

Aus der Abteilung Kardiologie und Pneumologie
(Prof. Dr. med. G. Hasenfuß)

im Zentrum Innere Medizin

der Medizinischen Fakultät der Universität Göttingen

Störgrößenanalyse der vergleichenden studentischen Selbsteinschätzung zur
Evaluation von Lehrveranstaltungen im Studium der Humanmedizin

INAUGURAL – DISSERTATION

zur Erlangung des Doktorgrades

der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von
Deborah Reinhardt

aus
Magdeburg

Göttingen 2013

Dekan: Prof. Dr. rer. nat. H. K. Kroemer

I. Berichtstatter: PD Dr. med. T. Raupach

II. Berichtstatter/in: Prof. Dr. Nicole Steinbüchel-Rheinwall

III. Berichtstatter/in: Prof. Dr. Martin Oppermann

Tag der mündlichen Prüfung: 01.07.2014

Die Ergebnisse dieser Arbeit wurden teilweise veröffentlicht in:

Schiekirka S., D. Reinhardt, T. Beißbarth, S. Anders, T. Pukrop, T. Raupach. "Estimating Learning Outcomes From Pre- and Posttest Student Self-Assessments: A Longitudinal Study." Acad Med, in press

und

Schiekirka S., D. Reinhardt, S. Heim, G. Fabry, T. Pukrop, S. Anders, T. Raupach (2012). "Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school." BMC Med Educ 12: 45

Inhaltsverzeichnis

Abbildungsverzeichnis.....	1
Tabellenverzeichnis.....	2
Abkürzungsverzeichnis.....	3
1 Einleitung.....	4
1.1 Evaluation der Hochschullehre.....	4
1.2 Evaluation im Studium der Humanmedizin.....	5
1.2.1 Lehre und Evaluation an der Universitätsmedizin Göttingen (UMG)	7
1.3 Störgrößen traditioneller Evaluationsinstrumente	8
1.4 Lernerfolg als Zielgröße der Evaluation.....	10
1.5 Abschätzung des Lernerfolgs durch studentische Selbsteinschätzungen.....	12
1.5.1 Das Konzept der vergleichenden Selbsteinschätzung.....	13
1.5.1.1 Vergleichende Selbsteinschätzung an der Universitätsmedizin Göttingen	14
1.6 Mögliche Störfaktoren der vergleichenden Selbsteinschätzung.....	16
1.7 Ziele dieser Arbeit.....	18
2 Methoden.....	19
2.1 Studiendesign.....	19
2.2 Probandenrekrutierung.....	19
2.3 Fokusgruppengespräche	20
2.4 Selbsteinschätzungen	20
2.4.1 Vorbefragung.....	20
2.4.2 Nachbefragung: retro-/ prospektiv	21
2.5 Objektive Leistungsparameter	21
2.6 Datensammlung	22
2.7 Datenanalyse.....	23
2.7.1 Berechnung der Lernzuwachsrate.....	23
2.7.2 Fokusgruppengespräche (Studienfrage 1)	24
2.7.3 Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (<i>individual-level</i> vs. <i>group-level</i>) (Studienfrage 2).....	24
2.7.4 Prospektiv vs. retrospektive Erhebung der initialen Selbsteinschätzungen (Studienfrage 3).....	25
2.7.5 Einfluss der Anzahl der Lernziele auf die Reliabilität der Lernzuwachsrechnung aus vergleichenden studentischen Selbsteinschätzungen (Studienfrage 4)	26
2.7.6 Einfluss des Selbstwertgefühls auf studentische Selbsteinschätzungen (Studienfrage 5)	26

2.7.7	Kriteriumsvalidität, Cut-Off für erfolgreich bzw. erfolglos gelehrt Inhalte (Studienfrage 6).....	26
2.8	Statistische Auswertung.....	28
3	Darstellung der eigenen Untersuchungen („Ergebnisse“).....	29
3.1	Fokusgruppengespräche (Studienfrage 1).....	29
3.2	Deskriptive Statistiken der Stichprobe und Rücklaufquote.....	31
3.3	Darstellung der Lernzuwachsdaten.....	31
3.4	Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (<i>individual-level</i> vs. <i>group-level</i>) (Studienfrage 2).....	36
3.5	Prospektive versus retrospektive Selbsteinschätzung des initialen Lernzuwachses (Studienfrage 3).....	38
3.6	Einfluss der Anzahl der Lernziele auf die Reliabilität der Lernzuwachsrechnung aus vergleichenden studentischen Selbsteinschätzungen (Studienfrage 4).....	39
3.7	Einfluss des Selbstwertgefühls auf die studentischen Selbsteinschätzungen (Studienfrage 5).....	41
3.8	Darstellung der Testleistungen.....	45
3.9	Kriteriumsvalidität (Studienfrage 6).....	48
3.9.1	Korrelation des Lernzuwachses errechnet aus Selbsteinschätzungen und Testatergebnissen auf <i>group-level</i>	48
3.9.2	Korrelation des Lernzuwachses errechnet aus Selbsteinschätzungen und Testatergebnissen auf <i>individual-level</i>	49
3.9.3	ROC-Analyse, Cut-Off-Wert.....	51
4	Besprechung der Ergebnisse im Zusammenhang mit den Angaben in der Literatur („Diskussion“).....	52
4.1	Wesentliche Ergebnisse der hier vorgestellten Studie.....	52
4.2	Wesentliche Limitationen der hier vorgestellten Studie.....	52
4.2.1	Limitationen der qualitativen Studie.....	52
4.2.2	Limitationen der quantitativen Studie.....	53
4.3	Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (Studienfrage 2).....	54
4.4	Pro-/Retrospektive Einschätzung (Studienfrage 3).....	55
4.5	Anzahl der Lernziele (Studienfrage 4).....	56
4.6	Selbstwertgefühl (Studienfrage 5).....	57
4.7	Kriteriumsvalidität (Studienfrage 6).....	58
4.8	Ausblick auf künftige Forschungsprojekte.....	59
5	Zusammenfassung.....	61
6	Anhang.....	63

6.1	Evaluationsbogen der UMG aus dem Wintersemester 06/07	63
6.2	Probandeninformation	66
6.3	Einverständniserklärung.....	68
6.4	Rosenbergskala.....	69
6.5	Prä-Evaluationsbogen.....	70
6.6	Post-Evaluationsbogen	75
7	Literaturverzeichnis.....	82

Abbildungsverzeichnis

Abbildung 1: Mittelwerte der Selbsteinschätzungen aller betrachteten Lernziele mit Standardabweichung.....	32
Abbildung 2: Mittelwerte der Lernzuwachsrate berechnet aus den pro Lernziel aggregierten studentischen Selbsteinschätzungen (sALZ) vor und nach dem Modul	34
Abbildung 3: Bland-Altman-Plot der Lernzuwachsrate auf sALZ- und oALZ-Niveau	37
Abbildung 4: Bland-Altman-Plot der Lernzuwachsrate berechnet mit prospektiv und retrospektiv erfassten Selbsteinschätzungen zu Modulbeginn sowie Selbsteinschätzungen zu Modulende	38
Abbildung 5: Darstellung der Schwankungsbreite je nach Anzahl der erfassten Lernziele mittels der sALZ-Werte	39
Abbildung 6: Darstellung der Mittelwerte der Testatergebnisse sowie der Standardabweichung pro Lernziel vor und nach dem Modul basierend auf den Testatergebnissen (oALZ)	45
Abbildung 7: Darstellung der Lernzuwachsrate pro Lernziel vor und nach dem Modul basierend auf den Testatergebnissen (oALZ)	46
Abbildung 8: Bland-Altman-Plot der sALZ- und oALZ-Werte (Korrelation zwischen dem prospektivem Lernzuwachs, errechnet aus Selbsteinschätzungen (sALZ), und dem prospektivem Lernzuwachs, errechnet aus Testatergebnissen (oALZ)).....	47
Abbildung 9: Darstellung der Korrelation zwischen subjektiv und objektiv eingeschätztem Lernzuwachs berechnet auf <i>individual-level</i> -Niveau	48
Abbildung 10: ROC-Analyse	49

Tabellenverzeichnis

Tabelle 1: Abkürzungsverzeichnis der möglichen Berechnungsarten des Lernzuwachses	23
Tabelle 2: Kernaussagen der Fokusgruppenteilnehmer	29
Tabelle 3: Lernzuwächse bezogen auf die im Testat verwendeten Lernziele – <i>group-level</i> (sALZ) und <i>individual-level</i> (sILZ)	36
Tabelle 4: Quartile nach Selbstwertgefühl	40
Tabelle 5: Selbsteinschätzungen gemittelt über die Selbstwert-Quartile pro Lernziel; Einschätzung zu Modulbeginn	41
Tabelle 6: Selbsteinschätzungen gemittelt über die Selbstwert-Quartile pro Lernziel; Einschätzung zu Modulende	43
Tabelle 7: Korrelationen der Lernzuwächse der vier Quartile	44
Tabelle 8: Mittelwerte des Selbstwertgefühls	55

Abkürzungsverzeichnis

Allgemeine Abkürzungen

bzw.	beziehungsweise
ca.	circa
evtl.	eventuell
ggf.	gegebenenfalls
MW	Mittelwert
SD	Standardabweichung
vs.	versus
z.B.	zum Beispiel

Spezielle Abkürzungen

oALZ	objektiver aggregierter Lernzuwachs
oILZ	objektiver individueller Lernzuwachs
sALZ	subjektiver aggregierter Lernzuwachs
sILZ	subjektiver individueller Lernzuwachs
UMG	Universitätsmedizin Göttingen

1 Einleitung

1.1 Evaluation der Hochschullehre

Evaluation beinhaltet die Erstellung, Auswertung und Interpretation von Informationen über die „Wirkung von ... Curriculumprojekten“ (Schaub and Zenke 1997), hier der Hochschullehre. Sie ist ein wesentlicher Bestandteil in deren Qualitätssicherung.

