

» Prüfungsstrukturen für eine erfolgreiche Studienreform

Yngve Falck-Ytter

Case Western Reserve University, USA

Zusammenfassung: Hintergrund: Seit der Einführung der Experimentierklausel sind alternative Evaluationsformen anstatt den bisherigen Prüfungen des IMPP notwendig geworden. Die Entwicklung und Auswahl von Prüfungsformen und Inhalten hat eine zentrale Bedeutung, da alle Prüfungsordnungen eine studiensteuernde Wirkung haben. Prüfungen sind das heimliche Curriculum einer Fakultät. Bei jedem Reformprojekt muss daher sorgfältig darauf geachtet werden, dass durch die Prüfungsgestaltung die Reformansätze nicht wieder zunichte gemacht werden. Ein Fehler in der Vergangenheit war die kritiklose Übernahme von MCQs in den 70er Jahren ohne begleitende wissenschaftliche Entwicklung. Ihr Scheitern führte paradoxerweise zur Wiedereinführung der unstrukturierten mündlichen Prüfung. **Ziel:** Prüfungen sollen weniger als Selektionsinstrument, als zur studentischen Standortbestimmung und Curriculumentwicklung eingesetzt werden. Doch wie können wir wissenschaftlich fundierte Prüfungsstrukturen entwickeln? Zunächst müssen drei Fragen gestellt werden: Warum wollen wir prüfen? Wollen wir formativ oder summativ prüfen? Was wollen wir prüfen? Wie kann man sinnvolle Lernziele definieren? Wie wollen wir prüfen? Welche Instrumente können wir einsetzen? **Ergebnisse:** Die Gefahr, Fehler zu wiederholen ist stets gegeben. Deshalb ist es notwendig, die Methoden im Lichte der historischen Entwicklung zu betrachten, wie z. B. OSCEs vs. Standardized Patient Tests etc. und ihre Vor- und Nachteile zu kennen. Die Auswirkungen der Prüfungen im Hinblick auf das Studium müssen berücksichtigt werden. Bei der Auswahl muss auf die Qualitätsmerkmale Validität und Reliabilität sowie Praktikabilität und Kosten geachtet werden. **Schlussfolgerungen:** 1. Eine einzelne, wenngleich auch geeignete Prüfungsmethode kann nicht das ganze Spektrum ärztlicher Kompetenz evaluieren. 2. Man erreicht mit einer Evaluationsmethode nur soviel, wie man bereit ist, in ihre kontinuierliche Entwicklung und Durchführung zu investieren. 3. Validierung der Testinstrumente wird zur wichtigsten Forschungsaufgabe wenn man anstatt normbezogener Kriterienbezogene Prüfungen anwendet.

Examinations – How to Successfully Approach a Reform in Medical Education in Germany: The introduction of the so-called „Experimentierklausel“, a new clause within the governmen-

tally fixed medical educational system in Germany, now allows individual medical schools to choose and organize their own curriculum and assessment system on an experimental basis. Therefore there is a great need for a reform of the current assessment system, which has largely relied on poorly designed multiple choice question (MCQ) examinations since this method was introduced in the 70's. Paradoxically, once it became clear, that these MCQ were poorly constructed and had a negative impact on student's way to study, unstructured oral examinations were reintroduced additionally instead of discussing ways to improve the MCQ or look for more suitable alternatives. **Aims:** Since assessment has a profound effect on the way students learn, careful planning is necessary. Examinations should be used less as a way to eliminate but rather to help students develop insight and to evaluate the curriculum. To start the development process, three questions should be asked in regard to assessment: Why to examine? Should we focus on summative or formative assessment? Should we create a hurdle jumping approach or focus on helping the student to do better. What to examine? How can we develop reasonable exit objectives? How to examine? Which assessment tools are available? Will they assess factual recall only or also include critical reasoning and communication skills? **Results:** In the past 20 years many methods of clinical examination have been described like the Objective Structured Clinical Examinations and Standardized Patient examinations with good reliability and far better validity than traditional methods (especially unstructured oral examinations). Practicability aspects have also been covered. It is, however, necessary to be familiar with advantages and disadvantages, as well as understand details about reliability and validity measures before deciding whether and how to implement new assessment techniques. **Conclusion:** (1) There is not a single method available to assess all aspects of clinical competence. (2) Continued development of assessment instruments is necessary to ensure proper results. (3) Validating examinations are an important task, especially if a criterion- instead of norm-referenced approach is used. Adequate standard-setting procedures need to be implemented.

