

## » Vergleich von Prüfungsmethoden in der medizinischen Ausbildung

Andrea Elmer<sup>1</sup>, J. Grifka<sup>2</sup>

<sup>1</sup> Abteilung Pneumologie, St. Hildegardis Krankenhaus, Mainz

<sup>2</sup> Orthopädische Klinik der Ruhr-Universität Bochum

**Zusammenfassung:** Prüfungen in der medizinischen Ausbildung unterliegen als Meßinstrumente den Testgütekriterien der Objektivität, Reliabilität und Validität. In der vorliegenden Arbeit werden schriftliche, mündliche und praktische Prüfungsverfahren vorgestellt und hinsichtlich ihrer Testgütekriterien miteinander verglichen. Es zeigt sich, daß mündliche und praktische Prüfungen einer einheitlichen Strukturierung bedürfen, um eine ausreichende Reliabilität und Objektivität zu erreichen. Multiple-Choice Fragen als eine weit verbreitete Form schriftlicher Prüfungen sind hinsichtlich ihrer Validität kritisch zu beurteilen. In verschiedenen Untersuchungen zur Qualität von Prüfungsverfahren konnte gezeigt werden, daß ein Prüfungssystem eine Kombination aus mündlichen, schriftlichen und praktischen Prüfungen darstellen muß, wenn es den genannten Testgütekriterien genügen will.

**Examinations in Medical Education:** Examinations in medical education, as testing-methods of clinical competence, must fulfil criteria of objectivity, reliability and validity. This survey compares popular written, oral and practical tests of knowledge with special reference to their testing quality. It can be shown that oral and practical tests should be standardized to improve objectivity and reliability. Multiple-choice examinations, which are widely used written tests of knowledge, have to be examined in detail. Several studies were done to examine the quality of these different testing-methods. The results strongly suggest that reliable and valid achievement tests in medical education should consist of a combination of oral, written and practical testing-methods.

**Key-words:** Medical education – Validity – Educational measurement/methods – Multiple choice questions

Die medizinische Ausbildung ist in den letzten Jahren verstärkt in das Bewußtsein der Öffentlichkeit gerückt und wurde zum Gegenstand kontroverser Diskussionen. In besonderer Weise ist davon auch das Prüfungssystem in der medizinischen Ausbildung betroffen, das vielfach Veränderungsprozessen und Reformbestrebungen unterworfen ist. Im vorliegenden Beitrag soll einführend kurz auf die Funktionen von Prüfungen eingegangen werden, die diese im Rahmen der Mediziner Ausbildung erfüllen. Es folgt eine Darstellung der

Testgütekriterien, denen Prüfungen als Meßinstrumente genügen sollten und an denen sich die Qualität von Prüfungen bemißt. In einer Übersicht werden die verschiedenen Prüfungsverfahren dargestellt und in ihren wichtigsten Vor- und Nachteilen miteinander verglichen. Im Anschluß folgt eine kurze Zusammenstellung der Ergebnisse einer eigenen Untersuchung, in der zwei schriftliche Prüfungsverfahren miteinander verglichen werden.

### Funktionen von Prüfungen

Die gegenwärtig stattfindende Diskussion zum Prüfungsweisen in der medizinischen Ausbildung ist keineswegs neu, vielmehr besteht sie ebenso lange, wie über Inhalt und Form der Ausbildung selbst nachgedacht wird. Prüfungen erfüllen wichtige didaktische und gesellschaftliche Funktionen [32,33], an denen sich die Konzeption von Prüfungen ausrichtet.

Trost [32] betont die besondere Bedeutung von Hochschulprüfungen für Studierende, Hochschulen und Gesellschaft. Die Funktionen von Hochschulexamina lassen sich somit in drei Punkten zusammenfassen: Sie sollen die Studierenden über ihren Leistungsstand informieren, den Fakultäten Rückmeldung über die Effektivität ihrer Ausbildung geben und gegenüber der Gesellschaft einen gewissen Grad fachlicher Kompetenz garantieren, indem bestimmtes Basiswissen und -können sichergestellt wird.

