

OSCE - praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung

OSCE - hands on instructions for the implementation of an objective structured clinical examination

• Christoph Nikendei¹ • Jana Jünger¹

Zusammenfassung:

Einleitung: Der OSCE (objective structured clinical examination) hat sich als reliables und valides Instrument zur Prüfung klinisch-praktischer Fertigkeiten erwiesen. In Hinblick auf die vermehrt praxisnahe Ausbildung im Medizinstudium wächst die Notwendigkeit der Durchführung einer solchen klinisch-praktischen Prüfung. Um eine Implementierung zu erleichtern wird nach einem Literaturüberblick der 12-Stationen-OSCE an der Medizinischen Universitätsklinik Heidelberg in Hinblick auf Konzeption, Bestimmung der Bestehensgrenze, Durchführung und Notenvergabe vorgestellt.

Methode: Im WS 2004 / 2005 nahmen 143 Medizinstudenten des interdisziplinären Blocks Innere Medizin an einem 12-Stationen-OSCE teil. Die Prüfungskonzeption erfolgte mit Hilfe eines Blueprint. Die Bestehensgrenze wurde mit einer modifizierten Angoff-Methode festgelegt, die Prüfungsleistung an einzelnen OSCE-Stationen von Experten mit Hilfe von eigens entwickelten Checklisten bestimmt.

Ergebnisse: Der durchgeführte OSCE weist bezüglich der Reliabilität ein Cronbach alpha von 0,748 auf. Aufgrund der im Standard Setting festgelegten Bestehensgrenze erreichten 142 Studenten (99,3%) Prozent das Prüfungsziel. Das Absolvieren der Prüfung am 1., 2. oder 3. Prüfungstag hatte keinen Einfluss auf die Benotung der studentischen Leistung.

Diskussion: Der OSCE an der Medizinischen Universitätsklinik Heidelberg erweist sich als reliables und gerechtes Prüfungsinstrument. Auch wenn der OSCE ein ressourcen- und zeitaufwendiges Prüfungsinstrument darstellt, erweist er sich bei guter Planung und Organisation als eine selbst bei großen Studentenzahlen in einem Regelstudiengang realisierbare Prüfungsform, welche eine hohe Zufriedenheit bei Studenten und Prüfern hinterlässt.

Schlüsselwörter: OSCE, klinisch-praktische Prüfung, Standard Setting, Reliabilität, Validität, Notenvergabe

Abstract:

Introduction: Objective structured clinical examinations (OSCEs) have proved to be a reliable and valid assessment tool for basic clinical skills. In regards to the more and more practical education of medical students there is an increasing need for the implementation of OSCEs. To facilitate this implementation we would like to give a review of literature and to present our 12 station OSCE at the Medical Hospital of the University of Heidelberg.

Methods: During winter term 2004 / 2005 143 medical students in their 3rd year of training participated the OSCE. For the conception of the assessment a blueprint was used. Standard setting was carried out with a modified Angoff method, students performance was assessed by experts using self-developed checklists.

Results: The OSCE at the Medical Hospital of the University of Heidelberg showed a reliability of 0,748. According to the results of the standard setting procedure 142 students (99,3%) passed the exam. Taking the OSCE on the first, second or third day of assessment did not have impact on students' performance.

Discussion: The OSCE at the Medical Hospital of the University of Heidelberg proves to be a reliable and fair assessment tool. Even the OSCE represents an assessment tool that needs resources and time it is well accepted in students and raters if it is planned and organised carefully.

Keywords: OSCE, undergraduate medical education, standard setting, reliability, validity

Einleitung

• Klinisch-praktische Prüfungen im Licht der Novellierung der Ärztliche Approbationsordnung

Klinisch-praktische Prüfungen in Form eines OSCE (objective structured clinical examination, [57]) haben in den vergangenen Jahren weltweit eine hohe Akzeptanz und Verbreitung gefunden. Mit dieser Prüfungsform werden die Medizinstudenten direkt auf der Stufe des "shows how" der Lernpyramide von Miller [34] geprüft, also die Durchführung prozeduraler Fertigkeiten und Fähigkeiten direkt beurteilt. Die Novellierung der Approbationsordnung

für Ärzte [3] betont eine praxisnahe Ausbildung von Medizinstudenten, womit in der Konsequenz auch die Durchführung solcher klinisch-praktischen Prüfungen an deutschen Medizinischen Fakultäten zur Leistungsbewertung und Notenvergabe einen wichtigen Stellenwert einnehmen wird. Dies impliziert die Notwendigkeit, eine gerechte, dem Wissensstand angepasste, reliable und valide OSCE-Prüfung zu gewährleisten. Hierbei werden die einzelnen Fakultäten mit organisatorischen und logistischen Schwierigkeiten konfrontiert. Um die Planung, Konzeption und Durchführung eines OSCE zu erleichtern, möchten wir den aktuellen Forschungsstand und die Konsequenzen für eine praxisnahe, ressourcensparende Umsetzung einer OSCE-Prüfung am Beispiel der Medizinischen Universitätsklinik Heidelberg skizzieren.

¹ Medizinische Universitätsklinik Heidelberg, Ludolf-Krehl-Klinik, Klinik für Allgemeine Klinische Medizin und Psychosomatik, Heidelberg, Deutschland

• **Was ist ein OSCE?**

Das Prüfungsformat eines OSCE wurde erstmals an der Universität Dundee in Schottland durchgeführt und von Harden et al. [19] publiziert. Die einzelnen Medizinstudenten rotieren beim OSCE durch einen Prüfungsparcours mit einer Serie von Prüfungsstationen, an denen sie definierte klinisch-praktische Fertigkeiten unter Beweis stellen müssen. An jeder Prüfungsstation wird die Prüfungsleistung durch einen Prüfer anhand einer Checkliste beurteilt. Die im Namen beinhaltete "Objektivität" und "Strukturierung" spiegeln sich in der hohen Anzahl von Prüfungsstationen und beteiligten Prüfern sowie in den standardisierten Aufgabenstellungen und inhaltlich definierten Checklisten wieder.

• **Validität des OSCE und Blueprint**

Die Validität von OSCEs konnte in vielen Fällen nachgewiesen werden (z.B. [22], [29], [6], [7]). Um eine hohe Inhaltsvalidität eines OSCE herzustellen, empfiehlt Newble [37] drei Schritte bei der Prüfungskonzeption: Erstens die Festlegung von Problemfeldern, in denen die Prüflinge kompetent sein sollen, zweitens die Definition von Aufgabenstellungen innerhalb dieser Problemfelder und - drittens - die Erstellung eines "Blueprint". Mit Hilfe des "Blueprint" werden die Prüfungsinhalte des OSCE mit den Zielen des der Prüfung zugrunde liegenden Curriculums abgeglichen. Der Blueprint gewährleistet, dass alle Curriculumsziele geprüft werden und die entscheidenden Schlüsselprobleme angemessen repräsentiert sind [14]. Im einfachsten Fall besteht ein Blueprint aus einer zweidimensionalen Matrix, bei der auf einer Achse allgemeine zu testende Kompetenzen (z.B. Anamneseerhebung, körperliche Untersuchung usw.) verzeichnet sind, auf der anderen die Problemstellungen, mit welchen diese Kompetenzen abgeprüft werden sollen [36]. Ein Beispiel für einen Blueprint wird in Tabelle 1 veranschaulicht.

