

Im Notfall bereit? OSCE in der studentischen Notfallausbildung

Zusammenfassung

Hintergrund: Multiple-Choice (MC)-Klausuren in der Notfallmedizin sind zum Überprüfen praktischer Fertigkeiten nicht geeignet. Trotzdem wird dieses Prüfungsformat noch häufig eingesetzt. An der Universität zu Lübeck besteht der interdisziplinäre Notfallkurs aus 12 Modulen, 6 werden anästhesiologisch betreut. Diese Studie beleuchtet die erstmalige Durchführung eines praktischen 6-Stationen-OSCE (Objective Structured Clinical Examination) „Notfallmedizin“ unter methodischen Gesichtspunkten und hinsichtlich studentischer Zufriedenheit.

Material und Methoden: Die OSCE-Prüfungsaufgaben wurden lernzielbasiert entwickelt und im Vorwege getestet. Die gesamte OSCE-Prüfung dauerte zwei Tage. Prüflinge wurden randomisiert auf Prüfungsgruppen verteilt und schrieben eine Woche später eine MC-Klausur mit jeweils zwei Fragen aus jedem Modul ($n=24$). Der OSCE wurde mittels standardisierten Fragebogens (87 Einzelfragen in 9 Frageblöcken) quantitativ und qualitativ evaluiert.

Ergebnisse: Es nahmen 109 Studierende (68,8% w, 31,2% m) des überwiegend 5. und 6. Semesters teil. Die Innere Konsistenz (Cronbach's α) für OSCE betrug ,672. MC-Klausur und OSCE korrelieren signifikant ($r=.310$, $p=.001$) bei geringer „Schnittmenge“ (aufgeklärte Varianz: 9,61%). Alle Prüfungsaufgaben trennen gut zwischen Studierenden mit niedriger/ hoher Punktzahl ($r' \geq 0,3$) bei akzeptabler

Prepared for an emergency? An OSCE in medical student training in emergency medical care

P. Iblher^{1,2} · M. Zupanic² · C. Härtel³ · A. Parbel¹ · M. Prüßmann¹ · H. Heinze¹ · M. R. Fischer² · P. Schmucker¹

Schwierigkeit der Aufgaben (.78 - ,85). Die Evaluation zeigte überwiegend positive Bewertungen der OSCE-Prüfung, in Freitext-Antworten wurden vor allem mehr praktische Übungsmöglichkeiten und mehr OSCE-Prüfungen gewünscht.

Schlussfolgerungen: Die OSCE-Prüfung wurde von den Teilnehmern positiv evaluiert mit dem Wunsch nach mehr praktischen Übungen und mehr OSCE-Prüfungen. Diese Untersuchung zeigt, dass bei Einhaltung eines strukturierten Reviewprozesses während der Aufgabenerstellung befriedigende Ergebnisse hinsichtlich methodischer Kriterien erreicht werden können.

Summary

Background: The use of Multiple Choice Questionnaires (MCQ) for examination purposes in emergency medical care is not appropriate for assessing practical competence. Nevertheless, this approach is still commonly employed. At the University of Luebeck, medical students have to participate in an interdisciplinary 12-module course in emergency medical care, in which 6 tutorials are supervised by anaesthetists. The aim of the present study was to examine the first-time implementation of a six-station Objective Structured Clinical Examination (OSCE) in this field with a focus on methodical aspects and student satisfaction.

Materials and Methods: Examination questions focusing on learning objectives were developed and tested previously. The entire OSCE assessment lasted two days. Medical students were randomly

¹ Klinik für Anästhesiologie
Universität zu Lübeck

² Institut für Didaktik und Bildungsfor-
schung im Gesundheitswesen (IDBG),
Private Universität Witten/Herdecke

³ Klinik für Kinder- und Jugendmedizin
Universität zu Lübeck

Danksagung

Allen beteiligten Studierenden, den Dozenten und Prüfern, den studentischen Schauspielern und Hilfskräften, allen OSCE-Entwicklern sowie Frau Richter aus dem Sekretariat der Klinik für Anästhesiologie und dem Studiendekanat der Universität zu Lübeck sei herzlich gedankt für Unter-stützung, Engagement, Kreativität und Konstruktivität!

Diese Publikation ist Teil der Dissertation von Altje Parbel und allen Kolleginnen und Kollegen gewidmet, die sich täglich engagiert für eine Verbesserung der Lehre einsetzen.

Schlüsselwörter

OSCE – Prüfung – Notfallmedi-
zin – Studierendenausbildung –
Evaluation

Keywords

OSCE – Assessment – Emergen-
cy Medicine – Undergraduate
Education – Evaluation

allocated to examination groups and, one week later, underwent MCQ-assessment involving two questions from each tutorial (n=24). Qualitative and quantitative evaluation of the OSCE was based on a standardised questionnaire (87 items in 9 question blocks).

Results: 109 students (68.8% f, 31.2% m) participated in the study. Cronbach's α was .672. MCQ and OSCE correlated significantly ($r=.310$, $p=.001$) 9.61% of variance was clarified. Corrected item-total correlation ($r' \geq 0.3$) was good, and item difficulty (.78 - .85) was acceptable. In the main, the OSCE met with approval, and students expressed a desire for more practical training and more OSCEs.

Conclusions: The OSCE was evaluated positively by the students, who would welcome more such tests, and more practice in their medical training. This study shows that a structured review process during OSCE-preparation would appear to help the student to achieve satisfactory results with regard to methodical criteria.

Einleitung

Einer aktuellen Umfrage zufolge organisieren die Kliniken für Anästhesiologie an 32 von insgesamt 38 deutschen Medizinischen Fakultäten die Ausbildung der Medizinstudierenden im „Querschnittsbereich Notfallmedizin“ [1]. Die Notfallmedizin stellt dabei einen Bereich der Medizin dar, in dem Handlungskompetenz und praktische Fertigkeiten sehr wichtig sind und die in den Grundlagen von allen Ärztinnen und Ärzten unabhängig von ihrer Fachdisziplin beherrscht werden sollten. Während die strukturierte und lernzielbasierte Kursplanung dabei essentiell ist für eine gute und einheitliche Lehre, kommt der Evaluation der Lernziele in einer Prüfung eine sehr hohe Bedeutung zu. Nur sie garantiert eine hohe Qualität in der medizinischen Ausbildung durch Erfassung des Ist/Soll-Zustandes von vermittelten und erwarteten Kompetenzen (Lernziele) und hat gleichzeitig einen Einfluss auf das Lernverhalten der Studierenden