Die Aufteilung in didaktische und gesellschaftliche Funktionen von Prüfungen wird von Van den Bussche [33] weiter differenziert. In ihrer didaktischen Funktion steuern Prüfungen die Lehre und das Lernen, indem Dozenten ihren Lehrplan und Studierende ihre Lernstrategien entsprechend ausrichten; für beide Gruppen haben Prüfungen Rückkopplungsfunktion. Auf gesellschaftlicher Ebene schaffen Prüfungen die Voraussetzungen für einen Kompetenznachweis und erfüllen somit die Funktion der Selektion und Statuszuweisung. Der enge Zusammenhang zwischen Prüfungsordnung und Status der einzelnen Disziplinen innerhalb der Fakultäten verdeutlicht die Funktion von Prüfungen im Sinne der Fächerlegitimation.

Die Bedeutung von Prüfungen und Examina reicht über den individuellen und universitären Bereich hinaus und ist auch gesellschaftlich von direkter und praktischer Relevanz. Deshalb darf die Forderung als berechtigt gelten, daß Prüfungen bestimmten Anforderungen hinsichtlich ihrer Qualität als Meßinstrumente genügen müssen.

### Testgütekriterien von Prüfungen

Als zentrale Gütekriterien von Prüfungen gelten Objektivität, Zuverlässigkeit bzw. Reliabilität und Validität [3,13,14,24,28]. Eine Prüfung hat dann einen hohen Grad an Objektivität, wenn zwei voneinander unabhängige Prüfer zu demselben Ergebnis kommen. Der subjektive Einfluß des Prüfenden soll in Durchführung und Auswertung der Prüfung möglichst klein gehalten werden. Reliabilität als ein Maß für die Meßgenauigkeit einer Prüfung läßt sich einerseits definieren als Stabilität der Ergebnisse, d.h. als Übereinstimmung der Ergebnisse desselben Kandidaten bei wiederholter Bearbeitung derselben Prüfung, andererseits als Paralleltest-Reliabilität, d.h. als Übereinstimmung der Ergebnisse desselben Kandidaten in verschiedenen Prüfungsversionen (inter-item reliability) [32].

Validität bezeichnet die Gültigkeit eines Prüfungsverfahrens. Die inhaltliche Gültigkeit oder Kontentvalidität einer Prüfung ist dann besonders hoch, wenn der Prüfungsstoff diejenigen Kenntnisse, Fähigkeiten oder Fertigkeiten stichprobenartig erfaßt, die in der Ausbildung vermittelt werden sollen. Nach Klauer [12,13] kann in diesem Zusammenhang auch von Lehrzielgültigkeit gesprochen werden. In der englischsprachigen Literatur wird zwischen drei Aspekten von Validität unterschieden [26,28,32,33]: "Content validity" entspricht der Kontentvalidität und bezeichnet das Maß, in dem eine Prüfung das mißt, was zuvor in der Ausbildung als Lehrziel galt und vermittelt wurde. Die Konstruktion einer Prüfung setzt hier einen genau definierten Lehrzielkatalog voraus. "Criterion validity" bezeichnet die Übereinstimmung der Ergebnisse einer Einzelprüfung mit den Ergebnissen der Gesamtprüfung. Sie bezeichnet die Gültigkeit, die einer Einzelprüfung innerhalb eines konstruierten Prüfungssystems zukommt. Unter "construct validity" versteht man die Gültigkeit, mit der ein Test die Fähigkeit oder Fertigkeit mißt, die er seiner Konstruktion zufolge zu messen vorgibt. Als weitere Testgütekriterien gelten Praktikabilität, Aufwand und Kosten eines Prüfungsverfahrens im Sinne einer ökonomischen Kosten-Nutzen-Analyse [24,32,33]. Angesichts hoher Studierendenzahlen und begrenzter finanzieller Möglichkeiten auf Seiten der Universitäten kommt diesem Kriterium eine nicht zu unterschätzende Bedeutung zu.

