die von FERMAN, LANCRENON und GREMY (1979) referiert wurden und teilweise auf der Methode des χ^2 -Testes und teilweise auf Korrelationsberechnung basieren, gibt es ein von HORST (1949) vorgeschlagenes Verfahren zur Ermittlung der 'Reliabilität des Validitätskriteriums'. Dieses Verfahren macht eine Wiederholung der Beurteilung überflüssig, indem es sich auf die innere Konsistenz der subjektiven Schätzwerte stützt, wobei die Reliabilität der Durchschnittsbeurteilungen berechnet wird. Danach ergeben sich z.B. für die Veranstaltung im April 1984 folgende Koeffizienten:

Relevanz r_{cc}	=	0,95
Lernerfolg r_{cc}	=	0,97
Gefallen r_{cc}	=	0,99

Diese Ergebnisse müssen gemessen am Maximalwert von +1 als sehr hoch bezeichnet werden und lassen den Schluß auf eine sehr hohe Raterübereinstimmung, d.h. Meßgenauigkeit zu. Ähnlich hohe Werte ergeben sich für die anderen Veranstaltungen. Bei der Überprüfung der Validität ergeben sich in der Praxis noch sehr viel größere Schwierigkeiten. Als valide wird ein Instrument dann ange-

sehen, wenn es das mißt, was es zu messen beansprucht. Diese Überprüfung erfolgt an einem Außenkriterium, von dem angenommen wird, daß es Gültigkeit besitzt. Als Außenkriterium benutzt man Expertenurteile, die bei kritischer Betrachtung jedoch auch nicht valide sein müssen. Immerhin ist diese Methode ein operationaler Weg. In unserem Falle wurde ebenfalls auf eine Expertenbewertung zurückgegriffen. Bei der Frage des Zusammenhanges zwischen den Kriterien der Teilnehmerbewertung und der Expertenbewertung ergibt sich aus Daten von 7 Veranstaltungen ab 1981 folgendes Bild:

Relevanz r_{TE}	=	0,65
Lernerfolg r_{TE}	=	0,65
Gefallen r_{TE}	=	0,77

Alle Koeffizienten zeigen hochsignifikante Zusammenhänge. Dennoch scheint das Kriterium 'Gefallen' die für das Expertenurteil gültigste Größe darzustellen.

Eine weitere Frage betraf den Zusammenhang der drei Kriterien untereinander. Sollte sich eine hohe Interkorrelation herausstellen, so ist zu fragen, ob nicht nur ein Kriterium die Qualität des Vortrages bzw. Vortragenden erfassen konnte. Entsprechend der These, wonach im ganzheitlichen Erleben eine gewisse Obereinstimmung bei den Einzelgliedern vorzufinden ist, müßte sich ein hinter den Kriterien stehender ganzheitlicher Faktor finden lassen, der die Einzelkriterien determiniert. Viele vermuten einen sogenannten Happiness-Index (BERTRAM 1977) als gemeinsamen Faktor. Empirisch zeigen sich folgende Ergebnisse (Abb. 2).

Folgendes Bild, in dem die Regressionsgeraden für alle drei Zusammenhänge dargestellt sind, mag diese veranschaulichen (Abb. 3).

ABB.2: Interkorrelation zwischen Relevanz, Lernerfolg, Gefallen April 84

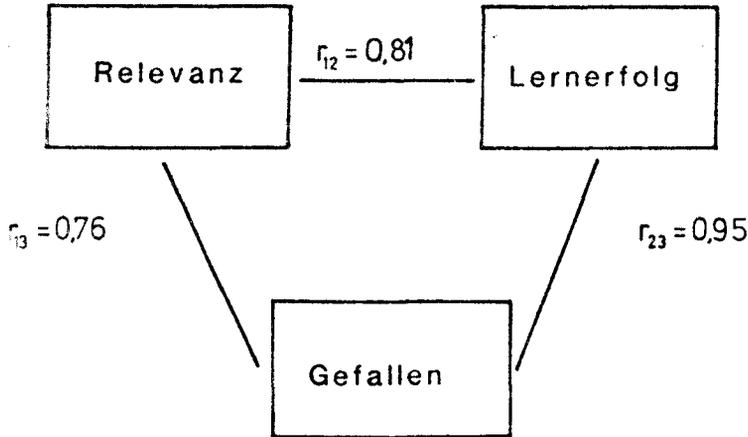
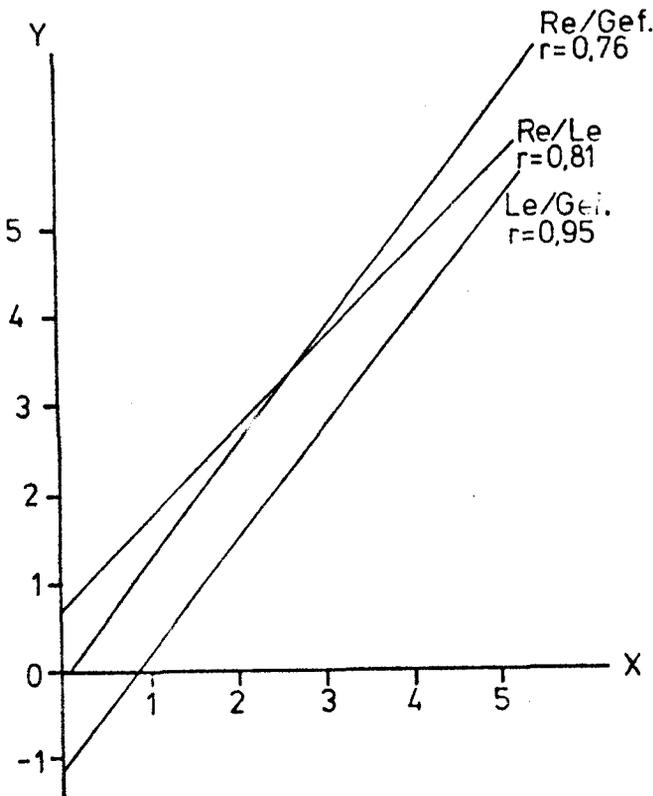


ABB.3

Zusammenhang zwischen Relevanz/Gefallen
Relevanz/Lernerfolg
Lernerfolg/Gefallen
April 1984

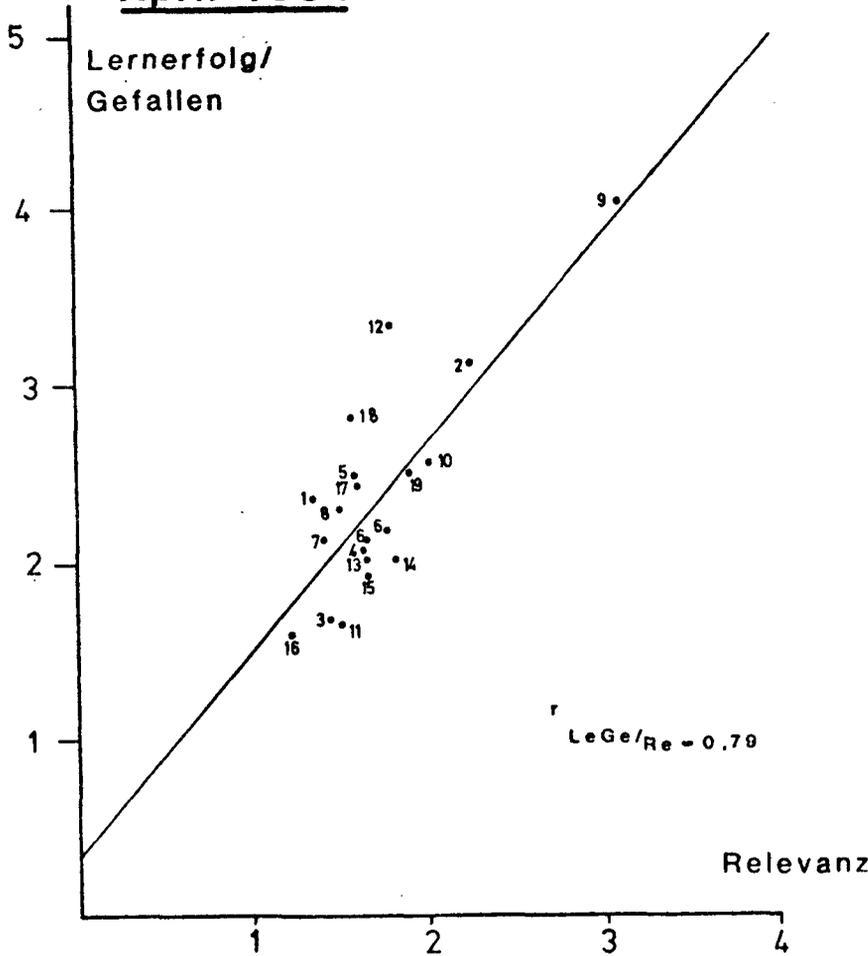


Obwohl alle drei Koeffizienten hochsignifikant sind, zeigt sich bei dem Zusammenhang Lernerfolg/Gefallen am ehesten ein gemeinsamer Faktor. Offensichtlich ist die Frage nach der Relevanz eines Themas unabhängig zu sehen. Inhaltlich ist dies plausibel. Die Relevanz eines Themas hat noch nichts mit dessen Lernerfolg zu tun. Andererseits wird ein Vortrag, bei dem viel gelernt wurde auch gefallen haben und umgekehrt. Insgesamt sprechen diese Ergebnisse für die Anwendbarkeit der vorliegenden Kriterien. An einem konkreten Beispiel von April 1984 mag das verdeutlicht werden (Abb. 4).

ABB.4

Zusammenhang zwischen
Lernerfolg/Gefallen und Relevanz

April 1984



Auf der y-Achse sind Gefallen und Lernerfolg rechnerisch über Mittelwert zusammengefaßt, auf der x-Achse die Relevanz.

Entsprechend der Aufgabe der Referentenevaluation wäre beispielsweise ein Referent entweder einer weiteren Analyse und Schulung zuzuführen, dessen Thema eine hohe Relevanzeinschätzung ergibt, der jedoch bei Lernerfolg und Gefallen schlecht abschneidet. Wie auf der Abbildung zu sehen ist, wären die Referenten 9, 12, 2, 18 zunächst einer weiteren Analyse zu unterwerfen, einer Schulung zuzuführen oder bei Mißerfolg dieser Schulung auszuschließen.

Damit wäre die Forderung erfüllt, die STUFFELBEAM in seiner Bestimmung der Evaluation genannt hat. Evaluation darf danach nicht Selbstzweck sein, sondern muß Informationen liefern, die Entscheidungen möglich machen.

Zusammenfassung

Ausgehend von der Bedeutung der Unterrichtsevaluation wurde als ein Bereich die Evaluation des Lehrers bzw. des Referenten ausgegliedert und eine Möglichkeit in der Teilnehmerbewertung gesehen.

Die von uns vorgeschlagenen Bewertungskriterien der Relevanz eines Themas gemessen am Ziel, des Lernerfolges und

Gefallens wurden anhand von empirischen Daten der Jahre 1981-1984 aus den Notfalldienstseminaren auf Reliabilität, Validität, Interkorrelation überprüft und für sehr brauchbar befunden. An einem konkreten Beispiel wurde ihre Funktion als Entscheidungsgrundlage bei der Evaluation aufgezeigt.

LITERATUR

1. ALPEROVITSCH, A., de BOMBAL, F.T., GREMY, F. (Ed.)
Evaluation of Efficacy of Medical Action
Amsterdam : North Holland Publishing Company, 1979
2. BERTRAM, D.A., BROOKS-BERTRAM, D.A.
The Evaluation of Continuing Medical Education: A Literature Review
Health Education Monographs 5, 1977
3. FERMANIAN, J., LANCRENON, S., GREMY, F.
Companion of Several Coefficients of Agreement between Two Judges Used in Medical Decision.-
In: Evaluation of Efficacy of Medical Action
ALPEROVITSCH, A., de BOMBAL, F.T., GREMY, F. (Ed.)
Amsterdam : North-Holland Publishing Company, 1979
4. FOLEY, R., SMILANSKY, J., BUCHMAN, E., SAJID, A.
Improving Lecture Skills, Instructional Guide
The Center for Educational Development, Univ. of Illinois at the Medical Center
Chicago : Aldine Publishing Company, 1976
5. HORST, A.P.
A Generalized Expression for the Reliability of Measures
Psychom. 14 : 21
6. LIENERT, G.A.
Testaufbau und Testanalyse
Weinheim : Verlag Julius Beltz, 1969
7. POPHAM, W.J.
Educational Evaluation
New Jersey : Prentice Hall, 1975
8. SOMMER, J., PETERMANN, Fr.
Deskriptive und präskriptive Aussagen in einem Fragebogen zur Beurteilung akademischer Lehrveranstaltungen
Zeitschrift für Entwicklungspsychologie und Päd. Psychologie 1978, Band X, Heft 4, S 336-346
9. STUFFELBEAM, D.L., FOLEY, W.J., GEPHART, W.J., GUBA, E.G., HAMMOND, R.L., MERRMAN, H.O., PROUNS, M.M.
Educational Evaluation and Decision Making
Itasca, I 11.: F.E. Peacock, 1971
10. STUFFELBEAM, D.L.
Alternative Approaches to Educational Evaluation
In: Evaluation in Education: Current Application
Berkeley, Calif. : Mc Cutchan Publishing , 1974
11. Audience Response Survey
Department of Postgraduate Medical Studies
Manchester, M60 7LP

Dr. G. B. Burkhard und Prof. Dr. H. E. Renschler, Institut für Didaktik der Medizin, Sigmund-Freud-Straße 25, D 5300 Bonn 1

EINSTELLUNGSÄNDERUNGEN BEI MEDIZINSTUDENTEN IM VERLAUF DES PSYCHIATRISCHEN UNTERSUCHUNGSKURSES

M. van der Ven; G. Buchkremer, Münster

Zusammenfassung:

In dieser Studie wird untersucht, ob sich die bei Medizinstudenten vermuteten negativen Einstellungen gegenüber "der Psychiatrie" durch den Psychiatrie-Unterricht abbauen lassen. Hierzu wurden die Teilnehmer des psychiatrischen Untersuchungskurses, der im Rahmen des "Kurses der allgemeinen klinischen Untersuchungen in dem nicht-operativen und operativen Stoffgebiet" angeboten wird, befragt. 186 Medizinstudenten füllten am Anfang und am Ende des Praktikums Fragebögen aus, auf denen sie verschiedene, psychisch Kranken zugeschriebene Eigenschaften bewerteten sowie zu wünschenswerten Psychiaterereigenschaften Stellung nahmen. Der Vergleich zeigt, daß der psychiatrische Untersuchungskurs dazu beitrug, negative Einstellungen gegenüber psychisch Kranken und Psychiatern abzumildern.

1. Einleitung

Obwohl die Weiterbildung in der Psychiatrie dem frisch approbierten Arzt derzeit relativ günstigere Berufschancen eröffnet als in den meisten anderen Gebieten, ist das Interesse an einer psychiatrischen Weiterbildung unter den Berufsanfängern vergleichsweise gering (Braune und Heitz 1972, Eagle und Marcos 1980, Mc Cormick und Voineskos 1978, Moos und Yalom 1966, Nielsen 1980, Walton 1969).

Hierfür dürfte eine Reihe von Gründen maßgeblich sein, u.a. auch die Erfahrungen des Studierenden im Psychiatrie-Unterricht. So besteht nach den Untersuchungen von Nielsen (1980) und Maxmen (1979) eine Beziehung zwischen der Qualität des Psychiatrie-Unterrichts und dem

späteren Wunsch nach Weiterbildung auf psychiatrischem Gebiet.

Nach Untersuchungen von Braune und Heitz (1972) wird der Psychiater von den Studenten als weltfern, außerordentlich geschickig und geschickt im Umgang mit Patienten charakterisiert. Außerdem werden ihm besondere Kontaktfähigkeit und Einfühlungsvermögen in interpersonelle Beziehungen bescheinigt. Diesen recht positiven Einschätzungen des Psychiaters steht aber die Ansicht von immerhin 38 % der Studenten gegenüber, daß fast jeder Psychiater "einen kleinen Stich" hätte, und 10 % vertraten die Ansicht, daß Psychiater "selbst einmal durchdrehen" würden. Auch nach Bruhn und Parsons (1964) sehen Medizinstudenten die Psychiater als an intellektuellen Problemen interessierte, emotional instabile und im Denken etwas verwirrte Menschen an. Moss und Yalom (1966) fanden eine positivere Einschätzung der Medizinstudenten gegenüber dem Psychiater: Psychiater seien risikofreudig, kompliziert, emotional, idealistisch, sensitiv, tolerant und unkonventionell. Nach dem Psychiatrie-Praktikum werden sie sogar noch positiver, nämlich als bedeutsam, freundlich, höflich, informiert, geachtet, friedfertig, humorvoll und entspannt bezeichnet. Ähnliche Erfahrungen beschrieben auch Mc Cormick und Voineskos (1978).

Eine Erklärung für das geringe Interesse an einer psychiatrischen Weiterbildung läßt sich möglicherweise in den von Moos und Yalom (1966) beschriebenen Studienbedingungen an den Hochschulen finden, die von Medizinstudenten eher Eigenschaften wie Selbstbewußtsein, Durchsetzungsvermögen, Maskulinität und Selbstvertrauen erfordern; die beschriebenen Psychiatermerkmale stehen dazu in deutlichem Kontrast. Zudem sind die Ansprüche der Studenten an einen "idealen Arzt" von eigenen Persönlichkeitsmerkmalen abhängig: Viele Medizinstudenten sind sehr "autoritär" eingestellt und halten es umso weniger für erforderlich, daß ein Arzt zwischenmenschliche Beziehungen versteht (Parker 1958). Schließlich läßt sich vermuten, daß die von Lorenzen (1981) beschriebene herbe öffentliche Kritik an der psychiatrischen Versorgung oftmals übertriebene und idealisierte Anforderungen an die Psychiatrie und damit auch an ihre Ärzte hervorruft. Die entsprechend hochgeschraubten Erwartungen

bezüglich wünschenswerter Psychiater-eigenschaften könnten sich somit als besonders hinderlich für einen späteren Wunsch nach psychiatrischer Weiterbildung erweisen.

Ein weiterer Hinderungsgrund für diese Berufswahl ergibt sich, wenn in einer Ärztebefragung (Koch 1975) der "ideale Patient" als "klar, geordnet, hilfsbereit, gelöst, konzentriert, friedlich, vergnügt, frisch" beschrieben wird. Psychisch Kranke hingegen dürften weitgehend mit den Gegensätzen dieser Adjektive in Verbindung gebracht werden. Nach Braune und Heitz (1972) meinen 40 % der befragten Medizinstudenten, daß Geisteskrankte eine Gefahr für die Umwelt seien. Ebenso viele Studenten wollten die Geisteskranken an ihrer Fortpflanzung hindern, und 25 % der befragten Studenten setzten sich dafür ein, daß ein mit Sicherheit Geisteskranker dauernd in einer psychiatrischen Anstalt verwahrt werden soll. Obwohl die Autoren selbst die bei der Befragung verwendete Formulierung "geisteskrank" nachher als unglücklich bezeichneten, sprechen diese Zahlen doch dafür, daß eine negative Einstellung zu psychisch Kranken unter Medizinstudenten weit verbreitet ist.

