

BENOTUNG DES ERFOLGES DER ÄRZTLICHEN AUSBILDUNG

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

Unter den durch Bafög geförderten Studenten müssen ab 01.01.1984 die ersten 30 % ermittelt werden, die in einem Kalenderjahr die ärztliche Prüfung ablegen (§18 b, Bafög v. 08.06.1983). Mit der 4. Verordnung zur Änderung der Approbationsordnung soll ein Verfahren eingeführt werden, nach dem die Ergebnisse der 3 Abschnitte der ärztlichen Prüfung jeweils getrennt benotet werden. Aus dem auf eine Hundertstel-Note errechneten Mittelwert wird eine Rangfolge der Absolventen erstellt, aus der die geforderten ersten 30 % abgezählt werden

Testtheoretische Grundlagen

Zur Zeit werden in der Bundesrepublik die ärztlichen Prüfungen noch nach dem Verfahren der klassischen Testtheorien gestaltet und ausgewertet. Dabei werden die Kandidaten im Bereich der Bestehensgrenze, die nach politischen Gesichtspunkten bei etwa 2 - 10 % der schlechtesten Kandidaten festgelegt wird, in möglichst kleine, aber gleichgroße Klassen mit unterscheidbarer Anzahl richtig gelöster Aufgaben eingeordnet.

Da die klassische Testtheorie eine Rangordnung der Kandidaten erstellt, ist sie für die geforderte Aufgabe geeignet. Als ungünstig erweist sich

allerdings, daß die Grenze zu den besten 30 % in dem Bereich liegt, in dem die einzelnen Leistungsklassen am größten sind und die größten Unterschiede in der Besetzung aufweisen. Kleine Ungenauigkeiten des Messens und der Berechnung haben daher große absolute und relative Auswirkungen. Der sich daraus in der Praxis ergebende Meßfehler berechtigt nicht die Einteilung der erfolgreichen Kandidaten in 400 Klassen, wie sie sich aus der Vergabe von Hundertstel-Noten ergibt.

Notengebung

Für die Umwandlung der Anzahl richtig gelöster Aufgaben in Noten sieht die klassische Testtheorie mehrere Verfahren vor. Diese müssen sich am Mittelwert der Rohwertverteilung orientieren. Die Standardabweichung wird dem Maß für die Notenskala zugrundegelegt. Bei den Prüfungen des National Board of Examiners in den USA erhält z.B. der Mittelwert den Wert 500, für jede Standardabweichung wird der Wert 100 addiert oder subtrahiert. Bei dem in Deutschland am stärksten verbreiteten Verfahren der Standardschulnoten mit den Noten 1 - 5 umfaßt die mittlere Note 3 den Mittelwert mit jeweils einer halben Standardabweichung nach oben und unten. Die Noten 2 und 4 umfassen den Bereich von 0,5 bis 1,5, Standardabweichung

chungen jeweils wieder in beiden Richtungen. Diese Rohwertverteilung und Notengebung sind in Abb. 1 zusammen mit den wichtigsten weiteren Verfahren dargestellt. In der deutschen pädagogischen Literatur werden die unterschiedlichen Bezüge zwischen der Häufigkeitsverteilung der Rohwerte und der Notengebung diskutiert (RAPP, HELLER). Häufig liegt bei den Rohwerten des IMPP keine Normalverteilung vor, da die Bemühungen um ein gutes Prüfungsergebnis eine asymmetrische Verschiebung zu höheren Punktzahlen bewirken. In solchen Fällen kann die Benotung nach der kumulativen Häufigkeit erfolgen.

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, MADDAUS).

Verwendet man das Verfahren der Standard-Schulnoten nach der Z-Skala (LIENERT), so erhalten bei Normalverteilung der Rohwerte 6,7 % der Kandidaten die Note 1 und auch genauso viele die Note 5. Eine Überprüfung an den Ergebnissen der letzten vier Prüfungstermine des 2. Abschnittes der ärztlichen Prüfung er-

gab, daß im Mittel 5,3 % der Teilnehmer die Note "Sehr gut" erhalten hätten (Standardabweichung 1,3 %). Nach dem im Entwurf der 4. Novelle vorgesehenen Verfahren würden nur 0,16 % (Standardabweichung 0.16 %) die Note 1 erhalten.