• **Reliabilität des OSCE: Anzahl und Dauer der Prüfungsstationen**

Zu den Einflussvariablen, die die Reliabilität [35] eines OSCE möglicherweise beeinflussen können, zählen die Anzahl der Prüfer, die Anzahl an Prüfungsstationen, die Strukturiertheit der Prüfung und die unterschiedlichen bei der Prüfung eingesetzten standardisierten Patienten [55]. Den gewichtigsten Faktor stellt allerdings die Anzahl der beim OSCE verwandten Prüfungsstationen dar. Aufgrund der Spezifität der Problemstellungen an den einzelnen Prüfungsstationen ist es notwendig, eine hohe Anzahl von unterschiedlichen Problemfällen und damit von Prüfungsstationen zum Einsatz zu bringen, um eine Reliabilität um 0,8 zu erzielen, wie für "high stakes" Examina gefordert [56], [50], [37]. Das Accreditation Council for Graduate Medical Education (ACGME) und das American Board of Medical Specialties (ABMS) [50] empfehlen die Implementierung von 14 - 18 Prüfungsstationen à 10 - 15 Minuten. Für eine einzelne Prüfungsstation mit einem standardisierten Patienten (SP) [4] zur Anamneseerhebung und Durchführung einer körperlichen Untersuchung ist eine Prüfungszeit von 15 Minuten einzukalkulieren [8]. Anzahl und Dauer der OSCE-Stationen werden in der Praxis ganz unterschiedlich festgelegt:

• Universität Dundee, Schottland: 25 - 35 Prüfungsstationen à 4½ Minuten [14]

• Medical Council, Canada: 20 Prüfungsstationen à 10 Minuten [44]

• Havard Medical School, USA: 16 Prüfungsstationen à 9 Minuten [18]

Die in der Literatur für die Durchführung von reliablen OSCEs veranschlagte Zeitdauer und Anzahl der Prüfungsstationen dürfte für die deutschen Fakultäten kaum realisierbar sein (vgl. [56], [50], [37]). Ziel der vorliegenden Untersuchung war es deshalb zu prüfen, inwieweit ein OSCE mit weniger Stationen und kürzerer Prüfungsdauer die geforderte Reliabilität erreicht.

• **Checklisten oder globales Rating?**

Traditionell werden beim OSCE Checklisten mit binären Items ("korrekt" / "nicht korrekt") verwandt, mit deren Hilfe die korrekte Durchführung von Teilschritten der Prüfungsaufgabe festgehalten wird [32], [49] (siehe Abbildung 1). Bei einem globalen Rating hingegen werden z.B. für insgesamt vier prozedurale Dimensionen jeweils fünf Punkte vergeben [21] (siehe Abbildung 1). Checklisten mit binären Items bergen die Gefahr, die Prüfer durch eine Fokussierung auf die Checklisten-Items abzulenken [15], [23], [47], so dass eine steigende Zahl von Items die Reliabilität und Validität des OSCE reduziert [59]. Auch insgesamt erweist sich ein globales Rating (siehe Abbildung 1) in Bezug auf Reliabilität, interne Konsistenz und Validität der Verwendung von Checklisten als ebenbürtig [22], [43], [42], [45], [54] und bildet interaktive Kompetenzebenen wie Empathie und ethische Haltungen besser ab [41], [53]. Auf der Ebene der Prüflinge führt ein globales Rating bei der Prüfung kommunikativer Fähigkeiten zu vermehrt offenen Fragen, ein Checklisten-Rating zu vielen fokussierten, geschlossenen Fragen [23]. Die Kombination aus globalem Rating und der Verwendung von Checklisten ist möglich und kann durchaus sinnvoll sein. Newble [37] empfiehlt bei Stationen mit vorrangig technischen Fertigkeiten eine Checkliste als adäquates Evaluationsinstrument, bei der Prüfung von kommunikativen Kompetenzen hingegen die Verwendung eines globalen Ratings. Ziel der vorliegenden Studie war es, ein eigenes Checklistenformat zu entwickeln, welches die Vorteile eines Checklisten-Ratings und einem globalen Rating vereint.

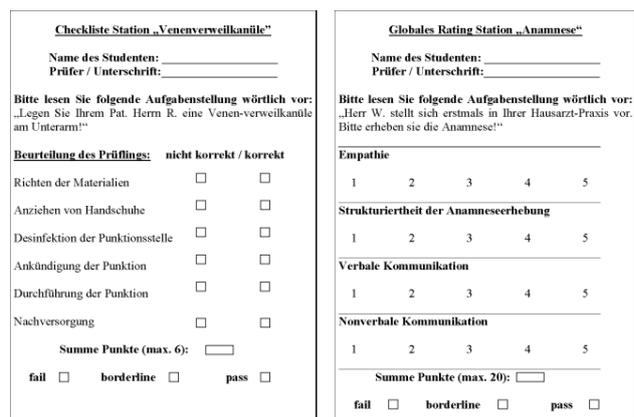


Abbildung 1: Checklisten-Rating (CR) / Globales Rating (GR) (vereinfachte Beispiele)

Tabelle 1: Prüfungsstationen des OSCE der Medizinischen Universitätsklinik Heidelberg Wintersemester 2004/2005: Blueprint (vereinfachtes Beispiel)

Station	Endokrinologie	Psychosomatik	Kardiologie	Pulmonologie	Gastroenterologie	Rheumatologie
Inhalt der Station	Labor und DD bei Diabetes mellitus	Anamnese bei Essstörungen	Befundung eines EKG	Befundung Röntgenbild	Management GI-Blutung	Diag. / Therapie Rheumatoide A.
Praktische Fertigkeiten	+		+	+	+	
Patienten-Untersuchung						
Patienten-Management	+	+	+	+	+	+
Kommunikative Fähigkeiten		+				
Haltungen, ethische Aspekte		+				
Treffen klinischer Entscheidungen	+	+	+	+	+	+
Verständnis wiss. Zusammenhänge	+	+	+	+	+	+
Station	Hämatologie	Allgemeinmedizin	Nephrologie	Körperliche Untersuchung	Braunüle	Klinische Chemie
Inhalt der Station	Diag. / Therapie Lymphome	Risikoabschätzung / Prävention CVR	Diag. / Therapie Niereninsuffizienz	Untersuchung des Herzens	Legen einer Braunüle	Interpretation von Laborparametern
Praktische Fertigkeiten		+		+	+	+
Patienten-Untersuchung				+		
Patienten-Management	+	+	+			+
Kommunikative Fähigkeiten		+		+	+	
Haltungen, ethische Aspekte		+		+		
Treffen klinischer Entscheidungen	+	+	+			+
Verständnis wiss. Zusammenhänge	+	+	+			+

• Der OSCE als formative und summative Prüfungsform / Feedback während des OSCE

OSCEs können mit unterschiedlichen Intentionen durchgeführt werden. In vielen Fällen wird der OSCE als sogenannte formative Prüfung eingesetzt [31]. Dies bedeutet, dass der OSCE in diesem Falle dazu dient, den Studenten ein konstruktives Feedback über ihren aktuellen Wissens- und Könnensstand geben zu können. Defizite in der Durchführung prozeduraler Fertigkeiten können somit identifiziert und korrigiert werden [52]. Bei summativen Prüfungen dient der OSCE dagegen vorrangig zur Entscheidung über das Vorhandensein notwendiger Qualifikationen zum Bestehen von definierten Abschnitten eines Curriculums.