Die Frage, ob sich derartige Einstellungen im Verlauf des psychiatrischen Unterrichts verändern lassen, wird von Meltzer und Grigorian (1972) bejaht. So hätten die Studenten Forderungen nach restriktiven Maßnahmen gegenüber Patienten nach Beendigung des Praktikums nur noch in abgeschwächter Form vorgetragen. Die Kennzeichnung der erstrebenswerten Arzt-Patient-Beziehung sei nachher weniger hierarchisch ausgefallen, und das Verständnis für die psychodynamischen Zusammenhänge der Krankheitsbilder habe beträchtlich zugenommen. Zu ähnlichen Ergebnissen kamen auch Crow et al. (1970) und Maurice et al. (1975). Maxmen (1979) konnte belegen, daß nach Beendigung eines Psychiatriepraktikums weitaus mehr Studenten als vorher glaubten, in der Psychiatrie könne den Patienten wirksam geholfen werden.

Die Einstellungsänderung der Medizinstudenten gegenüber den Patienten ist auch von der Patientenauswahl im psychiatrischen Unterricht abhängig (Eagle et al. 1979). Ein gesteigertes Interesse an der Psychiatrie zeigten nach dem Praktikum vor allem die Studenten, die mit Patienten konfrontiert wurden, die nicht zwangseingewiesen waren und die therapeutisch gut und schnell zu beeinflussen waren (Eagle u. Marcos 1979). Schließlich beschreiben Miller et al. (1979), daß die im Verlauf des Psychiatrie-Unterrichts festgestellten Einstellungsänderungen der Studenten bis zum Examen und auch danach bestehen blieben.

Durch unsere Untersuchung versuchten wir zwei Fragen zu klären:

1. Verändert sich die Einstellung der Medizinstudenten durch einen psychiatrischen Untersuchungskurs hinsichtlich wünschenswerter Eigenschaften des Psychiaters?
2. Verändert sich die Einstellung der Medizinstudenten hinsichtlich der bei psychisch Kranken vermuteten Eigenschaften?

2. Methode:

2.1. Der psychiatrische Untersuchungskurs

Im Rahmen des "Kurses der allgemeinen klinischen Untersuchungen", der zu den scheinpflichtigen Praktika der Medizinstudenten des 1. und 2. klinischen Semesters gehört, ist ein psychiatrischer Untersuchungskurs als fester Bestandteil nur an wenigen deutschen Universitäten eingeführt. An der Universität Münster ist der psychiatrische Teil des Untersuchungskurses - der nicht mit dem im späteren Verlauf des Studiums vorgesehenen "Praktikum der Psychiatrie" verwechselt werden sollte - folgendermaßen gegliedert:

Nach einer einmaligen, theoretischen Einführungsveranstaltung für alle Studenten findet der Unterricht in Gruppen von jeweils 10 bis 15 Studenten an vier Nachmittagen statt. Für jede Sitzung stehen 2 1/2 Unterrichtsstunden zur Verfügung, in denen von den Studenten mindestens ein Patient untersucht wird. Dabei ordnet sich die Studentengruppe in Zwiebelschalenform an (vgl. Schepank 1976). Im inneren Kreis befinden sich der Patient und die zwei bis vier Studenten, die das Interview führen. Jeder der teilnehmenden Studenten sollte während des Kurses mindestens einmal in den inneren Kreis kommen und somit Gelegenheit erhalten, aktiv ein Gespräch mit dem Patienten führen zu können. Der äußere Kreis beobachtet die Interaktionsprozesse, die einerseits unter den Studenten selbst, andererseits zwischen den Studenten und dem Patienten ablaufen. Dabei spielen Wahrnehmungsübungen vor allem der nonverbalen Interaktionsanteile eine besondere Rolle. Am Ende des Praktikums findet eine Abschlusveranstaltung statt, in der testpsychologische Verfahren und körperliche Untersuchungsverfahren vor-

gestellt werden. Außerdem wird ein Teil dieser Veranstaltung der Unterrichtskritik gewidmet.

Dieser psychiatrische Untersuchungskurs zielt weniger auf Wissenserwerb als auf Vermittlung emotionaler, sozialer und praktischer Fertigkeiten hinsichtlich der Arzt-Patient-Beziehung ab. Hierzu dienen die Einübung in den Umgang mit dem Kranken und in das ärztliche Gespräch (vgl. Dahme 1974) sowie die Anleitung zur Fremd- und Selbstwahrnehmung. Die Erhebung eines genauen psychischen Befundes - frei von persönlichen Interpretationen und Mutmaßungen - fällt vielen Studenten anfangs schwer (Finzen 1970). Der Untersuchungskurs schließt also auch Selbsterfahrung ein.

2.2. Fragebögen vor und nach dem Praktikum

Im ersten Teil des Fragebogens wird nach wünschenswerten Eigenschaften und Charakterisierungen eines Psychiaters gefragt, im zweiten Teil nach Eigenschaften und Charakterisierungen, die nach Ansicht der Medizinstudenten auf psychisch Kranke zutreffen. Der Student soll also darüber Auskunft geben, wie der Psychiater idealiter sein soll und wie er sich den psychisch Kranken realiter vorstellt.

In jedem Teil des Fragebogens wurden den Studenten jeweils 24 Adjektive vorgelegt, bei denen sie mit Hilfe einer Ordinalskala bewerten sollten, in welchem Maße die einzelnen Adjektive zutreffend bzw. unzutreffend waren [völlig (1) - nahezu (2) - in mittlerem Maße (3) - kaum (4) - überhaupt nicht (5)]. Dabei markiert also die Zahl 3 die Zahlenmitte, so daß sich bei der statistischen Auswertung für jedes Adjektiv ein Median errechnen ließ. Wir gingen bei der Auswahl der Adjektive auch von einigen Untersuchungen anderer Autoren aus (Fietkau 1976, Moos u. Yalom 1966, Stumme 1970, Stumme 1975) und selektierten in mehreren Vortests. Es verblieben schließlich 30 Adjektive, von denen insgesamt 16 bei beiden Teilen des Fragebogens vorkommen sollten (willensschwach, belastbar, empfindsam, mißtrauisch, tatkräftig, mit religiöser Bindung, schweigsam, ungepflegt, selbstsicher, aggressiv,

ordentlich, beeinflussbar, gefühlsbetont, verschlossen, geduldig und pessimistisch). Bei den "Patientenfragen" sollten zusätzlich die 8 Adjektive "intelligent, sachlich, kritisch, beruflich erfolgreich, friedlich, phantasiearm, träge und ichbezogen" bewertet werden. Bei den "Psychiaterfragen" kamen folgende 8 Adjektive hinzu: "seelisch kompliziert, fürsorglich, mit eigener psychischer Erkrankung, äußerlich gut aussehend, väterlich, vertrauens erweckend, politisch aktiv und selbstzufrieden".

Der Fragebogen wurde von jedem Studenten zweimal ausgefüllt - am ersten und am letzten Kurstag. In den Vortests äußerten die Studenten hinsichtlich der Fragebögen Bedenken, weil in ihnen sehr persönliche Einstellungen abgefragt würden. Deshalb wurden die Befragungen zunächst anonym durchgeführt. Für den statistischen Vergleich der ersten und der zweiten Befragung wurde aus diesem Grund der U-Test von Mann und Whitney als parameter-freier Rangtest für zwei unabhängige Stichproben herangezogen. Die in den Vortests geäußerten Bedenken hinsichtlich einer niedrigen Beteiligung an der Umfrage bestätigten sich bei der Hauptuntersuchung jedoch nicht. Deshalb wurden die Fragebögen im letzten Drittel der Umfrage mit Codewörtern versehen. Somit wurde ein direkter Vergleich der zwei Befragungen bei den einzelnen Studenten (und nicht nur der Gruppenmittel) möglich. Dementsprechend konnte der Wilcoxon-Test für paarige Stichproben angewandt werden. Außerdem stellte sich nach Abschluß der ersten Hälfte der Umfrage bei der Durchsicht der Fragebögen heraus, daß überraschend viele Studenten die Fragebögen mit persönlichen, z.T. sehr kritischen Anmerkungen versehen hatten. Aus diesem Grund wurden die Bögen bei der zweiten Hälfte der Umfrage mit dem ausdrücklichen Hinweis versehen, daß auf der freien Rückseite Raum für "Anmerkungen und Kritik" reserviert war.

2.3. Teilnehmer

Befragt wurden die Studenten des zweiten klinischen Semesters, in dem der psychiatrische Untersuchungskurs obligatorisch im Studienplan steht. Von 200 Studenten, die an dem Kurs teilnahmen, füllten 186 (93 %) bei der ersten Befragung die Bögen aus. Bei der zweiten Befragung waren es 159 Studenten (79,5 %). Bei 45 Studenten (22,5 %) konnte dem mit einem Codewort versehenen Fragebogen der ersten Befragung der entsprechende Fragebogen der zweiten Be-

fragung zugeordnet werden.

121 Studenten (65,1 % der 186 Probanden der ersten Befragung) waren männlich und 58 (31,2 %) weiblich. (7 Studierende versäumten die Angabe zum Geschlecht.) Dieses Verhältnis änderte sich bei der zweiten Befragung nur unwesentlich. Das mittlere Alter betrug 24,6 Jahre. Ein Krankenpflegepraktikum oder eine Famulatur im Bereich der Psychiatrie hatten 22 (11,8 %) der 186 Studenten absolviert.

3. Ergebnisse

3.1. Wünschenswerte Eigenschaften des Psychiaters

Auf die Frage, welche Eigenschaften und Charakterisierungen in den angeführten Adjektiven am ehesten ("völlig" bis "nahezu") bei Psychiatern erwünscht seien, wurden von den Studenten bei der ersten Befragung die Adjektive "geduldig, vertrauens-erweckend, belastbar, tatkräftig, empfindsam und fürsorglich" hervorgehoben. Zu den Adjektiven, die "kaum" bis "überhaupt nicht" zutreffen sollten, gehörten: mißtrauisch, mit religiöser Bindung, seelisch kompliziert, äußerlich gut aussehend, beeinflussbar, pessimistisch, verschlossen, aggressiv, mit eigener psychischer Krankheit, willensschwach und ungepflegt.

Bei der zweiten Befragung hatten sich die Einstufungen der Adjektive deutlich abgeschwächt, d.h., sie wurden mehr in der Mitte der Skala eingeordnet. In Tab. 1 sind die Adjektive aufgeführt, die sich in statistisch signifikanter Weise verändert hatten. Bei einem Vergleich der Teil-Stichproben mit bzw. ohne Codewort ergaben sich weitgehende Überschneidungen. Allerdings waren bei den Probanden mit Codewort im Wilcoxon-Test nur zwei Adjektive (tatkräftig und empfindsam) auf dem 5 %-Niveau signifikant verändert. Sowohl beim Vergleich der Antworten als auch der Teilnehmer

Wünschenswerte Psychiatereigen- schaften	Median 1./2. Befragung	Irrtums- wahrchein- lichkeit
geduldig	1.32 / 1.57	p < 0.05
tatkräftig	2.11 / 2.47	p < 0.01
willensschwach	4.62 / 4.26	p < 0.01
verschlossen	4.33 / 4.06	p < 0.01
beeinflussbar	3.99 / 3.81	p < 0.05
aggressiv	4.51 / 4.09	p < 0.01
vertrauens- erweckend	1.36 / 1.75	p < 0.001
ungepflegt	4.62 / 4.17	p < 0.001
mißtrauisch	3.64 / 3.35	p < 0.05

Tabelle 1:

Wünschenswerte Psychiatereigenschaften, denen bei der zweiten Befragung weniger Bedeutung beigemessen wurde, d.h., sie wurden signifikant häufiger in der Skalenmitte (bei 3,0; s. Text) eingestuft.

mit bzw. ohne Krankenpflegepraktikum oder Famulatur zeigten sich keine bedeutsamen Unterschiede.

3.2. Auf psychisch Kranke zutreffende Eigenschaften

Bei der Einstufung der Adjektive, die auf psychisch Kranke zutreffen sollten, wurde schon bei der ersten Befragung weitaus weniger bei "völlig" und "überhaupt nicht" angekreuzt als bei dem Psychiater-Teil der Umfrage. So meinten die Studenten bei der ersten Befragung, daß nur die Adjektive "empfindsam, gefühlsbetont, mißtrauisch, pessimistisch und beeinflussbar" in erheblichem Maße ("völlig" bis "nahezu") auf psychisch Kranke zuträfe. "Kaum" bis "überhaupt nicht" sollten nur für "phantasiarm, sachlich, tatkräftig, belastbar und selbstsicher" zutreffen.

Bei der zweiten Befragung hatte sich die Einstufung der Adjektive noch weit zur Mitte der Skala verschoben. Die statistisch signifikanten Veränderungen

Auf psychisch Kranke zutreffende Eigenschaften	Median 1./2. Befragung	Irrtumswahrscheinlichkeit
empfindsam	1.94 / 2.25	p < 0.001
mißtrauisch	2.14 / 2.53	p < 0.001
tatkräftig	3.84 / 3.56	p < 0.001
schweigsam	2.85 / 3.01	p < 0.05
intelligent	2.83 / 2.92	p < 0.05
selbstsicher	3.97 / 3.73	p < 0.01
sachlich	3.67 / 3.31	p < 0.01
beeinflußbar	2.23 / 2.58	p < 0,01
gefühlsbetont	2.13 / 2.39	p < 0.01
phantasiearm	3.65 / 3.32	p < 0.01
geduldig	3.23 / 3.05	p < 0.05

Tabelle 2:

Auf psychisch Kranke zutreffende Eigenschaften, denen bei der zweiten Befragung weniger Bedeutung beigegeben wurde, d.h., sie wurden signifikant häufiger in der Skalenmitte (bei 3,0; s. Text) eingestuft.

sind in Tab. 2 aufgeführt. 11 der 24 Adjektive wurden bei der zweiten Befragung von mehr als 50 % der Studenten in der Mitte der Skala eingestuft. Auch bei der Durchführung des Wilcoxon-Tests für die paarigen Stichproben ergaben sich 5 statistisch signifikante Veränderungen auf 5 %-Niveau (empfindsam, willensschwach, schweigsam, geduldig und tatkräftig). Beim Vergleich der männlichen mit den weiblichen Teilnehmern zeigten die weiblichen Teilnehmer häufiger Einstufungen der Adjektive in den Extrempositionen. Beim Vergleich der Teilnehmer mit bzw. ohne Krankenpflegepraktikum oder Famulatur ergaben sich keine bedeutsamen Ergebnisse.

3.3. Zusätzliche schriftliche Anmerkungen

Bei der ersten Befragung nahmen 22 von 186 Medizinstudenten (11,8 %) die Gelegenheit wahr, sich auf der Rückseite

des Fragebogens kritisch mit dem Inhalt der Fragen auseinanderzusetzen. Nach Abschluß der ersten Hälfte der Umfrage wurde jeweils ausdrücklich auf diese Möglichkeit hingewiesen (siehe oben). Bei der zweiten Befragung nahmen dann 47 von 159 Teilnehmern (29,6 %) zu der Umfrage Stellung. Dabei befaßte sich der weitaus größere Teil der Kommentare mit dem Patienten-Teil des Fragebogens. Die Studenten begründeten in diesen Kommentaren, warum sie psychisch Kranke nicht kategorisieren könnten. Tatsächlich hatten sich nämlich bei der zweiten Befragung 10 bis 15 % der Studenten hinsichtlich der Patientenfrage geweigert, bestimmte Adjektive in einem bestimmten Ausmaß den Patienten zuzuordnen. In den schriftlichen Anmerkungen wiesen viele von ihnen darauf hin, daß die Zuordnung der Adjektive auf psychisch Kranke Vorurteilen entspräche, die sie nicht mehr hätten.

4. Diskussion

Im Laufe des psychiatrischen Untersuchungskurses erfuhr die Einstellung der Medizinstudenten sowohl zu Psychiatern als auch zu psychisch Kranken eine deutliche Veränderung. Hierfür spricht nicht nur die Verschiebung der Einschätzungen im Fragebogen, sondern auch die Zunahme und die Art der schriftlichen Anmerkungen sowie die häufigeren Verweigerungen bei der zweiten Befragung. Nach dem Untersuchungskurs setzten sich mehr Studenten kritisch mit Form und Inhalt der Umfrageaktion auseinander. Hier wurde ein Lernprozeß erkennbar, der offensichtlich vom psychiatrischen Untersuchungskurs induziert wurde. Das gilt insbesondere für diejenigen Studenten, die sich bei der zweiten Befragung gänzlich weigerten, die Adjektive den psychisch Kranken zuzuordnen.

Bei den Fragen nach wünschenswerten

Eigenschaften des Psychiaters wurde deutlich, welche hohe Anforderungen die Studenten an einen Psychiater stellen. Bei der Zweitbefragung hingegen wurden diese Anforderungen deutlich niedriger gesteckt. Diese Veränderung ließ sich allerdings mit dem bei einer Teilstichprobe angewandten Wilcoxon-Test nur bei wenigen Adjektiven statistisch nachweisen. Insgesamt zeichnete sich jedoch auch hier die Tendenz weg von der überzeichneten Idealisierung des Psychiaters deutlich ab.

Bei der Einschätzung der Umfrageergebnisse könnte auch erwogen werden, ob die Studenten möglicherweise sozial erwünschte Antworten gegeben haben, die ohne konkrete Auswirkung auf den direkten Kontakt mit dem psychisch Kranken bleiben (Koch und Grippain 1978, Stumme 1975). Gegen diese Vermutung lassen sich jedoch die Anonymität der Umfrage und der hohe Prozentsatz kritischer Äußerungen zum Fragebogen anführen.

Ein Einwand gegen die Interpretation der Ergebnisse könnte darin bestehen, daß der "Fehler der Zentraltendenz" wirksam wurde (Sixtl 1979). Hiermit ist gemeint, daß von den Probanden extreme Merkmalsausprägungen vermieden werden oder daß unkritisch und wahllos angekreuzt wird. Dies könnte das gewünschte Ergebnis ("Trend zur Mitte") noch verstärkt haben. Daher sollte bei künftigen Untersuchungen geprüft werden, ob nicht besser andere Skalen zur Anwendung kommen können.

Obwohl einige methodische Probleme in dieser Arbeit noch nicht zufriedenstellend gelöst werden konnten, zeichnet sich bei unseren Ergebnissen deutlich ab, daß im Verlauf des psychiatrischen Untersuchungskurses Einstellungsänderungen der Medizinstudenten zum Psychiater und zum psychisch Kranken erfolgen. Offensichtlich trägt der psychiatrische Untersuchungskurs dazu bei, nega-

tive Einstellungen abzubauen. Während dieser Unterrichtsveranstaltung sind die Studenten durch die Begegnung in einer relativ kleinen Gruppe, durch die Anleitung im Umgang mit Patienten und in der Diskussion über das eigene Verhalten der Psychiatrie und dem seelisch Kranken näher gekommen. Wir wurden somit in der Zielrichtung bestätigt, den Psychiatrie-Unterricht mit einem praktischen Untersuchungskurs zu beginnen, in dem es als erstes nicht um Wissensvermittlung nach Art der Propädeutik geht, sondern um eine verbesserte Fremd- und Selbstwahrnehmung im Umgang mit dem psychisch Kranken. Die dadurch erreichten Einstellungsänderungen können über das Praktikum hinaus anhalten (Maxmen 1979, Miller et al. 1979).

