

Computerbasiert prüfen

Möglichkeiten und Grenzen

Computer-based Assessment

Potentials and Drawbacks

• Peter Frey¹

Zusammenfassung:

Computerbasierte Prüfungen im Medizinstudium eröffnen neue Möglichkeiten. Vorteile solcher Prüfungen liegen im sequentiellen oder adaptiven Prüfen, in der Integration von Bewegtbildern oder Ton, der raschen Auswertung und zentraler Verwaltung der Prüfungsfragen via Internet. Ein Einsatzgebiet mit vertretbarem Aufwand sind Prüfungen mit mehreren Stationen wie beispielsweise die OSCE-Prüfung. Computerbasierte formative Selbsttests werden im Bereiche e-learning häufig angeboten. Das hilft den Lernenden ihren Wissensstand besser einzuschätzen oder sich mit den Leistungen anderer zu vergleichen.

Grenzen zeigen sich bei den summativen Prüfungen beim Prüfungsort, da zuhause Betrug möglich ist. Höhere ärztliche Kompetenzen wie Untersuchungstechnik oder Kommunikation eigenen sich kaum für rechnergestützte Prüfungen.

Schlüsselwörter: Medizinische Ausbildung, Rechner, Prüfung, online, Selbsttest, Kompetenzen

Abstract:

Computer-based testing in medical education offers new perspectives.

Advantages are sequential or adaptive testing, integration of movies or sound, rapid feedback to candidates and management of web-based question banks. Computer-based testing can also be implemented in an OSCE examination. In e-learning environments formative self-assessment are often implemented and gives helpful feedbacks to learners.

Disadvantages in high-stake exams are the high requirements as well for the quality of testing (e.g. standard setting) as additionally for the information technology and especially for security.

Keywords: medical education, computer-based, online testing, self-assessment, high-stake examination, competence

Einleitung

Online Prüfungen scheinen verlockend - die Medizinischen Fakultäten müssen durch die neue Approbationsordnung für Ärzte (ÄappO) immer mehr Prüfungen bei sinkendem Budget durchführen. Computerbasierte Prüfungen scheinen ein effizientes Werkzeug zur rationalen Prüfungsabwicklung darzustellen.

Schriftliche Prüfungen werden schon heute häufig mit elektronisch lesbaren Fragebogen durchgeführt. Die Daten werden anschließend mittels Rechner ausgewertet. Unter computerbasierten Prüfungen werden nachfolgend solche Prüfungen verstanden, bei denen bereits die Dateneingabe am Rechner online erfolgt [1]. Die Daten werden über ein Netzwerk, meist das Internet, an einen zentralen Rechner übermittelt und dort ausgewertet.

Welches Potenzial bieten also computerbasierte Prüfungen? Werden die Möglichkeiten ausgenutzt und die Grenzen beachtet?

Grundsätzlich muss bei Prüfungen (assessment) zwischen formativer und summativer Prüfung unterschieden werden. Die *formativen Prüfungen* werden auch als Selbst-Beurteilung (self-assessment) bezeichnet und dienen der Kontrolle des Lernerfolges ohne sanktionierende Maßnahmen, also ohne Entscheid über Bestehen/Nichtbestehen. Viele web-basierte Lernprogramme (e-learning) [2] beinhalten bereits Selbstkontrollen, denen sich der Studierende unterziehen kann und die stimulierend für das weitere Lernen sein

sollen. Der Vorteil von formativen online Prüfungen ist eine korrekte und rasche Auswertung der Antworten. Idealerweise erhält der Lernende zusätzlich Begründungen der Antworten. Formative Prüfungen haben aber meist eine schlechte testmethodische Qualität. Es können irrelevante oder nicht repräsentative Inhalte geprüft werden. Meist werden nur Multiple Choice Questions (MCQ) mit Einfachwahlantworten verwendet, da diese schnell geschrieben werden können. Die Fragen werden selten durch einen Review-Prozess validiert. Auch wird dem Testkandidaten selten sein Leistungsniveau im Vergleich zu den übrigen Kandidaten aufgezeigt.

