

Judith Brockmann/Arne Pilniok (Hrsg.)

# Prüfen in der Rechtswissenschaft

Probleme, Praxis und Perspektiven



**Nomos**

# Theorie und Praxis: Prüfungen in der Medizin

*Daniel Bauer/Anja Görlitz/Johanna Huber/Martin R. Fischer/Matthias Holzer*

## *I. Ein Gesetz zum Schutz der Patientinnen und Patienten*

Auf die Frage nach dem Sinn und Zweck von Prüfungen in der Medizin gibt es mehrere Antworten. Die Ärztliche Approbationsordnung (ÄAppO) formuliert inhaltliche und formale Rahmenbedingungen für die Approbation als Ärztin bzw. Arzt und hat damit entscheidenden Einfluss auf die strukturelle und inhaltliche Gestaltung medizinischer Curricula. Nicht nur die zweiteilige Ärztliche Prüfung, sondern auch zahlreiche benotete und unbenotete studienbegleitende Leistungen sollen sicherstellen, dass nur gut ausgebildete Ärztinnen und Ärzte in das Berufsleben entlassen und zur Patientenbehandlung zugelassen werden. Es darf als ureigenes Interesse also nicht nur der Lehrenden an der Universität, sondern der gesamten Gesellschaft angesehen werden, dass Prüfungen qualitätssichernd das Studium begleiten und abschließen, um hohe Standards der medizinischen Versorgung in Deutschland sicherzustellen. Seit einigen Jahren gilt dies nicht mehr allein für die medizinische Ausbildung und die Weiterbildung zur Fachärztin bzw. zum Facharzt, sondern auch für die laufende Fortbildung im Berufsalltag.<sup>1</sup> Hierzu wurde von den Ärztekammern ein Zertifikatssystem geschaffen, das sicherstellen soll, dass auch Ärztinnen und Ärzte, die schon lange die Universität verlassen haben, auf dem aktuellen Stand von Wissenschaft und Forschung sind. In diesem Beitrag sollen dennoch schwerpunktmäßig universitäre Prüfungen beleuchtet werden, wobei geschilderte Verfahren und Begrifflichkeiten in ähnlicher Form auch auf das „lebenslange Prüfen“ anzuwenden sind.

## *II. Abstimmung von Lehren, Lernen und Prüfen*

Soll eine Prüfung wie besprochen dazu beitragen, die sichere und hochwertige ärztliche Patientenversorgung sicherzustellen, muss sie in Inhalt und Form auch die Anforderungen an zukünftige Ärztinnen und Ärzte widerspiegeln. Um Lehren, Lernen und Prüfen mit dem zu erwartenden Arbeitsalltag der Absolventinnen und Absolventen abzustimmen, ist es auf übergeordneter Ebene sinnvoll, sich das Ziel der Ausbildung zu vergegenwärtigen und die nicht immer einfach zu beantwor-

<sup>1</sup> Bundesärztekammer (Hrsg.), Empfehlungen zur ärztlichen Fortbildung, 3. überarb. Aufl., Köln 1993.

tende Frage zu stellen, welches Profil, welche Kompetenzen eine Absolventin bzw. ein Absolvent eigentlich haben soll. Hieran kann sich dann ein integriertes Prüfungskonzept für die gesamte Ausbildung orientieren, so dass auch Einzelprüfungen in einen übergeordneten Zusammenhang eingebettet sind. Analog dazu sollte eine Einzelprüfung sich dann an den relevanten Lernzielen einer Lehrveranstaltung orientieren. So genannte Prüfungsblaupausen („Prüfungs-Blueprints“) sollen hierfür eine Hilfe darstellen: Sie bieten eine multidimensionale Übersicht, inwiefern die geplanten Prüfungsaufgaben z.B. die Subdisziplinen des Faches, die Vielfältigkeit des Patientenspektrums, die Dringlichkeit der Beratungsanlässe, die klinische Relevanz sowie die notwendigen Behandlungsschritte (Diagnostik, Therapie) abdecken und zu diesen Aspekten formulierte Lernziele widerspiegeln. Solche Blueprints sind besonders dann hilfreich, wenn der Unterricht zu einer Prüfung von mehreren Dozentinnen und Dozenten geleistet wird und entsprechend auch die Prüfungsaufgaben mit allen Beteiligten, ggf. auch mit Dozentinnen und Dozenten anderer Fachbereiche, abgesprochen werden müssen.

## 1. Lernsteuernde Wirkung von Prüfungen

Man möge sich an die eigene Ausbildung erinnern und sich vergegenwärtigen, welchen enormen Einfluss Prüfungen auf das eigene Lernverhalten hatten („Assessment drives learning“). Diesen Einfluss gilt es bei der Erarbeitung eines Prüfungskonzeptes, aber auch einer Einzelprüfung zu bedenken und ggf. gezielt zu nutzen.

Verschiedene Faktoren können das Lernverhalten zur Prüfungsvorbereitung von Studentinnen und Studenten beeinflussen. Ein individueller Lernplan kommt durch die komplexe, in der Regel unbewusste Verrechnung unterschiedlich gewichteter Faktoren zustande: Aus der prognostizierten Schwierigkeit der Prüfung, der zeitlichen Entfernung zur Prüfung, der Menge des bereits absolvierten Lernstoffes und dem persönlichen Verständnis der Inhalte resultiert die Lernleistung, die erbracht werden müsste, um diese Prüfung zu bestehen bzw. das individuelle Risiko des Nichtbestehens. Weiterhin ist von Bedeutung, welche Konsequenzen ein Nichtbestehen hätte, was der eigene Anspruch an die Prüfungsleistung ist, wie ausgeprägt das eigene, intrinsische Interesse an der Materie ist und welche Ressourcen zur Verfügung stehen (u. a. auch, ob andere Prüfungen um die begrenzte Lernzeit konkurrieren). Der Druck von Kommilitoninnen und Kommilitonen kann als extrinsischer Motivator verstärkend hinzukommen.<sup>2</sup> Mit welchem Aufwand, mit welcher

2 *Cilliers/Schuwirth/Adendorff/Herman/van der Vleuten*, The mechanism of impact of summative assessment on medical students' learning, in: *Advances in Health Sciences Education* 15 (2010), Nr. 5, S. 695-715.