Literatur

- Braune, P. u. U. Heitz: Die Einstellung von Heidelberger und Mannheimer Medizinstudenten zu Fragen der Ausbildung und Praxis im Fach Psychiatrie. Med. Dissertation, Heidelberg, 1972.
- Bruhn, J.G. u. O.A. Parson: Medical student attitudes toward four medical specialties, in: Journal of Medical Education 39 (1964) S. 40-49.
- Crow, C.M., Mowbray, R.M. u. S. Bloch: Attitudes of medical students to mental illness, in: Journal of Medical Education 45 (1970) S. 594-599.
- Dahme, B.: Psychologische Dimensionen der Gesprächsführung, in: Medizinische Psychologie 1 (1974) S. 34-49.
- Eagle, P.F. u. L.R. Marcos: The psychiatric clerkship: Its effect on medical students, in: American Journal of Psychiatry 39 (1979) S. 85-89.
- Eagle, P.F., Marcos, L.R. u. R. Cancro: Medical students' attitudinal changes associated with the psychiatric clerkship, in: Journal of Psychiatric Education 3 (1979) S. 180-188.
- Eagle, P.F. u. L.R. Marcos: Factors in medical students' choice of psychiatry, in: American Journal of Psychiatry 137 (1980) S. 423-427.
- Fietkau, H.J.: Einstellungen und Wertehaltungen von Psychotherapeuten. Philos. Dissertation, Berlin 1976.

Finzen, A.: Ergebnisse einer Umfrage bei Tübinger Medizinstudenten über ihre Einstellung zur Psychiatrie und zum psychiatrischen Unterricht, in: Der Nervenarzt 41 (1970), S. 299-300.

Koch, U.: Wie Ärzte sich selbst und ihre Patienten sehen, in: Psychologie heute, Heft 1 (1975) S. 13-17 u. 76.

Koch, U. u. P. Grippain: Das Arzt-Patienten-Verhältnis aus der Sicht des Arztes, in: Psycho 4 (1978) S. 404-413.

Lorenzen, D.: Psychiatrie heute: Schlimmer als ihr Ruf?, in: Psychologie heute, Heft 7 (1981) S. 42-49.

Maurice, W.L., Klonoff, H., Miles, J.E. u. R. Krell: Medical student change during a psychiatric clerkship: Evaluation of a program, in: Journ. of Med. Education 50 (1975) S. 181-189.

Maxmen, J.S.: Student attitude changes during "psychiatric medicine" clerkships, in: General Hospital Psychiatry 1 (1979) S. 98-103.

McCormick, W.O. u. G. Voineskos: Medical students' views of psychiatry. Does teaching have any effect? In: Canadian Psychiatric Association Journal 23 (1978) S. 541-547.

Meltzer, M.L. u. H.M. Grigorian: Effect of psychiatric education on attitudes of medical students to mental illness, in: Psychiatry 35 (1972) S. 195-204.

Miller, S.I., Lenkoski, L.D. u. D. Weinstein: Enduring attitude change in medical students, in: Journal of Psychiatric Education 3 (1979) S. 171-179.

Moos, R.H. u. I.D. Yalom: Medical students' attitudes toward psychiatry and psychiatrists, in: Mental Hygiene 50 (1966) S. 246-256.

Nielsen, A.C.: Choosing psychiatry: The importance of psychiatric education in medical school, in: American Journal of Psychiatry 137 (1980) S. 428-431.

Parker, S.: Personality factors among medical students as related to their predisposition to view the patient as a "whole man", in: Journ. of Med. Education 33 (1958) S. 736-744.

Schepank, H.: Modell eines Praktikums der Psychosomatischen Medizin und Psychotherapie, in: Zeitschrift für Psychosomatische Medizin u. Psychoanalyse 22 (1976) S. 224-239.

Sixtl, F.: Meßmethoden der Psychologie. 2. Aufl., Weinheim, Beltz, 1979.

Stumme, W.: Was heißt Geisteskrankheit?, in: Der Nervenarzt 41 (1970) S. 294-298.

Stumme, W.: Psychische Erkrankungen im Urteil der Bevölkerung, in: Fortschritte der Sozialpsychologie 1. München-Berlin-Wien, Urban & Schwarzenberg, 1975.

Walton, H.J.: Personality correlates of a career interest in psychiatry, in: British Journal of Psychiatry 115 (1969) S. 211-219.

Dr. M. van der Ven, Priv.Doiz. Dr. G. Buchkremer, Psychiatrische und Nervenambulanz der Universität Münster, Albert-Schweitzer-Str. 11, D 4400 Münster

DIE TRIAGE SOLLTE AUS DEM ENTWURF DER ÄRZTLICHEN APPROBATIONSORDNUNG HERAUSGENOMMEN WERDEN

K. Schimmelpfennig, Berlin

Der Entwurf der Fünften Verordnung zur Änderung der Approbationsordnung für Ärzte wird in diesen Monaten im Bundesministerium für Jugend, Familie und Gesundheit vorbereitet. Zu den neuen Gebieten des Prüfungstoffes der 2. Ärztlichen Prüfung, die nach dem 2. klinischen Abschnitt abgenommen wird, gehört die Triage. Damit wird an den Universitäten ein neues Lehr- und Lerngebiet eingeführt, das wenigen Studenten gut bekannt ist, Hochschullehrer kennen diesen Begriff kaum. Nur in den neu aufgelegten Nachschlagewerken findet man eine treffende Definition.

Untrennbar ist die Triage von der Katastrophenmedizin; Erklärungen der Triage finden sich daher auch unter diesem Stichwort. Der Begriff Katastrophenmedizin wird in der vorgesehenen Approbationsordnung aber nicht erwähnt. Das Prüfungsgebiet der Triage liegt verborgen in dem neuen Prüfungstoff: "Grundlagen der ärztlichen Versorgung bei einem Massenanfall von Kranken oder Verletzten" (1). Zur Definition der Triage: "Unter Triage faßt man alle ärztlichen Maßnahmen zusammen, die geeignet sind, die Überlebensdringlichkeit, die Transportdringlichkeit und die Überlebenswahrscheinlichkeit eines Katastrophenopfers zu beurteilen" (2). "Unter Katastrophenbedingungen kommt der Triage, d.h. der Patientensortierung nach Art und Schwere der Verletzung, aber auch unter Berücksichtigung der Prognose, unter gegebenen logistischen und medizinisch infrastrukturellen Umständen, die größte Bedeutung zu.... In der Katastrophensituation wird die Effektivität eines Sanitätsdienstes ... an der Zahl der Überlebenden bewertet... Die Triage ist die wichtigste und schwierigste Aufgabe des Arztes unter Katastrophenbedingungen.

Der Begriff Hoffnungslose ist relativ und hängt nicht allein von der Verletzung, sondern auch von den gegebenen äußeren Umständen ab..."(3).

Aufgabe der Lehre der Triage im Medizinstudium ist somit:

- der Student muß definieren können, was eine Triage ist.
- es müssen Kriterien vermittelt werden, die bei der Triage zu berücksichtigen sind.
- Was bedeutet Triage ?
- Auswertung eines kasuistischen Beitrages (Hamburger Überschwemmung, Attentat auf der Oktoberwiese, Bologna.)

Der Verfasser dieser Zeilen will im Folgenden nicht die Katastrophenmedizin diskutieren! Nur um Mißverständnissen vorzubeugen sei erwähnt, daß der Autor für die Triage Verständnis hat und die Katastrophenmedizin nicht in Frage stellt.

Das hier angesprochene Problem bezieht sich allein auf die Lehre und Prüfung der Triage im Medizinstudium.

Was bedeutet die Triage ?

Triage bedeutet, daß bei ärztlichen Entscheidungen individualmedizinische ärztliche Aufgaben zugunsten anderer Behandlungsprioritäten aufgegeben werden. Ein neuer Grundsatz bestimmt die ärztliche Handlungsweise: möglichst viele Menschen zu retten. Dies gelingt mit einer konzeptierten Sortierung von Verletzten hinsichtlich der Überlebenswahrscheinlichkeit und Behandlungsdringlichkeit. Zur Frage der Sichtung von Verletzten bei einer Katastrophe, d.h. dem Auswahlprinzip, schreibt der Moraltheologe Böckle in einem Sammelband der Deutschen Gesellschaft für Katastrophenmedizin unter dem Titel: Ethik ärztlichen Handelns in der

Katastrophe : "Der Massenansturm von Verletzten bei einer begrenzten Zahl von Ärzten macht es leider nicht möglich, jedem Betroffenen nach individualmedizinischen Grundsätzen zugleich zu helfen. Eine Sichtung und Entscheidung nach Dringlichkeitsstufen der Behandlung wird hier zur schmerzlichen Pflicht. Sie folgt dem für die austeilende Gerechtigkeit entsprechenden Grundsatz: Größtmöglicher Nutzen für möglichst viele "(4).
Einleitend betont Böckle: "Bereits der ärztliche Alltag erfordert vom Arzt immer wieder Entschlüsse, die nicht nur solides medizinisches Wissen voraussetzen, sondern auch einen aus der Erfahrung wachsenden Mut zum Gewissensentscheid verlangen."

Hilfe zur voraussetzungslosen Planung von Entscheidungen ?

Hier muß gefragt werden, ob das von Böckle geforderte "solide medizinische Wissen" bei den Medizinstudenten als verantwortliche Grundlage für die Entscheidung der Triage vorausgesetzt werden kann. Die Grundlage wäre hier das Wissen um die Notfallmedizin - auch um die Notfallmedizin mit nicht-technischen Hilfsmitteln. Dazu gehören vereinfachte und schematisierte Behandlungen von Verletzungen, wie sie bei einer Katastrophe zu erwarten sind: Polytraumen, Vergiftungen, Verbrennungen und offene Verletzungen. Gelernt werden müssen die schrittweisen diagnostischen und therapeutischen Sofortmaßnahmen nach Dringlichkeitsstufen bei einem verletzten Menschen.-
Erst wenn dieses solide medizinische Wissen als Grundvoraussetzung nachgewiesen - also nachgeprüft - ist, dürfen Entscheidungen für eine größere Anzahl von Verletzten geltend gemacht werden, weil dann erst ein Bezug zu den äußeren Umständen (3) gefunden werden kann. Zu den äußeren Umständen gehört die

Funktionsmöglichkeit einer Krankenversorgung an einem ferneren Ort. Ohne die Berücksichtigung dieser Möglichkeiten - unter individualmedizinischen Aspekten - sind Entscheidungen am Katastrophenort theoretische Gedankenkonstruktionen ohne praktischen Bezug, also von einer verantwortungsethischen Haltung nicht getragen.

Vor Gewissensentscheidungen müssen verantwortungsethische Fragen geklärt sein. Bevor an einer Universität den Studenten im 3. bis 6. klinischen Semester Handlungsmaximen vermittelt werden, bei denen ethische Fragen bedacht werden müssen, brauchen die Dozenten Voraussetzungen, um sich mit formalen Theorien und Werturteilen gründlich auseinanderzusetzen, um das Gleiche von den Studenten fordern zu können. Es darf nicht akzeptiert werden, daß ohne diese Voraussetzung Stoffgebiete gelehrt und staatlich geprüft werden. Fragen des ärztlichen Gewissens im ärztlichen Alltag gehören bisher nicht zum Lehrstoff der medizinischen Ausbildung, daher kann darauf nicht zurückgegriffen werden. Es stellt sich sogar die Frage, ob bei der Triage das Spannungsfeld zwischen Pflicht und Gewissen dem neuen Problem angemessen ist.-

Zur Frage der Vorbereitung eines verantwortlichen ärztlichen Handelns im Medizinstudium.

"Die wichtigste Eigenschaft eines gebildeten Menschen ist seine Fähigkeit verantwortungsvolle Entscheidungen zu treffen und den Dialog über die Grenzen der Fächer zu ermöglichen" (Aus der Präambel von Harvard). Folgen wir dem Weg über die Grenzen rein medizinischer Fragen in den Aufgabenbereich verantwortungsvoller ärztlicher Entscheidungen, so kann das traditionelle Ethos der sittlichen Grundbestimmung der Pflicht heute nicht frag-

los hingenommen werden. Ein neuer Elementarwert ist heute die Verantwortung (5), dieser Begriff begegnet uns z.B. in der Verantwortungsethik. Nach Schwarztländer wird "der Begriff der Pflicht heute vielfach abgelehnt, sofern mit ihm die Vorstellung von obrigkeitlichem Befehl, allseitiger Reglementierung, Unterdrückung des eigenen Willens und der spontanen Verhaltensweisen verbunden ist, kurz, die Vorstellung von einem äußeren und inneren Zwang, der der Freiheit des eigenen Willens und der Spontanität des Miteinanderlebens entgegengesetzt ist"(5).- Wir müssen versuchen, die Frage zu klären, ob ein ärztlich verantwortliches Handeln begründbar ist, das vom Nahhorizont geleitet wird: ärztliche Fürsorge und Mitleid für die Nächsten bei einem Massenansturm von Verletzten auf Kosten vieler anderer. Das Durchdenken von Handlungsnormen in einer Grenzsituation wie der Katastrophenmedizin, ist heute im Lichte der Verantwortung schwerer, da auf die prinzipielle Bedeutung der Gegenseitigkeit im Nahhorizont und auf das personale Miteinander, also auf subjektive Bestimmungen zum Nächsten, nicht verzichtet werden kann zu Gunsten von Allgemeinen Bezügen, wie sie bei der Sortierung von Menschen in Behandlungskategorien zu tätigen sind. Nur auf dieser Ebene der subjektiven Bezüge sowie des personalen Bezuges lassen sich wohl Analogien finden mit ärztlichen Gewissensentscheidungen der Individualmedizin: Die ärztliche Aufgabe verstanden als Gegenseitigkeit im Sinne von Dasein für..., Handeln für.... Eine Gegenseitigkeit wird aufgehoben bei Allgemeinen Bezügen wie Anzahl oder Rangordnung. Für die ethische Besprechung der Triage im Medizinstudium brauchen wir hauptamtliche Professoren an den Universitäten, die für diese Fragen vorbereitet sind, um verantwortungsethisches Handeln - wie es

die Triage erfordert - zu begründen. Wenn dafür durchdachte Voraussetzungen noch nicht gegeben sind, so braucht die Frage der personellen Vertretung an einer Universität nicht gestellt zu werden.- Eine Gefahr, diese Fragen in absehbarer Zeit mit Studenten zu besprechen, ist nicht gegeben, wenn schlimme Kritik an der Triage geübt wird, sondern die Gefahr ist, daß diese Kritik in den Lehrveranstaltungen unterdrückt wird.- Eine systematische und eventuell qualifizierte Kritik ist aber von jenen Studenten zu erwarten, die sich derzeit besonders um die vernachlässigten Probleme unserer Gesellschaft bemühen, eine Gesellschaft, die eine "technische Logistik" unserer Zivilisation zu wenig berücksichtigt.

Wird mit der Triage ein politisch-ideologisches Element in den Universitäten eingeführt ?

Nur mit großem Bemühen gelingt es - wie hier - die Triage nicht in einem militärärztlichen Zusammenhang zu besprechen. Wenn man glaubwürdig bleiben will, kann man diesen Aspekt in Lehrveranstaltungen nicht umgehen. Es gehört aber nicht zur Aufgabe der Universitäten, zu parteipolitischen Fragen Stellung zu beziehen. Es gehört jedoch zum Selbstverständnis jener Gruppen, die der Friedensbewegung zugeordnet werden, diese politischen Fragen wiederholt zu thematisieren.

Katastrophenmedizin nur im Medizinstudium ?

Wenn die Triage als Prüfungsstoff zwingend zur Notfallmedizin bei einem Massenansturm von Kranken und Verletzten gehört, dann sollte aus Gründen der Offenheit bei Anhörungen zur vorgesehenen Novellierung der ÄAppO auf den Begriff der Katastrophenmedizin nicht verzichtet werden.

Der Gesetzgeber muß außerdem auch überprüfen, ob er eine hohe ärztliche Verantwortlichkeit bei der Triage einkalkulieren darf, wenn es auf diesem Sektor keine Pflichtweiterbildung gibt.

Wenn die Triage aus dem Entwurf der ÄAppO herausgenommen wird, so wird die Freiheit der Lehre zu dieser Frage selbstverständlich nicht beeinträchtigt.

Literatur:

- 1) Bundesminister für Jugend, Familie und Gesundheit. Gz: 315-4331-2/7.- 9.11.83. Entwurf einer Fünften Verordnung zur Änderung der Approbationsordnung für Ärzte. Anlage 11 : Prüfungsstoff für den zweiten Abschnitt der ärztlichen Prüfung (Anlage 16 zu § 29 Abs.2 Satz 2).
- 2) Kirchhoff, R. (Hrsg) (1984) Triage im Katastrophenfall. Band 9, Notfallmedizin. perimed. Fachbuch-Verlagsgesellschaft mbh Erlangen.
- 3) Lanz, R. und Rossetti, M. (1984) Katastrophenmedizin. Ferdinand Enke Verlag Stg.
- 4) Peter, K., Heberer, G., Rebentisch, E., Linde H. J., Kirchhof, R. (Hrsg) (1984) Katastrophenmedizin. (1. Tagung der Deutschen Gesellschaft für Katastrophenmedizin e.V.) J. F. Bergmann Verlag München.
- 5) Schwarztländer, J., (1973): Verantwortung. In: Krings, H. - Baumgartner, H. - M., Wild, C., (Hrsg) Handbuch philosophischer Grundbegriffe. S. 1577. Band 6. Kösel-Verlag München
- 6) Protokoll des Außerordentlichen Medizinischen Fakultätentages am 13.1.84. - S. 9.

Prof. Dr. med. K. Schimmelpfennig
Institut für Toxikologie der F.U., Garystr. 9
D 1000 Berlin-33

DIE AMERIKANISCHEN PRÜFUNGEN FÜR AUSLÄNDISCHE ÄRZTE

H. E. Renschler, Bonn und D. Habeck, Münster

Zusammenfassung:

Die in gleicher Form weltweit durchgeführten Prüfungen der amerikanischen Kommission für ausländische Ärzte (Educational Commission for Foreign Medical Graduates, ECFMG) bieten sich als Vergleichsmaßstab für die Qualität der Ausbildung in verschiedenen Ländern an. Als Instrument für die Regelung des Zuganges der im Ausland ausgebildeten Ärzte zur Berufsausübung in den USA sind die Prüfungen allerdings dem Einfluß des Wechsels zwischen Mangel und Überschuß an Ärzten in den USA ausgesetzt.

Die kurze und relativ einfache ursprüngliche Prüfung ("ECFMG") ermöglichte in den ersten Jahrzehnten nach dem Krieg eine Einwanderung oder einen Studienaufenthalt ausländischer Staatsangehöriger als Ärzte unter kontrollierten Bedingungen. Sie wurde mit der Ankündigung eines Ärzteüberschusses in den USA 1977 zunächst für Ausländer durch eine strengere Prüfungsform (Visa Qualifying Examination, VQE) ersetzt. Ab 1985 müssen außer den Ausländern auch US-Bürger, die in immer größerer Zahl im Ausland eine ärztliche Qualifikation erwerben, die neu eingeführte "Prüfung in den medizinischen Wissenschaften" ablegen. Diese soll sich nach Umfang und Schwierigkeit den Prüfungen angleichen, die die Absolventen der amerikanischen Medizinschulen ablegen. Es werden die verschiedenen ärztlichen Prüfungen mit Teilnehmerzahlen und globalen Erfolgsquoten vor dem Hintergrund der nationalen Änderungen des Ärztebedarfes der USA dargestellt.