(„Assessment drives Learning“) [2]. Darüber hinaus ist die Prüfung die letzte Chance, den Studierenden noch etwas beizubringen beziehungsweise gravierende Fehler zu korrigieren, sofern ein direktes Feedback möglich ist. Es ist also notwendig, insbesondere in der Notfallmedizin Prüfungsformate zu verwenden, die relevante praktische Kompetenzen überprüfen. Beckers et al. zeigten aber in ihrer Arbeit, dass 2009 noch immer an 89% der deutschen Fakultäten Lernziele der Notfallmedizin mittels MC-Klausur abgeprüft wurden [1]. MC-Klausuren sind in der Entwicklung und Umsetzung zwar mit relativ wenig Personalaufwand und Kosten verbunden und werden daher gerne als Prüfungsform verwendet. Die wichtige Frage aber, ob die Studierenden notwendige klinisch-praktische Fertigkeiten tatsächlich beherrschen, kann dieses Prüfungsformat niemals beantworten, da es lediglich auf der untersten Kompetenzebene (Level I bis II nach Miller) Fakten- und Handlungswissen überprüft

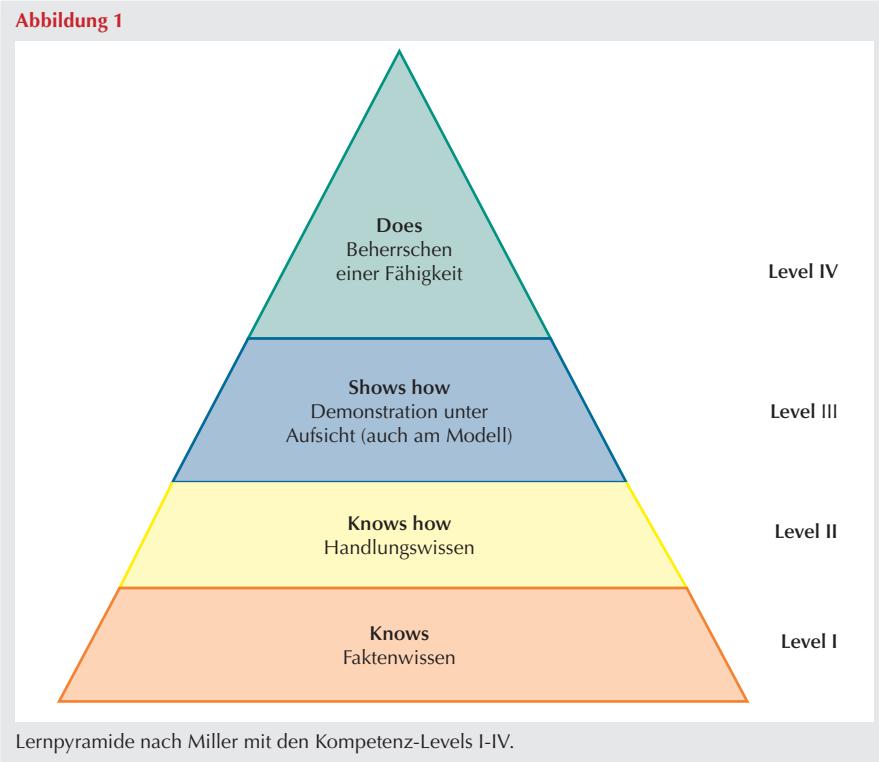
(Abb. 1) [3]. Auch mündliche Prüfungen sind letztendlich nur zur Abfrage von Wissen geeignet, obwohl hier direkt die Leistung mit Fehlern rückgemeldet werden könnte. Wünschenswert wären also Prüfungsformate, die mindestens auf der Kompetenzebene III („Shows how“) praktische Fertigkeiten überprüfen (Abb. 1), beziehungsweise eine Kombination verschiedener Prüfungsformen [2]. Hier hat sich international die Durchführung von OSCE- (Objective Structured Clinical Examination) Prüfungen bewährt. Dieses Prüfungsformat stammt ursprünglich von der schottischen Universität Dundee [4] und ist mittlerweile weltweit als zuverlässige Prüfungsmethode insbesondere in der Ausbildung von Medizinstudierenden anerkannt und etabliert [5]. Die Studierenden absolvieren hier einen Prüfungsparcours mit mehreren Prüfungsstationen, an denen definierte klinisch-praktische Aufgaben des Lernzielkatalogs zum Teil mit Schauspielpatienten zu lösen sind. An jeder Station wird die Prüfungsleistung durch einen Prüfer anhand einer Bewertungsliste beurteilt. Die „Objektivität“ und „Strukturierung“, die dieser Prüfungsform ihren Namen gibt, soll dabei durch die standardisierte Aufgabenstellung (Durchführungsobjektivität), die standardisierte Auswertung (Auswertungsobjektivität) und den Erwartungshorizont, der für alle Prüflinge gleich ist (Interpretationsobjektivität), sowie eine hohe Anzahl von Prüfungsstationen erreicht werden.

Fragestellung

Hinsichtlich der oben dargestellten Notwendigkeit eines praktischen Prüfungsformats wurde an der Universität Lübeck ein anästhesiologischer Sechs-Stationen-OSCE für den Bereich Notfallmedizin neu konzipiert und eingesetzt. Diese Arbeit untersucht die Qualität dieser Prüfung mit Hilfe von testtheoretischen Analysen und qualitativer und quantitativer studentischer Evaluation unter Verwendung eines standardisierten Fragebogens.

Methodik

Das Versuchsprotokoll wurde der Ethikkommission der Universität zu Lübeck



zur Begutachtung vorgelegt und durch diese ein zustimmendes Votum erteilt. Die teilnehmenden Studierenden erklärten im Vorwege ihr Einverständnis. An der Universität zu Lübeck wird die Lehre im Querschnittsbereich Notfallmedizin interdisziplinär unter anästhesiologischer Leitung durchgeführt. Die Studierenden absolvieren dort in halber Semesterstärke (n~120) den einsemestrigen Kurs „Notfallmedizin“ im 1./2. klinischen Semester. Die Lehrveranstaltung besteht dabei aus zwölf Modulen, von denen sechs Module anästhesiologisch und jeweils ein Modul gynäkologisch, kinderchirurgisch, neurologisch, neurochirurgisch, internistisch und unfallchirurgisch betreut werden.

OSCE-Prüfungsvorbereitungen

Im Rahmen der Sondierungs- und Planungsphase wurde festgelegt, im Sommersemester 2009 zunächst eine OSCE-Prüfung thematisch aus den sechs anästhesiologisch betreuten Modulen des „Querschnittsbereichs Notfallmedizin“ (Tab. 1) durchzuführen, um im Verlauf die anderen beteiligten Kliniken nach

und nach an den OSCE heranzuführen und so mittelfristig einen 12-Stationen-Parcour zu etablieren. Dies war insofern notwendig, da dieses Prüfungsformat zu diesem Zeitpunkt noch gar nicht an der Universität zu Lübeck genutzt wurde und entsprechend unbekannt war.