Eine eineinhalbminütiges Feedback im Anschluss an die Absolvierung einzelner OSCE-Stationen wird von den Medizinstudenten als hilfreich und wenig störend empfunden [1]. Ein solches Feedback führt nachgewiesenermaßen nachfolgend zu einer Verbesserung der studentischen Fertigkeiten [20].

• Standard Setting - Ermittlung der Bestehensgrenze eines OSCE

Mit einem Standard Setting wird die Bestehensgrenze eines OSCE festgelegt, die "kompetente" von "nicht-kompetenten" Studenten trennen soll. Man unterscheidet je nach Vorgehensweise kriteriumsorientierte, normorientierte und holistische Methoden (siehe Tabelle 2). Die willkürliche Festlegung einer Bestehensgrenze (z.B. 60%), wie es das Vorgehen bei der holistischen Methode vorsieht, weist eine ungenügend begründbare Entscheidungsbasis über das Bestehen oder Nicht-Bestehen einer Prüfung auf [17]. Aber auch

normorientierte Verfahren haben Schwächen, die in unter anderem in der relativen Bewertung der Studenten untereinander zu sehen sind [28]. Die kriteriumsorientierten Methoden bieten hingegen gute Zugänge zur Ermittlung der Bestehensgrenze und sollen hier kurz skizziert werden.

Tabelle 2: Formen des Standard-Setting

Absolute, kriteriumsorientierte Methoden
<u>Testzentriert (rational):</u> Angoff-Methode bzw. modifizierte Angoff-Methode [3] Methoden nach Ebel [16], Nedelsky [35], Jäger [24]
<u>Prüflingszentriert (empirisch):</u> Borderline-Methode / Kontrast-Gruppen-Methode [30]
Relative, normorientierte Methoden
Bestimmung des Cut-off-Point mit Wjinen-Methode [57] Bestimmung des Cut-off-Point anhand der Subgruppe der besten Studenten [10]
Holistische Methode
Prozentsatz derer, die bestehen wird vor der Prüfung definiert [28]

Bei der Methode nach Angoff, die beispielsweise in den USA (ECFMG/CSA) [5] verwandt wird, wird die Bestehensgrenze bereits vor der Prüfung durch eine Gruppe von Experten festgelegt. Diese Experten sollten über die Prüfungsstandards informiert und mit der Prüfungsmethode vertraut sein und den Kenntnisstand der Medizinstudenten einschätzen können. Zudem sollten sie sich über die Schwierigkeit der Prüfungsaufgabe im Klaren sein, was durch die Demonstration von Videos oder die eigene Durchführung der Prüfungsaufgabe unter examensähnlichen Bedingungen unterstützt werden kann [5], [17]. Searle [48] verwandte für die Angoff-Methode eine Anzahl von drei Experten.

Friedman [17] empfiehlt eine modifizierte Angoff-Methode ("modified Angoff") mit einem Vorgehen in vier Schritten in dessen Zentrum die Frage steht, wie viele Punkte ein sogenannter

"Borderline-Prüfungskandidat" an einzelnen Prüfungsstation erzielen würde. Nach einem ersten Rating der Experten werden die Einschätzungen untereinander präsentiert und diskutiert, bevor ein zweites, endgültiges Rating erfolgt. Ein "Borderline-Prüfungskandidat" wird dabei definiert als ein Prüfungsteilnehmer, der die Experten weder als qualifiziert, noch als unqualifiziert einschätzen [5]. Der Mittelwert dieser Einschätzungen gibt die Bestehensgrenze für die einzelnen Stationen bzw. die Mittelwertsbildung über alle Prüfungsstationen die Bestehensgrenze für die gesamte OSCE-Prüfung an. Es ist dabei zu berücksichtigen, dass die Ratings immer in Form der Anzahl von erreichten Punkten erfolgen sollten und nicht in Prozentwerten, da sich gezeigt hat, dass sich die Experten von gängigen Vorstellungen von Bestehensgrenzen (z.B. 50% oder 70%) beeinflussen lassen. Nachteile der Angoff-Methode sind in der Zeitaufwendigkeit und der hypothetischen Herangehensweise zur Festlegung der Bestehensgrenze zu sehen [17].

Bei der Borderline-Methode, die beispielsweise vom Medical Council von Canada verwandt wird [45], wird die Bestehensgrenze hingegen empirisch durch eine direkte Beobachtung während der Prüfung festgelegt, so dass keine zusätzliche Zeit zur Bestimmung der Bestehensgrenze aufgewandt werden muss [37]. Nachteil dieser Methode ist es, dass die Bestehensgrenze nicht wie in den meisten Studienordnungen gefordert vor der Prüfung festgelegt werden kann. Die Einschätzung, ob ein Prüfling ein "Borderline-Prüfungskandidat" ist, wird durch eine zusätzliche Skala auf der Checkliste der einzelnen Stationen festgelegt (siehe Abbildung 1), die in der Praxis auf unterschiedliche Weise erfolgen kann:

- Rating auf einer Skala bestehend aus: pass / borderline / fail [46]
- Rating auf einer Skala bestehend aus: fail, borderline, pass, above expected standard [13]

Der Mittelwert der an den einzelnen Stationen von den "Borderline-Prüfungskandidaten" erreichten Punktzahlen entspricht wiederum der Bestehensgrenze der OSCE-Prüfung.

Eine Modifikation der Borderline-Methode stellt die Kontrast-Gruppen-Methode dar, bei der keine Gruppe von "Borderline-Prüfungskandidaten" definiert wird, sondern lediglich eine Unterscheidung in "kompetent" vs "nicht-kompetent" erfolgt und anhand dieser Festlegung die Errechnung der Bestehensgrenze erfolgt [5]. Sowohl die Borderline-Methode als auch die Kontrast-Gruppen-Methode bieten beide konsistente und realistische Standards [12].

Sowohl bei der Methode nach Angoff als auch bei der Borderline-Methode stellt es sich für die Experten bzw. die Prüfer häufig als schwierig dar, sich einen "Borderline-Prüfungskandidaten" vorzustellen. Ziel war es dementsprechend eine modifizierte Angoff-Methode zu entwickeln, die es erlaubt eine Standard Setting innerhalb kurzer Zeit effizient durchzuführen.