In seinen "Empfehlungen zur Qualitätsverbesserung von Lehre und Studium" hielt der Wissenschaftsrat im Jahr 2008 fest, dass Evaluationen auf der Grundlage studentischer Einschätzungen dazu dienen können, den Bedarf an Lehrangeboten zu ermitteln und – vor allem – Veränderungen in der Lehrpraxis herbeizuführen. Jedoch seien die häufig durchgeführten studentischen Lehrevaluationen bisher nur wenig standardisiert (Wissenschaftsrat 2008), was die Vergleichbarkeit erschwert. Zudem können sich Fehler ergeben, die die Evaluationsergebnisse verfälschen und zu nicht korrekten Schlussfolgerungen führen. Eine wissenschaftliche Überprüfung der eingesetzten Evaluationsmethoden scheint hier angebracht.

Nationalen und internationalen Vergleichen können Hochschulen heute nur standhalten, wenn sie in der Lage sind, die Qualität ihrer Lehre nachzuweisen und stetig zu verbessern. Dies bedeutet, dass die Hochschulen ein Instrument benötigen, das die Qualität ihrer Lehre realistisch widerspiegelt, um Schwächen zu erkennen, zu beheben und sie in Stärken zu verwandeln.

Welche Konsequenzen die einzelnen Hochschulen aus den durch die Evaluation gewonnenen Erkenntnissen ziehen, variiert stark. Öffentliche Konsequenzen ergeben sich zum Beispiel im Hochschul-Ranking, in dem Hochschulen bewertet werden und somit an Attraktivität für potentielle Studierende gewinnen oder verlieren. Auch die Orientierung der leistungsorientierten Mittelvergabe (LOM) an Evaluationsergebnissen ist denkbar.

Wie oben ausgeführt können Evaluationen die Grundlage zur Verbesserung der Lehre bilden. Zu den wesentlichen Voraussetzungen für diese Funktion von Evaluationen gehören die Reliabilität und Validität der erhobenen Daten. Die Reliabilität stellt hierbei die Zuverlässigkeit eines Erhebungsinstruments dar, ein Merkmal, das angibt, wie stabil bei wiederholter Anwendung das gleiche Ergebnis erreicht wird. Die Validitätsprüfung dagegen zeigt, ob ein Instrument auch das misst, was es zu messen angibt, also im Fall der Lehrevaluation die Qualität der Lehre.

Das Evaluationsinstrument ist abhängig vom jeweilig gelehrten Fachgebiet. Unter anderem werden Multiple-Choice-Fragebögen, offene Fragebögen, Dozentenfeedback etc. eingesetzt. Je nachdem, welche Fähigkeiten in einem bestimmten Studiengang gefördert werden, muss die Art der Evaluation auf die Lerninhalte abgestimmt sein.

Die Evaluation nimmt eine zentrale Stellung innerhalb eines jeden Lehrcurriculums ein. Zur Illustration dieses Sachverhalts soll an dieser Stelle kurz auf das sechsstufige Konzept der Curriculumentwicklung nach Kern eingegangen werden (Kern et al. 1998): Auf die allgemeine und lernerzentrierte Bedarfsanalyse (Schritt 1 und 2) folgt die Definition geeigneter Lernziele (Schritt 3). Anschließend müssen Lehrformate identifiziert werden, die sich zu den angestrebten Lernzielen kongruent verhalten (Schritt 4). Auf die Implementierung, d.h. die praktische Umsetzung des Curriculums (Schritt 5), folgt schließlich die Evaluation (Schritt 6), deren Ergebnisse in die

kontinuierliche Bedarfsanalyse eingehen, wodurch sich der Kern'sche Zyklus schließt. Auf diese Weise wird eine immerwährende Weiterentwicklung und Anpassung der Lehre an die aktuellen Gegebenheiten ermöglicht. Die von Kern in diesem Entwicklungsprozess geforderte Kongruenz zwischen Lernzielen, Lehr- und Prüfungsformaten konnte durch die Einführung innovativer Lehr- (Schmidt 1983) und Prüfungsmethoden (Harden et al. 1975) in den vergangenen vier Dekaden deutlich verbessert werden. Viele aktuell genutzte Evaluationsverfahren weisen jedoch keine ausreichende Kongruenz zu den anderen Aspekten der betrachteten Curricula auf [Raupach et al. 2011; Raupach et al. GMS Z 2012]. Diesem Mangel sollte mit einem neuen Evaluationsinstrument begegnet werden, das in vorliegender Studie untersucht wurde.

1.2 Evaluation im Studium der Humanmedizin

Um Evaluationsergebnisse bewerten zu können, muss zunächst festgelegt werden, welches Konstrukt mit einem Evaluationsinstrument erfasst wird. Wenn Evaluation zur Abschätzung der Qualität im Medizinstudium dienen soll, müssen demzufolge entsprechende Qualitätsindikatoren definiert werden. Zu diesen können einerseits Strukturen und Prozesse im Studienablauf zählen, andererseits kann auch das Studienergebnis ("Outcome") oder die Leistung eines/r einzelnen Dozenten/in als Zielgröße, das zu untersuchende Konstrukt, definiert werden.

Die Wahl des Evaluationsinstruments ist davon abhängig, welches Konstrukt abgebildet werden soll. Beispiele für Zielgrößen, die an deutschen medizinischen Fakultäten verwendet werden, finden sich im „Positionspapier der nordrhein-westfälischen Studiendekane zur hochschulvergleichenden leistungsorientierten Mittelvergabe für die Lehre“ (Herzig et al. 2007), das auch einen Algorithmus für die leistungsorientierte Mittelvergabe (LOM) in der Lehre vorschlägt. In diesem Artikel beschreiben die Autoren, dass die LOM-Gelder in der Vergangenheit anhand der Ergebnisse der Staatsexamina verteilt wurden, was nach der Änderung der ärztlichen Approbationsordnung 2003 nun nicht mehr sinnvoll erscheint. Es sind durch einen Arbeitskreis, bestehend aus Vertretern des Ministeriums für Wissenschaft und Forschung sowie Angehörigen der beteiligten Fakultäten, neue Kriterien für eine Verteilung der Mittel erarbeitet worden. Der Arbeitskreis hat dabei drei Qualitätsebenen berücksichtigt: die Struktur-, die Prozess- und die Ergebnisqualität. Die Struktur wird hierbei durch materielle und personelle Ressourcen sowie gesetzliche und organisatorische Regelungen umschrieben. Der Prozess beinhaltet die Gestaltung von Lehrveranstaltungen und Prüfungen. Das Ergebnis umfasst den Wissenserwerb bzw. die Berufsfähigkeit.

Die Strukturqualität wird von vielen herkömmlichen Evaluations-Fragebögen erfasst. Da die Prozessqualität nicht unwesentlich durch die Aktivitäten der Dozenten beeinflusst wird, empfiehlt es sich, im Rahmen von Evaluationen auch die Leistung der Dozenten beurteilen zu lassen. Als Parameter für die Messung der Ergebnisqualität schlagen die Autoren studentische Leistungen in Staatsexamina vor (Herzig et al. 2007).

Hier ergibt sich die Frage, ob Punktzahlen und Noten in Staatsexamina das Studienergebnis tatsächlich valide abbilden. Es erscheint sinnvoll, als wesentliches „Ergebnis“ des Medizinstudiums die Befähigung zur Ausübung der ärztlichen Tätigkeit zu definieren. Entsprechend sollte ein Evaluationsinstrument genau diese Befähigung – oder ihre Teilaspekte – valide erfassen. Die in der Medizin relevanten Lernziele beinhalten drei verschiedene Dimensionen. Im Vordergrund stehen

zunächst die kognitiven Lernziele. Psychomotorische Lernziele nehmen im Verlauf des Studiums an Wichtigkeit zu. Sie beinhalten die praktischen Fertigkeiten des späteren Berufsalltags. Zuletzt sind noch die affektiven Lernziele zu berücksichtigen, die eine persönliche Einstellung und Wertvorstellungen umfassen.

Alle drei Dimensionen sollten in einer Evaluation erfasst werden.

In dem Bestreben, aktuelle Evaluationspraktiken an deutschen medizinischen Fakultäten zu erfassen, führte der Medizinische Fakultätentag eine Befragung der Lehrverantwortlichen durch. Hiernach wurden im Februar 2012 an mindestens 33 der 36 medizinischen Fakultäten Evaluationen durchgeführt, jedoch in unterschiedlicher Art und Weise. Sowohl die Fragenart und -anzahl der Evaluationen, als auch die dafür verantwortlichen Ämter oder Gremien variierten. So ist zum Beispiel in Bochum die Evaluationskommission der Fakultät zuständig, in Erlangen-Nürnberg dagegen das Studiendekanat.

Eine semesterbezogene Evaluation wurde an 2/3 der deutschen Universitäten durchgeführt. Einige Universitäten, wie zum Beispiel München, ließen die Studenten wöchentlich evaluieren; in Einzelfällen ist die Evaluation nur jährlich erhoben worden. Häufig werden Globalratings mit Fragebögen auf der Grundlage von Schulnotenskalen verwendet.

Evaluationsfragen zu einzelnen Veranstaltungen bzw. Dozenten wurden an knapp 1/3 (11 von 36) der deutschen Universitäten regelhaft erhoben, mehrfach bestand jedoch die Möglichkeit der freiwilligen Evaluation, wenn der Dozent dies wünschte (MFT 2012).

Zwei deutsche Universitäten bezogen bisher die Ergebnisse des Progress-Tests Medizin in ihre Evaluationsergebnisse ein. Hierbei handelt es sich um einen nach jedem Semester durchgeführten Test auf Absolventen-Niveau, der den kontinuierlichen Wissenszuwachs während des Studiums abbildet. Der Test besteht aus 200 Multiple-Choice-Fragen mit Fragestellungen aus allen medizinischen Fachgebieten (Osterberg et al. 2006). Auch Klausuren bzw. Prüfungsergebnisse fanden Eingang in die Evaluation der Lehre. An der Universität in Frankfurt am Main zum Beispiel stellten Prüfungsergebnisse des klinischen Abschnittes knapp ein Viertel des Evaluationsergebnisses der medizinischen Fakultät dar.

Ebenso wie die oben beschriebene Durchführung sind auch die sich ergebenden Konsequenzen der Evaluationsergebnisse sehr uneinheitlich. Sie erstrecken sich von Einzelgesprächen bis hin zur Vergabe von Geldpreisen. In Erlangen-Nürnberg erhielten die drei besten Dozenten klinikgebunden bis zu 30.000 Euro. In Rostock würden Konsequenzen aus den Evaluationsergebnissen durch Gespräche mit den Fachbereichen gezogen (MFT 2012).

Zusammenfassend ist trotz aktuell bestehender Differenzen in der praktischen Durchführung festzustellen, dass das Studium der Humanmedizin hinsichtlich der zugrundeliegenden Strukturen, der darin ablaufenden Prozesse, des Lernerfolgs und der unterrichtenden Dozenten/innen evaluiert werden kann. Ziel der Lehrevaluation ist letztlich eine Beurteilung und Verbesserung der Lehrqualität. Da das Studienergebnis gerade im Studiengang Humanmedizin von außerordentlicher Bedeutung ist, erscheint es gerechtfertigt, diesem Aspekt (im Vergleich zur Struktur- und Prozessqualität) bei der Bewertung von Lehrveranstaltungen besonders großes Gewicht zu verleihen.

1.2.1 Lehre und Evaluation an der Universitätsmedizin Göttingen (UMG)

An der medizinischen Fakultät der Universität Göttingen wird die Lehre im klinischen Abschnitt in einem Modulsystem umgesetzt. Die Module dauern zwischen 2 und 7 Wochen. Pro Semester finden 3-4 Module statt. Jedes Modul hat einen thematischen Schwerpunkt, zum Beispiel die Erkrankungen des Herz-Kreislauf-Systems und der Lunge (Modul 3.1). Dieser Schwerpunkt bestimmt, welche Fachdisziplinen und Querschnittsbereiche in einem Modul vertreten sind. So kann ein Fach, zum Beispiel die allgemeine Pathologie, in nur einem Modul vertreten sein. In diesem Modul werden dann alle Inhalte des betreffenden Faches gelehrt, es wird später thematisch nicht wieder aufgegriffen. Andere Fächer, wie zum Beispiel die Hygiene, Chirurgie oder spezielle Pathologie, sind an vielen Modulen beteiligt, da ihre Inhalte für verschiedene Themen von Bedeutung sind.

Für jedes Modul wurden die zu unterrichtenden Inhalte in Form von Lernzielen festgelegt. Der daraus entstandene Lernzielkatalog (http://www.med.uni-goettingen.de/de/media/G1-2_lehre/lernzielkatalog.pdf) beschreibt das Wissen, die Fertigkeiten und die professionellen Qualitäten, die die Studierenden im Laufe des Studiums erwerben sollen. Die hier angegebenen Lernziele sind jedoch nicht operationalisiert, sondern stellen lediglich Stichpunkte dar (z.B.: „Akuter Myokardinfarkt“).

Der Göttinger Lernzielkatalog ist in vier Teile unterteilt. Im ersten Teil sind 89 Symptome und Befunde angegeben, mit deren Differentialdiagnostik Studierende am Ende des Studiums vertraut sein müssen. Insofern sind hier sowohl kognitive Inhalte als auch praktische Fertigkeiten abgebildet. In Teil Zwei sind Gesundheitsstörungen aufgelistet, deren Ätiologie, Pathogenese, Symptomatik, Diagnostik und Therapie von Absolventen auf unterschiedlichen Kompetenzniveaus beherrscht werden muss; insofern bildet der zweite Katalogteil in erster Linie Faktenwissen ab.

Bei der Konstruktion des Kataloges wurden Prioritäten bezüglich häufiger und wichtiger Erkrankungen gesetzt. Diese Priorisierung wird in Form von Kompetenzniveaus angegeben. Für Erkrankungen, die auf dem Kompetenzniveau 1 verankert sind, müssen Studierende über ein umfassendes Wissen verfügen. Erkrankungen, die mit Kompetenzniveau 2 beschrieben werden, erfordern dagegen Grundkenntnisse. Solche mit Kompetenzniveau 3 sollen nur erkannt werden (Göttinger-Lernzielkatalog 2008). Der dritte Katalogteil umfasst die affektiven Lernziele, zu denen zusätzliche grundlegende Kenntnisse, Einstellungen und Methoden gehören. Im vierten Teil sind die ärztlichen Basisfertigkeiten beschrieben.

Grob vereinfacht werden die Inhalte des dritten Katalogteils zusammen mit den ärztlichen Basisfertigkeiten (Katalogteil Vier) in den ersten beiden klinischen Semestern gelehrt; die umfangreichen Inhalte des ersten und zweiten Teils bilden die Grundlage der Ausbildung in den klinischen Semestern 3 bis 5. Im letzten Semester des klinischen Studienabschnitts findet eine symptombezogene Lehre statt, so dass hier nochmals vor allem Inhalte des ersten Katalogteils betont werden. Praktische Fertigkeiten (Katalogteil Vier) werden in den Blockpraktika und im Praktischen Jahr trainiert. Das Curriculum ist so aufgebaut, dass spezifische Inhalte nicht nur einmal, sondern in wachsender Komplexität mehrfach angesprochen werden. So werden die Grundlagen der Pathophysiologie des Herz-Kreislauf-Systems bereits im ersten klinischen Semester diskutiert. Hier erlernen die Studierenden auch basale Untersuchungstechniken wie die Auskultation des Herzens. Eine detailliertere Auseinandersetzung mit den Erkrankungen des Herz-Kreislauf-Systems erfolgt dann im dritten klinischen Semester (Modul 3.1), und entsprechende praktische Fertigkeiten werden im anschließenden Blockpraktikum trainiert. Im sechsten Semester werden die bereits behandelten Erkrankungen des Herz-Kreislauf-Systems im Kontext anderer Erkrankungen und ausgehend von

Schlüsselsymptomen wiederholt. Dieses wiederholte Ansprechen gleicher Inhalte in wachsender Komplexität ist ein wesentliches Merkmal von Spiralcurricula (Harden 1999). Dieser Aspekt wird an dieser Stelle betont, weil bei der Beurteilung des Lernerfolgs der jeweilige Leistungsstand der Studierenden zu Beginn eines klinischen Moduls berücksichtigt werden muss (siehe Kapitel 1.5.1, Seite 13).

In der Vergangenheit wurde die Lehre an der Universitätsmedizin Göttingen mit Hilfe eines von den Studierenden ausgefüllten Fragebogens (siehe 6.1: Evaluationsbogen der UMG aus dem Wintersemester 06/07, Seite 65) evaluiert. Die Zufriedenheit der Studierenden mit den ihnen angebotenen Lehrveranstaltungen wurde so erfasst. Die Bewertung erfolgte anhand eines Schulnotenprinzips. Hierbei spielten sowohl die Organisation der verschiedenen Lehrveranstaltungsformen als auch die Frage, inwieweit diese zum Wissenszuwachs beigetragen haben, eine Rolle. Erfasst wurden somit nur die Struktur- und Prozessqualität. Die Ergebnisqualität konnte in der Vergangenheit nicht widerspiegelt und somit keine Rückschlüsse auf den Lernerfolg gezogen werden.

1.3 Störgrößen traditioneller Evaluationsinstrumente

Eine Problematik in der Evaluation besteht darin, dass die Validität der erhobenen Daten im Kontext des zugrundeliegenden Konstrukts beurteilt werden muss. Jedes Instrument misst etwas, aber es ist entscheidend, die richtige Zielgröße zu erfassen. Bei der Interpretation der Evaluationsdaten muss man sich des zugrundeliegenden Konstrukts bewusst sein, um zu erfassen, was gemessen wurde, und dann zu hinterfragen, ob dieses Messergebnis auch die ursprüngliche Fragestellung bearbeitet. Zum Beispiel muss definiert sein, inwiefern Lehrqualität durch die Strukturqualität dargestellt wird. Wäre eine gute Struktur der Lehre mit einer hohen Lehrqualität gleichzusetzen, so würden Fragebögen, die nur die Struktur erfassen, völlig ausreichen und die traditionelle Lehrevaluation könnte einfach weitergeführt werden wie bisher. Da jedoch auch andere Dimensionen wie die Prozess- und Ergebnisqualität in die Bewertung der Lehre einfließen, ist die Entwicklung eines Evaluationsinstruments, das auch diese Aspekte erfasst, durchaus sinnvoll.

Jede Messung ist Einflussgrößen, also Faktoren, die das Messergebnis beeinflussen, unterworfen. Ist diese Einflussgröße gewollt und in die Überlegungen/ Berechnungen miteinbezogen, so kann das Ergebnis valide sein. Wird die Evaluation dagegen durch einen Faktor beeinflusst, der das Ergebnis unbeabsichtigt, aber systematisch verändert oder die Fragestellung modifiziert, so ist von einer Störgröße zu sprechen. Störgrößen können folglich als Faktoren verstanden werden, die die studentische Evaluation beeinflussen, aber nicht im Zusammenhang mit der Lehreffektivität stehen (Marsh 1984). Um die Einflussgrößen von den Störgrößen trennen zu können, ist es notwendig, das Konstrukt der Evaluation genau zu kennen. Ein möglichst weitgehendes Ausschalten von Störgrößen ist für die Konstruktvalidität eines Evaluationsinstruments notwendig. Es wird also eine Messung benötigt, die weder durch systematische Fehler noch durch andere Konstrukte verfälscht wird.

In der Literatur wird auf verschiedene mögliche Einflussfaktoren auf Evaluationen hingewiesen. Zum Beispiel scheint das Interesse der Studierenden am Thema der Veranstaltung Einfluss auf die

Bewertung der Lehrveranstaltung zu haben (Spiel and Gössler 2000). Weiterhin wurde beobachtet, dass die Bewertungen der Lehreffektivität verschiedener Unterrichtsformen bei Lehrern und Studierenden gegebenenfalls nicht übereinstimmen. So präferieren Studierende Veranstaltungen, die zu einem passiven „Konsum“ der dargebotenen Inhalte einladen (zum Beispiel Vorlesungen), während es aus didaktischer Sicht sinnvoller wäre, aktivierende Methoden einzusetzen – diese jedoch werden von Studierenden nicht immer positiv bewertet (McKeachie 1997).

Traditionelle Evaluationsinstrumente weisen viele Einflussfaktoren auf. Wie oben bereits erwähnt, stellen Fragebögen auf der Grundlage von Schulnotenskalen eine zeitsparende Evaluationsmöglichkeit dar. Diese Messgrößen sind jedoch sehr störanfällig, und die von ihnen generierten Daten weisen eine hohe Varianz auf (Greenwald 1997; Germain and Scandura 2005).

Viele traditionelle Evaluationsformen scheinen bei der Bestimmung der Lehrqualität eher die Zufriedenheit der Studierenden mit zum Beispiel Organisation und Struktur als wirkliches Wissen oder Fähigkeiten zu messen (Theall and Franklin 2001; Braun and Leidner 2009). Auch Sympathie und organisatorische Aspekte in der Lehre können eine große Rolle in der Bewertung von Lehrveranstaltungen spielen (Gibson et al. 2008). Weiterhin kann der Ruf des Dozenten einen Einfluss auf die Evaluation haben. Studierende, die vor dem Kurs etwas Positives über den Dozenten gehört hatten, schätzten diesen im Mittel besser ein. Dieses Phänomen konnte von Griffin in einer Studie beobachtet werden. Die Studenten wurden hierbei je nach Vorinformation über den Dozenten in drei Gruppen eingeteilt – keine, positive oder negative Informationen. Diejenigen Studenten die vorab positive Informationen erhielten, schätzen Kurs und Dozenten positiver ein als die Vergleichsgruppen (Griffin 2001). Ebenfalls kann die Motivation für einen Kurs als Variable in der Evaluation verstanden werden (Howard and Maxwell 1980; Cashin 1989).

Der Dr. Fox-Effekt, die Reflektion von Charakteristika der Lehrer, besagt, dass Studierende zum Beispiel eine günstigere Evaluation aufgrund des Enthusiasmus des Lehrers abgeben, auch wenn fast nichts gelernt wurde. Dieser wurde in Evaluationen von verschiedenen Forschern beobachtet (Marsh and Roche 1997; Germain and Scandura 2005). Zudem konnte Germain bei der Auswertung verschiedener Studien beobachten, dass die Notengebung eine Rolle für die Ergebnisse von Lehrevaluation spielt. Es wird eine wechselseitige Beziehung angenommen (Germain and Scandura 2005).

Marsh stellte zudem fest, dass allgemeine Items in einer Evaluation der Multidimensionalität der Lehre nicht gerecht werden und vermutlich anfälliger für Biasfaktoren sind (Marsh 1984). In zwei Studien konnte Marsh 4 Hintergrundvariablen als Haupterklärungspunkte einer hohen Varianz in den Evaluationen herausfiltern: Interesse, bessere erwartete Noten, Arbeitsaufwand, höherer Prozentsatz an interessierten Studierenden (Marsh and Roche 1997; McKeachie 1997) Eine weitere Biasvariable könnte die Begeisterung des Lehrenden (Naftulin et al. 1973) sein. Ein Einfluss des Arbeitsaufwandes, je mehr Aufwand, desto schlechter die Bewertung in der Evaluation, wird auch von Greenwald und Gillmore vermutet (Greenwald 1997).

Traditionelle Evaluationen in Form von globalen Fragebögen sind ebenfalls Verzerrungen unterworfen. Es ist bei der Verwendung dieser Fragebögen festzustellen, dass durchaus etwas gemessen wird. Es ist jedoch nicht zuverlässig möglich, aus ihren Ergebnissen auf die Ergebnisqualität der Lehre zu schließen, da diese nicht zwangsläufig und ausschließlich erfasst wird. Zu viele Biasvariablen können die Antworten verzerren und Antworten auf Fragen liefern, die in der

Evaluation gar nicht gestellt wurden. Die verwendeten Bögen sind meist nicht auf Lernziele bezogen, sodass eine inhaltsbezogene Datenerhebung nicht zu erwarten ist. Folglich liegt traditionellen Evaluationsinstrumenten am ehesten nicht das Konstrukt „Ergebnisqualität“ zugrunde. Aus oben genannten Gründen erscheint es sinnvoll, diesen Aspekt der Evaluation stärker zu betonen.

Es ist nicht sicher, ob die Studierenden bei der Beantwortung von Evaluationsfragen die gleichen Absichten verfolgen wie diejenigen, die diese Fragen gestellt haben. Nur wenige Studien haben sich bisher mit den gedanklichen Hintergründen der Studierenden bei der Teilnahme an einer Evaluation beschäftigt. Eine dieser Studien wurde von Billings -Gagliardi et al. durchgeführt. Medizinstudenten/innen wurden aufgefordert, ihre Gedanken während des Ausfüllens eines typischen Evaluationsfragebogens auszusprechen. Es zeigten sich bei den Studierenden andere Prioritäten als bei den Auswertern des Fragebogens (Billings-Gagliardi et al. 2004). Eine nähere Betrachtung der studentischen Hintergründe scheint hier angebracht.

Welche Parameter letztlich als Störgrößen aufgefasst werden, hängt vom Konstrukt des Evaluationsinstruments ab. Es ist aber auch von großer Bedeutung, wie das Konstrukt von den evaluierenden Studierenden aufgefasst wird. Gegebenenfalls können bei gleichem Verständnis des Konstrukts Störfaktoren minimiert werden. Eine Erfassung der studentischen Sicht zur Lehrevaluation und ein besseres Verständnis dessen, was Studierende unter der Evaluation verstehen, tragen möglicherweise zur Minimierung von Störungen bei.

1.4 Lernerfolg als Zielgröße der Evaluation

Zur Messung des Lernerfolgs ist es im Prinzip möglich, Klausur- oder Examensdaten zu verwenden. Doch dieses Vorgehen ist – zumindest auf der Ebene einzelner Fakultäten – sehr aufwendig und häufig wenig valide. Eine Umfrage im Herbst 2007 an allen deutschen medizinischen Fakultäten ergab, dass nur in ca. 40% der Fälle ein Aufgabenreview vor der Prüfung bzw. eine statistische Ergebnisauswertung vorgenommen wurde. Die Maßnahmen zur Qualitätssicherung von Leistungserfassungen müssen an verschiedenen Fakultäten noch gefördert werden (Möltner 2010). Die Überprüfung der Qualität der Lehre anhand von Klausuren enthält viele Fehlermöglichkeiten. Examensnoten zum Beispiel entsprechen nicht unbedingt den wirklichen Fähigkeiten der Studierenden (Spiel et al. 2006). Wenn Prüfungsergebnisse trotzdem genutzt werden sollen, um den Leistungsstand von Studierenden zu bewerten, müssten jedoch zwei Tests durchgeführt werden – jeweils ein Test vor und ein Test nach der Absolvierung einer Lehrveranstaltung. Nur dann könnte der Wissenszuwachs erfasst werden. Dieser Prüfungsaufwand ist jedoch kaum durchführbar. Zudem entsprechen fakultätsinterne Prüfungen den aktuellen Erkenntnissen zufolge nicht den internationalen Standards (Möltner 2010). Hinzu kommt, dass nur wenige Prüfungen alle drei Domänen des Medizinstudiums (kognitives Wissen, praktische Fertigkeiten und professionelle Einstellungen) abbilden können. Aus diesen Gründen sind die Ergebnisse der Modul- bzw. Kursabschlussprüfungen zur Beurteilung des multidimensionalen Lernzuwachses von Studierenden während eines Moduls bzw. Kurses ungeeignet.

Die hier angesprochene Problematik soll am Beispiel von Multiple-Choice-Fragen (MC) erläutert werden: Bei dieser Prüfungsform werden den Studierenden verschiedene Antwortmöglichkeiten präsentiert, von denen in der Regel nur eine richtig ist. Der Vorteil dieser Methode liegt in der einfachen und objektiven Auswertung, die bei entsprechenden Formularen und der passenden Software maschinell vorgenommen werden kann. Durch den geringen Aufwand ist zwar ggf. das Qualitätskriterium (Schuwirth and van der Vleuten 2004) der Kosteneffektivität erfüllt, alle weiteren Kriterien jedoch nicht unbedingt. Nachteilig wirkt sich aus, dass in vielen Fällen nur Faktenwissen und keine komplexen kognitiven Fertigkeiten geprüft werden.

Das taxonomische Niveau ist in der Regel bei MC-Fragen eher niedrig. Nach der Lernpyramide nach Miller wird nur das kognitive Wissen abgefragt, die Verhaltensebene die gleichwohl auch zum Kompetenzerwerb gehört, bleibt unerreicht (Miller and Geraci 2011). Nach Krathwohl wird nur die erste Ebene „Erinnerung“, evtl. noch die zweite Ebene „Verstehen“, einbezogen. Die dritte Ebene „Anwendung“ fließt meist nicht in MC-Fragen ein. Es zeigt sich somit, dass hier nur ein Teil der für die Medizin relevanten Kompetenzen geprüft werden kann (Krathwohl 2002).

Die Testweisheit, zu der nach Rogers zum Beispiel Zeiteinteilungs-, Fehlervermeidungs- und Schlussfolgerungsstrategien gehören (Rogers 1996), ist zudem trainierbar und kann helfen, die richtige Antwort angeben zu können, auch ohne die Lösung zu wissen. Multiple-Choice-Fragen stellen so in der Wissensüberprüfung eine Variante dar, die mit vielen potentiellen Fehlern behaftet ist (Shumway and Harden 2003; Epstein 2007). Es ist auch bei Nicht-Wissen bei zum Beispiel 5 Antwortmöglichkeiten mit einer Wahrscheinlichkeit von 20% eine richtige Antwort durch einfaches Raten möglich (Möltner 2006). Aus dem Ergebnis einer einzigen Prüfung am Kursende kann somit nicht valide auf den Kompetenzgewinn im Laufe des Kurses geschlossen werden.

Neben den möglichen Fehlerquellen der verschiedenen Leistungsüberprüfungsmethoden müssen zusätzlich noch weitere Aspekte bedacht werden. Es sollte berücksichtigt werden, dass die vielleicht sogar sehr guten Leistungen der Studierenden nicht zwangsläufig auf der guten Lehrqualität beruhen. Auch durch einen hohen Aufwand mit vielen Eigenleistungen ist ein deutlicher Wissenszuwachs zu erwarten (Wikipedia ; McKeachie 1997). Dieses Selbststudium kann gewollt sein oder als Ausgleich einer schlechten Lehre verstanden werden.

Da Prüfungsergebnisse als alleinige Quellen zur Beurteilung der Ergebnisqualität im Medizinstudium nicht geeignet erscheinen, werden andere Instrumente benötigt, die diese wesentliche Dimension der Lehrqualität valide abbilden können. Im Folgenden wird ein solches Instrument vorgestellt.

1.5 Abschätzung des Lernerfolgs durch studentische Selbsteinschätzungen

Eine Möglichkeit der Erfassung des studentischen Leistungsstands besteht darin, die Studierenden um Selbsteinschätzungen hinsichtlich spezifischer Lernziele zu bitten. In der Regel werden hierzu punktuelle Selbsteinschätzungen erhoben. Dies bedeutet, dass die Einschätzung der eigenen Leistung nur zu einem bestimmten Zeitpunkt vorgenommen wird. Ein Vergleich mit der eigenen Leistung zu zum Beispiel einem anderen Zeitpunkt findet nicht statt.

Jedoch unterliegen auch studentische Selbsteinschätzungen vielfachen Einflussfaktoren, die - je nach betrachtetem Konstrukt – als Störgrößen interpretiert werden müssen. Es scheint zum Beispiel einen inversen Zusammenhang zwischen Selbstsicherheit und Leistung, aber einen positiven Zusammenhang zwischen Selbstsicherheit und Selbsteinschätzung zu geben (Leopold et al. 2005).

In der Literatur sind bereits Versuche beschrieben, verschiedene Charakteristika der Studierenden und der Kurse vor der Auswertung zu kontrollieren, um mögliche Einflüsse erfassen zu können (d'Apollonia and Abrami 1997; Marsh and Roche 1997; McKeachie 1997; Theall and Franklin 2001). Es gibt Hinweise darauf, dass Personen sich tendenziell eher überschätzen. Dies zeigt sich besonders bei jenen die in objektiven Prüfungen als eher schlecht eingeschätzt werden. Kruger und Dunning führten mehrere Studien durch, die diese Vermutung bestätigten. Zudem konnte gezeigt werden, dass eine Überschätzung oft mit einem Mangel an metakognitiven Fähigkeiten, also der Fähigkeit über sich selbst zu urteilen, einhergeht (Kruger and Dunning 1999). Ein Training könnte die Ergebnisse der Selbsteinschätzung durch Verbesserung der metakognitiven Fähigkeiten realistischer gestalten (Kruger and Dunning 1999).

Inwiefern das Geschlecht eine Rolle bei der Selbsteinschätzungsfähigkeit spielt, ist unklar. Männer scheinen sich aber stärker zu überschätzen bzw. positiver einzuschätzen (Colthart et al. 2008).

Welche Auswirkungen das Selbstwertgefühl, also die persönliche Einschätzung der eigenen Fähigkeiten und Einstellungen, auf die Selbsteinschätzungsfähigkeit hat, ist nicht klar. Studien haben aber gezeigt, dass es einen solchen Einfluss zu geben scheint (Gordon 1991).

In Analogie zu der oben für punktuelle Prüfungen beschriebenen Problematik sind auch punktuelle studentische Selbsteinschätzungen zu Kursende nicht geeignet, um den Zugewinn an Wissen, Fertigkeiten und Professionalität während einer Lehrintervention abzubilden. Hierzu wären Informationen über den Leistungsstand der Studierenden vor Beginn der Lehrintervention erforderlich.

Die Selbsteinschätzungsfähigkeit war bereits Gegenstand verschiedener Studien. Davis et. al analysierten verschiedene Studien (Davis 2006), die sich mit Selbsteinschätzungsfähigkeiten von Ärzten befassten. Sie gelangten zu der Erkenntnis, dass Selbsteinschätzungen oft nicht mit Fremdeinschätzungen übereinstimmen. Weiterhin seien mehrere Studien methodisch fehlerhaft und somit nicht gut beurteilbar gewesen.

Fitzgerald dagegen erkannte die studentische Fähigkeit zur Selbsteinschätzung des Leistungsstandes auch über längere Zeiträume als stabil an (Fitzgerald et al. 2003).

Ob zum Beispiel das Selbstwertgefühl einen Einfluss auf die Selbsteinschätzungsfähigkeit der Studierenden hat, soll in vorliegender Studie untersucht werden.

1.5.1 Das Konzept der vergleichenden Selbsteinschätzung

2008 unternahmen Thompson & Rogers erstmals den Versuch, die studentische Lernkurve anhand des Verlaufs von Selbsteinschätzungen darzustellen (Thompson and Rogers 2008). Entscheidend war hierbei, dass nicht nur punktuelle Selbsteinschätzungen erhoben wurden, sondern dass die Studierenden ihre eigenen Fähigkeiten zu verschiedenen Zeitpunkten beurteilten. Studierende, die sich zum Zeitpunkt der Studie in vier aufeinanderfolgenden Studienjahren befanden, wurden aufgefordert ihr Wissen zu Beginn und Ende des Studienjahres einzuschätzen. Somit wurden Selbsteinschätzungen über das Wissen in allen vier Jahren gleichzeitig gewonnen. Wenngleich diese Studie methodische Einschränkungen aufwies, ergaben sich aus den Ergebnissen Hinweise darauf, dass studentische Selbsteinschätzungen – im Gegensatz zu allen anderen bisher üblichen Messverfahren – eine ressourcenschonende Abschätzung des Lernerfolgs durch den Vergleich der Einschätzungen vor und nach Absolvierung der Lehre hinsichtlich psychomotorischer und affektiver Lernziele zulassen.

D'Eon untersuchte die Möglichkeit, vergleichende Selbsteinschätzungen zur Bewertung von Workshops einzusetzen. Die Teilnehmer wurden aufgefordert, sowohl vor als auch nach dem Workshop ihre Leistung, die auf Video aufgezeichnet wurde, selbst einzuschätzen. Zusätzlich wurden sie um eine retrospektive Einschätzung gebeten. Retrospektiv bedeutet hier, dass die eigene Leistung vor dem Training im Nachhinein eingeschätzt wird. Als prospektiv wird die Einschätzung der eigenen Leistung zum Zeitpunkt vor dem Training verstanden (Leistung vor dem Training, Einschätzung vor dem Training). Die Videoaufzeichnungen wurden von zwei unabhängigen Bewertern angeschaut. Sowohl die Teilnehmer als auch die Bewerter schätzten die Leistung nach dem Workshop als höher ein. Die retrospektiven Selbsteinschätzungen unterschieden sich nicht signifikant von den prospektiven. In dieser Studie wurden ausschließlich aggregierte Gruppendaten verwendet. Ein paarweiser Vergleich von Einschätzungen und Leistungen einzelner Teilnehmer fand nicht statt. Individuelle Aussagen sind somit nicht möglich. Alle Ausreißer, *under-* und *overachiever* (Personen die sich unter- beziehungsweise überschätzten) wurden eingerechnet. Zudem wurde nicht der Lernzuwachs berechnet, sondern nur die Korrelationen zwischen Experten- und Teilnehmerratings zu den jeweiligen Erhebungszeitpunkten. Trotzdem schien die Studie generell die vergleichende Selbsteinschätzung als kostengünstige Evaluationsmethode zu stützen (D'Eon et al. 2008). 2009 wurde D'Eons Studie von Lam und Eva kommentiert. Hierbei wurde die Einrechnung von balancierten Über- und Unterschätzungen als nicht sinnvoll erachtet, da dies die Ergebnisse verzerren würde. Lam kritisiert die Annahme D'Eons, aggregierte Gruppendaten, also die Berechnung auf *group-level-Niveau*, zu verwenden, um Ausreißer zu kaschieren (Lam 2009). In der *group-level-Berechnung* werden die Ergebnisse aller Teilnehmer zu einer Fragestellung zusammengerechnet und gemittelt, sodass nur ein gemeinsamer Wert entsteht. Dieses Verfahren wird vor und nach dem Training angewandt, sodass beide Gruppenwerte dann miteinander verglichen werden können, um einen Unterschied festzustellen. Die alternative Berechnung ist die auf *individual-level-Niveau*, in der die Einschätzungen der einzelnen Kandidaten vor und nach dem Training für jede einzelne

Fragestellung miteinander verglichen werden müssen und erst die ermittelten Unterschiede einen gemittelten Wert ergeben.

Ein weiterer Kritikpunkt in der Analyse Lams war die direkte Zurückführung der Leistung der Teilnehmer auf das Training. Zudem dürften die Selbsteinschätzungsergebnisse nicht ohne weiteres als auf andere Situationen übertragbar angenommen werden. Die Zurückweisung der Selbsteinschätzung als sinnvolle Evaluationsmethode beruht hier auf der Annahme, dass Selbsteinschätzungsdaten nicht valide sind und die Ergebnisse der Teilnehmer nicht unbedingt auf das Training zurückgeführt werden können.

Die hier beschriebenen Ansätze zur Abschätzung des Lernerfolgs durch zwei- oder mehrzeitige (d.h. vergleichende) Selbsteinschätzungen waren unter anderem der Limitation unterworfen, dass das Vorwissen der Lernenden bei der Berechnung des Zuwachses nicht adäquat berücksichtigt wurde.

1.5.1.1 Vergleichende Selbsteinschätzung an der Universitätsmedizin Göttingen

Im Jahr 2006 wurde an der Universitätsmedizin Göttingen ein Beratungsprozess begonnen, in dessen Rahmen ein neues, ergebnisorientiertes Evaluationsinstrument entwickelt werden sollte. Vor dem Hintergrund der oben beschriebenen Limitationen punktueller Leistungsnachweise bzw. Selbsteinschätzungen und der Frage nach der Konstruktvalidität traditioneller Evaluationsbögen, wurde ein System entwickelt und pilotiert, das auf der Grundlage wiederholter studentischer Selbsteinschätzungen eine Abschätzung des Lernzuwachses gestattet.

Den Studierenden wurden Aussagen über Fachwissen oder spezielle Fertigkeiten präsentiert, zu denen sie sich auf einer sechsstufigen Skala von „Trifft voll zu“ bis „Trifft überhaupt nicht zu“ selbst einschätzen. So können nicht nur das kognitive Wissen, sondern die Leistungen aller drei Domänen erfasst werden. Die Studierenden haben zu Modulbeginn und am Modulende jeweils sechs Tage Zeit, die Evaluationsbögen online zu bearbeiten.