In zwei zweitägigen Workshops wurden im März 2009 sechs OSCE-Prüfungsaufgaben durch Mitarbeiter der Klinik für Anästhesiologie lernzielbasiert entworfen und getestet. Zeitgleich wurden Drehbücher für Schauspielpatienten geschrieben, die im Rahmen von OSCE-Szenarien notwendig waren. Anfang Juni 2009 wurden Mitglieder der studentischen Theatergruppe als Schauspieler (n=6) und Stationshelfer (n=4) rekrutiert und geschult. Die Erstattung der Kosten wurde durch das Studiendekanat zugesagt. Ende Juni wurde zunächst eine OSCE-Generalprobe mit allen sechs Prüfungsstationen durchgeführt und in einer sich anschließenden Review-Konferenz mit Prüfern, „Prüflingen“ und Schauspielpatienten zum einen die Qualität der Prüfungsstationen beurteilt und Verbesserungsvorschläge erarbeitet, zum

Tabelle 1

Übersicht über die sechs OSCE-Stationen mit entsprechenden Prüfungslehrzielen (P=Puppe, SP=Schauspielpatient).

Station	Inhalt	Prüfungslehrziele Der Studierende...	Modell
1	Basic Life Support (BLS)	<ul style="list-style-type: none"> demonstriert den diagnostischen Block am Modell demonstriert die Herzdruckmassage am Modell und benennt vier häufige Fehler erläutert den BLS-Algorithmus des ERC für Erwachsene erläutert häufige Ursachen für Herzkreislaufstillstand und deren Therapien beim Erwachsenen 	P
2	Advanced Cardiac Life Support (ACLS)	<ul style="list-style-type: none"> erläutert den ACLS-Algorithmus demonstriert die Anlage von EKG-Elektroden und erkennt im EKG „Kammerflimmern“ demonstriert eine sichere Defibrillation am Modell benennt fünf reversible Ursachen des Kreislaufstillstandes und deren Therapien erläutert Abbruchkriterien für eine Reanimation 	P
3	Traumamanagement	<ul style="list-style-type: none"> führt eine orientierende körperliche Untersuchung durch und benennt Verdachtsdiagnosen erläutert mögliche Verletzungen und deren Folgen bei einem Motorradunfall erläutert dem Patienten das weitere Vorgehen im Rahmen der Notfallversorgung demonstriert Helmabnahme und Anlage eines Stiff-neck 	SP
4	Kindernotfälle	<ul style="list-style-type: none"> demonstriert die korrekte Durchführung einer kardio-pulmonalen Säuglingsreanimation (BLS) erläutert Unterschiede in der Reanimation von Säuglingen und Erwachsenen benennt vier Kriterien für den Beginn einer Neugeborenen-Reanimation 	P
5	Fallbeispiel „Herzinfarkt“	<ul style="list-style-type: none"> erhebt eine gezielte Notfallanamnese und führt die Erstversorgung durch kennt therapierelevante Differentialdiagnosen bei akutem Brustschmerz kennt 3 Therapieprinzipien des Akuten Coronarsyndroms (ACS) und kann 3 Maßnahmen zur Notfallbehandlung nennen 	SP
6	Atemwegsmanagement	<ul style="list-style-type: none"> demonstriert das Zusammenstellen und Überprüfen des zur Intubation notwendigen Materials nennt fünf Möglichkeiten zur Kontrolle des Intubationserfolges und bewertet diese hinsichtlich ihrer Sicherheit demonstriert die Maskenbeatmung am Modell und erläutert Risiken im Gegensatz zur Intubation 	P

anderen wurden die Bestehenskriterien mittels modifizierter Angoff-Methode festgelegt [5]. Die Stationsbewertungen bestanden aus drei bis fünf Teilaufgaben, die summiert immer 25 Punkte pro Station ergaben, so dass insgesamt bei sechs Stationen 150 Punkte zu erreichen waren.

Prüfungsregime

Die Studierenden wurden randomisiert einem der zwei aufeinanderfolgenden Prüfungstage und einer Prüfungsgruppe zugeordnet und schrieben eine Woche nach der OSCE-Prüfung eine MC-Klausur mit jeweils zwei Fragen aus jedem Modul ($n_{\text{gesamt}}=24$). Der OSCE wurde mittels Fragebogen quantitativ und qualitativ evaluiert.

OSCE-Prüfungsablauf

Die Studierenden erhielten zu Beginn des Semesters ein Skript mit Lernzielen und Kursdetails und wurden im Vorfeld der Prüfung schriftlich über Prüfungstermin, -ort, -ablauf und organisatorische

Besonderheiten des OSCE informiert. Eine komplette Rotation durch die sechs Prüfungsstationen dauerte insgesamt 36 Minuten (jeweils fünf Minuten Prüfung, eine Minute Wechselzeit) (Abb. 2). Der Prüfungsparcour war doppelt aufgebaut, so dass pro Prüfung zwölf Studierende den OSCE-Parcour absolvieren konnten. Die zwölf Prüfer, von denen niemand über Prüfungserfahrung im OSCE verfügte, wurden eine Stunde vor Prüfungsbeginn im Sinne eines Prüfertrainings standardisiert instruiert und in den Prüfungsablauf eingewiesen [5]. Jedem Prüfer lag ein standardisierter Bewertungsbogen zu den Stationsaufgaben vor (mit wechselweise Checklisten- und Global-Rating) [6]. Während Checklisten Items aufführen, die objektiv „abgehakt“ und mit Punkten bewertet werden (genannt/nicht genannt), ermöglicht das globale Rating eine semi-standardisierte Beurteilung unter Vorgabe von Items, die unter dem subjektiven Eindruck des Prüfers hinsichtlich vermittelter Kompetenz und Souveränität bewertet werden.

Qualitative und quantitative Evaluation der OSCE-Prüfung

Zur Evaluation der OSCE-Prüfung wurde ein Fragebogen mit neun Frageblöcken (Skalen) und insgesamt 87 Einzelfragen (Items) konstruiert, der gegliedert war in:

- die einzelnen Prüfungs-Stationen mit jeweils sechs Items (Skala I-VI), einer Schulnotenbewertung und jeweils zwei Freitext-Items „Was könnte man Ihrer Meinung nach an dieser Station verbessern?“ und „Was würden Sie sich wünschen, damit Sie noch besser auf diese OSCE-Station vorbereitet sind?“,
- Skala VII „Selbstzweifel“ mit 5 Items zum Selbstwert [7] und 3 Items zur beruflichen Selbstwirksamkeit [8],
- Skala VIII mit 8 Items zum OSCE-Parcours im Ganzen,
- Freitextoption hinsichtlich Anregungen, Lob oder Kritik zu Lehrveranstaltung oder Prüfung und
- Skala IX mit 3 Items zum Fragebogen.