### Verschiedene Prüfungsmethoden

Prüfungsverfahren lassen sich einteilen in schriftliche, mündliche und praktische Prüfungen. In Deutschland kommen im Rahmen der medizinischen Ausbildung vor allem Multiple-Choice-Prüfungen, mündliche und praktische Examensprüfungen zur Anwendung. Diese gängigen Verfahren wurden in den letzten Jahren zunehmend durch neu entwickelte Modelle ergänzt (Modified Essay Question, Triple Jump, Objective Structured Clinical Examination). Einige dieser neuen Prüfungstechniken sind besonders im anglo-amerikanischen Raum vielfach bereits fest in das Prüfungssystem integriert, finden aber auch vermehrt in deutschen Prüfungssystemen Anwendung.

#### Mündliche Prüfungen

Mündliche Prüfungen haben in der medizinischen Ausbildung eine lange Tradition: vor Einführung der Approbationsord-

nung im Jahr 1970 waren alle Staatsprüfungen mündliche Prüfungen. Sie ermöglichen dem Prüfer einen Einblick in den Denkprozeß des Kandidaten; ein wesentlicher Vorzug von mündlichen Prüfungen liegt in der direkten Rückmeldung. Herkömmliche, unstrukturierte mündliche Prüfungen zeichnen sich jedoch aus durch einen Mangel an Objektivität und Reliabilität, bedingt durch den subjektiven Einfluß des Prüfenden. Um diesem Mangel entgegenzuwirken, wurden Modelle für strukturierte und wiederholbare Prüfungen entwickelt, so die standardisierte mündliche Prüfung. Diese Prüfung verläuft nach einem vorgegeben Schema; die Leistungen des Kandidaten werden anhand einer vorgegebenen Bewertungsskala durch den Prüfer beurteilt [24]. Eine andere Form einer strukturierten mündlichen Prüfung ist der Triple Jump. Diese in drei Schritten verlaufende Prüfung (Problemstellung und Hypothesengenerierung, Lernphase, Auswertung) hat eine höhere Reliabilität als die unstrukturierte mündliche Prüfung [30] und ist gekennzeichnet durch eine hohe Kontentvalidität [24], da sie den klinischen Problemlösungsprozeß testet.

#### Praktische Prüfungen

Praktische Prüfungen erfassen die praktischen Fertigkeiten eines Kandidaten, denen in der ärztlichen Tätigkeit eine besondere Bedeutung zukommt. Die herkömmlichen, unstrukturierten praktischen Prüfungen haben eine geringe Objektivität und Reliabilität, da erhebliche Unterschiede bestehen bezüglich des Prüfungsverlaufs, des Schwierigkeitsgrades, des Patientenguts und der abschließenden Beurteilung. Wie schon für die mündlichen Prüfungen, wurden auch hier Modelle für strukturierte praktische Prüfungen entworfen, von denen die "Objective Structured Clinical Examination" (OSCE) vorgestellt werden soll.

Das OSCE-Verfahren wurde 1975 von HARDEN et.al. [8,9] eingeführt als eine Methode zur Einschätzung klinischer Fertigkeiten. Die Gesamtprüfung besteht aus ca. 20 Einzelprüfungen von jeweils 10–20 Minuten Dauer. Die Kandidaten werden am Krankenbett geprüft und von einem Prüfer begleitet, der anhand vorgegebener Checklisten die Fertigkeiten des Kandidaten einschätzt. Es kommen standardisierte Patienten zum Einsatz, die sich die Krankengeschichten realer Patienten aneignen. Testgütekriterien und Praktikabilität des OSCE-Verfahrens wurden in verschiedenen Studien getestet [2,8,9,25]. Gegenüber herkömmlichen praktischen Prüfungen zeichnet sich die OSCE durch eine höhere Objektivität und Reliabilität aus. Die Kandidaten haben die gleichen Aufgaben zu bewältigen und werden von den gleichen Prüfern geprüft. Die Strukturierung betrifft Inhalt, Ablauf und Bewertung der Prüfung. Indem die Kandidaten bis zu zwanzig Teststationen durchlaufen, werden Stichprobenfehler verringert.

Das OSCE-Verfahren gilt als sehr valide [25]. Gegenüber mündlichen oder schriftlichen Prüfungen ist es jedoch kosten- und zeitintensiver [3,24]. In der Kosten-Nutzen-Analyse muß allerdings bedacht werden, daß nur mit dieser Methode klinische Fertigkeiten strukturiert erfaßt werden können.