Einleitung

Die von der "Kommission für ausländische Ärzte" der Vereinigten Staaten (Educational Commission for Foreign Medical Graduates, ECFMG) durchgeführten Prüfungen finden weltweit Beachtung. Die Erfolgsquoten, besonders der bekanntesten, der "ECFMG-Prüfung", werden als Meßinstrument

nicht nur für das Können und Wissen des einzelnen Arztes, sondern auch kollektiv für die Qualität der Ausbildung oder der Auswahl der Medizinstudenten in den verschiedenen Teilen der Welt diskutiert (1-4). Von Seiten der Amerikanischen Medizinischen Fakultäten war diese Prüfung aber wegen ihrer geringen Anforderungen starken Angriffen ausgesetzt und mußte 1984 ganz eingestellt werden (5). Da die Prüfungen Auswirkungen des dynamischen Wandels des Bedarfes und der Ausbildung von Ärzten in den Vereinigten Staaten von Nordamerika aufweisen, ist es auch für eine Betrachtung, die sich nur auf den begrenzten Zeitraum von 1977-1983 erstrecken soll, notwendig, den gesamten Hintergrund des Prüfungssystemes zu berücksichtigen. Die folgende Übersicht wurde besonders aus den Darstellungen des Systemes der Prüfungen (6) und der Ausbildung für Ärzte in den USA (7) sowie aus den Veröffentlichungen der ECFMG extrahiert. Im Rahmen des seit 1945 weltweit zunehmenden Einflusses von nationalen Organisationen und von Staatsregierungen müssen Einzelschicksale infolge der immer enger werdenden persönlichen Freiheiten für unabhängige Entscheidungen außer Betracht bleiben.

Ärztebedarf der Nachkriegszeit

Nach dem 2. Weltkrieg unterstützte das FULBRIGHT-Programm den internationalen Austausch. Es setzte noch die jahrhundertealte Tradition fort, nach der Studenten und junge Wissenschaftler sich in freier Entscheidung an die besten Ausbildungsstätten begeben. 1948 erhielten junge Ärzte ein besonderes Visum (J) für einen Aufenthalt in den USA. Im Jahre 1950 wanderten 1.878 Ärzte als Immigranten in die USA ein, 308 erwarben die in den USA von den Staatsregierungen nach entsprechender Prüfung vergebene Lizenz zur ärztlichen Berufsausübung.

Schwierigkeiten tauchten auf, da in den USA

das Bestehen der Abschlußprüfung nach der (universitären) medizinischen Grundausbildung die Voraussetzung für die Zulassung zur Staatsprüfung ist. Der amerikanische Ärzteverband (American Medical Association, AMA) erstellte daher 1950 eine Liste der Medizinschulen im Ausland, deren Absolventen zur Staatsprüfung vorgeschlagen und zugelassen werden konnten. Der Erfolg dieser Liste wurde durch die Prüfungsergebnisse bestätigt:

während 68 % der Kandidaten, die Absolventen einer empfohlenen Schule waren, die Prüfung bestanden, waren es nur 46 % der übrigen. Die ausländischen Ärzte wurden als "Bedrohung für das Wohl der Patienten" angesehen. Daher wurde 1957 eine besondere Organisation, der Educational Council for the Foreign Medical Graduate, ECFMG, gegründet. Zweck war, den Krankenhäusern bei der Auswahl qualifizierter Ärzte zu helfen. 1958 wurden die ersten im Auftrag des ECFMG von der nationalen Prüfungsorganisation (National Board of Medical Examiners, NBME) ausgeführten Prüfungen abgehalten, 51 Prozent der 5.000 Teilnehmer bestanden. Die Krankenhäuser fanden in den im Ausland ausgebildeten Ärzten billige Arbeitskräfte, um dem nach dem Krieg fühlbaren Ärztemangel in den USA abzuhelpen. Der Anteil von Ausländern an den in den Krankenhäusern der USA tätigen Ärzten überschritt 1957 20 % und erreichte im Jahre 1969 mit 35 % seinen Gipfel. Bis zum 31.12.1960 mußten alle ausländischen Ärzte, die eine Krankenhaustätigkeit in den USA ausübten, die Prüfung des ECFMG bestanden haben. Die Zahl der Kandidaten betrug im Jahre 1960 nahezu 15.000. Die Gesamtzahl der jährlich in die USA einwandernden Ärzte stieg bis 1972 auf 7.144 an und reduzierte sich danach bis 1979 um mehr als die Hälfte. Die Anzahl der Ärzte, die vorübergehend für einen Forschungsaufenthalt oder zur Weiterbildung in die USA kommen, zeigte nach der Höchstzahl von 5.700 im Jahre 1968 einen noch stärkeren Rückgang auf 420 im Jahre 1979 und 370 im Jahre 1983.

Von 1950 bis 1979 wanderten 92.626 ausländische Ärzte in die USA ein. In derselben Zeit wurden 258.374 in den USA ausgebildeten Ärzten die Lizenz zur Berufsausübung erteilt. Nimmt man an, daß beide Gruppen die Gesamtzahl der zur Zeit

in den USA tätigen Ärzten darstellen, so haben 26 % davon ihre Grundausbildung im Ausland ohne Belastung der amerikanischen Medizinschulen und ihrer Geldgeber erhalten.

Sozialisierung der Medizin in den USA und ihre Auswirkung auf den Ärztebedarf

Die gesamten Bedingungen der Medizin änderten sich grundlegend in den USA in den sechziger Jahren, in denen die USA die Sozialisierung der gesundheitlichen Betreuung der ganzen Bevölkerung einführten. Mit den 1965 erlassenen Gesetzen erhielten alle alten und bedürftigen Amerikaner eine Berechtigung zur ärztlichen Versorgung und für einen erforderlichen Krankenhausaufenthalt. Diese -im Gegensatz zu Deutschland- in den USA bundesweite Einheitsversicherung (MEDICAID und MEDICARE) umfaßte 1980 183 Millionen Einwohner, ihr Umsatz machte mit 914 Milliarden US Dollar 43 % aller Gesundheitskosten aus.

Maßnahmen zur Behebung des Ärztemangels in den USA

Die parallel mit der Sozialisierung diskutierten Maßnahmen zur Erweiterung der Ausbildung der Ärzte führten 1968 zur gesetzlichen Unterstützung der medizinischen Ausbildungsstätten auf Bundesebene. Die Anzahl der ärztlichen Approbationen sollte von 1970-1978 um 60 % gesteigert werden. Insgesamt nahmen zwischen 1960 und 1980 40 neu gegründete Medizinschulen die Ausbildung von Ärzten auf. 1976 konnten, nachdem das Erreichen des ursprünglichen Zieles in Aussicht stand, die Verringerung der Bundeszuschüsse und das Ende des Ärztemangels in den USA angekündigt werden. Die im Gesetz vorgesehene Unterstützung von Medizinstudenten, die als US-Staatsbürger aus dem Ausland zurückkehren wollten und dazu seit 1972 privat Unterstützung erhalten hatten ("Fifth Pathway", COTRANS), war aber schon nicht mehr durchsetzbar und mußte 1978 widerrufen werden.

Entwicklung eines Ärzteüberschusses in den USA

Ende der sechziger Jahre setzte sich bei Wirtschaftswissenschaftlern und Gesundheitspolitikern die Erkenntnis durch, daß die Anzahl der Ärzte die nicht mehr kontrollierbaren Kosten des Gesundheitssystems der USA übermäßig belasten würde. Es wurde errechnet, daß durch einen Arzt jährlich Kosten in der Höhe von 350.000 US Dollar entstehen. Das Defizit der Gesundheitskosten in den USA stellt bis heute das größte medizinische Problem der USA dar. 1980 wurde daher von der Weiterbildungskommission empfohlen, die Neuzulassungen zum Medizinstudium um 17 % zu reduzieren. Die Zahl der Studienanfänger hat sich seit dem Höchststand im Jahre 1980/81 von 17.320 auf 16.395 im Studienjahr 1984/85 reduziert. Im Jahre 1984/85 waren einschließlich der Wiederholer 16.997 Studenten im 1. Studienjahr an den 126 Medizinschulen der USA eingeschrieben. Auf die starke Expansion der Mediziner Ausbildung in den USA zwischen 1960 und 1980 folgt bereits jetzt eine wirkungsvolle Reduktion der Studienplätze um 5 %.

Eine nur in Einzelheiten angezweifelte Vorausberechnung des Bureau of Health Manpower ergibt eine Reduktion des Verhältnisses von Bevölkerung/ Arzt von dem Gipfel von 754 im Jahre 1960 auf 376 im Jahre 2000 – oder aber einen Überschuß von 145 000 - 200 000 Ärzten allein in den USA. Berechnet man für 1982 die Zahl der Einwohner pro lizenzierten Arzt, so ergibt sich für die gesamten Vereinigten Staaten bei einer Zahl von 440 000 Ärzten ein Mittelwert von 514 Einwohnern pro Arzt. Fünf Staaten haben schon jetzt weniger als 400 und nur noch zehn Staaten hatten 1982 mehr als 750 Einwohner pro Arzt.

Infolge der Attraktivität des Arztberufes und den hohen Anforderungen an die Zulassung zum Medizinstudium in den USA nahmen zahlreiche Amerikaner ein Medizinstudium im Ausland auf. Vor allem in Mittelamerika entstand nach 1970 eine eigene "Industrie" auf dieser Basis. Die Qualität dieser Ausbildung erhielt jedoch 1981 nach einer ausführlichen Begutachtung durch eine Kommission des Verbandes der amerikanischen Medizinschulen ein vernichtendes Urteil (5). Da die Zahl der Rückkehrer inzwischen auf etwa ein

Viertel der Zahl der Absolventen der amerikanischen Medizinschulen angewachsen ist, sind die amerikanischen Staatsbürger mit einer im Ausland erfolgreich abgeschlossenen medizinischen Ausbildung ein Politikum geworden, mit dem sich bedeutende Politiker, wie Edward Kennedy, Andrew Young oder Califano befaßt haben. Unter Vernachlässigung der traditionellen wissenschaftlichen Interessen eines internationalen Erfahrungsaustausches und der Moral dominiert diese innerpolitische Zielsetzung (8).

Die weitere Entwicklung der ärztlichen Prüfungen wird gemeinsam von Gesundheitspolitikern, Ärzteverbänden, den Medizinschulen und den Prüfungsorganisationen vertreten. Die gegenwärtig eingeführten Veränderungen der Prüfungen haben das gesundheitspolitisch definierte Ziel, bei höchster Qualität der Ärzte deren Gesamtzahl in den USA möglichst gering zu halten. Trotzdem beteiligten sich bis 1984 überraschend viele ausländische Staatsbürger an den ECFMG-Prüfungen.

In der Zwischenzeit liegen die ersten Ergebnisse der neueingeführten Prüfung in den medizinischen Wissenschaften "Foreign Medical Graduate Examination in the Medical Sciences" (FMGEMS) vor, die von allen im Ausland qualifizierten Ärzten – unabhängig von der Staatsangehörigkeit – abgelegt werden muß, wenn eine ärztliche Tätigkeit zur Weiterbildung in den USA aufgenommen werden soll (s.u.Tab. 2). Unter den 7.423 Teilnehmern der ersten Durchführung der zweitägigen Prüfung konnten 20 % der Nicht-US-Bürger, aber nur 3,8 % der im Ausland ausgebildeten US-Bürger die Prüfung bestehen (9).

Die Prüfungen der amerikanischen Kommission für ausländische Ärzte (ECFMG)

Obersicht

Die 1958 erstmals durchgeführten Prüfungen des ECFMG haben sowohl in den USA als auch international eine große Bedeutung erlangt und wurden dementsprechend sorgfältig methodisch weiterentwickelt. Die Prüfungsergebnisse werden seit 1972 als eigene Drucksache des ECFMG veröffentlicht

und weltweit verbreitet. Für die USA ermöglichten die Prüfungen in den ersten Nachkriegsjahren eine Selektion der besten Ärzte der Welt, die danach zur gesundheitlichen Betreuung der eigenen Bevölkerung eingesetzt werden konnten. Für die Ausländer bot sich damit die Möglichkeit der Einwanderung in die USA mit einer Verbesserung der Arbeitsbedingungen unter Vernachlässigung der Auswirkung auf das Mutterland. Der Begriff des "brain drain" charakterisiert diese Situation. Für Einzelpersonen, Universitäten und ganze Nationen ergab sich die Möglichkeit einer "Eichung" des eigenen Wissens und der ärztlichen Ausbildung an dem anerkannten Standard der USA. So gab es außerhalb der USA neue Medizinschulen, die von ihren Absolventen verlangten, die Prüfung des ECFMG zu bestehen, um damit den eigenen Ausbildungsstandard hochzuhalten. Nach den 1977 eingeführten Änderungen der Prüfungen (Aufteilung in ECFMG und VQE) blieben die Prüfungsbedingungen bis zum Übergang in die später darzustellende einheitliche Prüfungsform für alle im Ausland ausgebildeten Ärzte konstant. Die 1984 abgeschaffte Visumsprüfung (VQE) wird später beschrieben, es soll zunächst die ECFMG-Prüfung kurz dargestellt werden.

ECFMG-Prüfung

Die ausschließlich schriftliche Prüfung bestand aus etwa 360 (oder zeitweilig 420) Fragen mit Auswahlantworten, die an einem Tag in etwa 6 1/2 Stunden beantwortet werden mußten. Amerikanische Medizinstudenten müssen die Fragen in etwa 70 % der den Ausländern zugebilligten Zeit beantworten. Die Prüfungsfragen stammen von der nationalen ärztlichen Prüfungsstelle (NBME) und wurden an einer vorbestimmten Zahl von mindestens 5.000 nordamerikanischen Studenten standardisiert. Die Prüfungsfragen bezogen sich auf das gesamte Gebiet der theoretischen und klinischen Medizin und wurden laufend erneuert.

Die Bestehensgrenze lag bei einem Punktwert ("Scaled Score") von 75, was einem Standardwert von 290 entsprach. Dies war die Bestehensgrenze des zweiten Teils des amerikanischen Staatsexamens, die nach der standardisierten Verteilung von 1,8 % der Kandidaten nicht erreicht wird. Bei den Teilnehmern an den ECFMG-Prüfungen liegen die Durchfallquoten mit einer sehr großen Streu-

breite im Mittel zwischen 60 und 70 %. Durch die Wiederholungsprüfungen reduzierte sich allerdings bisher die gesamte Durchfallsquote auf etwas mehr als 30 %.

Der in die Prüfung einbezogene einstündige Test der englischen Sprache erhielt 1974 eine neue Form. Das Bestehen des Sprachtestes, das für die Erteilung des ECFMG-Zertifikates erforderlich ist, wurde ab 1977 bei der Veröffentlichung aller nach Ländern und Universitäten aufgegliederten Prüfungsergebnisse in die Bestehensquote einbezogen. Die veröffentlichten Bestehensquoten liegen daher ab 1977 niedriger (1977=20,2 %), als dies den früheren Mitteilungen entsprochen hätte: für 1977 wären dies 33,3 % gewesen. Aus den Veröffentlichungen der Ergebnisse der ECFMG ist jedoch der Anteil der 1977 erstmals durchgeführten Visa Qualifying Examination für Ausländer nicht zu erkennen. So gibt die ECFMG für 1983 eine Gesamtzahl von 27.162 abgenommenen ECFMG-Prüfungen an, das NBME eine identische Zahl von 27.165 und zusätzlich dazu 7.449 VQE, deren Ergebnisse aber nicht im Bericht des ECFMG enthalten sind. Die große Anzahl von 27.000 Teilnehmern an der ECFMG-Prüfung auch im Jahre 1983 ist überraschend, da diese Prüfung nur für die daran teilnehmenden 7.311 US-Bürger praktische Bedeutung hat. Die 20.000 Nicht-US-Bürger erzielen mit dieser Prüfung nur eine Bewertung ihres Wissens nach international festgelegten und am Wissen der US-Studenten standardisierten Kriterien. Das Bestehen der ECFMG-Prüfung oder der FLEX-Prüfung wird für die Erteilung eines Einreisevisums in die USA nicht anerkannt !

Entwicklung neuer Prüfungen: Federation Licensing Examination (FLEX)

Die erst nach erfolgreichem Abschluß des Studiums an einer Medizinschule anerkannte Staatsprüfung ("National Board"), deren ersten beiden Teile allerdings schon während des Medizinstudiums abgelegt werden, verfolgt neben einer Messung des Studienerfolges mit dem Schwerpunkt in den Grundlagenfächern den Zweck des Nachweises der ärztlichen Kompetenz als Grundlage der Berufsausübung. Die Prüfungsämter der einzelnen Bundesstaaten, die die zeitlich begrenzte Li-

zenz zur ärztlichen Berufsausübung erteilen, führten daher 1968 eine zweite Prüfungsform "Federation Licensing Examination" (FLEX) ein, die mehr als das bisherige US-Staatsexamen die ärztliche Kompetenz bewerten soll. Sie wird ebenfalls vom National Board (NBME) im Auftrag der Vereinigung der Landesprüfungsämter (Federation of State Medical Boards, FSMB) ausgeführt und benutzt überwiegend praxisrelevante Fragen, die in den üblichen Staatsexamen vorge-testet worden sind. Die Fragen werden vor allem unter dem Gesichtspunkt der Bewertung der ärztlichen Kompetenz ausgewählt und müssen testtheoretischen Forderungen entsprechen. Die Bestehensbedingungen werden von den einzelnen Prüfungsämtern unabhängig voneinander unter Anpassung an die örtlichen Bedingungen festgelegt. Wegen der geringeren theoretischen Anforderungen werden sie von vielen Kandidaten bevorzugt. 1983 beteiligten sich jeweils 13 bis 15.000 Kandidaten an den drei Teilen des Staatsexamens ("National Board"), aber 16.600 an FLEX (10).

Durch eine stratifizierte Analyse der Tätigkeit von 10.000 Ärzten in 24 Spezialgebieten wurden der Inhalt und die Ziele von 102.705 Arzt/Patienten-Interaktionen bestimmt und zur Grundlage einer Neugestaltung der FLEX 1 und FLEX 2 genannten Prüfungen gemacht, die 1985 erstmals eingesetzt werden soll. An 2 1/2 Tagen besteht die Prüfung aus Fragen mit Auswahlantworten, am letzten, sechsten Halbtage müssen 10-15 klinische Fallsimulationen (Patient Management Problems, PMP) bearbeitet werden (11). Ausländische Ärzte können an dieser Prüfung ebenfalls teilnehmen. Beim Erreichen eines Punktwertes von 75 wird FLEX neben FMGEMS, ECFMG und VQE als eine der medizinischen Prüfungen anerkannt, deren Bestehen Voraussetzung für das Erteilen des ECFMG-Zertifikates ist. In Zukunft wird besonders für Ausländer nur noch das Bestehen von FMGEMS relevant sein.