Als Kontrollvariablen wurden a) Kursgruppen-Nr., b) Prüfungsgruppen-Nr., c)

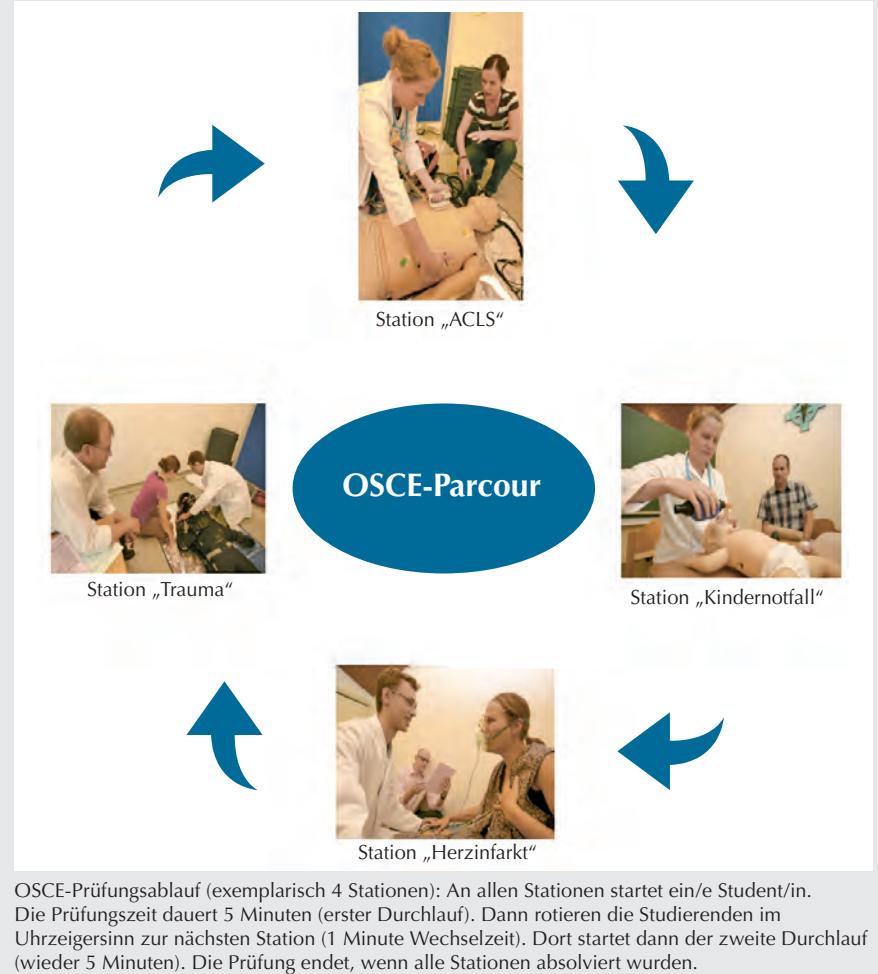
Alter, d) Fachsemester und e) Geschlecht sowie f) die Frage: „Sind Sie Sanitäter/Rettungssanitäter/-assistent?“ mitgeführt.

Alle Fragen bis auf die Freitextantworten und die Fragen zur Bewertung der Station mittels Schulnote (1-5) basierten auf einer 5-stufigen verbalisierten Likert-Skala (1 = trifft nicht zu, 2 = trifft eher nicht zu, 3 = teils/teils, 4 = trifft eher zu, 5 = trifft voll zu). Die Einzelitems der Skalen I-VIII sind in Tabelle 2 dargestellt. Die Evaluation erfolgte unmittelbar nach Ende der Prüfung in einem separaten Raum.

Statistische Auswertung

Alle Prüfungsergebnisse (OSCE und MC-Klausur) und Evaluationsdaten wurden mit SPSS 18.0 erfasst und ausgewertet. Die Normalverteilung der Daten wurde mittels Kolmogorov-Smirnov-Test überprüft. Es wurden deskriptive Datenanalysen, Analysen zur internen Konsistenz (Cronbach's α) nach z-Standardisierung sowie die Berechnungen zur Aufgabenschwierigkeit, zur Trennschärfe (korrigierte Pearson-Korrelation r' und Spearman) und des Diskriminationsindex der Prüfungsaufgaben durchgeführt. Die Zuverlässigkeit eines Prüfungsverfahrens ist ein entscheidendes Gütekriterium, das durch die Reliabilität beschrieben wird und die über die Berechnung der Internen Konsistenz (Cronbach's α) geschätzt werden kann. Interne Konsistenzen sind Ausdruck der Messgenauigkeit eines Instruments (hier der OSCE-Prüfung). Hohe Werte bedeuten, dass alle Einzelfragen (Items) eines Frageblocks (Skala) das gleiche messen, eine Skala sozusagen „homogen“ ist. Während für Gruppenvergleichs-Testungen eine Reliabilität ab 0,6 als ausreichend bezeichnet wird [14], sollte diese für medizinische Prüfungen mindestens 0,8 betragen [15]. Die Aufgabenschwierigkeit der Prüfungsaufgaben sollte im Bereich von 0,4-0,85 streuen, da Aufgaben mit höheren Werten als zu leicht gelten, bei niedrigeren Werten dagegen als zu schwer. Ein wesentliches Kriterium der testtheoretischen Auswertung stellt auch die Trennschärfe als Korrelationsmaß dar. Sie zeigt an, ob die Aufgaben eine Unterscheidung zwischen guten und schlechten Prüfungskandidaten ermöglicht. Dabei gelten

Abbildung 2



OSCE-Prüfungsablauf (exemplarisch 4 Stationen): An allen Stationen startet ein/e Student/in. Die Prüfungszeit dauert 5 Minuten (erster Durchlauf). Dann rotieren die Studierenden im Uhrzeigersinn zur nächsten Station (1 Minute Wechselzeit). Dort startet dann der zweite Durchlauf (wieder 5 Minuten). Die Prüfung endet, wenn alle Stationen absolviert wurden.

korrigierte Trennschärfen r' von über 0,3 als gut, von 0,2-0,3 als akzeptabel, von 0,1-0,2 als marginal und Werte unter 0,1 als schlecht [18]. Eine schlechte Trennschärfe bedeutet beispielsweise konkret, dass gute Studierende weniger Punkte erreichen als schlechte. Während die Trennschärfe r' eher die „Schärfe“ der Unterscheidung zwischen guten und schlechten Prüfungskandidaten aufzeigt, drückt der Diskriminationsindex nDI als Vergleichsmaß zwischen guten und schlechten Kandidaten eher die „Größe“ des Unterschiedes aus [18].