Sorgfalt und mit welchen Materialien und Lernformen eine Prüfung vorbereitet wird, kann also in starkem Ausmaß gesteuert werden.<sup>3</sup>

Auch beim reflektierten und didaktisch motivierten Einsatz von Prüfungen im Medizinstudium sollte man sich bewusst bleiben, dass ein respektabler Anteil von bis zu 10% der Medizinstudentinnen und -studenten in Deutschland unter Prüfungsangst leiden<sup>4</sup> und die didaktisch eingesetzte Prüfung für diese Stress bedeuten kann. Jenseits von Aufklärung und unterstützenden Angeboten zum Thema Prüfungsangst gilt es daher, Transparenz bzgl. der Prüfungen und deren Machbarkeit herzustellen.

## 2. Feedback für alle

Eine regelmäßige Rückmeldung über den eigenen Kenntnisstand ist dem kontinuierlichen Lernen von Studentinnen und Studenten zuträglich und meldet ihnen sowohl Inhalte, die bereits gut beherrscht werden als auch solche, bei denen eine Intensivierung des Lernens angemessen wäre. Solch eine rückkoppelnde (formative) Wirkung haben alle Prüfungen ebenso wie informelle Rückmeldungen durch Dozentinnen und Dozenten. *Hattie* beschreibt 2008 in einer Metaanalyse von über 50.000 Lehr-/Lernstudien, dass die Rückmeldung des eigenen Kenntnisstands mit einer Effektstärke von im Mittel 1,13 die effektivste Einzelintervention zur Lernförderung ist.<sup>5</sup>

Der *Progress Test Medizin*, der auch an mehreren Fakultäten in Deutschland eingesetzt wird, ist ein eindrückliches Beispiel für eine rein formative Prüfung. Regelmäßig wird ab dem ersten Semester ein 200 Aufgaben umfassender Multiple-Choice-Test auf Niveau des Abschlussexamens angeboten, der nicht bestehensrelevant ist, nicht benotet wird und auf den bewusst nicht gelernt werden soll, der aber detailliertes Feedback über das Ausmaß des spontan abrufbaren Wissens bietet und die persönliche Lernkurve zeigt, auch im Vergleich zur eigenen Semesterkohorte.<sup>6</sup>

3 *van der Vleuten*, The assessment of professional competence: Developments, research and practical implications, in: *Advances in Health Sciences Education* 1 (1996), Nr. 1, S. 41-67.

4 *Schaefer/Mattheß/Pfizer/Köhle*, Seelische Gesundheit und Studierenerfolg von Studierenden der Medizin mit hoher und niedriger Prüfungsängstlichkeit, *Psychotherapie, Psychosomatik*, in: *Medizinische Psychologie* 57 (2007), Nr. 7, S. 289-297.

5 *Hattie*, *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Oxford 2008.

6 *Verwijnen/Imbos/Snellen/Stalenoef/Pollemans/v. Luyk/Sprooten/v. Leeuwen/v.d. Vleuten*, The Evaluation System at the Medical School of Maastricht, *Assessment & Evaluation in Higher Education* 7 (1982), Nr. 3, S. 225-244; *Osterberg/Köbel/Brauns*, Der Progress Test Medizin: Erfahrungen an der Charité Berlin, in: *GMS Zeitschrift für Medizinische Ausbildung* 23 (2006), Nr. 3, S. 62-66.

Prüfungsergebnisse bedeuten jedoch auch Rückmeldung an Dozentinnen und Dozenten. Wurden Prüfungsaufgaben zu einer bestimmten Veranstaltung unerwartet gut oder schlecht bearbeitet, kann dies auch die Lehre spiegeln und z. B. besonders gute oder verbesserungswürdige Lehre kennzeichnen beziehungsweise auf ausgefallene Unterrichtsveranstaltungen oder Lücken im Curriculum hinweisen. Zu guter Letzt werden Prüfungsergebnisse auch genutzt, um inter fakultäre Vergleiche anzustellen. Inwiefern ein solches Benchmarking sinnvoll ist, hängt stark von der Vergleichbarkeit der Prüfungen in Form und Inhalt ab und sollte daher vorsichtig interpretiert werden.

### 3. Gute Prüfungen

Wie kann man also die Qualität einer Prüfung messen? Hierzu werden für studienbegleitende Prüfungen im Medizinstudium regelmäßig die allgemeinen Testgütekriterien der psychologischen Testtheorie herangezogen, insbesondere die drei Hauptgütekriterien der Objektivität, Reliabilität und Validität.<sup>7</sup>

Die *Objektivität* eines Testes gibt an, inwiefern eine Messung von der Untersuchungsleiterin oder dem Untersuchungsleiter (also die Prüferin bzw. der Prüfer) und äußeren Prüfungsbedingungen unabhängig ist. Hierbei können die drei Aspekte der Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden werden. Solange die Bewertung von Teilaufgaben eindeutig festgelegt ist, z. B. in Multiple Choice-Aufgaben, und für alle Kandidatinnen und Kandidaten möglichst ähnliche Prüfungsbedingungen herrschen, ist die Objektivität in allen drei Aspekten leicht zu gewährleisten. Komplizierter wird es, wenn qualitative Informationen innerhalb der Prüfung bewertet werden sollen (z. B. verschiedene Aspekte von Kommunikation, Professionalität). Mangelnde Objektivität wird in der Regel von den Prüflingen bemerkt („unfaire Prüfung“)<sup>8</sup> und kann nach einer Prüfung Konsequenzen bis hin zu ihrer Annullierung haben. Zur Objektivität (wie auch zur Reliabilität) tragen die Schulung von Schauspielpatientinnen bzw. Schauspielpatienten und Prüferinnen und Prüfern, standardisierte Checklisten zur Bewertung studentischer Leistungen sowie angemessene Information der Studentinnen und Studenten bezüglich *Organisation und Inhalt der Prüfung* bei.

Die *Reliabilität* einer Prüfung ist ein Maß dafür, wie zuverlässig gemessen wurde. Einem Axiom der klassischen Testtheorie zufolge müsste eine Messung (hier: Prüfung) bei Messwiederholung möglichst dasselbe Ergebnis (hier: Bestehen, No-

<sup>7</sup> Bortz/Döring, *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4. Aufl. Berlin 2006.

