

DAS OBJEKTIVE STRUKTURIERTE KLINISCHE EXAMEN (OSKE) IN DER CHIRURGIE

CHRISTOPH BECKMANN, YNGVE FALCK-YTTER, DR. PAED INGRID STIEGLER,
PROF REINER LABITZKE, PROF. FLORIAN EITEL, PROF. JOHANNES BIRCHER

Zusammenfassung

Die Ausbildung von Studierenden der Medizin in klinisch-praktischen Fertigkeiten weist in Deutschland erhebliche Mängel auf, weil unter anderem die Prüfungen des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) nur aus Multiple-Choice-Fragen (MC) bestehen und die klinisch-praktischen Prüfungen mangels Standardisierung wenig objektiv bzw. zuverlässig sind. Wir haben deshalb in der Chirurgie ein objektives klinisches Examen (OSKE) eingeführt. 16 Studierende wurden gleichzeitig im 5-Minutentakt an 20 Stationen geprüft, welche entweder eine Anamnese,

eine körperliche Untersuchung oder Kurzantworten zum Inhalt hatten. Das Examen machte den Studierenden Freude. Die Beurteilung sowohl der Leistung der Studierenden wie auch des vorangegangenen Unterrichts war unter den von uns organisierten Gegebenheiten durchführbar. Mit Hilfe dieser bei uns praktizierten Prüfungsform erhalten Studierende Hinweise zur Verbesserung ihres Lernstils und Lehrende zur Änderung ihres Unterrichts. Die Resultate dokumentieren die Nützlichkeit von OSKE zeigen aber auch, daß wenigstens am Anfang ziemlich viel Aufwand erforderlich ist.

Einleitung

Die Vermittlung und Beherrschung klinisch-praktischer Fertigkeiten wird in der medizinischen Ausbildung als besonders wichtig angesehen [1]. Eine Evaluation dieser Fertigkeiten ist mit den bei uns etablierten MC-Prüfungen nicht möglich. Die gängigen mündlichen Prüfungen sind zu diesem Zweck auch nur eingeschränkt aussagekräftig. Die in den siebziger Jahren von R. Harden und F. Gleeson [2] entwickelte Prüfungsform des *Objektiven Strukturierten Klinischen Examins* (OSKE) bietet hierzu eine Alternative. Diese Prüfungsform ist in einem hohen Ausmaß valide, zuverlässig (reliabel) [3] und einigermaßen praktikabel. International wird dieses Verfahren überwiegend sowohl von Hochschullehrern als auch von Studierenden als eine sehr geeignete Prüfungsform zur Evaluation klinisch-praktischer Fertigkeiten beurteilt. Sie wird im chirurgischen Fachbereich aber selten praktiziert. Da die Chirurgie ein ausgesprochen hohes Maß klinisch-praktischer Fertigkeiten verlangt, haben wir mit der Frage, wie diese Prüfungsform dort angewandt werden könnte, ein OSKE entwickelt. Dabei kam es uns darauf an, in dieser Prüfung häufige und alltagsrelevante Problemstellungen aus dem chirurgischen Arbeitsbereich zu thematisieren.

Methode

Die Prüfung wurde von 16 Studierenden des 5. Semesters zum Abschluß ihres ersten vierwöchigen klinischen Blockpraktikums [4] (vier Tage pro Woche ganztägig in einer chirurgischen Klinik) durchlaufen. Es wurden 20 praktische Aufgabenstellungen konzipiert, um eine einigermaßen gute Erfassung der chirurgischen Arbeitsbereiche zu erreichen. Die Auswahl der Themen erfolgte in Zusammenarbeit mit unseren vier auszubildenden chirurgischen Abteilungen (PD Dr. R. Dietz, Witten, Prof G.H. Engelhardt, Wuppertal; Dr. U. Garde, Iserlohn, Prof Labitzke, Schwerte) und mit Prof F. Eitel (LMU München). Diese wurden von den Studenten in einer Rotation durchlaufen. Jede Aufgabe fand in einem eigenen Raum statt. Es gab drei Arten von Aufgaben: Anamnese (5 Aufgaben), körperliche

Untersuchung (7 Aufgaben), Kurzantwort (8 Aufgaben). Die Inhalte sind als Stichworte der Abb. 1 zu entnehmen. Ein Beispiel für eine Aufgabenstellung ist in Tabelle 1 gegeben. Die Aufgaben wurden so konzipiert, daß zu ihrer Lösung ein Zeitraum von fünf Minuten ausreichend war.