#### Schriftliche Prüfungen

Von den schriftlichen Prüfungsverfahren sollen hier die Multiple-Choice-Prüfung (MC-Prüfung), Short-Answer-Que-

stion, Essay und Modified-Essay-Question näher erläutert werden.

Das Multiple-Choice-Testverfahren wurde in den USA entwickelt und kommt dort seit 1950 als Prüfungsmethode in der medizinische Ausbildung zur Anwendung, seit 1973 in Deutschland. Bei den MC-Fragen handelt es sich um Auswahlfragen mit vorgegebenen Antwortmöglichkeiten, aus denen der Kandidat die richtige Antwort auszuwählen hat. Je nach der Anzahl der vorgegebenen Möglichkeiten und der Antwort als Einfach-, Mehrfach- oder Kombinationsauswahl werden verschiedene MC-Fragetypen unterschieden [11]. Über die hohe Objektivität und Reliabilität von MC-Prüfungen herrscht in der Literatur generell Einigkeit [3,14,16,24]. Alle Prüflinge erhalten dieselben Fragen, die Auswertung erfolgt nach einem vorgegebenen Schema per Computer. Aufgrund des großen Stichprobenumfangs pro Zeiteinheit zeigen MC-Prüfungen eine hohe Reliabilität, die diejenige anderer Prüfungsarten übertrifft [16]. Ein weiterer Vorteil von MC-Fragen liegt in ihrer Praktikabilität; in der Durchführung gestalten sie sich kostengünstig und ökonomisch [4,21,23,24,34].

Bezüglich der Validität von MC-Prüfungen ist eine differenzierte Betrachtung notwendig. MC-Fragen haben eine hohe Kontentvalidität, wenn sie inhaltlich das Wissensspektrum erfassen, das in der Ausbildung vermittelt wurde. Untersuchungen zur "criterion validity" ergaben, daß die Ergebnisse von MC-Prüfungen stärker mit den Beurteilungen kognitiver Fähigkeiten wie Abruf von Wissen und Sammeln von Informationen korrelieren als mit den Beurteilungen praktischer Fertigkeiten oder affektiver Eigenschaften (Interaktion mit den Patienten, Kollegialität) [16]. Arbeiten zur "construct validity" von MC-Prüfungen untersuchen, welche Aspekte klinischer Kompetenz von MC-Prüfungen erfaßt werden. Zunächst ergibt sich die Schwierigkeit, die einzelnen Teilaspekte zu definieren, aus denen die klinische Kompetenz eines Arztes oder Mediziners besteht. Auf dieses Problem soll hier jedoch nicht näher eingegangen werden. In einer Faktorenanalyse wurde ermittelt, daß MC-Fragen eine enge Korrelation aufweisen zu einem kognitiven Faktor, der auch das Abrufen von Wissen beinhaltet, wohingegen zwischen MC-Fragen und anderen Faktoren (Schlußfolgern, Ausdrucksfähigkeit, klinisches Urteilen) kein signifikanter Zusammenhang besteht [16]. Als weitere Nachteile von MC-Fragen gelten außer der umstrittenen Validität die unerwünschten Auswirkungen auf das Lernverhalten [27,35,36] und die durch die Wahlantworten bedingten Ratefaktoren und Antworthinweise [7,10,15,17,18].

Offene Fragen sind dadurch gekennzeichnet, daß im Gegensatz zu MC-Fragen die Antwort frei zu formulieren ist, je nach Fragetyp mehr oder weniger ausführlich. So handelt es sich bei den "Short-Answer-Questions" um sehr spezifische Fragen, die in Stichworten innerhalb eines bestimmten Umfangs beantwortet werden sollen. Beim "Essay" handelt es sich dagegen um relativ weit gefaßte Fragen, die in Form eines kurzen Aufsatzes zu beantworten sind. Gegenüber MC-Fragen verursachen Offene Fragen höhere Kosten und einen größeren zeitlichen Aufwand [4,5], sie enthalten jedoch keine Antworthinweise und Ratefaktoren [34]. Zwar ist die Objektivität von Offenen Fragen geringer als die von MC-Fragen [14,24], die Validität (i. S. v. criterion und construct validity) wird jedoch besser als die von MC-Fragen beurteilt [1,6,22,26]. Bezüglich

der Reliabilität von Offenen Fragen gibt es unterschiedliche Aussagen, überwiegend wird diese jedoch nicht niedriger als die von MC-Fragen beurteilt, wobei sich die "inter-item reliability" als Funktion der Prüfungszeit darstellt [31].