Key words: Evaluation – Curriculum – Tests – MCQ – OSCE

Hintergrund

Obwohl seit über 10 Jahren an einer Reform des Medizinstudiums gearbeitet wird, ist es kürzlich durch die Einführung einer sog. Experimentierklausel erstmals gelungen, eine Voraussetzung für wirkliche Änderungen in der Ausbildung und Evaluation von Medizinstudierenden zu schaffen. Dabei kommt der Evaluation schon durch den „experimentellen“ Charakter eine besondere Bedeutung zu, da nun auch nachvollziehbar werden soll, wie Qualitätsmängel behoben werden können. Eine historische Betrachtung zeigt uns, dass durch die unbedachte Einführung einer Methodik (multiple choice questions – MCQ) über Jahrzehnte hinweg wissenschaftlich ungenügende Prüfungen bis heute Anwendung finden¹. Dabei kommt es nicht darauf an, unter wessen Verantwortung diese Tests durchgeführt wurden und werden; es soll vielmehr aufgezeigt werden, dass unangemessene politische und institutionelle Verfahren solche Fehler über Jahrzehnte im System zulassen, ohne dass ausreichende Qualitätskontrollen zum Einsatz kommen. Als auf politischer Ebene deutlich wurde, dass etwas geändert werden musste, kam es anstatt wissenschaftlicher Auseinandersetzungen über die MCQ-Methode, paradoxerweise zur Wiedereinführung der unstrukturierten mündlichen Prüfung. Durch die Vorgabe einer Experimentierklausel ist nun die Möglichkeit gegeben, wissenschaftlich sinnvolle Prüfungen zu entwickeln und einzusetzen.

Die Entwicklung und Auswahl von Prüfungsformen und Inhalten hat eine zentrale Bedeutung, da alle Prüfungsordnungen eine studiensteuernde Wirkung haben. Die Realität zeigt uns: Prüfungen sind das heimliche Curriculum einer Fakultät. Bei jedem Reformprojekt muss daher sorgfältig darauf geachtet werden, dass durch die gewählten Prüfungen die Reformansätze nicht wieder zunichte gemacht werden.

Ziel

In den vergangenen Jahrzehnten ist es deutlich geworden, dass Curriculumplanung nicht ohne eine effektive Evaluationsentwicklung geleistet werden kann. Sie bildet den Schlüssel zur dauerhafter Weiterentwicklung, auch nach erfolgreicher Im-

plementierung eines Reformcurriculums. Deshalb kommt der Evaluation eine dauerhaft zentrale Stellung zu. Die meisten Reformuniversitäten haben eine ganze Abteilung, die nur dieser anspruchsvollen Aufgabe gewidmet ist.

Prüfungen sollten nicht nur als Selektionsinstrument, sondern zur studentischen Standortbestimmung und Curriculumentwicklung eingesetzt werden. Zunächst muss daher definiert werden, welches Konstrukt wir versuchen zu evaluieren. Die ärztliche Kompetenz ist nicht einfach zu definieren. In Abb. 1 ist ein Versuch dazu gezeigt. Man erkennt die Stufen der Kompetenzentwicklung, wie sie im englischen Sprachraum geläufig sind [6]. Das zur Zeit durch die MCQs abgefragte reine Faktenwissen ist nur ein Teil der Kompetenzleistung. Höhere kognitive sowie soziale und manuelle Fähigkeiten dürfen daneben in ihrer Bedeutung nicht vernachlässigt werden.