• Der OSCE als kompensatorische oder non-kompensatorische Prüfungsform?

Bei einem kompensatorischen Ansatz muss der Prüfling lediglich eine bestimmte Gesamtpunktzahl beim OSCE erreichen - eine unzufriedenstellende Leistung an einer einzelnen OSCE-Station kann durch eine bessere Leistung an einer beliebigen anderen Station ausgeglichen werden. Bei einem non-kompensatorischen

OSCE müssen die Studenten einzelne oder im Extremfall an allen OSCE-Stationen die Bestehensgrenze erreichen, um auch den gesamten OSCE erfolgreich zu bestehen. Es existiert keine falsche oder richtige Antwort auf die Frage, ob ein kompensatorischer oder non-kompensatorischer Ansatz präferiert werden sollte und welcher Zugang der validere ist [37]. Für die Befürwortung einer non-kompensatorischen Vorgehensweise spricht, dass die Prüflinge in allen Bereichen ihre Fähigkeiten unter Beweis stellen müssen. Die Gefahr dieses Modells ist jedoch, dass eine schlecht konzipierte Station eventuell schlecht zwischen "kompetenten" und "nicht-kompetenten" Studenten trennt und nicht nur die "nicht-kompetenten" Studenten die Leistungserwartung an dieser Station nicht erfüllen. Außerdem kann es auch für gute Studenten zulässig sein, Schwierigkeiten bei der Absolvierung von ein oder zwei Stationen aufzuweisen [9], [17]. Als schwierig erweist sich bei non-kompensatorischen OSCE-Stationen auch die Wiederholung von nicht bestandenen OSCE-Stationen. Da der Prüfling meist nur wenige Stationen nicht bestanden hat, müsste er nur diejenigen wiederholen, was in einem laufenden OSCE nicht praktikabel ist.

Im Folgenden soll die praktische Umsetzung einer klinisch-praktischen Prüfung am Beispiel des OSCE im Wintersemester 2004 / 2005 an der Medizinischen Universitätsklinik Heidelberg in Hinblick auf Konzeption, Bestimmung der Bestehensgrenze, Durchführung und Notenvergabe dargestellt werden.

Methoden

• Der OSCE an der Medizinischen Universitätsklinik Heidelberg

An der Medizinischen Universitätsklinik Heidelberg werden OSCEs seit dem Wintersemester 2001/2002 durchgeführt (vgl. [26]). Die OSCEs finden im 6. bzw. 7. Studiensemester am Ende des 14-wöchigen Blockkurses "Innere Medizin" statt, während dem sich die Studenten klinisch-praktisch Kompetenzen beim Kommunikationstraining mit standardisierten Patienten [25][40], beim Skills-Lab-Training [38], [39] und beim Einsatz auf Station erwerben können. Die Prüfungsinhalte werden jeweils mit Hilfe eines Blueprint auf die Lernziele des Blockes Innere Medizin abgestimmt (siehe Tabelle 1). Zu Beginn ihres Einsatzes in der Medizinischen Universitätsklinik werden die Studenten über den Ablauf und die Formalien des OSCE informiert. Die OSCE-Prüfung in Heidelberg ist eine kompensatorische Prüfung. Bei Nicht-Bestehen muss die gesamte Prüfung ein Semester später wiederholt werden.

• Design der Stationen und Checklisten / Standard Setting und Notenvergabe

Die einzelnen Prüfungsstationen des 12-Stationen-OSCE wurden von Vertretern der jeweiligen internistischen Teildisziplin in Übereinstimmung mit dem entworfenen Blueprint entwickelt. Als Grundlage diente ein definiertes Checklistenformat mit einem modifizierten Rating mit fünf Teilaufgaben, bei denen jeweils 5 Punkte zu erzielen waren. Insgesamt konnte somit eine Maximalpunktzahl von 25 pro Prüfungsstation erreicht werden. Das Checklistenformat beinhaltet sowohl Anteile des Checklisten-Ratings als auch des globalen Ratings: So werden z.B. für die Teilaufgabe 1 und 2 der Checkliste in Abbildung 2 nur die maximal zu erreichende Punktzahl vorgegeben, so dass der Gesamteindruck

des Prüfers bezüglich der Gesprächsführung und Empathie mit in die globale Bewertung einfließen kann.

STATION Herzinsuffizienz

Setting: In der internistischen Ambulanz stellt sich erstmalig ein 57-jähriger Patient (standardisierter Patient) vor.

Aufgabe 1 + 2: Bitte erheben Sie die aktuelle Anamnese!	
Symptombezogene Anamnese	Punkte (max. 5)
<ul style="list-style-type: none"> → Akute Beschwerdesymptomatik? Dyspnoe, Angina pectoris, Knöchel-/Beinödeme, Gewichtsveränderung, Husten, Auswurf? → Entwicklung der Beschwerden (innerhalb welcher Zeit)? → Beschwerden bei Belastung? → Sonstige Symptome/Auffälligkeiten? 	
Vegetative Anamnese, Risikofaktoren, sonstige Anamnese	Punkte (max. 5)
<ul style="list-style-type: none"> → Fieber, Leistungsminde- rung, Nykturie, bestimmte Schlafposition? → Risikofaktoren? → Frühere Erkrankungen? Begleiterkrankungen? → Medikamentenanamnese? Allergien? → Soziale Anamnese: Beruf, Lebenssituation, Familienanamnese (Herzkrankheiten)? 	
Aufgabe 3: Bitte erläutern Sie dem Patienten Ihre Verdachtsdiagnose! Welche Untersuchungen zur Sicherung Ihrer Verdachtsdiagnose veranlassen Sie als nächstes?	Punkte (max. 5)
Verwendung einer einfachen für den Patienten verständlichen Sprache: Verdachtsdiagnose: → Erklärt dem Patienten ohne Fremdwörter die Verdachtsdiagnose (Globale Herzinsuffizienz mit links- und rechtskardialer Dekompensation)(2 Pkt.); Untersuchungen zur Diagnosesicherung: → Erklärt dem Patienten in verständlicher Sprache welche Maßnahmen als nächstes sinnvoll sind (Körperliche Untersuchung, EKG, Blutdruckmessung, Blutentnahme, Röntgen-Thorax, Echokardiographie)(3 Pkt.)	
Aufgabe 4: Bitte erläutern Sie dem Patienten den Zusammenhang zwischen Ihrer Verdachtsdiagnose und dem Auftreten seiner klinischen Beschwerden!	Punkte (max. 5)
Verwendung einer einfachen für den Patienten verständlichen Sprache: → Reizhusten, Dyspnoe, verringerte Belastbarkeit, Beinödeme, Nykturie als Zeichen von Vorwärtsversagen und Rückwärtsversagen des Herzens (je verständlich und sinnvoll erklärtem Symptom 1 Pkt.)	
Aufgabe 5: Bitte erklären Sie dem Patienten, worauf er in Zukunft achten sollte, um den Verlauf seiner Erkrankung selbst beurteilen und sich ggf. rechtzeitig wiedervorstellen zu können!	Punkte (max. 5)
Verwendung einer einfachen für den Patienten verständlichen Sprache: → Auf Veränderungen z.B. eine Abnahme der körperlichen Belastbarkeit achten (1 Pkt.) → Tägliches Messen von Blutdruck und Puls (Blutdruckänderungen? Puls rhythmisch? Frequenz?) (1 Pkt.) → Täglich wiegen (Gewichtszunahme von mehr als 1,5–2 kg?) (1 Pkt.) → Symptome wie Knöchel- und/oder Beinödemen, Husten (insbesondere trockenen Husten) beobachten (1 Pkt.) → Patiententagebuch führen und bei Auftreten der genannten Symptome wiedervorstellen (1 Pkt.)	
Gesamtpunktzahl (max. 25)	