Aus den auf freiwilliger Basis erhobenen studentischen Daten wird mittels einer in Göttingen entwickelten Formel (Raupach et al. 2011) der lernzielspezifische Lernzuwachs (in %) errechnet. Zunächst wird die Differenz der Prä- und Post-Werte ermittelt. Der Lernzuwachs ergibt sich aus dem Quotienten dieser Differenz mit dem Prä-Wert minus 1. Zur Darstellung in Prozent wird die ermittelte Zahl mal 100 genommen.

$$\cdot \text{Lernzuwachs [\%]} = \frac{\overline{X_{prä}} - \overline{X_{post}}}{\overline{X_{prä}} - 1} \times 100$$

X_{prä} = durchschnittliche initiale Selbsteinschätzung

X_{post} = durchschnittliche Selbsteinschätzung nach dem Kurs

Im Unterschied zu einer einfachen Wissensüberprüfung am Ende eines Moduls berücksichtigt die Berechnung des Lernzuwachses auch den Kenntnisstand der Studierenden vor Modulbeginn und kann die Ergebnisqualität der Lehre erfassen.

Seit dem Wintersemester 08/09 wird das Instrument der vergleichenden Selbsteinschätzung für die ergebnisorientierte Evaluation im Studium der Humanmedizin an der UMG eingesetzt. Für jedes der 21 klinischen Module wurden 15 operationalisierte Lernzielaussagen formuliert, die den

Studierenden zu Modulbeginn und –ende vorgelegt werden. Auf der Basis der vergleichenden Selbsteinschätzung wird für jedes Lernziel ein prozentualer Zuwachs berechnet. Der Mittelwert aus diesen 15 Zuwächsen wird als mittlerer Lernzuwachs im Modul betrachtet.

Das neue lernzielbezogene und somit ergebnisorientierte Evaluationssystem wurde in zwei Studien eingehend untersucht:

1) Raupach et al. 2011:

In Göttingen wurde von 2008 bis 2010 eine Studie zur vergleichenden Selbsteinschätzung der Studierenden durchgeführt (Raupach et al. 2010). Die aus dieser Studie abgeleiteten Daten wurden verwendet, um weitere Erkenntnisse zur vergleichenden Selbsteinschätzung als Evaluationsinstrument zu gewinnen. Alle Medizinstudenten/innen der klinischen Semester der Universitätsmedizin Göttingen, einschließlich vier Kohorten des Kardiologiemoduls (Modul 3.1), wurden um Selbsteinschätzungen ihrer Leistung vor und nach jedem Modul in Bezug auf 15 curriculumskonforme Lernziele gebeten. Die Studienteilnahme im Kardiologiemodul betrug 100%. Diese Studierenden erhielten zusätzlich zur Selbsteinschätzungsfrage „Ich kann ein EKG interpretieren.“ eine praktische Aufgabe – drei EKG-Interpretationen. Auch diese praktische Aufgabe wurde vor und nach dem Modul geprüft.

Der Lernzuwachs wurde anhand der Selbsteinschätzungen sowohl gemittelt über die gesamte Kohorte als auch als individueller Lernzuwachs pro Student/in errechnet. Berechnet wurde der Lernzuwachs mittels der auch in vorliegender Studie verwendeten Formel (siehe Kapitel 1.5.1.1 oben).

Im Modul 3.1 fand außerdem entweder eine formative oder eine summative Prüfung statt. Formative Prüfungen sollen dabei das zukünftige Lernen lenken und Erfolge bestätigen, summative dagegen bieten ein umfassendes Urteil über Kompetenz, praktische Fähigkeiten oder die Qualifikation für den Fortschritt zu höheren Ebenen der Verantwortlichkeit (Epstein 2007). Auch die Intensität und Organisation der Lehre unterschied sich in den vier Studienkohorten.

In dieser Studie wurden Alternativen zum prozentualen Lernzuwachs, reine Prä-Post-Differenzen bzw. Effektstärken, betrachtet. Es zeigte sich, dass sich große Mittelwertunterschiede sowohl in den Effektstärken als auch im errechneten Lernzuwachs abbildeten. Das weitere Ranking war allerdings in allen Methoden verschieden. Letztlich schien der Lernzuwachs aus der vergleichenden Selbsteinschätzung über die Studierendenkohorten reproduzierbare Ergebnisse zu liefern, im Gegensatz zu Effektstärken, die zwischen den Kohorten variierten. Es ergab sich außerdem eine signifikante Korrelation einer hohen Selbsteinschätzung mit der praktischen Durchführung der EKG-Interpretation (Raupach et al. 2011).

Ob die bisher verwendeten 15 Lernziele ausreichen oder ob mehr Lernziele nötig sind um reliable Daten zu erhalten, ist in vorliegender Studie untersucht worden.

2) Raupach et al. 2012:

In dieser Arbeit wird die vergleichende Selbsteinschätzung der Globalbewertung als traditionellem Evaluationsinstrument und weiteren Evaluationsparametern gegenübergestellt. Im Wintersemester 08/09 wurden alle Studierenden des klinischen Studienabschnitts gebeten, vor und nach Absolvierung der Module an einer Evaluation teilzunehmen. Die Evaluation beinhaltete Selbsteinschätzungsfragen zu 15 auf die Module projizierten Lernzielen. Zusätzlich wurden vor dem Modul drei Fragen zur Motivation (zum Beispiel „Ich freue mich schon auf dieses Modul.“) und nach dem Modul fünf Fragen zur Struktur (zum Beispiel „Bezogen auf meine berufliche Zukunft schätze ich

meinen Lernzuwachs in diesem Modul als sehr hoch ein.“) sowie eine Frage zur Globalbewertung („Bitte bewerten Sie das Modul mit einer Gesamtnote.“) gestellt.

Es wurden zwei Ranglisten erstellt, die auf dem Mittelwert der Globalbewertungen beziehungsweise auf dem Mittelwert des Lernzuwachses, der mittels der vergleichenden Selbsteinschätzung errechnet wurde, beruhten. Zudem wurden Korrelationen zwischen verschiedenen Evaluationsparametern gebildet. Es konnte in dieser Studie gezeigt werden, dass die Rangreihenfolge der Module bezogen auf die Globalbewertung nicht mit der Reihenfolge bezogen auf den errechneten Lernzuwachs übereinstimmt. Dagegen ergab sich, dass die studentischen Erwartungen zu Modulbeginn (Motivation) mit der Bewertung der Struktur nach dem Modul stark positiv korrelierten, ebenso mit der Globalbewertung. Auch zeigte sich eine hohe Korrelation zwischen der strukturellen und der Globalbewertung. Die Korrelation des Lernzuwachses war dagegen sowohl mit der subjektiven Einschätzung, viel gelernt zu haben ($r=0,44$), als auch mit der Globalbewertung ($r=0,42$) eher schwach (Raupach et al. 2012).

Die beiden Studien ergaben zusammenfassend, dass die vergleichende studentische Selbsteinschätzung als gutes Evaluationsinstrument betrachtet werden kann. Allerdings ist noch weitere Forschung zum besseren Verständnis nötig.

1.6 Mögliche Störfaktoren der vergleichenden Selbsteinschätzung

Auch die vergleichende Selbsteinschätzung kann Einflüssen unterworfen sein, die die Ergebnisse verzerren. Z.B. werden lernzielbezogene studentische Selbsteinschätzungen durch eine retrospektive Erhebung evtl. verändert. In einer Studie von Skeff et.al (Skeff et al. 1992) wurde untersucht, ob eine Einschätzung der eigenen Leistungen vor einem Training besser pro- oder retrospektiv erfolgen sollte, um eine möglichst große Übereinstimmung mit externen Einschätzungen zu erzielen. Die Einschätzungen der befragten Teilnehmer zeigten retrospektiv eine größere Übereinstimmung mit externen Ergebnissen. Von Autoren wurde dies erklärt mit einem besseren Verständnis der zu bewertenden Dimensionen durch das Training.

Sinnvoll scheint also eine retrospektive Selbsteinschätzung, wenn es wahrscheinlich ist, dass die Interpretation der Dimensionen durch Teilnahme am Training beeinflusst wird (Skeff et al. 1992). Andererseits wäre es möglich, dass Studierende ihr Wissen vor dem Modul in der Retrospektive selbstwertförderlich etwas höher einstufen (Baumeister et al. 2003). Inwiefern diese Beobachtung bestätigt werden kann, soll durch die vorliegende Studie gezeigt werden.

Frühere Studien ergaben, dass die Genauigkeit von studentischen Selbsteinschätzungen über die Ausbildung leicht ansteigt (Fitzgerald et al. 2003).

Eine weitere Schwierigkeit stellt die Anzahl der gestellten Fragen dar. Um den Aufwand für die teilnehmenden Studierenden überschaubar zu halten, wurde bisher nur eine kleine Auswahl von Fragen genutzt, anhand derer die Studierenden ihr Wissen selbst einschätzen sollten. Natürlich ist es möglich, dass durch eine solche Auswahl Fehler entstehen, da vorhandenes Wissen möglicherweise nicht erfasst, Wissenslücken dagegen überproportional bewertet werden. Bei der Anzahl der gestellten Fragen sollte aber bedacht werden, dass die Chancen, möglichst viele Studierende zur

Teilnahme an der Evaluation zu motivieren, umso größer sind, je weniger Antworten für ein reliables Ergebnis benötigt werden.

Bisher wurden die Studierenden jeweils vor Modulbeginn und am Modulende angehalten, die Evaluationsbögen auszufüllen. Leider stimmte die Zahl derer, die am Anfang teilnahmen, nur selten mit der Teilnehmerzahl am Ende überein. Somit wurde eine Vergleichbarkeit der Daten erschwert. Es ist möglich, dass zu den unterschiedlichen Erhebungszeitpunkten besonders leistungsstarke bzw. besonders leistungsschwache Studierende teilnahmen und in einer Berechnung des Lernzuwachses aus diesen Selbsteinschätzungen verfälschte Ergebnisse resultierten. Es ergibt sich die Notwendigkeit, eine Zuordnung der zu Modulbeginn und am Modulende bearbeiteten Evaluationsbögen in Hinsicht auf die Teilnehmer vorzunehmen, um sicherzustellen, dass der berechnete Lernzuwachs aus vergleichbaren Daten resultiert. Zudem könnte nur dann ein individueller paarweiser Vergleich erfolgen, wenn Selbsteinschätzungen desselben/ derselben Studierenden vor und nach dem Modul vorliegen.

Die vergleichende Selbsteinschätzung erfordert von den Studierenden eine Mitarbeitsbereitschaft. Doch trotz großer Motivation kann die Korrektheit der vergleichenden Selbsteinschätzung durch Störfaktoren beeinflusst werden. Eine Minimierung der Störfaktoren kann erreicht werden, indem ermittelt wird, welches Konstrukt die Studierenden der Evaluation zugrunde legen.

1.7 Ziele dieser Arbeit

Aufgrund der noch offenen Fragen sind weitere Studien nötig, um Klarheit in Bezug auf studentische Selbsteinschätzungen sowie deren Möglichkeiten und Einschränkungen zu schaffen.

Unklar ist zum Beispiel ob die Berechnung auf *individual-level* andere Ergebnisse zeigt, als die *group-level*-Daten der gesamten Kohorte. Weiterhin ist es möglich, dass das Selbstwertgefühl der Teilnehmer die Selbsteinschätzungen beeinflusst. Zudem gibt es die Möglichkeit, den initialen Leistungsstand erst retrospektiv, nach Absolvierung eines Moduls, zu erfassen. Dabei ist nicht klar, ob es einen Unterschied zwischen den so ermittelten Werten und den prospektiv, also vor dem Modul erhobenen Werten gibt. Weiterhin ist eine Ermittlung der Sensitivität und Spezifität des neuen Verfahrens notwendig, um die aus den Selbsteinschätzungen gewonnenen Daten auch adäquat bewerten zu können. Inwiefern die neue Evaluationsmethode eine akzeptable Kriteriumsvalidität besitzt, bedarf noch weiterer Klärung.

Die Untersuchungen in vorliegender Studie sollen helfen, Antworten auf noch offene Fragen geben zu können:

1. Welchen Stellenwert hat die Evaluation für die Studierenden?
2. Liefert ein paarweiser Vergleich studentischer Selbsteinschätzungen andere Ergebnisse als die Kalkulation des Lernzuwachses aus den Mittelwerten einer gesamten Studentenkohorte?
3. Liefert eine prospektive Datenerhebung bei der Lernzuwachsevaluation andere Ergebnisse als eine retrospektive Erhebung des initialen Leistungsstands?
4. Wie viele Lernziele pro Modul müssen abgefragt werden, um reliable Daten zu erhalten?
5. Wird das Ergebnis der vergleichenden Selbsteinschätzung durch psychometrische Größen wie das Selbstwertgefühl beeinflusst?
6. Besitzt das Evaluationsverfahren eine akzeptable Kriteriumsvalidität? Lässt sich für die Lernzuwachs-Evaluation ein Cut-Off für die Unterscheidung erfolgreich gelehrter Inhalte von erfolglos gelehrten Inhalten definieren?

2 Methoden

2.1 Studiendesign

Zur Beantwortung der in Kapitel 1.7 genannten Studienfragen wurden qualitative und quantitative Methoden zur Datenerhebung genutzt. Es wurden Fokusgruppengespräche als ein qualitativer Ansatz gewählt, um zu eruieren, welches Konstrukt den studentischen Bewertungen von Lehrveranstaltungen zugrundeliegt (Studienfrage 1). Die Beantwortung der übrigen Studienfragen fand im Rahmen einer prospektiven Studie statt, in der bezüglich spezifischer operationalisierter Lernziele in einem klinischen Modul sowohl studentische Selbsteinschätzungen (subjektive Leistungsparameter) als auch Prüfungsleistungen (objektive Leistungsparameter) erhoben wurden. Die Analyse stützte sich dabei nicht nur auf aggregierte Daten („*group-level*“). Durch die Verwendung individueller Codes war erstmals auch ein paarweiser Datenabgleich auf individueller Ebene („*individual-level*“) möglich (Studienfrage 2). Zudem wurden Selbsteinschätzungen des Leistungsniveaus zu Modulbeginn nicht nur vor (prospektiv), sondern auch nach (retrospektiv) der Lehrexpotion erhoben (Studienfrage 3). In der prospektiven Studie konnten – anders als aktuell an der UMG üblich – nicht nur 15 Lernziele, sondern alle 46 auf das betrachtete Modul bezogenen Lernziele berücksichtigt werden. Auf diese Weise wurde eine Abschätzung des mittleren Lernzuwachses auf der Grundlage unterschiedlich großer Lernziel-Stichproben möglich (Studienfrage 4). Zusätzlich zu den subjektiven und objektiven Leistungsparametern fand in der prospektiven Studie eine Erhebung zum Selbstwertgefühl der Studierenden statt (Studienfrage 5). Schließlich wurde die Kriteriumsvalidität des Instruments der vergleichenden Selbsteinschätzung für 33 kognitive Lernziele anhand des Vergleichs zwischen subjektiven und objektiven Zuwachsdaten untersucht. Die Fähigkeit des neuen Evaluationsinstruments, zwischen erfolgreich und erfolglos gelehrten Lernzielen zu unterscheiden, wurde mittels einer *Receiver-Operating Characteristic* (ROC)-Analyse getestet (Studienfrage 6).

2.2 Probandenrekrutierung

Zur Teilnahme an den Fokusgruppengesprächen (Studienfrage 1) wurden die Studierenden des 1., 3. und 4. klinischen Semesters per E-Mail eingeladen. Alle Studierenden, die sich für ein Fokusgruppengespräch interessierten, hatten die Möglichkeit teilzunehmen. Um die Motivation für eine Teilnahme zu erhöhen, erhielt jede/r Studierende einer Fokusgruppe nach dem Gespräch einen Büchergutschein über 25 Euro.

Die prospektive Studie zur Beantwortung der Studienfragen 2 bis 6 wurde im Modul 3.1 des klinischen Curriculums an der UMG durchgeführt. Alle Studierenden, die sich zur Teilnahme am Modul 3.1 angemeldet hatten, wurden per E-Mail über die geplante Studie mit Hilfe einer umfassenden Probandeninformation (s. Anhang 3: Einverständniserklärung) in Kenntnis gesetzt. Die Studierenden erhielten Informationen über Sinn und Zweck der Studie, sowie die geplante Umsetzung in ihrem Semester. Auch welche Daten wann erhoben wurden, konnten die Studierenden der Probandeninformation entnehmen. Am ersten Modultag fand zudem eine mündliche Aufklärung

mit der Möglichkeit, Fragen zu stellen, statt. Das Verfahren sowie die Durchführung der Studie wurden eng mit den Semestersprechern abgestimmt.

Jede/r an der Studie beteiligte Studierende erhielt eine Einverständniserklärung, die er ausgefüllt dem Studienleiter zukommen ließ und sich so mit der Verwendung seiner Daten im Rahmen der Studie einverstanden erklärte.

Um die Motivation der Studierenden zu steigern und die Wahrscheinlichkeit einer hohen Rücklaufquote der Evaluationsbögen zu erhöhen, fand eine Verlosung von 30 Büchergutscheinen im Wert von je 25 Euro unter allen Studienteilnehmern statt.

Das Vorgehen wurde in einem Antrag (Nummer 27/3/11 An) der Ethikkommission der Universitätsmedizin Göttingen vorgestellt. Diese erklärte sich für nicht zuständig. Die mit der Rekrutierung einhergehende Erstellung eines Codes für jeden Teilnehmer wurde mit dem Datenschutzbeauftragten der UMG abgestimmt.

2.3 Fokusgruppengespräche

Zur Beantwortung der ersten Studienfrage wurden Fokusgruppengespräche durchgeführt. Die Gespräche fanden kurz nach Ende des Moduls 3.1 statt. Es erschienen zu den Terminen, die nach dem 1., 3. und 4. klinischen Fachsemester getrennt durchgeführt wurden, jeweils fünf bis sieben Teilnehmer. In jedem der genannten Fachsemester wurde ein Fokusgruppengespräch durchgeführt. Die Treffen fanden außerhalb der Lehrveranstaltungen statt und dauerten 59-75 Minuten. Die Fokusgruppenleiterin stellte offene Fragen zur Diskussionsanregung bezüglich verschiedener Themen. Die weitere Gesprächsführung wurde dann durch die Studierenden gestaltet. Somit war es möglich, die Gedanken der Studierenden zum Thema Evaluation in Erfahrung zu bringen und neue Anregungen zu erhalten.

Die standardisierten Leitfragen sind hier wiedergegeben:

- 1) Was ist Ihrer Meinung nach der Sinn von Evaluation im Medizinstudium?
- 2) Wie würden Sie gute Lehre definieren?
- 3) Was denken Sie über die an unserer Institution aktuell durchgeführte Evaluation?
- 4) Wie gelangen Sie zu einer Gesamteinschätzung eines Kurses?
- 5) Welche Konsequenzen wünschen Sie sich aus Ihrer Evaluation?

2.4 Selbsteinschätzungen

2.4.1 Vorbefragung

Zur Beantwortung der Studienfrage 2 war es notwendig einen individuellen Prä-Post-Vergleich von Selbsteinschätzungen und Testatergebnissen durchzuführen. Bisher war nur ein Vergleich der Mittelwerte über die Kohorte möglich. Um den individuellen Vergleich durchzuführen, war es nötig, die Prä- und Post-Evaluationsbögen eines Studierenden jeweils miteinander in Verbindung zu bringen. Diese Zuordnung geschah über einen durch die Studierenden selbst generierten sechsstelligen Code, bestehend aus vier Buchstaben und zwei Zahlen. Die Studierenden wurden

gebeten, diesen Code in jedem Teil der Studie anzugeben. Die Anonymität der Studierenden blieb dadurch gewahrt, dass bei der Auswertung der Studie nur die Codes, nicht jedoch die Namen benutzt wurden. Eine Zuordnung der Namen zu den Codes wurde auf einer gesonderten Liste vorgenommen, die nur für den Zweck der Büchergutscheinverlosung verwendet wurde. Diese Zuordnung wurde durch die Studierenden selbst ermöglicht, da diese beim Eingangstestat auf einem Extra-Zettel ihren Namen zusammen mit dem Code angaben.