Bivariate Korrelationen wurden nach Spearman-Rho bestimmt und dienen grundsätzlich der Untersuchung, ob statistische Größen miteinander zusammenhängen beziehungsweise über eine gemeinsame „Schnittmenge“ (aufgeklärte

Varianz) verfügen. Zur Auswertung und Berechnung der Evaluationsergebnisse wurden für jeden Frageblock (Skala) die Summenskalen berechnet (Erläuterung in Tab. 4). Die Berechnung zur notwendigen Anzahl von Prüfungsstationen bei konstanter Reliabilität erfolgte nach Spearman-Brown.

Ergebnisse

Es nahmen 109 Studierende (68,8% w, 31,2% m) des überwiegend 5. (16,5%) und 6. (76,1%) Semesters teil. Der Altersdurchschnitt betrug im Mittel 25,4 ± 3,4 Jahre. Sieben Teilnehmer (6,4%) verfügten über die Qualifikation Sanitäter/Rettungssanitäter/ Rettungsassistent. Die Rücklaufquote der Evaluationsbogen betrug 100%.

Tabelle 2

Darstellung der abgefragten Einzelitems in der OSCE-Evaluation (Skala I-VIII).

Stationen 1 - 6 (Skala I - VI)	Skala VII: „Selbstzweifel“ Wie sehr treffen die folgenden Aussagen ihrer Meinung nach auf Sie zu?	Skala VIII Der OSCE-Parcour im Ganzen...
Bei der Prüfung habe ich mich sicher gefühlt	Alles in allem bin ich mit mir selbst zufrieden	...hat mir Spaß gemacht
Die Prüfungsfragen der Station waren verständlich und eindeutig	Manchmal neige ich dazu, mich für einen Versager/eine Versagerin zu halten	...motiviert mich, mich näher mit dem Lernthema zu befassen
Die Prüfungsinhalte der Station wurden mir im Unterricht vermittelt	Manchmal denke ich, dass ich nicht viel wert bin	...war sehr anspruchsvoll
Die Aufgaben der Station waren zeitlich gut zu bewältigen	Ich habe eine Reihe von guten Eigenschaften	...gibt mit die Möglichkeit zu zeigen, was ich drauf habe
Ich war gut auf die Inhalte der Station vorbereitet	Ich bin selten stolz auf mich selber	...sollte als Prüfungsform öfter verwendet werden
Die Prüfungsstation hat mir keine Probleme bereitet	Immer wenn ich versuche, im Studium voranzukommen, kommt etwas dazwischen, was mich hindert	...ist eine tolle Sache
Im Ganzen gebe ich dieser Station die Schulnote:	Auch wenn ich mich noch so anstrenge, schaffe ich trotzdem häufig nicht, was andere ohne Mühe können	...hilft mir zu erkennen, was ich kann und was nicht
Was könnte man Ihrer Meinung nach an dieser Station verbessern? (Freitext)	Es hat für mich wenig Sinn, dass ich mich im Studium anstrenge, weil ich doch nicht so viel erreichen kann	...ist eine sinnvolle Form der Prüfung
Was würden Sie sich wünschen, damit Sie noch besser auf die OSCE-Prüfung vorbereitet sind? (Freitext)		

Tabelle 3

Teststatistische Analysen und Skalenkennwerte der OSCE-Prüfungsaufgaben.

MW Mittelwert; **Min** Minimum; **Max** Maximum; **SD** Standardabweichung; **AS** Aufgabenschwierigkeit; **nDI** normierter Diskriminationsindex; **TS** Trennschärfe (nach Spearman); **r'** corrected item - total correlation= part-whole korrigierte TS).

	MW	Min	Max	SD	Cronbach's α nach z-Stand	Cronbach's α (wenn item gelöscht)	AS	nDI	TS	r'
1) BLS	19.45	12	25	2.80	0.650	0.641	0.778	0.14	0,578	.351
2) ACLS	21.02	7	25	2.92	0.673	0.625	0.841	0.16	0,656	.404
3) Traumamanagement	19.51	11	25	3.40	0.767	0.590	0.781	0.22	0,722	.494
4) Kindernotfälle	18.04	10	25	3.40	0.544	0.572	0.722	0.21	0,679	.542
5) Fallbeispiele	20.93	11	25	4.07	0.824	0.628	0.837	0.24	0,636	.402
6) Atemwegsmanagement	21.35	11	25	3.67	0.668	0.685	0.854	0.14	0,406	.224
OSCE gesamte Punktzahl (Schulnote) (max. 150 Pkt.)	119.67 (2.48)	74 (5)	147 (1)	13.63 (0.94)	0.672					
MC gesamte Punktzahl (Schulnote) (max. 24 Pkt.)	20.20 (2.28)	13 (5)	24 (1)	1.95 (0.73)						

Prüfungsergebnisse OSCE und MC-Klausur

Die Ergebnisse der OSCE-Prüfung erwiesen sich als unabhängig von Alter und Geschlecht und sind in Tabelle 3 und Abbildung 3 dargestellt.

Zwischen MC- und OSCE-Ergebnissen konnten signifikante Korrelationen für die Summenwerte (Punkte) und die daraus resultierenden Noten für die Studierenden belegt werden (Punkte: $r=.310$, $p=.001$, Noten: $r=.205$, $p=.032$), wobei mit den Punkten 9,61% der Varianz aufgeklärt wurden.

Evaluationsergebnisse quantitativ

Die Ergebnisse der Evaluation sind in Tabelle 4 dargestellt. Es wurde eine bivariate Korrelationsanalyse zwischen den Skalen durchgeführt. Es zeigten sich signifikante Korrelationen zwischen Schulnoten und Skalen der jeweiligen OSCE-Station (I-VI), weiterhin eine signifikante Korrelation zwischen Skala VIII („OSCE-Parcour im Ganzen“), Schulnoten und Skalen I-VI sowie nega-

tive, signifikante Korrelationen zwischen Skala VII („Selbstzweifel“) und jeweils den Skalen I-V.