Es bestehen seit Jahren bis in die Gegenwart hinein Meinungsverschiedenheiten zwischen den medizinischen Fakultäten und den Prüfungsämtern über den Bezug der Prüfungen auf die wissenschaftlichen Grundlagen der Medizin und der Betonung der Praxisorientierung mit einer Verlagerung des Schwerpunktes auf die ärzt-

liche Kompetenz. Letzteres wird als Bedrohung der Förderung wissenschaftlicher Forschungen angesehen und von den Medizinschulen daher abgelehnt.

Ärztliche Prüfung für Aufenthaltsgenehmigungen Visa Qualifying Examination (VQE)

Für Ausländer mit einer im Ausland erworbenen ärztlichen Qualifikation wurde von der Bundesregierung 1976 eine besondere Prüfung (Visa Qualifying Examination, VQE) gefordert, die zur Ausübung der ärztlichen Tätigkeit in einem Krankenhaus berechtigt. Sie wurde von der unabhängigen nationalen Prüfungsorganisation National Board of Medical Examiners (NBME) entwickelt und 1977 erstmals für 4611 Teilnehmer an 27 Prüfungsstellen auf der ganzen Welt abgehalten. Sie sollte den höheren Qualitätsansprüchen des Staatsexamens für Absolventen der nordamerikanischen Medizinschulen ("National Board") entsprechen. Mit der daraus resultierenden höheren Durchfallquote, die bei etwa 73 % gegenüber 60 % bei den ECFMG-Prüfungen lag, ging dann auch die Zahl der in die USA einwandernden Ärzte zurück.

Die Prüfung wurde an zwei Tagen abgehalten, am ersten Tag waren 500 Fragen mit Auswahlantworten aus den Grundlagenfächern und am zweiten Tag 450 Fragen aus den klinischen Fächern zu bearbeiten. Die letzten Teilnehmerzahlen waren: 1981: 5.574; 1982: 7.401; 1983: 7.449. Detaillierte Prüfungsergebnisse der VQE fanden wir nicht in den Veröffentlichungen des NBME oder ECFMG.

Die Tabelle 1 gibt eine Übersicht über die amerikanischen ärztlichen Prüfungen, die auf Bundesebene durchgeführt werden. Das National Board of Medical Examiners, NBME, ist eine unabhängige Organisation, die außer von der Vereinigung der ärztlichen Prüfungsämter der Bundesstaaten, Federation of State Medical Boards (FSMB), von sechs weiteren ärztlichen Organisationen eingerichtet wurde. Ihr ursprünglicher Zweck war die Durchführung der Prüfungen für die Studenten der anerkannten amerikanischen Medizinschulen, die in drei Teilprüfungen abgelegt

Tab. 1:

Ärztliche Prüfungen in den USA

Prüfung	Dauer in Tagen	Anzahl der Fragen		
		Grundlagen fächer	klinische Fächer	Gesamtzahl
"NBME"	5	1.000	1.660 ^x	2.660
FLEX	3	540	1.300 ^x	1.840
VQE	2	500	450	950
FMGEMS	2	500	450	950
"ECFMG"	1	60	300	360
"AAppO"	9	540	830	1.370

^x Zusätzlich Patient Management Problems

"NBME" Prüfungen (Teil I, II und III) des National Board of Medical Examiners für Studenten der amerikanischen Medizinschulen

"AAppO" Prüfungen nach der Approbationsordnung für Ärzte

Obrige Abkürzungen siehe Text

werden. Sie haben heute nur noch einen Anteil von 22 % an den vom NBME durchgeführten Prüfungen. Etwa drei Viertel der Absolventen der amerikanischen Medizinschulen beteiligen sich an den Prüfungen des "National Board", die übrigen legen die Prüfung des Bundesstaates, in dem sie praktizieren wollen oder die Prüfung der Föderation der "Landesprüfungsämter" (FLEX) ab. Alle in der Tabelle 1 aufgeführten Prüfungen werden vom NBME durchgeführt. Vergleichsweise ist der Umfang der Prüfungen nach der Approbationsordnung in der Tab. 1 angeführt.

Beteiligung an den ECFMG-Prüfungen

Die alte Prüfung des ECFMG blieb nur noch für Amerikaner mit einer im Ausland erworbenen ärztlichen Qualifikation von Bedeutung. Dies zeigte Auswirkungen auf die Beteiligung an den Prüfungen der ECFMG. Die Zahl der erstmalig an der Prüfung teilnehmenden Ärzte hatte 1975 einen Gipfel mit 20.415. Mit der Verschärfung der Prüfung einschließlich der Sprachprüfung trat ein Rückgang um etwa die Hälfte ein. Die Beteiligung steigt jedoch seit 1978 wieder kontinuierlich an und liegt zwischen 1981 und 1983 bei etwa jährlich 13.000 Erstteilnehmern, von denen aber die US-Bürger einen immer größer werdenden Anteil darstellen. Einschließlich der Wiederholer stieg die absolute Zahl der an der Prüfung teilnehmenden US-Bürger von 2.750 im Jahre 1977 auf 7.311

im Jahre 1983 an, der Prozentsatz von 11 auf 27. Im Jahr 1984 haben 3.342 US-amerikanische Staatsbürger, die außerhalb Nordamerikas an einer von der Verbindungskommission für ärztliche Ausbildung der amerikanischen Ärzteschaft, Liaison Committee for Medical Education (LCME), nicht anerkannten Medizinschule eine ärztliche Qualifikation erworben haben und die Prüfungen des ECFMG bestanden haben, sich um eine Weiterbildungsstelle bei der Nationalen Vermittlungsstelle beworben. Zusätzlich dazu hat eine unbekannte Zahl unabhängig davon mit eigener Initiative eine Stelle gefunden (4). 51 % der Bewerber mit amerikanischer Staatsbürgerschaft und ausländischem Studienabschluß konnten nicht vermittelt werden, bei den Ausländern waren es 75 %. Von den Absolventen nordamerikanischer Medizinschulen, unter denen weniger als 1 % Nicht-US-Bürger sind, konnten 8 % nicht vermittelt werden.

Auch nach mindestens dreijähriger Weiterbildung in den USA sind die Bestehensquoten zwischen den Absolventen der amerikanischen Medizinschulen, den ausländischen Ärzten und den US-Bürgern mit einer im Ausland erworbenen ärztlichen Qualifikation nachweisbar unterschiedlich. Zwischen 1975 und 1978 lagen die Bestehensquoten bei der Facharztprüfung für Innere Medizin bei den drei Gruppen um 80 %, 40 % und 20 %. Danach trat eine relative Verbesserung bei den im Ausland qualifizierten US-Bürgern ein, 1980 betragen die Bestehensquoten der drei genannten Gruppen 82 %, 45 % und 38 % (12).

Die neue Prüfung der amerikanischen Kommission für ausländische Ärzte (ECFMG)

Foreign Medical Graduate Examination in the Medical Sciences (FMGEMS)

1982 wurde auf der Grundlage einer Verlagerung der Interessen eine Änderung der Prüfungen für ausländische Ärzte eingeleitet. Die ab 1984 verfügbare "Prüfung für Ausländer in den medizinischen Wissenschaften" (Foreign Medical Graduate Examination in the Medical Sciences, FMGEMS) soll in der Schwierigkeit den Prüfungen entsprechen, die vom NBME für die Absolventen der zur Ärzteausbildung zugelassenen nordamerika-

nischen Medizinschulen durchgeführt werden.

Die nach langen Diskussionen und Vorbereitungen 1984 eingeführte Prüfung ersetzt sowohl die eintägige ECFMG-Prüfung für US-Bürger als auch die zweitägige Visa Qualifying Examination (VQE), deren Bestehen für nicht-amerikanische Ärzte eine der Voraussetzungen für die Erteilung einer Aufenthaltsgenehmigung in den USA war. Die Vorbereitung und Durchführung der Prüfung liegt bei der nationalen Prüfungsstelle für Arzt (National Board of Medical Examiners). Es werden dazu Fragen verwendet, die bei Prüfungen des National Board of Medical Examiners (NBME) verwendet und evaluiert wurden. Die als Ergänzung der schriftlichen Prüfung geforderten praktischen Prüfungen am Patienten können wegen der damit verbundenen technischen und logistischen Schwierigkeiten nicht eingeführt werden.

Teile der Prüfung

Die Prüfung besteht aus zwei Komponenten, die an getrennten Tagen abgenommen werden. Die Prüfung in den sieben Grundlagenfächern (Anatomie, Verhaltenswissenschaften, Biochemie, Mikrobiologie, Pathologie, Pharmakologie, Physiologie) besteht aus etwa 500 Fragen mit Auswahlantworten, die in 6 1/2 Stunden zu beantworten sind (Zeit pro Frage= 47 Sekunden).

Die zweite, klinische Komponente betrifft die Fächer Innere Medizin, Geburtshilfe und Frauenheilkunde, Kinderheilkunde, Vorsorgemedizin und öffentliches Gesundheitswesen, Psychiatrie und Chirurgie, die ebenfalls etwa gleichmäßig bei der Aufstellung der etwa 450 Fragen berücksichtigt werden. Für die Beantwortung einer Frage stehen im Mittel 48 Sekunden zur Verfügung. Voraussetzung für die Zulassung zum ersten Teil (Grundlagenfächer) ist der Abschluß der ersten beiden Studienjahre einer in der Liste der Weltgesundheitsorganisation aufgeführten Medizinschule, zum zweiten Teil (Klinik) der in nicht mehr als 12 Monaten anstehende oder bereits erreichte Abschluß des sogenannten didaktischen Teils des Medizinstudiums an einer anerkannten Medizinschule.

Der englische Sprachtest muß an demselben Tag wie die klinische Komponente abgelegt werden. Er prüft Verständnis anhand von Tonbandeinpielungen sowie Struktur und Vokabeln der englischen Umgangssprache ohne besonderen medizinischen Inhalt. Er verliert nach zwei Jahren seine Gültigkeit, falls in dieser Zeit keine klinische Tätigkeit in den USA aufgenommen wird und kann dann durch eine Wiederholung oder durch das Ablegen des TOEFL (Test of English as a Foreign Language) erneuert werden.

Prüfungsbewertung

Die Ergebnisse werden getrennt für die beiden Komponenten berichtet, bei beiden muß ein Punktwert von 75 erreicht werden. Die Umwandlung der richtigen Antworten in den Punktwert ("scaled score") stellt keine Funktion des Prozentsatzes richtiger Antworten unter Bezug auf die Rohwertverteilung der Prüfgruppe dar. Der Punktwert wird vielmehr an der Leistung amerikanischer und kanadischer Kandidaten standardisiert und sichert eine Vergleichbarkeit der Beurteilung über die gesamte Einsatzdauer der Fragen. Durch die vor Jahren erfolgte Umstellung der Prüfungsauswertung durch das NBME ist es möglich, die Beurteilung der Leistung des einzelnen Kandidaten von der Schwierigkeit der Fragen und der Leistung der jeweilig geprüften Gruppe unabhängig zu machen und ermöglicht somit einen über Jahre konstanten Bewertungsmaßstab. Das Bestehen dieser Prüfung ist neben dem Bestehen des Sprachtests Voraussetzung für die Erteilung des Zertifikates der Ausbildungskommission für ausländische Ärzte (ECFMG). Zusätzlich ist für die Nicht-US-Bürger der Nachweis der uneingeschränkten Berechtigung zur Ausübung einer ärztlichen Tätigkeit im Heimatland erforderlich.

Die Ergebnisse der Durchführung der ersten Prüfung zeigen erneut die unterschiedlichen Schwierigkeiten der einzelnen Komponenten, obwohl ein ursprünglich angegebenes Ziel der Neuentwicklung gerade die Aufhebung der unterschiedlichen Schwierigkeiten zwischen den Grundlagenfächern und der Klinik war. Wie die nachfolgende Tabelle 2 zeigt, bestehen auch unterschiedliche Bestehensquoten zwischen den US-Bürgern und den an den Prüfung teilnehmenden anderen Staatsangehörigen. Diese bezieht

Tab. 2:

Prüfung für ausländische Ärzte in den medizinischen Wissenschaften
Foreign Medical Graduate Examination in the Medical Sciences
Ergebnisse des ersten Prüfungstermines 1985

Teilnehmer	US-Bürger		Nicht-US-Bürger		Alle Teilnehmer	
	N	%	N	%	N	%
Teilnehmer	1196		6227		7423	
Mit Erfolg bestanden:						
Grundlagenfächer	47	4,0	1274	20,5	1321	17,8
Klinische Fächer	255	21,4	3129	50,2	3384	45,6
Gesamtprüfung	45	3,8	1246	20,0	1291	17,4

sich vor allem auf die unterschiedliche Bestehensquote in den Grundlagenfächern.

Ziel der Prüfung

Wenn das politisch definierte Ziel dieser neuen Prüfung ist, die im Ausland als Ärzte qualifizierten US-Bürger von der Ausübung des Arztberufes in den USA abzuhalten, erfüllt die neue Prüfungsform diese Anforderung in bewundernswerter Weise. Der Unterschied in der Bestehensquote der Gesamtprüfung zwischen den 1196 US-Bürgern mit 3,8 % und den übrigen 6227 Teilnehmern mit 20 % ist statistisch hochsignifikant ($\chi^2 = 184,3$; $p \ll 0,001$). Da vor der Aufnahme einer ärztlichen Tätigkeit in den USA von den Nicht-US-Bürgern noch zahlreiche weitere Hürden zu überwinden sind, spielt die höhere Bestehensquote bei diesen keine so übertragende Bedeutung. Ausländer lassen sich notfalls durch weitere Reglementierungen der Einreisebedingungen oder durch andere Erschwernisse, etwa bei der erforderlichen Wiederholung des Sprachtestes, von der Aufnahme der ärztlichen Tätigkeit in den USA abhalten!

Die globale Aussage, daß die im Ausland ausgebildeten US-Bürger bei den neuen Prüfungen so viel schlechter als ihre Kollegen ohne US-Staatsangehörigkeit abschneiden, erfordert jedoch eine differenzierte Analyse. Dazu müssen außer den Durchfallsquoten die Persönlichkeitsmerkmale und die Studienländer in die Betrachtung einbezogen werden. Dies ist ohne Zugang zu den Daten des ECFMG allein aus den mitgeteilten Ergebnissen nicht möglich. Grundsätzlich haben in einer Diskussion vor einigen Jahren Vertreter des ECFMG

ihre Bereitschaft erklärt, den Zugang zu den Daten für weitere wissenschaftliche Untersuchungen zu ermöglichen.

Literaturverzeichnis

1. J. EIGLER
Kritik als Anregung - Aspekte medizinischer Ausbildung in den achtziger Jahren
Internist 21 : 123-126 (1980)
2. A. McGUINNESS; H. MASON
Career Destiny of 550 Americans Several Years After Graduating From A Foreign Medical School
3. B. STIMMEL
Medical Education and the Regents of the State of New York
N Engl J Med 304 : 602-607 (1981)
4. B. STIMMEL; J. GRAETTINGER
Medical Students Trained Abroad And Medical Manpower
N Engl J 310 : 230-235 (1984)
5. ASSOCIATION OF AMERICAN MEDICAL COLLEGES
Quality of Preparation for the Practice of Medicine In Certain Foreign-Chartered Medical Schools
J Med Educ 56 : 962-979 (1981)
6. T. SAMPH; B. TEMPLETON
Evaluation in Medical Education
Past, Present, Future
Cambr. Mass. : Ballinger Publ. Comp. 1979
7. J.R. SCHOFIELD
New and Expanded Medical Schools, Mid-Century to the 1980s
Washington : Association of American Medical Colleges and San Francisco : Jossey-Bass Inc., 1984
8. B. STIMMEL; J. GRAETTINGER
Graduates of Foreign Medical Schools and Medical Manpower
N Engl J Med 310 : 1541 (1984)
9. J. COOPER
ECFMG Reports Examination Results
President's Weekly Activities Report, 85-17, April 1985
10. NATIONAL BOARD OF MEDICAL EXAMINERS
Annual Report 1983
Philadelphia : National Board of Medical Examiners, 1984
11. A. LaDUCA; D. TAYLOR; I. HILL
The Design of a New Physician Licensure Examination

Evaluation & The Health Professions,

7 : 115-140 (1984)

12. J. BENSON; J. MESKAUSKAS; L. GROSSO

Performance of US-Citizen-Foreign Medical

Graduates on Certifying Examinations in

Internal Medicine

Amer J Med 71 : 270-273 (1981)

Prof. Dr. H. E. Renschler, Institut für Didaktik
der Medizin, Sigmund-Freud-Straße 25, D 5300 Bonn 1
und Prof. Dr. D. Habeck, Albert-Schweitzer-Str. 11
D 4400 Münster

ERGEBNISSE DER ECFMG-PRÜFUNG IM INTERNATIONALEN VERGLEICH

D. Habeck, Münster und H. E. Renschler, Bonn

Zusammenfassung:

Aus den Mitteilungen der Ergebnisse der ECFMG-Prüfungen wurden die Bestehensquoten als Mittelwert aus den Prüfungen der Jahre 1977-1983 berechnet. Sie wurden getrennt nach Ländern für die Staatsangehörigen der Vereinigten Staaten und für die aller übrigen Länder ermittelt, die mehr als insgesamt 50 Teilnehmer hatten. Unter dem Aspekt des Erfassens des Ausbildungserfolges in der Medizin wurde das Ergebnis der Sprachprüfung nicht berücksichtigt.

Für die US-Staatsbürger zeigte sich eine unterschiedliche Häufigkeitsverteilung zwischen den europäischen Ländern und denen von AuBereuropa. Eine bimodale Verteilung, die jedoch um etwa 10-15 % niedriger liegt, fand sich auch für die nach Ländern berechneten Bestehensquoten der Teilnehmer, die nicht die amerikanische Staatsbürgerschaft besitzen. Die schlechtere Gruppe verteilt sich um den Mittelwert von etwa 30 %, die bessere um den Mittelwert von etwa 70 %. Die Position der Teilnehmer, die ihr Studium in der Bundesrepublik absolviert haben, liegt mit 41,5 % für die US-Bürger und mit 53,3 % für die übrigen Teilnehmer auf der Grenze zwischen den beiden Populationen. Es wird auf die zahlreichen Einschränkungen der Bedeutung der Ergebnisse und ihrer Interpretation verwiesen, die daher zunächst nur als Hypothesen verstanden werden sollen.

Einleitung

Von der Kommission für ausländische Ärzte der Vereinigten Staaten (Educational Commission for Foreign Medical Graduates, ECFMG) werden weltweit einheitliche Prüfungen abgehalten. An diesen können Ärzte teilnehmen, die ihre Qualifikation außerhalb der USA an einer Medizinschule erworben haben, die in der von der Weltgesundheitsorganisation geführten Liste ärztlicher Ausbildungsstellen enthalten ist. Die Teilnehmerzahlen und Ergebnisse der Prüfungen werden mit vielen Einzelheiten seit 1972 veröffentlicht (1). Sie bieten sich als Indikator

für die Qualität nationaler Ausbildungssysteme an. In den folgenden Ausführungen wird eine gegliederte Darstellung der Summe der Ergebnisse der Jahre 1977 bis 1983 gegeben, um Interessierten die Möglichkeit einer eigenen Urteilsbildung zu ermöglichen. Diese wird allerdings schon durch die Art der Darstellung und durch die Auswahl in bestimmte Richtungen gelenkt. Um jedoch eine möglichst objektive Betrachtung zu ermöglichen, werden die Ergebnisse aus allen Staaten systematisch einbezogen. Eine Auswahl erfolgt nur unter statistischen Gesichtspunkten.