Abbildung 2: Checklistenformat an der Medizinischen Universitätsklinik Heidelberg (aus: Jünger und Nikendei, 2005 [26])

Zur Ermittlung der Bestehensgrenze wurde ein modifiziertes Vorgehen nach Angoff [2] entwickelt. Ein Vertreter jeder internistischen Teildisziplin (n=10) nahm als Experte am Standard Setting teil und gab ein Rating zur "minimal zu demonstrierenden Kompetenz" in Form der minimal zu erreichenden Punktzahl für jede einzelne Stationen ab. Der Mittelwert der Einschätzung aller Experten ergab die Bestehensgrenze für die entsprechende Station, der Mittelwert über alle Stationen hinweg die Bestehensgrenze für den OSCE. Da in Deutschland die Notwendigkeit der Notenvergabe besteht, wurden die Punktwerte über der Bestehensgrenze in vier gleichgroße Abschnitte unterteilt, entsprechend den Noten eins bis vier.

• Ablauf des OSCE

Alle 12 Stationen des OSCE im Wintersemester 2004 / 2005 waren mit der Stationsnummer beschildert. Maximal zwei Stationen waren - abgetrennt durch einen Sichtschutz - in einem Prüfungsraum lokalisiert. Pro Station stand den Medizinstudenten eine Prüfungszeit von fünf Minuten zur Verfügung mit einer sich anschließenden einminütigen Wechselzeit. Der Beginn und das Ende der Prüfungszeit wurde jeweils mit einem lauten Signalton angezeigt. Ein einmaliger Durchlauf des 12-Stationen-OSCE nahm somit 72 Minuten in Anspruch. Bis zum nachfolgenden OSCE-Durchlauf wurde jeweils eine Pause von 18 Minuten eingelegt. Über drei Tage hinweg wurden jeweils zwei Mal nacheinander zwei parallele OSCE-Prüfungen durchgeführt, somit also pro Tag maximal vier Gruppen à 12 Studenten geprüft. Insgesamt wurden

an den drei Prüfungstagen 143 Studenten der OSCE-Prüfung unterzogen.

Um eine Informationsweitergabe zwischen den Studentengruppen zu vermeiden, wurden die Prüflinge eines Prüfungstages gemeinsam in einem Warteraum gesammelt. Zwischen den einzelnen Prüfungstagen wurden die OSCE-Parcours inhaltlich leicht modifiziert, um ebenfalls eine Informationsweitergabe zu erschweren. Die Identifikation der Medizinstudenten erfolgte mittels Personalausweis. Die von den Medizinstudenten mitzubringenden Utensilien umfassten Arztkittel, Stethoskop und Reflexhammer. Bevor die Studenten den Parcours betraten, wurden sie mittels eines Parcours-Planes über die Lage der Stationen informiert und in den zeitlichen und räumlichen Ablauf unterwiesen.

An den Stationen selbst wurde zuerst der Namen der Studenten erfragt und dieser auf der Checkliste notiert. Mit dem ersten Signalton wurde die Aufgabenstellung vom Prüfer wörtlich vorgelesen. Diese stand den Studenten zusätzlich in schriftlicher Version an den einzelnen Prüfungsstationen zur Verfügung. Mit dem zweiten Signalton endete die Prüfungszeit und die erreichte Gesamtsumme wurde vom Prüfer notiert. Der Medizinstudent erhielt ein kurzes Feedback (ca. 30s) zu seiner Prüfungsleistung ohne die Nennung des erreichten Punktwertes, bevor er zur nächsten Station gebeten wurde. Anschließend wurde die Checkliste vom Prüfer unterschrieben und in die vorgesehene Ablage gelegt.

• Die Prüfer

Für jede Prüfungsstation des OSCE wurde ein Prüfer aus dem jeweiligen Fachgebiet festgelegt. Alle Prüfer erhielten direkt vor dem Beginn der Prüfung eine 30-minütige Einführung in den Ablauf der Prüfung, das Procedere an den Prüfungsstationen sowie die erwarteten Lösungen der Aufgabenstellungen. Alle Prüfer prüften zwei OSCE-Durchläufe und wurden in der Pause mit Essen und Trinken versorgt. Da mit zwei parallelen Parcours geprüft wurde, wurden pro Tag 24 Prüfer über eine Zeit von ca. 3,5 Stunden (inklusive Einführung und Pause) benötigt, um 48 Studenten zu prüfen.

• Die Helfer und standardisierten Patienten

Für den Aufbau der beiden 12-Stationen-OSCEs war die Hilfe von vier Personen über ca. drei Stunden erforderlich. Während des OSCE war für beide Parcours ein "Zeitgeber" für die Einhaltung der Prüfungszeiten und die Abgabe des Signaltons sowie eine Person als zentrale Aufsicht und als Ansprechpartner bei auftretenden Problemen verantwortlich. Eine weitere Person pro Parcours war für das Einsammeln der Checklisten und deren Kontrolle auf Vollständigkeit zuständig, darüber hinaus achtete eine zusätzliche Person pro Parcours darauf, dass die Studenten problemlos zur nächstfolgenden Prüfungsstation fanden. Eine Person über dieselbe Zeit diente der Kontrolle, Einführung und Betreuung der Studenten vor ihrem jeweiligen Prüfungsbeginn, so dass insgesamt vier Helfer zum Aufbau sowie sieben Helfer während der eigentlichen Prüfungszeit benötigt wurden. Dass lediglich ein "Zeitgeber" und eine zentrale Aufsicht zum Einsatz kommen musste lag an der räumlichen Nähe der beiden Parcours. An jeweils zwei Prüfungsstationen waren standardisierte Patienten (SP) [4] für die Prüfung kommunikativer Fähigkeiten und an einer Station ein SP für die Prüfung körperlicher Untersuchungstechniken eingesetzt. Nicht berücksichtigt bleibt in dieser Aufstellung der Zeitaufwand für die

Erstellung der Prüfungsstationen (je nach Vorerfahrung ca. 3 Stunden pro Station), für die im Vorfeld notwendige Prüfungsorganisation, für das Standard Setting (ca. 2,5 Stunden) sowie für die Auswertung der Prüfungsergebnisse.