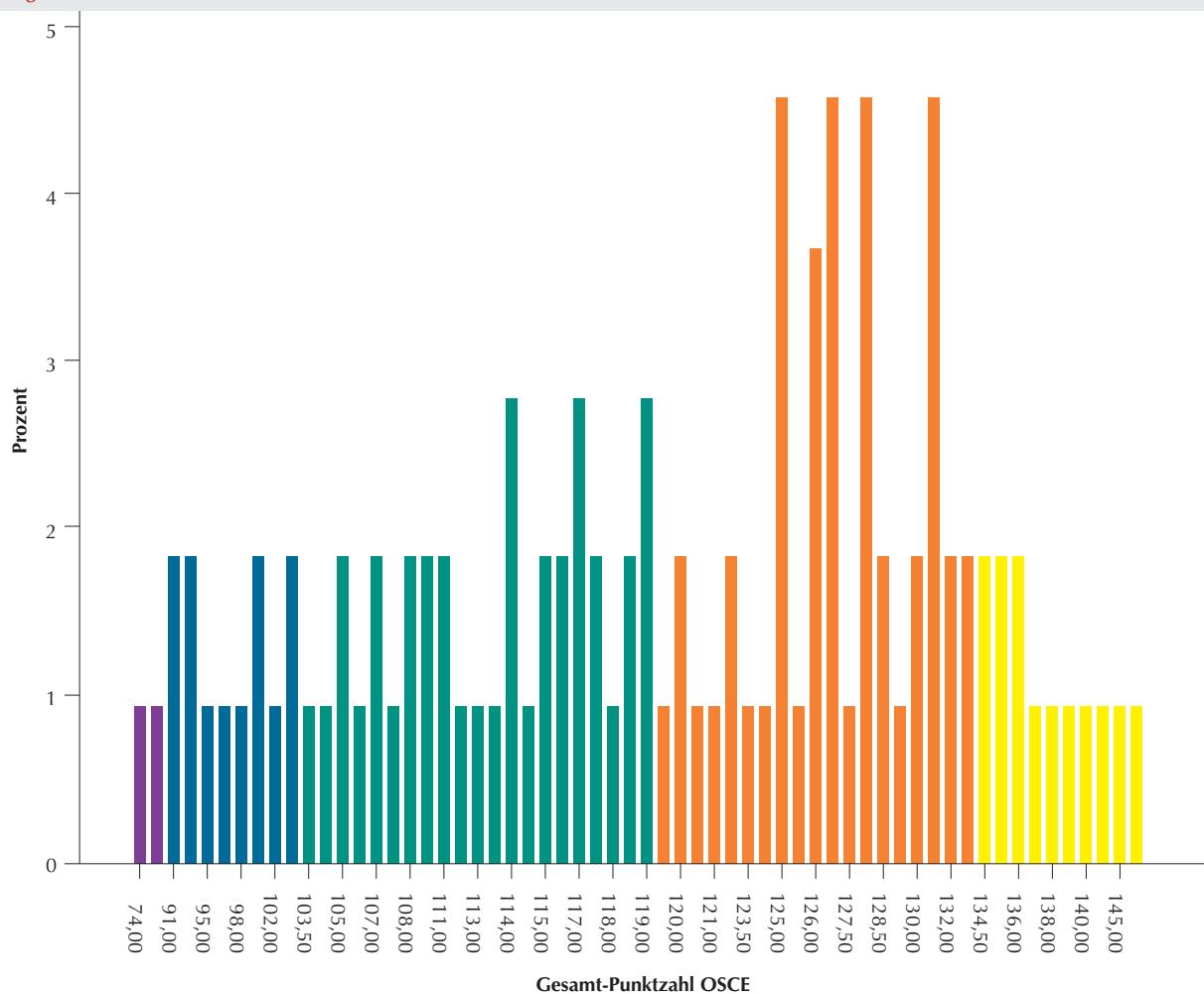
Evaluationsergebnisse qualitativ

Die Ergebnisse der Freitextantworten aus den Einzelstationen und der Gesamtbeurteilung von Prüfung und Kurs (80 Einzelnennungen) konnten im wesentlichen in den drei Hauptkategorien 1) „Prüfungsvorbereitung“, 2) „Prüfungsformat“ und 3) „Prüfungsorganisation und -durchführung“ zusammengeführt werden. Die Ergebnisse nach den vier am häufigsten genannten Wünschen sind in Abbildung 4 dargestellt. Hinsichtlich der Prüfungsvorbereitung wünschten sich 40% der Studierenden in der Gesamtbeurteilung mehr praktische Vorbereitungsmöglichkeiten, in den Freitextantworten der Einzelstationen machte dieser Wunsch nahezu durchgehend 90% der Verbesserungsvorschläge aus. Das Prüfungsformat wurde überwiegend positiv beurteilt, verbunden mit dem Wunsch, dieses Prüfungsformat auszubauen und

dafür die MC-Klausur wegzulassen. Zur Prüfungsorganisation und -durchführung wurde durch die Studierenden die gute Leistung der Schauspielpatienten mehrfach ausdrücklich hervorgehoben, zwei Nennungen wünschten sich den Einsatz von Schauspielpatienten im Unterricht. Kritisch angemerkt wurde von mehreren Studierenden, dass sie 45 Minuten vor Prüfungsbeginn zur Anmeldung erscheinen mussten. Die Organisation der Prüfung wurde überwiegend positiv bewertet, allerdings wünschten sich einige Studierende mehr Zeit für ein Feedback nach jeder Prüfungsstation. Die Relevanz des Querschnittsbereichs Notfallmedizin für die ärztliche Tätigkeit wurde durch einige Studierende ausdrücklich hervorgehoben.

Diskussion

In der neuen ÄAppO von 2002 wurde der Anspruch formuliert, Prüfungen durchzuführen, die die erforderlichen Fertigkeiten der Ärztinnen und Ärzte

Abbildung 3

Darstellung der OSCE-Prüfungsergebnisse nach erreichten Gesamtpunktzahlen und Noten
(Schulnoten: lila="5" (1,8%), blau="4" (11,9%), grün="3" (32,1%), orange="2" (40,4%), gelb="1" (13,8%); maximal erreichbare Punktzahl: 150 Pkt., n_{Prüflinge} = 109).

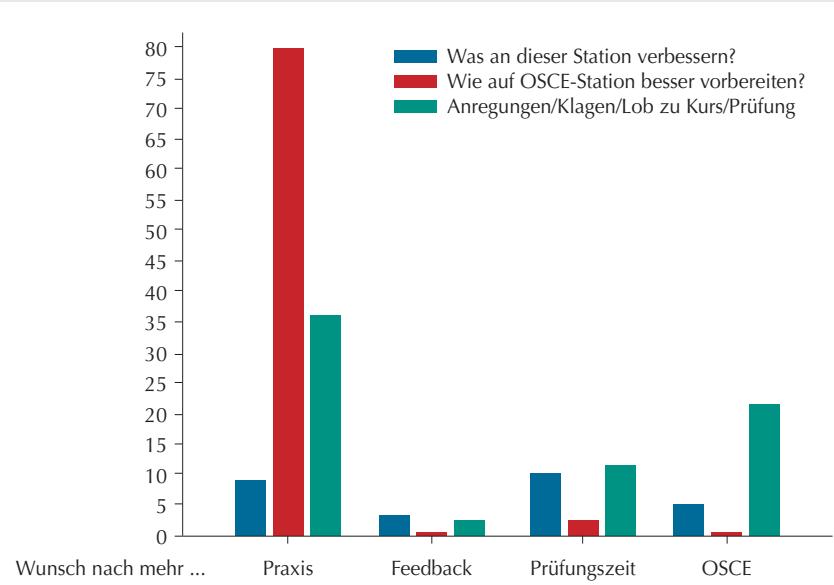
Tabelle 4

Ergebnisse der OSCE-Evaluation: Die (Einzel-)Items der acht Frageblöcke (Skalen) wurden inhaltlich beschreibend zusammengefasst (Addition der Items und anschließende Division durch Anzahl der Items) und den zusätzlich für die jeweilige Station vergebenen Schulnoten gegenübergestellt

(Skalen: 1=trifft nicht zu, 2=trifft eher nicht zu, 3=teils/teils, 4=trifft eher zu, 5=trifft voll zu),

(Schulnoten: 1=sehr gut, 2=gut, 3=befriedigend, 4=ausreichend, 5=mangelhaft). Skala 7 und 8 wurden nicht benotet.