Tabelle 1: Beispiel einer Aufgabenstellung

OSKE-Station 3

Diese junge Frau hat gestern einen Unterschenkelgips erhalten.

Aufgabe:

- Führen Sie eine Gipskontrolle durch.
- Erläutern Sie, worauf Sie achten.
- Bei dieser Station müssen Sie die Untersuchung mit einer fokussierten Anamnese verbinden.

Die Prüfung war so ausgelegt, daß alle Studierende gleichzeitig an irgendeiner Station anfangen und im Takt weiterrotieren konnten. Zwischen den Aufgaben wurden zwei Minuten zum Übergang zur nächsten Aufgabe gewährt. Die Zeitintervalle wurden durch einen Gongschlag angezeigt und genau eingehalten, um damit einen reibungslosen Ablauf und Chancengleichheit zu gewährleisten. Die totale Prüfungszeit betrug somit zwei Stunden und zwanzig Minuten. Während der Übergangszeiten zwischen den Aufgaben konnte der Studierende die Aufgabenstellung der nächsten Prüfungsstation, welche an der Tür außen angebracht war lesen (siehe Beispiel einer Aufgabenstellung) und schon einmal seine Gedanken zu dieser Aufgabe sammeln. Beim Gongschlag wurde das Zimmer betreten. Im Zimmer traf der Studierende auf einen Simulanten, der einen Patienten darzustellen hatte, und einen Prüfer. Jeder Simulant hatte einen definierten Symptomenkomplex vor der Prüfung einstudiert. Er war instruiert,

bei Anamneseaufgaben nur auf gezielte Fragen seitens der Studierenden Auskünfte zu erteilen, bzw. bei Untersuchungsstationen nur gezielten Instruktionen der Studierenden Folge zu leisten. Der Prüfer hatte auf einem strukturierten Protokoll im Sinne einer Checkliste, welche die zur Fragestellung relevanten Aspekte enthielt, die Leistungen der Studierenden zu dokumentieren. Die Folge ihrer Aktionen wurde mit Zahlen codiert, um die Systematik der Vorgehensweise der Studierenden bei der Auswertung nachvollziehen zu können. Der Prüfer sollte möglichst in keiner Weise in den Prüfungsablauf eingreifen, z.B. durch didaktische Fragen, um Ungleichheiten für die Studierenden zu vermeiden. Sollte dies doch einmal geschehen sein, dann mußte dies auch vom Prüfer mit dokumentiert werden. Im Falle einer Anamneseerhebung der Studierende die relevante Anamnese, bei einer Untersuchungsstation führte er die körperliche Untersuchung durch und erläuterte dabei jeden seiner Schritte. Zusätzliche Leistungen wurden auch notiert, gingen aber nicht in die Auswertung mit ein. Sie können, falls erwünscht, zu einem persönlichen Feedback genutzt werden. Außerdem gab der Prüfer Urteile über die Interaktion zwischen Studenten und Simulanten und die Güte und Systematik des Studierenden ab, welche zur Korrelation mit der objektiven Leistung des Studierenden herangezogen werden konnten, aber nicht mit in die Auswertung eingingen. Bei den Kurzantwortstationen wurden klinisch relevante Aufgaben gestellt, wie z.B. die Beschreibung und Beurteilung eines Röntgenbildes. Diese wurden im Anschluß an die Prüfung ebenfalls mit Hilfe einer Checkliste ausgewertet.

Vor der Prüfung beurteilten die verantwortlichen Hochschullehrer die Aufgaben hinsichtlich ihrer Relevanz. Die Prüfer und Simulanten trafen sich direkt im Anschluß an die Prüfung und tauschten ihre Erfahrungen aus und dokumentierten ihre Verbesserungsvorschläge und aufgetauchte Probleme auf vorbereiteten Evaluationsbögen. Nachdem alle Prüfungsstationen von allen Studierenden durchlaufen worden waren, wurden die Stationen mit körperlicher Untersuchung von Prüfern und Simulanten einmal für alle Studierenden in der richtigen Weise demonstriert, da erfahrungsgemäß die Lernmotivation direkt im Anschluß an eine solche Prüfung am höchsten ist. Die Prüfung selbst wurde von den Studierenden bezüglich folgender Fragen auf einer 9-Punkteskala beurteilt.