Summative Prüfungen (high stake exams) beinhalten einen Bestehen-/Nichtbestehens-Entscheid mit sanktionierenden Maßnahmen wie Wiederholung der einzelnen Prüfung, eines ganzen Studienjahres oder Ausschluss aus dem Studium. Summative Prüfungen müssen daher bezüglich der Prüfungsinhalte relevant und valide sein und eine hohe Verlässlichkeit im Sinne der Reliabilität aufweisen [3]. Zudem darf kein Betrug durch die Studierenden möglich sein. Dies gilt auch für summative computerbasierte Prüfungen, auf die im Folgenden etwas näher eingegangen werden soll.

Methoden

• Anwendungsgebiete von rechnergestützten Prüfungen

Miller [4] teilt die ärztlichen Kompetenzen in vier Bereiche (siehe Abbildung 1).

¹ Universität Bern, Institut für Medizinische Lehre IML, Abteilung für Unterricht und Medien (AUM), Bern, Schweiz

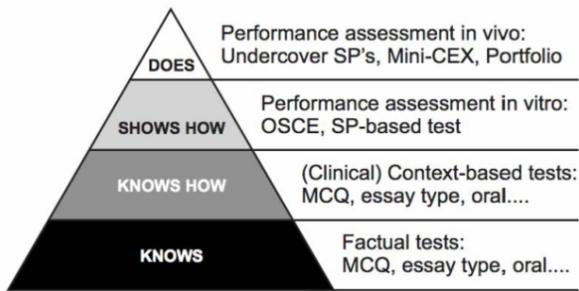


Abbildung 1: Kompetenzen nach Miller und geeignete Prüfungen

Der Arzt

- weiß es (knows)
- weiß wie (knows how)
- zeigt wie (shows how)
- handelt (does).

Jedem Kompetenzbereich können geeignete Prüfungsmethoden zugeordnet werden. Für die *Bereiche Wissen (knows) und Handlungswissen (knows how)* können computerbasierte Prüfungen sinnvoll sein. Für das reine Faktenwissen sind beispielsweise MCQ-Prüfungen am sinnvollsten, da sie sehr zuverlässig messen und mit vertretbarem Aufwand hergestellt werden können. Handlungswissen im Sinne der klinischen Entscheidungskompetenz (clinical reasoning) wird meist als klinische Situationsbeschreibung (,Vignette') mit Mehrfachantworten abgefragt. Dazu wurden das Prüfungsverfahren [5] der Patient Management Problems (PMP) und die ,Schlüsselprobleme' (key features) entwickelt. Die PMP haben sich für summative Prüfungen als nicht geeignet erwiesen [3], obwohl sie sich gut als computerbasierte Prüfungen eignen würden. Deshalb liegt die Hoffnung nun bei den Fragetypen der key features.

Für Handlungskompetenzen (shows how, does) wie Anamnese, Untersuchungstechnik oder gar Kommunikation eignen sich computerbasierte Prüfungen weniger. Dazu bieten Prüfungsmethoden mit Schauspielpatienten wie objective structured clinical examination (OSCE) oder klinik-orientierte Methoden [6] wie die Mini clinical evaluation exercise [7] (Mini-CEX) oder Portfolio bessere Voraussetzungen (siehe Tabelle 1).

Tabelle 1

Mehrwert computerbasierter Prüfungen

- Sequentielles oder adaptives Prüfen
- Online Verwaltung der Fragen
- Rasche Auswertung
- Integration von Multimedia wie Ton, Bewegtbildern oder Simulation
- Prüfgegenstand ist selber online (Pubmed)

Computerbasierte online Prüfungen können theoretisch folgende Vorteile gegenüber den schriftlichen, computerlesbaren Formularen bieten.