	Anzahl der Items/Frageblock	Median	Interquartilsabstand (IQR)	Schulnoten (Median/IQR)
Skala 1 – „BLS-Station ist gut“	6	3,83	1,16	2/2
Skala 2 – „ACLS-Station ist gut“	6	4,17	1	2/1
Skala 3 – „Trauma-Station ist gut“	6	3,67	1,33	2/1
Skala 4 – „Kindernotfall-Station ist gut“	6	3,92	1,12	2/2
Skala 5 – „Herzinfarkt-Station ist gut“	6	4	1,34	2/1
Skala 6 – „Atemwegs-Station ist gut“	6	4	1,59	2/1
Skala 7 – „Ich habe Selbstzweifel“	8	1,75	0,72	
Skala 8 – „OSCE gutes Prüfungsformat“	8	4,13	0,75	

Abbildung 4

Ergebnisse der qualitativen Evaluation. Dargestellt sind die vier dominanten Wünsche farblich nach den drei Freitext-Fragen (siehe Text) und nach 1) mehr praktischer Vorbereitung (Praxis), 2) mehr Feedback in der Prüfung (Feedback), 3) mehr Zeit für die Prüfungsaufgaben (Prüfungszeit) und 4) mehr OSCE als Prüfung (OSCE).

besser abbilden. Insbesondere die Handlungskompetenz stellt dabei einen Bereich dar, der in „klassischen“ schriftlichen und mündlichen Prüfungen kaum überprüfbar ist. Dies führte fachübergreifend zu Bemühungen, adäquate Prüfungsformate, wie beispielsweise das OSCE, zu etablieren [9,10,11,12]. Nach Fischer et al. „sind alle curricularen Reformen und Innovationen wirkungslos, wenn das Curriculum nicht in einer qualitativ hochwertigen Prüfungskonzeption abgebildet wird“ [13]. Dies setzt zunächst eine lernzielbasierte Prüfungsplanung anhand eines „Blueprint“ voraus, durch den der Abgleich zwischen Prüfungs-inhalten (Aufgaben) und Lernzielen des Curriculums möglich ist. Dieses Vorgehen wurde in der Testkonstruktion des OSCE-Parcours berücksichtigt und im Rahmen eines ausführlichen Review-prozesses überprüft.

Die Ergebnisse unserer Untersuchung zeigten hinsichtlich der Zuverlässigkeit (Reliabilität) der durchgeführten OSCE-Prüfung mit sechs Prüfungsstationen eine mittlere interne Konsistenz (Cronbach's α) von 0,672. Dies ist insofern nicht

überraschend, da die Verwendung von mehreren standardisierten klinischen Fällen als Voraussetzung für eine reliable Prüfung beschrieben wurde [16,17]. Nach der Formel von Spearman-Brown kann die notwendige Anzahl von Stationen bei gegebenem Cronbach's α kalkuliert werden, die notwendig ist, um eine Reliabilität von 0,8, wie für medizinische Prüfungen empfohlen [15], zu erreichen. Hier zeigte sich für unsere Daten die Notwendigkeit eines 12-Stationen-OSCE, was durchaus realistisch umsetzbar ist und mittelfristig unter Einbeziehung der nicht-anästhesiologischen Kliniken geplant ist. Die Prüfungsoptimierung sollte auch durch Elimination von Aufgaben erfolgen, die die Reliabilität der Gesamtprüfung erniedrigen. Aufgabe 6 der hier vorgestellten Notfall-OSCE müsste daher in Folgeprüfungen verbessert werden, da Cronbach's α der Gesamtprüfung ohne diese Aufgabe größer ist (Cronbach's α - „wenn item gelöscht“) (Tab. 3). Hinsichtlich der Aufgabenschwierigkeit sind die Aufgaben dieser OSCE-Prüfung als eher leicht zu bewerten. Die korrigierten

Trennschärfen r' sind als gut, bei der Station 6 „Atemwegsmanagement“ als akzeptabel zu bezeichnen.

Die Korrelationen zwischen OSCE und MC-Klausur sind hinsichtlich der Beurteilung der Prüfungsqualität relevant, denn zeigten sich dort hohe Korrelationen mit großen „Schnittmengen“ (hohe aufgeklärte Varianz) wäre durchaus kritisch anzumerken, ob die OSCE-Prüfung denn wirklich notwendig ist, wenn ein guter MC-Kandidat auch ein guter OSCE-Kandidat wäre. Andererseits könnten hohe Korrelationen auch ein Hinweis für eine mangelhafte OSCE-Testkonstruktion sein, da ja erwartungsgemäß andere (praktische) Kompetenzen überprüft werden sollen als in der MC-Klausur. OSCE- und MC-Ergebnisse korrelieren hier signifikant, wobei die mäßige Ausprägung und die Aufklärung von lediglich 9,61% der Varianz schlüssig damit zu erklären ist, dass im MC-Test zum einen 50% der Prüfungsfragen ($n=12$) von nicht OSCE-relevanten Modulen stammen, zum anderen im OSCE neben kognitiven Aspekten auch psychomotorische Kompetenz (Praxiskompetenz) überprüft wird. Die geringe Varianz entspräche somit der geringen Schnittmenge zwischen OSCE- und MC-Prüfung.