Eine Übersicht über die Prüfungen für Ausländer wird in der vorausgehenden Arbeit (Renschler und Habeck) gegeben. 1977 wurden durch die Einführung einer besonderen Prüfung für die ausländischen Ärzte, die eine Aufenthaltsgenehmigung für die USA beantragen wollten (Visa Qualifying Examination, VQE) und durch die Einbeziehung des Sprachtestes in die mitgeteilten Ergebnisse neue Bedingungen geschaffen. Diese blieben bis zur Einführung einer neuen Prüfung (Foreign Medical Graduate Examination in the Medical Sciences, FMGEMS) im Jahre 1984 konstant. Es erscheint daher sinnvoll, die Ergebnisse der Jahre 1977 bis 1983 zusammenzufassen. Unter Vernachlässigung geringer zeitlicher Änderungen erfolgte aus den Ergebnissen eine Mittelwertbildung. Damit ergeben sich bei einer hinreichend großen Anzahl von Staaten genügend große Zahlen, die Vergleiche gestatten. In die eigentlichen Betrachtungen werden mit geringen Ausnahmen nur die Staaten einbezogen, die von 1977 bis 1983 insgesamt mehr als 50 Teilnehmer jeweils in den Kategorien US-Bürger und Nicht-US-Bürger hatten.

Da die US-Bürger, die in zunehmender Zahl ein Medizinstudium außerhalb der Vereinigten Staaten absolvieren, sich nicht nur in der Sprache, sondern wahrscheinlich in weiteren Persönlichkeitsmerkmalen von den Angehörigen anderer Staaten, die sich an den Prüfungen beteiligen, unterscheiden, sollen sie getrennt dargestellt werden.

Unter dem für die Beurteilung der ärztlichen Ausbildung wichtigen Gesichtspunkt des Beherrschens medizinischen Wissens spielen die englischen Sprachkenntnisse eine geringe, wenn auch nicht zu vernachlässigende Rolle, so daß bei den Nicht-US-Bürgern die Bestehensquote des medizinischen Teiles ohne Berücksichtigung des Versagens im englischen Sprachtest sinnvoll erscheint. Hier bieten die US-Bürger den Vorteil der Einheitlichkeit, da bei ihnen der Sprachtest irrelevant ist.

Einschränkungen der Auswertung

Für alle Analysen und Ergebnisse gelten jedoch die folgenden Einschränkungen:

1. Wie in der voranstehenden Arbeit ausgeführt, wurden die ECFMG-Prüfungen mit dem Ziel eingeführt, zur Zeit des Ärztemangels möglichst viele ausländische Ärzte für die Tätigkeit in den Krankenhäusern in den USA zu gewinnen. Die Kriterien stellen daher nur die unterste Grenze des erforderlichen Wissens bei einem Maximum an tolerierbaren Lücken dar. Dies ergibt sich aus dem Vergleich der Durchfallsquoten, die aus der Festlegung der Bestehensgrenze bei einem Standardwert von 290 resultieren. Bei Absolventen amerikanischer Medizinschulen hätte die Durchfallsquote nur 1,8 % betragen, bei den Teilnehmern an den ECFMG-Prüfungen lag sie bei 60-70 %.
2. Die Prüfungen messen ausschließlich kognitive Anteile, wobei die mit den klinischen Fallsimulationen (PMP) erfaßbaren Fähigkeiten klinischen Problemlösens nicht in die Prüfung einbezogen worden sind. Psychomotorische Fertigkeiten und ärztliche Einstellungen werden nicht geprüft.
3. Es ist unbekannt, was die Nicht-US-Bürger zur Teilnahme an den ECFMG-Prüfungen veranlaßt hat. Da für das Erlangen einer Einreise genehmigung nach 1977 eine andere Prüfung eingeführt worden ist, bleibt das Ergebnis der ECFMG-Prüfung für Nicht-US-Bürger unverbindlich. Es ist nicht ersichtlich, warum die Kommission Nicht-US-Bürger überhaupt noch zugelassen hat, die nur für die relativ kleine Anzahl der US-Bürger nach 1977 noch

Bedeutung hat. Dienten die Nicht-US-Bürger nur dem Zweck, die Prüfungsorganisation im Sinne der amerikanischen Staatsbürger noch aufrecht zu erhalten? Ist die selbstbestimmte Teilnahme der Ausländer positiv im Sinne einer freiwilligen Unterziehung einer relativ strengen Prüfung zu werten oder haftet der Meldung zur Prüfung einer amerikanischen Einrichtung ein negatives Urteil an? Das auffällig schlechte Abschneiden der Nicht-Amerikaner in den sozialistischen Staaten legt zumindest für diese eine solche Interpretation nahe.

4. Besondere Bedingungen gelten für die US-Bürger. Bei ihnen ist anzunehmen, daß sie vollzählig an den Prüfungen teilnehmen, soweit ihnen nicht durch andere Mittel eine Rückkehr in die USA noch vor Ende der Grundausbildung möglich geworden ist. Da sie sich wohl mit nur geringen Ausnahmen um einen Studienplatz an einer Medizinschule in Nordamerika beworben hatten, ist eine den Aufnahmebedingungen entsprechende relativ einheitliche Vorbildung und Motivation anzunehmen. Da sie trotz einer im Durchschnitt mehrfachen Bewerbung keinen Studienplatz bekommen haben, ist jedoch anzunehmen, daß bei ihnen die Begabungshöhe nach oben begrenzt ist. Nur wenige Amerikaner der auf die Einwanderung folgenden Generationen beherrschen trotz starker ethnischer Bindungen noch die Sprache des Herkunftslandes der einwandernden Generation. Das Meistern der Sprache des Gastlandes in dem - mit Ausnahme der englischsprachigen Länder - der Unterricht stattfindet, spricht dagegen dafür, daß ein nicht geringes Maß an Lernfähigkeit bei den US-Bürgern vorhanden ist. Es kann daher davon ausgegangen werden, daß die US-Bürger eine in sich einheitlichere Gruppe darstellen als die Nicht-US-Bürger. Ihre Ergebnisse werden daher getrennt dargestellt. Den Aussagen der Interpretation der Ergebnisse kann eine größere Bedeutung zugemessen werden.

Wenn auch anzunehmen ist, daß fast alle der in einem Land sich den Prüfungen unterziehenden Nicht-Amerikaner Bürger dieses Landes sind, ist bei der Freizügigkeit, besonders in Europa, mit

einem gewissen Anteil fremder Staatsbürger in den jeweiligen Ländern zu rechnen. Es wird daher die Beziehung "Nicht-US-Bürger" für alle nicht als US-Bürger ausgewiesenen US-Bürgern verwendet.

Methodik

Aus den Mitteilungen der Ergebnisse der Prüfungen der ECFMG der Jahre 1977 bis 1983 wurde getrennt nach Ländern die Zahlen der Teilnehmer und derer, die die gesamte Prüfung sowie den medizinischen Teil bestanden haben, zusammengestellt. Die Aufstellungen werden getrennt für Europa und die außereuropäischen Staaten sowie für die US-Bürger und Nicht-US-Bürger gemacht. Israel wird hier den europäischen Staaten zugerechnet, da es aufgrund seiner Kultur stärker mit Europa verbunden sein dürfte als mit den benachbarten arabischen Staaten und dementsprechend auch der europäischen Organisation für die medizinische Ausbildung (Association for Medical Education in Europe, AMEE) angehört.

Ergebnisse

1. EUROPA

1.1 Prüfungsergebnisse der US-Bürger in europäischen Ländern.

Die Ergebnisse von 19 europäischen Ländern sind in Tab. 1 in der Reihenfolge der größten Quoten aufgelistet. Da insgesamt nur 1,2 % der amerikanischen Staatsbürger den Sprachtest nicht bestanden haben, sind die Ergebnisse der Gesamtprüfung von ausreichender Bedeutung. Die Teilnehmerzahl von 14 Ländern war größer als 50. Die Ergebnisse der Länder mit weniger als insgesamt 50 Teilnehmern in den Jahren 1977-1983 sind in Klammern gesetzt. Zur Spitzengruppe mit den hohen Bestehensquoten von mehr als 80 % gehören außer den englischsprachigen Ländern Großbritannien und Irland auch Israel und Belgien. Infolge der kleinen Zahl von 16 Teilnehmern ist die Anführung von Schweden, von denen alle 16 dort studierenden US-Bürger die

Tab. 1 Prüfungsergebnisse der US-Bürger in Europa (ECFMG-Prüfungen 1977-1983)

Staat	Anzahl: Teilnehmer	Bestehens- quote (%) in Prozent	Anzahl: erfolgreiche Teilnehmer
1. Schweden	16	(100,0)	16
2. Israel	176	98,3	173
3. Großbritannien	87	93,1	81
4. Irland	91	83,5	76
5. Belgien	329	80,5	265
6. Frankreich	275	66,2	182
7. Schweiz	77	64,9	50
8. Niederlande	26	(50,0)	13
9. Polen	320	46,6	149
10. Italien	1.913	43,8	838
11. BRD	284	41,5	118
12. Österreich	112	39,3	44
13. Portugal	23	(34,8)	8
14. Jugoslawien	39	(33,3)	13
15. Ungarn	71	31,0	22
16. Griechenland	332	29,5	98
17. Rumänien	185	26,5	49
18. Türkei	26	(23,1)	6
19. Spanien	2.587	14,5	376

Tab. 2 Prüfungsergebnisse der Nicht-US-Bürger aus europäischen Staaten bei den ECFMG-Prüfungen 1977-1983

	Staat	Anzahl Prüfungs- teilnehmer	Bestehens- quote (%) für med. Teil insgesamt	Bestehens- quote (%) für Gesamt- prüfung	Anzahl erfolg- reicher Teilnehmer
1.	Großbritannien	1.642	87,0	86,5	1.421
2.	Schweden	392	77,3	71,9	282
3.	Dänemark	567	82,5	68,6	389
4.	Norwegen	151	76,2	66,9	101
5.	Irland	805	64,1	63,9	514
6.	Niederlande	574	66,9	60,8	349
7.	Israel	782	78,6	48,0	375
8.	Finnland	185	63,8	40,0	74
9.	Schweiz	769	74,9	39,4	303
10.	Bundesrep. D	2.835	55,3	37,4	1.059
11.	Belgien	861	62,8	35,4	305
12.	Portugal	173	39,9	28,9	50
13.	Österreich	324	36,7	24,7	80
14.	Frankreich	1.594	38,7	21,0	335
15.	DDR	56	17,9	14,3	8
16.	Italien	2.952	29,9	13,7	403
17.	Ungarn	382	33,8	12,6	48
18.	CSSR	395	25,9	12,2	48
19.	Griechenland	1.165	27,8	12,1	141
20.	Türkei	860	25,9	11,4	98
21.	Spanien	3.145	24,7	11,0	345
22.	Polen	1.818	25,9	9,5	172
23.	Rumänien	1.140	27,2	8,4	107
24.	Jugoslawien	850	19,0	8,4	72
25.	UDSSR	3.346	34,3	7,7	258
26.	Bulgarien	214	17,7	7,0	214

Prüfung bestanden haben, schwierig. Die Bundesrepublik liegt mit 41,5 % in der Zehnprozentgruppe mit der größten Häufigkeit. Trotz relativ niedriger Bestehensquoten für Italien mit 43,8 % und den schlechtesten Prüfungsergebnissen für Spanien (14,5 %) stellen diese beiden Staaten mit 838 bzw. 376 die größte Anzahl der erfolgreichen Kandidaten dar.

1.2 Prüfungsergebnisse der Nicht-US-Bürger in europäischen Ländern.

In Tab. 2 sind die Anzahl der Teilnehmer aus 26 europäischen Staaten, die nicht Staatsangehörige der Vereinigten Staaten sind und ihre Prüfungsergebnisse aufgeführt. Zusätzlich zu den Quoten des Bestehens der gesamten Prüfungen sind die Bestehensquoten für den medizinischen Teil der Prüfung bei gleichzeitigem Nicht-Bestehen des englischen Sprachtestes angegeben.

Die nach der Quote des Bestehens der gesamten Prüfung angeordnete Rangliste zeigt eine bimodale Verteilung mit einer Lücke zwischen den ersten sechs Staaten, die eine Bestehensquote von mehr als 60 % aufweisen. Nimmt man die Staaten hinzu, deren Teilnehmer den medizinischen Teil, nicht aber den Sprachtest bestanden haben, rücken Israel, Finnland, die Schweiz, Belgien und die Bundesrepublik in die Spitzengruppe auf, von denen mehr als 50 % der Teilnehmer im medizinischen Teil erfolgreich waren. Diese Verhältnisse sind für 16 ausgewählte Länder in Abb. 1 wiedergegeben.

Mit der Ausnahme der Niederlande liegen die Bestehensquoten der US-Bürger bei diesen Staaten deutlich über denen der übrigen Teilnehmer aus den betreffenden Staaten. Diese Verhältnisse sind in Abb. 2 dargestellt.

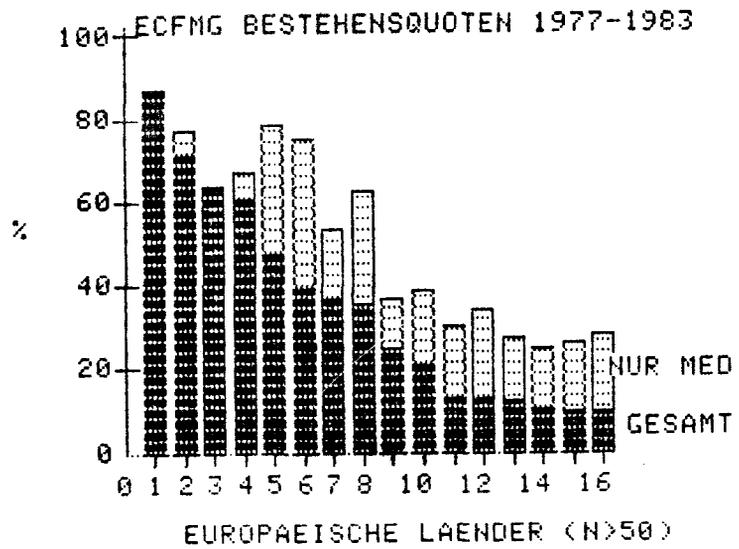


Abb. 1 Dargestellt sind für 16 ausgewählte europäische Länder die Bestehensquoten für die Teilnehmer an der Prüfung, die Nicht-US-Bürger sind. Die ausgefüllten Teile der Balken repräsentieren die Kandidaten, die die gesamte Prüfung, die schraffierten oberen Teile die Kandidaten die den medizinischen Teil der Prüfung, nicht aber die englische Sprachprüfung bestanden haben.

Alle Teilnehmer aus Großbritannien (Nr. 1) und aus Irland (Nr. 3) haben den Sprachtest bestanden.

Liste der dargestellten Länder:

1	Großbritannien	5	Israel	9	Österreich	13	Griechenland
2	Schweden	6	Schweiz	10	Frankreich	14	Spanien
3	Irland	7	BRD	11	Italien	15	Polen
4	Niederlande	8	Belgien	12	Ungarn	16	Rumänien

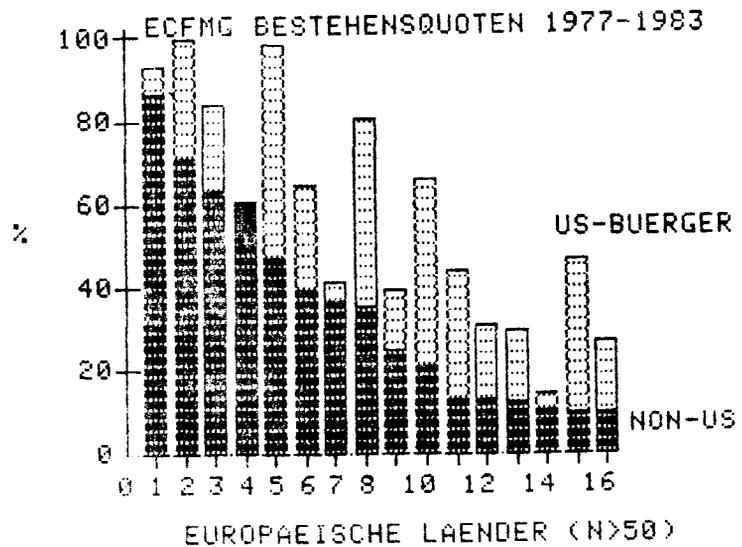


Abb. 2 Dargestellt sind die Bestehensquoten von 16 ausgewählten europäischen Ländern für die Nicht-US-Bürger (ausgefüllter Teil der Säulen) und für die Bürger der Vereinigten Staaten (schraffierter Teil). Mit der Ausnahme von Schweden (Nr. 2, N = 16) und den Niederlanden (Nr. 4, N = 26) haben jeweils mehr als 50 US-Bürger in den Jahren 1977 bis 1983 an der Prüfung teilgenommen. Trotz der geringen Teilnehmerzahl sind jedoch die Unterschiede der Quoten für US-Bürger und Nicht-US-Bürger zwischen den beiden Ländern hochsignifikant ($p = 0,993$).

Liste der dargestellten Länder:

1	Großbritannien	5	Israel	9	Österreich	13	Griechenland
2	Schweden	6	Schweiz	10	Frankreich	14	Spanien
3	Irland	7	BRD	11	Italien	15	Polen
4	Niederlande	8	Belgien	12	Ungarn	16	Rumänien

Bemerkenswert sind die fünf Länder Israel, Schweiz, Belgien, Frankreich und Polen, bei denen die US-Bürger Bestehensquoten von nahezu 50 % oder sehr viel besser haben und große Differenzen zu den anderen Teilnehmern aus diesen Ländern bestehen. Offensichtlich vermögen die US-Bürger unter den Bedingungen dieser Länder sehr viel besser zu lernen als die einheimischen Staatsbürger. Diese Differenz bleibt mit Ausnahme der Schweiz bestehen, wenn zur Bestehensquote der Gesamtprüfung die des medizinischen Teiles addiert wird.

2. AUSSEREUROPÄISCHE LÄNDER

2.1 Prüfungsergebnisse der US-Bürger

In Tabelle 3 sind die Anzahl der Teilnehmer mit US-Staatsangehörigkeit mit ihren Ergebnissen

aufgelistet. Zum Vergleich sind die Bestehensquoten der Nicht-US-Bürger mit angeführt. Die Bestehensquoten der Länder mit weniger als 50 Teilnehmern sind in Klammern gesetzt; sie sollen bei den weiteren Überlegungen außer Betracht bleiben. Infolge der kleinen Zahl ist es wenig sinnvoll, ausführliche statistische Verfahren anzuwenden.

Die Häufigkeitsverteilung weist nach der Übertragung in das Wahrscheinlichkeitsnetz mit visueller Beurteilung nahezu Normalverteilung auf. Nur die Ergebnisse von Grenada liegen mit 82,1 % weit außerhalb der Streubreite. Der Abstand vom Mittelwert von 41,2 beträgt 2,7 Standardabweichungen, was eine Wahrscheinlichkeit von 0,0035 für die Zugehörigkeit zur übrigen Population gibt.