Als besondere Form der schriftlichen Prüfung soll die "Modified-Essay-Question" (MEQ) vorgestellt werden. In diesem strukturierten Test erfolgt zunächst die Darstellung eines klinischen Falls, zu dem Fragen gestellt werden, die von der Struktur her sowohl Offene Fragen als auch MC-Fragen sein können. Die Objektivität der MEQs ist höher als die der Essays, jedoch niedriger als die von MC-Fragen [24]. Die Reliabilität der MEQs hängt im wesentlichen ab von der Qualität der Modellantworten, die vor Beginn der Prüfung einige Testzyklen durchlaufen [5]. Den MEQs wird eine gute Validität bescheinigt [3,4].

### **Eigene Untersuchungen: Vergleich von MC-Fragen und Offenen Fragen als Prüfungsmethoden in der orthopädischen Ausbildung**

An einem Kollektiv von 543 Studierenden und Ärzten wurde die Anwendbarkeit und Akzeptanz von MC-Fragen und Offenen Fragen mittels eines Fragebogens getestet, der aus 16 MC-Fragen und 4 Offenen Fragen zu orthopädischen Themen bestand. Die Untersuchung wurde an der Ruhr-Universität Bochum durchgeführt. Befragt wurden Studierende verschiedener Semester, die Teilnehmer des Orthopädie-Praktikums (6. klinisches Semester) wurden vor und nach dem Praktikum erfaßt. Die Auswertung der Fragebögen erfolgte für die MC-Fragen nach einem Lösungsschema, für die Offenen Fragen mit einem Modell-Antwortbogen und Punktvergabeschema.

Es wurde gezeigt, daß sich die Prüfungsergebnisse sowohl für die MC-Fragen als auch für die Offenen Fragen mit zunehmendem Ausbildungsniveau verbessern; in allen Gruppen sind die Ergebnisse jedoch besser für die MC-Fragen als für die Offenen Fragen. Der positive Wissenszuwachs ist insgesamt am größten im Praktikum der Orthopädie, wie der Vergleich von Prä-Test und Post-Test zeigt. Für die Lernenden als auch für die Lehrenden stellt diese Art der Lernzielkontrolle eine sinnvolle Rückmeldung dar. Es zeigte sich, daß die Fragen mit Zunahme des Ausbildungsstandes als weniger schwierig empfunden werden, wobei die Einschätzung des Schwierigkeitsgrades für die einzelnen Gruppen nicht unbedingt mit dem Ergebnis korreliert. In allen Gruppen werden die Offenen Fragen schwieriger eingestuft als die MC-Fragen.

### **Schlußfolgerung**

Der Vergleich verschiedener Prüfungsverfahren hinsichtlich der Testgütekriterien zeigt die Vor- und Nachteile eines jeden Verfahrens. Durch den Einsatz eines einzigen Verfahrens ist es nicht möglich, ärztliche Kompetenz in all ihren Dimensionen vollständig zu erfassen. Ein Prüfungssystem, daß den oben genannten Testgütekriterien genügen soll, muß eine Kombination aus unterschiedlichen Prüfungsmethoden darstellen, in der strukturierte mündliche, schriftliche und praktische Prüfungen parallel eingesetzt werden. Zu diesem Ergebnis kommen verschiedene Arbeiten, die sich mit dem Vergleich unterschiedlicher Prüfungsverfahren befassen [3,16,19,20,24,26,29].