1. Die Prüfung hat Spaß gemacht.
 2. Die Prüfung kann meine Leistung gut differenzieren.
 3. Die Prüfung war fair.
 4. Fünf Minuten pro Station war viel zu wenig und hat mich unter unangenehmen Druck gesetzt.
 5. Zusammenfassende Beurteilung der Prüfung.
- Zusätzlich konnten sie frei zur Prüfung Stellung nehmen.

Die Auswertung erfolgte anhand der Checklisten. Jeder Aspekt der Checkliste hatte einen zuvor festgelegten Wert, in den meisten Fällen 1 Punkt. Die erreichten Punkte wurden durch die insgesamt erreichbaren Punkte für diese Aufgabe geteilt und das Ergebnis als Prozentwert notiert. Dieser gibt Auskunft über den

Anteil des maximal erreichbaren Limits und kann durchaus zum Beispiel mit 60% ein für diesen Ausbildungsstand sehr gutes Ergebnis widerspiegeln. Aus den Ergebnissen aller Prüflinge wurde der Mittelwert errechnet und dieser willkürlich gleich 500 Punkte und die Standardabweichung gleich 100 gesetzt. Somit wurde der Leistungsstand jedes Studierenden im Verhältnis zu seiner Prüfungsgruppe ausgedrückt. Vier Tage nach der Prüfung fand mit den Studierenden eine Nachbesprechung der Prüfung und die Ergebnismittelung statt.

Resultate:

Von den 12 eingegangenen Fragebögen beurteilten zehn Studierenden die gestellten Fragen positiv. Die Prüfung hat den meisten Studierenden Spaß gemacht. Zwei Studierende bemängelten, daß die Interaktion zwischen Studenten und Simulanten durch den auf 5 Minuten begrenzten Zeitraum als sehr künstlich erlebt wurde und auch, daß die Anamnesen sehr direktiv gehalten werden mußten. Allerdings hatten die meisten Studierenden Verständnis dafür, weil die Prüfung dann entweder wesentlich länger dauern müßte, oder nur ein sehr eingegrenztes Gebiet abgefragt werden könnte. Ein mündliches Feedback direkt im Anschluß an die Prüfung ergab, daß die Prüfung von vielen Prüfern als für den Ausbildungsstand des 5. Semesters zu schwer eingeschätzt wurde. Weiterhin wurde von einigen Prüfern festgehalten, daß der Aufwand für eine solche Prüfung mit 24 Prüfern und 5 Studenten höherer Semester als Simulanten relativ hoch sei.

Die Ergebnisse sind in Abbildung 1 zusammengefaßt. Sie zeigt, daß gewisse Aufgaben relativ gut (z.B. Untersuchung bei Verdacht auf Appendizitis) und andere schlecht gelöst worden sind (z.B. Untersuchung bei venöser Insuffizienz, Röntgenbild einer distalen Radiusfraktur). 13 Aufgaben lagen zwischen 40 und 60%, d.h. in einem mittleren Bereich. Eine Aufteilung der Ergebnisse in die Untergruppen ergab, daß bei Anamnese- und Untersuchungsaufgaben im Mittel je 51% und Kurzantwortaufgaben im Mittel 43% des Limits erreicht wurden. Die Verteilung des Leistungsniveaus innerhalb der Prüfungsgruppe ist auf Abbildung 2 dargestellt. Es zeigt sich eine gleichmäßige Verteilung des Leistungsspektrums.