• Bestehensgrenze, Leistungsbewertung und statistische Analyse

Die Bestehensgrenze des OSCE wurde anhand des zuvor beschriebenen Standard Setting bestimmt. Die Punktwerte über der Bestehensgrenze wurden in vier gleich große Anteile geteilt und den Notenwerten 1 bis 4 zugeteilt. Die Reliabilität des OSCE wurde mittels Cronbach alpha auf Basis der an den einzelnen Prüfungsstationen erreichten Punktwerte bestimmt. Zudem wurde für die einzelnen Stationen die durchschnittlich erreichte Punktzahl, die Standardabweichung, die Trennschärfe (Korrelation der Punktzahl mit der Gesamtpunktzahl an *allen* Stationen) und die korrigierte Trennschärfe (Korrelation der Punktzahl mit der Summe *aller anderen* Stationen) errechnet. Das Abschneiden an den einzelnen Prüfungstagen wurde mittels Varianzanalyse verglichen.

Ergebnisse

• Stichprobenbeschreibung

Im Wintersemester 2004 / 2005 nahmen n = 143 Medizinstudenten des 6. Studiensemesters des interdisziplinären Blocks Innere Medizin (53 Männer, 88 Frauen) mit einem Durchschnittsalter von 24,7 Jahren (Männer 25,4 J., Frauen 24,2 J.) teil.

• Ergebnisse des Standard Setting und Bestehensgrenze

Die Bestehensgrenze für die einzelnen OSCE-Stationen wird aus Tabelle 3 ersichtlich. Die Mittelwertbildung über die 12 Stationen ergab für den OSCE des WS 2004 / 2005 eine Bestehensgrenze von 188 Punkten.

Tabelle 3: Prüfungsstationen des OSCE der Medizinischen Universitätsklinik Heidelberg Wintersemester 2004/2005: Bestehensgrenze nach Standard Setting, durchschnittliche erreichte Punktzahl, Standardabweichung, Trennschärfe und korrigierte Trennschärfe (n=143).

Station	Bestehensgrenze nach Standard Setting	Durchschnittlich erreichte Punktzahl	Standardabweichung	Trennschärfe	Korrigierte Trennschärfe
Endokrinologie	15,4	21,04	2,82	0,59	0,49
Psychosomatik	15,7	20,43	2,57	0,53	0,43
Kardiologie	17,9	21,63	3,51	0,58	0,45
Pulmonologie	16,1	21,93	2,68	0,51	0,41
Gastroenterologie	16,5	21,10	3,48	0,66	0,54
Rheumatologie	12,5	21,41	3,36	0,55	0,42
Hämatologie	13,3	20,71	3,53	0,44	0,28
Allgemeinmedizin	17,0	19,30	2,42	0,25	0,14
Nephrologie	16,4	20,22	3,54	0,68	0,57
Körperliche U.	15,2	21,62	2,43	0,39	0,28
Braunüle	15,7	22,75	2,63	0,45	0,34
Klinische Chemie	15,7	17,46	5,30	0,56	0,34

• Prüfungsergebnisse und Notenvergabe

Die im Durchschnitt erreichten Punktzahlen an den 12 Prüfungsstationen ist ebenfalls Tabelle 3 zu entnehmen. Im Durchschnitt wurde von den Studenten eine Gesamtpunktzahl von 249,6 (SD = 20,2; min. = 172, max. = 284) erreicht. Studenten, die am ersten, zweiten und dritten Prüfungstag den OSCE durchliefen unterschieden sich in ihren Prüfungsergebnissen nicht voneinander (Varianzanalyse: $F(2, 140) = 0,844$, $p = 0,432$). Daten zur Aufgabenschwie-

rigkeit und Trennschärfe [35] finden sich ebenfalls in Tabelle 3. Die Reliabilität des OSCE belief sich auf ein Cronbach alpha von 0,748.

Die prozentuale Verteilung der Noten entsprechend der erreichten Gesamtpunktzahlen ist aus Tabelle 4 ersichtlich. Ein Student (0,7%) blieb unterhalb der Bestehensgrenze und erreichte das Prüfungsziel nicht.

Tabelle 4: Daten zur Notenvergabe OSCE Wintersemester 2004/2005 (n=143)

Punktebereich	Note	n (%) der Studenten mit entsprechender Note
272 – 300	1	19 (13,3%)
244 – 271	2	79 (55,2%)
216 – 243	3	37 (25,9%)
188 – 215	4	7 (4,9%)
0 – 187	nicht bestanden	1 (0,7%)

Diskussion

Der OSCE stellt eine Prüfungsmethode dar, die sich sehr gut zur Prüfung von wesentlichen klinisch-praktischen und kommunikativen Kompetenzen zukünftiger Ärzte eignet. Vor allem in Folge der Novellierung der Approbationsordnung für Ärzte [3] wird die Durchführung einer klinisch-praktischen Prüfung zunehmend an Bedeutung gewinnen. Allerdings ist die OSCE-Prüfung bei den Studentenzahlen deutscher Fakultäten mit organisatorischen und logistischen Herausforderungen verknüpft. Das Anliegen der vorliegenden Arbeit war eine zeit- und ressourcenschonende Implementierung an anderen Fakultäten in Hinblick auf Konzeption, Durchführung und Notenvergabe zu erleichtern.

Die Bestehensgrenze für den an der Medizinischen Universitätsklinik Heidelberg durchgeführten OSCE wurde mit einem modifizierten Standard Setting Verfahren nach Angoff [2], [17] festgelegt. Wenngleich dieses zusätzliche Zeit- und Personalressourcen beansprucht und im Vergleich zur Borderline-Methode [30] einen hypothetischen Zugang zur Festlegung einer Bestehensgrenze präferiert [17], bietet sie deutliche Vorteile gegenüber der Borderline-Methode: So kann der Notwendigkeit der Ankündigung der Bestehensgrenze vor Prüfungsbeginn leichter Rechnung getragen werden und die Bildung von Expertenteams zur Festlegung der Bestehensgrenze dient darüber hinaus der Curriculumentwicklung und dem Einbezug von Kollegen in verantwortliche Entscheidungsprozesse.

In der Konzeption des OSCE entschieden wir uns für ein neues Checklistenformat, welches die Vorteile eines Checklisten-Ratings und eines globalen Ratings integriert, um höhere Kompetenzebenen mit zu erfassen [41][53] sowie für die Durchführung des OSCE in einer kompensatorischen Form. Hiermit sollte vermieden werden, dass schlechter konzipierte Stationen zum Nachteil für "kompetente" Studenten werden (vgl. [9], [17]), zumal bisher unklar ist, welcher Zugang der validere ist [37]. Der vorgestellte OSCE beinhaltete sowohl Prüfungsstationen mit hoher Trennschärfe, die gut zwischen "kompetenten" und "nicht-kompetenten" Studenten unterschieden, als auch Stationen, die elementares Grundwissen mit geringerer Trennschärfe abprüften.