<sup>8</sup> Duffield/Spencer, *A survey of medical students' views about the purposes and fairness of assessment*, in: *Medical Education* 36 (2002), Nr. 9, S. 879-886.

ten) liefern. Da dieses Qualitätskriterium in der Praxis nicht durch Prüfungswiederholung bestimmt werden kann (Erinnerung an erste Prüfung als konfundierender Faktor), hat sich im medizinischen Prüfungsalltag das Modell der internen Konsistenz durchgesetzt: Die Prüfung wird hierbei in zwei Teilprüfungen getrennt, wobei die Korrelation der Testwerte aller möglichen Testhalbierungen dann einer Untergrenze der Reliabilität entspricht (sog. Cronbachs  $\alpha$ ; Maximum bei 1, logisches Minimum bei 0). Die Reliabilität steigt demzufolge mit steigender Stichprobe (Anzahl der Aufgabenstellungen) sowie steigender Interaufgabenkorrelation. Für summative Prüfungen, also bestehens- und ggf. notenrelevante Prüfungen, wird eine Reliabilität von  $\alpha \geq 0,8$  gefordert.<sup>9</sup> Abbildung 1 visualisiert die Varianz des Notenspektrums unterschiedlich reliabler, sonst aber identer Prüfungen.

a) Verteilung der Noten bei Reliabilität $\alpha = 0,65$		Note Prüfung 2				
		1	2	3	4	5
Note Prüfung 1	1	5	4	2		
	2	4	8	8	3	
	3	2	8	2	8	2
	4		3	8	8	4
	5			2	4	5

b) Verteilung der Noten bei Reliabilität $\alpha = 0,80$		Note Prüfung 2				
		1	2	3	4	5
Note Prüfung 1	1	7	4	1		
	2	4	11	7	1	
	3	1	7	14	7	1
	4		1	7	11	4
	5			1	4	7

Abb. 1: Theoretische Verteilung der Noten derselben Kandidatinnen und Kandidaten in zwei gleichwertigen („parallelen“) aufeinanderfolgenden Prüfungen bei a) einer Testreliabilität von  $\alpha = 0,65$  und b) einer Testreliabilität von  $\alpha = 0,80$ . Man beachte, dass im Fall a) 14 Kandidatinnen bzw. Kandidaten im zweiten Test eine um zwei Notenstufen abweichende Note haben, im Fall b) „nur noch“ 6. Nach *Andreas Möltner*, Kompetenzzentrum für Prüfungen in der Medizin Baden-Württemberg an der Universität Heidelberg.

Eine hohe Reliabilität ist in der Prüfungsroutine im Medizinstudium speziell in kleineren Fächern ein oftmals nur schwer einzuhaltendes Qualitätsmerkmal und wird von Multiple Choice-Prüfungen mit guter Aufgabenqualität im Allgemeinen erst ab einer Aufgabenzahl von 50 bis 60 erreicht. Da diese Zahl im Prüfungsalltag in einigen Prüfungen nicht gewährleistet ist, z. B. wegen eines zu geringen Stoffumfangs, ist darauf zu achten, dass zumindest in „high stakes“-Prüfungen, z. B. in

<sup>9</sup> *Möltner/Schellberg/Jünger*, Grundlegende quantitative Analysen medizinischer Prüfungen, in: GMS Zeitschrift für Medizinische Ausbildung 23 (2006), Nr. 3, S. 12-22.

Staatsprüfungen, ausreichend viele Prüfungsinformationen zusammengefasst werden, um ein reliables Ergebnis sicher zu stellen.

Die *Validität* einer Prüfung entspricht dem Grad an Genauigkeit, mit dem die Prüfung das misst, was bei der Prüfungserstellung zu messen beabsichtigt war.<sup>10</sup> Hierbei beschreibt die Inhaltsvalidität, wie sehr die Stichprobe an Wissen in der Prüfung die Gesamtheit des Wissens in dem jeweiligen Fach repräsentiert. Im Prüfungsalltag wird dies durch die Nutzung einer Prüfungsblaupause erleichtert. Faktoren der Kriteriumsvalidität, wie gut also Prüfungsleistungen z. B. Aussagen über spätere klinische Tätigkeit zulassen, sind für Einzelprüfungen im Medizinstudium bzgl. fachdomänenspezifischer Merkmale nur schwer zu finden. Nichtsdestotrotz sollte bei der Prüfungserstellung auf eine praxisnahe Aufgabenstellung, in der Medizin z. B. durch Patientenfälle, Wert gelegt werden. Auch die Konstruktvalidierung, z. B. indem unterschiedliche Prüfungsleistungen verschiedener Formate im selben oder in verwandten Fächern zusammenpassen, wird im Prüfungsalltag in der Regel nicht thematisiert, kann aber bei bekannten Korrelationskoeffizienten der Prüfungen erfolgen (Multitrait-Multimethod-Methode).

Zu den beschriebenen Hauptgütekriterien kommen noch andere sekundäre Gütekriterien: Die Testfairness und Unverfälschbarkeit sollen sicherstellen, dass Prüflinge nicht aufgrund persönlicher Eigenschaften benachteiligt werden oder durch Benutzung unerlaubter Hilfsmittel eine bessere Prüfungsleistung vorspiegeln können. Sie kann als Aspekt der Objektivität verstanden werden. Die Testökonomie umfasst die Zumutbarkeit für Prüferinnen bzw. Prüfer und Prüflinge sowie eine angemessene Kosten-Nutzen-Relation. Die Transparenz, z. B. Nachvollziehbarkeit von Prüfungsinhalten und Notengebung, fließt in hohem Maße in die subjektive Bewertung der Prüfung durch die Prüflinge mit ein. Auf die lernsteuernde Wirkung von Prüfungen ist bereits eingegangen worden.

#### 4. Gute Aufgaben

Neben den Gütekriterien für gesamte Prüfungen gibt es gemäß der klassischen Testtheorie auch numerische Gütekriterien für einzelne Aufgaben (engl: Items),<sup>11</sup> die deren Bewertung im Hinblick auf ihren Beitrag zur Erfassung des Prüflingswissens ermöglichen. Hier wäre zum einen die *Schwierigkeit* der Items zu nennen, die als mittlere erreichte Punktzahl aller Kandidaten bei dem betreffenden Item errechnet wird. Die Schwierigkeit eines Testitems ist auf einen Wert

10 *Lienert/Raatz*, Testaufbau und Testanalyse, 6. Aufl. Weinheim, 1998.