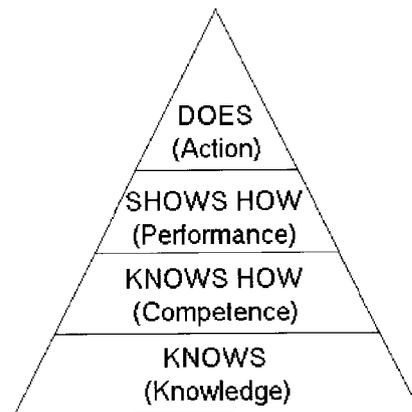


Abb. 1 Die Pyramide verdeutlicht die Stufen klinischer Kompetenz [6].

Wie geht man an die Umsetzung von wissenschaftlich angemessenen Prüfungen heran, ohne die gleichen Fehler zu wiederholen, die in der Vergangenheit gemacht wurden?

Die fünf Schritte der Evaluationsspirale [7]:

1. Planung und Vorbereitung: Die studiensteuernde Wirkung von Prüfungen ist schon erwähnt worden. Welche Effekte haben geplante mündliche oder MCQ-Prüfungen auf die Art und Weise wie Studierende lernen werden (z.B. ausgerichtet auf Vorlesungsprotokolle oder „Schwarze Reihe“ (Fragen pauken) anstatt Lehrbücher und medizinische Fachzeitschriften)? Daher muss in dieser Phase geklärt werden, wie Lernvorgänge im Curriculum geplant worden sind. Handelt es sich z.B. um ein problemorientiertes Curriculum mit studentenzentriertem Lernen, so müssen Prüfungen entsprechend gestaltet werden. Sodann können durch eine Evaluationsstrategie wesentliche Dimensionen der geplanten Prüfungsstrukturen entwickelt werden:

Warum prüfen wir: Wollen wir eher summativ oder formativ prüfen? Ist es das primäre Ziel, Lehr- und Lerngemeinschaften optimal zu fördern, werden Prüfungen bevorzugt formativ, d.h. lernprozessbegleitend, eingesetzt. Summative Prüfungen werden benötigt, wenn Mindestkompetenzen evaluiert werden und somit unzureichende Leistungen erkannt werden sollen.

¹ Obwohl die staatlichen MCQ-Prüfungen eine hohe Konsistenz (Reliabilität) aufweisen, ist im Allgemeinen die Validität von solchen „Papier-und-Bleistift-Tests“ begrenzt, wie Lienert [1] konstatiert, und im speziellen Fall der MCQ-Prüfungen des Instituts für Medizinische und Pharmazeutische Prüfungsfragen (IMPP) aus verschiedenen Gründen verringert: 1. Fragen mit negativer biserialer Korrelation (Trennschärfe) oder ungenügender Korrelation (Trennschärfen um 0) werden nicht eliminiert; 2. Fragen, bei denen die Alternativantwort eine höhere biserialer Korrelation aufweisen als die korrekte Antwort, und Fragen, bei denen die Alternativantworten nicht eine negative biserialer Korrelation aufweist, können zunächst nicht optimiert werden, da kein Vortesten der Fragen erfolgt, wie es z.B. in den USA üblich ist [2]; 3. die sog. „Gleitklausel“ regelt zudem, dass je schlechter die Studierenden gesamthaft abschneiden, desto mehr Studierende den Test bestehen – eine in sich unlogische und problematische Bedingung, die den Grad der Genauigkeit der Prüfung herabsetzt; 4. in Deutschland wird ausschließlich normbezogen getestet; ein Kriterienstandard ebenso wie standardisierende Verfahren wurden bisher nicht angewendet. Diese Verfahren sind z.B. in den USA üblich geworden und erhöhen besonders bei der Festlegung von minimalen Kompetenzstandards die Validität eines Tests bedeutend [3–5].

Was prüfen wir: Sinnvolle Lernziele zu entwickeln bedeutet nicht unbedingt umfangreiche „Gegenstandskataloge“ zu erstellen, zumal sie schnell veralten. Vielmehr sollte man sich die Frage stellen, was der Studierende am Ende seines Studiums an klinischen bzw. wissenschaftlichen Problemstellungen bewältigen können muss (sog. Exit Objectives).