99,3% der Medizinstudenten erreichten die festgelegte Bestehensgrenze unseres OSCE. Äquivalent zu Berichten in der Literatur (vgl. [33]) lagen die erreichten Punktwerte der Prüflinge im oberen

Wertebereich. Eine differenzierte Notenvergabe war problemlos möglich. Obwohl sich die gesamte Prüfung über drei Tage hinzog, unterschieden sich die studentischen Prüfungsleistungen an den unterschiedlichen Prüfungstagen nicht. Die Erfahrung zeigt, dass eine Informationsweitergabe mit nachfolgenden Täuschungsversuchen eher zu einem schlechteren Abschneiden führt, da die Studenten auf eine Modifikation der Stationen nicht vorbereitet sind und in dem Versuch zuvor Gehörtes zu reproduzieren nicht mehr so ungedrungen und frei auftreten. Das informelle Feedback von studentischer Seite und von Prüfern zeigt, dass der OSCE als Prüfung sowohl den Prüflingen als auch den Prüfern Spaß bereitet und auf beiden Seiten Zufriedenheit hinterlässt. Diese in Fokusgruppen zusammengetragene Resonanz bezieht sich dabei sowohl auf die Art der Prüfungsform, die Prüfungsinhalte als auch auf das Feedback während der Prüfung selbst.

Aufgrund der hohen Anzahl von zu prüfenden Studenten war in Hinblick auf die Personalintensität des OSCE notwendig, Prüfer mit einzubeziehen, die noch nicht genügend Erfahrung mit dieser Prüfungsform sammeln konnten. Hilfreich war hier das obligatorische Prüfertraining direkt vor dem OSCE. Aufgrund der unmittelbaren zeitlichen Nähe zur Prüfung und den direkten Bezug zur zu prüfenden Station, erwies sich dieses als sehr effizient.

Der vorgestellte OSCE (Abbildung 3) erweist sich als reliables und faires Prüfungsinstrument, welches wichtige ärztliche Basisfertigkeiten direkt prüft, Medizinstudenten und Prüfern Spaß bereitet und nicht zuletzt durch das integrierte Feedback einen wichtigen Lerneffekt für die Studenten beinhaltet.

Unsere Empfehlung für die Implementierung eines OSCE:

- Erstellung eines Blueprints
- 10 – 14 Prüfungsstationen
- Kurze Prüfungsstationen mit ca. 5 Minuten Prüfungsdauer, 30s Feedback und 30s Wechselzeit
- Kombinierte Checklisten mit Checklisten-Rating und Globalem Rating
- Modifiziertes Standard Setting mit Festlegung der minimalen Kompetenz
- Durchführung des OSCE als kompensatorische Prüfung
- Wiederholung des OSCE bei Nicht-Bestehen mit dem nächsten Semester
- 30 Minuten Einführung für die Prüfer direkt vor der Prüfung.

Abbildung 3: Empfehlung

Korrespondenzadresse:

• Dr. med. Christoph Nikendei, Medizinische Universitätsklinik Heidelberg, Ludolf-Krehl-Klinik, Klinik für Allgemeine Klinische Medizin und Psychosomatik, Im Neuenheimer Feld 410, 69120 Heidelberg, Deutschland, Tel.: 06221/56-38663, Fax: 06221/56-5988
christoph_nikendei@med.uni-heidelberg.de

Literatur:

- [1] Allen R, Heard J, Savidge M, Bittergele J, Cantrell M, Huffmaster T. Surveying Students' Attitudes During the OSCE. *Adv Health Sci Educ Theory Pract.* 1998;3:197-206.
- [2] Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL, editor. *Educational measurement* (2nd ed.), Washington, DC: American Council on Education; 1971; p. 508-600.

- [3] Approbationsordnung für Ärzte, Beschluss des Bundesrates vom 26.4.2002. Bonn: Bundesanzeiger Verlagsgesellschaft mbH, Drucksache 316/02; 2002.
- [4] Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med.* 1993;68(6):443-453.
- [5] Boulet JR, de Champlain AF, McKinley DW. Setting defensible performance standards on OSCEs and standardized patient examinations. *Med Teach.* 2003;25:245-249.
- [6] Campos-Outcalt D, Watkins A, Fulginiti J, Kutob R, Gordon P. Correlations of family medicine clerkship evaluations and Objective Structured Clinical Examination scores and residency directors' ratings. *Fam Med.* 1999;31:90-94.
- [7] Cerilli GJ, Merrick HW, Staren ED. Objective Structured Clinical Examination technical skill stations correlate more closely with postgraduate year level than do clinical skill stations. *Am Surg.* 2001;67:323-326.
- [8] Chambers KA, Boulet JR, Gary NE. The management of patient encounter time in a high-stakes assessment using standardized patients. *Med Educ.* 2000;34:813-817.
- [9] Chesser AM, Laing MR, Miedzybrodzka ZH, Brittenden J, Heys SD. Factor analysis can be a useful standard setting tool in a high stakes OSCE assessment. *Med Educ.* 2004;38:825-831.
- [10] Cohen-Schotanus J, van der Vleuten CPM, Bender W. Een betere cesuur bij tentamens: de beste studenten als referentiepunt [A better cutoff for examinations: the best students as a reference point]. In: Ten Cate TJ, Dikkers JH, Houtkoop E, Pollemans MC, Pols J, Smal JA, (Hrsg.) *Gezond Onderwijs [Health Education]*, Vol. 5. Houten/ Diegem, The Netherlands: Bohn Stafleu van Loghum. 1996;83-88.
- [11] Colliver JA, Willis MS, Robbs RS, Cohen DS, Swartz MH. Assessment of empathy in a standardized-patient examination. *Teach Learn Med.* 1998;10:8-11.
- [12] Cusimano MD, Rothman AI. Consistency of standards and stability of pass/fail decisions with examinee-based standard-setting methods in a small-scale objective structured clinical examination. *Acad Med.* 2004;79:25-27.
- [13] Dauphinee WD, Blackmore DE, Smee SM, Rothman AI, Reznick RK. Using the judgments of physician examiners in setting the standards for a national multi-center high stakes OSCE. *Adv Health Sci Educ Theory Pract.* 1997;2:201-211.
- [14] Davis MH. OSCE: the Dundee experience. *Med Teach.* 2003;25:255-261.
- [15] Dreyfus HL, Dreyfus SE. *Mind over Machine.* New York: Free Press; 1986.
- [16] Ebel RL. *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice Hall; 1972.
- [17] Friedman Ben-David M. Standard setting in student assessment. *AMEE Med. Educ. Guide No. 18.* Dundee: AMEE; 2000.
- [18] Hamann C, Volkan K, Fishman MB, Silvestri RC, Simon SR, Fletcher SW. How well do second-year student learn physical diagnosis? Observational study of an objective structured clinical examination (OSCE). *BMC Med Educ.* 2002;2:1.
- [19] Harden RMCG, Stevenson M, Wilson Downie W, Wilson GM. Assessment of Clinical Competence using Objective structured Examination. *Brit Med J.* 1975;1:447-451.
- [20] Hodder R, Rivington R, Calcutt LE, Hart IR. The effectiveness of immediate feedback during the OSCE. *Med Educ.* 1989;23:184-188.
- [21] Hodges B, McIlroy JH. Analytic global OSCE ratings are sensitive to level of training. *Med Educ.* 2003;37:1012-1016.
- [22] Hodges B, Regehr G, Hanson M, McNaughton N. Validation of an objective structured clinical examination in psychiatry. *Acad Med.* 1998;73:910-912.
- [23] Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. Checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74:1129-1134.
- [24] Jaeger RM. An interactive structures judgement process for establishing standards on competency tests. *Theory and application. Educ Eval Policy Anal.* 1982;4:461-476.
- [25] Jünger J, Köllner V. Integration eines Kommunikationstrainings in die klinische Lehre. *Psychother Psychosom Med Psychol.* 2003;53:56-64.
- [26] Jünger J, Nikendei C. *OSCE Prüfungsvorbereitung Innere Medizin.* Stuttgart: Thieme; 2005.
- [27] Jünger J, Schäfer S, Schellberg D, Roth C, Ben-David M, Nikendei C. Effects of Basic Clinical Skills Training on OSCE performance among medical students: a group control design study. *Med Educ.* 2005;39:1015-1020.
- [28] Kaufman DM, Mann KV, Muijtjens AMM, van der Vleuten CPM. A Comparison of Standard-setting Procedures for an OSCE in Undergraduate Medical Education. *Acad Med.* 2000;75:267-271.