11 *Möltner/Duelli/Resch/Schultz/Jünger*, Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten, in: *GMS Zeitschrift für Medizinische Ausbildung* 27 (2010), Nr. 3, Doc44.

zwischen 0 (kein Prüfling hat korrekt geantwortet) und 1 (alle Prüflinge haben korrekt geantwortet) normiert. Aufgaben, die von keinem oder allen Prüflingen korrekt beantwortet werden, liefern keine Information zur Einschätzung des individuellen Prüflingswissens im Vergleich zur Vergleichskohorte. Daher sollte die Schwierigkeit der Aufgaben in einem Bereich streuen, z. B. zwischen 0,5 und 0,9. Wie sehr ein Item gute und schlechte Prüflinge unterscheiden kann, gibt die *Trennschärfe* wieder. Die Trennschärfe ist eine Korrelation der Ergebnisse eines Items mit den Ergebnissen des gesamten Tests. Sie kann als Korrelation Werte im Bereich von -1 bis 1 annehmen und sollte über 0,3 liegen. Eine sehr kleine Trennschärfe drückt aus, dass der Erfolg im gesamten Test nicht mit dem Erfolg beim betreffenden Item korreliert war. Eine negative Trennschärfe (schlechte Prüflinge sind in dieser Aufgabe besser als gute) kann ein Hinweis auf eine fehlerhafte Expertenlösung sein und sollte eine genaue Betrachtung der Aufgabenstellung möglichst noch vor der Ergebnisveröffentlichung zur Folge haben.

Zur Optimierung aller Gütekriterien empfehlen sich Reviewverfahren sowohl zur Optimierung der Aufgaben im Vorfeld der Prüfung als auch im Nachgang unter Berücksichtigung der Itemkennwerte im Hinblick auf die Minimierung studentischer Einsprüche und auf eine spätere Wiederverwendung der Aufgaben.

### *III. Die Toolbox: verschiedene Prüfungsformate in der Medizin*

Die zur Verfügung stehenden Methoden werden entsprechend ihrer ökologischen Validität von *Miller* dargestellt (s. Abb. 2).<sup>12</sup> Auf der untersten Ebene („knows“) ist die Prüfung des kognitiven, z.B. biomedizinischen Grundlagenwissens repräsentiert, welches als Voraussetzung für klinisches Denken gesehen werden kann.<sup>13</sup> Geeignete Formate, diese Art von Wissen zu prüfen, sind vor allem schriftliche Formate wie Wahlantwortaufgaben (Multiple Choice Questions, MCQ) und Kurzantwortaufgaben (Short Answer Questions, SAQ) oder auch Strukturierte Mündliche Prüfungen (SMP).

Auf der nächsten Stufe („knows how“) stehen Aufgaben- und Prüfungsformate, bei denen ein Prüfling demonstrieren muss, dass sie oder er unterschiedliche Wissensseinheiten sinnvoll miteinander verknüpfen und dieses Wissen anwenden kann. Dies wird z. B. von entsprechend konstruierten schriftlichen (modifizierte Aufsatzfragen (Modified Essay Questions, MEQ), MCQ, SAQ) und mündlichen Formaten gewährleistet. Zwei neuere schriftliche Formate, die dies ebenfalls erlauben,

12 *Miller*, The assessment of clinical skills/competence/performance, in: *Medical Education* 65 (1990), Nr. 9, Suppl. S. 63-67; *Wass/van der Vleuten/Shatzer/Jones*, in: *Assessment of clinical competence*, in: *The Lancet* 357 (2001), Nr. 9260, S. 945-949.

13 *Woods*, Science is fundamental: the role of biomedical knowledge in clinical reasoning, in: *Medical Education* 41 (2007), Nr. 12, S. 1173-1177.

sind die Key Feature (KF) Prüfung<sup>14</sup> und der Skriptkonkordanztest (Script Concordance Test, SCT).<sup>15</sup> Diese Formate messen v. a. klinisches Denken angesichts limitierter zur Verfügung stehender klinischer Informationen.

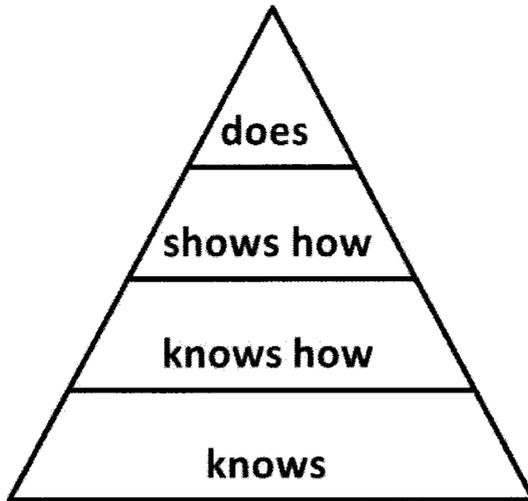


Abb. 2: Rahmenwerk für klinische Prüfungen, vier Kompetenzebenen nach *Miller*, 1990 (sog. „Miller-Pyramide“)

Zur Messung der Kompetenzen auf der dritten Ebene der *Millerschen* Pyramide („shows how“) stehen Performanzprüfungen zur Verfügung, wie das in Deutschland inzwischen weit verbreitete objektive, strukturierte klinische Examen (Objective Structured Clinical Examination, OSCE)<sup>16</sup> und der objektive, strukturierte, lange Examinationsreport (Objective Structured Long Examination Record, OSLER).<sup>17</sup> Hier müssen die Prüflinge, z. B. unter Zuhilfenahme sogenannter „standardisierter Patientinnen und Patienten“ oder „Schauspielpatientinnen und -patienten“, ihre Fähigkeiten in klinischer Gesprächsführung unter Beweis stellen oder an einem Phantom Untersuchungstechniken demonstrieren. Der OSLER kommt der mündlich-klinischen Prüfung im Zweiten Abschnitt der Ärztlichen Prüfung

14 *Fischer/Herrmann/Kopp*, Answering multiple-choice questions in high-stakes medical examinations, in: *Medical Education* 39 (2005), Nr. 9, S. 890-894.

15 *Charlin/van der Vleuten*, Standardized assessment of reasoning in context of uncertainty: The Script Concordance Test approach, in: *Evaluation and the Health Professions* 27 (2004), Nr. 3, S. 304-319.

16 *Harden/Gleeson*, Assessment of clinical competence using an objective structured clinical examination (OSCE), in: *Medical Education* 13 (1979), Nr. 1, S. 39-54.

17 *Gleeson*, Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER), in: *Medical Teacher* 19 (1997), Nr. 1, S. 7-14.

wahrscheinlich am nächsten und kann auch an realen Patientinnen und Patienten stattfinden.