Tab. 3 Prüfungsergebnisse der US-Bürger aus außereuropäischen Staaten bei den ECFMG-Prüfungen 1977-1983

Staat	Anzahl Prüfungs- teilnehmer	Bestehens- quote (%) für Gesamtprüfung	Anzahl erfolgreicher Teilnehmer	jeweilige Bestehensquote (%) für Nicht-US-Bürger
1. Australien	14	(100,0)	14	95,6
2. Grenada	800	82,1	657	52,9
3. Indien	75	54,7	41	25,0
4. Nigerien	22	(54,5)	12	35,5
5. Chile	17	(52,9)	9	19,2
6. Philippinen	1.015	48,1	488	17,9
7. Dominika	451	48,1	217	26,0
8. Mont Serrat	1.055	47,5	501	28,7
9. Argentinien	70	(45,7)	32	13,4
10. Rep. Korea	74	43,2	32	1,2
11. Jamaika	35	(42,9)	15	69,9
12. Mexiko	11.875	38,6	4.583	22,9
13. Costa Rica	36	(33,3)	12	22,9
14. Kolumbien	115	33,0	38	10,6
15. Brasilien	49	32,7	16	13,1
16. Santa Lucia	130	32,3	42	(15,2)
17. Arab.Rep.Ägypt.	81	32,1	26	9,0
18. Guatemala	112	29,5	33	9,2
19. Iran	17	(23,5)	4	7,0
20. Rep.Chin.(Taiwan)	63	22,2	14	9,5
21. Dominik.Rep.	8.014	19,7	1.576	9,5
22. Haiti	20	(15,0)	3	3,2
23. Ekuador	35	(11,4)	4	7,1
24. Kuba	20	(-)	-	7,1
gesamte Welt	31.405	35,1	11.031	18,9

Tab. 4

Bestehensquoten der Nicht-US-Bürger aus außereuropäischen Staaten
bei den ECFMG-Prüfungen 1977-1983

Staat	Anzahl Prüfungs- teilnehmer	Bestehens- quote (%) für med. Teil insgesamt	Bestehens- quote (%) für Gesamt- prüfung	Anzahl erfolg- reicher Teilnehmer
1. Australien	429	95,6	95,6	410
2. Hongkong	146	95,9	88,4	129
3. Singapur	130	92,3	86,9	113
4. Südafrika	1241	84,7	84,5	1049
5. Malaysia	157	82,8	79,0	124
6. Kenia	92	78,3	78,3	72
7. Zambia	89	70,8	70,8	63
8. Jamaica	682	70,2	69,9	477
9. Ghana	177	65,5	65,5	116
10. Libanon	848	71,0	57,4	487
11. Grenada	153	56,8	52,9	81
12. Sri Lanka	311	62,0	46,6	145
13. Nigeria	677	38,3	35,3	239
14. Uruguay	105	56,2	31,4	33
15. Mont Serrat	317	36,0	28,7	91
16. Dominica	131	30,6	26,0	34
17. Indien	21485	38,0	25,0	5365
18. Honduras	87	34,2	11,0	9
19. Costa Rica	175	44,0	22,9	40
20. Panama	246	43,1	22,0	54
21. Jordan	187	78,0	21,9	41
22. Algerien	57	29,8	19,3	11
23. Chile	406	52,7	19,2	78
24. Philipinen	16155	20,1	17,9	2889
25. Irak	877	44,3	17,7	155
26. Burma	641	30,6	17,2	110
27. Santa Lucia	33	(15,2)	15,2)	5
28. Pakistan	5839	21,8	14,2	830
29. Thailand	360	64,2	13,9	50
30. Argentinien	1336	33,0	13,4	179

Tab. 4 Fortsetzung

Staat	Anzahl Prüfungs- teilnehmer	Bestehens- quote (%) für med. Teil insgesamt	Bestehens- quote (%) für Gesamt- prüfung	Anzahl erfolg- reicher Teilnehmer
31. Saudiarabien	197	37,1	13,2	26
32. Brasilien	1.499	33,7	13,1	197
33. Mexiko	7.502	20,8	11,0	827
34. Kolumbien	1.000	31,9	10,6	106
35. Venezuela	595	26,4	10,3	61
36. Bangladesch	638	23,5	9,6	61
37. Rep.Chin.(Taiw.)	3.438	48,9	9,5	325
38. Dominik. Rep.	3.271	19,0	9,5	312
39. Guatemala	574	24,2	9,2	53
40. Arab.Rep.Ägypt.	6.662	27,7	9,0	601
41. Peru	952	29,1	8,7	83
42. Vietnam	1.811	44,4	7,7	140
43. El Salvador	211	42,2	7,1	15
44. Iran	1.648	33,1	7,0	116
45. Indonesien	327	17,1	6,4	21
46. Arab.Rep.Syrien	1.694	37,3	6,1	104
47. Ekuador	584	18,8	7,1	34
48. Bolivien	388	15,8	5,7	22
49. Nicaragua	265	24,9	5,3	14
50. Japan	4.634	48,4	4,8	224
51. Haiti	2.065	24,5	3,2	67
52. Afghanistan	394	17,2	2,5	10
53. Volksrep.China	804	15,5	2,1	17
54. Kuba	1.180	14,9	1,8	21
55. Rep. Korea	5.851	63,1	1,2	71
gesamte Welt	130.686	36,2	18,9	24.663

2.2 Prüfungsergebnisse der Nicht-US-Bürger

Diese sind in der Tab. 4 dargestellt. In der ersten Spalte der Bestehensquoten sind diese für den medizinischen Teil dargestellt, das Ergebnis des Sprachtests wird dabei nicht berücksichtigt. Die Quoten für das Bestehen der Gesamtprüfung, die für die Erteilung des Zertifikates entscheidend ist, zeigt die folgende Spalte, nach der auch die Reihenfolge erstellt ist. Aus der Differenz, bzw. deren Fehlen, ergibt sich der Anteil der Versager, der durch den Sprachtest bedingt ist. Dieser fehlt z.B. bei Australien, aber auch bei einigen afrikanischen Staaten, bei denen Englisch die offizielle Landessprache ist. Dabei fällt die große Anzahl von 20 Staaten auf, bei denen das Ergebnis des Englischtests die Quote unter 10 % drückt. Betrachtet man das Ergebnis des Bestehens des medizinischen Teiles allein, zeigt sich deutlich wieder eine bimodale Verteilung mit einem ersten Gipfel zwischen 20 % und 40 % und einem zweiten Gipfel zwischen 70 und 80 %.

3. WELTWEITE ERGEBNISSE

Berechnet man die Häufigkeitsverteilung unter Einschluß der europäischen Staaten, bei denen ebenfalls 14 eine Beteiligung von mehr als 50 US-Bürgern aufweisen, wird eine bimodale Verteilung der Bestehensquoten deutlich. In der Abb. 3.1 und 3.2 ist die Verteilung aus allen 28 Ländern mit hellen Balken dargestellt. In der oberen Abbildung 3.1 sind die außereuropäischen Länder, in der unteren die europäischen Länder als schwarzer Teil der Säulen hervorgehoben. Zur Spitzengruppe zählen Großbritannien, Israel, Irland, Belgien, Frankreich und die Schweiz sowie von allen außereuropäischen Staaten nur Grenada. Die drei folgenden Staaten, Polen, Italien und die Bundesrepublik mit Bestehensquoten von 47 %, 44 % und 42 % sind der durch die übrigen Staaten bestimmten Gruppe zuzuweisen, deren Modus zwischen 40 und 50 % Bestehensquote liegt.

Aus der Analyse der Häufigkeitsverteilungen der Bestehensquoten ergibt sich die Hypothese einer bimodalen Verteilung. Dies wird durch die Eintragung der kumulativen Häufigkeit in das Wahrscheinlichkeitsnetz unterstützt. Es lassen sich

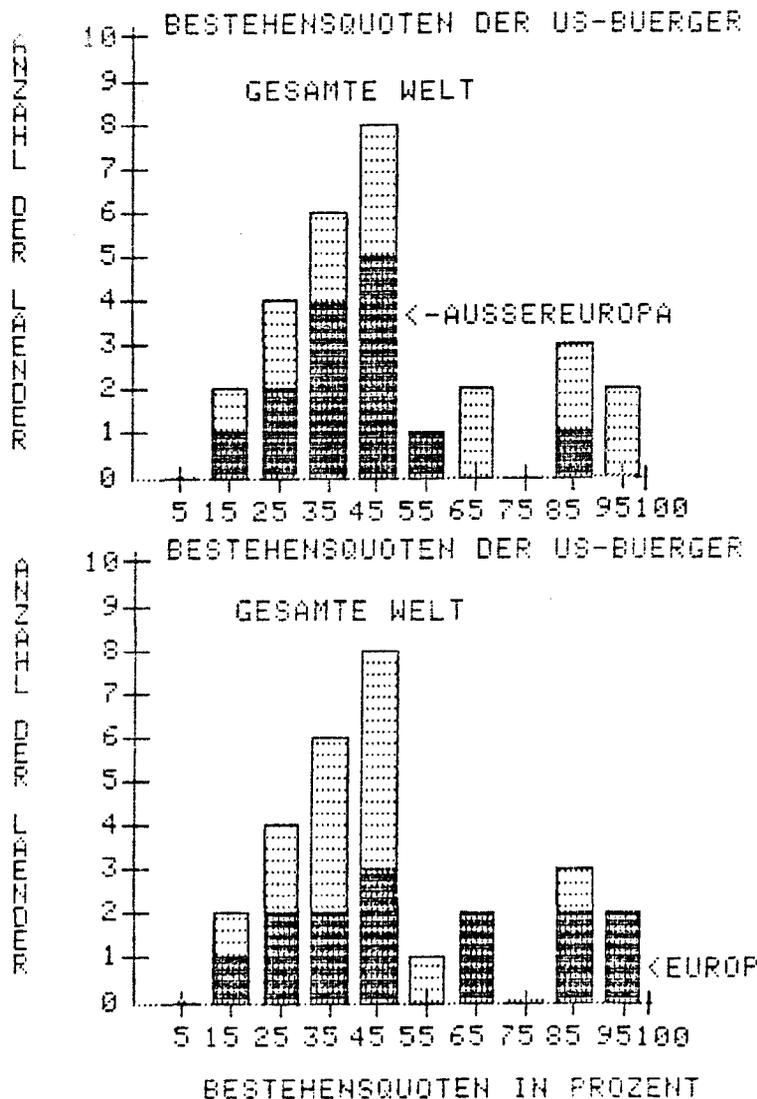


Abb. 3: Die Häufigkeitsverteilung umfaßt die Bestehensquoten der US-Bürger in den Ländern, in denen von 1977-1983 jeweils mehr als 50 US-Bürger teilgenommen haben. In der oberen Abbildung (3.1) sind die außereuropäischen Länder, in der unteren (3.2) die europäischen Länder als schwarzer Teil

so zwei unterschiedliche Populationen mit unterschiedlichen Mittelwerten und unterschiedlicher Breite der Basis vermuten. Wie bereits angeführt und in der Abb. 3.2 ersichtlich, liegen der Ergebnisse der US-Bürger mit Ausnahme von Grenada in der Population, deren größte Häufigkeit und deren Mittelwert zwischen Bestehensquoten von 40 und 50 % liegen. Die zeichnerisch ermittelte Standardabweichung beträgt 15 %, rechnerisch ergeben sich 15,7 %. Die vermutete bessere Population mit einem Gipfel um 80 % weist eine breitere Basis auf, entsprechend kann zeichnerisch eine Standardabweichung von 38 % abgelesen werden.

Es interessiert, ob sich für die Verteilung der Bestehensquoten der Nicht-US-Bürger ebenfalls die Hypothese einer bimodalen Verteilung aufstellen läßt. Da die englischen Sprachkenntnisse für diese Frage weniger relevant sind, wurden nur die Ergebnisse des medizinischen Prüfungsteils bei den Teilnehmern, die nicht US-Bürger sind, verwendet. Es sollen dazu nur die Staaten mit mehr als 50 Teilnehmern von 1977-1983 in die Untersuchung einbezogen werden.

In der Abb. 4 wurde die Häufigkeitsverteilung der Bestehensquoten der Teilnehmer aus 80 Ländern der Erde aufgetragen, die nicht die US-Staatsangehörigkeit besitzen (schraffierter Teil der Säulen). Es zeigt sich ein deutlicher Einschnitt zwischen 50 und 60 %, wobei der untere Gipfel mit 68 Ländern mit 20 - 30 % sehr viel höher ist als der obere, dessen größte Häufigkeit (N = 9) bei 70 bis 80 % liegt. Die Hypothese einer bimodalen Verteilung wird durch die Darstellung im Wahrscheinlichkeitsnetz unterstützt (Abb. 5). Die Steigungen, die durch die Standardabweichungen bestimmt werden, sind weniger verschieden als bei den US-Bürgern. Dies betrifft vor allem die Spitzengruppe, bei der eine Standardabweichung von 24 % abgelesen werden kann. Vergleichsweise wurden die Verteilung der Ergebnisse der US-Bürger aus 28 Staaten der Welt als schwarzer Teil der Säulen mit eingetragen.

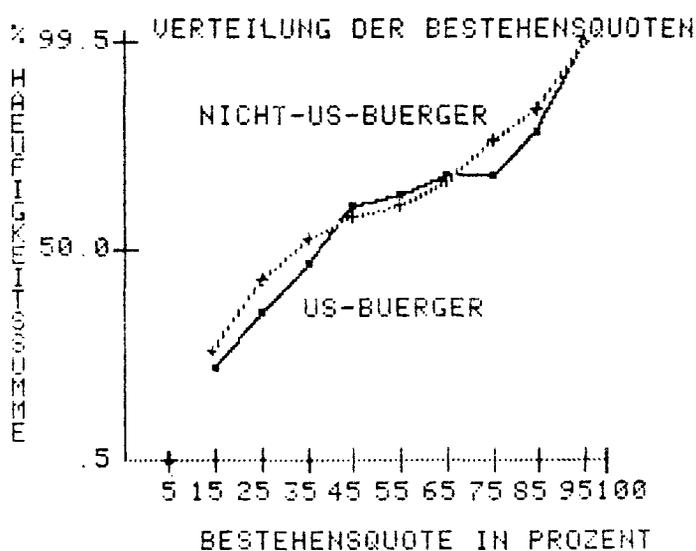
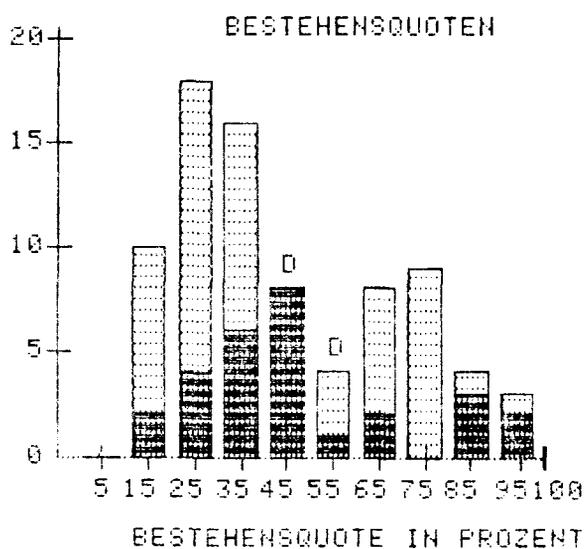


Abb. 4 Häufigkeitsverteilung der Bestehensquoten in allen Ländern der Welt. Die schraffierten Säulen stellen die Ergebnisse der Teilnehmer dar, die nicht Staatsangehörige der USA sind, die schwarzen Teile repräsentieren die Ergebnisse der US-Bürger in allen Ländern der Welt. Der Buchstaben 'D' markiert die Position der Teilnehmer, die in der Bundesrepublik studiert haben.

Abb. 5 Kumulative Häufigkeitsprozente im Wahrscheinlichkeitsnetz. Dargestellt sind die Werte für die US-Bürger (ausgefüllte Quadrate) und die Angehörigen der übrigen Staaten (Kreuze).

4. POSITION DER BUNDESREPUBLIK

Letztlich gewinnt die Frage an Bedeutung, wo sich die Ergebnisse der Teilnehmer einordnen lassen, die ihr Studium in der Bundesrepublik absolviert haben. Auf der Grundlage dieser aus 108 Ergebnissen erstellten Verteilung läßt sich die Position der Bundesrepublik bestimmen. Die Bestehensquote für die US-Bürger war 41,5 %, die der Nicht-US-Bürger für den medizinischen Prüfungsteil 53,3 %. Die Position ist durch den Buchstaben 'D' in der Abb. 4 angezeigt. Danach liegen die Ergebnisse der Teilnehmer aus der Bundesrepublik zwischen den beiden Populationen. Genauer ist sie für die US-Bürger eher mit "am obersten Rand der unteren Populationen" für die Nicht-US-Bürger eher am "unteren Rand der Spitzengruppe" zu beschreiben.

SCHLUSSBETRACHTUNG

Wie bereits einleitend einschränkend festgestellt wurde, läßt die Zuordnung einzelner Staaten keinen zwingenden Schluß auf die Qualität der Ausbildung zu. Außer dem Lehrplan und der Unterrichtsmethode, die in erster Linie die Gestaltung der Ausbildung bestimmen, wird ihr Ergebnis in sehr großem Umfang von anderen Faktoren beeinflusst. Dazu gehören Persönlichkeitsmerkmale der zum Studium zugelassenen Studenten und deren weitere Motivation, das Engagement der Hochschullehrer, das methodische und wissenschaftliche Niveau der ärztlichen Tätigkeit, die Verfügbarkeit von Hilfsmitteln usw.

Verbindliche Aussagen über die Qualität verschiedener Ausbildungssysteme lassen sich auf der Grundlage unserer Untersuchungen im internationalen Vergleich nicht machen. Das gilt besonders für die Position der Bundesrepublik, die sowohl für die US-Bürger als auch für die Angehörigen der übrigen Staaten an der Grenze zwischen den beiden Gruppen der Medizinschulen liegt. Die zunächst hypothetisch durchgeführte Zweiteilung der Medizinschulen kann jedoch die Richtung weiterer Untersuchungen bestimmen.

Literatur

EDUCATIONAL COMMISSION FOR FOREIGN MEDICAL GRADUATES
Results of 1977-1983 ECFMG Examinations,
Philadelphia

Prof. Dr. D. Habeck, Albert-Schweitzer-Str. 11
D 4000 Münster und Prof. Dr. H. E. Renschler,
Institut für Didaktik der Medizin, Sigmund-
Freud-Straße 25, D 5300 Bonn 1

BESPRECHUNGEN VON LEHRFILMEN

K. Schimmelpfennig, Berlin

Auf der Medikinale International 84 in Marburg wurden 140 Filme vorgestellt. Den Titeln der Filme und dem Rang der gewonnenen Medaillen ist nicht sicher anzusehen, welche Filme - als Videokassetten - für eine Studenten-Mediothek geeignet sind oder ob sie eher in die Weiter- bzw Fortbildung gehören.

Zahlreiche Firmen stiften diese und andere Filme - sehr großzügig - einer studentischen Mediothek.- Die Werbung hat in den vergangenen Jahren deutlich abgenommen, denn Reklame reduziert Medaillenchancen drastisch!

Wir wollen die Videoangebote besonders unter dem Aspekt des Selbststudiums von Medizinstudenten kritisch betrachten.