- [29] Lieberman ME, Volkan K, Shaffer K, Novelline R, Lang EV. Clinical competence assessment in radiology: introduction of an objective structured clinical examination in the medical school curriculum. *Acad Radiol.* 2001;8:74-81.
- [30] Livingston SA, Zieky MJ. *Passing scores: A manual for setting standards of performance on educational and occupational tests.* Princeton, New Jersey: Educational Testing Services; 1982.
- [31] Mavis B, Cole B, Hoppe R. A survey of student assessment in U.S. medical schools: The balance of breadth versus fidelity. *Teach Learn Med.* 2001;13:74-79.
- [32] McLroy JH, Hodges B, McNaughton N, Regehr G. The effect of candidates' perceptions of the evaluation method on reliability of checklist and global rating scores in an objective structured clinical examination. *Acad Med.* 2002;77:725-728.
- [33] McLachlan JG, Whiten SG. Marks, scores and grades: scaling and aggregating student assessment outcomes. *Med Educ.* 2000;34:788-797.
- [34] Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* 1990;65:63-67.
- [35] Möltner A. Die quantitative Analyse einer Prüfung. *GMS Z Med Ausbild.* 2006;23(3): Doc. 52.
- [36] Nedelsky L. Absolute grading standards for objective tests. *Educ Psychol Meas.* 1954;14:13-19.
- [37] Newble D. Techniques for measuring clinical competence: objective structured clinical examinations. *Med Educ.* 2004;38:199-203.
- [38] Nikendei C, Schilling T, Nawroth P, Hensel M, Ho AD, Schwenger V, Ritz E, Herzog W, Schellberg D, Katus HA, Dengler T, Stremmel W, Müller M, Jünger J. Integriertes Skills-Lab-Konzept für die studentische Ausbildung in der Inneren Medizin. *Dtsch Med Wochenschr.* 2005a;130:1133-1138.
- [39] Nikendei C, Zeuch A, Dieckmann P, Roth C, Schäfer S, Völkl M, Schellberg D, Herzog W, Jünger J. Role-playing for a More Realistic Technical Skills Training. *Med Teach.* 2005b;27(2):122-126.
- [40] Nikendei C, Zipfel S, Roth C, Löwe B, Herzog W, Jünger J. Kommunikations- und Interaktionstraining im Psychosomatischen Praktikum: Einsatz von standardisierten Patienten. *Psychother Psychosom Med Psychol.* 2003;53(11):440-445.
- [41] Norman GR, van der Vleuten CPM, de Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ.* 1991;25:119-126.
- [42] Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalisability of OSCE measures across content domains. *Acad Med.* 1999;74:1320-1322.
- [43] Regehr G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med.* 1998;73:993-997.
- [44] Reznick R, Blackmore D, Dauphinee WD, Rothman A, Smees S. Large-scale high-stakes testing with an OSCE. Report from the Medical Council of Canada. *Acad Med.* 1996;71:19-21.
- [45] Reznick R, Regehr G, Yee G, Rothman A, Blackmore D, Dauphinee D. Process rating forms versus task specific checklists in an OSCE for medical licensure. *Acad Med.* 1998;73:97-99.
- [46] Rothman AI, Blackmore D, Dauphinee WD, Reznick R. The use of global ratings in OSCE station scores. *Adv Health Sci Educ Theory Pract.* 1997;1:215-219.
- [47] Schmidt HG, Norman GR, Boshuizen E. A cognitive perspective on medical expertise: theory and implications. *Acad Med.* 1990;65:611-621.
- [48] Searle J. Defining competency - the role of standard setting. *Med Educ.* 2000;34:363-366.
- [49] Simon SR, Volkan K, Hamann C, Duffey C, Fletcher SW. The relationship between second-year medical students' OSCE scores and USMLE Step 1 scores. *Med Teach.* 2002;24:535-539.
- [50] Swanson DB, Norman GR, Linn RL. Performance-based assessment: Lessons learnt from the health professions. *Educ Res.* 1995;24:5-11.
- [51] Swing S, Bashook PG. Toolbox of Assessment Methods. ACGME Outcomes Project. Accreditation Council for Graduate Medical Education (ACGME) & American Board of Medical Specialties (ABMS); Version 1.1; 2000. Zugänglich unter: <http://www.acgme.org/outcome/assess/toolbox.pdf>.
- [52] Townsend AH, McIlvenny S, Miller CJ, Dunn EV. The use of an objective structured clinical examination (OSCE) for formative and summative assessment in a general practice clinical attachment and its relationship to final medical school examination performance. *Med Educ.* 2001;35:841-846.
- [53] Van der Vleuten CPM, Norman GR, de Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ.* 1991;25:110-118.
- [54] Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med.* 1990;2:58-76.
- [55] Van der Vleuten CPM. Reliability of OSCEs. University of Maastricht: AMEE conference, 29th August - 1st September 2002. Maastricht: AMEE; 2002. Zugänglich unter: <http://www.fdg.unimaas.nl/educ/cees/amee>.
- [56] Van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ.* 1996;1:41-67.
- [57] Wass V, van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet.* 2001;357:945-949.
- [58] Wijnen WHFW. *Onder of Boven de Maat [Missing or Hitting the Mark].* PhD dissertation. Groningen, The Netherlands: University of Groningen; 1971.
- [59] Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. *Acad Med.* 2003;78:219-223.