Die ökologisch validesten Prüfungen (Ebene „does“) finden in einer realen Umgebung statt, also an realen Patientinnen und Patienten. Hier demonstrieren die Prüflinge, dass sie klinisches Denken in diagnostische und therapeutische Konzepte umsetzen können. In einem kleinen klinischen Examen (Mini Clinical Evaluation Exercise, Mini-CEX)<sup>18</sup> wird dies in strukturierter Manier umgesetzt. Letztlich ist jede (auch formative) Evaluation im realen klinischen Setting auf dieser obersten Stufe der Prüfungspyramide einzuordnen.

Dass nicht in jeder der studiumsbegleitenden Prüfungen alle Formate zum Einsatz kommen können, ist der Prüfungsökonomie geschuldet. Während qualitativ hochwertige schriftliche Prüfungen vor allem in der Erstellung aufwändig sind, sind praktische Prüfungen am Prüfungstag selbst von hohem personellem Einsatz vonseiten der durchführenden Institution gekennzeichnet. Gleichmaßen sind je nach Inhalt der Prüfung einige Formate geeigneter, andere weniger sinnvoll. Grundsätzlich gilt es, die klinische Gesamtbefähigung der Prüflinge durch Einsatz verschiedener Formate und Auswahl geeigneter Inhalte im Sinne einer Triangulation zu approximieren.

Auf alle Aufgabenformate kann in diesem Text nicht eingegangen werden, stattdessen sollen als Vertreter der schriftlichen Formate Multiple Choice- und Freitextaufgaben besprochen werden sowie die OSCE-Prüfung und arbeitsplatzbasiertes Prüfen und schlussendlich die strukturierte mündliche Prüfung.

## 1. Schriftliche Formate: Multiple Choice und Freitext

Schriftliche Prüfungsformate haben den Vorteil, dass die Durchführungsobjektivität in der Regel gewährleistet ist, indem alle Kandidatinnen und Kandidaten die gleiche Aufgabenstellung erhalten. Auch Auswertungsobjektivität und Interpretationsobjektivität lassen sich in Zweifelsfällen nachträglich überprüfen. Die schriftlichen Aufgabenformate lassen sich in geschlossene (Antworten sind vorgegeben), halboffene (Freitextantwort mit niedrigem Elaborationsgrad) und offene (Freitextantwort mit hohem Elaborationsgrad) Formate unterteilen, die alle in medizinischen Prüfungen verwendet werden können, wobei aus praktischen Gründen häufig geschlossene und halboffene Formate gewählt werden. In der Abwägung zueinander bleibt festzuhalten, dass geschlossene Aufgaben eine höhere Objektivität bieten und offene Formate unter Umständen eine höhere Validität haben, da sie

18 *Norcini/Blank/Duffy/Fortna*, The Mini-CEX: A Method for Assessing Clinical Skills, in: *Annals of Internal Medicine* 138 (2003), Nr. 6, S. 476-481.

auch komplexe praktische Fragestellungen ermöglichen auf Kosten von Interpretationsobjektivität und Generalisierbarkeit.<sup>19</sup>

<p>Bei einer Patientin ist die Erythrozytenzahl pro <math>\mu\text{L}</math> Blut vermindert, aber die Hämoglobinkonzentration im Normbereich (= MCH erhöht).</p> <p>Für welche der folgenden Störungen spricht diese Befundkonstellation am ehesten?</p> <p>(A) Folsäuremangel (B) Darmblutung (C) Eisenmangel (D) verkürzte Erythrozytenlebensdauer (E) Vitamin-D-Mangel</p>
--

Abb. 3. Allgemeiner Aufbau einer Wahlantwortaufgabe: Informationstext, Aufgabenstellung, vorformulierte Antwortmöglichkeiten (nach: IMPP, vgl. Fn. 20).

Seit vielen Jahren sind in Prüfungen im Medizinstudium Wahlantwortaufgabenformate (Multiple Choice Questions, MCQ) im Einsatz, welche sich einen festen Platz in der Überprüfung von Fakten- und teilweise Anwendungswissen erobert haben. Im Gegensatz zu offenen Aufgabenformaten („constructed response format“), bei denen die Prüflinge selbständig den Antworttext auf eine Aufgabenstellung formulieren, muss man bei MC-Formaten („selected response format“) unter vorgegebenen Antwortalternativen auswählen. Der Aufbau einer MC-Aufgabe ist in Abb. 3 beispielhaft wiedergegeben. Auf einen Informationstext (bei klinischen Aufgabenstellungen oftmals mit Patientenvignette) folgt die Aufgabenstellung. Die Prüflinge haben dann von den gegebenen Antworten eine i. d. R. vorgegebene Anzahl an Richtigerantworten auszuwählen.<sup>20</sup>

Ein Vorteil der MC-Aufgaben bei Prüfungen in der Medizin liegt u. a. darin, dass im Sinne der Prüfungsökonomie ein breites Wissensspektrum bei einer großen Anzahl von Prüflingen erfasst werden kann.<sup>21</sup> MC-Prüfungen sind dabei vergleichsweise objektiv und standardisiert.<sup>22</sup> Die Nutzung von computerlesbaren Antwortbögen minimiert zusätzlich den Korrekturaufwand, erhöht die Zuverlässigkeit im Vergleich zur händischen Auswertung mit Schablone und beschleunigt

19 *Krebs*, Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung. Institut für Medizinische Lehre IML, Abteilung für Ausbildungs- und Examensforschung AAE, Bern 2004.

20 Institut für medizinische und pharmazeutische Prüfungsfragen, Praktische Hinweise zur Durchführung der schriftlichen Prüfungen nach der Approbationsordnung für Ärzte, 46. überarbeitete Auflage, Mainz 2009.

21 *Elstein*, Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence, in: *Academic Medicine* 68 (1993), Nr. 4, S. 244-249.

22 *Epstein*, Assessment in Medical Education, in: *New England Journal of Medicine* 356 (2007), S. 387-396.

das Auswerteverfahren, sodass die Prüflinge rasch und zuverlässig eine Rückmeldung über ihre jeweiligen Leistungen erhalten.

Der Aufwand, gute MC-Aufgaben zu erstellen, ist dabei dennoch relativ hoch. Aufgabenautorinnen und -autoren müssen darin geschult sein,

- welche MC-Formate überhaupt genutzt werden sollen,
- relevante, dem Blueprint und Lernzielkatalog entsprechende Inhalte zu definieren,
- Aufgabenvignetten zu verfassen,
- beantwortbare Aufgabenstellungen zu formulieren,
- plausible Falschantworten (Distraktoren) auszuwählen und
- unbeabsichtigte Lösungshinweise (Cues) zu vermeiden.<sup>23</sup>

Nachteilig kann sich allerdings der Einsatz von MC-Prüfungen auf das Lernverhalten der Studentinnen und Studenten auswirken. Eine Studie von *Scouller* zeigte, wie sich bei Studentinnen und Studenten der Erziehungswissenschaften eher oberflächliches Lernen zwar positiv auf das Abschneiden in einer primär Wissen messenden MC-Prüfung ausgewirkt hatte, diese dafür aber in einem Prüfungsaufsatz schlechter abschnitten.<sup>24</sup> Die Prüflinge dagegen, die tiefenverarbeitende Lernstrategien verfolgt hatten, schnitten in der MC-Prüfung schlechter ab, waren dafür aber im Aufsatz besser. Diesen Hinweis, dass Studentinnen und Studenten sich MC-Prüfungen eher mit oberflächlicherem Lernen nähern als Prüfungen in Freitextformaten, konnte *Cilliers* auch für Studentinnen und Studenten der Medizin nachweisen.<sup>25</sup> Aufgrund dieser den MC-Formaten immanenten Stärken und Schwächen wird die Kombination mit anderen Aufgabenformaten empfohlen (Triangulation).

Die Kurzantwortfrage (SAQ) ist ein halboffenes Antwortformat, das die Prüflinge in der Regel mit wenigen Wörtern beantworten sollen. Die Fragen können auch klinische Fallvignetten beinhalten. Durch die eigenständige Textformulierung der Prüflinge werden Rate- und Cueing-Effekte vermindert. Zudem wird angenommen, dass bei der aktiven Antwortgenerierung speziell bei der Überprüfung höherer Wissens Ebenen andere mentale Prozesse ablaufen als bei der Beantwortung von geschlossenen Fragen. Allerdings wurde auch gezeigt, dass die Ergebnisse in Tests mit halboffenen und geschlossenen Fragen gut miteinander korre-

23 *Haladyna/Downing*, Validity of a Taxonomy of Multiple-Choice Item-Writing Rules, in: *Applied Measurement in Education* 2 (1989), Nr. 1, S. 51-78; *Krebs*, Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen für die ärztliche Ausbildung (Fn. 19).

24 *Scouller*, The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay, in: *Higher Education* 35 (1989), S. 453-472.

25 *Cilliers et al.*, The mechanism of impact of summative assessment on medical students' learning (Fn. 2).

lieren. Solche Ergebnisse sind jedoch stets mit Vorsicht zu betrachten, da sie stark von den im Experiment verwendeten Prüfungen und Kontext abhängen.<sup>26</sup>

Ein vorher festgelegtes strukturiertes Bewertungsschema für die Freitextantworten erhöht die Objektivität der Bewertung. Doch muss ein solches Schema beim ersten Einsatz zumeist noch ergänzt und erweitert werden, da häufig, insbesondere bei unscharfer Fragestellung, sehr kreative Lösungsvorschläge von Prüflingen generiert werden. Bei der Korrektur im Team sollte daher eine Absprache hierzu vor der abschließenden Bewertung vorgesehen werden.

Im Vergleich zu Langantwort-Formaten kann das Problem der mangelnden Generalisierbarkeit von medizinischen Fällen bei Kurzantwortfragen durch eine breitere Abdeckung des geforderten Wissens, d. h. mehr unterschiedliche klinische Situationen, gelöst werden.<sup>27</sup>

## 2. Klinisch-praktisches Prüfen: OSCE und arbeitsplatzbasiertes Prüfen

Um auch die dritte Kompetenzstufe der Miller-Pyramide „shows how“ (siehe Abb. 2) bei Studentinnen und Studenten erfassen und messen zu können, wurde 1975 das objektive, strukturierte klinische Examen (Objective Structured Clinical Examination, der OSCE)<sup>28</sup> eingeführt, welches inzwischen international als reliables und valides Format der Performanzprüfung eingesetzt wird. In einigen Ländern, z. B. den USA und der Schweiz, werden derartige Prüfungsparcours auch im Rahmen von summativen Staatsprüfungen verwendet. Idealerweise besteht ein OSCE aus mindestens zehn einzelnen Prüfungsstationen. Bei diesen Stationen werden in acht bis zehn Minuten reale Situationen nachgestellt und klinisch-praktische Aufgabenstellungen gelöst. In der Medizin ist es üblich, hierzu auch sogenannte Schauspielpatientinnen und -patienten sowie Modelle und Mannequins einzusetzen. Übliche Prüfungsinhalte sind dann beispielsweise das Führen eines Anamnese- oder Aufklärungsgesprächs, dem auch die Demonstration einer bestimmten Untersuchungstechnik oder klinischen Prozedur, wie dem Legen eines peripher-venösen Zugangs, folgen kann. Die Bewertung erfolgt anhand von sogenannten Checklisten, die entweder sehr detailliert oder eher global sein können: Detaillierte Checklisten ermöglichen auch relativ unerfahrenen Prüferinnen und Prüfern das standardisierte Bewerten bestimmter Fähigkeiten und Fertigkeiten. Globale Bewertungskriterien hingegen erfordern ein ausführliches Prüferinnen- und Prüfertraining,

26 *Norman/Swanson/Case*, Conceptual and methodological issues in studies comparing assessment formats, in: *Teaching and Learning in Medicine* 8 (1996), Nr. 4, S. 208-216.

27 *Elstein/Shulman/Sprafka*, *Medical Problem Solving: An Analysis of Clinical Reasoning*, Cambridge 1978.

28 *Harden et al.*, Assessment of clinical competence using an objective structured clinical examination (OSCE) (Fn. 16).

können dafür aber z. B. kommunikative und soziale Aspekte besser erfassen.<sup>29</sup> Trotz einer relativ aufwendigen Organisation (Räume für Prüfungsparcours, Arbeitszeit der Prüferinnen und Prüfer) und vergleichsweise hohen Kosten hat sich dieses Format in der Medizin nicht zuletzt wegen der starken lernsteuernden Wirkung durchgesetzt.

Unter der Annahme, dass für die Bewältigung einer bestimmten, praxisrelevanten Aufgabe das *Wissen* über Grundlagen und Prozesse eine Voraussetzung ist, und das tatsächliche *Ausführen und Demonstrieren* eine andere, vielleicht auch höhere Kompetenzstufe erfordert als die reine Reproduktion von Faktenwissen,<sup>30</sup> ist dieses Prüfungsformat eine wichtige Ergänzung zu den üblichen, im universitären Kontext verwendeten Formaten der Wissensabfrage. Bei der Erfassung impliziterer Inhalte, wie beispielsweise Einstellungen und Haltungen, oder auch einer tatsächlichen komplexeren Kompetenz<sup>31</sup> sind auch diesem Prüfungsformat Grenzen gesetzt.

Neben dem OSCE werden immer häufiger Methoden verwendet, mit denen die reale und tatsächliche Performanz am Arbeitsplatz erfasst und bewertet werden kann. Diese sogenannten arbeitsplatzbasierten Prüfungsmethoden werden zunehmend für die formative Bewertung, nur sehr selten für die summative Bewertung von Studentinnen und Studenten, eingesetzt. Auch in der ärztlichen Weiter- und Fortbildung werden arbeitsplatzbasierte Prüfungsmethoden genutzt, um gezielt und strukturiert Feedback geben und lernsteuernd wirken zu können.<sup>32</sup> In der Medizin haben sich verschiedene Formate etabliert, die sich oftmals nur in Details unterscheiden. Den meisten gemein ist eine strukturierte, mehr oder weniger detaillierte Checkliste für die Beobachtung und Bewertung vordefinierter Kriterien bestimmter Tasks in unterschiedlichen, realen Situationen.<sup>33</sup> So wird beispielsweise bei dem DOPS (Direct Observation of Procedural Skills) das Hauptaugenmerk auf die Performanz bei praktischen Fertigkeiten gelegt, beim Mini-CEX (Mini-Clinical Evaluation Exercise) hingegen vor allem auf strukturiertes Vorgehen an Patientinnen und Patienten und Gesprächsführung. Der Vorteil der Prüfungsmethoden am Arbeitsplatz ist die Möglichkeit der Beobachtung unter realen Bedingungen (ökologische Validität) – dies jedoch zugleich auf Kosten der Standardisierung und Vergleichbarkeit gemessener Leistungen. Dies ist auch der

29 Cohen/Rothman/Poldre/Ross, Validity and generalizability of global ratings in an objective structured clinical examination, in: *Academic Medicine* 66 (1991), Nr. 9, S. 545 ff.

30 Anderson/Krathwohl/Airasian (Hrsg.), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*, Boston 2000.

31 Hodges, Validity and the OSCE, in: *Medical Teacher* 25 (2003), Nr. 3, S. 250-254.

32 Hattie/Timperley, The Power of Feedback, in: *Review of Educational Research* 77 (2007), Nr. 1, S. 81-112.

33 Norcini/Burch, Workplace-based assessment as an educational tool: AMEE Guide No. 31, in: *Medical Teacher* 29 (2007), Nr. 9, S. 855-871.

Grund, warum die Methoden nur selten für summative Prüfungen heran gezogen werden.

### 3. Strukturierte mündliche Prüfung

Das mündliche Prüfungsgespräch ist ein traditionelles Prüfungsformat, das in der Medizin sowohl studienbegleitend wie auch in der Ärztlichen Prüfung und in Facharztprüfungen eingesetzt wird. Eine mündliche Prüfung als Prüferin bzw. Prüfer gut abzunehmen, darf u. a. wegen der knappen Wissensstichprobe einer zeitlich begrenzten mündlichen Prüfung als Herausforderung verstanden werden. Mehr als bei anderen Prüfungsformaten gilt hier: „We get what we assess, and if we don't assess it, we don't get it“.<sup>34</sup> Im Sinne der Prüfungsökonomie ist die knappe Zeit also sinnvoll aufzuteilen und sind Prüfungsinhalte sowie Prüfungsaufgaben detailliert vorzubereiten. Die Flexibilität und Dynamik eines Prüfungsgesprächs kann als Möglichkeit verstanden werden, im Sinne einer ökologisch validen Messung die Kompetenz der Prüflinge zu erfassen, gleichermaßen aber auch als Gefahr, das Prüfungsgespräch über Gebühr von den Prüflingen lenken zu lassen. Da für die Abfrage von Faktenwissen andere, ökonomischere Formate (z. B. Multiple Choice) existieren, lohnt es sich, die mündliche Prüfung für die Abfrage von Verständnis- und Begründungswissen zu reservieren, wofür schriftliche, offene Formate in puncto Interpretationsobjektivität keinen Vorteil bieten.

Von Prüferinnen und Prüfern wird während der Prüfung eine gewisse metakognitive Kompetenz gefordert, um sich in der Interaktion mit den Prüflingen möglicher Sympathien und Antipathien bewusst zu sein, um selbstreflektiert und -kritisch keinem Persönlichkeits-Bias aufzusitzen und Prüflinge weder wegen ihres sozialen Status noch in Bezug auf Geschlecht, ethnische Herkunft o. ä. zu diskriminieren. Soweit möglich, sollte aus Gründen der Justiziabilität die Anwesenheit einer Beisitzerin bzw. eines Beisitzers, besser aber eine weitere aktive Prüferin bzw. ein aktiver Prüfer gewährleistet werden und das Prüfungsgespräch dokumentiert werden. Dies gilt vor allem bei summativen Prüfungen und insbesondere, wenn Prüflinge den Anforderungen nicht genügen und somit das Prüfungsziel nicht erreichen. Die Definition von konkreten Aufgabenstellungen, Bewertungs- und Benotungskriterien erhöht die Transparenz und Aussagekraft der Prüfungsergebnisse.

34 *Lauren Resnick*, zitiert nach: *Wiggins*, The case for authentic assessment: Practical Assessment, Research & Evaluation 2 (1990), Nr. 2, <http://PAREonline.net/getvn.asp?v=2&n=2> (letzter Zugriff 18.12.2012).

Wie für alle anderen Prüfungseinsätze gilt auch für mündliche Prüfungen, dass Prüferinnen und Prüfer entsprechende Unterstützung und Schulungen durch ihre Fakultät erhalten sollten.<sup>35</sup>

#### 4. Evaluation und Qualitätssicherung

Zu guter Letzt sollten auch Prüfungen kontrolliert werden, und zwar vor der Durchführung durch inhaltlichen und formalen Aufgabenreview, während der Durchführung durch Anwesenheit geschulten Aufsichtspersonals und ggf. prüfungsbegleitende Evaluation durch die Prüflinge, und nach einer abgehaltenen Prüfung durch Korrektur der gestellten Aufgaben nach im Vorfeld der Prüfung definierten, transparenten Kriterien. Bei Freitextaufgaben oder Aufsatzarbeiten ist hierzu eine Verblindung der Bewerterinnen und Bewerter anzustreben, das heißt die inhaltskompetente, korrigierende Dozentin bzw. der inhaltskompetente korrigierende Dozent sollte die Identität der Prüflinge nicht kennen. Die Ergebnisse der Korrektur werden dann anhand der psychometrischen Kenndaten (Reliabilität, Aufgabenschwierigkeit, Trennschärfe), der studentischen Evaluation sowie Beobachtungen der korrigierenden Gutachterinnen und Gutachter validiert und ggf. korrigiert.