In der Evaluation wurden signifikante Korrelationen zwischen Schulnoten und Summenskalen der jeweiligen OSCE-Station offensichtlich und können als Hinweis für die Zuverlässigkeit (Reliabilität) der Evaluation gewertet werden. Die signifikante Korrelation zwischen Skala VIII („OSCE-Parcour im Ganzen“), Schulnoten und Skalen I-VI ist dahingehend zu interpretieren, dass die Studierenden, die die einzelnen OSCE-Stationen positiv bewerteten auch den Prüfungsparcour positiv beurteilten. Dies geht einher mit negativen, signifikanten Korrelationen zwischen Skala VII („Selbstzweifel“) und jeweils den Skalen I-V als Ausdruck des Zusammenspiels von Prüfungsleistung und Selbstsicherheit („Je besser die Stationen, desto selbstsicherer bin ich“). Allerdings konnte die tatsächliche Prüfungsleistung nicht mit der Skala „Selbstzweifel“ korreliert werden, da Prüfungen und Evaluationen nicht verknüpft waren. Dies sollte in Folgeuntersuchungen

unbedingt erfolgen. Generell wurde das praxisorientierte Prüfungsformat OSCE überwiegend positiv evaluiert mit dem Wunsch, diese Prüfungsform häufiger beziehungsweise anstelle von MC-Klausuren einzusetzen. Die bereits erfolgte Änderung des Lernverhaltens seitens der Studierenden durch das neue Prüfungsformat wurde dahingehend deutlich, dass eine hohe Anzahl von Studierenden mehr praktische Übungsmöglichkeiten einforderten.

Die Umsetzung eines OSCE-Formates muss allerdings auch hinsichtlich finanzieller und personeller Ressourcen betrachtet werden. So müssen Schauspielpatienten und studentische Hilfskräfte für die Prüfungsdurchführung rekrutiert werden. Insbesondere bei Prüfungsdurchführung unter der Woche stellt sich die Frage, wie eine ausreichende Anzahl von Prüfern aus der Patientenversorgung freigestellt werden können, ohne den OP-Betrieb zu beeinträchtigen. Nahezu alle Prüfer beurteilten dieses für sie neue Prüfungsformat informell sehr positiv und waren mit der Kursorganisation und -durchführung zufrieden. Eine ökonomische Gesamtevaluation von OSCE-Prüfungen ist erforderlich, um die wahren Kosten im Vergleich zu anderen Prüfungsformen beurteilen zu können. Generell könnte der kombinierte Einsatz verschiedener Prüfungsformen möglicherweise ressourcenschonend eine valide und reliable Prüfung ermöglichen. Auch mögliche Unterschiede zwischen Prüfungsleistungen in Abhängigkeit vom Zeitpunkt der Prüfung sollten in Folgestudien erfasst werden, da Prüfungen allen Prüflingen gleiche Chancen gewährleisten sollen (Test-Fairness).

Schlussfolgerung

Diese Untersuchung zeigt, dass bei lernziel- und Blueprint-basierter Aufgaben-gestaltung und Einhaltung eines struktu-rierten Reviewprozesses befriedigende Ergebnisse hinsichtlich methodischer Kriterien auch mit relativ wenigen Sta-tionen erreicht werden können. Die Ergebnisse der Korrelationen von MC-Klausur und OSCE unterstreicht dabei die Hypothese, dass diese Prüfungsfor-

mate jeweils unterschiedliche Kompe-tenzbereiche abbilden. Das Prüfungs-format OSCE wurde dabei sowohl durch die Prüfer als auch die Studierenden po-sitiv beurteilt. Bereits unmittelbare Aus-wirkungen der OSCE-Prüfung auf das Lernverhalten der Studierenden wurden in dem Wunsch nach mehr praktischen Übungsmöglichkeiten offensichtlich.

Literatur

- Beckers SK, Timmermann A, Müller MP, Angstwurm M, Walcher F. Undergraduate medical education in emergency medical care: A nationwide survey at German medical schools. *BMC Emerg Med* 2009;9:7.
- Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357:945-949.
- Miller GE. The assessment of clinical skills/ competence/performance. *Acad Med* 1990;9 Suppl:S63-7.
- Harden RMCG, Stevenson M, Wilson Downie W, Wilson GM. Assessment of Clinical Competence using Objective Structured Examination. *Brit Med J* 1975; 1:447-451.
- Nikendei C, Jünger J. OSCE-praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung. *GMS Z Med Ausbild* 2006;23(3):Doc 47.
- McIlroy JH, Hodges B, McNaughton N, Regehr G. The Effect of candidates' Perception of the Evaluation Method on Reliability of Checklist and Global Rating Scores in an Objective Structured Clinical Examination. *Acad Med* 2002; 77:725-728.
- Rosenberg, M. Society and the adolescent self-image. Princeton, NJ: Princeton University Press; 1965.
- Fend H, Berger F, Grob U. 1527 „Lebensgeschichten“ von der späten Kindheit ins Erwachsenenalter. Konzept und Durchführung der LifE-Studie. In: Fend H, Berger F, Grob U (Hrsg.). Lebensverläufe, Lebensbewältigung, Lebensglück. Ergebnisse der LifE-Studie. Wiesbaden: VS; 2009:9-34.
- Frobenius W, Bingold F, Beckmann MW. Die neue Ärztliche Approbationsordnung (ÄAppO)-Teil II: Perspektiven für praxis-nahe Unterrichts- und Prüfungsmethoden in Gynäkologie und Geburtshilfe. *Geburtsh Frauenheilk* 2007;67:959-965.
- Heckmann JG, Rauch C, Düttsch M, Lang C, Weih M, Schwab S. Erlanger Neurologisches Examen-Strukturiert (ERNEST). *Nervenarzt* 2008;79:676-683.
- Jäger J, Köllner V, Graf N. Der OSCE (Objective Structured Clinical Examination) Allgemeinmedizin: eine kompetente praktische Prüfung nach dem Blockpraktikum. *Z Allg Med* 2008;84:375-381.
- Timmermann A, Roessler M, Barwing J, Blaschke S, Bräuer A, Eich C, Hirn A, Klockgether-Radke A, Nickel E, Russo S, Kettler D, Saur P. Neue Wege der studentischen Lehre - Erste Erfahrungen im Querschnittsbereich Notfall- und Intensivmedizin. *Anästhesiol Intensivmed Notfallmed Schmerzther* 2005;40:536-543.
- Fischer MR. Gesellschaft für Medi-zinische Ausbildung, GMA-Ausschuss Prüfungen & Kompetenzzentrum Prüfungen Baden-Württemberg: Leitlinie für Fakultäts-interne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier. *GMS Z Med Ausbild* 2008;25(1):Doc 74.
- Bortz J, Döring N. Hypothesengewin-nung und Theoriebildung. In: Bortz J, Döring N (Hrsg). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 3. Aufl. Berlin Heidelberg New York Tokio: Springer; 2006:355-396.
- Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004;38:1006-12.
- Newble D. Techniques for measuring clinical competence: objective structur-ed clinical examinations. *Med Educ* 2004;38:199-203.
- Van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
- Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild* 2006;23(3):Doc53.

Korrespondenz-adresse

**Dr. med.
Peter Iblher, MME**



Klinik für Anästhesiologie
Universität zu Lübeck
Ratzeburger Allee 160
23538 Lübeck, Deutschland
Tel.: 0451 5002766
Fax: 0451 5003405
E-Mail: peter.iblher@uk-sh.de