Wie prüfen wir: Welche Evaluationsinstrumente man einsetzt hängt von Punkt 1 ab. In den vergangenen Jahrzehnten wurde dazu viel entwickelt, wie z.B. Objective Structured Clinical Examinations (OSCE oder „standardized patient“-based tests) [8 – 11], Progress Test, Triple Jump Tests [12], um nur einige zu nennen. Es geht über das Ziel dieser Darstellung hinaus, auf einzelne Tests näher einzugehen [13].

2. Entwicklung und Durchführung: Zunächst muss aus den erwähnten globalen Lernzielen eine aufgeschlüsselte Inhaltsmatrix („assessment blueprint“) erstellt werden, die regelmäßig auf den neuesten Stand der wissenschaftlichen Entwicklung gebracht werden muss. Sie sollte die globalen Dimensionen Wissen, Fertigkeiten und Haltungen beinhalten. Klassische Fächeraufteilungen können vor allem in den Grundlagenfächern vermieden werden, um die Medizinrelevanz deutlicher in den Vordergrund zu rücken (siehe dazu die United States Medical Licensing Examination, Part one) [2]. Wenn summativ geprüft wird, muss das sorgfältige Setzen von Kompetenzstandards vorgenommen werden. Der normbezogene Ansatz (der Kompetenzstandard richtet sich nach dem Mittelwert und Standardabweichung der geleisteten Prüfung) ist einfacher durchzuführen, jedoch werden international immer mehr kriterienbezogenes Testen vorgezogen, da sie besser definieren, welches Wissen und welche Fertigkeiten am Ende des Studiums de facto beherrscht werden müssen. Um jedoch zu kriterienbezogenem Prüfen zu kommen, müssen standardisierende Verfahren entwickelt werden [4], die in den letzten Jahren auch für OSCEs angewandt wurden [14]. In diesem Entwicklungsschritt muss die Praktikabilität bedacht werden. Gute Prüfungen sind nicht immer billig und eine Prioritätensetzung muss erfolgen.

3. Ergebnisse präsentieren: Werden Prüfungsergebnisse den Studierenden mitgeteilt, so ist es hilfreich, Folgendes zu beachten: Formative Prüfungen dienen dem Studierenden als Hilfe und somit steht das unmittelbare Feedback im Vordergrund. Direktes Feedback im Anschluss an einzelne Stationen in einem OSCE hat sich bewährt und kann das Feedback verbessern.

Wenn Noten vergeben werden sollen, muss die Sicherheit in der Notenvergabe, also der Standardmessfehler stets bekannt

sein, um den Studierenden eine Einschätzung der Irrtumswahrscheinlichkeit zu geben.

Dem Studierenden muss Feedback darüber gegeben werden, was diese Prüfung nicht evaluieren konnte. Dies sind häufig Bereiche wie klinische Kompetenz, über die der Studierende zwar annimmt, geprüft worden zu sein, die durch eine ausschließliche MCQ Prüfung aber nur ungenügend evaluiert werden kann.

Prüfungen sollten hinreichend in ihrem Inhalt und ihrer Methodik dokumentiert und begründet werden, so dass Schlussfolgerungen aus den Prüfungsergebnissen zuverlässig sind (z.B. „kompetent“ im Sinne einer „selbständigen Patientenbetreuung“).

4. Prüfungen evaluieren: Dieser Schritt in der Evaluationsentwicklung ist am wichtigsten, doch wird er häufig aus Zeit- und Kostengründen unterlassen. Hier stellen sich die grundlegenden Fragen von Reliabilität und Validität. Die Reliabilität beschreibt die Sicherheit in der Richtigkeit der Beobachtung über die momentane Prüfung hinaus, also die Wiederholbarkeit (innere Konsistenz). Es ist manchmal nicht leicht, die Reliabilität von OSCEs auf ein akzeptables Niveau zu steigern [15]. Doch sind schon nationale Prüfungen in Kanada und den USA mit ausreichender Reliabilität im Einsatz. MCQ-Prüfungen haben eine bekannt hohe Reliabilität von über 0,9, was häufig als Hauptargument für ihre Güte herangezogen wird. Bei der Gesamtbetrachtung müssen jedoch weitere Aspekte berücksichtigt werden, wie man der Tab. 1 entnehmen kann.