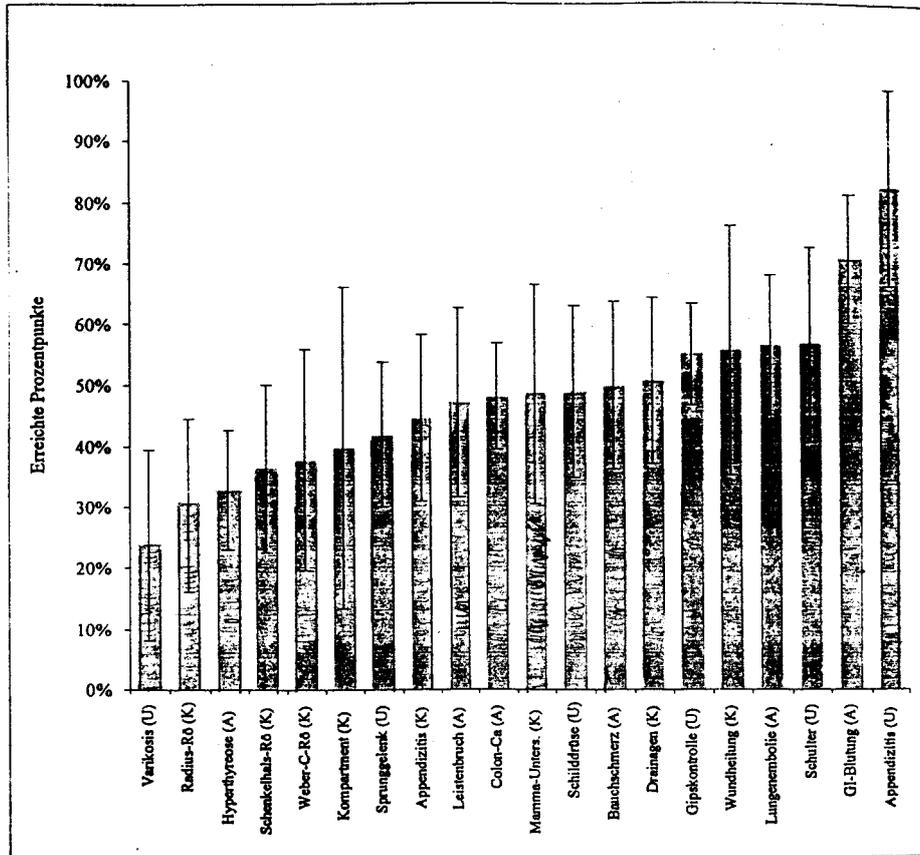


Abbildung 1: Mittelwerte und Standardabweichungen der Ergebnisse der einzelnen Aufgaben. (A) bedeutet Anamnese, (U) körperliche Untersuchung und (K) Kurzantwortstation. Wie oben beschrieben fallen einige Stationen mit hohen bzw. niedrigen Mittelwerten auf, was bedeutet, daß sie das Leistungsniveau nicht gut differenzieren. Die Station (U) Appendizitis ist mit 82% zu leicht. Die Stationen (U) Varikosis, (K) Radiusfraktur-Röntgenbild, (A) Hyperthyreose, (K) Schenkelhalsfraktur-Röntgenbild, (K) Weber C Fraktur-Röntgenbild lagen im Mittelwert unter 40% und waren für diesen Ausbildungsstand zu schwer.

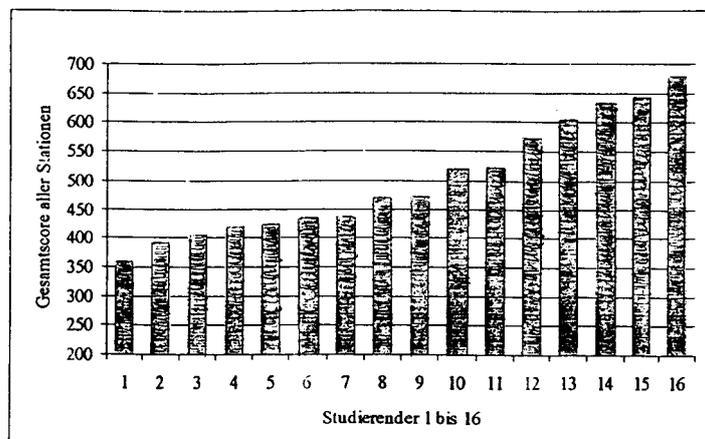


Abbildung 2 zeigt vom unteren bis zum oberen Bereich eine gleichmäßige Verteilung des Leistungsspektrums der 16 Studierenden.

Diskussion:

Ein selbstverantwortetes Studium, wie es an unserer Universität von den Studierenden erwartet wird, erfordert periodische Rückmeldungen des Lernerfolgs. In diesem Sinne verstehen wir auch dieses OSKE. Das Ergebnis dieser Prüfung hat keinen Einfluß auf Scheinvergabe oder das Weiterstudieren. Das OSKE ermöglicht den Studierenden eine fundierte Standortbestimmung über ihre klinisch-praktischen Fähigkeiten [5] und damit Änderungen ihres Lernstils noch während der Ausbildungszeit. Wenn solchermaßen das Grunderlebnis der Studierenden in dieser Prüfung nicht von Leistungsdruck geprägt ist, sondern vornehmlich dazu dient ein Bewußtsein über seine persönlichen Stärken und Schwächen zu bekommen, dann ist damit viel gewonnen und es verwundert nicht, daß die Prüfung den meisten Studierenden Spaß gemacht hat. Damit ist die Zielsetzung der Prüfung für die Studierenden weniger ihr Endergebnis, als vielmehr eine Lernhilfe. Für die Ausbilder eröffnet das OSKE die Möglichkeit, zu erfassen, welchen Teil ihrer Ausbildung bei den Studierenden angekommen ist. Sie können ihren Unterricht entsprechend ändern und erfolgreicher gestalten.