Für die Qualitätssicherung von Prüfungen stehen die gleichen Verfahren bzw. Erhebungsmethoden wie für die Lehrevaluation zur Verfügung, also online-basierte Evaluationen, papierbasierte Evaluationen und leitfadengestützte Gruppendiskussionen (Feedbackgespräche) mit den Studentinnen und Studenten sowie den jeweiligen Verantwortlichen. Während im Rahmen von online- und papierbasierten Evaluationen überwiegend quantitative bzw. semi-quantitative Fragenbogendesigns (Items auf der Basis einer Likert<sup>36</sup>- oder Schulnotenskaliierung, teilweise ergänzt durch kurze offene Fragen zu Verbesserungspotenzialen) eingesetzt werden, können mittels strukturierter Feedbackgespräche spezifische bzw. mehrschichtige Probleme beispielsweise bei der Implementierung neuer Prüfungsformate im Detail erörtert werden.<sup>37</sup> Da strukturierte Gruppendiskussionen jedoch einen deutlich erhöhten Ressourcenverbrauch (Zeit und Personal) beinhalten, sollte dieses Instrument mit entsprechender Zurückhaltung verwendet werden, z. B. bei der Einführung neuer Prüfungsformate oder ganzer Prüfungskonzepte.

35 *Davis/Karunathilake*, The place of the oral examination in today's assessment systems, in: *Medical Teacher* 27 (2005)9, Nr. 4, S. 294-297.

36 Z. B. ein skaliertes Kontinuum von „Aussage X trifft voll zu“ bis „Aussage X trifft überhaupt nicht zu“.

37 *Kern/Thomas/Hughes*, Curriculum Development for Medical Education: A Six-Step Approach, Baltimore 2009.

Die Qualitätskriterien sollten vor der Prüfung festgelegt und in geeignete Fragebogenitems/Qualitätsindikatoren überführt werden.<sup>38</sup> Folgende Kriterien können unter anderem zur Qualitätssicherung einer Prüfung herangezogen werden: Organisation/Administration, Verfügbarkeit von Ansprechpartnern und Informationen zur Prüfung, Kongruenz von Lernzielen mit Prüfungsinhalten, Fairness der Prüfung, Zufriedenheit mit der Prüfung.

#### *IV. Künftige Entwicklungen, Kompetenzbasiertes Prüfen, NKLM*

Zurzeit werden fast alle Prüfungen in der Medizin mittels Multiple Choice, mündlicher Prüfungen und OSCE abgenommen.<sup>39</sup> In der näheren Zukunft könnten der vermehrte und gezieltere Einsatz formativer, lernsteuernder Prüfungen sowie arbeitsplatzbasierte Prüfungen die immer professionelleren Prüfungskonzeptionen weiter ergänzen. Solche Weiterentwicklungsprozesse sollten gemeinsame, standortübergreifende Qualitätskriterien berücksichtigen und eine evidenzbasierte Weiterentwicklung der bereits existierenden Leitlinie für fakultätsinterne Leistungsnachweise während des Medizinstudiums der Gesellschaft für Medizinische Ausbildung nach sich ziehen.<sup>40</sup>

Unterstützung erhalten Prüferinnen und Prüfer von Fakultätsentwicklungsprogrammen, die auf lokaler Ebene bereits zu zunehmender Prüfungsexpertise geführt haben. Interuniversitäre Kooperationen und eine zunehmend bessere informationstechnologische infrastrukturelle Ausstattung bieten neue Möglichkeiten und erlauben inzwischen auch computerunterstützte Prüfungen im summativen Kontext und den Austausch von Prüfungsaufgaben zwischen Fakultäten. Die Nutzung moderner Medien ist jedoch nicht auf schriftliche Prüfungsformate beschränkt: So sind u. a. auch computergestützte OSCE-Stationen beschrieben worden<sup>41</sup> und an der Medizinischen Fakultät der Universität Heidelberg werden computerunterstützte Key Feature-Prüfungen durchgeführt.<sup>42</sup> Auch Variationen der etablierten

38 *Fabry*, *Medizindidaktik - Ein Handbuch für die Praxis*, Bern 2008; *Kern et al.*, *Curriculum Development for Medical Education: A Six-Step Approach* (Fn. 37).

39 *Möltner et al.*, *Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten* (Fn. 11).

40 *Gesellschaft für Medizinische Ausbildung/Kompetenzzentrum Prüfungen Baden-Württemberg/Fischer*, *Leitlinie für Fakultäts-interne Leistungsnachweise während des Medizinstudiums*, in: *GMS Zeitschrift für Medizinische Ausbildung* 25 (2008), Nr. 1, Doc74.

41 *Koca/Hege/Fischer*, *Herz- und Lungenauscultation: Computergestütztes Lernen als Vorbereitung auf eine klinisch-praktische Prüfung*, in: *Kundt/Bernauer/Fischer/Haag/Klar/Leven/Matthies/Puppe* (Hrsg.), *eLearning in der Medizin und Zahnmedizin*, Aachen 2007, S. 188-197.

42 *Kompetenzzentrum für Prüfungen in der Medizin/Baden-Württemberg - Komp Med* <http://www.medizinische-fakultaet-hd.uni-heidelberg.de/Computergestuetzte-Pruefungen.100385.0.html> (letzter Zugriff 15.12.2012).

Formate, z. B. OSVE (objektivierte, strukturierte, videobasiertes Examen) finden immer mehr Verbreitung.

Die Ausarbeitung eines nationalen, kompetenzbasierten Lernzielkataloges Medizin<sup>43</sup> könnte eine wesentliche Vorarbeit auch für zukünftige, kompetenzorientierte Prüfungen in der Medizin sein und die Debatte nach den eigentlichen Ausbildungszielen humanmedizinischer Studiengänge an vielen Standorten anstoßen.

43 *Hahn/Fischer*, Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT), in: GMS Zeitschrift für Medizinische Ausbildung 26 (2009), Nr. 3: Doc35.