## Literatur

- <sup>1</sup> Anbar, M.: Comparing assessments of student's knowledge by computerized open-ended and multiple-choice tests. *Academic Medicine* 66 (1991) 420–422
- <sup>2</sup> Black, N. M. I., R. M. Harden: Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination (OSCE). *Medical Education* 20 (1986) 48–52
- <sup>3</sup> Das Netzwerk: Positionspapier zur Reform des Prüfungswesens im Medizinstudium. Korrespondenzadresse: Prof. Dr. med F. Eitel, Chirurgische Klinik Innenstadt LMU, 80336 München (1994)
- <sup>4</sup> Elstein, A. S.: Beyond multiple-choice questions and essays: The need for a new way to assess clinical competence. *Academic Medicine* 68 (1993) 244–249
- <sup>5</sup> Feletti, G. I., C. E. Engel: The modified essay question for testing problem-solving skills. *The Medical Journal of Australia* 1 (1980) 79–80
- <sup>6</sup> Frary, R. B.: Multiple-choice versus free-response: A simulation study. *Journal of Educational measurement* Vol. 22, NO. 1 (1985) 21–31
- <sup>7</sup> Harden, R., G. Mc, R. Lever, G. M. Wilson: Two systems of marking objective examination questions. *Lancet* i (1969) 40
- <sup>8</sup> Harden, R. M., F. A. Gleeson: Assessment of clinical competence using an objective structured clinical examination. *Medical Education* 13 (1979) 41–54
- <sup>9</sup> Harden, R. M., M. Stevenson, W. W. Downie, G. M. Wilson: Assessment of clinical competence using objective structured examination. *British Medical Journal* I (1975) 447–451
- <sup>10</sup> Harden, R., G. Mc, R. A. Brown, L. A. Biran, W. P. Dallas Ross, R. E. Wakeford: Multiple-choice questions: to guess or not to guess. *Medical Education* 10 (1976) 27–32
- <sup>11</sup> Hubbard, J. P., V. C. William: Multiple-Choice examinations in medicine. Lea & Febiger, Philadelphia (1968)
- <sup>12</sup> Klauer, K. J.: Kriteriumsorientierte Tests. Lehrbuch der Theorie und Praxis lehrzielorientierten Messens. Hogrefe, Göttingen, Toronto, Zürich (1987)
- <sup>13</sup> Klauer, K. J.: Kontenvalidität von Prüfungen. In: Westhoff, K. (Hrsg.): Erstes Symposium zu Prüfungen in der Medizin: Multiple-Choice. Möglichkeiten und Grenzen von Multiple-Choice-Prüfungen in der Medizin. Papst, Lengerich (1995) 78–88
- <sup>14</sup> Kommission der Europäischen Gemeinschaften: Bericht über die Beurteilungsmethoden in der medizinischen Grundausbildung, Beratender Ausschuss für die ärztliche Ausbildung, III/D1643/7/83-DE, Brüssel (1987)
- <sup>15</sup> Lennox, B.: Marking Multiple-Choice Examinations. *British Journal of Medical Education* 1 (1967) 203–211
- <sup>16</sup> Levine, H. G., C. H. Mc Guire, L. W. Nattress: The validity of multiple-choice achievement tests as measures of competence in medicine. *American Educational Research Journal*, Vol. 7, NO. 1, January (1970) 69–82
- <sup>17</sup> McCarthy, W. H., M. B. S., F. R. A. C. S.: An assessment of the influence of cueing items in objective examinations. *Journal of Medical Education* 41 (1966) 263–266
- <sup>18</sup> McCloskey, D. I., R. A. B. Holland: A comparison of student performances in answering essay-type and multiple-choice questions. *Medical Education* 10 (1976) 382–385
- <sup>19</sup> Newble, D. I., A. Baxter, R. G. Elmslie: A comparison of multiple-choice tests and free-response tests in examinations of clinical competence. *Medical Education* 13 (1979) 263–268
- <sup>20</sup> Nnodim, J. O.: Multiple-choice testing in anatomy. *Medical Education* 1992 Jul. 26(4) (1992) 301–309
- <sup>21</sup> Norcini, J. J., D. B. Swanson, L. J. Grosso, G. D. Webster: Reliability, validity and efficiency of multiple choice questions and the patient management problem item formats in assessment of clinical competence. *Medical Education* 19 (1985) 238–247
- <sup>22</sup> Norman, G. R., E. K. M. Smith, A. C. P. Powles, P. J. Rooney, N. L. Henry, P. E. Dodd: Factors underlying performance on written tests of knowledge. *Medical Education* 21 (1987) 297–304
- <sup>23</sup> Pauli, H. G., IAE (Institut für Ausbildungs- und Examensforschung) der Medizinischen Fakultät Bern: Die Multiple-Choice-Methode in der Evaluation medizinischer Ausbildungsresultate. *Schweizerische Ärztezeitung*, Bd. 66, Heft 26 (1985) 1193–1201
- <sup>24</sup> Planungsgruppe Medizin, Hardegg, W., M. Schäfer, A. Nelle, A. Richterich: Gutachten über das Prüfungswesen in einer neuen Approbationsordnung für Ärzte (ÄAppO) anhand einer Untersuchung ausländischer Prüfungssysteme im Studiengang Medizin im Zusammenhang mit den Diskussionen zur neuen Approbationsordnung für Ärzte. Erstellt im Auftrag des Bundesministeriums für Gesundheit. Medizinischer Fakultätentag, Heidelberg, Febr. 1995
- <sup>25</sup> Reznick, R. et al.: An Objective Structured Clinical Examination for the Licentiate: Report of the Pilot Project of the Medical Council of Canada. *Academic Medicine* 67 (8) (1993) 487–494
- <sup>26</sup> Rothman, A. I., N. Kerényi: The assessment of an examination in pathologie consisting of multiple-choice, practical and short essay questions. *Medical Education* 14 (1980) 341–344
- <sup>27</sup> Schimmelpfennig, K.: Empfehlungen zur gesetzlichen Neugestaltung der Ärztlichen Vorprüfung und der Ärztlichen Prüfung. *Med. Ausbildung* 8:5–14 und *Dtsch. Ärztebl.* 89 (5)B (1991) 204–206
- <sup>28</sup> Schumacher, C. F.: Reliability, Validity and Standard Setting. In: Hubbard, J. P.: *Measuring Medical Education*. Lea & Febiger, Philadelphia (1978) 59–71
- <sup>29</sup> Schwartz, R. W., M. B. Donnelly, D. A. Sloan, B. Young: Knowledge gain in a problem-based surgery clerkship. *Academic Medicine* 69 (1994) 148–151
- <sup>30</sup> Smith, R. M.: The Triple Jump Examination as an Assessment Tool in the Problem-Based Medical Curriculum at the University of Hawaii. *Academic Medicine* 68 (5) (1993) 366–372
- <sup>31</sup> Stalenhoef-Halling, B. F., C. P. M. Van Der Vleuten, T. A. M. Jaspers, J. F. B. M. Fiolet: The feasibility, acceptability and reliability of open-ended questions. In: Bender W. et al. (ed.): *Teaching and assessing clinical competence*. Boek Werk, Groningen (1990) 552–557
- <sup>32</sup> Trost, G.: Evaluation of examinations-methods, results, conclusions. Invited paper, presented at the International Symposium on Examinations in Medical Education and their influence on teaching and learning, Mainz, 29./30. 6. 1995
- <sup>33</sup> Van den Bussche, H.: Prüfungen in der ärztlichen Ausbildung – Funktionen, Möglichkeiten und Grenzen. In: Habeck, Schagen, Wagner: *Reform der Ärzteausbildung*. Blackwell Wissenschaft (1993) 91–107
- <sup>34</sup> Whitby, L. G.: Marking systems for multiple-choice examinations. *Medical Education* 11 (1977) 216–220
- <sup>35</sup> Wirsching, M.: Fünf Thesen zur Reform der ärztlichen Ausbildung. *Dtsch. Ärztebl.* 85 (1988) 13–17
- <sup>36</sup> Wissenschaftsrat: Leitlinien zur Reform des Medizinstudiums. Eigenverlag, Bremen, 3. 7. 1992

Andrea Elmer

Abteilung Pneumologie  
St. Hildegardis-Krankenhaus  
Hildegardstr. 2  
55131 Mainz