Sequentielles Prüfen wird möglich. Alle Prüflinge bekommen zuerst einen gemeinsamen Pool von Fragen. Die schlechten und die guten Prüflinge bekommen keine zusätzlichen Fragen und gelten als bestanden/nicht bestanden. Kandidaten im Mittelfeld, um die Bestehensgrenze, bekommen nochmals Fragen mit mittlerem Schwierigkeitsgrad. So lässt sich eine bessere Trennung erzielen und einige Prüflinge haben eine kürzere Prüfungszeit.

Bei *adaptiven Tests* werden die Fragen den Leistungen der Kandidaten individuell angepasst. Die Entwicklung solcher zweistufiger Prüfungen ist aufwändig und der Gewinn gegenüber einstufigen Prüfungen mit vielen Fragen (200 im MCQ-Stil) verschwindend klein.

Ein weiterer Vorteil kann die *online Redaktion der Fragen* sein. Dieser Vorteil gilt allerdings auch für die datenbankbasierte Verwaltung von Fragen (,itembank') für Prüfungen auf Papier. Die online Verwaltung von Fragen könnte dazu führen, der Qualität der Fragen ungenügend Beachtung zu schenken. Ein seriöser Validierungsprozess der Fragen wird den Prüfungsverantwortlichen auch im Internetzeitalter nicht erspart. Selbstverständlich müssen ein hoher Datenschutz und Sicherheit des Zugriffs nur für Berechtigte gewährleistet sein.

Computerbasierte Prüfungen können *theoretisch sofort ausgewertet* werden nachdem alle Kandidaten die Antworteingabe abgeschlossen haben. Die unmittelbare Rückmeldung über Bestehen/Nichtbestehen bedingt allerdings, dass die Bestehensgrenzen im Voraus festgelegt worden sind (standard setting). Die Festlegung eines starren minimalen Prozentsatzes richtiger Antworten (z.B. 60 %) ist für summative Prüfungen ebenso fragwürdig wie eine auf die Gruppenleistung bezogene Grenzsetzung (z.B. eine Standardabweichung unterhalb des Mittelwertes). In beiden Fällen hängt die Erfolgchance eines Kandidaten nicht ausschließlich von seiner Leistungsfähigkeit ab, sondern ist im ersten Fall abhängig von der Prüfungsschwierigkeit und im zweiten Fall von der Leistungsfähigkeit der übrigen Kandidaten. Das Problem ist nur mit einem inhaltsorientierten Verfahren zu lösen (Angoff, Ebel, Nedelsky) [8]. Leider sind solche Verfahren ziemlich aufwändig und werden deshalb noch viel zu selten eingesetzt. Je nach Verfahren ist eine unmittelbare Auswertung gar nicht möglich.

Weitere Einsatzgründe computerbasierter Prüfungen bestehen in der *Integration von Bewegtbildern* (Video, Animation, Simulation) oder *Ton* (Herz- oder Lungengeräusche). Dies ist auf Papier nicht möglich. In Bern geschieht dies im Rahmen einer OSCE-Prüfung im dritten Studienjahr mit insgesamt 20 Stationen, wobei 14 Stationen mit Schauspielpatienten bestückt sind (,shows how'). Ein Vorteil von OSCE-Prüfstationen ist die kleine Anzahl Rechner, die eingesetzt werden muss, da die Kandidaten von Posten zu Posten rotieren. Für die Herzauskultation wurde zuerst ein umfangreiches CD-ROM-basiertes Lernprogramm im Rahmen einer Dissertation erstellt. Davon ausgehend wurde ein Auskultations-Trainer mit Prüfmodul der 10 wichtigsten Pathologien für das Betriebssystem MacOS erstellt. Der Studierende beurteilt Schritt für Schritt Töne und Geräusche und stellt eine Diagnose. In 8 Minuten mussten zwei Befunde interpretiert werden (siehe Abbildung 2).

Der Auskultations-Trainer

Fahren Sie mit der Maus über die Liste mit den Auskultationsbefunden. Im Kopfhörer hören Sie sofort das entsprechende Tonbeispiel. Studieren Sie die graphische Beschreibung. Wenn Sie glauben, die Herz- und Geräusche erkennen zu können, gehen Sie zum Prüfungsmodus des Programmes und versuchen Sie sich daran.