Hier stellen wir modellhaft eine Filmbesprechung in ausführlicher und auch in kurzer Fassung vor.- Derartige Besprechungen können eine Entscheidungshilfe für Anschaffungen sein, sie können aber auch Studenten ermuntern sich einen Film anzuschauen.

Wir bitten hiermit Kollegen, die in der Lehre tätig sind, Besprechungen von Film- und Ton-Dia-Angeboten für uns zu schreiben. Unsere Redaktion wird diese Besprechungen veröffentlichen, wenn ein habilitierter Dozent des jeweiligen Fachgebietes zustimmt. Falls genügend Einsendungen kommen, so werden wir ein Sonderheft " Mediothek " herausgeben.

Ausführliche Besprechung

der Video Cassetten "Das Ovarial-Karzinom Nr.I und II"

Hersteller:

Die beiden Cassetten wurden von der Firma Ciba-Geigy hergestellt. Für den Inhalt verantwortlich ist eine Onkologen-Gruppe in Berlin unter der Leitung von Herrn Chefarzt Dr.Imholz, die in Berlin eine multizentrische prospektive Studie über das Ovarial-Karzinom durchgeführt hat. Der Film wurde während des Ärztekongresses 1981 in Berlin hergestellt, und zwar im Rahmen einer Fortbildungsveranstaltung.

Zielgruppe:

Der Film richtet sich in erster Linie an Ärzte. Er ist auch geeignet zur Fortbildung von anderem medizinischen Personal sowie von Medizin-Studenten.

Lernziel:

Das Thema wird auf zwei Cassetten dargestellt. Länge der ersten Cassette 50 Min., Länge der zweiten Cassette 30 Min.. Auf der ersten Cassette wird die Diagnostik und die op. Therapie dargestellt, auf der zweiten Cassette die zytostatische Therapie sowie die Nachsorge. Diagnostik: Ambulante Diagnostik mit Erhebung und Anamnese, klin. Untersuchung und Ultraschalluntersuchung. Stationäre Diagnostik Röntgen-IV-Pylogramm, Kontrasteinlauf, Computer-Tomogramm, gegebenenfalls Laparoskopie oder auch explorative Laparotomie. Stadieneinteilung nach FIGO. Makroskopische und mikroskopische Morphologie. Histologisches Grading. Operative Therapie: Grundzüge der op. Therapie unter Mitnahme des großen Netzes..

Zytostatische Therapie: Prä- und Postoperative Zytostase. Remissionsraten. Nebenwirkungen. Laborüberwachung. "Second-Look"-Operation.
Nachsorge: Klinische Kontrollen. Röntgen-Thorax, Leber-Szintigramm, CT, event. IV-Polygramm. Soziale Hilfen. "Second-Look"-Operation.

Didaktik:

Die Wissensvermittlung ist eingebettet in Gespräche. Zum einen in Arzt - Patientengespräche, zum anderen in Gesprächsgruppen-Expertenrunde. Die Stadieneinteilung wird sehr übersichtlich und einprägsam anhand eines Schemas mit Hilfe einer Magnettafel dargestellt. Im Rahmen der röntgenologischen Diagnostik wird ein Computer-Tomogramm erläutert. Stellvertretend für die op. Therapie wird ein kurzer Streifen einer Laparotomie mit Längsschnitt bei einem großen Ovarialkystom eingeblendet. Die Morphologie wird anhand makroskopischer und mikroskopischer Präparate ausführlich erläutert. Auf der zweiten Cassette wird die zytostatische Therapie sowie die Nachsorge nur in Form von Gesprächen dargestellt.

Kritische Beurteilung:

Die inhaltliche Darstellung ist korrekt. Auch heute, 3 Jahre nach der Entstehung des Filmes, ist der Film auf dem neusten Stand. Die heute notwendige erweiterte präoperative und intraoperative Diagnostik wird gut herausgearbeitet (Ultraschall, Computer-Tomogramm, Stadieneinteilung). Insbesondere wird auch die heutige deutlich verbesserte therapeutische Möglichkeit der Chemo-Therapie beim Ovarial-Karzinom richtig gewichtet. Für Ärzte, zur deren Fortbildung dieser Film im wesentlichen gedreht worden ist, ist die Darstellung der Problematik des Ovarial-Karzinoms gut geeignet. Für Medizinstudenten ist sie nicht umfassend genug. So ist die klinische Symptomatik zu kurz dargestellt; es fehlen z.B. die Komplikationen wie die Stieldrehung des Ovarialtumors, die Ruptur des Kystoms oder auch die Verjauchung des Ovarial-

Tumors. Desgleichen fehlt die Darstellung der Spätkomplikationen beim Ovarial-Karzinom mit dem Hinweis auf die häufige Ileussyptomatik im Spätstadium. Andererseits sind die Pathologen im Film der üblichen Gefahr erlegen, ihr Gebiet zu breit darzustellen. Die op. Therapie wird für die Studenten zu knapp geschildert. Hier fehlen für ihn schematische Darstellungen. So geht z.B. die Wichtigkeit der obligaten Netzresektion bei der verbalen Darstellung unter. Von der didaktischen Darstellung her, wirken die Patienten-Arzt-Gespräche zum Teil wegen ihrer Länge ermüdend. Schon aus diesem Grunde wären abschnittsweise Zusammenfassungen wünschenswert gewesen. Uneingeschränkt zu loben sind die schematischen Darstellungen an der Magnettafel.

Die beiden Cassetten Ovarial-Karzinom I und II sind mit Einschränkung für den Unterricht von Medizinstudenten geeignet

Prof. Dr. W. Busch

Univ.-Frauenklinik und Poliklinik
Berlin - Charlottenburg

Kurze Besprechung

der Video-Cassetten "Das Ovarial-Karzinom Nr. I und II"

Hersteller: Ciba-Geigy mit Onkologengruppe um Dr. Imholz, Berlin 1981.

Zielgruppe: Ärzte in Fortbildung. Medizinstudenten mit Einschränkung.

Lernziel: Erweiterte Diagnostik, operative und zytostatische Therapie mit Nachsorge.

Kritische Beurteilung: Der Film ist heute noch auf dem neusten Stand. Für Ärzte ist die Darstellung der Problematik gut geeignet. Für Medizinstudenten ist die klin. Symptomatik mit möglichen Komplikationen zu kurz oder lückenhaft dargestellt. Auch die operative Therapie ist zu knapp, zudem fehlen hier schematische Darstellungen. Aber die sonstigen

schematischen Darstellungen an der Magnet-
tafel sind gut.

Prof. Dr. Schimmelpfennig

Ausführliche Besprechung
der Video-Cassetten 349 A, B, C, "Die
natürliche Geburt"

Hersteller:

Die vorliegenden drei Cassetten wurden
im Auftrage des NDR hergestellt. Die
Filme wurden 1980 im allgemeinen Fern-
sehen gesendet. Die Produzenten sind
"Wissenschafts-Journalisten", - also
keine Mediziner.

Zielgruppe:

Die Produzenten richten sich an die all-
gemeine Bevölkerung, also an Laien.
Die Filme sind nicht hergestellt zur
Information von Medizinstudenten oder
Ärzten.

Lernziel:

Anhand der "Natürlichen Geburt" soll die
Gefahr der Inhumanisierung der sogenann-
ten Apparatmedizin bewußt gemacht werden.
Verschiedene Beispiele natürlicher Ge-
burten sollen zeigen, wie es möglich sein
kann, die Geburt nicht als das Leben und
die Gesundheit von Mutter und Kind be-
drohenden Akt zu sehen, sondern primär
als ein normales wenn auch außergewöhn-
lich intensives Erlebnis im Lebensablauf
der Frau.

Lerninhalt:

Auf den Cassetten 394 A und B werden
unterschiedliche Geburtsszenen darge-
stellt; - auf der Cassette 394 C wird
eine Diskussionsrunde von Experten ver-
schiedener Disziplinen zu diesem Film
gezeigt.

Cassette 394 A:

Auf der Cassette sind zwei Hauptteile
zu unterscheiden:

1. Geburtsszenen aus einer französ-
ischen Klinik,

2. Hausentbindung in West - Berlin

Länge des 1. Teiles: 24. Min.; Länge des
2. Teiles: 18 Min.

Stichwortartiger Inhalt von Teil 1

(französische Klinik):

Austreibungsphase der Geburt. Erster
Mutter-Kind-Kontakt. Baden des Kindes.
Abnabeln. Nachgeburt. Eröffnungsphasen,
einmal durch Spaziergehen im Klinik-
garten oder auf dem Klinikflur, zum
anderen Entspannung durch Schmerzverar-
beitung im Bad. Gruppengespräche von
Schwangeren und Entbundenen in der Vor-
sorge. Geburt im Wasser.

Stichwortartiger Inhalt von Teil 2

(Hausentbindung in Berlin):

Vorsorgeuntersuchung durch die Hebamme
in eigener Wohnung. Austreibungsphase
im Kreise der Familie. Abnabeln, Erst-
untersuchung des Neugeborenen, Stillen,
Wochenbettbesuch. Hausentbindung in
Holland durch Hebamme und Wochenbett-
pflgererin.

Cassette 394 B:

Auch diese Cassette hat zwei große Haupt-
teile, wobei der zweite Teil identisch
ist mit dem zweiten Teil der Cassette
394 A (Hausgeburt in Berlin). Im
ersten Teil der Cassette wird abwechselnd
eine Geburt in Dachau und eine Geburt in
Hamburg dargestellt. Die Geburt in Dachau
steht repräsentativ für den Versuch der
Symbiose einer natürlichen Geburt und
den modernen Überwachungsmöglichkeiten
des Kindes. Die Geburt in Hamburg soll
beispielhaft sein für eine programmierte
Geburt. Länge des 1. Teiles 28 Min.

Cassette 394 C:

Von Experten wird in einer Live-Fernseh-
runde das Für und Wider der modernen
Geburtsmedizin in Gegenüberstellung zur
natürlichen Geburt diskutiert. Teil-
nehmer sind: Geburtshelfer, Kinderarzt,
Neurologe, Psychiater, Hebamme, Mütter.

Kritische Beurteilung:

Die Filme sind nicht geeignet, Medizin-
Studenten ein Wissen über den Geburtsab-

lauf zu vermitteln. Der Student kann an den Filmen einen Eindruck gewinnen, über die Auswirkung der Trias Angst-Spannung-Schmerz auf die Geburt. Er kann auch Lernen, wie diese Reaktionskette durchbrochen werden kann. Hauptanliegen der Filme ist jedoch die Anprangerung der Apparatemedizin in der Geburtshilfe und die Verherrlichung der natürlichen Geburt. Dies geschieht nicht objektiv. Die Kommentare sind tendenziös und häufig dazu angelegt, um aggressive Emotionen zu schüren. Dazu werden auch filmerische Mittel benutzt, Hintergrundgeräusche, Farbgebungen, Licht- und Schattenverteilung. Einiges wird nicht korrekt wiedergegeben. So wird im Zusammenhang mit der programmierten Geburt behauptet, daß "viele" Frauen nach einer Periduralanästhesie über "Wochen" an Kopfschmerzen leiden. Richtig ist, daß bis zu 1% nach Periduralanästhesie Kopfschmerzen auftreten können und dies nicht länger als drei Tage. Die Filme sind daher am ehesten noch für geburtshilfliches Fachpersonal geeignet. Sie können für Hebammen und Geburtshelfer einen Anstoß darstellen, ihr tägliches Tun zu überdenken. Nur der Erfahrene wird sich von der tendenziösen Darstellung der Filme freimachen können.

Prof. Dr. W. Busch
Univ. Frauenklinik und Poliklinik
Berlin - Charlottenburg

Kurze Besprechung

der Video-Cassetten 394 A, B, C "Die natürliche Geburt"

Hersteller: "Wissenschafts-Journalisten"
-keine Mediziner- im Auftrag des NDR.

Zielgruppe: Laien.

Lernziel: Die Gefahr der Inhumanisierung durch die sogenannte Apparatemedizin soll bewußt gemacht werden.

Lerninhalt: Es werden Geburtsszenen dargestellt, anschließend nimmt eine Expertenrunde dazu Stellung.

Kritische Beurteilung: Hauptanliegen ist die Anprangerung der Apparatemedizin in der Geburtshilfe und die Verherrlichung der natürlichen Geburt. Die tendenziösen Kommentare schüren aggressive Emotionen, die mit Hintergrundgeräuschen sowie Farb- Licht- und Schattengebung untermauert werden. Falsch sind Aussagen über die Periduralanästhesie. Nur der Erfahrene kann sich mit dem Film auseinandersetzen. Für Medizinstudenten wird kein Wissen über den Geburtsablauf vermittelt.

Prof. Dr. med. K. Schimmelpfennig
Institut für Toxikologie der F.U., Garystr. 9,
D 1000 Berlin-33

RESOLUTION DER GESELLSCHAFT FÜR MEDIZINISCHE AUSBILDUNG ZU STUDIENAUFENTHALTEN DEUTSCHER MEDIZINSTUDENTEN IM AUSLAND

Die Gesellschaft für Medizinische Ausbildung (Deutsche Sektion der Association for Medical Education in Europe) hat sich während ihrer Arbeitssitzungen am 21. Juni 1984 in Homburg/Saar und am 30. November 1984 in Köln ausführlich mit zeitweiligen Studienaufenthalten deutscher Medizinstudenten in ausländischen Staaten befaßt. Die Gesellschaft begrüßt das in den letzten Jahren angestiegene Interesse für Studienaufenthalte im Ausland als eine sehr wünschenswerte Entwicklung, da hierdurch der allgemeine Horizont, die persönliche Urteilskraft und die medizinische Erfahrung der künftigen Ärzte eine wesentliche Erweiterung und Bereicherung erfahren. Im Bereich der Europäischen Gemeinschaft fördern sie darüber hinaus die europäische Integration.

Wichtigste Voraussetzungen für einen erfolgreichen Studienaufenthalt bilden einschlägige Sprachkenntnisse und Vertrautheit mit den wichtigsten historischen, kulturellen und gesellschaftlichen Gegebenheiten des Gastlandes sowie eine Beherrschung der klinisch-theoretischen und klinisch-praktischen Grundlagen, die bis zum Abschluß des Ersten Abschnitts der Ärztlichen Prüfung und durch eine mindestens einmonatige patientennahe Famulatur gewonnen werden können.

Bei Vorliegen dieser Voraussetzungen empfiehlt die Gesellschaft nachdrücklich eine verstärkte Förderung von Studienaufenthalten. Sie appelliert

- an die medizinischen Fakultäten und einzelne Professoren, bestehende Auslandskontakte zu diesem Zweck zu vertiefen und zu erweitern,
- an den DAAD, Studienstiftungen und entsprechende andere Einrichtungen, Auslandsstudienaufenthalte vermehrt zu fördern,
- an den Gesetzgeber und die Landesprüfungsämter Hürden, die einer Anerkennung von im Ausland erbrachten Studienleistungen im Wege stehen, abzubauen.

MITTEILUNGEN

Vom 10. bis 12. September 1985 findet in Jerusalem in der Van Leer Jerusalem Foundation (46 Jabotinske Street) die diesjährige **Jahreskonferenz der Association for Medical Education in Europe** statt. Das Thema der Konferenz lautet: Tradition and Change in Medical Education. Die Teilnahmegebühr beträgt 190 US Dollar, für Studenten 100 US Dollar (mit Mahlzeiten) bzw. 55 Dollar (ohne Mahlzeiten). Nähere Einzelheiten können beim Geschäftsführer unserer Gesellschaft, Herrn Priv.Doz. Dr. E. Doppelfeld, Bundesärztekammer, Hasdenkampstr. 1, 5000 Köln 41, erfragt werden.

BILD(UNG) UND MEDIZIN

ZUM TITELBILD: DR. MED. MARIE HEIM-VÖGTLIN (1845 - 1916) MIT DEM
PFLEGEKIND HANNELI, 1892

Christa Lange und Johanna Geyer-Kordesch, Münster

Ich denke viel an meine Zukunft;... Am liebsten denke ich mich immer in die Berge hinein, - aber ich denke manchmal, wenn ich Aussicht hätte, in einer Stadt viel Arbeit zu bekommen - so wäre es vielleicht doch besser, dies zu wählen - nicht um der Annehmlichkeit willen, sondern weil ich so sehr wünsche, das Prinzip, dass Frauen zu diesem Berufe tüchtig sind, etabliert zu sehen-; verberge ich mich an einem ganz abgeschlossenen Ort, so werde ich bald verschollen sein. Hätte ich nur an mein eigenes Leben zu denken,... - aber ich wünsche für andere die Bahn zu brechen - wird mir das gelingen?"

Der Weg Marie Heim-Vögtlins war der einer Pionierin für das Medizinstudium der Frau. Sie war die erste schweizerische Medizinstudentin, die erste Schweizerin, die den Dr. med. erwarb und eine Praxis eröffnete. Dabei wurde ihr dieser Weg durch viele Hindernisse erschwert. Mit 23 Jahren "ertrotzte" sie sich die Erlaubnis, das Medizinstudium aufzunehmen und immatrikulierte sich 1868 in Zürich.

"Es braucht wohl die Überzeugung, dass mein ganzes Leben dran hängt, um mir in diesen Wirrsalen und Ungewissheiten immer den Mut aufrecht zu erhalten".

Und sie behielt den Mut; nach erfolgreich abgeschlossenem Studium und beschwerlichen Assistenzarztzeiten in Leipzig und Dresden ließ sie sich 1874 in Zürich als erste Ärztin der Schweiz nieder. Im darauffolgenden Jahr heiratete sie mit 29 Jahren den Geologieprofessor Albert Heim. Marie Heim-Vögtlin führte ein aussergewöhnlich erfülltes Leben als hochgeschätzte und vielgesuchte Ärztin - insbesondere für Frauenheilkunde und Geburtshilfe - als Ehefrau und Mutter zweier Kinder sowie eines Pflegekindes. Ihr soziales Engagement war herausragend,

nicht nur auf direkter caritativer Ebene, 1901 gründete sie mit ihrer Kollegin Anna Heer die Pflegerinnenschule in Zürich und leitete die dortige Kinderabteilung.

Die erste Ärztin der Schweiz war keine politische Agitatorin, sie verstand es, durch ihr gelebtes Beispiel in ihrer Zeit den Begriff Ärztin denkbar zu machen, ein Vorbild zu schaffen für künftige Ärztinnengenerationen.

"Das eigene Leben hat Sinn, wenn es als Baustein gilt für die kommenden Geschlechter".

(Zitate aus: Johanna Siebel, Das Leben von Frau Dr. Marie Heim-Vögtlin der ersten Schweizer Ärztin, Zürich, 1920, S. 76; S. 55; S. 149).

Christa Lange und Johanna Geyer-Kordesch
Institut für Theorie und Geschichte der
Medizin, Waldeyerstr. 27, D 4400 Münster

Sehr geehrte Leserin,
sehr geehrter Leser,

um Anhaltspunkte für die künftige Auflagenhöhe zu erhalten, bitte ich Interessenten an dieser Zeitschrift um Mitteilung von Abonnementswünschen für die nächsten Hefte.

D. Habeck

Herrn Prof. Dr. Habeck
Albert-Schweitzer-Str. 11
D 4400 Münster

Hiermit bestelle ich die nächsten Hefte der "Medizinischen Ausbildung".

Absender:

