

Aus der Abteilung Allgemeinmedizin
(Prof. Dr. med. M. M. Kochen, MPH, FRCGP)
im Zentrum Innere Medizin
der Medizinischen Fakultät der Universität Göttingen

**Studentische Tutoren als Prüfer in einer
„objective structured clinical examination“ (OSCE):
Evaluation ihrer Bewertungsleistungen**

INAUGURAL - DISSERTATION
zur Erlangung des Doktorgrades
der Medizinischen Fakultät
der Georg-August-Universität zu Göttingen

vorgelegt von

Alexandra Koch

aus

Bremerhaven

Göttingen 2008

D e k a n: Prof. Dr. med. C. Frömmel

I. Berichterstatter: Prof. Dr. disc. pol. W. Himmel

II. Berichterstatter/in:

III. Berichterstatter/in:

Tag der mündlichen Prüfung:

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Abkürzungsverzeichnis	VI
1 Einleitung	1
1.1 Einführung der neuen Approbationsordnung	1
1.2 „Kursus für ärztliche Basisfertigkeiten“	2
1.3 Die OSCE (objective structured clinical examination)	4
1.3.1 Struktur der Prüfungsform „OSCE“	4
1.3.2 Gründe für die Einführung der OSCE	5
1.3.3 Entwicklung einer OSCE.....	6
1.3.4 Testgütekriterien der OSCE.....	7
1.3.5 Vor- und Nachteile der OSCE	8
1.3.6 Einfluss von ethnischer Herkunft und Geschlecht.....	10
2 Fragestellungen und Ziele	12
3 Material und Methoden	13
3.1 Pilotierung der OSCE	13
3.2 Die OSCE des Wintersemesters 2004/05	14
3.3 Studentische und ärztliche Prüfer	17
3.4 Simulationspatienten	18
3.5 OSCE-Teilnehmer	18
3.6 Evaluation	18
3.7 Statistische Auswertungen.....	19
3.7.1 Vergleich zwischen den Prüfern	19
3.7.2 Vergleich zwischen den Studierenden	20
4 Ergebnisse	21
4.1 Bewertungen der Tutoren und Ärzte im Vergleich	21
4.1.1 Vergleich der Checklistennoten.....	21
4.1.2 Vergleich der Globalbeurteilungen.....	22
4.1.3 Vergleich der Gesamtnoten	23
4.1.4 Vergleich der Checklisten-, Global- und Gesamtnoten	24
4.1.5 Einzelitemanalyse der vier Stationen.....	25

4.2	Deutsche und Ausländer im Vergleich	29
4.2.1	Vergleich der Checklistennoten	29
4.2.2	Vergleich der Globalbeurteilungen	30
4.2.3	Vergleich der Gesamtnoten	30
4.3	Männliche und weibliche Studierende im Vergleich	32
4.3.1	Vergleich der Checklistennoten	32
4.3.2	Vergleich der Globalbeurteilungen	33
4.3.3	Vergleich der Gesamtnoten	33
4.4	Evaluation der OSCE des WS 2004/05 aus studentischer Sicht	35
4.4.1	Akzeptanz der OSCE	35
4.4.2	Vergleich der Bewertungsweise von Tutoren und Ärzten	37
4.4.3	Einsatz von studentischen Prüfern	37
5	Diskussion	39
5.1	Beurteilung des Einsatzes von studentischen Tutoren als Prüfer	39
5.1.1	Zusammenfassung der Ergebnisse	39
5.1.2	Kritische Betrachtung der Ergebnisse	39
5.1.3	Vergleich mit bisherigen Publikationen	41
5.2	Ethnische Herkunft als Einflussfaktor auf die OSCE	43
5.2.1	Zusammenfassung der Ergebnisse	43
5.2.2	Kritische Betrachtung der Ergebnisse	44
5.2.3	Vergleich mit bisherigen Publikationen	45
5.3	Geschlecht als Einflussfaktor auf die OSCE	46
5.3.1	Zusammenfassung der Ergebnisse	46
5.3.2	Kritische Betrachtung der Ergebnisse	46
5.3.3	Vergleich mit bisherigen Publikationen	47
5.4	Akzeptanz der OSCE im Vergleich mit bisherigen Publikationen	48
5.5	Schlussfolgerung und Forschungsausblick	49
6	Zusammenfassung	51
7	Literaturverzeichnis	52
8	Anhang	64
8.1	Bewertungsbogen für die CVR-Station	64
8.2	Bewertungsbogen für die AU-Station	65
8.3	Bewertungsbogen für die Depressionsstation	66

8.4	Bewertungsbogen für die EKG-Station.....	67
8.5	Bewertungsbogen für die Dermatologiestation	68
8.6	Bewertungsbogen für die Reanimationsstation	69
8.7	Bewertungsbogen für die Hygienestation.....	70
8.8	Bewertungsbogen für die Rezeptstation	71
8.9	Studierendenanleitung für die CVR-Station.....	72
8.10	Studierendenanleitung für die AU-Station	72
8.11	Studierendenanleitung für die Depressionsstation.....	73
8.12	Studierendenanleitung für die EKG-Station.....	73
8.13	Studierendenanleitung für die Dermatologiestation	74
8.14	Studierendenanleitung für die Rezeptstation (Version 1).....	74
8.15	Studierendenanleitung für die Rezeptstation (Version 2).....	75
8.16	Evaluationsbogen.....	76

Abbildungsverzeichnis

Abbildung 1: System zur Beurteilung von Fertigkeiten (nach Miller 1990).....	4
Abbildung 2: Aufbau der Pilot-OSCE im WS 2003/04.....	13
Abbildung 3: Aufbau der OSCE im WS 2004/05.....	15
Abbildung 4: Checklistennoten der Tutoren und Ärzte.....	21
Abbildung 5: Globalbeurteilungen der Tutoren und Ärzte.....	22
Abbildung 6: Gesamtnoten der Tutoren und Ärzte	23
Abbildung 7: Checklisten-, Global- und Gesamtnoten zusammengefasst	24
Abbildung 8: Arzt-Gesamtnoten für deutsche und ausländische Studierende	31
Abbildung 9: Tutoren-Gesamtnoten für deutsche und ausländische Studierende	31
Abbildung 10: Arzt-Gesamtnoten für männliche und weibliche Studierende.....	34
Abbildung 11: Tutoren-Gesamtnoten für männliche und weibliche Studierende	34
Abbildung 12: Beliebteste Prüfungsform	36
Abbildung 13: Angemessenheit der studentischen Bewertung	37
Abbildung 14: Zutrauen der Studierenden zum Selberprüfen	38

Tabellenverzeichnis

Tabelle 1: Vergleich des Aufwandes für verschiedene Prüfungsformen	9
Tabelle 2: Übersicht der Klasseneinteilung der Kappa-Werte (nach Altmann 1991) ..	20
Tabelle 3: Überprüfung auf Normalverteilungen	20
Tabelle 4: Prüfer-Übereinstimmung bei den Checklistennoten.....	22
Tabelle 5: Prüfer-Übereinstimmung bei den Globalbeurteilungen.....	23
Tabelle 6: Prüfer-Übereinstimmung bei den Gesamtnoten	24
Tabelle 7: Prüfer-Übereinstimmung bei den Stationszusammenfassungen	25
Tabelle 8: Item-Übereinstimmungen der Tutoren und Ärzte an der CVR-Station.....	26
Tabelle 9: Item-Übereinstimmungen der Tutoren und Ärzte an der AU-Station.....	27
Tabelle 10: Item-Übereinstimmungen der Tutoren und Ärzte an der DEP-Station	28
Tabelle 11: Item-Übereinstimmungen der Tutoren und Ärzte an der EKG-Station.....	28
Tabelle 12: Checklistennoten für deutsche und ausländische Studierende	29
Tabelle 13: Globalbeurteilungen für deutsche und ausländische Studierende	30
Tabelle 14: Checklistennoten für männliche und weibliche Studierende.....	32
Tabelle 15: Globalbeurteilungen für männliche und weibliche Studierende	33
Tabelle 16: Akzeptanz der OSCE.....	36

Abkürzungsverzeichnis

ÄAppO	=	Ärztliche Approbationsordnung
AU	=	Arbeitsunfähigkeitsanamnese-Station
CVR	=	kardiovaskuläre-Risikoanamnese-Station
DEP	=	Depressionsanamnese-Station
DERMA	=	Dermatologiestation
EKG	=	EKG-Station
HYG	=	Hygienestation
KI	=	Konfidenzintervall
MCQ	=	Multiple Choice Questions
n	=	Gesamtmenge
OSCE	=	objective structured clinical examination
REA	=	Reanimationsstation
Rezept	=	Rezeptstation
SAS	=	Statistical Analysis System (amerikanisches Statistikprogramm)
SD	=	standard deviation (Standardabweichung)
SP	=	Simulationspatient
SS	=	Sommersemester
WS	=	Wintersemester

1 Einleitung

1.1 Einführung der neuen Approbationsordnung

Die Approbationsordnung für Ärzte (ÄAppO) (Bundesministerium für Gesundheit 2005) regelt, welche praktischen und theoretischen Kenntnisse die Medizinstudierenden für den ärztlichen Beruf erwerben sollen, und bestimmt somit den Verlauf und die Zielsetzung des Medizinstudiums.

Das Medizinstudium gliedert sich in zwei Abschnitte. Auf die Vorklinik von vier Semestern folgt der klinische Abschnitt für die restlichen sechs Semester mit einem anschließenden praktischen Jahr. In der Vorklinik werden die Grundlagen für den klinischen Teil des Studiums geschaffen, in dem die Krankheitsbilder einzelner Fachgebiete mit zugehöriger Diagnostik und Therapie vorgestellt werden. Die endgültige Gestaltung der Praktika und Seminare innerhalb des Semesters obliegt der jeweiligen Universität.

In den 90er Jahren wurde vermehrt Kritik über mangelnde Vermittlung von praktischen Fertigkeiten im Medizinstudium geäußert (Pabst 1995). Der Schwerpunkt des Studiums lag auf der Vermittlung theoretischer Kenntnisse und es gab nur wenige Angebote, praktische Fertigkeiten und Erfahrungen im direkten Patientenkontakt zu erwerben. Daher wurde eine neue ÄAppO gefordert, die die Praxis in den Vordergrund stellt und somit die Medizinstudierenden besser auf die Zukunft als eigenverantwortliche und selbstständige Ärzte vorbereitet.

Der Ausschuss „Ausbildung zum Arzt, Hochschule und medizinische Fakultät“ der Bundesärztekammer begleitet die Inhalte und Möglichkeiten der ärztlichen Ausbildung fortlaufend kritisch und hat ebenfalls einen Änderungsbedarf hinsichtlich der Qualität der Lehre erkannt (Bundesärztekammer 2004). Somit wurde die ÄAppO aus dem Jahr 1997 novelliert, am 27.06.2002 mit Zustimmung des Bundesrates verabschiedet und mit dem 01.10.2003 in Kraft gesetzt. Stellungnahmen der Bundesärztekammer haben Eingang in den neuen Gesetzesentwurf gefunden.

Mit der bundesweiten Einführung der neuen ÄAppO steht nun das praxisorientierte Lernen im Vordergrund. Durch Kleingruppen-Seminare erhöht sich der persönliche Kontakt zwischen Studierenden und Lehrpersonal. Während nach der alten Studienordnung die ersten Patientenkontakte und Einblicke in den ärztlichen Alltag vor allem in

den Famulaturen und im praktischen Jahr stattfanden, wird nun neben dem problemorientierten Lernen auch Unterricht am Krankenbett durchgeführt. Der praktische Anteil im Medizinstudium wurde gestärkt, um den Medizinstudierenden klinische Kompetenz zu vermitteln.

1.2 „Kursus für ärztliche Basisfertigkeiten“

Klinische Kompetenz umfasst Wissen, Fertigkeiten und Interaktion mit den Patienten (Campos-Outcalt et al. 1994, Lai et al. 2007). Dazu gehören z. B. die Befähigung zur zwischenmenschlichen Beziehung und Kommunikation, zur Anamneseerhebung und zur manuellen oder apparativen Untersuchung (Newble 1992). Auch die Planung und Interpretation der Diagnostik, die Therapiegestaltung sowie die Patientenaufklärung gehören zur klinischen Kompetenz (Harden and Gleeson 1979).

Um die klinische Kompetenz der Medizinstudierenden zu fördern, wurde an der Georg-August-Universität Göttingen u. a. der „Kursus für ärztliche Basisfertigkeiten“ in das erste klinische Semester des Medizinstudiums eingeführt. Der Kurs ist aus einer Fusion des bisher fakultativen „Spritzenkurses“ der Abteilung Allgemeinmedizin und des „Anamnesekurses“ der Abteilung Psychosomatik und Psychotherapie entstanden. Der „Spritzenkurs“ bot jeweils 24 Studierenden die Möglichkeit, Erfahrungen bei Venenpunktionen, Injektionen, Nahttechniken und Verbänden zu sammeln. In dem „Anamnesekurs“ lernten 30 Studierende möglichst angstfrei eine Arbeitsbeziehung mit einem Patienten aufzubauen und mit schwierigen Gesprächssituationen umzugehen. Beide Kurse waren in den zurückliegenden Semestern regelmäßig überlaufen, so dass nicht alle interessierten Studierenden teilnehmen konnten.

Im WS 2003/04 wurde zunächst ein Probedurchlauf des Kurses „Ärztliche Basisfertigkeiten“ (Pilotkurs) für 16 Studierende angeboten, die durch ein Losverfahren ermittelt wurden. Der Zeitpunkt für diesen Kurs ist günstig, da den Studierenden die nötigen klinischen Grundfertigkeiten für die erste Famulatur gelehrt werden. Die Abteilungen Allgemeinmedizin, Psychosomatik/Psychotherapie sowie Arbeits- und Sozialmedizin vermitteln kommunikative und praktische Fertigkeiten des medizinischen Alltags (Fischer et al. 2005, Simmenroth-Nayda et al. 2004).

Die strukturierte Anamneseerhebung wird in Kleingruppen an den ersten Kurstagen in Rollenspielen und gegen Ende des Kurses mit Simulationspatienten (SP) geübt. Für die SP finden vorab Schulungen statt, damit sie sich auf ihre Rollen vorbereiten können und die Anamnese an Authentizität gewinnt. Das Verhalten der SP ähnelt in der Regel dem von realen Patienten, daher gelten SP als Gewinn für die medizinische Lehre (Simmenroth-Nayda et al. 2007, Stillman and Swanson 1987, Stillman et al. 1987). Einige Studierende vergessen sogar, dass der Patient nicht „echt“ ist (Barrows 1968). Bei den Anamneseübungen ist immer ein studentischer Tutor oder ein Dozent gegenwärtig, der im Anschluss an das Gespräch die Diskussion innerhalb der Kleingruppe leitet und zusammen mit dem SP Feedback gibt.

Der praktische Teil des Kurses beinhaltet nach kurzen thematischen Einführungen die korrekte Handhabung von Injektionen und das Legen von Venenverweilkanülen, die richtige Durchführung von Lungenfunktion und Blutzuckermessungen sowie das Anlegen und Interpretieren eines EKGs auf Basisniveau. Des Weiteren erhalten die Studierenden einen Einblick in die Physiotherapie und eine Einführung in das Nähen.

Sowohl von Seiten der Studierenden als auch von den Dozenten und SP gibt es eine sehr positive Rückmeldung bezüglich des neuen Kurses, dem ein hoher Lerneffekt zugesprochen wird (Simmenroth-Nayda et al. 2004). Seit dem Sommersemester 2004 wird an der Universität Göttingen der Kurs für das gesamte Semester als Pflichtkurs durchgeführt.

Am Ende des Basisfertigkeitenkurses findet eine praktische Prüfung statt, die *Objective Structured Clinical Examination* (OSCE). Diese Prüfung soll die während des Kurses erworbene klinische Kompetenz der Studierenden beurteilen (Sloan et al. 1995). Schriftliche und mündliche Prüfungen sind dazu nur begrenzt bzw. gar nicht geeignet, da sie vor allem Faktenwissen und Theorie prüfen (Mavis 2000). Die OSCE prüft sowohl theoretisches Wissen als auch klinische Fähigkeiten (Coovadia and Moosa 1985); ihre Ergebnisse können besser als schriftliche Tests klinische Kompetenz voraussagen (Campos-Outcalt et al. 1994, Mavis 2000).

Miller zeigte 1990 in einer Hierarchie der klinischen Kompetenz zu jeder Kompetenzebene die entsprechende Prüfungsform (Abb. 1).

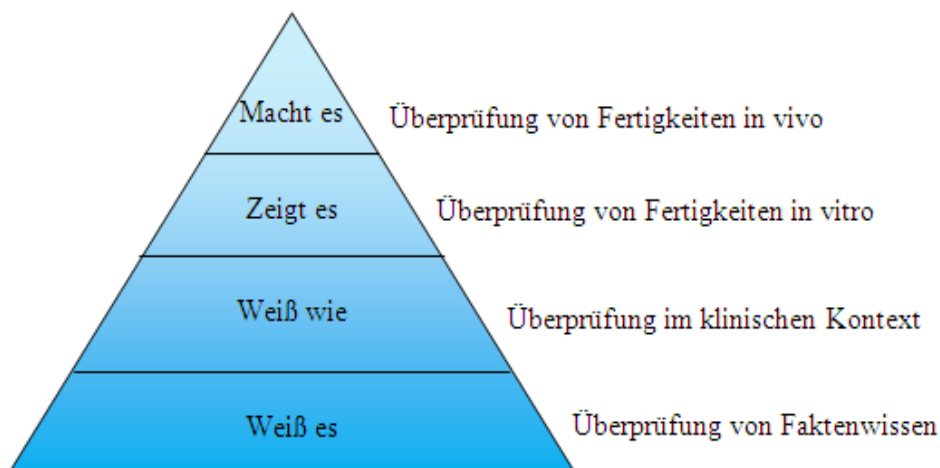


Abbildung 1: System zur Beurteilung von Fertigkeiten (nach Miller 1990)

Die beiden unteren Bereiche der Kompetenzpyramide werden mittels *Multiple Choice Questions* (MCQ) und mündlichen Prüfungen geprüft. In dem dritten Bereich wird die OSCE eingesetzt und die Pyramidenspitze könnte mit „Undercover“-Patienten oder Videos überprüft werden.

1.3 Die OSCE (objective structured clinical examination)

1.3.1 Struktur der Prüfungsform „OSCE“

Die OSCE wurde erstmals 1975 von Harden et al. in Schottland eingeführt. Sie besteht aus einem Parcours von mehreren „Stationen“, an denen bestimmte vorgegebene Aufgaben ausgeführt werden. Dem Studierenden steht für jede Station ein festgelegter Zeitraum zur Verfügung, um seine Aufgabe unter Beobachtung und Bewertung mindestens eines Prüfers zu erfüllen. An jeder Station erhält der Studierende eine kurze und präzise schriftliche Instruktion. Nach Ablauf der vorgegebenen Zeit wird zu der nächsten Station gewechselt. Dieser Parcours kann von mehreren Studierenden gleichzeitig bestritten werden, da der Stationswechsel einheitlich erfolgt.

Das Spektrum an Prüfungsthemen ist auf Grund der Fallspezifität sehr breit (Elstein et al. 1978), von daher kann die OSCE in vielen Bereichen des Medizinstudiums durchgeführt werden. An den Stationen können neben SP auch Puppen bzw. Attrappen, Bilder oder klinische Befunde zur Interpretation verwendet werden (Duffield and Spencer 2002, Mavis et al. 2002).

Die SP sind schon seit den 60er Jahren fester Bestandteil der OSCE (Barrows 1968). Sie können gesunde Menschen sein oder Patienten mit zu befundenen Beschwerden. Sie sind trainiert, einen speziellen Fall aus dem klinischen Alltag konstant und konsistent darzustellen. Die SP werden genau deshalb in der Fallvorführung echten Patienten gegenüber bevorzugt, damit gerechterweise jeder Prüfling dasselbe Problem lösen muss (Martin et al. 1996). Im Vergleich zu realen Patienten haben die SP weitere Vorteile, z.B. müssen die Studierenden keine Angst haben, den Patienten bei der körperlichen Untersuchung Schmerzen zuzufügen oder die Erkrankung durch falsche Handgriffe zu verschlimmern. Außerdem kann über alle möglichen Komplikationen der dargestellten Erkrankung frei diskutiert werden; ein objektives Feedback ist ebenfalls möglich (Barrows 1968).

Es besteht die Möglichkeit, kurze MCQs einzugliedern, in denen Fragen zu einer vorhergehenden Station gestellt werden. Die OSCE kann eine gute Aussage über zukünftige klinische Fähigkeiten machen, wenn sie mit einem schriftlichen Teil kombiniert wird (Wilkinson and Frampton 2004).

Die Beurteilung der Leistung der Studierenden erfolgt meist durch Ärzte, die klar definierte Bewertungsvorgaben haben. Mit einer Checkliste beurteilen sie Teilaspekte der zu erfüllenden Aufgabe; hierbei haben sie im Wesentlichen die Auswahl zwischen „erfüllt“ und „nicht erfüllt“. Abschließend geben die Prüfer noch eine Globalbeurteilung ab. Die Globalbeurteilung ist separat und umfasst den Patientenumgang, die Sprache, die Körperhaltung des „Arztes“ und soll den Gesamteindruck mit einer Note von zu meist 1 bis 5 wiedergeben. Die Beurteilung der klinischen Kompetenz von Studierenden wird somit in einer umfassenden, konsistenten und strukturierten Weise durchgeführt, die Objektivität gewährleisten soll (Harden 1988).

1.3.2 Gründe für die Einführung der OSCE

Mangelnde Empathie, unzulängliche Anamneseführung sowie Inkompetenz der Studierenden in den Basisfähigkeiten waren Gründe für die Einführung der OSCE (Waterston et al. 1980). Die OSCE ist als Beurteilungssystem wichtig für die Qualitätskontrolle, Entwicklung und Sicherung von klinischen Fähigkeiten, was mit den herkömmlichen Multiple-Choice-Prüfungen nicht gewährleistet wird (Stillman et al. 1987). Sie ist das ideale Prüfungsformat, um praktische Fertigkeiten zu prüfen und durch die Änderung

der Prüfungsform wird das Lernverhalten der Studierenden beeinflusst (Elton and Laurillard 1979). OSCEs fördern das Lernen von praktischen, kommunikativen und zwischenmenschlichen Fähigkeiten und bewirken dadurch eine Verbesserung der klinischen Kompetenz und ein größeres Ausmaß an realistischer Selbsteinschätzung (Schoonheim-Klein et al. 2006).

Obwohl die OSCE eine international anerkannte und etablierte Prüfungsform ist, hat sie an den deutschen Universitäten bisher nur in kleineren Formaten in Düsseldorf, Göttingen, Hannover, Münster, Ulm und Witten-Herdecke stattgefunden (Chenot und Ehrhardt 2003). Bisher wurden im deutschen Medizinstudium überwiegend Prüfungen in MCQ-Form durchgeführt. Lange Zeit gab es nahezu keine systematische Beurteilung von praktischen Fertigkeiten und von Fähigkeiten zum Aufbau von zwischenmenschlichen Beziehungen und zur Kommunikation. Nur eine praktische Prüfung kann eine Aussage machen, ob der Studierende die Kenntnisse und Fertigkeiten für das spätere Berufsleben beherrscht (Falck-Ytter und Stiegler 1993). Die Studierenden, die ihr Medizinstudium nach alter Approbationsordnung ablegten, konnten vor ihrem Staatsexamen ihre klinischen Fertigkeiten nicht unter Beweis stellen und bewerten lassen.

1.3.3 Entwicklung einer OSCE

Die OSCE ist eine arbeits- und zeitaufwendige Prüfungsform (Falck-Ytter und Stiegler 1993, Harden et al. 1975). Es müssen zunächst klinisch relevante Prüfungsstationen für die OSCE konzipiert werden. Dafür werden Checklisten für die Stationen erstellt, die dem Prüfer verschiedene Aspekte der zu prüfenden Handlung darlegen, die entweder vom Prüfling erfüllt werden oder nicht (Andrew 1977). Hierzu wird Fachliteratur zu Hilfe genommen und genau überlegt, was die wichtigen Schritte bei der Durchführung der Stationsaufgabe sind (O'Conner and McGraw 1997). In diesem Zusammenhang können vorab Minimalanforderungen definiert und die Prüfung kann an das Leistungsniveau angepasst werden. Die Bewertungsbögen sollen ein übersichtliches Design haben und leicht in der Handhabung sein. Aus diesen Gründen werden verständliche dichotome Items bevorzugt (Andrew 1977).

Die Checklisten werden Spezialisten vorgelegt und deren Meinung und Verbesserungsvorschläge eingeholt. Daraufhin folgt die Testung mit Prüfern und Studierenden in einem kleinen Rahmen, um zu kontrollieren, ob die Checklisten in angemessener Zeit

auszufüllen sind und ob die Items die wesentlichen Handlungsaspekte berücksichtigen. Vor der OSCE wird eine mindestens 30-minütige Checklistenführung für alle Prüfer empfohlen; dann sollten auch unerfahrene Prüfer keine Schwierigkeiten bei der Bewertung mit den Listen haben (O'Conner and McGraw 1997). Sind die Checklisten entworfen, werden für die Studierenden präzise schriftliche Instruktionen verfasst, die an den OSCE-Stationen während der Prüfung ausliegen (Falck-Ytter und Stiegler 1993).

1.3.4 Testgütekriterien der OSCE

Eine ideale Prüfung erfüllt die Kriterien: Validität, Reliabilität, Objektivität und Durchführbarkeit (Harden and Gleeson 1979).

Das entscheidende Gütekriterium ist die Validität, nämlich die Frage, ob der Test tatsächlich überprüft, was er vorgibt zu prüfen. Ob das Prüfungsergebnis tatsächlich mit dem Handeln in konkreten Situationen korreliert, ist methodisch jedoch nur schwer zu untersuchen (z.B. durch „Undercover“-Patienten). Allerdings wird der OSCE bei gut entwickelten Stationen eine hohe Augenschein-Validität zugesprochen. Die Ergebnisse von MCQ und praktischen Prüfungen korrelieren nur gering, da jeweils unterschiedliche Parameter - theoretisches Wissen gegenüber praktischen Fertigkeiten - erfasst werden (Coovadia and Moosa 1985).

Die Reliabilität ist ein Maß für die Zuverlässigkeit der Prüfung und gibt zusätzlich Auskunft über deren Objektivität. Unter Zuverlässigkeit versteht man die Konsistenz der Genauigkeit der Ergebnisse. Im Falle von OSCEs unterscheidet man zwischen Interrater-Reliabilität (= die Prüferkorrelation) und der Interstation-Reliabilität, die Auskunft über die Leistung eines Prüflings an den verschiedenen Stationen gibt (Chenot und Ehrhardt 2003). Die Retest-Reliabilität macht eine Aussage über die Leistung von Studierenden bei einer sich wiederholenden Prüfung. Je mehr Stationen in einer OSCE integriert sind, desto höher wird die Reliabilität. Eine Methode, um die Reliabilität zu messen, ist die Übereinstimmung zwischen Prüferbeobachtungen am gleichen Prüfling (Van der Leuten et al. 1991), die als „Cronbachs Alpha“ berechnet wird; für die OSCE werden sehr gute Werte zwischen 0,77 und 0,93 angegeben (Doig et al. 2000, Sloan et al. 1995, Stillman et al. 1987). Eine Erhöhung der Itemanzahl an den Stationen verringert sowohl Reliabilität als auch Validität, da die Prüfer sich weniger auf den Prüfling und dessen Darbietung, sondern stärker auf die Checklisten konzentrieren müssen (Wilkin-

son et al. 2003b). Die Varianz in der einzelnen Stationsbeurteilung gleicht sich mit zunehmender Länge der OSCE aus. Es werden mindestens zehn Stationen als notwendig angesehen; ab 18 Stationen sinkt der Reliabilitätskoeffizient bereits wieder (Chenot und Ehrhardt 2003). Durch die Genauigkeit der Checklistenitems wird eine bessere Reliabilität gewährleistet, vor allem wenn die Prüfer bei der Erstellung der Stationen und Checklisten integriert werden (Newble et al. 1980, Wilkinson et al. 2003b).

Studierende ändern ihr Verhalten in der Prüfung entsprechend der von ihnen erwarteten Bewertungsweise der Prüfer (McIlroy et al. 2002); wenn sie annahmen, nur mit Checklisten beurteilt zu werden, zeigten sie bessere Leistungen in diesem Bewertungsteil. Der alleinige Gebrauch von Checklisten wird kritisiert, da diese wegen ihrer Kürze nicht alle Aspekte berücksichtigen können (Waterston et al. 1980). Globalbeurteilungen erfassen ein breiteres Spektrum, während die Checklisten sehr inhaltspezifisch sind, daher ist eine Kombination sinnvoll, um die Reliabilität zu erhöhen. Insgesamt aber wird die OSCE als hoch valide und reliabel angesehen (Elliot et al. 1994, Sloan et al. 1995, Watson et al. 1982).

Die Objektivität ist auf Grund der Benutzung von stark standardisierten Checklisten gegeben. Etwaige Prüfererfahrungen können von daher allenfalls die Globalbeurteilung beeinflussen (Wilkinson et al. 2003b). Es gibt dabei vor allem Übereinstimmungen bei guten Leistungen, aber nicht bei einer schlechteren Leistung. Deshalb empfiehlt es sich, zwei Prüfer zur Bewertung heranzuziehen, so dass sich diskrepante Beurteilungen ausgleichen können (Burchard et al. 1995).

Hinsichtlich der Durchführbarkeit ist zu beachten, dass ausreichend Prüfer anwesend und die entsprechenden Räumlichkeiten für die OSCE gegeben sind, damit genügend Studierende teilnehmen können (Harden and Gleeson 1979).

1.3.5 Vor- und Nachteile der OSCE

Durch die Einführung der OSCE hat die Lehre und das Erlernen der klinischen Basisfähigkeiten eine neue und wachsende Bedeutung erhalten - jedenfalls im Ausland. Die Ärzte können erkennen, wo die Defizite bei den Studierenden liegen (Kowlowitz et al. 1991). Gleichzeitig erhalten die Dozenten direkt ein Feedback zu ihrer Lehre und die Studierenden werden auf ihre Schwächen aufmerksam gemacht (Watson et al. 1982).

Sowohl Studierende als auch Lehrende äußern sich positiv zur OSCE und stufen die Prüfung als gerecht ein (Stillman et al. 1987, Watson et al. 1982). Die zeitliche Begrenzung auf fünf Minuten pro Station wird allerdings als Nachteil angesehen, denn es ist nicht möglich, den Patienten als ganze Persönlichkeit in dieser Zeit kennen zu lernen. Es wird dadurch eine „potenziell artifizielle Situation“ geschaffen (Falck-Ytter und Stiegler 1993, S.54).

Im Vergleich zu den schriftlichen Prüfungen gilt die OSCE als die bessere Prüfungsform für praktische Fertigkeiten (Waterston et al. 1980). Der Aufwand für eine OSCE ist im Verhältnis zu den obligaten Prüfungen jedoch deutlich größer (Chenot et al. 2004a; Tab. 1).

Tabelle 1: Vergleich des Aufwandes für verschiedene Prüfungsformen

<i>Aufwand für...</i>	<i>OSCE</i>	<i>Offene Fragen schriftlich</i>	<i>Multiple Choice Prüfung</i>	<i>Mündliche Prüfung</i>
Entwicklung	++++	+++	++	+++
Durchführung	+++	+	+	+++
Auswertung	++	++	+	+

Es wird eine sorgfältige Schulung der Prüfer vor der OSCE gefordert (Falck-Ytter und Stiegler 1993), damit jüngere Ärzte ohne Prüfererfahrung genauso korrekt bewerten können wie erfahrene Ärzte (Ogden et al. 2000, Reiter et al. 2004, Van der Vleuten et al. 1989). Dann können sogar Fachfremde durch Einweisungen in die Handhabung von Checklisten zur Bewertung in der OSCE herangezogen werden. Bisher wurden jedoch überwiegend Ärzte als Prüfer in der OSCE eingesetzt; es ist noch nicht untersucht worden, ob Studierende in einer OSCE prüfen können.¹ Es wird jedoch auch die Ansicht vertreten, dass es entweder konsistente Prüfer oder nicht-konsistente Prüfer gibt. Letztere ließen sich auch durch ein Training nicht beeinflussen. Daher müsse man in einer Art Test-OSCE die Eignung der Prüfer feststellen (Newble et al. 1980).

Die praktische Prüfung ist in vielen klinischen Fachgebieten einsetzbar. Diese Fachgebiete können zusammen eine OSCE erstellen, so wie es im Göttinger Basisfertigkeiten-

¹ Persönliche Mitteilung am 01.04.2006 von Prof. Cees van der Vleuten, Professor of Education Department of Educational Development and Research University of Maastricht:

„I am not aware of any literature on students ratings in OSCE“

kurs des ersten klinischen Semesters der Fall ist. Ansonsten kann die praktische Prüfung auch in eine bereits bestehende Semesterabschlussprüfung integriert werden.

Die OSCE ist jedoch wegen des hohen Personal- und Materialaufwands eine sehr kostspielige Prüfungsform. Die Angaben für die Kosten sind sehr unterschiedlich und hängen mit der Größe der OSCE sowie mit deren Entwicklung, Pilotierung und den Schulungen der Prüfer und SP zusammen. Eine OSCE von vier Stunden Dauer für 120 Studierende würde ungefähr zwischen 60.000 und 100.000 \$ kosten (Elliot et al. 1994). Ohne die Aufwandsentschädigungen für das Personal belaufen sich die Kosten für eine OSCE auf ungefähr 90 bis 100 \$ pro Teilnehmer (Carpenter 1995). Die Universität Hannover gab nach einer Kostenanalyse aus dem Jahr 2000 einen Betrag von ca. 130 DM / Studierenden an (Blume et al. 2000).

Obwohl die OSCE im Vergleich zu den herkömmlichen Prüfungsformen vergleichsweise kostenaufwendig ist, gibt es kaum Berichte, wonach die Einführung einer OSCE aus Kostengründen gescheitert sei. Dies hängt einerseits mit der überwiegenden Meinung zusammen, dass sich der finanzielle Aufwand lohnt, andererseits könnte sich darin auch ein Publikationsbias ausdrücken (Almeida Troncon 2004).

1.3.6 Einfluss von ethnischer Herkunft und Geschlecht

Eine objektive Prüfung sollte Menschen auf Grund ihrer ethnischen Herkunft oder ihres Geschlechts nicht benachteiligen. Es gibt bereits viele internationale Untersuchungen darüber, ob ausländische und/ oder männliche Studierende einen Nachteil durch die Prüfungsform OSCE haben könnten. In den meisten Studien erzielten weibliche hellhäutige Studierende bessere Prüfungsergebnisse als die männlichen und ausländischen Studierenden (Haq et al. 2005, Woolf et al. 2007). Ausländische Studierende wurden vor allem an kommunikativen Stationen einer OSCE schlechter bewertet, unabhängig davon, ob sie mit der Sprache des Studienlandes oder mit ihrer eigenen Heimatsprache aufgewachsen sind (Fernandez et al. 2007, Schoonheim-Klein et al. 2007). Diese Literatur stammt überwiegend aus den USA, Großbritannien und den Niederlanden. Auf Grund der Bevölkerungszusammensetzung dieser Länder befassen sich die Studien hauptsächlich mit Unterschieden zwischen den Prüfungsergebnissen von farbigen, asiatischen und hellhäutigen Studierenden. Die Besonderheiten der Nachkriegszeit und die Immigration von Gastarbeitern aus süd- und osteuropäischen Ländern und Vorderasien

haben in Deutschland zu einer anderen Zusammensetzung des Ausländeranteils geführt (Statistisches Bundesamt Deutschland 2006a und b). Bisher gibt es nur sehr wenige Untersuchungen zu möglichen Problemen von ausländischen Studierenden in einer deutschen OSCE (Chenot et al. 2007).

Es liegen bereits einige Studien zum Geschlechtervergleich in einer OSCE vor. Danach wurden weibliche Studierende vor allem in den kommunikativen Teilen der OSCE besser bewertet (Dewhurst et al. 2007, Wiskin et al. 2004). Als mögliche Gründe werden genannt, dass Frauen besser zuhören können (Clack and Head 1999), einen größeren Sinn für das Patientenwohl haben (Zaharias et al. 2004) und leichter eine gute Zusammenarbeit in einem Patienten-Gespräch herstellen können (Skelton and Hobbs 1999). In dem Kurs „Ärztliche Basisfertigkeiten“ der Universität Göttingen werden in Kleingruppen von vier bis sechs Studierenden kommunikative Fähigkeiten intensiv gelehrt und nach jeder Anamnese wird von den studentischen Tutoren oder Dozenten und den SP ein umfangreiches Feedback gegeben. Der Schwierigkeitsgrad der Anamneseerhebung steigt mit jedem Kurs; es wird auch das Überbringen von schlechten Nachrichten und das erforderliche Einfühlungsvermögen thematisiert. Die Kleingruppen erlauben es, sich verstärkt um Studierende mit schwächeren kommunikativen Fähigkeiten zu kümmern und eine umfassende Hilfestellung zu geben. Männliche und weibliche Studierende werden dabei gleichermaßen behandelt. Am Ende des Kurses werden die kommunikativen Fähigkeiten mit der OSCE überprüft. Die Ergebnisse können möglicherweise Hinweise auf eine notwendige Verbesserung der Kursstruktur in Form von einer noch intensiveren Betreuung der kommunikationsschwächeren Studierenden geben.

2 Fragestellungen und Ziele

Bisher wurden studentische Tutoren vor allem als Hilfskräfte in der Lehre eingesetzt; bei Prüfungen beschränkten sich ihre Aufgaben auf die Klausuraufsicht (Harris et al. 1990, Heylings and Stefani 1997). Für die breitere Einführung der OSCE wäre es von Vorteil zu wissen, ob sich Studierende höherer Semester für eine Leistungsüberprüfung bei jüngeren Semestern eignen. Die Planung der OSCE würde dadurch erheblich erleichtert werden, da Lehrärzte nicht in so großer Anzahl und so flexibel zur Verfügung stehen wie die Studierenden. Somit könnten auch kleinere Abteilungen Stationen für eine OSCE entwickeln oder sogar eine eigene OSCE durchführen. Als wissenschaftliche Hilfskraft erhalten Studierende von der Georg-August-Universität ein Bruttogehalt von 10,52 € / Stunde, während Ärzte je nach Stand der Aus- und Weiterbildung einen Lohn zwischen 19,71 € und 30,12 € / Stunde² bzw. zwischen 16,83 € und 26,05 € / Stunde³ erhalten (Tarifgemeinschaft deutscher Länder 2006). Demnach könnten die Kosten durch eine vermehrte Integration von studentischen Prüfern in die OSCE reduziert werden. Die Hauptfrage dieser Dissertation ist, ob die Bewertungen der studentischen Tutoren und der ärztlichen Prüfer übereinstimmen. Zusätzlich wird die Akzeptanz der Bewertung durch studentische Tutoren evaluiert.

Es sollte in der OSCE eine hohe Objektivität und Gerechtigkeit gegenüber allen Studierenden gegeben sein. Als Nebenaspekt der Dissertation werden die Prüfungsergebnisse auf mögliche Unterschiede zwischen deutschen und ausländischen Studierenden untersucht und überprüft, ob ein Geschlecht durch die Prüfungsform bevorteilt wird.

² Ärzte in der Patientenversorgung bei einer 42 Stundenwoche (182,62 Stunden/Monat)

³ Ärzte ohne Patientenversorgung bei einer 38,5 Stundenwoche (167,4 Stunden/Monat)

3 Material und Methoden

3.1 Pilotierung der OSCE

Eine OSCE wurde in Göttingen das erste Mal im WS 2003/04 als Probedurchlauf („Pilot“-OSCE) im Rahmen des Kurses „Ärztliche Basisfertigkeiten“ durchgeführt. Sie sollte Erfahrungen in Organisation und Durchführung liefern und die Akzeptanz durch die Studierenden ermitteln.

An der OSCE nahmen 22 Studierende teil, die sich aus 16 Kursteilnehmern und 6 freiwilligen Studierenden anderer Semester zusammensetzten. Es wurden neun Stationen von jeweils fünf Minuten Dauer entwickelt (Abb. 2). Sechs Abteilungen waren beteiligt: Allgemeinmedizin, Anästhesiologie/Notfallmedizin, Arbeits- und Sozialmedizin, Dermatologie, Hygiene und Rechtsmedizin. Die Prüfungsinhalte wurden im Kurs „Ärztliche Basisfertigkeiten“ und anderen Kursen vermittelt. Nur die Aufgabe, einen Totenschein auszufüllen (Rechtsmedizin), war nicht Lehrinhalt des ersten klinischen Semesters. Sie wurde unter dem Aspekt der Eignung für die neue Prüfungsform eingegliedert.

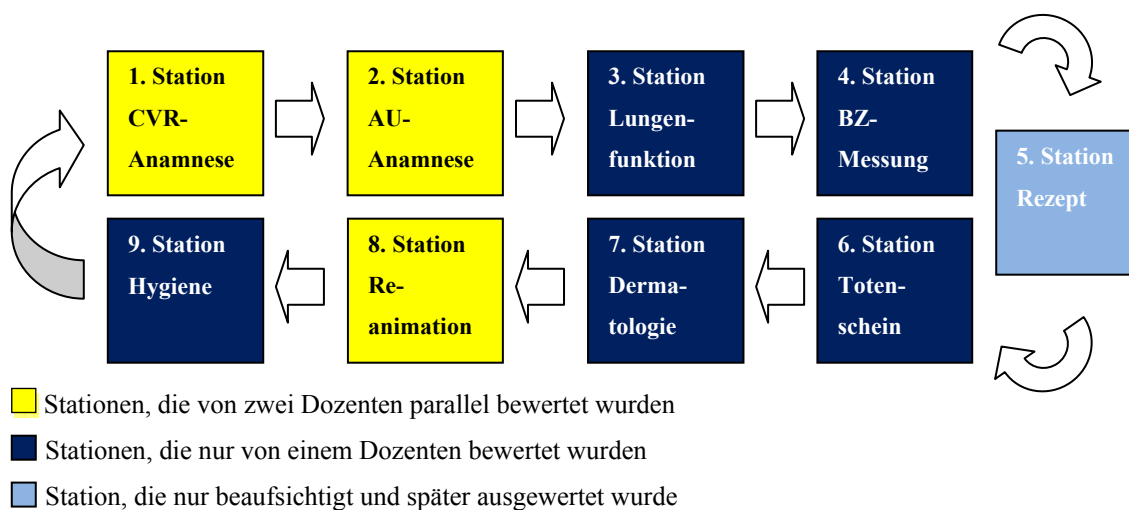


Abbildung 2: Aufbau der Pilot-OSCE im WS 2003/04

Die Stationen der Allgemeinmedizin und der Arbeits- und Sozialmedizin (1.- 5. Station, Abb. 2) wurden vorab mit den studentischen Tutoren und Dozenten des Kurses pilotiert und optimiert. Es fand ein Prüfertraining für die Ärzte statt und die SP wurden für die Anamnesestationen geschult.

Zumindest an drei Stationen konnten zwei Prüfer eingesetzt werden (Abb. 2); die ermittelten Interrater-Reliabilitäten waren mit 0,72 bis 0,86 gut (Chenot et al. 2004b). Die Rezeptstation befand sich unter Aufsicht einer studentischen Hilfskraft. An der Rezept- und Hygienestation vergaben die Prüfer keine Globalbeurteilungen.

Im Anschluss an die Pilot-OSCE evaluierten die Teilnehmer die Prüfung und es fand eine gemeinsame Abschlussbesprechung statt. Die Studierenden und Lehrenden empfanden die OSCE als angemessen und gerecht. Kritik an einigen Aspekten (zu viele und teilweise missverständliche Checklisten-Items, zu schwierige Differenzierung bei drei Beurteilungsmöglichkeiten pro Item [Likert-Skalierung], zu lange Instruktionen der Studierenden bei 5 Minuten Prüfungszeit) führte zu einer Verbesserung der folgenden OSCEs.

3.2 Die OSCE des Wintersemesters 2004/05

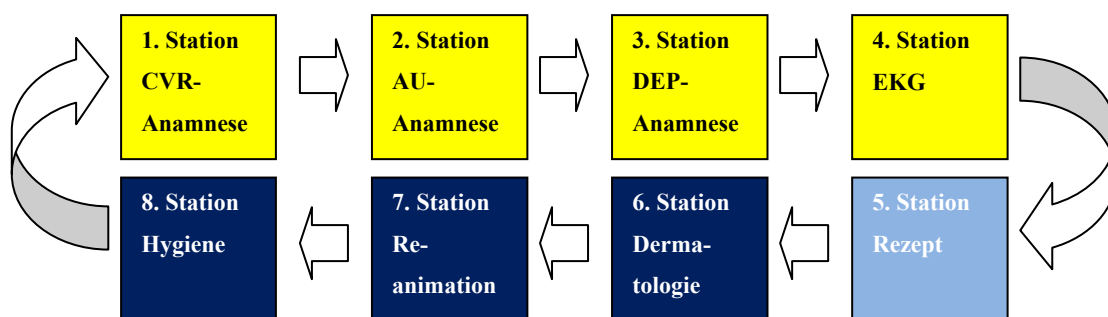
Seit dem SS 2004 findet an der Universität Göttingen die OSCE im Anschluss an den Kurs „Ärztliche Basisfertigkeiten“ für das gesamte erste klinische Semester statt. Die OSCE des WS 2004/05 hatte die gleiche Zusammensetzung der Stationen sowie den identischen Ablauf wie die OSCE des SS 2004. Studierende erzielten im Regelfall keine besseren Ergebnisse, wenn sich die Stationen innerhalb eines Jahres wiederholen und die Prüfungsthemen durch den studentischen Austausch bereits vor der Prüfung bekannt sind (Niehaus et al. 1996). Es sollten jedoch nicht über mehrere Jahre die gleichen Stationen verwendet werden, da es sonst auf Grund von Kenntnissen speziellerer Prüfungsinhalte einen Effekt in Form von besseren Endergebnissen für die Studierenden gäbe (Cohen R et al. 1993). Studierende, die innerhalb eines Jahres die gleiche OSCE durchliefen, zeigten in der Diagnosestellung eine bessere Leistung, aber in der Anamneseerhebung gab es keine signifikanten Unterschiede zu den vorherigen Ergebnissen (Petruša et al. 1990).

Der Themenkomplex der Stationen wurde vorab an den letzten Kurstagen und im Internet bekannt gegeben. Die Bekanntgabe der Themen dürfte keinen Effekt auf die Ergebnisse haben (Wilkinson et al. 2003a). Ein Informationsaustausch unter den Studierenden verschiedener Gruppen mit späteren Prüfungszeiten lässt sich sowieso nicht vermeiden. Rutala et al. (1991) hatten trotz eines Austausches zwischen den ersten und letzten OSCE-Gruppen keine signifikanten Ergebnisveränderungen festgestellt.

Die OSCE setzte sich aus acht interdisziplinären Stationen von den Abteilungen Allgemeinmedizin, Anästhesiologie/Notfallmedizin, Arbeits- und Sozialmedizin, Dermatologie, Hygiene und Psychosomatik zusammen. Es gab drei Anamnesestationen mit SP: bei kardiovaskulärem Risiko, Arbeitsunfähigkeit und Depression. Des Weiteren wurde an den fünf praktischen Stationen das Anlegen und Auswerten eines EKGs, das Ausfüllen eines Rezeptes mit Hilfe der Roten Liste, die Durchführung einer Reanimation sowie das Beschreiben von Hauteffloreszenzen und das hygienische Legen einer Venenverweilkanüle geprüft.

Für jede Station waren fünf Minuten und für den Stationswechsel eine Minute vorgesehen. Ein Pfiff mit einer Trillerpfeife signalisierte den Anfang und das Ende der Stationszeit; die letzte Minute wurde mit lauter Stimme angekündigt.

Im WS 2004/05 wurden an vier von acht Stationen neben den Dozenten auch studentische Tutoren als zweite Prüfer eingesetzt; die schriftliche Rezeptstation wurde allein von den Tutoren betreut (Abb. 3). Die OSCE ist eine summative Prüfung; es wurden von den Prüfern keine Hilfestellungen gegeben und die Leistung der Studierenden wurde auch nicht abschließend kommentiert. Dies lässt sich jedoch nicht an allen Stationen der OSCE vermeiden; die Dermatologie-, Reanimations- und Hygienestation haben einen formativen Charakter und wurden deshalb nur von Lehrärzten der entsprechenden Abteilung geleitet (Abb.3).



■ Stationen des Basisfertigkeitenkurses, Dozenten und studentische Tutoren prüfen

■ Station des Basisfertigkeitenkurses, studentische Tutoren prüfen

■ Stationen eines Semesterkurses, Lehrärzte der entsprechenden Abteilung prüfen

Abbildung 3: Aufbau der OSCE im WS 2004/05

An allen Stationen wurden zur Bewertung maschinenlesbare Checklisten mit Einzelitems verwendet (**Anhang 8.1 – 8.8**). Bis auf die Rezeptstation wurden zusätzlich Glo-

balbeurteilungen abgegeben (**Anhang 8.1 - 8.7, jeweils am Ende der Checkliste**). Die Gesamtnoten pro Station setzten sich aus jeweils 50% der Checklisten- und Globalbeurteilung zusammen; an der Rezeptstation ergab sich die Gesamtnote nur aus der Checklistenbewertung. Den Studierenden war bekannt, dass die Tutoren als zweite Prüfer an den Stationen des Kurses „Ärztliche Basisfertigkeiten“ bewerteten, aber ihre Noten nicht berücksichtigt wurden.

Die Checklisten wurden - basierend auf den Erfahrungen der Pilot-OSCE - bereits vor der OSCE des SS 2004 überarbeitet und präzisiert. Um das Ausfüllen zu vereinfachen und Interpretationsspielraum zu minimieren, wurden die Items dichotomisiert (**Anhang ab 8.1**). An der Depressionsstation und der Dermatologiestation kam eine Likert-Skalierung für jedes Item zum Einsatz, wobei eine höhere Punktzahl einer besseren Leistung entsprach (**Anhang 8.3 und 8.5**). Außerdem wurden die Instruktionen für die Studierenden an den Stationen gekürzt, um Verständnisschwierigkeiten zu vermeiden (**Anhang 8.9 - 8.15**).

Die OSCE wurde in der Osthalle des Göttinger Universitätsklinikums an vier Nachmittagen durchgeführt. Um die Auswertung zu erleichtern, trugen die maschinenlesbaren Bewertungsbögen bereits den Namen und die Matrikelnummer des Studierenden (**Anhang ab 8.1**). Jeder Prüfer erhielt eine Nummer, die auf den Checklisten eingetragen wurde, damit die Bewertungen anschließend den Prüfern zugeordnet werden konnten. Ein Prüfer durfte maximal vier Runden hintereinander bewerten und wurde danach abgelöst.

Im WS 2004/05 nahmen 213 Studierende an der OSCE teil. Da acht Studierende gleichzeitig geprüft werden konnten, bestand die OSCE aus 27 Runden zu jeweils ca. 47 Minuten. Die Studierenden sollten während der Prüfung weiße Arztkittel tragen, an denen mittels Klebestreifen ihre Nachnamen befestigt wurden. Es durften weder Notizen während der Anamnese angefertigt, noch an den übrigen Stationen Hilfsmittel benutzt werden; nur an der EKG-Station konnten die Studierenden ein EKG-Lineal für die Interpretation verwenden.

3.3 Studentische und ärztliche Prüfer

Die studentischen Prüfer waren Studierende ab dem siebten Semester aufwärts, die als Tutoren des Basisfertigkeitenkurses von den Abteilungen Allgemeinmedizin und Psychosomatik eingestellt worden waren. Die ärztlichen Prüfer waren die Kursdozenten der beiden Abteilungen und erfahrene Hausärzte, die als Lehrärzte in der Abteilung Allgemeinmedizin tätig sind. Sie wurden zusammen mit den studentischen Tutoren an den Stationen des Basisfertigkeitenkurses eingesetzt. An den übrigen Stationen bewerteten nur die Ärzte der entsprechenden Abteilungen.

Zwei Wochen vor der OSCE fand ein Prüfertraining mit Videobeispielen statt. Es wurden vorab die kardiovaskuläre Risikoanamnese und Arbeitsunfähigkeitsanamnese mit verschiedenen Leistungsniveaus von Lehrärzten und SP dargestellt und gefilmt. Anschließend beurteilten fünf Lehrärzte die Videos und erstellten eine „Masterauswertung“ für das Prüfertraining. Die Checklistenbeurteilungen wurden gemeinsam überprüft und Unklarheiten in der Formulierung verbessert.

Für das Prüfertraining standen für die beiden dargestellten Stationen jeweils fünf Video-Anamnesen mit den Noten von 1 bis 5 zur Verfügung. Alle Prüfer, die an den Anamnestationen eingesetzt werden sollten, nahmen an dem Training teil. Zu Beginn wurde kurz die Prüfung „OSCE“ vorgestellt. Die anwesenden Prüfer wurden darüber informiert, dass an vier der acht OSCE-Stationen ein Arzt und ein Tutor zusammen als Prüfer eingesetzt werden und gebeten, keine Absprachen über die Leistungsbewertung zu machen. Die Prüfer erhielten eine Erläuterung der Checklisten und bewerteten daraufhin die Videoanamnesen und wurden dann nach Schwierigkeiten in der Handhabung der Bewertungsbögen und nach Verständnisproblemen gefragt. Danach wurden die Item-Bewertungen nacheinander im Plenum verglichen und diskutiert. Anschließend wurden noch einmal die Kriterien für die Globalbeurteilung der Anamnese präzisiert, da laut Hodges and McIlroy (2003) die Validität durch präzise Kriterien steigt. Von jeder Anamnese wurden insgesamt drei bis vier Varianten vorgeführt, damit die Prüfer ein Gefühl für die Checkliste entwickeln und die Unterschiede in den Leistungen des dargestellten Arztes erkennen konnten. Die übrigen Stationen des Basisfertigkeitenkurses wurden nur anhand ihrer Checklisten erklärt. Wir haben uns bei diesem Vorgehen an Doig et al. (2000) orientiert.

3.4 Simulationspatienten

Die SP waren Freiwillige, die sich durch ein Zeitungsinserat oder Aushänge bei niedergelassenen Ärzten und im Klinikum angesprochen fühlten, sich gegen ein Honorar an der medizinischen Ausbildung zu beteiligen. Sie wurden während des Semesters in den Kursen und an den drei Anamnesestationen der OSCE eingesetzt.

Die SP wurden vor der OSCE für eine standardisierte Darstellung der Beschwerden bei der Anamneseerhebung trainiert, um eine einheitliche Prüfungssituation für jeden Studierenden zu schaffen (orientiert an Falck-Ytter und Stiegler 1993, Newble et al. 1980).

3.5 OSCE-Teilnehmer

Ausländische und deutsche Studierende wurden anhand des Namens und des Geburtsortes identifiziert (Vor- und Nachname hatten erste Priorität). Somit waren Studierende mit eindeutig ausländischem Namen „ausländisch“, auch wenn sich der Geburtsort in Deutschland befand.

Das Geschlecht der Studierenden wurde entweder durch eindeutigen Vornamen oder - in unklaren Fällen - mit Hilfe des Studierendensekretariats der Georg-August-Universität bestimmt.

3.6 Evaluation

Die Inhalte der Evaluation wurden im Voraus mit Lehrärzten der Abteilung Allgemeinmedizin besprochen und festgelegt, um spezifische und klar verständliche Fragen stellen zu können. Als Vorlagen dienten die Evaluationen von Pierre et al. (2004) und Weißer (2000). Die Antwortmöglichkeiten waren drei- bis vierstufig oder fragespezifisch; Freitext war zusätzlich möglich (orientiert an Duffield and Spencer 2002). Durch den Freitext sollten die Studierenden Möglichkeiten haben, Vorschläge und Meinungen zum Verfahren zu äußern.

Die Evaluation bestand aus drei Abschnitten (**Anhang 8.16**).

1. Die Akzeptanz der OSCE befasste sich mit allgemeinen Fragen zu der Prüfung. In einem ersten Teil, bestehend aus neun Fragen, wurden Informationen über das direkte Erleben der OSCE in Form einer Likert-Skalierung gewonnen. In einem zweiten Teil folgten Fragen zu dem Niveau der OSCE, früherer Teilnahme an

einer OSCE und Länge der Vorbereitungszeit. Abschließend ging es um den Beliebtheitsgrad der OSCE im Vergleich zu anderen Prüfungsformen.

2. Zwei Fragen zur Bewertungsweise und Objektivität (mit jeweils drei Antwortmöglichkeiten) zielten auf den Vergleich zwischen Tutoren und Ärzten.
3. Zwei Fragen galten dem Einsatz von studentischen Prüfern (vierstufiges Antwortschema).

Da die Ergebnisse einer freiwilligen Evaluation oftmals unspezifisch, unsystematisch und in ihrer Wertung negativ sind (Morris 1976), sollte jeder Prüfungsteilnehmer am Ende der OSCE einen Evaluationsbogen ausfüllen. Die Evaluation war anonym.

3.7 Statistische Auswertungen

Die Abteilung Medizinische Informatik der Georg-August-Universität erstellte die maschinenlesbaren Bewertungsbögen. Diese wurden nach der OSCE eingescannt, wobei Fehler beim Ausfüllen der Bewertungsbögen angezeigt wurden. Die auf den Checklisten zu markierenden Felder mussten zu mindestens 50% ausgefüllt und es durfte nur ein Feld pro Zeile angestrichen sein. In den seltenen Fällen, wo die Markierung nicht eindeutig einer Bewertungskategorie zugeordnet werden konnte, wurde die bestmögliche Bewertung angenommen. Die Daten wurden in einer Tabelle zusammengeführt und die statistischen Auswertungen mit SAS, Version 9.1.3 durchgeführt.

Die Evaluation wurde manuell ausgewertet, in Form absoluter und relativer Häufigkeiten (orientiert an Morris 1976).

3.7.1 Vergleich zwischen den Prüfern

Die Durchschnittsnoten (mit Standardabweichung) für die Checklistennoten, Globalbeurteilungen und Gesamtnoten von Tutoren und Ärzten wurden an den drei Anamnese-Stationen und der EKG-Station verglichen. Es wurde die rohe Prüfer-Übereinstimmung (Prozentangaben) und das Kappa (mit 95%-Konfidenzintervall) ermittelt. Außerdem wurde an den vier Stationen für jedes Item die Übereinstimmung zwischen den Tutoren und Ärzten errechnet.

Die Übereinstimmung zwischen zwei Prüfern wird mit dem Kappa-Test beurteilt (Viera and Garrett 2005). Der Übereinstimmungsgrad der verglichenen Bewertungen wird dabei auf einer Scala von 0 bis 1 (0 bis 100%) angegeben (Cohen JA 1960). Ein niedriger Wert bedeutet eine geringe Übereinstimmung zwischen den Prüfern und markiert Items bzw. Stationen, deren Genauigkeit und Verständlichkeit überprüft werden sollten (O’Conner and McGraw 1997). Zur Interpretation der Kappa-Werte wird üblicherweise und in dieser Arbeit die von Altmann (1991) modifizierte Einteilung nach Landis und Koch (1977) verwendet (Tab. 2).

Tabelle 2: Übersicht der Klasseneinteilung der Kappa-Werte (nach Altmann 1991)

<i>Kappa- Wert</i>	<i>Intensität der Übereinstimmung</i>
≤ 0,20	schlecht
0,21- 0,40	gering
0,41- 0,60	mäßig
0,61- 0,80	gut
0,81- 1,00	sehr gut

3.7.2 Vergleich zwischen den Studierenden

Die Durchschnittsnoten der deutschen und ausländischen Studierenden und der männlichen und weiblichen Studierenden wurden zunächst mit dem Shapiro-Wilk-Test auf eine Normalverteilung hin überprüft (Tab. 3). Da keine Normalverteilung vorlag, erfolgten die Vergleiche mit dem Wilcoxon-Rangtest auf dem 5%- Signifikanzniveau.

Tabelle 3: Überprüfung auf Normalverteilungen

<i>Überprüfung auf Normalverteilung der...</i>	<i>p-Wert des Shapiro-Wilk-Tests</i>
Deutschen und Ausländer	<0.0001
Männer und Frauen	<0.0001

4 Ergebnisse

4.1 Bewertungen der Tutoren und Ärzte im Vergleich

In der OSCE des WS 2004/05 prüften 20 studentische und 42 ärztliche Prüfer. Die studentischen Tutoren wurden an vier Stationen zusammen mit den Ärzten als Prüfer eingeteilt: bei den drei Anamneseerhebungen und beim Anlegen und Auswerten eines EKGs. Im Folgenden werden zunächst die Checklistennoten (4.1.1), Globalbeurteilungen (4.1.2) und Gesamtnoten (4.1.3) von studentischen und ärztlichen Prüfern verglichen und danach die Einzelitems analysiert.

4.1.1 Vergleich der Checklistennoten

Die Checklistenbewertungen der Ärzte fielen - mit Ausnahme der AU-Station - im Durchschnitt geringfügig schlechter als die der studentischen Tutoren aus (Abb. 4). Die Übereinstimmung der Prüfer war bei der CVR-Station mit einem Kappa von 0,64 am höchsten. An den übrigen Stationen war die Übereinstimmung mittelmäßig (Tab. 4).

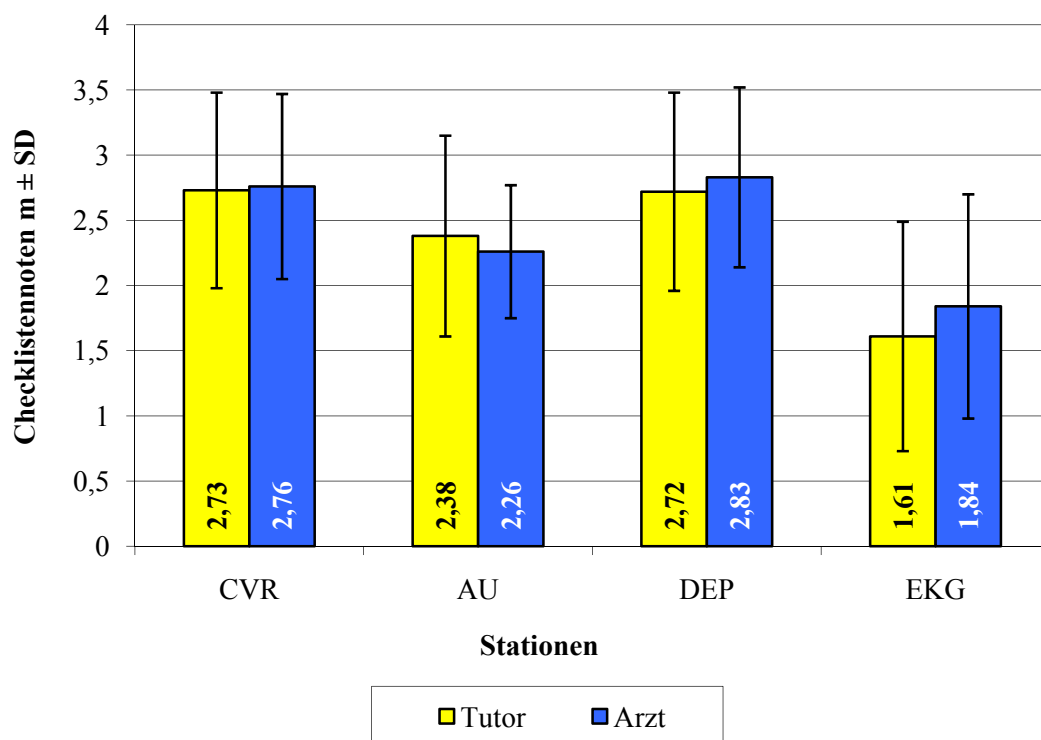


Abbildung 4: Checklistennoten der Tutoren und Ärzte

Tabelle 4: Prüfer-Übereinstimmung bei den Checklistennoten

Station	Rohe Übereinstimmung	Kappa
	(%)	K (95%-KI)
CVR	75,1	0,64 (0,56-0,73)
AU	73,7	0,42 (0,30-0,54)
Depression	63,9	0,48 (0,37-0,59)
EKG	56,8	0,39 (0,29-0,50)

4.1.2 Vergleich der Globalbeurteilungen

Im Durchschnitt fielen die Globalbeurteilungen der Ärzte - wiederum bis auf die AU-Station - etwas schlechter als die der studentischen Tutoren aus (Abb. 5). Die Prüfer-Übereinstimmung war mit Werten zwischen 0,46 und 0,57 gut (Tab. 5).

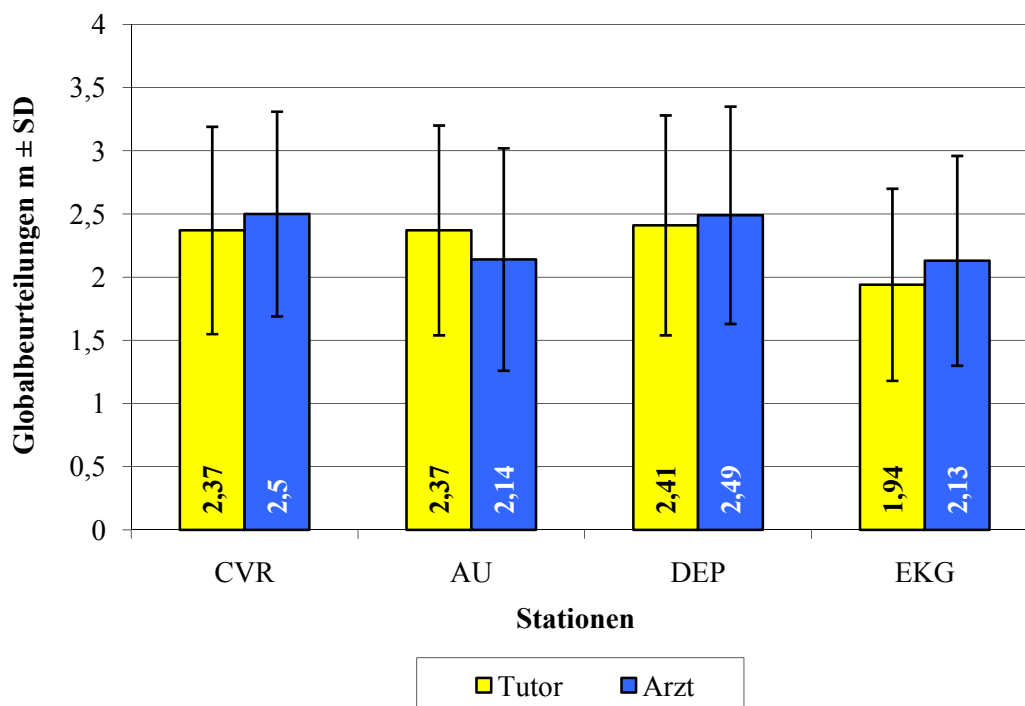


Abbildung 5: Globalbeurteilungen der Tutoren und Ärzte

Tabelle 5: Prüfer-Übereinstimmung bei den Globalbeurteilungen

Station	Rohe Übereinstimmung	Kappa
	(%)	K (95%-KI)
CVR	61,5	0,57 (0,48-0,65)
AU	61,0	0,57 (0,48-0,65)
Depression	53,1	0,48 (0,40-0,59)
EKG	55,0	0,46 (0,37-0,55)

4.1.3 Vergleich der Gesamtnoten

Die Gesamtnoten der Ärzte fielen an fast allen Stationen im Durchschnitt etwas schlechter aus als diejenigen der studentischen Tutoren, maximal um 0,22 am EKG (Abb. 6). Die Prüfer- Übereinstimmung war mit Kappa-Werten zwischen 0,41 und 0,64 wiederum gut (Tab. 6).

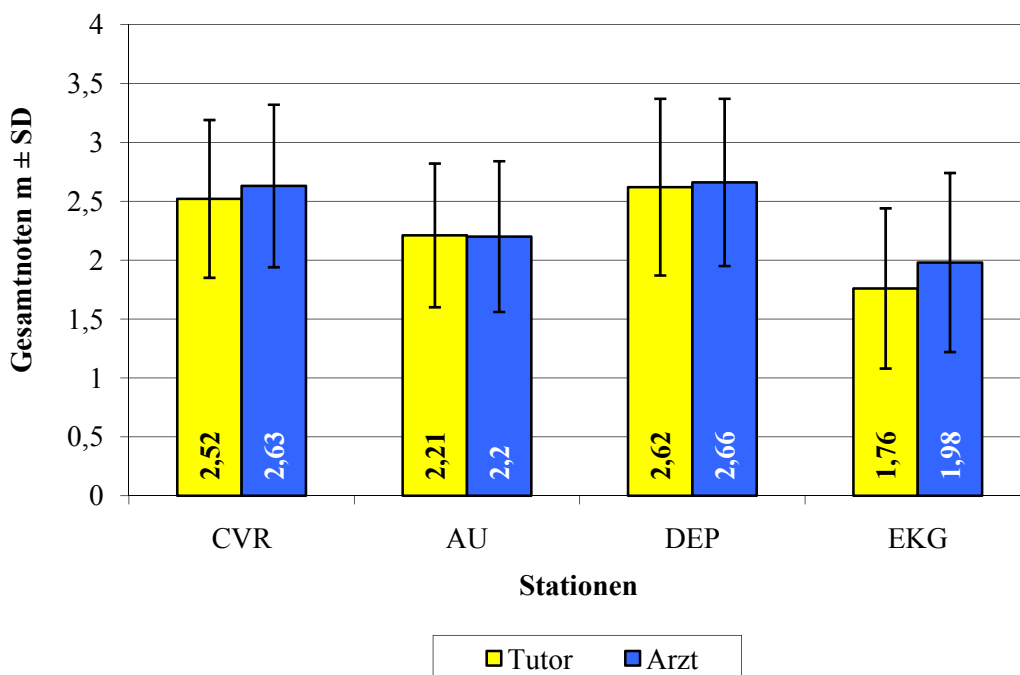


Abbildung 6: Gesamtnoten der Tutoren und Ärzte

Tabelle 6: Prüfer-Übereinstimmung bei den Gesamtnoten

Station	Rohe Übereinstimmung	Kappa
	(%)	K (95%-KI)
CVR	54,9	0,64 (0,57-0,71)
AU	49,3	0,51 (0,43-0,59)
Depression	37,1	0,49 (0,41-0,57)
EKG	36,6	0,41 (0,33-0,50)

4.1.4 Vergleich der Checklisten-, Global- und Gesamtnoten

Im Folgenden wurden die Checklisten-, Global- und Gesamtnoten der vier Stationen zu einer Note zusammengefasst (Abb. 7). Die Tutoren bewerteten im Durchschnitt etwas besser als die Ärzte, maximal lag die Differenz aber bei 0,13. Man sieht auch sehr deutlich, dass bei den Globalbeurteilungen die Gesamtnoten für die Studierenden - sowohl von Ärzten als auch studentischen Tutoren - besser ausfielen. Der dritte Säulenkomplex zeigt die Durchschnittsnote aller Gesamtnoten: wenn Tutoren bewerteten, erzielten die Studierenden an den vier Stationen eine durchschnittliche Note von 2,26 und bei ärztlichen Prüfern von 2,37. Die Prüferübereinstimmung war bei der Summe aller Gesamtnoten relativ gut, mit einem Kappa von 0,58 (Tab. 7).

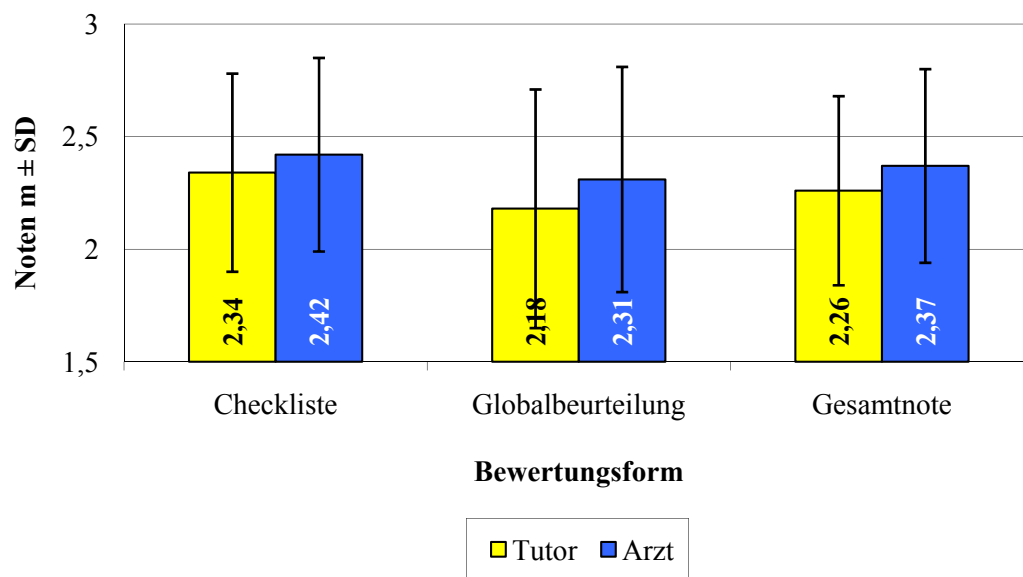


Abbildung 7: Checklisten-, Global- und Gesamtnoten zusammengefasst

Tabelle 7: Prüfer-Übereinstimmung bei den Stationszusammenfassungen

<i>Bewertungsform</i>	<i>Rohe Übereinstimmung</i>	<i>Kappa</i>
	<i>(%)</i>	<i>K (95%-KI)</i>
Checkliste	33,0	0,48 (0,41-0,55)
Globalbeurteilung	17,1	0,30 (0,24-0,37)
Gesamtnote	24,4	0,58 (0,53-0,64)

4.1.5 Einzelitemanalyse der vier Stationen

Die Einzelitemanalyse ermöglichte einen genaueren Vergleich zwischen den studentischen und ärztlichen Prüfern. Items mit einer geringen Übereinstimmung sprechen für eine unterschiedliche Interpretation durch die Prüfer und müssten vor der nächsten OSCE nochmals präzisiert werden (Tab. 8 ff.).

Bei der CVR-Station war die Übereinstimmung zwischen studentischen und ärztlichen Prüfern im allgemeinen Teil der Checkliste im mittleren Bereich (Tab. 8). Die „Begrüßung“ zeigte eine gute Übereinstimmung mit einem Kappa von 0,55 (rohe Übereinstimmung ca. 90%). Beim „Patientenkontakt“ dagegen zeigten sich deutliche Differenzen im Urteil von studentischen und ärztlichen Prüfern mit einem Kappa von nur 0,3 (rohe Übereinstimmung ca. 70%). Bei den themenbezogenen Items war die Übereinstimmung sehr gut mit Kappa-Werten von überwiegend $>0,8$, bei dem „Nikotinkonsum“-Item sogar 1,0 (rohe Übereinstimmungen $>90\%$).

Tabelle 8: Item-Übereinstimmungen der Tutoren und Ärzte an der CVR-Station⁴

<i>CVR- Items</i>	<i>Rohe Übereinstimmung</i>	<i>Kappa</i>
	<i>n (%)</i>	<i>K (95%-KI)</i>
Begrüßung	188 (90,4)	0,55 (0,36-0,73)
Sprache	161 (77,4)	0,38 (0,24-0,52)
Fachwörter	172 (82,7)	0,43 (0,28-0,57)
Patientenkontakt	145 (69,7)	0,31 (0,18-0,43)
RR-Wert	199 (96,1)	0,89 (0,82-0,97)
Frühere Messung	182 (87,5)	0,75 (0,65-0,84)
Diabetes	201 (96,6)	0,91 (0,84-0,97)
Nierenerkrankung	200 (96,2)	0,82 (0,71-0,94)
Familienanamnese	190 (91,4)	0,55 (0,37-0,73)
vor dem 60. Lj	188 (90,4)	0,70 (0,57-0,82)
Nikotinkonsum	208 (100)	1,00 (1,00-1,00)
Dauer Nikotin	196 (94,2)	0,87 (0,80-0,94)
Menge Nikotin	203 (97,6)	0,87 (0,76-0,98)
Abstinenzversuche	193 (92,8)	0,85 (0,78-0,92)
Alkoholkonsum	207 (99,5)	0,99 (0,97-1,00)
Menge Alkohol	196 (94,2)	0,88 (0,81-0,95)
Körperliche Aktivität	199 (95,7)	0,91 (0,86-0,97)
Cholesterinmessung	206 (99,0)	0,96 (0,91-1,00)
Cholesterinwert	201 (96,6)	0,90 (0,83-0,97)
Alter der Patientin	199 (95,7)	0,91 (0,86-0,97)
Beruf	204 (98,1)	0,94 (0,89-1,00)
Familienstand	195 (93,8)	0,87 (0,81-0,94)
Stressoren	192 (92,3)	0,80 (0,71-0,89)
Medikamente	203 (97,6)	0,94 (0,89-0,99)
Allergie	197 (94,7)	0,85 (0,77-0,94)

⁴ Es fehlen fünf studentische Wertungen

An der AU-Station (Tab. 9) war die Übereinstimmung zwischen den studentischen und ärztlichen Prüfern bei den allgemeineren Aspekten wie z.B. der Bewertung der Sprache vergleichsweise gering (Kappa von 0,26). Währenddessen zeigten die restlichen themenspezifischen AU-Items überwiegend eine Übereinstimmung von Kappa 0,52 bis 0,88 auf (rohe Übereinstimmung bei allen Items meist >80%).

Tabelle 9: Item-Übereinstimmungen der Tutoren und Ärzte an der AU-Station⁵

<i>AU-Items</i>	<i>Rohe Übereinstimmung</i>	<i>Kappa</i>
	<i>n (%)</i>	<i>K (95%-KI)</i>
Begrüßung	183 (90,2)	0,36 (0,11-0,60)
Sprache	175 (86,2)	0,26 (0,70-0,44)
Fachbegriffe	198 (97,5)	0,28 (-0,16-0,72)
Patientenkontakt	160 (78,8)	0,36 (0,20-0,51)
Blick auf den Unterarm	192 (94,6)	0,87 (0,79-0,94)
Schmerzursache	190 (93,6)	0,61 (0,42-0,81)
Schmerzlokalisierung	174 (85,7)	0,55 (0,40-0,69)
Schmerzausstrahlung	170 (83,7)	0,65 (0,54-0,76)
Schmerzdauer	179 (88,2)	0,23 (0,02-0,44)
Funktionseinschränkung	160 (78,8)	0,52 (0,40-0,65)
Händigkeit	189 (93,1)	0,85 (0,77-0,93)
Aktuelle Tätigkeit	202 (99,5)	0,89 (0,67-1,00)
Besondere Belastung	154 (75,9)	0,34 (0,18-0,50)
Spezielle Beanspruchung	158 (77,8)	0,54 (0,42-0,66)
Arbeitsatmosphäre	174(85,7)	0,68 (0,58-0,79)
Andere Belastungen	176 (86,7)	0,69 (0,58-0,80)
Keine AU	187 (92,1)	0,60 (0,42-0,77)
AU für 2-3 Tage	186 (91,6)	0,70 (0,57-0,83)
Begründung	169 (83,3)	0,28 (0,11-0,46)

⁵ Es fehlen zehn studentische Wertungen

Bei der Depressionsstation (Tab. 10) lag die Übereinstimmung im geringen bis mittleren Bereich bis maximal 0,48 (rohe Übereinstimmungen 50-60%).

Tabelle 10: Item-Übereinstimmungen der Tutoren und Ärzte an der DEP-Station⁶

<i>DEP-Items</i>	<i>Rohe Übereinstimmung</i>	<i>Kappa</i>
	<i>n (%)</i>	<i>K (95%-KI)</i>
Kommunikation	128 (62,1)	0,45 (0,35-0,56)
Aktuelle Symptomatik	97 (47,1)	0,31 (0,21-0,41)
Depressive Symptomatik	105 (51,0)	0,48 (0,39-0,56)
Auslösende Situation	113 (54,9)	0,48 (0,39-0,57)

Die „Begrüßung“ an der EKG-Station (Tab. 11) ergab eine sehr geringe Übereinstimmung zwischen studentischen und ärztlichen Prüfern mit einem Kappa von nur 0,22 (rohe Übereinstimmung ca. 68%). Die besten Kappa-Werte lagen bei der Bestimmung des „Lagetypen“ (0,74) und der „Frequenz“ (0,62) vor (rohe Übereinstimmungen >90%).

Tabelle 11: Item-Übereinstimmungen der Tutoren und Ärzte an der EKG-Station⁷

<i>EKG-Items</i>	<i>Rohe Übereinstimmung</i>	<i>Kappa</i>
	<i>n (%)</i>	<i>K (95%-KI)</i>
Begrüßung	142 (67,6)	0,22 (0,1-0,36)
Extremitätenelektroden	206 (98,1)	0,59 (0,23-0,95)
Brustwandelektroden	194 (92,4)	0,44 (0,21-0,66)
Erklärt flüssig und verständlich	185 (88,1)	0,01 (-0,13-0,16)
Rhythmus	196 (93,3)	0,53 (0,31-0,74)
Frequenz	193 (91,9)	0,62 (0,46-0,79)
Lagetyp	196 (93,3)	0,74 (0,62-0,87)

⁶ Es fehlen sieben studentische Wertungen

⁷ Es fehlen drei studentische Wertungen

4.2 Deutsche und Ausländer im Vergleich

Von den 213 OSCE-Teilnehmern waren 26 Studierende (13%) Ausländer und 180 Studierende (87%) Deutsche. Sieben Studierende konnten nicht eindeutig zugeordnet werden und wurden für die Auswertungen außer Acht gelassen.

4.2.1 Vergleich der Checklistennoten

Die ausländischen Studierenden erhielten nur an der Hygienestation eine bessere durchschnittliche Bewertung als die deutschen Studierenden. An der EKG-Station schnitten sie im Vergleich zu den Deutschen am schlechtesten ab, besonders in der Bewertung der Ärzte, die hier die Deutschen im Durchschnitt um 0,74 besser beurteilten. Es lagen signifikante Unterschiede an der AU- (ärztliche Bewertung), EKG-, Dermatologie- und Reanimationsstation vor (Tab. 12).

Tabelle 12: Checklistennoten für deutsche und ausländische Studierende

Station; Prüfer	<i>Studierende; Checkliste</i>		Differenz	p-Wert
	<i>Deutsche</i>	<i>Ausländer</i>		
	<i>Mittelwert (±SD)</i>	<i>Mittelwert (±SD)</i>		
CVR; Arzt	2,73 (±0,64)	3,08 (±0,65)	0,35	0,16
CVR; Tutor	2,73 (±0,77)	2,79 (±0,58)	0,06	0,61
AU; Arzt	2,25 (±0,50)	2,37 (±0,57)	0,12	0,02
AU; Tutor	2,36 (±0,72)	2,54 (±1,10)	0,18	0,46
Depression; Arzt	2,79 (±0,71)	3,08 (±0,50)	0,29	0,07
Depression; Tutor	2,66 (±0,75)	3,17 (±0,71)	0,51	0,12
EKG; Arzt	1,76 (±0,77)	2,50 (±1,21)	0,74	0,002
EKG; Tutor	1,56 (±0,82)	2,08 (±1,13)	0,52	0,02
Dermatologie	2,49 (±0,81)	2,87 (±0,79)	0,38	0,0003
Reanimation	2,08 (±1,03)	2,58 (±1,17)	0,50	0,02
Hygiene	1,51 (±0,83)	1,45 (±0,65)	0,06	0,3

4.2.2 Vergleich der Globalbeurteilungen

Die ausländischen Studierenden erhielten von den studentischen und ärztlichen Prüfern an allen Stationen im Durchschnitt deutlich schlechtere Globalbeurteilungen als die deutschen Studierenden, am deutlichsten an der Dermatologiestation (Notenunterschied: 0,79), am geringsten an der Hygienestation (Notenunterschied: 0,14). Statistisch signifikant waren die Unterschiede an der AU- (ärztliche Bewertung), Depressions-, EKG-, Dermatologie-, Reanimations- und Hygienestation (Tab. 13).

Tabelle 13: Globalbeurteilungen für deutsche und ausländische Studierende

Station; Prüfer	Studierende; Globalbeurteilung		Differenz	p-Wert
	Deutsche	Ausländer		
	Mittelwert (\pm SD)	Mittelwert (\pm SD)		
CVR; Arzt	2,48 (\pm 0,82)	2,66 (\pm 0,76)	0,18	0,69
CVR; Tutor	2,33 (\pm 0,83)	2,65 (\pm 0,71)	0,32	0,28
AU; Arzt	2,09 (\pm 0,86)	2,50 (\pm 0,97)	0,41	0,0004
AU; Tutor	2,13 (\pm 0,83)	2,28 (\pm 0,84)	0,15	0,13
Depression; Arzt	2,42 (\pm 0,75)	3,04 (\pm 0,95)	0,62	0,05
Depression; Tutor	2,47 (\pm 0,86)	3,00 (\pm 0,85)	0,53	0,02
EKG; Arzt	2,07 (\pm 0,80)	2,66 (\pm 0,91)	0,59	0,004
EKG; Tutor	1,89 (\pm 0,72)	2,33 (\pm 0,96)	0,44	0,05
Dermatologie	2,25 (\pm 0,97)	3,04 (\pm 1,08)	0,79	0,0001
Reanimation	1,90 (\pm 0,88)	2,41 (\pm 1,05)	0,51	0,002
Hygiene	1,94 (\pm 0,97)	2,08 (\pm 0,82)	0,14	0,04

4.2.3 Vergleich der Gesamtnoten

Die ausländischen Studierenden erhielten deutlich schlechtere Gesamtnoten sowohl von den ärztlichen als auch studentischen Prüfern (Abb. 8 und 9). An der Rezeptstation lag die größte Notendifferenz (1,17) zugunsten der deutschen Studierenden vor. Statistisch signifikante Unterschiede gab es an der AU- (ärztliche Bewertung), EKG-, Depressions-, Dermatologie-, Reanimations- und Rezeptstation (*).

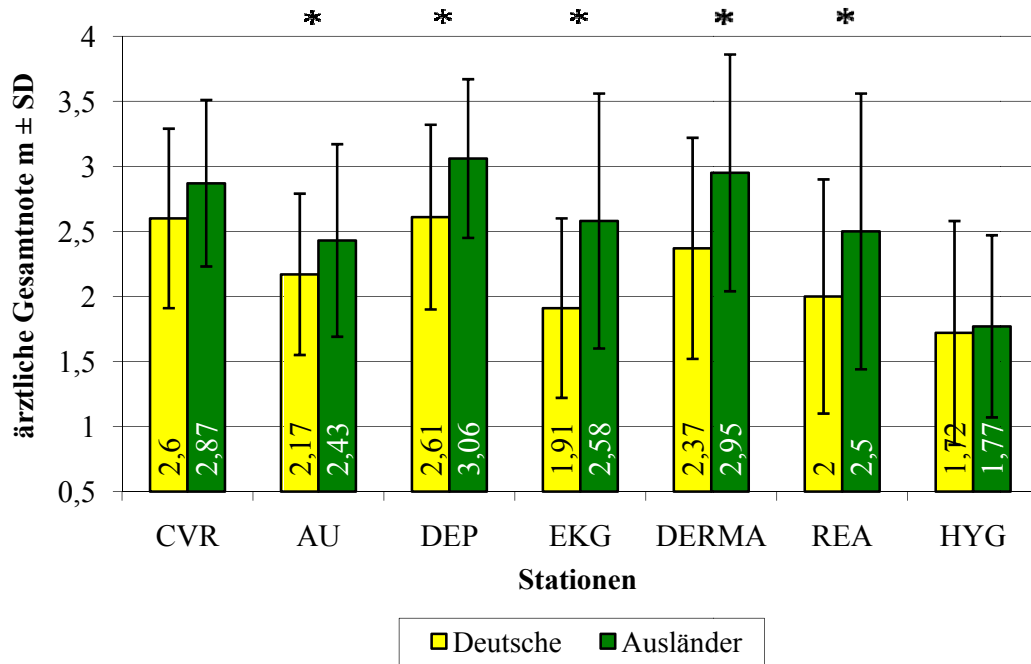


Abbildung 8: Arzt-Gesamtnoten für deutsche und ausländische Studierende

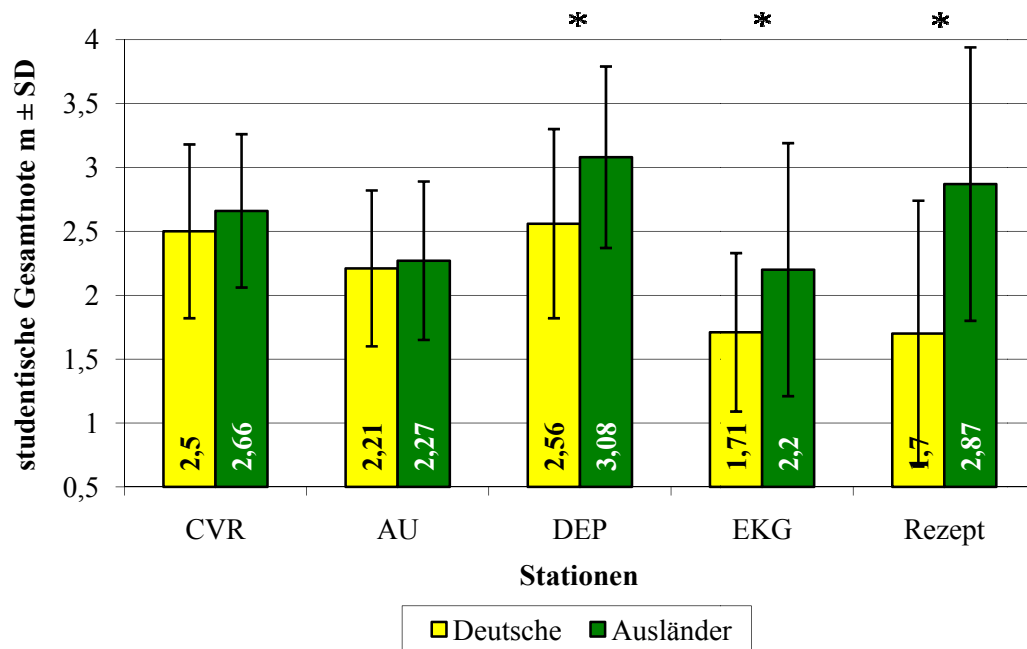


Abbildung 9: Tutoren-Gesamtnoten für deutsche und ausländische Studierende

4.3 Männliche und weibliche Studierende im Vergleich

Insgesamt waren von den 213 Studierenden in der OSCE 110 Männer (52%).

4.3.1 Vergleich der Checklistennoten

Die männlichen Studierenden erhielten von den Prüfern bis auf die Reanimationsstation im Durchschnitt etwas schlechtere Checklistennoten als die weiblichen Studierenden. Insgesamt waren die Notenunterschiede jedoch sehr gering (maximal 0,39) und nur an zwei Stationen (Depression (studentische Bewertung) und Hygiene) signifikant (Tab. 14).

Tabelle 14: Checklistennoten für männliche und weibliche Studierende

Station; Prüfer	Studierende; Checkliste		Differenz	p-Wert
	Männer	Frauen		
	Mittelwert (\pm SD)	Mittelwert (\pm SD)		
CVR; Arzt	2,95 (\pm 0,73)	2,56 (\pm 0,63)	0,39	0,2
CVR; Tutor	2,89 (\pm 0,80)	2,56 (\pm 0,66)	0,33	0,55
AU; Arzt	2,33 (\pm 0,54)	2,19 (\pm 0,46)	0,14	0,38
AU; Tutor	2,46 (\pm 0,82)	2,29 (\pm 0,70)	0,17	0,55
Depression; Arzt	2,91 (\pm 0,74)	2,73 (\pm 0,62)	0,18	0,1
Depression; Tutor	2,80 (\pm 0,76)	2,63 (\pm 0,75)	0,17	0,05
EKG; Arzt	1,91 (\pm 0,91)	1,75 (\pm 0,81)	0,16	0,19
EKG; Tutor	1,70 (\pm 0,95)	1,52 (\pm 0,79)	0,18	0,68
Dermatologie	2,60(\pm 0,77)	2,46 (\pm 0,84)	0,14	0,51
Reanimation	2,12 (\pm 1,09)	2,17 (\pm 1,03)	0,05	0,26
Hygiene	1,60 (\pm 0,91)	1,40 (\pm 0,69)	0,20	0,05

4.3.2 Vergleich der Globalbeurteilungen

Die Globalbeurteilungen der studentischen und ärztlichen Prüfer fielen an allen Stationen wiederum für die Männer im Durchschnitt etwas schlechter als für die Frauen aus. Es gab aber nur an der CVR-Station eine größere Notendifferenz (0,55 und 0,47) und nur einen statistisch signifikanten Unterschied an der Hygienestation (Tab. 15).

Tabelle 15: Globalbeurteilungen für männliche und weibliche Studierende

Station; Prüfer	Studierende; Globalbeurteilung		Differenz	p-Wert
	Männer	Frauen		
	Mittelwert (\pm SD)	Mittelwert (\pm SD)		
CVR; Arzt	2,77 (\pm 0,79)	2,22 (\pm 0,74)	0,55	0,11
CVR; Tutor	2,60 (\pm 0,8)	2,13 (\pm 0,78)	0,47	0,33
AU; Arzt	2,25 (\pm 0,91)	2,01 (\pm 0,85)	0,24	0,93
AU; Tutor	2,23 (\pm 0,85)	2,07 (\pm 0,81)	0,16	0,44
Depression; Arzt	2,60 (\pm 0,91)	2,36 (\pm 0,79)	0,24	0,41
Depression; Tutor	2,63 (\pm 0,88)	2,43 (\pm 0,86)	0,2	0,09
EKG; Arzt	2,22 (\pm 0,88)	2,03 (\pm 0,77)	0,19	0,95
EKG; Tutor	2,01 (\pm 0,81)	1,88 (\pm 0,70)	0,13	0,84
Dermatologie	2,42 (\pm 1,07)	2,25 (\pm 0,94)	0,17	0,59
Reanimation	2,01 (\pm 0,97)	1,97 (\pm 0,88)	0,04	0,27
Hygiene	2,11 (\pm 0,99)	1,81 (\pm 0,92)	0,3	0,02

4.3.3 Vergleich der Gesamtnoten

Die Männer erhielten von den Prüfern eine im Durchschnitt etwas schlechtere Bewertung als die Frauen, der Notenunterschied war an der CVR-Station mit 0,47 und 0,37 am größten (Abb. 10 und Abb. 11). Es gab nur an der Depressions- (studentische Bewertung) und Hygienestation einen signifikanten Unterschied (*).

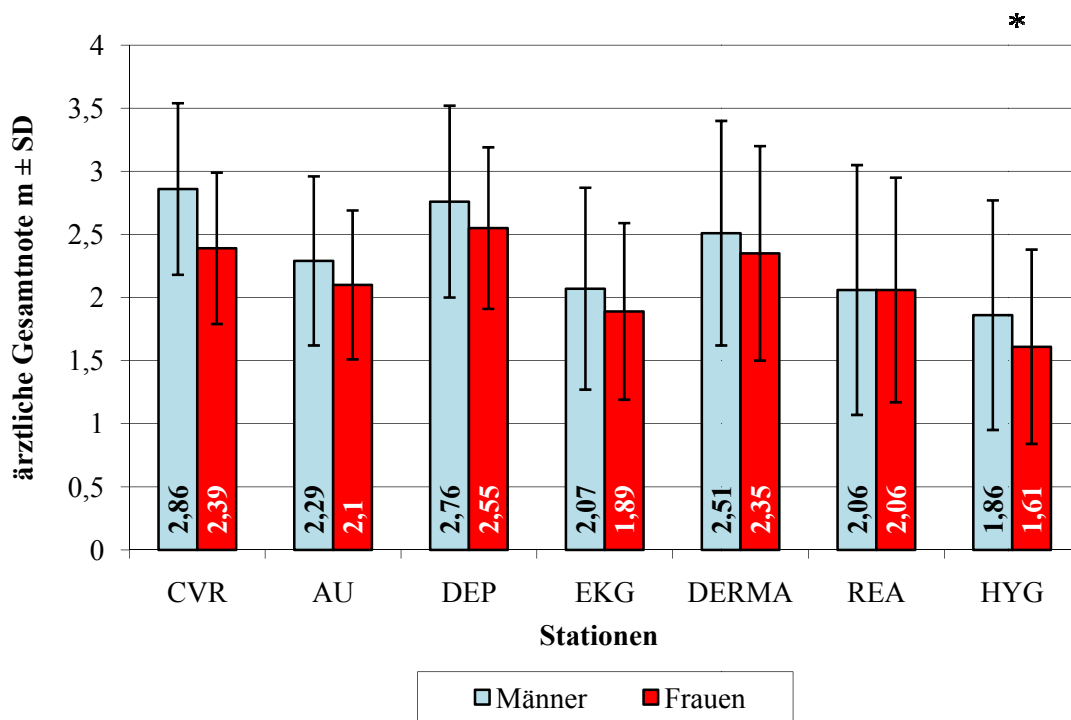


Abbildung 10: Arzt-Gesamtnoten für männliche und weibliche Studierende

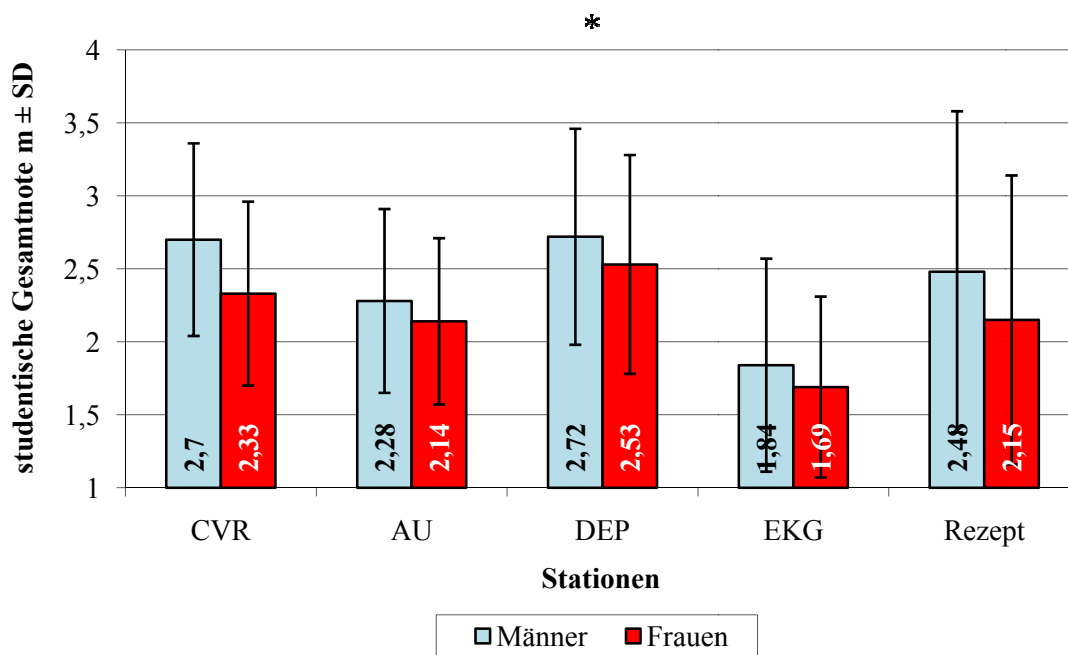


Abbildung 11: Tutoren-Gesamtnoten für männliche und weibliche Studierende

4.4 Evaluation der OSCE des WS 2004/05 aus studentischer Sicht

Nach der OSCE gaben 195 der 213 Studierenden (92%) ihren ausgefüllten Evaluationsbogen an die Prüfungsleitung zurück. Die Evaluation lässt sich in folgende Abschnitte gliedern: Akzeptanz der OSCE (4.4.1), Vergleich der Bewertungsweise von studentischen und ärztlichen Prüfern (4.4.2) sowie Einsatz von studentischen Prüfern (4.4.3).

4.4.1 Akzeptanz der OSCE

Die meisten Studierenden waren der Ansicht, dass mit der OSCE praktische Fertigkeiten (94%) und theoretisches Wissen (77%) überprüft werden könnten. Viele Studierende (82%) glaubten, dass sie ihre Leistung mit der neuen Prüfungsform einschätzen könnten und die OSCE demnach einen positiven Lerneffekt hätte. Es würden sogar mehrere Studierende (65%) die Kursthemen wiederholen. Nur wenige Prüflinge (15%) hatten keinen Spaß an der OSCE, während sich die Mehrheit (85%) positiv dazu äußerte. Einige Studierende (40%) fühlten sich durch die OSCE verunsichert und empfanden die Prüfungsatmosphäre als angespannt. Eine Minderheit (12%) schätzte die Prüfung als nicht objektiv ein (Tab.16).

Die Mehrheit (95% bei n=195) hatte das erste Mal an einer OSCE teilgenommen und von fast allen Studierenden (97% bei n=188) wurde das Niveau der OSCE als gerade richtig angesehen; nur 2 Studierende haben es als zu hoch empfunden. Es bereiteten sich nur wenige Studierende gar nicht (2,5%) oder bis zu einer Stunde (12%) auf die OSCE vor, während der Großteil der Studierenden die Prüfungsthemen für 1-3 Stunden (56%) und länger (29,5%) wiederholte (n=195). In der Beliebtheitsfrage schnitt die OSCE als beste Prüfungsform (41%) ab. Schriftliche Prüfungen sowie Hausarbeiten oder Referate waren mit 6% und 3% am wenigsten beliebt (n=190; Abb. 12).

Tabelle 16: Akzeptanz der OSCE

Dimensionen	n	Antwortmöglichkeiten			
		Stimme zu	Stimme teilweise zu	Stimme eher nicht zu	Stimme nicht zu
		n (%)	n (%)	n (%)	n (%)
Praktische Fertigkeiten	195	124 (63,6)	58 (29,7)	8 (4,1)	5 (2,6)
Theoretisches Wissen	195	59 (30,3)	91 (46,7)	36 (18,5)	9 (4,6)
Eigene Leistung	194	73 (37,6)	86 (44,3)	30 (15,5)	5 (2,6)
Positiver Lerneffekt	195	101 (51,8)	60 (30,8)	23 (11,8)	11 (5,6)
Wiederholen der Themen	194	66 (34,0)	60 (30,9)	46 (23,7)	22 (11,3)
Spaß	195	87 (44,6)	78 (40,0)	24 (12,3)	6 (3,1)
Verunsicherung	193	16 (8,3)	62 (32,1)	64 (33,2)	51 (26,4)
Angespannte Atmosphäre	194	23 (11,9)	55 (28,4)	60 (30,9)	56 (28,9)
Fair / Objektiv	191	70 (36,7)	99 (51,8)	13 (6,8)	9 (4,7)

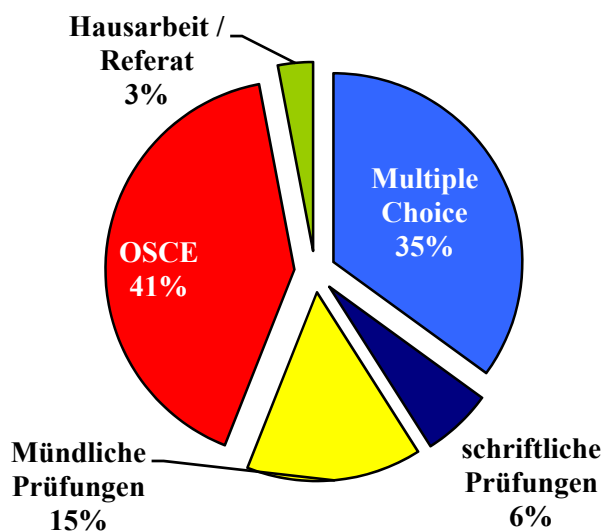


Abbildung 12: Beliebteste Prüfungsform

4.4.2 Vergleich der Bewertungsweise von Tutoren und Ärzten

Die Mehrheit der Studierenden (64%) war der Ansicht, dass es keinen Unterschied zwischen studentischen und ärztlichen Prüfern gäbe. Nur sehr wenige Studierende (9%) glaubten, dass sie durch die studentischen Prüfer eher schlechter bewertet wurden, 27% eher besser (n=177).

Die Mehrheit (69%) vertrat die Ansicht, dass die studentischen Prüfer genauso objektiv bewerten würden wie die ärztlichen Prüfer. Es glaubte jeweils nur eine Minderheit der Studierenden, dass die Tutoren weniger objektiv (14%) oder objektiver (17%) bewerten würden (n=173).

4.4.3 Einsatz von studentischen Prüfern

Die Mehrheit der Studierenden (94% bei n=158) stimmte ganz oder teilweise der Ansicht zu, dass die studentischen Prüfer angemessen bewerten würden. Nur wenige Studierende fühlten sich durch die Tutoren in der OSCE nicht angemessen beurteilt (Abb. 13).

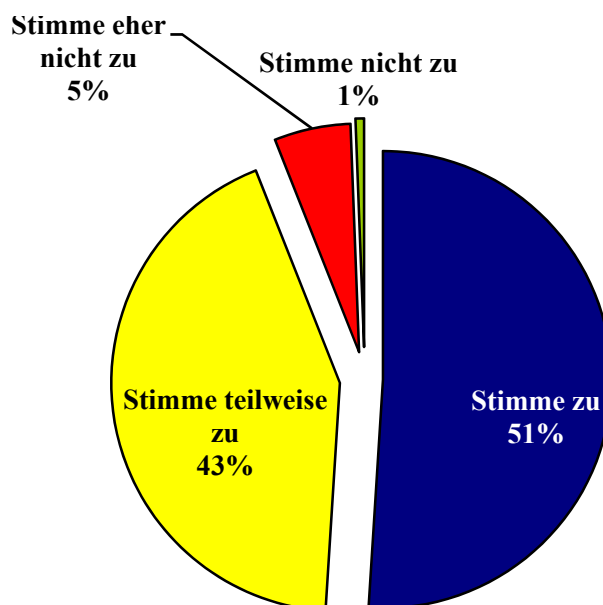


Abbildung 13: Angemessenheit der studentischen Bewertung

Es konnten sich viele Studierende (84% bei n=184) vorstellen, selbst als Prüfer zu agieren. Nur wenige Studierende würden nicht gerne selber prüfen (Abb. 14).

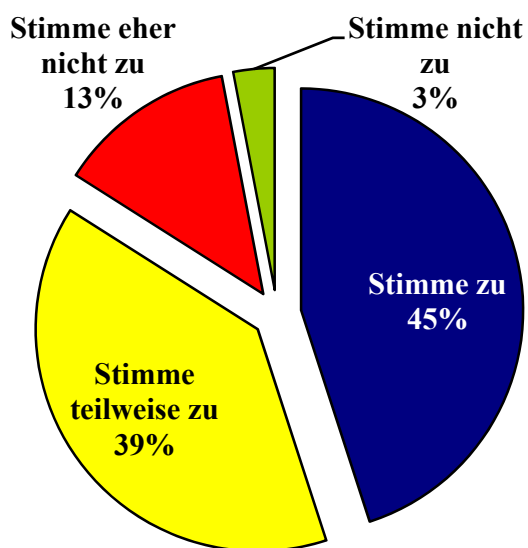


Abbildung 14: Zutrauen der Studierenden zum Selberprüfen

5 Diskussion

5.1 Beurteilung des Einsatzes von studentischen Tutoren als Prüfer

5.1.1 Zusammenfassung der Ergebnisse

Studentische Tutoren und Ärzte unterschieden sich nicht wesentlich in der Bewertung von Studierenden in der OSCE (maximaler Notenunterschied: 0,23). Insgesamt vergaben die Studierenden bei den einzelnen Checklisten- und Globalbeurteilungen im Durchschnitt etwas bessere Noten als die Ärzte; der Notenunterschied betrug bei der Gesamtnote der vier Stationen jedoch nur 0,11 auf einer Bewertungsskala von 1 bis 5. Die Übereinstimmung zwischen den Prüfern bei der Beurteilung der medizinischen Basisfertigkeiten war mit Kappa-Werten bis 0,64 gut.

Wenn sich allerdings die Bewertung der Prüfer auf kommunikative Aspekte (z.B. Sprache der Studierenden) bezog, war die Übereinstimmung zwischen Tutoren und Ärzten mit Kappa-Werten von 0,22 bis 0,55 deutlich geringer. Bei den themenbezogenen Aspekten vor allem an der CVR- und AU-Station waren sich beide Prüfergruppen mit einer rohen Übereinstimmung von überwiegend 80-90% und guten Kappa-Werten (CVR-Station: 0,8 bis 1,0; AU-Station: meist >0,6 bis 0,89) weitaus einiger. An der EKG-Station fielen die Übereinstimmungen etwas schwächer aus, der beste Kappa-Wert war 0,74. An der Depressionsstation gab es die niedrigste Übereinstimmung, die sich im gleichen Rahmen wie die allgemeinen Kommunikationsaspekte der Checklisten befand.

Die Mehrheit der Studierenden nahm an, dass die studentischen Prüfer in ihrer Bewertungsweise genauso adäquat (64%) und objektiv (69%) waren wie die Ärzte. Fast alle Studierenden (94%) fühlten sich angemessen beurteilt; sie konnten sich auch vorstellen, in einer OSCE selbst zu prüfen (84%).

5.1.2 Kritische Betrachtung der Ergebnisse

An vier Stationen wurden alle Studierenden simultan von studentischen Tutoren und Ärzten beurteilt. Die OSCE war eine summative Prüfung, die an einigen Stationen auch einen formativen Charakter hatte. Diese Stationen waren für den Einsatz von studentischen Prüfern nicht geeignet, da die Leistungen der Studierenden von den Prüfern direkt

kommentiert und gegebenenfalls korrigiert wurden, was eine höhere Fachkompetenz voraussetzt.

Zwischen den Stationen variierte der Grad der Übereinstimmung. Die höchsten Prüfer-Übereinstimmungen wurden für die themenspezifischen Checklisten-Items der CVR- und AU-Station ermittelt (CVR-Station: 0,8 bis 1,0; AU-Station: meist $>0,6$ bis 0,89). An diesen zwei Stationen waren die Checklisten durch Dichotomisierung der Antwortmöglichkeiten offensichtlich anwenderfreundlicher, verständlicher und einfacher zu bedienen als komplexe Antwortschemata (Andrew 1977). Somit hatten selbst unerfahrene Prüfer wie die studentischen Tutoren keine oder nur geringe Schwierigkeiten, die Studierenden objektiv zu bewerten. Die schwächsten Übereinstimmungen mit Kappa-Werten zwischen 0,22 bis 0,55 zeigten der allgemeine Teil zum Patientenkontakt in Form einer dreistufigen Ordinal-Skala („gut“, „mittel“, „schlecht“) und die Items der Depressionsstation mit einer fünfstufigen Likert-Skala von 0 („trifft nicht zu“) bis 4 („trifft zu“). Diese Items sollten auf Grund des Interpretationsfreiraums genauer erläutert oder ebenfalls dichotomisiert werden, um bessere Übereinstimmungen zu erzielen.

An der EKG-Station gab es bei den Gesamtnoten die größte Notendifferenz (0,22) und die Übereinstimmungen zwischen den Prüfern fielen ebenfalls etwas geringer aus (maximaler Kappa-Wert: 0,74). Die Bewertung mit der Checkliste war an der EKG-Station erschwert, da die meisten Studierenden ein EKG richtig auswerten konnten und die Prüfer nur zwischen den Kategorien „erfüllt“ und „nicht erfüllt“ entscheiden konnten. In diesem Fall war die dichotome Form der Checkliste von Nachteil, weil kleine Abweichungen bei der Elektrodenplatzierung und Auswertung des EKGs nicht berücksichtigt waren und es nicht genau genug festgelegt worden war, wie damit umzugehen ist. Deshalb sollten diese Checklisten um die Kategorie „fast richtig“ erweitert werden. Die Unterschiede in der globalen Bewertung ergaben sich hauptsächlich aus der Geschwindigkeit und Sicherheit bei der EKG-Interpretation. Die Ärzte bewerteten hier etwas strenger als die studentischen Tutoren und erwarteten von den Studierenden teilweise mehr als die im Kurs gelehrt Basisanforderungen. Dies lag vor allem daran, dass an dieser Station Lehrärzte eingesetzt worden waren, die nur am Prüfertraining teilgenommen haben. Die bessere Bewertung der studentischen Tutoren kann demnach auf deren aktive Beteiligung am Unterricht und der daraus resultierenden besseren Leistungseinschätzung oder auf einer möglichen Solidarität unter den Studierenden beruhen.

Gute Übereinstimmungen zwischen zwei Prüfern sind von vielen Faktoren abhängig wie Prüfertraining, Konstruktion der Checklisten und Beteiligung der Prüfer an der Stationsentwicklung (Wilkinson et al. 2003b). Eine stetige Verbesserung der Faktoren zur Optimierung der Übereinstimmungen zwischen den Prüfern ist sowohl kosten- als auch zeitintensiv. Der Aufwand für eine hohe Prüferübereinstimmung muss im Verhältnis zu den dazu notwendigen Bemühungen und Kosten stehen. Die bisher ermittelten Prüferübereinstimmungen zwischen zwei Ärzten (z.B. O’Conner and McGraw 1997) erreichten ähnliche Kappa-Werte (0,65-0,71) wie in dieser Studie. Von daher können die Übereinstimmungen zwischen den studentischen Tutoren und Ärzten als gut bewertet werden.

Die vorliegenden Ergebnisse sind nicht auf jede OSCE übertragbar, da die Interrater-Reliabilität für jede OSCE-Station neu ermittelt werden muss. Unter diesem Vorbehalt hat diese Studie gezeigt, dass ältere Medizinstudierende, die in die Lehre von Basisfertigkeiten integriert waren, Studierende an OSCE-Stationen mit klar formulierten Checklisten-Items reliabel bewerten konnten. Dies dürfte nicht für Studierende zutreffen, die an der Lehre unbeteiligt sind.

5.1.3 Vergleich mit bisherigen Publikationen

Dies ist vermutlich die erste Studie in der Humanmedizin zur Reliabilität und Akzeptanz von studentischen Tutoren als Prüfer in einer OSCE. Nur in der Zahnmedizin wurde bisher der Einsatz von studentischen Prüfern in einer OSCE beschrieben (Ogden et al. 2000). Dort wurden insgesamt drei Studierende aus dem letzten Studienjahr zusammen mit einem Zahnarzt an drei OSCE-Stationen als Prüfer eingesetzt. Die Studierenden hatten zuvor als Tutoren in einer Semesterwoche zahnmedizinische Basis-Untersuchungen in Kleingruppen unterrichtet. Die Bewertungen erfolgten anhand von Checklisten mit einer dreistufigen Likert-Skala („nicht korrekt“= 0 Punkte, „fast korrekt“ = 0,5 Punkte, „korrekt“ = 1 Punkt). Bei 116 von 125 Studierenden bewerteten die studentischen und ärztlichen Prüfer nur um einen Punkt anders. Die Prüfer-Übereinstimmungen waren sehr gut, es muss jedoch einschränkend erwähnt werden, dass an allen Stationen jeweils immer das gleiche Prüferpaar bewertete. In unserer Untersuchung war die Prüferanzahl größer (20 Studierende und 42 Ärzte) und es wurde zwischen den Stationen gewechselt, so dass immer wieder neue Prüferpaare aufeinander

trafen. Von daher waren die Prüfer-Übereinstimmungen unabhängig von der Prüferzusammensetzung an den jeweiligen Stationen, so dass - im Vergleich zur Studie von Ogden et al. (2000) - unsere Ergebnisse weitaus weniger durch informelle Absprachen der Prüfer-Paare kontaminiert sind.

Bisher wurden an den medizinischen Fakultäten amerikanischer Universitäten ältere Studierende in einer formativen OSCE vor allem als SP eingesetzt, die Studierenden ein mündliches Feedback geben (Harris et al. 1990). Für unseren Kurs „Ärztliche Basisfertigkeiten“ und die anschließende OSCE haben wir ältere SP für die authentischere Gestaltung der Stationen bevorzugt. Da die Anamnesestationen der OSCE eine summative Prüfung ergaben, wurde den Studierenden keine Rückmeldung über ihre Leistung gegeben. An den Stationen, wie z.B. bei der Reanimation, wurde es den Ärzten überlassen, ein kurzes Feedback zu geben.

In einer weiteren amerikanischen Studie arbeiteten sechs Studierende kurz vor dem Examen als Prüfer an zwei von zehn OSCE-Stationen, die mit einer siebenstufigen Skala bewertet wurden (Reiter et al. 2004). Unter Aufsicht eines zweiten ärztlichen Prüfers teilten die studentischen Prüfer den Studierenden am Ende der Station die Note mit und kommentierten die Leistung. Mit dieser Studie wurde überprüft, ob studentische Prüfer fähig sind, ein gutes mündliches Feedback zu geben. Am Ende der OSCE erfolgte eine Evaluation mit Fragen zum generellen Einsatz von studentischen Prüfern und zur Fähigkeit, ein arztähnliches Feedback geben zu können. Die Beurteilung der studentischen Prüfer durch die Studierenden fiel genauso positiv wie in Göttingen aus. Als mögliche Erklärung nannten Reiter et al. die guten Noten, die den Studierenden von den studentischen Prüfern während der OSCE gegeben worden sind. Zum Zeitpunkt der Göttinger Evaluation konnten die Studierenden höchstens vermuten, dass die Bewertung durch die Tutoren besser ausfallen würde, so dass die positive Bewertung der studentischen Prüfer in unserer Studie nicht auf einen Noteneffekt beruhen dürfte. Konsequenterweise glaubten auch die meisten Studierenden, dass es keinen Unterschied in der Bewertungsweise und der Objektivität zwischen studentischen und ärztlichen Prüfern gab (64% und 69%).

In einer anderen Studie gaben studentische Prüfer in einem Kommunikations- und Anamnesekurs deutlich bessere Beurteilungen als die ärztlichen Dozenten (Rudy et al. 2001). Dies wurde mit einer Freundschaft unter den Studierenden und der daraus resul-

tierenden Scheu begründet, untereinander schlechte Bewertungen zu geben. Studierende haben in der Tat ambivalente Gefühle, wenn sie sich gegenseitig beurteilen sollen (Arnold et al. 2005) und sind nicht der Ansicht, klinische Fertigkeiten ihrer Kommilitonen objektiv bewerten zu können (Asch et al. 1998). Unsere anfänglichen Befürchtungen, dass die Beurteilungen der studentischen Tutoren ebenfalls deutlich besser ausfallen würden, bestätigten sich nicht: die Notenunterschiede zu den ärztlichen Prüfern waren sehr gering (maximal 0,23). Viele studentische Bewertungen fielen zu Gunsten der Studierenden aus, oft aber auch zu Ungunsten der Studierenden. Eine systematische Bevorzugung der Studierenden durch studentische Prüfer kann in dieser Studie ausgeschlossen werden. Im Gegensatz zu der Studie von Asch et al. (1998) vertraten bei unserer Evaluation die meisten Studierenden (84%) die Ansicht, dass sie jüngere Studierende in der OSCE bewerten könnten.

In zwei weiteren Studien über die Bewertung von Studierenden durch Kommilitonen im höheren Semester zeigten sich Unterschiede in der Fähigkeit, Leistungsniveaus in Blockpraktika zu differenzieren. Studierende mit einem schwach ausgeprägten zwischenmenschlichen Verhalten neigten eher zu negativen Bewertungen (Lurie et al. 2006a und b). Deshalb ist die Auswahl von geeigneten Studierenden als Prüfer für die OSCE entscheidend. Generell sollte bei jedem Einsatz von studentischen Prüfern beachtet werden, dass Studierende nicht den Wissensstand und die gleiche Auffassungsgabe wie Ärzte haben und die Schwerpunkte nicht selbstständig setzen können. Daher benötigen sie als Prüfer in der OSCE klare Kriterien für die Relevanz der Bewertung (Fry 1990). Da in unserer Studie die studentischen Prüfer im Kurs „Ärztliche Basisfertigkeiten“ aktiv an der Lehre teilgenommen haben, waren ihnen die Erwartungen bekannt. Die OSCE gibt zudem einen strukturierten Rahmen für die Bewertung vor, so dass persönliche Eigenschaften bei der Bewertung keine so große Rolle spielen dürften.

5.2 Ethnische Herkunft als Einflussfaktor auf die OSCE

5.2.1 Zusammenfassung der Ergebnisse

Ausländische Studierende erhielten an sechs von acht Stationen signifikant schlechtere Gesamtnoten als deutsche Studierende. Diese Unterschiede wurden sowohl an kommunikativ als auch praktisch ausgerichteten Stationen beobachtet.

5.2.2 Kritische Betrachtung der Ergebnisse

Der Vergleich von ausländischen und deutschen Studierenden diente der Überprüfung der Objektivität in der OSCE. Die beobachteten Unterschiede können sowohl auf einer systematischen Benachteiligung beruhen als auch auf allgemein schlechteren Leistungen ausländischer Studierender (Van den Bussche et al. 2006), sprachlichen Problemen, oder einer unzureichenden Vorbereitung auf die Prüfung (Chenot et al. 2007). Bei den sprachrelevanten Stationen könnten die deutlichen Unterschiede zu den deutschen Studierenden auch auf kulturell anderen Kommunikationsstilen und Umgangsformen beruhen. Die meisten ausländischen Studierenden waren sog. „Bildungsinländer“ und sprachen fließend Deutsch. Ihre Sprachkenntnisse wurden nicht formal beurteilt. Jede Sprache hat individuelle subtile und nonverbale Kommunikationsaspekte; freies Sprechen über Gefühle, persönlicher Einbezug des Patienten in die Behandlung, Vermeiden von Unsicherheit, vermehrtes Lächeln, angemessene räumliche und (non-)verbale Distanz zwischen den Gesprächspartnern sind kulturell determiniert. Dadurch ergeben sich kulturelle Unterschiede in der Kommunikation, die sich auch in einem Arzt-Patienten-Gespräch manifestieren können (Fernandez et al. 2007). In unserer OSCE gab es keine ausländischen Prüfer, die studentischen Tutoren, Ärzte und SP waren alle deutsch. Interessant wären daher für die Zukunft vergleichende Untersuchungen mit dem Einsatz von Prüfern und SP ausländischer Herkunft.

Eine systematische Benachteiligung der ausländischen Studierenden ist unwahrscheinlich, da die studentischen Tutoren und Ärzte unabhängig voneinander die Ausländer schlechter als die Deutschen bewerteten. Die Checklistenbewertung ist auf Grund der klaren Struktur sehr objektiv, nur die Globalbeurteilung hat einen subjektiven Charakter. Es gab jedoch bei der Checklistenbewertung und bei der Globalbeurteilung gleichermaßen signifikante Unterschiede zwischen den Studierenden. Der größte Notenunterschied (1,17) zeigte sich an der Rezeptstation, die anhand der ausgefüllten Rezeptvorlagen ohne direkten Kontakt mit den Studierenden beurteilt wurde. Die Herkunft hatte von daher keinen Einfluss auf die Auswertung.

Insgesamt bildeten die ausländischen Studierenden eine deutliche Minderheit in der OSCE (26/206 Studierenden). Deshalb können diese Ergebnisse nicht als allgemeingültig angesehen werden. Sie geben eine Tendenz an, die eventuell durch eine weitere Studie mit ausgeglichenen Gruppengrößen überprüft werden sollte.

5.2.3 Vergleich mit bisherigen Publikationen

Im englischsprachigen Ausland wurden die möglichen Benachteiligungen ausländischer Studierender in der OSCE bereits untersucht. Es wurden vor allem hell- und dunkelhäutige Studierende miteinander verglichen, die alle fließend Englisch sprechen konnten (Haq et al. 2005, McManus et al. 1996). Dies entspricht unserer Untersuchung, bei der fast alle ausländischen Studierenden mit der deutschen Sprache aufgewachsen sind. Die Ausländer in unserer OSCE hatten jedoch eine andere ethnische Herkunft; ihre Familien stammten überwiegend aus Süd- oder Osteuropa und Vorderasien.

Nur wenige vergleichende Untersuchungen der Bewertung in der OSCE fanden keine Unterschiede zwischen einheimischen und ausländischen Studierenden (Bienenstock et al. 2000, Campos-Outcalt et al. 1994). In den meisten Untersuchungen erhielten ausländische Studierende signifikant schlechtere Bewertungen (Dillner 1995, Lumb and Vail 2004). Insbesondere an den Anamnesestationen wurden ausländische Studierende schlechter beurteilt, was mit Unterschieden in der Kommunikationsfähigkeit begründet wurde (Wass et al. 2003). In den meisten Studien gab es keine sprachlichen Probleme, da die ausländischen Studierenden - wie oben angedeutet - die Sprache des Studienlandes fließend sprechen konnten. Es handelte sich eher um ein spezielles Problem der kulturellen Interpretation und des Verstehens (Dewhurst et al. 2007). Die verbale Kommunikation eines Ausländers ist von seiner Lebenssituation, Lernweise und Integration abhängig. Nonverbale Kommunikation ist kulturell determiniert und kann sich in Gestik, Emotionen und Verhalten zu derjenigen des Studienlandes unterscheiden. Van Zanten et al. (2004) haben in diesem Zusammenhang festgestellt, dass sich die Zufriedenheit von SP mit Ärzten unterschiedlicher ethnischer Herkunft zwar nicht signifikant unterscheidet, sie sich dennoch bei Ärzten gleicher Herkunft wohler fühlen. In einer anderen Studie zu diesem Thema wurden hellhäutige Studierende von hellhäutigen SP im Durchschnitt etwas besser als von dunkelhäutigen SP bewertet (Colliver et al. 2007). In unserer OSCE erhielten die ausländischen Studierenden sowohl an den sprachrelevanten als auch an den praktischen Stationen signifikant schlechtere Bewertungen. Von daher kann als Grund - neben dem kulturellen Kommunikationsunterschied - auch eine schlechtere Vorbereitung auf die OSCE und eine insgesamt schlechtere Leistung vermutet werden.

In den Niederlanden erfolgte ein Vergleich der OSCE-Bewertungen für niederländische und ausländische Studierende, die zu 60% mit ihrer Heimatsprache und zu 40% mit der

niederländischen Sprache aufgewachsen waren (Schoonheim et al. 2007). In dieser Studie wurde untersucht, ob eine Verlängerung der Stationszeit sich vorteilhaft auf die Bewertung der ausländischen Studierenden auswirkt. Die ausländischen Studierenden wurden jedoch auch in dieser OSCE schlechter bewertet. Deshalb plädieren die Autoren für mehr Unterstützung und Übungsmöglichkeiten für ausländische Studierende in der medizinischen Ausbildung.

5.3 Geschlecht als Einflussfaktor auf die OSCE

5.3.1 Zusammenfassung der Ergebnisse

Es gab keine relevanten Unterschiede zwischen den weiblichen und männlichen Studierenden. Männer erhielten an allen Stationen im Durchschnitt etwas schlechtere Gesamtnoten als Frauen. Die Notenunterschiede waren insgesamt jedoch sehr gering, bis auf eine vergleichsweise große Differenz von 0,47 an der CVR-Station. Es lagen nur zwei signifikante Unterschiede vor: Depressionsstation (studentische Bewertung) und Hygienestation.

5.3.2 Kritische Betrachtung der Ergebnisse

Mit dem Geschlechter-Vergleich wurde ebenfalls die Objektivität der OSCE überprüft. Da diese OSCE mit drei Anamnesen, einer beschreibenden Dermatologie- und Hygienestation vor allem aus sprachrelevanten Stationen bestand, könnte man zunächst einen Nachteil für die männlichen Studierenden vermuten, da Frauen im Durchschnitt als sprachbegabter und patientenorientierter gelten (Krupat et al. 1999, Ramsbottom-Lucier et al. 1995). Dies wurde im Rahmen der PISA-Studie 2000 ebenfalls festgestellt: Mädchen sind stärker sozial orientiert und zeigen mehr Empathie und Hilfsbereitschaft; Jungen neigen dagegen zu individualistischen und aggressiven Orientierungen und sind weniger bereit, Verantwortung zu übernehmen (Deutsches PISA-Konsortium 2001). Die Geschlechterrollen werden gesellschaftlich unterschiedlich geprägt, woraus sich verschiedene Interessen, Bedürfnisse, Kompetenzen und Lebenserfahrungen von Frauen und Männern, Mädchen und Jungen ergeben (Artzt 2004). Stereotype Vorstellungen vom „starken“ Mann und der „sanften“ Frau schlagen sich in unterschiedlichen Erziehungsstilen und Sozialisationserfahrungen nieder. Nach der Geburt bzw. schon vorher

haben Eltern ein genaues idealtypisches Bild für Mädchen und Jungen und beginnen sofort mit einer (unbewussten) geschlechtsspezifischen Behandlung und Erziehung (Scheu 1977). Obwohl unsere OSCE überwiegend sprachlich gewichtet war, gab es keinen relevanten Unterschied zwischen Männern und Frauen. Dies lässt auf eine gute Lehre und Prüfungsvorbereitung schließen, da gute Kommunikationsfertigkeiten nicht angeboren sind, sondern unterrichtet werden können (Laidlaw et al. 2006). Leistungen hängen auch von der Motivation und dem Interesse für das jeweilige Fachgebiet ab (Artzt 2004).

Das Geschlecht der Prüfer und der SP kann das Verhalten der Studierenden und deren klinischen Entscheidungen beeinflussen (Wiskin et al. 2004). In unserer OSCE war die Mehrheit der Prüfer und SP weiblich, es wurde jedoch kein Geschlecht deutlich besser bewertet. In weiteren Studien könnte mit SP unterschiedlichen Geschlechts dieser Einflussfaktor genauer untersucht werden.

5.3.3 Vergleich mit bisherigen Publikationen

In den meisten Studien, bei denen die Leistung in einer OSCE zwischen den Geschlechtern verglichen wurde, waren die männlichen Studierenden schlechter (Ferguson et al. 2002, Lumb and Vail 2004, McManus et al. 1996). Es wird dabei aber nicht von einer Diskriminierung der männlichen Studierenden ausgegangen, sondern es wird angenommen, dass die Prüfungsform der OSCE eher „frauenorientiert“ ist, da der Schwerpunkt auf den kommunikativen Fähigkeiten der Prüflinge liegt. In naturwissenschaftlich ausgelegten OSCEs erhielten dagegen die männlichen Studierenden bessere Beurteilungen (Rothman et al. 1995). Eine Geschlechterdifferenz in unterschiedlichen Fachbereichen wurde in der PISA-Studie 2000 ebenfalls festgestellt; die Mädchen zeigen eine höhere Lese- und Sprachkompetenz, während die Jungen über höhere Kompetenzen im Bereich Mathematik verfügen (Deutsches PISA-Konsortium 2001). In unserer OSCE erhielten die männlichen Studierenden zwar im Durchschnitt etwas schlechtere Bewertungen, doch waren die Notenunterschiede insgesamt sehr gering. Somit kann davon ausgegangen werden, dass die Männer trotz der überwiegend sprachlich gewichteten Stationen in dieser OSCE nicht benachteiligt wurden.

5.4 Akzeptanz der OSCE im Vergleich mit bisherigen Publikationen

Die Evaluation der OSCE zeigte eine hohe Akzeptanz dieser Prüfungsform bei den Studierenden. Die hohe Rücklaufquote (92%) spricht für die Repräsentativität der Daten.

In vergleichenden Studien wurde die OSCE als effektive und objektive Methode zur Überprüfung von klinischen Fähigkeiten evaluiert (Newble et al. 1978) und gegenüber den schriftlichen und mündlichen Prüfungen bevorzugt (Wassermann et al. 1982). Es wurde der Wunsch geäußert, vermehrt OSCEs in das Curriculum zu integrieren (Pierre et al. 2004). Diese Publikationen stehen beispielhaft für viele weitere Veröffentlichungen über die OSCE, die durch unsere Evaluation bestätigt werden konnten.

Während unserer OSCE empfanden 40% der Studierenden die Prüfungsatmosphäre angespannt und wurden etwas verunsichert. Aber nur eine Minderheit der Studierenden (12%) schätzte die Prüfung als nicht fair ein. In anderen Studien wurde die Zeit von fünf Minuten pro Station als zu kurz und als Grund für diese negativen Äußerungen angesehen (Lazarus and Kent 1983, Pierre et al. 2004). Denn durch die geringe Zeit wird der Prüfungsstress größer, was einen nachteiligen Effekt auf die Leistung hat (Duffield and Spencer 2002).

Die Mehrheit der Studierenden (68%) bereitete sich nach eigenen Aussagen bis zu drei Stunden für die OSCE vor und ein Drittel lernte mehr als drei Stunden. Nur eine Minderheit (2%) hatte die Themen gar nicht wiederholt. In einer anderen Studie über die Vorbereitungszeit und -weise für eine OSCE betrug die Hauptstudienzeit ebenfalls ungefähr 3,5 Stunden und zu 20% wurden neben Textbüchern und Aufzeichnungen auch körperliche Untersuchungen geübt (Mavis 2000).

Ein Feedback im Anschluss an die Prüfung ist für die Studierenden motivierend, weil Stärken und Schwächen aufgezeigt werden können (Pierre et al. 2004). Da die Göttinger OSCE an vier Nachmittagen durchgeführt wurde und ein Informationsaustausch unter den Studierenden vermieden werden sollte, blieb das Feedback - bis auf die Dermatologie-, Reanimations- und Hygienestation - sehr allgemein. Dennoch waren viele Studierende (82%) der Ansicht, dass die OSCE einen positiven Lerneffekt hatte und sie ihre Leistung nun einschätzen konnten.

5.5 Schlussfolgerung und Forschungsausblick

Medizinstudierende höherer Semester, die an der Lehre aktiv beteiligt waren, können als Prüfer in einer summativen OSCE objektiv und reliabel Studierende bewerten. Die Übereinstimmungen zwischen den studentischen und ärztlichen Bewertungen waren gut. Eine höhere Übereinstimmung wurde bei klar formulierten dichotomen Checklisten beobachtet. Es gab keine Hinweise auf eine systematisch bessere Bewertung der Studierenden durch die studentischen Tutoren. Zwar waren die Bewertungen der studentischen Tutoren tendenziell etwas besser, aber die meist minimalen Unterschiede (maximal 0,23) hatten keine praktische Bedeutung. Die Mehrheit der Studierenden akzeptierte die studentischen Prüfer (94%) und konnte sich vorstellen, selber in einer OSCE zu prüfen (84%).

Die OSCE ist eine hoch anerkannte Prüfungsform, ihr breiter Einsatz ist aber wegen des kosten- und zeitintensiven Aufwands und hohen Personalbedarfs limitiert. Durch die Verwendung von älteren Studierenden zur Unterstützung der professionellen Prüfer kann der personelle Aufwand an Ärzten reduziert und die Organisation der OSCE vereinfacht werden, da Studierende zahlreich zur Verfügung stehen und flexibel einsetzbar sind. Deshalb sind unsere Ergebnisse sowohl für kleine Abteilungen mit wenig Personal als auch für größere Abteilungen, die sich durch den Aufwand überfordert fühlen, von Bedeutung. Zukünftige Studien sollten untersuchen, welche klinischen Fertigkeiten für die Beurteilung durch studentische Prüfer geeignet sind.

Ausländische Studierende wurden an sechs von acht Stationen signifikant schlechter beurteilt. Es gab keinen Hinweis auf eine systematische Benachteiligung der ausländischen Studierenden; die schlechteren Beurteilungen dürften überwiegend auf eine schlechtere Leistung und an den sprachrelevanten Stationen zusätzlich auf einen kulturell determinierten Kommunikationsstil zurückzuführen sein.

Obwohl die meisten Stationen sprachlich gewichtet waren und die Frauen im Durchschnitt als sprachbegabter und patientenorientierter gelten, gab es in der Bewertung von männlichen und weiblichen Studierenden keine wesentlichen Unterschiede. Dies lässt auf eine gute Lehre und Prüfungsvorbereitung schließen.

Die allgemeine Evaluation der OSCE ergab eine hohe Akzeptanz dieser Prüfungsform. Die Mehrheit der Studierenden (95%) hatte zum ersten Mal an einer OSCE teilgenommen. Es waren ca. 80% der Evaluierenden der Ansicht, dass sie durch die OSCE ihre

Leistungen besser einschätzen könnten, indem Schwächen aufgezeigt werden. Es wurde jedoch auch bemängelt, dass die OSCE sehr stressbelastet sei und eine angespannte Atmosphäre vermutlich wegen des Zeitmangels vorherrschen würde. Dennoch wurde sie als die beliebteste Prüfungsform gewählt.

6 Zusammenfassung

Hintergrund: Einer Integration von OSCEs in das medizinische Curriculum steht der hohe personelle, zeitliche und finanzielle Aufwand entgegen. Der Einsatz von Medizinstudierenden als Prüfer könnte diese Hindernisse teilweise überwinden. Wie jede Prüfung sollte auch die OSCE Menschen auf Grund ihrer ethnischen Herkunft oder ihres Geschlechts nicht benachteiligen.

Methoden: Im Frühjahr 2005 wurden 213 Studierende des ersten klinischen Semesters an der Universität Göttingen nach Abschluss des Semesterkurses „Ärztliche Basisfertigkeiten“ mit einer summativen OSCE geprüft. An vier OSCE-„Stationen“ bewerteten jeweils ein studentischer Kurs-Tutor und ein ärztlicher Prüfer simultan die Studierenden mit Checklisten und einer Globalbeurteilung (Kappa-Statistik als Maß der Übereinstimmung). Nach der Prüfung wurden die Studierenden gefragt, inwieweit sie studentische Prüfer akzeptieren. Mögliche Unterschiede in den Prüfungsergebnissen zwischen deutschen und ausländischen Studierenden sowie Frauen und Männer wurden mit dem Wilcoxon-Rangtest auf Signifikanz überprüft.

Ergebnisse: Die studentischen Prüfer vergaben etwas bessere Noten als die ärztlichen Prüfer (Unterschiede zwischen 0,03 und 0,23 auf einer 5-Punkte-Likert-Skala). Die Übereinstimmung zwischen den studentischen und ärztlichen Prüfern war für Checklisten und Globalbeurteilungen mäßig bis gut (Kappa 0,39 bis 0,64); insgesamt lag die Übereinstimmung an allen vier Stationen bei (Kappa) 0,58. Die meisten Studierenden sahen keinen Unterschied zwischen studentischen und ärztlichen Prüfern in der Bewertungsweise (64%) und attestierten den studentischen Tutoren Objektivität (69%). Fast alle (84%) konnten sich vorstellen, jüngere Studierende in einer OSCE zu prüfen. Die ausländischen Studierenden wurden trotz guter deutscher Sprachfähigkeiten an sechs von acht OSCE-Stationen signifikant schlechter bewertet. Zwischen den Geschlechtern gab es kaum Unterschiede.

Schlussfolgerung: Studentische Tutoren können in einer summativen OSCE als Prüfer für klinische Basisfertigkeiten eingesetzt werden. Ihre geringfügig besseren Bewertungen haben keine praktische Bedeutung. Studierende akzeptieren studentische Tutoren als Prüfer in der OSCE. Die schlechteren Bewertungen der Ausländer sind nicht auf Benachteiligung, sondern eher auf eine schlechtere Leistung zurückzuführen.

7 Literaturverzeichnis

Almeida Troncon LE (2004):

Clinical skills assessment: limitations to the introduction of an OSCE “objective structured clinical examination” in a traditional Brazilian medical school.

Sao Paulo Med J 122, 12-17

Altman DG:

Practical statistics for medical research.

Chapter 14: Some common problems in medical research.

Chapman & Hall, London 1991, 396-435

Andrew BJ (1977):

The use of behavioral checklists to assess physical examination skills.

J Med Educ 52, 589-590

Arnold L, Shue CK, Kritt B, Ginsberg S, Stern DT (2005):

Medical students' views on peer assessment of professionalism.

J Gen Intern Med 20, 819-824

Artzt I (2004):

Reflektierte Koedukation nach PISA?

Konsens, Information des Deutschen Akademikerinnenbundes e.V. 3, 20-24

Asch E, Saltzberg D, Kaiser S (1998):

Reinforcement of self-directed learning and the development of professional attitudes through peer- and self-assessment.

Acad Med 73, 575

Barrows HS (1968):

Simulated patients in medical teaching.

Can Med Assoc J 98, 674-676

Bienenstock JL, Tzou WS, Martin SA, Fox HE (2000):

Effect of student ethnicity on interpersonal skills and objective standardized clinical examination scores.

Obstet Gynecol 96, 1011-1013

Blume JH, Perleth M, Busse R (2000):

Was kostet die Objective Structured Clinical Examination?

GMS Z Med Ausbild 17, 12-17

Bundesärztekammer:

Bundesärztekammerordnung: Ausschuss „Ausbildung zum Arzt, Hochschule und medizinische Fakultät“

<http://www.bundesärztekammer.de/05/30TB/10TB2004/06.pdf>, 2004

(Datum des letzten Zugriffs: 28.10.2005)

Bundesministerium für Gesundheit:

Approbationsordnung für Ärzte

<http://www.bmgbund.de/download/gesetze/gesundheitsberufe/approbation.pdf>

(Datum des letzten Zugriffs: 28.10.2005)

Burchard KW, Rowland PA, Coe NPW, Garb JL (1995):

A surgery oral examination: interrater agreement and the influence of rater characteristics.

Acad Med 70, 1044-1046

Campos-Outcalt D, Rutala PJ, Witzke DB, Fulginiti JV (1994):

Performances of underrepresented-minority students at the University of Arizona College of Medicine, 1987-1991.

Acad Med 69, 577-582

Carpenter JL (1995):

Cost analysis of objective structured clinical examinations.

Acad Med 70, 828-832

Chenot JF, Ehrhardt M (2003):

Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine Alternative zur Klausur.

ZFA 79, 437-442

Chenot JF, Altiner A, Erhardt M, Gulich M, Schmittziel L (2004a):

Seminar Didaktik und Lehre der Vereinigung der Hochschullehrer und Lehrbeauftragten für Allgemeinmedizin.

ZFA 80, 82-84

Chenot JF, Fischer T, Simmenroth-Nayda A, Fassheber S, Hummers-Pradier E, Aut B, Kernbach-Wighton G, Emmert S, Müller D, Küntzel H, Klockgether-Radke AP, Kochen MM (2004b):
Interdisziplinärer Pilot-OSCE „Medizinische Basisfähigkeiten“.
ZFA 80, 503-506

Chenot JF, Simmenroth-Nayda A, Simmen A, Scherer M (2007):
Medizinstudierende ausländischer Herkunft in Deutschland.
GMS Z Med Ausbild 24, 70

Clack GB, Head JO (1999):
Gender differences in medical graduates' assessment of their personal attributes.
Med Educ 33, 101-105

Cohen JA (1960):
A coefficient of agreement for nominal scales.
Educ Psychol Meas 20, 37-46

Cohen R, Rothman AI, Ross J, Poldre P (1993):
Security issues in standard-patient examinations - Impact of repeated use of objective structured clinical examination stations.
Acad Med 68, 73-75

Colliver JA, Swartz MH, Robbs RS (2007):
The effect of examinee and patient ethnicity in clinical-skills assessment with standardized patients.
Eur J Dent Educ 11, 222-229

Coovadia HM, Moosa A (1985):
A comparison of traditional assessment with the objective structured clinical examination (OSCE).
S Afr Med J 67, 810-812

Deutsches PISA-Konsortium. Baumert J, Kleine E, Neubrand M, Prenzel M, Schiefele U, Schneider W, Stanat P, Tillmann KJ, Weiß M:
PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.
Leske + Budrich, Opladen 2001

Dewhurst NG, McManus C, Mollon J, Dacre JE, Vale AJ (2007):
Performance in the MRCP (UK) Examination 2003–4: analysis of pass rates of UK graduates in relation to self-declared ethnicity and gender.
BMC Med 5, 8

Dillner L (1995):

Manchester tackles failure rate of Asian students.

BMJ 310, 209

Doig CJ, Harasym PH, Fick GH, Baumber JS (2000):

The effects of examiner background, station organization, and time of exam on OSCE scores assessing undergraduate medical students' physical examination skills.

Acad Med 75, 96-98

Duffield KE, Spencer JA (2002):

A survey of medical students' views about the purposes and fairness of assessment.

Med Educ 36, 879-886

Elliot DL, Fields SA, Keenen TL, Jaffe AC, Toffler WL (1994):

Use of a Group objective structured clinical examination with first-year medical students.

Acad Med 69, 990-992

Elstein A, Shulman L, Sprafka S:

Medical problem solving.

Harvard University Press, Cambridge, MA 1978

Elton LRB, Laurillard DM (1979):

Trends in research on student learning.

Stud High Educ 4, 87-102

Falck-Ytter Y, Stiegler I (1993):

Beurteilung klinischer Fertigkeiten mit dem OSCE-Verfahren.

GMS Z Med Ausbild 10, 48-55

Ferguson E, James D, Madeley L (2002):

Factors associated with success in medical school: systematic review of the literature.

BMJ 324, 952-957

Fernandez A, Wang F, Braveman M, Finkas LK, Hauer KE (2007):

Impact of student ethnicity and primary childhood language on communication skill assessment in a clinical performance examination.

J Gen Intern Med 22, 1155-1160

Fischer T, Chenot JF, Kleiber C, Kochen MM, Simmenroth-Nayda A, Staats H, Herrmann-Lingen C (2005):

Kurs "ärztliche Basisfähigkeiten" - Evaluation eines primärärztlich orientierten Unterrichtskonzepts im Rahmen der neuen Approbationsordnung.

GMS Z Med Ausbild 22, Doc 59

Fry SA (1990):

Implementation and evaluation of peer marking in higher education.

Assess Eval High Educ 15, 177-189

Haq I, Higham J, Morris R, Dacre J (2005):

Effect of ethnicity and gender on performance in undergraduate medical examinations.

Med Educ 9, 1126-1128

Harden RM (1988):

What is an OSCE?

Med Teach 10, 19-22

Harden RM, Gleeson FA (1979):

Assessment of clinical competence using an objective structured clinical examination (OSCE).

Med Educ 13, 41-54

Harden RM, Stevenson M, Downie WW, Wilson GM (1975):

Assessment of clinical competence using objective structured examinations.

BMJ 1975,1, 447-451

Harris IB, Miller WJ, Edwards JC (1990):

Ideas for medical education: Feedback in an objective structured clinical examination by medical students serving as patients, examiners, and teachers.

Acad Med 65, 433-434

Heylings DJ, Stefani LAJ (1997):

Peer assessment feedback marking in a large medical anatomy class.

Med Educ 31, 281-286

Hodges B, McIlroy JH (2003):

Analytic global ratings are sensitive to level of training.

Med Educ 37, 1012-1016

Kowlowitz V, Hoole AJ, Sloane PD (1991):

Implementing the objective structured clinical examination in a traditional medical school.

Acad Med 66, 345-347

Krupat E, Hiam CM, Fleming MZ, Freeman P (1999):

Patient-centeredness and its correlates among first year medical students.

Int J Psych Med 29, 347-56

Lai NM, Sivalingam N, Ramesh JC (2007):

Medical students in their final six months of training: progress in self-perceived clinical competence, and relationship between experience and confidence in practical skills.

Singapore Med J 48, 1018-1028

Laidlaw TS, Kaufman DM, MacLeod H, van Zanten S, Simpson D, Wrixon W (2006):

Relationship of resident characteristics, attitudes, prior training and clinical knowledge to communication skills performance.

Med Educ 40, 18-25

Landis JR, Koch GG (1977):

The measurement of observer agreement for categorical data.

Biometrics 33, 159-174

Lazarus J, Kent AP (1983):

Student attitudes towards the objective structured clinical examination (OSCE) and conventional methods of assessment.

S Afr Med J 64, 390-394

Lumb B, Vail A (2004):

Comparison of academic, application form and social factors in predicting early performance on the medical course.

Med Educ 38, 1002-1005

Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein RM (2006a):

Longitudinal trends in peer assessment amongst medical students.

Med Educ 40, 840-847

Lurie SJ, Nofziger AC, Meldrum S, Mooney C, Epstein RM (2006b):

Effects of rater selection on peer assessment among medical students.

Med Educ 40, 1088-1097

Martin JA, Reznick RK, Rothmann A, Tamblyn RM, Regehr G (1996):
Who should rate candidates in an objective structured clinical examination?
Acad Med 71, 170-175

Mavis BE (2000):
Does studying for an objective structured clinical examination make a difference?
Med Educ 34, 808-812

Mavis BE, Ogle KS, Lovell KL, Madden LM (2002):
Medical students as standardized patients to assess interviewing skills for pain evaluation.
Med Educ 36, 135-140

McIlroy JH, Hodges B, McNaughton N, Regehr G (2002):
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 77, 725-728

McManus IC, Richards P, Winder BC, Sproston KA (1996):
Final examination performance of medical students from ethnic minorities.
Med Educ 30, 195-200

Miller GE (1990):
The assessment of clinical skills/ competence/ performance.
Acad Med 65, 563-567

Morris VD (1976):
A positive approach to the utilization of student feedback in medical education.
J Med Educ 51, 541-545

Newble DI (1992):
Assessing clinical competence at the undergraduate level.
Med Educ 26, 504-511

Newble DI, Elmslie RG, Baxter A (1978):
A problem-based criterion-referenced examination of clinical competence.
J Med Educ 53, 720-726

Newble DI, Hoare J, Sheldrake PF (1980):
The selection and training of examiners for clinical examinations.
Med Educ 14, 345-349

Niehaus AH, DaRosa DA, Markwell SJ, and Folse R (1996):

Is test security a concern when OSCE stations are repeated across clerkship rotations?

Acad Med 71, 287-289

O'Conner HM, McGraw RC (1997):

Clinical skills training: developing objective assessment instruments.

Med Educ 31, 359-363

Ogden GR, Green M, Ker JS (2000):

The use of interprofessional peer examiners in an objective structured clinical examination: Can dental students act as examiners?

Br Dent J 189, 160-164

Pabst R (1995):

Medical Education and reform initiatives in Germany.

Int Med Educ 70, 1006-1011

Petrusa ER, Blackwell TA, Ainsworth MA (1990):

Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents.

Arch Intern Med 150, 573-577

Pierre RB, Wierenge A, Barton M, Branday J, Christie CD (2004):

Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica.

BMC Med Educ 4, 22

Ramsbottom-Lucier M, Johnson MM, Elam CL (1995):

Age and gender differences in students' preadmission qualifications and medical school performances.

Acad Med 70, 236-239

Reiter HI, Rosenfeld J, Nandagopal K, Eva KW (2004):

Do clinical clerks provide candidates with adequate formative assessment during objective structured clinical examinations?

Adv Health Sci Educ Theory Pract 9, 189-199

Rothman AI, Cohen R, Ross J, Poldre P, Dawson B (1995):

Station gender bias in a multiple-station test of clinical skills.

Acad Med 70, 42-46

Rudy DW, Fejfar MC, Griffith CH 3rd, Wilson JF (2001):

Self- and peer assessment in a first-year communication and interviewing course.

Eval Health Prof 24, 436-445

Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ (1991):

Sharing of information by students in an objective structured clinical examination.

Arch Intern Med 151, 541-544

Scheu U:

Wir werden nicht als Mädchen geboren - wir werden dazu gemacht. Zur frühkindlichen Erziehung in unserer Gesellschaft.

1. Auflage; Fischer Verlag, Frankfurt 1977

Schoonheim-Klein ME, Habets LL, Aartman IH, van der Vleuten CP, Hoogstraten J, van der Velden U (2006):

Implementing an Objective Structured Clinical Examination (OSCE) in dental education: effects on students' learning strategies.

Eur J Dent Educ 10, 226-35

Schoonheim-Klein ME, Hoogstraten J, Habets LL, Aartman IH, van der Vleuten CP, Manogue M, van der Velden U (2007):

Language background and OSCE performance: a study of potential bias.

Eur J Dent Educ 11,222

Simmenroth-Nayda A, Fischer T, Hermann-Lingen C, Wetzel D, Chenot JF, Staats H, Kochen MM (2004):

Praxis Lernen: Medizinische Basisfähigkeiten- ein Unterrichtskonzept im Rahmen der neuen Approbationsordnung.

Niedersächs Ärztebl 1, 22-24

Simmenroth-Nayda A, Chenot JF, Fischer T, Scherer M, Stanske B, Kochen MM (2007):

Mit Laienschauspielern das ärztliche Gespräch trainieren.

Dtsch Ärztebl 104, A847-852

Skelton JR, Hobbs FD (1999):

Descriptive study of cooperative language in primary care consultations by male and female doctors.

BMJ 318, 576-579.

Sloan DA, Donnelly MB, Schwartz RW, Strodel WE (1995):

The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance.

Ann Surg 222, 735-42

Statistisches Bundesamt Deutschland (2006a):

Ausländische Bevölkerung in Deutschland (Stand vom 31.12.2006)

[http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Bevoelkerung/](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Bevoelkerung/AuslaendischeBevoelkerung/Tabellen/Content50/TOP10,templateId=renderPrint.psml)

[AuslaendischeBevoelkerung/Tabellen/Content50/TOP10,templateId=renderPrint.psml](http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Bevoelkerung/AuslaendischeBevoelkerung/Tabellen/Content50/TOP10,templateId=renderPrint.psml)

(Datum des letzten Zugriffs: 16.11.2007)

Statistisches Bundesamt Deutschland (2006b):

Datenreport 2006: Zahlen und Fakten über die Bundesrepublik Deutschland.

Kapitel 14: Zuwanderer und Ausländer in Deutschland.

bpb Auszug aus Teil II, 565-572

Stillman PL, Swanson DB (1987):

Ensuring the clinical competence of medical school graduates through standardized patients.

Arch Intern Med 147, 1049-1052

Stillman PL, Regan MB, Swanson DB, and the Fourth-Year Performance Assessment Task Force Group (1987):

A diagnostic fourth-year performance assessment.

Arch Intern Med 147, 1981-1985

Tarifgemeinschaft deutscher Länder (2006):

Ärzte mit Patientenversorgung:

<http://www.tdl.bayern.de/Aerzte/Tabellen/Tabelle%20West.pdf>

Ärzte ohne Patientenversorgung:

<http://www.tdl.bayern.de/Tabellen/tabelle%20west%201.pdf>

(Datum des letzten Zugriffs: 14.11.2007)

Van den Bussche H, Wegscheider K, Zimmermann T (2006):

Medizinische Fakultäten: Der Ausbildungserfolg im Vergleich (II).

Dtsch Arztebl 103, A2225-2228

Van der Leuten CP, Norman GR, de Graaff E (1991):

Pitfalls and pursuit of objectivity: issues of reliability.

Med Educ 25, 110-118

Van der Vleuten CPM, Maastricht: persönliche Mitteilung 2006

Van der Vleuten CPM, van Luyk SJ, van Ballegooijen AMJ, Swanson DB (1989):
Training and experience of examiners.
Med Educ 23, 290-296

Van Zanten M, Boulet JR, McKinley DW (2004):
The influence of ethnicity on patient satisfaction in a standardized patient assessment.
Acad Med 79, 15-17

Viera AJ, Garrett JM (2005):
Understanding interobserver agreement: the Kappa statistic.
Fam Med 37, 360-363

Wass V, Roberts C, Hoogenboom R, Jones R, van der Vleuten C (2003):
Effect of ethnicity on performance in a final objective structured clinical examination: qualitative and quantitative study.
BMJ 326, 800-803

Wassermann HP, Slabbert BR, van Zyl JJW (1982):
Die objektif-struktureerd kliniese eksamen (OSKE).
S Afr Med J 61, 325-330

Waterston T, Cater JI, Mitchell RG (1980):
An objective undergraduate clinical examination in child health.
Arch Dis Child 55, 917-922

Watson AR, Houston IB, Close GC (1982):
Evaluation of an objective structured clinical examination.
Arch Dis Child 57, 390-398

Weißer FO:
Konzeption, Durchführung und Evaluation einer Objective Structured Clinical Examination (OSCE): eine neue Prüfungsform in der notfallmedizinischen Ausbildung und im deutschen medizinischen Regelstudium.
Med. Diss. Hannover 2000

Wilkinson TJ, Frampton CM (2004):
Comprehensive undergraduate medical assessments improve prediction of clinical performance.
Med Educ 38, 1111-1116

Wilkinson TJ, Fontaine S, Egan T (2003a):

Was a breach of examination security unfair in an objective structured clinical examination? A critical incident.

Med Teach 25, 42-46

Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T (2003b):

Objectivity in Objective Structured Clinical Examinations: Checklists are no substitute for examiner commitment.

Acad Med 78, 219-223

Wisikin CMD, Allan TF, Skelton JR (2004):

Gender as a variable in the assessment of final year degree-level communication skills.

Med Educ 38, 129-137

Woolf K, Haq I, McManus IC, Higham J, Dacre J (2007):

Exploring the underperformance of male and minority ethnic medical students in first year clinical examinations.

Adv Health Sci Educ Theory Pract, [Epub ahead of print] – abstract

Zaharias G, Piterman L, Liddell M (2004):

Doctors and patients: gender interaction in the consultation.

Acad Med 79, 148-155

8 Anhang

8.1 Bewertungsbogen für die CVR-Station

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
 Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Allgemeinmedizin / Psychosomatik (F1/F19)**
 Semester: **WS 04/05**
 Station: **Kardiovaskuläre Risikoanamnese**
 Modul: **1.1**

Matrikelnummer:

Zuname:

Vorname:

Rater	Runde
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
0	<input type="checkbox"/>

- | | Gut | Mittel | Schlecht |
|--|--------------------------|--------------------------|--------------------------|
| 1. Kommunikation mit dem Simulationspatienten: | | | |
| - Begrüßung und Vorstellung | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Klare Sprache, verständlicher Ausdruck | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Keine Fachwörter | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Guter Patientenkontakt | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Blutdruckanamnese, relevante Erkrankungen: | trifft zu | trifft nicht zu | |
| - Letzter gemessener Wert | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Frühere Blutdruckmessungen | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Diabetes bekannt | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Nierenerkrankungen | <input type="checkbox"/> | <input type="checkbox"/> | |
| 3. Familienanamnese: | | | |
| - Kardiovaskuläre Erkrankungen in der Familie | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Vor dem 60. LJ kardiovaskulär erkrankte Angehörige | <input type="checkbox"/> | <input type="checkbox"/> | |
| 4. Raucheranamnese: | | | |
| - Nikotinkonsum | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Seit wann | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Wie viele Zigaretten pro Tag | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Abstinenzversuche oder Absicht aufzuhören | <input type="checkbox"/> | <input type="checkbox"/> | |
| 5. Alkohol, körperliche Aktivität: | | | |
| - Alkoholkonsum | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Menge | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Körperliche Aktivität | <input type="checkbox"/> | <input type="checkbox"/> | |
| 6. Cholesterinanamnese: | | | |
| - Cholesterinmessung | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Gemessener Wert | <input type="checkbox"/> | <input type="checkbox"/> | |
| 7. Kurze Sozialanamnese: | | | |
| - Alter | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Beruf | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Familienstand | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Stressoren (beruflich, privat) | <input type="checkbox"/> | <input type="checkbox"/> | |
| 8. Medikamentenanamnese: | | | |
| - Medikamenteneinnahme | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Allergie | <input type="checkbox"/> | <input type="checkbox"/> | |

Globalbeurteilung	Note	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		1	2	3	4	5
Die Leistung des Studierenden war		sehr gut	gut	befriedigend	ausreichend	mangelhaft

8.2 Bewertungsbogen für die AU-Station

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Arbeitsmedizin, Sozialmedizin (F3)**

Semester: **WS 04/05**

Station: **Arbeitsanamnese und AU**

Modul: **1.1**

Matrikelnummer:

Zuname:

Vorname:

Rater	Runde
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
0	<input type="checkbox"/>

- | | Gut | Mittel | Schlecht |
|---|--------------------------|--------------------------|--------------------------|
| 1. Kommunikation mit dem Simulationspatienten: | | | |
| - Begrüßung und Vorstellung | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Klare Sprache, verständlicher Ausdruck | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Keine Fachwörter | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| - Guter Patientenkontakt | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. Beschwerdeanamnese: | trifft zu | trifft nicht zu | |
| - Blick auf den Unterarm | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Ursache des Schmerzes | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Schmerzlokalisierung | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Schmerzausstrahlung | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Schmerzdauer | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Funktionseinschränkung | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Frage nach Händigkeit | <input type="checkbox"/> | <input type="checkbox"/> | |
| 3. Arbeitsanamnese: | | | |
| - Aktuelle Tätigkeit | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Besondere Belastungen am Arbeitsplatz | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Spezielle Beanspruchungen der linken Hand am Arbeitsplatz | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Allgemeine Arbeitsatmosphäre | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Andere Belastungen, häusliche Belastungen | <input type="checkbox"/> | <input type="checkbox"/> | |
| 4. AU-Entscheidung: | | | |
| - Zunächst keine AU-Entscheidung | <input type="checkbox"/> | <input type="checkbox"/> | |
| - AU-Bescheinigung maximal für 2-3 Tage | <input type="checkbox"/> | <input type="checkbox"/> | |
| - Verständliche Begründung/Erklärung für den Patienten | <input type="checkbox"/> | <input type="checkbox"/> | |

Globalbeurteilung	Note	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		1	2	3	4	5
Die Leistung des Studierenden war		sehr gut	gut	befrie- digend	ausrei- chend	mangel- haft

8.3 Bewertungsbogen für die Depressionsstation

Georg-August-Universität Göttingen • Stiftung Öffentliches Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Allgemeinmedizin / Psychosomatik (F1/F19)**

Semester: **WS 04/05**

Station: **Depression**

Modul: **1.1**

Matrikelnummer:

--	--	--	--	--	--	--	--	--	--

Zuname:

Vorname:

Rater

1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>
7	<input type="checkbox"/>	<input type="checkbox"/>
8	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	<input type="checkbox"/>
0	<input type="checkbox"/>	<input type="checkbox"/>

Runde

1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>
7	<input type="checkbox"/>	<input type="checkbox"/>
8	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	<input type="checkbox"/>
0	<input type="checkbox"/>	<input type="checkbox"/>

	trifft nicht zu		trifft zu		
	0	1	2	3	4
1. Kommunikation mit dem Simulationspatienten: Begrüßung, klare Sprache, verständliche Sprache, Patientenkontakt, Beendigung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Aktuelle Symptomatik: Aktuelles Anliegen, aktuelle Beschwerden, Verlauf seit der letzten Untersuchung, Wissen über zurückliegende Befunde,	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Depressive Symptomatologie: Allgemeine Stimmung, Ängste, soziale und familiäre Einschränkungen, Schlafstörungen und Appetitverlust, Suizidgedanken	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Auslösende Situationen: Aktuelle oder zurückliegende Belastungen, Familiensituation, berufliche Situation, länger bestehende oder anstehende Lebensveränderungen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Globalbeurteilung

Note

	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	1	2	3	4	5
Die Leistung des Studierenden war	sehr gut	gut	befrie- digend	ausrei- chend	mangel- haft

8.4 Bewertungsbogen für die EKG-Station

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Allgemeinmedizin / Psychosomatik (F1/F19)**
Semester: **WS 04/05**
Station: **EKG**
Modul: **1.1**

Matrikelnummer:

Zuname:

Vorname:

Rater	Runde
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
0	<input type="checkbox"/>

- | | | | |
|--|--|--|---|
| 1. Kommunikation mit dem Simulationspatienten:
Begrüßung, klare Sprache, verständlicher Ausdruck,
professionelle Gesprächsführung | schlecht
<input type="checkbox"/>

nein
<input type="checkbox"/>

nein
<input type="checkbox"/>

nein
<input type="checkbox"/>

nein
<input type="checkbox"/> | mittel
<input type="checkbox"/>

ja
<input type="checkbox"/>

ja
<input type="checkbox"/>

ja
<input type="checkbox"/> | gut
<input type="checkbox"/>

<input type="checkbox"/> |
| 2. Legt Extremitätenableitungen in korrekter Position an: | <input type="checkbox"/> | <input type="checkbox"/> | |
| 3. Legt Brustwandelektroden in korrekter Position an: | <input type="checkbox"/> | <input type="checkbox"/> | |
| 4. Erklärt flüssig und verständlich | <input type="checkbox"/> | <input type="checkbox"/> | |
| 5. Einfache Analyse des EKGs
- Rhythmus richtig angeben
- Frequenz richtig angeben
- Lagetyp richtig angeben | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> | |

Globalbeurteilung	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Die Leistung des Studierenden war	1 sehr gut	2 gut	3 befrie- digend	4 ausrei- chend	5 mangel- haft

8.5 Bewertungsbogen für die Dermatologiestation

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Dermatologie (F6)**
Semester: **WS 04/05**
Station: **Hauteffloreszenzen**
Modul: **1.1**

Matrikelnummer:

Zuname:

Vorname:

Rater	Runde
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
0	<input type="checkbox"/>

	trifft nicht zu			trifft zu	
	0	1	2	3	4
1. Erkennen der Erosio: Beschreibung einer Sekundäreffloreszenz nach Art und Form.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
2. Erkennen der Schuppe: Beschreibung einer Sekundäreffloreszenz nach Art und Lage.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
3. Erkennen der Kruste: Beschreibung einer Sekundäreffloreszenz nach der Art und Entstehungsursache.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
4. Benennung der Primäreffloreszenz: Begründete Herleitung der Primäreffloreszenz aus den obigen Sekundäreffloreszenzen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Beschreibung der Hauteffloreszenz der Abbildung 2: Beschreibung einer Primäreffloreszenz nach Art, Form und Farbe sowie einer Sekundäreffloreszenz.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Beschreibung der Verteilung dieser Effloreszenzen am Körper: Grundsätzliche und besondere Körperverteilung von einer Hauteffloreszenz.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
7. Erstellen einer Verdachtsdiagnose: Herleitung einer Diagnose mit Hilfe der dermatologischen Deskription	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Globalbeurteilung	Note	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		1	2	3	4	5
Die Leistung des Studierenden war		sehr gut	gut	befriedigend	ausreichend	mangelhaft

8.6 Bewertungsbogen für die Reanimationsstation

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Notfallmedizin (Q8)**

Semester: **WS 04/05**

Station: **Reanimation**

Modul: **1.1**

Matrikelnummer:

--	--	--	--	--	--	--	--	--	--

Zuname:

Vorname:

Rater

1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>
7	<input type="checkbox"/>	<input type="checkbox"/>
8	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	<input type="checkbox"/>
0	<input type="checkbox"/>	<input type="checkbox"/>

Runde

1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>
6	<input type="checkbox"/>	<input type="checkbox"/>
7	<input type="checkbox"/>	<input type="checkbox"/>
8	<input type="checkbox"/>	<input type="checkbox"/>
9	<input type="checkbox"/>	<input type="checkbox"/>
0	<input type="checkbox"/>	<input type="checkbox"/>

Der/die Studierende

	falsch/ fehlt	teilw. richtig	richtig
1. führt einen korrekten diagnostischen Block durch (Atmung, Puls, Bewußtsein)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. setzt bei Erkennen der Reanimationssituation einen Notruf ab	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. führt die Atemspende korrekt durch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. führt die Herzdruckmassage korrekt durch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. koordiniert den zeitlichen Ablauf der Maßnahmen korrekt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Globalbeurteilung

Note

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	2	3	4	5
sehr gut	gut	befrie- digend	ausrei- chend	mangel- haft

8.7 Bewertungsbogen für die Hygienestation

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Hygiene (F10)**

Semester: **WS 04/05**

Station: **Anlage einer peripheren Venenverweilkanüle**

Modul: **1.1**

Rater

Runde

1
2
3
4
5
6
7
8
9
0

1
2
3
4
5
6
7
8
9
0

Matrikelnummer:

Zuname:

Vorname:

	erfüllt	nicht erfüllt
1. Hygienische Händedesinfektion (HD) (Durchführung auch nach Punkt 2 möglich) <i>1</i>	<input type="checkbox"/>	<input type="checkbox"/>
2. Der Studierende sucht sich die entsprechend benötigten Gegenstände und legt sie sich griffbereit zurecht (max. 2 mal "nachgreifen")	<input type="checkbox"/>	<input type="checkbox"/>
3. Hautdesinfektion <i>2</i>	<input type="checkbox"/>	<input type="checkbox"/>
4. Anziehen von Schutzhandschuhen (Durchführung auch vor Punkt 2 oder 3 möglich)	<input type="checkbox"/>	<input type="checkbox"/>
5. Punktion der Vene <i>3</i>	<input type="checkbox"/>	<input type="checkbox"/>
6. Entfernen der Stahlkanüle: sofort in den Abwurf für spitze Gegenstände	<input type="checkbox"/>	<input type="checkbox"/>
7. Steriles Einsetzen des Mandrains, d.h. der Mandrain darf nicht mit der Haut des Patienten oder Anderem in Berührung kommen	<input type="checkbox"/>	<input type="checkbox"/>
8. Hygienische Händedesinfektion	<input type="checkbox"/>	<input type="checkbox"/>

Globalbeurteilung	Note	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		1	2	3	4	5
		sehr gut	gut	befriedigend	ausreichend	mangelhaft

- 1 erfüllt: 6 Schritte ?, 30 sec. ? (kann zeitlich verkürzt werden, wenn Student sagt, dass die HD mind. 30 sec. lang durchzuführen ist)

2 erfüllt: 1. Sprühen ?, 2. mit sterilem Tupfer im desinfizierten Bereich wischen ?, 3. Sprühen ?, 4. 30 sec. einwirken lassen ?, oder Wischdesinfektion: 1. Sprühen ?, 2. mit feuchtem sterilen Tupfer wischen ?, 3. 30 sec. einwirken lassen ?

3 erfüllt: Ohne Nachtasten der Vene im desinfizierten Bereich ?, ohne Berührung der Kanüle ?

8.8 Bewertungsbogen für die Rezeptstation

Georg-August-Universität Göttingen • Stiftung Öffentlichen Rechts
Bereich Humanmedizin • Universitätsklinikum • Medizinische Fakultät



OSCE Bogen

Abteilung: **Allgemeinmedizin / Psychosomatik (F1/F19)**

Semester: **WS 04/05**

Station: **Rote Liste**

Modul: **1.1**

Matrikelnummer:

Zuname:

Vorname:

Rater	Runde
1	<input type="checkbox"/>
2	<input type="checkbox"/>
3	<input type="checkbox"/>
4	<input type="checkbox"/>
5	<input type="checkbox"/>
6	<input type="checkbox"/>
7	<input type="checkbox"/>
8	<input type="checkbox"/>
9	<input type="checkbox"/>
0	<input type="checkbox"/>

Umgang mit der Roten Liste

	gar nicht	mit Verzögerung	zügig
1. Findet die Substanz im Wirkstoffteil (blau):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Findet die gesuchte Substanz im Präparateteil (weiß):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Findet die Gegenanzeige/Anwendungsbeschränkung im Signaturteil (orange):	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Auf dem Rezept wurde notiert:

	nein	ja
1. Name, Geburtsdatum, Anschrift des Patienten	<input type="checkbox"/>	<input type="checkbox"/>
2. Krankenkasse des Patienten	<input type="checkbox"/>	<input type="checkbox"/>
3. Medikament (Wirkstoff oder Handelsname)	<input type="checkbox"/>	<input type="checkbox"/>
4. Dosisangabe	<input type="checkbox"/>	<input type="checkbox"/>
5. Applikationsform	<input type="checkbox"/>	<input type="checkbox"/>
6. Packungsgröße N1-3 (N1)	<input type="checkbox"/>	<input type="checkbox"/>
7. Datum	<input type="checkbox"/>	<input type="checkbox"/>
8. Unterschrift	<input type="checkbox"/>	<input type="checkbox"/>
9. Wegstreichen freien Platzes	<input type="checkbox"/>	<input type="checkbox"/>

8.9 Studierendenanleitung für die CVR-Station

Setting:

Sie arbeiten in einer Allgemeinarztpraxis. Der/ die Patient/in ist Ihnen bisher nicht bekannt. In der Apotheke ist ein erhöhter Blutdruck gemessen worden. Sie/ er stellt sich nun bei Ihnen zum ersten Mal vor.

Ihre Aufgabenstellung:

Erheben Sie eine fokussierte kardiovaskuläre Risikoanamnese die alle wesentlichen Risikofaktoren einschließlich einer kurzen Sozialanamnese erfasst. Denken Sie bitte auch an Medikamente.

Das Gespräch wird mit der Ankündigung, den Blutdruck zu messen beendet.

Zeitvorgabe: 5 Minuten!, das erfordert ein zügiges Vorgehen!

Viel Erfolg!

8.10 Studierendenanleitung für die AU-Station

Setting:

Eine 36-jährige Patientin stellt sich in Ihrer Praxis erstmals vor. Sie klagt über starke, ziehende Schmerzen im linken Unterarm.

Die Patientin wünscht von Ihnen eine **Krankschreibung für 1 Woche**.

Ihre Aufgabenstellung:

- Gezielte Erhebung der Beschwerdeanamnese (Schmerzen, Funktion) und gezielte Erhebung der Arbeitsanamnese im Hinblick auf den Wunsch der Patientin nach einer Krankschreibung.
- Entscheiden Sie, ob die Ausstellung einer Arbeitsunfähigkeitsbescheinigung gerechtfertigt ist.
- Erklären Sie der Patientin Ihre Entscheidung nachvollziehbar.

Zeitvorgabe: 5 Minuten!

Viel Erfolg!

8.11 Studierendenanleitung für die Depressionsstation

Setting:

Sie arbeiten in einer Allgemeinarztpraxis. Der/ die Patient/in ist Ihnen bisher nicht bekannt. Er/ Sie ist von Ihrem Praxiskollegen, der augenblicklich im Urlaub ist, wegen unspezifischer Oberbauchbeschwerden bis vor 3 Wochen intensiv untersucht worden. Eine Gastroskopie ergab keinen Befund. Insgesamt fand sich kein Hinweis auf eine körperliche Erkrankung.

Ihre Aufgabenstellung:

Erheben Sie eine Anamnese und einen Befund zur psychischen Befindlichkeit der Patientin/des Patienten.

Zeitvorgabe: 5 Minuten

Viel Erfolg!

8.12 Studierendenanleitung für die EKG-Station

Setting:

Sie arbeiten alleine in einer Notaufnahme. Ein junger Mann kommt zu Ihnen und klagt über Herzrasen.

Ihre Aufgabenstellung:

- Legen Sie ein EKG an. Erklären Sie was sie tun! Leiten Sie das EKG ab.
- Geben Sie Rhythmus, Frequenz und Lagetyp an.

Die Station wird mit einer kurzen Ergebnismitteilung an den Patienten beendet.

Zeitvorgabe: 5 Minuten!, das erfordert ein zügiges Vorgehen!

Viel Erfolg!

8.13 Studierendenanleitung für die Dermatologiestation

Ihre Aufgabenstellung:

- Bitte beschreiben Sie detailliert alle Sekundäreffloreszenzen, die Sie auf der Abbildung 1 sehen, nach Art, Form und Lage. Aus welcher Primäreffloreszenz könnten diese Sekundäreffloreszenzen entstanden sein?
- Bitte beschreiben Sie ebenso die Hautveränderung auf dem Detailbild der Abbildung 2. Wenn Sie die Verteilung dieser Hautveränderungen auf dem Übersichtsbild der Abbildung 2 betrachten, was fällt Ihnen auf? Können Sie eine Hauterkrankung nennen, die typischerweise solche Effloreszenzen in dieser Verteilung zeigt?

Zeitvorgabe: 5 Minuten

Viel Erfolg!

8.14 Studierendenanleitung für die Rezeptstation (Version 1)

Setting:

Sie sind auf Hausbesuch und müssen ein Kassen-Rezept von Hand ausfüllen. Hierzu haben Sie vorgestempelte Blankorezepte. Frau Müller hat einen Harnwegsinfekt und ist im 3.Monat schwanger. Sie sind sich nicht sicher, ob dass von Ihnen ausgesuchte Antibiotikum in der Schwangerschaft verordnet werden kann.

Aufgabe:

Stellen Sie ein Rezept für **Cefixim** an. Nehmen Sie dabei die „Rote Liste“ zur Hilfe.

Erläutern Sie was sie tun.

- Schlagen Sie zunächst unter dem Wirkstoff nach!
- Wählen Sie ein Präparat
- Schlagen Sie nach, ob **Cefixim** in der Schwangerschaft kontraindiziert ist.
- Stellen Sie ein Rezept für **Cefixim** an

Patientendaten:

Maria Müller, Hainweg 1, 37073 Göttingen, geb. 8.05 1975, AOK Göttingen

Sie haben 5 Minuten Zeit.

Viel Erfolg!

8.15 Studierendenanleitung für die Rezeptstation (Version 2)

Setting:

Sie sind auf Hausbesuch und müssen ein Kassen-Rezept von Hand ausfüllen. Hierzu haben Sie vorgestempelte Blankorezepte. Herr Meyer hat eine eitrige Tonsillitis. Er hat ein Krampfleiden und nimmt regelmäßig Carbamazepin ein. Sie sind sich nicht sicher, ob das von Ihnen ausgesuchte Antibiotikum Wechselwirkungen mit dem Antikonvulsivum hat.

Aufgabe:

Stellen Sie ein Rezept für den Wirkstoff **Phenoxymethylpenicillin** (Penicillin V) auf. Nehmen Sie dabei die „Rote Liste“ zur Hilfe.

Erläutern Sie, was sie tun.

- Schlagen Sie zunächst unter dem Wirkstoff nach!
- Wählen Sie ein Präparat
- Schlagen Sie nach, ob **Phenoxymethylpenicillin** Wechselwirkungen mit Carbamazepin hat.
- Stellen Sie ein Rezept für **Phenoxymethylpenicillin** Tabletten auf

Patientendaten:

Manfred Meyer, Eichenweg 2, 37073 Göttingen, geb. 8.05 1952, BKK Heilberufe

Sie haben 5 Minuten Zeit.

Viel Erfolg!

8.16 Evaluationsbogen

Gesamtevaluation der OSCE

Diese Prüfung...	Stimme zu	Stimme teilweise zu	Stimme eher nicht zu	Stimme nicht zu
... ist geeignet, praktische Fertigkeiten zu überprüfen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... ist geeignet, theoretisches Wissen zu überprüfen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... erlaubt mir, meine eigene Leistung abzuschätzen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... hat einen positiven Lerneffekt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... hat mich zum Wiederholen der Kurs-/ Semesterthemen angeregt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... hat Spaß gemacht	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... hat mich verunsichert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... war atmosphärisch angespannt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
... erscheint mir fair / objektiv	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Das Niveau der Prüfung war	Zu hoch <input type="checkbox"/>	Gerade richtig <input type="checkbox"/>	Zu niedrig <input type="checkbox"/>	
Haben Sie schon vorher einmal an einer praktischen Prüfung wie der OSCE teilgenommen?	Ja <input type="checkbox"/>	Nein <input type="checkbox"/>		
Wie haben Sie sich auf die OSCE vorbereitet?	<input type="checkbox"/> gar nicht <input type="checkbox"/> nicht länger als 1h <input type="checkbox"/> 1-3h <input type="checkbox"/> mehr als 3h			
Welche Prüfungsform bevorzugen Sie? (Einfachauswahl)	<input type="checkbox"/> Multiple Choice <input type="checkbox"/> Sonstige schriftliche Prüfungen <input type="checkbox"/> Mündliche Prüfungen <input type="checkbox"/> Praktische Prüfungen (OSCE) <input type="checkbox"/> Hausarbeit / Referat			
Im Vergleich zu Ärzten bewerten Studenten als Prüfer	Eher besser <input type="checkbox"/>	Kein Unterschied <input type="checkbox"/>	Eher schlechter <input type="checkbox"/>	
Im Vergleich zu Ärzten bewerten Studenten als Prüfer	objektiver <input type="checkbox"/>	Kein Unterschied <input type="checkbox"/>	weniger objektiv <input type="checkbox"/>	
	Stimme zu	Stimme teilweise zu	Stimme eher nicht zu	Stimme nicht zu
Ich fühle mich durch die Studierenden angemessen bewertet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich würde mir selbst zutrauen jüngere Studierende in einer Prüfung zu bewerten	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Danksagung

Ich danke Herrn Prof. Dr. Michael M. Kochen, MPH, FRCGP und meinem Doktorvater Prof. Dr. Wolfgang Himmel für die freundliche Überlassung des Themas der Dissertation.

Mein ganz besonderer Dank gilt Prof. Himmel und Dr. Jean François Chenot für die ausgezeichnete Betreuung und die vielen wichtigen Anregungen. Ich bedanke mich ebenso bei der gesamten Abteilung Allgemeinmedizin, die mir bei den Organisationsfragen und bei der Durchführung der OSCE hilfsbereit zur Seite stand.

Ich möchte mich außerdem bei Dr. Thomas Kleinoeder aus der Abteilung Medizinische Informatik für die Erstellung der maschinenlesbaren Bewertungsbögen und bei der Abteilung Medizinische Statistik für die Beratung bedanken.

Lebenslauf

Ich, Alexandra Koch, wurde am 10.05.1982 in Bremerhaven geboren und bin in Schiffdorf (Niedersachsen) aufgewachsen. Ich bin das dritte Kind meiner Eltern, Herbert (01.10.1941) und Petra Koch, geb. Koch (28.11.1946), die mittlerweile beide pensionierte Lehrer sind. Meine Geschwister Stephanie Klingbeil, geb. Koch (11.02.1974), und Kerstin Grotheer, geb. Koch (27.05.1976), sind ebenfalls Lehrerinnen.

Ich wurde 1988 in die Grundschule Schiffdorf eingeschult, von 1992-1994 besuchte ich die Orientierungsstufe Schiffdorf und wechselte 1994 auf das Kreisgymnasium Wesermünde, wo ich meine Schullaufbahn im Sommer 2001 mit dem Abitur beendete.

Im Wintersemester 2001/02 begann mein Medizinstudium an der Georg-August-Universität Göttingen und im Sommersemester 2003 absolvierte ich dort das Physikum. Den klinischen Teil des Studiums verbrachte ich ebenfalls in Göttingen. Das praktische Jahr habe ich im Sommersemester 2006 begonnen und war für das Tertial Innere Medizin an den Asklepios Harzkliniken Goslar, das Chirurgie-Tertial verbrachte ich am Kantonsspital St. Gallen (Schweiz) und das Gynäkologie-Tertial absolvierte ich am Spitalzentrum Biel/Bienne (Schweiz). Im Frühjahr 2008 werde ich voraussichtlich mein zweites Staatsexamen absolvieren.

Von dem Wintersemester 2003/04 bis zum Wintersemester 2005/06 habe ich als studentische Hilfskraft im Kurs „Ärztliche Basisfertigkeiten“ für die Abteilung Allgemeinmedizin der Universität Göttingen gearbeitet.