An der Medizinischen Fakultät der Universität Bern erfolgt die "richtige Prüfung" genau mit diesem Programm.

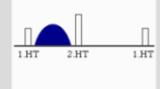
Normbefund (Herzbasis)			valvuläre Pulmonalstenose (VPS) 
3. HT laut	Auskultationsort: ZL		
Aorteninsuffizienz	1. Herzton - Lautstärke: normal - Spaltung: normal (= keine oder enge)	Systole Extraton - kein Extraton	Diastole Extraton - kein Extraton
Aortenstenose	2. Herzton - Lautstärke: normal - Spaltung: keine	Geräusch - früh-mesosystolisch - spritzförmig - leifrequent - 4/6	Geräusch - kein Geräusch
Aorteninsuff. & -stenose			
Mitralinsuffizienz			
Mitralklappenprol. & MI			
Mitralstenose			
Mitralsten. & -insuffiz.			
Pulmonalsten. (valvular)			
offener Ductus Botalli			
Perikardreiben			

Abbildung 2: Auskultationstrainer als Grundlage für computerbasierte Prüfungsstation im OSCE

Die pädiatrische Untersuchungstechnik im Rahmen eines OSCE-Parcours bei 150 Kandidaten kann kaum an Kleinkindern durchgeführt werden, da Kleinkinder rasch ermüden oder die Kooperation schwierig ist. Als Ersatz für echte Patienten wurden den Kandidaten acht Videoclips gezeigt. Bei jeder gezeigten Untersuchung hatte der Kandidat zu beurteilen, ob diese richtig oder falsch durchgeführt wurde und falls inkorrekt, kurz zu begründen was falsch war. Der Computer diente allerdings nur der Wiedergabe der Videoclips. Die Beurteilung wurde auf einem rechnerlesbaren Formular erfasst und die Freitexte nachträglich von Experten bewertet (siehe Abbildung 3).



Abbildung 3: OSCE-Station mit Videoclips zur pädiatrischen Untersuchungstechnik

Die *online Literatursuche* in Pubmed ist für jedermann möglich geworden. Die Literatursuche mit Pubmed, also Kompetenzen, die auch in der Realität nur im Umgang mit Rechnern erfolgen, sind weitere geeignete Anwendungen von online Prüfungen. Die wichtigsten Werkzeuge von Pubmed wie limits, history, clipboard, MeSH database und clinical queries werden deshalb in einem zweistündigen hands-on Kurs im 3. Studienjahr in Bern trainiert und am Semesterende an einer OSCE-Station geprüft. Die Kandidaten bekommen eine klinische Fragestellung und sollen dazu in acht Minuten Literatur mit hoher Evidenz (RCT, Systematic Reviews oder Metaanalysen) suchen. Dies geschieht online mit Zugriff auf die Datenbank der National Library of Medicine (NML).

Verbindungsunterbrüche sind selten (einmal täglich) und dauern nur 1-2 Minuten. Der Studierende drückt die gefundenen Artikel der ersten Seite aus. Die Bewertung erfolgt nach Kriterien wie: korrekte englische Suchbegriffe, gefundene Artikel oder Evidenz.

• Schwierigkeiten von summativen computerbasierten Prüfungen

Eine Herausforderung bei *summativen Prüfungen* stellt sicher die Verhinderung des Plagiates dar. Wie können 100-300 Kandidaten geprüft werden ohne Möglichkeiten zum Betrug?

Bei beispielsweise 140 Studierenden müsste die Prüfung an einem *zentralen Prüfungsort* mit 70 Rechnern mehrmals durchgeführt werden. Um Betrug zu verhindern, müssten die Fragen ausgetauscht oder die Weitergabe durch Abschottung der Kandidatengruppen minimiert werden. Ein weiteres Problem stellt das Netzwerk dar: Es dürfen keine längeren Ausfälle auftreten, sonst muss die Prüfung wiederholt oder verlängert werden. Die Daten müssen auch sicher übermittelt werden.

Ein grosser zentraler Rechnerraum wäre für Schulungen ungeeignet: Die PC's würden die meiste Zeit ungenutzt herumstehen. Studierenden lernen heute lieber mit eigenem Notebook mit Funkkontakt (WiFi) zum Internet. Sie sind zunehmend weniger auf große Rechner-Räume angewiesen. Wegen des raschen Technologiewandels müssten die PC's nach zwei bis drei Jahren bereits wieder ersetzt werden. Mit computerlesbaren schriftlichen Fragebogen ist man flexibler in der Wahl des Prüfungsortes.

Wie sieht die Situation in den USA aus?

Die United States Medical Licensing Examination (USMLE) führen summative computerbasierte Prüfungen in mehreren privaten Computer-Zentren mit proprietärer, verschlüsselter Software durch [9][10][11][12][13]. Die Resultate werden aber erst nach einigen Wochen verschickt. Die Prüfungsfragen werden dabei zwar über das Internet verteilt, sind aber nicht zugänglich, so dass eher von einem Intranet gesprochen werden müsste. Soll das nun als online Prüfung verstanden werden?

Als Alternative scheint ein *dezentrales Prüfen zuhause* am eigenen Rechner verlockend. Trotz Passwortkontrolle kann damit aber ein *Betrug nicht verhindert* werden. Erstens kann nicht sicher überprüft werden, wer die Antworten am heimischen Rechner eingibt. So können die Kandidaten sich Hilfe bei älteren Kommilitonen oder anderen Experten holen. Zweitens hätten die Kandidaten freien Zugang zu Internet (Wissensabfrage) und e-mail. Es müsste auch sichergestellt werden, dass alle Rechner die gleichen *minimalen technischen Anforderungen* wie Hardware, Software und Verbindungsgeschwindigkeit zum Internet erfüllen.

Einfacher sieht die Situation im Rahmen von *Prüfungen mit Stationen wie dem OSCE* aus. Bei zwei parallelen Parcours genügen lediglich zwei Rechner, sowie ein bis zwei Ersatzgeräte.

Folgende Computerstationen waren bisher in Bern beim OSCE im Einsatz:

- Herz- oder Lungenauskultation (Töne, siehe Abbildung 2)
- Histologie (1./2. Jahr) oder Hämatologie (3. Jahr)

- Pädiatrische Untersuchungstechnik (Video, siehe Abbildung 3)
- Literatursuche online mit Pubmed.

Wie bei allen Prüfungen sollte man bei der Planung computerbasierter Prüfungen immer auch ein Auge auf mögliche *Prüfungssprachen* richten. Ist die Prüfung inhaltlich und technisch für alle Studierenden gerecht? Gab es bei computerbasierten Prüfungen keine zusätzlichen Hürden [14] oder dann entsprechende Hilfestellungen wie beispielsweise online Mustertests?

• Formative Selbsttests (Self-Assessment)

Für *formative Tests* sieht die Situation ganz anders aus. Selbstbezug ist kein Thema, respektive jedem Studierenden selber überlassen. Die Hochschulen haben sich zum Ziel gesetzt, Studierende zu lebenslangen Lernern (life long learners) auszubilden.

Dazu gehören zwei Fähigkeiten [15]

- Selbstbeurteilung (self-assessment) von Lücken der eigenen Kompetenz
- Selbstständiges Lernen (self-directed learning) zur Verbesserung der Kompetenz.

Umfangreiche Studien [15][16][17] kommen zu einem ernüchternden Resultat. Weder Studierende noch praktizierende Ärzte können ihre Lerndefizite respektive ihren Wissensstand realistisch einschätzen. In der Studie an der McMaster-Universität [15] wurde die Selbsteinschätzung der Studierenden bezüglich eines dreimal jährlich stattfindenden Progress-Testes evaluiert. Die Studierenden konnten ihre Leistungen in den Tests nicht realistisch einschätzen respektive voraussagen. Erstaunlicherweise wurde dies auch nach zwei Jahren nicht besser, obwohl die Studierenden dann bereits sechs solcher Tests absolviert hatten und ein Trainingseffekt in der Selbstbeurteilung erwartet werden dürfte.

Welche Schlüsse für computerbasierte Selbsttests können wir daraus ziehen?

Selbstkontrollen (auch auf Papier) sind wichtig und helfen den Studierenden, ihre Leistung zu objektivieren. Es ist deshalb sicher sinnvoll, den Lernenden ein Instrument zur Beurteilung ihrer Leistung in die Hand zu geben. Es stellt sich dann die Frage, wie hoch die *Motivation* zur Teilnahme an solchen Lernkontrollen ist. Am höchsten ist sie, wenn die Selbsttests dasselbe Format (Inhalte, ‚Medium‘) wie die echten Prüfungen aufweisen. Das Institut für Medizinische Lehre in Bern (Schweiz) bietet seit Jahren einen begrenzten online Zugang zu einem Teil der echten MC-Prüfungsfragen der Schweizerischen Staatsexamina an, im Sinne des *Self-Assessments* (<http://www.iawf.unibe.ch/self-assessment/>). Die Aktivitäten der 1000 Kandidaten sind auf dem Server unmittelbar vor den Prüfungen so hoch, dass dieser in den vergangenen Jahren auch schon mal seine Funktion aufgab. Für dieses Self-Assessment werden ausschließlich Fragen verwendet, die bereits am Staatsexamen eingesetzt wurden. Damit kommen nur validierte Fragen zum Einsatz, und die Teilnehmenden können ihre Leistung mit der mittleren Leistung früherer Staatsexamenskandidaten vergleichen. Sie können sich dabei auch nur in Teilbereichen oder mehrmals testen. Die Fragen werden vom Server jeweils zufällig zusammengestellt. Falsch beantwortete Fragen können nochmals

angeschaut werden. Die korrekte Lösung wird aber bewusst nicht bekannt gegeben, um oberflächlichem Auswendiglernen isolierter Einzelfakten entgegen zu wirken (siehe Tabelle 2).

Tabelle 2

Formative computerbasierte Selbsttests werden geschätzt, wenn sie

- der Vorbereitung auf summative Prüfungen dienen
- (prüfungs)relevante Themen abdecken
- bildorientierte Fragen beinhalten (Dermatologie, Radiologie, Ophthalmologie, Pathologie)
- guten Feedback mit Begründung geben
- ein spielerisches Format aufweisen (Simulation, hotspot, drag and drop)

Wenig Interesse besteht, wenn sowohl das Prüfungsformat wie die Inhalte von formativen und summativen Prüfungen weit auseinander liegen. Selbsttests mit MC-Fragen (Detailwissen) dienen kaum der Vorbereitung auf mündliche Prüfungen mit Einbezug von Patienten.

Alle heute eingesetzten Lernplattformen für e-learning, auch learning management systems (LMS) genannt, bieten auch dem technischen Laien die Möglichkeit, Selbsttests zu implementieren. Der Aufwand für die Dozierenden und der Nutzen für die Studierenden sollte dringend näher erforscht werden.

Noch besser als Selbsttests sind regelmäßige summative Prüfungen mit verschiedenen Prüfungsformen, da so ein klarer Anreiz zum vielseitigen Lernen gegeben wird.

Schlussfolgerungen

Computerbasierte Prüfungen können in formative und summative, in lokal netzwerkbasierte oder web-basierte eingeteilt werden. Für rechnergestützte Prüfungen eignen sich vorwiegend ärztliche Kompetenzen im Bereiche des Wissens (knows, knows how).

Für summative Prüfungen mit Konsequenzen stellen sich hohe Qualitätsansprüche sowohl an die Prüfung selber als auch zusätzlich an die Technologie. Der Aufwand des computerbasierten Staatsexamens (UMSLE) in den USA ist eindrücklich bis erschreckend.

Ein Einsatzgebiet mit vertretbarem Aufwand sind Prüfungen mit mehreren Stationen wie die OSCE-Prüfung. Bern hat damit gute Erfahrungen gemacht.

Elektronische Selbsttests für Studierende machen pädagogisch Sinn. Studien haben gezeigt, dass die Studierenden und Ärzte ihren Wissensstand schlecht einschätzen können. Die Lernplattformen der Universitäten bieten dazu geeignete Lösungen an. Unklar bleibt, wieweit die Masse der Lernenden diese Angebote auf freiwilliger Basis auch tatsächlich nutzt.

Danksagung

Das Manuskript wurde freundlicherweise durchgesehen und ergänzt von Dipl. Psych. René Krebs, Abteilung für Assessment and Evaluation (aae) des Institutes für Medizinische Lehre der Universität Bern.

Korrespondenzadresse:

• Dr. med. et. MME Peter Frey, Universität Bern, Institut für Medizinische Lehre IML, Abteilung für Unterricht und Medien (AUM), Inselspital 38, CH-3010 Bern, Tel.: +41 31 632 25 13, Schweiz
frey@iml.unibe.ch

Literatur:

- [1] Cantillon P, Irish B, Sales D. Using computers for assessment in medicine. *BMJ*. 2004;329(7466):606-609.
- [2] Hammoud MM, Barclay ML. Development of a Web-based question database for students' self-assessment. *Acad Med*. 2002;77(9):925.
- [3] Schuwirth LW, van der Vleuten CP. ABC of learning and teaching in medicine: Written assessment. *BMJ*. 2003;326(7390):643-645.
- [4] Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):63-67.
- [5] Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. *Med Educ*. 2003;37(Suppl 1):65-71.
- [6] Norcini JJ. Work based assessment. *BMJ*. 2003;326(7392):753-755.
- [7] Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. 2003;138(6):476-481.
- [8] Bloch R, Hofer D, Krebs R. Handbuch 'Kompetent prüfen'. Bern: Abteilung für Ausbildungs- und Examensforschung (AEE). Zugänglich unter <http://www.iawf.unibe.ch/>.
- [9] Clauser BE, Margolis MJ, Swanson DB. An examination of the contribution of computer-based case simulations to the USMLE step 3 examination. *Acad Med*. 2002;77(10 Suppl):80-82.
- [10] Guagnano MT, Merlitti D, Manigrasso MR, Pace-Palitti V, Sensi S. New medical licensing examination using computer-based case simulations and standardized patients. *Acad Med*. 2002;77(1):87-90.
- [11] Dillon GF, Boulet JR, Hawkins RE, Swanson DB. Simulations in the United States Medical Licensing Examination (USMLE). *Qual Saf Health Care*. 2004;13(Suppl 1):41-45.
- [12] Dillon GF, Clyman SG, Clauser BE, Margolis MJ. The introduction of computer-based case simulations into the United States medical licensing examination. *Acad Med*. 2002;77(10 Suppl):94-96.
- [13] Robert Galbraith, Peter Scoles, Stephen Clyman. National Board of Medical Examiners: Online Assessment in Medical Education: Emerging Experience in the United States Conference, eLearning Results 2004'. Sestri: Levante; 2004. Zugänglich unter: <http://www.elearningresults.com/agenda4.html>.
- [14] Peterson MW, Gordon J, Elliott S, Kreiter C. Computer-based testing: initial report of extensive use in a medical school curriculum. *Teach Learn Med*. 2004;16(1):51-59.
- [15] Eva KW, Cunnington JP, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract*. 2004;9(3):211-224.
- [16] Fitzgerald JT, White CB, Gruppen LD. A longitudinal study of self-assessment accuracy. *Med Educ*. 2003;37(7):645-649.
- [17] Day SC, Norcini JJ, Webster GD, Viner ED, Chirico AM. The effect of changes in medical knowledge on examination performance at the time of recertification. *Proceedings of the Annual Conference of Research*. *Med Educ*. 1998;27:139-144.