Gesamtnote

Da sich die Bewertung des nach der klassischen Testtheorie ermittelten Prüfungsergebnisses für jeden einzelnen nach der Verteilung in der jeweiligen Prüfungsgruppe ausrichtet, muß dieses Verhältnis besonders dann berücksichtigt werden, wenn aus mehreren Teilprüfungen ein gewichteter Mittelwert gebildet werden soll. Ohne diese Standardisierung der Noten ist eine Mittelwertbildung methodisch falsch. Dieser Fehler hat dieselbe Bedeutung wie ein Fehler in der Anwendung des kleinen Einmaleins oder der Rechtschreibung. Der Unterschied besteht nur in der unterschiedlichen Einsicht in die Grundlagen. Diese sollten bei Politikern und Sozialwissenschaftlern gegeben sein. An jedem Wahltag wird der Wert stochastischer Rechenverfahren bewiesen. Schon zwei Stunden nach Schluß der Wahlurnen weicht die Vorhersage des Endergebnisses nur um Zehntel-Prozente mit einer Standardabweichung von weniger

als 1 % vom amtlichen Endergebnis ab und dieses wird schon in weniger als 3 Stunden exakt vorhergesagt.

Gültigkeit der Prüfungsergebnisse

Die Gesamtnote kann nur dann als gültig für die Bewertung des Studienerfolges angesehen werden, wenn sie den gesamten Inhalt und alle Ziele der Ausbildung zum Arzt erfasst. Eine Beurteilung der Kompetenz für die ärztliche Tätigkeit, die davon zu unterscheiden ist, ist mit einer ausschließlich nach den Regeln der klassischen Testtheorie gestalteten Prüfung ohnehin nicht möglich.

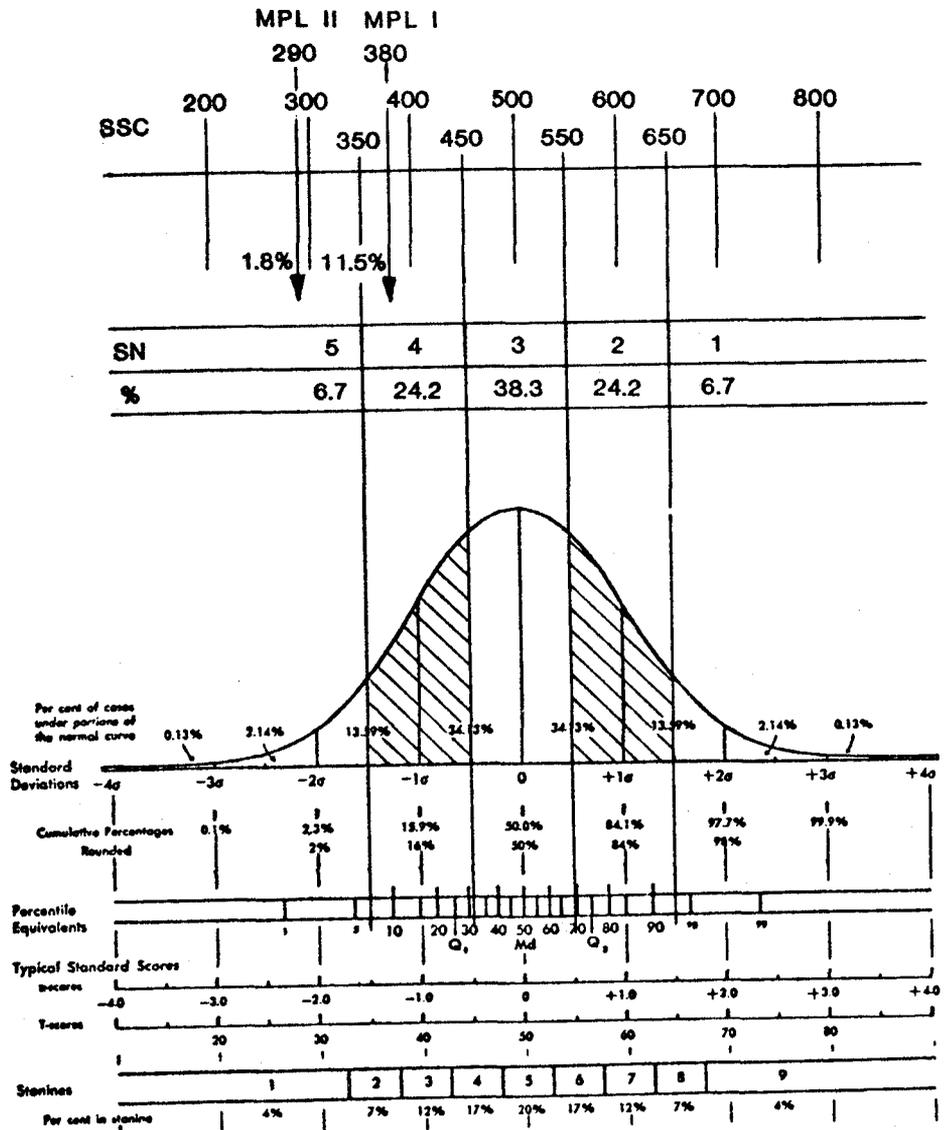
Da nur etwa 20 - 40 % der Ausbildungsziele des Medizinstudiums mit den Fragen mit Auswahlantworten erfaßt werden können, ist ein Einbezug weiterer Prüfungsleistungen in angemessenem Umfang erforderlich. Dazu gehört außer dem praktischen Können und dem Gestalten zwischenmenschlicher Beziehungen auch die Methodik der Wissenschaft. Minister Geißler hat in einer Fernsehdiskussion am 06.05.83 gefordert, daß die bisher überproportionale Gewichtung von Fragen mit Auswahlantworten zugunsten anderer, mindestens ebenso wichtiger Verfahren verringert werden sollte. Nach der im

Entwurf vorgesehenen Gewichtung der mündlichen Prüfung mit 20 % ist bei durchschnittlichen Ergebnissen in den schriftlichen Teilprüfungen (entsprechend der Note 3) eine Verbesserung der Gesamtnote, selbst durch hervorragende Leistungen im mündlichen Teil des 3. Abschnittes nicht möglich. Durch die beim Erlaß vom 19.12.1983 erfolgte größere Gewichtung der mündlichen Prüfung mit 25 % ist dies auch nur dann möglich, wenn mindestens ein schriftlicher Teil mit der Note 2 oder besser bestanden wird. Ein solch gutes Ergebnis ist bei etwa einem Sechstel der Kandidaten zu erwarten.

Bewertung der wissenschaftlichen Grundlagen der ärztlichen Ausbildung

Um der Forderung nach einer Einbeziehung des Erwerbs von Arbeitsmethoden in die Gesamtnote zu entsprechen, sollten die Ergebnisse der Vorprüfung einbezogen werden. Da die Vorprüfung eine Universitätsprüfung zum Steuern des Studienfortschrittes und keine berufsqualifizierende Prüfung nach der Bundesärzterordnung ist, würde dies einen stärkeren Einbezug der Universitäten bedeuten (SCHIMMELPFENNIG). Außerdem würde damit dem drohenden Abbau von Grundlagenforschung an den medizinischen Einrichtungen vorgebeugt werden.

Beziehung zwischen Rohwertverteilung und Notengebung



SN : Standardschulnoten(nach LIENERT 1969)

SSC : Standard Scores (NBME)

MPL I,II: minimum passing score
Partl, Partil u. III

Wiedergegeben sind verschiedene Verfahren der Notengebung in Bezug zu der als Normalverteilung dargestellten Rohwertverteilung. Auf der obersten Skala (SSC), die vom National Board of Examiners benutzt wird, sind die zum Bestehen erforderlichen Werte für den Teil 1 (MPL I) sowie für den Teil 2 und 3 (MPL II) und beispielsweise zwei Durchfallsquoten eingetragen. SN stellt das in Deutschland am häufigsten verwendete Verfahren der Standard-

schulnoten sowie die auf die 5 Noten fallenden prozentualen Anteile beim Vorliegen einer Normalverteilung dar. Im unteren Teil der Abbildung sind verschiedene, besonders in Nordamerika verwendete Skalen dargestellt. Von Bedeutung ist die Prozentrangskala Percentile Equivalents, die bei 20 Skalenschritten von jeweils 5 Einheiten eine hinreichend differenzierte Beurteilung ermöglicht.

Die deutsche Medizin ist in zunehmendem Umfang auf die Übernahme neuer Verfahren aus dem Ausland angewiesen. Diese dürfen nicht unkritisch übernommen werden, sondern müssen vor der Einführung einer wissenschaftlich begründeten Analyse und Anpassung unterzogen werden. Es erscheint daher sinnvoll, eine schriftlich wissenschaftliche Arbeit, wie dies für die Diplomprüfungen anderer Fächer üblich ist, oder auch die Note der Dissertation in die Benotung miteinzubeziehen. Trotz der in Deutschland schon im Jahre 1685 eingeführten Nachkontrolle der Universitätsprüfung durch den Staat wird ein Einbezug einer von der Universität erteilten Teilnote in die Gesamtnote empfohlen. In England erbringt auch heute noch die Prüfung, die auf die ärztliche Kompetenz und nicht auf den alleinigen Besitz von Wissen ausgerichtet ist, durch die Universität oder die Standesorganisationen allein die Qualifikation für die ärztliche Berufstätigkeit. In Nordamerika ist das Bestehen der Prüfung durch die von den Standesorganisationen American Medical Association und Association of American Medical Colleges anerkannten Medizinschulen Voraussetzung für die Zulassung zu den Staatsprüfungen. Dabei gewinnt in den letzten Jahren zunehmend eine auf die ärztliche Kompetenz ausgerichtet neue Prüfungsform (FLEX) gegenüber der den Studienerfolg messenden des National Board of Examiners an Bedeutung.

Auswirkungen der Notengebung

Über die Auswirkungen auf das Lernen der Studenten durch diese Änderungen der Prüfung und den zu erwartenden Einfluß der Gesamtnote auf die Beschäftigungschancen in der Zukunft lassen sich nur Vermutungen anstellen. Unsere Untersuchungen haben gezeigt, daß das Bild, das sich Bildungspolitiker und auch Hochschullehrer über das Lernen der jetzigen Studentengeneration machen nicht mit der Wirklichkeit übereinstimmt (BURKHARD, RENSCHLER und Mitarbeiter). Aber auch die von uns aus einer Studentenbefragung abgeleiteten Aussagen müssen durch unmittelbare Beobachtung des Verhaltens der Studierenden ergänzt werden.

Literatur:

- BLOOM, B.S., HASTINGS, J.T.,
MADAUS, G.F.:
Handbook an Formative and Summative Evaluation of Student Learning
New York, McGraw-Hill 1971
- HELLER, K.
Leistungsbeurteilung in der Schule
Heidelberg, Quelle & Meyer 1978

KLAUER, K.J.

Ein kriteriumsorientiertes Zensierungsmodell

Z Entwicklungspsychol. Pädag. Psychol.
14, 65-79

LIENERT, G.A.

Testaufbau und Testanalyse
Weinheim, Beltz 1969

RAPP, G.

Messung und Evaluierung von Lernergebnissen in der Schule
Bad Heilbrunn/Obb., Klinkhardt 1975

SCHIMMELPFENNIG, K.

Ein Vorschlag zur gesetzlichen Neuregelung
der ärztlichen Vorprüfung
(des sogenannten "Physikums")
Protokoll des Medizinischen Ordentlichen Fakultätentages
Göttingen 1983, S. 142-146

Prof. Dr. H.E. Renschler, Institut für Didaktik der
Medizin der Medizinischen Fakultät, Sig-
mund-Freudstr. 25, D 5300 Bonn 1