Es wird deutlich, dass es keine optimale Prüfung gibt, die jedem Anspruch gerecht wird. Besonders bedenklich wird es bei den meisten (unstrukturierten) mündlichen Prüfungen, wenn nicht nur die Reliabilität, sondern auch die Validität durch die häufig nicht vorhandene Themenbreite unter Niveau fällt (hoher Stichprobenfehler). Wird diese Prüfung, wie z.B. in den mündlichen Anteilen der Staatsexamina, summativ eingesetzt, so ist sie wissenschaftlich unzureichend.

Bei der Validität handelt es sich um das wichtigste Gütekriterium und zwar um die Begründbarkeit der Richtigkeit der Beobachtung (oder Messung). Um Bereiche der Validität besser zu beurteilen, kann man diese in unterschiedliche Güteklassen unterteilen:

- **Augenscheingültigkeit (Face Validity):** Das OSCE z.B. erscheint dem Betrachter intuitiv als eine „valide“ Form des Testens von klinischer Kompetenz – sie hat somit eine hohe Augenscheingültigkeit. Dies ist zwar von Vorteil (und er-

Tab. 1 Vergleich verschiedener Prüfungen. Siehe dazu auch [13].

	Validität	Reliabilität	Praktikabilität	Studiensteuerung
MCQ	+	++	++	-
OSCE	+	+	-	+
Triple Jump	+	-	+/-	+
mündlich	-	--	+/-	- oder +

Abkürzungen: MCQ = Multiple Choice Questions (im Sinne der United States Medical Licensing Examinations); OSCE = Objective Structured Clinical Examination; mündliche Prüfung (im Sinne der unstrukturierten mündlichen Examen der deutschen Staatsexamina). + = zufriedenstellend/positiv (z.B. Reliabilität über 0,8); - = ungenügend/negativ (z.B. Reliabilität unter 0,7).

- zeugt eine hohe Akzeptanz unter den Studierenden) kann aber weitere Formen der Validitätsprüfung nicht ersetzen.
- **Inhaltsvalidität:** Inhaltsmatrizen haben sich, als Standard für die Gewährleistung einer angemessenen Auswahl von Lernzielen, als hilfreich erwiesen. Weitere Möglichkeiten ergeben sich durch statistische Verfahren: Welche Erkrankung ist z. B. „häufig“ oder „wichtig“ (z. B. im Sinne von Ausmaß der Fehlerquote in ihrer Behandlung) oder „dringend“ (im Sinne ihrer zeitlichen Dringlichkeit in der Behandlung), um nur einige zu nennen.
 - **Kriterienbezogene Validität:** Insbesondere während des Validierungsverfahren eines Testinstruments können Außenkriterien, wie z. B. ähnliche Prüfungen, die das gleiche zu messen vorgeben, Eingang finden. Man muss jedoch vorsichtig sein, wenn verschiedene Konstrukte evaluiert werden: so ist es weniger sinnvoll reine Wissenstests (z. B. MCQ) mit sog. Performanz-Tests (z. B. OSCE) zu korrelieren (es sei denn, man möchte eine sog. „discriminant validity“ beweisen).
 - **Vorhersage Validität:** Es ist ein grundlegendes Ziel von summativen Prüfungen, unangemessene Patientenversorgung durch defizitäre Leistungen im späteren Arztberuf zu verhindern. Somit kommt den Prüfungen eine bedeutende Aufgabe zu: die Gültigkeit der Vorhersage. Obwohl diese Art von Validität oft nicht einfach zu testen ist, so kann sie wertvolle Ergebnisse liefern. Vor mehreren Jahrzehnten war z. B. das „Patient-Management-Problems“-Format in den USA sehr populär, bis man feststellen musste, dass erfahrene Ärzte schlechter abschnitten als Berufsanfänger. Der Grund: durch die Art der Methode wurde übermäßige Gründlichkeit überbewertet.

5. **Curriculum verändern:** Wie in Abb. 2 zu sehen ist, schließt sich die Evaluationsspirale mit dem fünften Punkt: die Möglichkeit, das Curriculum so zu beeinflussen, das es sich effizienter dem Curriculumsziel annähert. Damit wird die Grundlage für ständige curriculare Entwicklung geschaffen, und nicht, wie sie mit der bisherigen Ärztlichen Approbationsordnungen auf Jahrzehnte hinweg verhindert wurde.

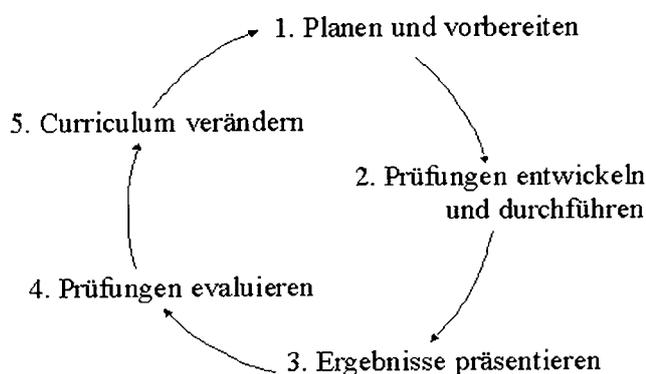


Abb. 2 Die Evaluationsspirale.

Probleme

Jede Prüfungsmethode, die nicht ständig in ihrem Zweck hinterfragt wird, kann sich durch die „Routine“ des Ausbildungsalltags mit der Zeit qualitativ verschlechtern:

- Fragen werden nicht analysiert oder auf den neuesten Stand gebracht,
- Aufgrund der benötigten hohen Prüfer- und standardisierter Patientenanzahl werden Stationen im OSCE ad hoc ohne Inhaltsmatrix zusammengestellt,
- neuere Prüfungsmethoden werden ungeprüft übernommen und eingesetzt, ohne ihren oft engen Einsatzbereich zu kennen,
- Evaluationsbereiche wie Wissen, Fertigkeiten und Haltungen werden häufig zu wenig getrennt und zusammen evaluiert, ohne ihrem jeweiligen Einflussbereich bewusst zu sein.

Diese Punkte repräsentieren lediglich die häufigsten Fehlerquellen. Es ist notwendig zu verstehen, dass sog. Performanz-Tests je nach ihrer Einsatzweise nicht immer wahre klinische Kompetenz messen, sondern häufig nur Fertigkeiten prüfen, und ähnlich vielen MCQs, nur Erinnertes messen. Will man die integrierte Anwendung von Wissen, Fertigkeiten und Haltungen messen (und damit die Performanz), so steigt der Aufwand erheblich. Viele Fakultäten haben selbst in den USA nicht die Kapazitäten oder Ressourcen dies durchzuführen, obwohl Barrows [8] zeigen konnte, wie diese Methode erfolgreich einzusetzen ist.

Schlussfolgerungen

Sorgfältige Planung ist notwendig in der Entwicklung von angemessenen Evaluationsformen und ihrer methodischen Integration in das Curriculum. Die Prüfungsordnung stellt das zentrale Messinstrument dar, an dem Erfolg oder Misserfolg gemessen wird. Dies nicht nur in bezug auf den einzelnen Studierenden sondern auch in Bezug auf eine erfolgreiche Curriculumsentwicklung. Dabei sind insbesondere drei Punkte wichtig:

- Eine einzelne, wenngleich auch geeignete Prüfungsmethode kann nicht das ganze Spektrum ärztlicher Kompetenz evaluieren, wie sie durch die Trias Wissen, Fertigkeiten und Haltungen beschrieben werden kann.
- Man erreicht mit einer Evaluationsmethode nur soviel, wie man bereit ist, in ihre kontinuierliche Entwicklung und Durchführung zu investieren.
- Validierung der Testinstrumente wird zur wichtigsten Forschungsaufgabe, insbesondere wenn man sich entscheidet, anstatt normbezogene kriterienbezogene Prüfungen durchzuführen.

Literatur

- ¹ Lienert GA, Raatz U (Hrsg). Testaufbau und Testanalyse. Weinheim: Beltz, Psychologie Verlags Union, 1994
- ² O'Donnel MJ, Obenshain SS, Erdmann JB. Background Essential to the Proper Use of Results of Step 1 and Step 2 of the USMLE. Acad Med 1993; 68: 734-739
- ³ Hoffman K. The USMLE, the NBME Subject examinations, and assessment of individual academic achievement. Acad Med 1993; 68: 740-747
- ⁴ Shea JA, Reshetar R, Norcini JJ, Dawson B. Sensitivity of the modified Angoff Standard-Setting Method to variations in item content. Teach Learn Med 1994; 6: 288-292
- ⁵ Swanson DB, Case SM, Waechter D, Veloski J, Habrouck C, Friedman M, Carline J, MacLaren C. A preliminary study of the validity of scores and pass/fail standards for USMLE Steps 1 and 2. Acad Med 1993; 68 (Suppl 10): S19-S21

- ⁶ Miller GE. Conference summary. Proceedings of the AAMC's Consensus Conference on the use of Standardized Patients in the Teaching and Evaluation of Clinical Skills. *Acad Med* 1993; 68: 471 – 474
- ⁷ Fowell SL, Southgate LJ, Bligh JG. Evaluating assessment: the missing link? *Med Edu* 1999; 33: 276 – 281
- ⁸ Barrows HS. An overview of the uses of standardized patients for teaching and evaluation clinical skills. Proceedings of the AAMC's Consensus Conference on the use of Standardized Patients in the Teaching and Evaluation of Clinical Skills. *Acad Med* 1993; 68: 443 – 451
- ⁹ Harden RM, Gleeson FA. Assessment of medical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; 13: 41 – 54
- ¹⁰ van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teach Learn Med* 1990; 2: 58 – 76
- ¹¹ Falck-Ytter Y, Stiegler I. Beurteilung klinischer Fähigkeiten mit dem OSCE-Verfahren. *Medizinische Ausbildung* 1993; 10/1: 48 – 55
- ¹² Stiegler I, Falck-Ytter Y. Erste Erfahrungen mit der „Triple Jump“ Prüfung. *Medizinische Ausbildung* 1993; 10/1: 41 – 47
- ¹³ Falck-Ytter Y, Bircher J. Vergleich verschiedener Prüfungsformen in der Medizin. In: Westhoff K, Lengerich (Hrsg): 1. Symposium zu Prüfungen in der Medizin: Multiple Choice. Möglichkeiten und Grenzen von Multiple-Choice Prüfungen in der Medizin. Berlin, Riga, Scottsdale, Wien, Zagreb: Pabst Science Publishers, 1995: 57 – 67
- ¹⁴ Morrison H, McNally H, Wylie C, McFaul P, Thompson W. The passing scope in the Objective Structured Clinical Examination. *Med Edu* 1996; 30: 345 – 348
- ¹⁵ Falck-Ytter Y, Bircher J. Use of an Objective Structured Clinical Examination in Internal Medicine. In: Majoor GD, van der Vleuten CPM, Vluggen PMJ, Hansen PA (eds): *MedEd-21: An Account of Initiatives for Change in Medical Education in Europe for the 21st Century*. Amsterdam: Thesis Publisher, 1997: 197 – 204

Yngve Falck-Ytter, MD

Case Western Reserve University, USA
MetroHealth Medical Center
2500 MetroHealth Drv.
Cleveland, OH, 44109, USA

E-mail: falck-ytter@bigfoot.com