Die willkürliche Festsetzung des Mittelwertes auf 500 Punkte und der Standardabweichung auf 100 Punkte wurde aus der Prüfungspraxis anderer Reformuniversitäten übernommen, die seit einigen Jahren regelmäßig mit dieser Methode arbeiten. Ein Ergebnis von 350 Punkten wird als untere Grenze der erwartbaren Leistung angenommen. Der Mittelwert ist somit prüfungsgruppenabhängig, was im Extremfall bedeuten kann, daß bei einer schweren Aufgabe, welche von niemandem gelöst wird das Ergebnis bei 0% einem Punktwert von 500 entspricht, Gleichermaßen ergibt eine Aufgabe, welche von allen Prüfungsteilnehmern vollständig gelöst wird, einen Mittelwert von 500 Punkten. Da uns wichtig ist, das Leistungsspektrum möglichst gut zu differenzieren, wird versucht die Aufgabenstellungen entsprechend dieser Maßgabe auszuwählen. In diesem Fall entsprachen 13 Aufgaben dem Kriterium einer guten Differenzierung. Die zu leichten Aufgaben körperliche Untersuchung bei Verdacht auf Appendizitis und Anamnese zur akuten oberen gastrointestinalen Blutung wären demnach für diese Prüfungsgruppe überflüssig gewesen. Die zu schweren Aufgaben wären entweder zu einem späteren Ausbildungszeitpunkt angebracht oder aber, wenn bei den ausbildenden Chirurgen Einigkeit darüber besteht, daß es sich um eine wichtige zu vermittelnde Fähigkeit handelt, muß während der praktischen Ausbildung stärker auf das entsprechende Thema eingegangen werden und die Aufgabe behält ihren Platz in der Prüfung.

Die Prüfung wurde von einigen Prüfern als für das 5. Semester zu schwer eingeschätzt. Diese Einschätzung beruht auf dem möglichen Mißverständnis, daß die Vorgaben auf den Checklisten im Sinne eines zu erreichenden Maßstabes interpretiert wurden. Dieser Maßstab kann aber nur durch eine Vielzahl solcher Prüfungen empirisch gefunden werden und konnte also für eine erste Prüfung dieser Art noch nicht zur Anwen-

dung kommen. Empirisch sind Aufgaben, die im Bereich von 40-70% korrekt gelöst werden, das angestrebte Ziel. Die im Mittel unter 40% gelösten Aufgaben sind in der Tat zu schwer, was zugleich bedeuten kann, daß diese Inhalte den Studierenden bisher noch nicht vermittelt wurden. Es gehörten z.B. alle Röntgenbildaufgaben dazu, bei denen die meisten zwar eine Diagnose zu stellen in der Lage waren, aber häufig andere wesentliche Aspekte nicht erwähnt wurden, was auf eine fehlende Systematik bei der Beurteilung eines Röntgenbildes hinweist. Außerdem differenzieren bei schlechten, wie auch bei sehr guten Ergebnissen die Aufgaben nicht mehr nach dem Leistungsniveau innerhalb der Prüfungsgruppe. Insofern wäre es natürlich denkbar, möglichst nur Aufgaben zu verwenden, bei denen ein mittleres Ergebnis erwartet werden kann. Die meisten Studierenden, welche schon Erfahrungen mit dieser Prüfungsform hatten, fanden die Prüfung nicht zu schwer.

Die Leistungsbeurteilung des gleichen OSKE bei unterschiedlichen Prüfungsgruppen ist möglich und kann im Prinzip bei ganz unterschiedlichem Niveau zu gleichen Scores führen. Ein Ergebnisvergleich ist aber mit Hilfe der erreichten Prozentwerte möglich und wird in Zusammenarbeit mit der Chirurgischen Klinik Innenstadt der LMU München angestrebt.

Neben der Bewertung der Leistung jedes einzelnen Studierenden erlaubt das OSKE auch eine Evaluation der vorausgegangenen Ausbildung. Unsere Chirurgen haben offenbar in ihren Seminaren die Appendizitis und die gastrointestinale Blutung ausführlich besprochen und z.B. die Varikosis vernachlässigt. Das werden sie angesichts der Prüfungsergebnisse in Zukunft korrigieren können.

Die Reliabilität der Prüfung ist insbesondere von den Variablen Prüfer und Simulant/Patient abhängig. Der Prüfer, welcher die Checkliste fährt, sollte diese möglichst genau kennen, um wirklich jede Leistung des Studierenden zuverlässig erfassen zu können. Der Simulant muß vor der Prüfung seine Rolle gut einstudiert haben, damit er sich gegenüber den verschiedenen Studierenden immer gleich verhalten kann. Auf diese Weise kann die intrinsische Variabilität der Prüfer minimiert, damit praktisch nur noch die Variabilität der Studierenden zum Tragen kommt. Die Möglichkeit zur weitgehenden Standardisierung dieser Variablen macht einen großen Vorzug des OSKE gegenüber herkömmlichen mündlichen und praktischen Prüfungen aus. Allerdings sind dies nicht die einzigen Einflußgrößen in dieser Prüfung. Sie ist auch abhängig von einer ausgesprochen übersichtlichen und straffen Organisation. Es werden viele Räume benötigt, die nah beieinander liegen sollten. Die Abfolge der Stationen muß für die Studierenden mühelos zu finden sein. Abendstunden sind deshalb besonders geeignet. Der Zeitablauf muß genauestens eingehalten werden. Wenn irgendein Aspekt nicht berücksichtigt wird, z.B. daß ein Prüfer die Zeit überzieht, dann entsteht Chaos. Dies ist eine der Varianzquellen dieses Prüfungssystems, die von den Organisatoren behandelt werden muß.

Der organisatorische und personelle Aufwand für diese Prüfung wurde als relativ hoch angesehen. Dies liegt zu einem guten Teil daran, daß diese in der Chirurgie jetzt zum ersten Mal durchgeführt wurde. Erfahrungen mit dieser Prüfungsform in der Inneren Medizin [6], wo wir schon mehrere OSKE durchgeführt haben, zeigen daß der in der Anfangsphase große organisatorische Aufwand sich von Mal zu Mal verringert, wenn die Infrastruktur steht d.h. ein genügend großer Pool an Aufgabenstellungen und Material angelegt ist. Dann ist es auch bei großen Studentenzahlen möglich, OSKE durchzuführen.

Ein interessanter Punkt ist die subjektive Beurteilung des Vorgehens der Studierenden durch die Prüfer. Gelegentlich wurde das Vorgehen eines Prüflings als unsystematisch erlebt, obwohl die Systematik, an der Checkliste nachvollziehbar, objektiv korrekt war. Hier zeigt sich das Problem des subjektiven Eindrucks des Prüfers, der unter Umständen täuschen kann z.B. wenn ein Student unsicher wirkt. Nur durch Planung einer Checkliste, entsprechende Dokumentation und Vergleich aller Studierenden untereinander, kann die tatsächliche Leistung der Studierenden korrekt bewertet werden. In dieser Beziehung ist das OSKE den üblichen mündlichen Prüfungen deutlich überlegen.

Literatur:

1. Irby, D.M.; F.G. Lippert III, D.C. Schaad: Psychomotor Skills for the General Professional Education of the Physician. *Teaching and Learning in Medicine* Vol. 3, No. 1 (1991) 2-5
2. Harden, R.M., F.A. Gleeson: Assessment of medical competence using an objective structured clinical examination (OSCE). *asme medical education booklet*, number 8, (1979)
3. Newble, D.1.- J. Hoare, R. G. Eimslic: The validity and reliability of a new examination of the clinical competence of medical students. *Medical Education* 15 (1981) 46-52
4. Falck-Ytter, Y. u.a. @ Kursbuch für die klinische Ausbildung an der medizinischen Fakultät der Universität Witten/Herdecke. Medizinische Fakultät. Vierte erweiterte Auflage. Oktober 1994
5. Newble, D.1.: Assessing clinical competence at the undergraduate level. *Medical Education* 26 (1992) 504-511
6. Falck-Ytter, Y.; 1. Stiegler-. Beurteilung klinischer Fähigkeiten mit dem OSCE-Verfahren. *Medizinische Ausbildung* 10/1 (1993) 48-55

Kontaktadresse:

Prof J. Bircher, Dekan
Medizinischen Fakultät der Universität Witten/Herdecke
Alfred-Herrhausen-Str. 50
D-58448 Witten