In die Verlosung der Büchergutscheine wurden nur Codes von denjenigen Studierenden einbezogen, die an allen studienbezogenen Befragungen und formativen Prüfungen teilgenommen hatten. Nach Ziehung der Gewinner wurden die Namen der Studierenden mit Hilfe der gesonderten Zuordnungsliste identifiziert.

Zur Beantwortung der 4. Studienfrage wurde die an der UMG bisher eingesetzte Prä-Evaluation dahingehend modifiziert, dass die Studierenden um Selbsteinschätzungen hinsichtlich aller im Modul 3.1 relevanten 46, nicht nur der bisher gefragten 15, Lernziele gebeten wurden.

Um die Studienfrage 5 zu beantworten, wurden die Studierenden zusätzlich aufgefordert, Angaben zum Selbstwertgefühl zu machen. Hierbei kam die teilrevidierte Skala zum Selbstwertgefühl von Rosenberg (Collani and Herzberg 2003) zum Einsatz (s. Anhang 4: Rosenbergskala, Seite 71). Dieses Instrument erfasst mit 10 Items, die auf einer vierstufigen Skala zu beantworten sind, sehr ökonomisch das globale Selbstwertgefühl. Der Maximalwert, der einem hohen Selbstwertgefühl entsprach, lag hier bei 30 Punkten, das Minimum bei null. Des Weiteren sollten mögliche Effekte durch ein sozial erwünschtes Antwortverhalten durch Integration von zwei Kontrollitems berücksichtigt werden.

2.4.2 Nachbefragung: retro-/ prospektiv

Zur Beantwortung der dritten Studienfrage wurde die Post-Evaluation dahingehend modifiziert, dass zusätzlich zur Einschätzung der Fähigkeiten nach Absolvierung des Moduls eine retrograde Einschätzung des Kompetenzniveaus zu Modulbeginn erhoben wurde. Auf diese Weise konnten mögliche Abweichungen zwischen den zwei zeitlich auseinander liegenden Prä-Einschätzungen untersucht werden.

2.5 Objektive Leistungsparameter

Zur Beantwortung der sechsten Studienfrage wurden die subjektiven Selbsteinschätzungen mit den objektiv ermittelten Testatergebnissen verglichen. Das Testat als Außenkriterium ist als formative Prüfung angelegt worden. Die Fragen und Aufgaben wurden so eng wie möglich an den einzelnen Lernzielen orientiert, um eine bestmögliche Vergleichbarkeit zwischen Selbsteinschätzung und objektiven Werten zu erhalten (Kern et al. 1998).

Das Eingangstestat wurde am ersten Modultag, das Abschlusstestat in der letzten Modulwoche durchgeführt. Die Aufgaben waren in beiden Testaten exakt dieselben.

Die Testate bestanden aus jeweils 30 fünffachen True-False-Items (150 Einzel-Items) und vier offenen Fragen (20 Rohpunkte). Eine Frage zum Thema Hyperlipidämie musste aufgrund eines Fehlers in der Fragestellung aus der Wertung genommen werden. Bei dieser Frage waren auf den Testatfragebögen

die Antworten bereits angegeben. Nach Herausnahme dieser Frage verblieben 33 lernzielbezogene Fragen zur Analyse. Entsprechend basiert die Teststatistik auf den Daten der 145 (29 x 5) auswertbaren True-False-Items sowie der 20 (4 x 5) offenen Items.

Bei der Auswertung der Fragen wurden nur die richtigen Antworten mit einem Punkt honoriert. Falsche oder nicht gegebene Antworten wurden mit null Punkten bewertet. Die Studierenden erhielten allerdings die Information, dass alle Antworten in die Bewertung einfließen würden und somit die *negative-marking*-Methode (Anderson 2004) zur Anwendung käme. Für richtige Antworten erhalten die Studierenden bei dieser Methode einen Punkt, für falsche wird ein Punkt abgezogen. Dieses Vorgehen - richtige Antworten: +1 Punkt, falsche Antworten: -1 Punkt, „weiß nicht“-Antworten: 0 Punkte - beugt laut Anderson dem Raten vor und gibt so eher das wirkliche Wissen wieder. Können die Studierenden nur Punkte bekommen, aber keine verlieren, benachteiligt dies die vorsichtigen Studierenden (Anderson 2004), die im Zweifel eher „weiß nicht“ ankreuzen, anstatt durch Raten der richtigen Lösung evtl. einen, nicht durch Wissen erlangten, Punkt zu erhalten. Durch die Ankündigung der *negative-marking*-Methode sollten die Studierenden in der Studie darin unterstützt werden, nur die Fragen, bei denen sie nicht zu raten brauchten, zu beantworten und unsichere Fragen eher auszulassen.

Ebenso wurde bei der Nennung relevanter Punkte in den offenen Fragen verfahren. Nur die Nennung dieser Antworten, keiner zusätzlichen, wurde mit jeweils einem Punkt honoriert, sodass auch hier pro Frage fünf Punkte erreicht werden konnten.

2.6 Datensammlung

Das Ausfüllen der Evaluationsbögen wurde den Studierenden durch ein per E-Mail zugeschicktes Passwort ermöglicht. Aufgrund technischer Schwierigkeiten konnten mehrere Studierende die Internetseite nicht öffnen. Deshalb erhielten alle Studienteilnehmer die zum Zeitpunkt des Testats noch keine Selbsteinschätzung durchgeführt hatten, den jeweiligen Evaluationsbogen als gedrucktes Exemplar.

Jede(r) Studierende trug den von ihm/ihr selbst generierten Code in die Formulare ein.

Der zu Modulbeginn und Modulende eingesetzte Fragebogen beinhaltete zunächst allgemeine Angaben, so den Code, das Geschlecht und das Alter. Die zu den Lehrinhalten passenden Lernzielfragen wurden gestellt, in denen die Studierenden um eine Einschätzung ihres eigenen Wissens gebeten wurden. Mit einem entsprechenden Hinweis versehen, wurde ein fast identischer Fragebogen am Modulende zweimal, für das retrospektiv eingeschätzte Vorwissen sowie das aktuelle Wissen nach Absolvierung des Moduls, ausgefüllt.

In den Prä-Evaluationsbögen wurde zudem das Selbstwertgefühl anhand der Rosenbergkala erhoben. In den Post-Evaluationsbögen wurden zusätzlich eine Gesamtmodulnote sowie die wöchentliche Arbeitszeit erfragt.

Die Evaluationsbögen sind in der Form, in der sie den Studierenden vorgelegt bzw. online präsentiert wurden, im Anhang (s. Anhang 5: Prä-Evaluationsbogen, Seite 70; Anhang 6: Post-Evaluationsbogen, Seite 75) dieser Arbeit zu finden.

Die ermittelten Testatpunkte wurden computergestützt in ein Statistikprogramm (SPSS Statistics®) eingegeben.

2.7 Datenanalyse

2.7.1 Berechnung der Lernzuwachsrate

Die Berechnung der Lernzuwachsrate erfolgte mittels einer in Göttingen entwickelten Formel. In der hier gezeigten Vorstellung der Studie wurden für die unterschiedlichen Berechnungsarten der Lernzuwachsrate folgende Abkürzungen verwendet:

Tabelle 1: Abkürzungsverzeichnis der möglichen Berechnungsarten des Lernzuwachses

ABKÜRZUNG	ERKLÄRUNG
oALZ (objektiver aggregierter Lernzuwachs)	Lernzuwachs berechnet aus den Gruppen-Mittelwerten der formativen Testatergebnissen
sALZ (subjektiver aggregierter Lernzuwachs)	Lernzuwachs berechnet aus den Gruppen-Mittelwerten der studentischen Selbsteinschätzungen
oILZ (objektiver individueller Lernzuwachs)	Lernzuwachs berechnet aus den individuellen Testatergebnissen
sILZ (subjektiver individueller Lernzuwachs)	Lernzuwachs berechnet aus den individuellen studentischen Selbsteinschätzungen

Die Formel zur *group-level*-Analyse wurde angewendet, um den aggregierten Lernzuwachs (ALZ) zu ermitteln. Hierzu wurden die Selbsteinschätzungen bzw. Testatergebnisse bezogen auf ein Lernziel gemittelt über die gesamte Kohorte erfasst.

$$\text{aggregierter Lernzuwachs [\%]} = \frac{X_{\text{prä}} - X_{\text{post}}}{X_{\text{prä}} - 1} \times 100$$

$X_{\text{prä}}$ = initiale Selbsteinschätzung oder Testergebnis für ein Lernziel gemittelt über die gesamte Kohorte

X_{post} = Selbsteinschätzung oder Testergebnis für dasselbe Lernziel nach dem Kurs gemittelt über die gesamte Kohorte

(Quelle (Raupach et al. 2011))

Pro Lernziel wurden die Selbsteinschätzungen bzw. Testatergebnisse der gesamten Kohorte erfasst und ein Mittelwert sowohl der Prä- als auch der Post-Werte gebildet. Nach Einsetzen der so errechneten Zahlen in die oben stehende Formel ergab sich pro Lernziel ein prozentualer Zuwachs. Insgesamt wurden 46 Prozentzahlen, die den aggregierten Lernzuwachs (ALZ) für jedes Lernziel zeigten, ermittelt. Bei Verwendung der subjektiven Selbsteinschätzungen ergibt sich der sALZ-, bei objektiven Testatergebnissen der oALZ-Wert.

Für die *individual-level*-Analyse wurde die folgende Formel verwendet, woraus sich für jede/n Studierende/n für jedes Lernziel durch paarweisen Vergleich der Prä- und Post-Werte ein subjektiver

individueller Lernzuwachs (sILZ) berechnen ließ. Es ergaben sich für jede/n Studierende/n 46 Prozentzahlen.

$$\text{individueller Lernzuwachs [\%]} = \frac{Y_{prä} - Y_{post}}{Y_{post} - 1} \times 100$$

$Y_{prä}$ = initiale Selbsteinschätzung oder Testatergebnis in einem Lernziel für einen Studierenden

Y_{post} = Selbsteinschätzung oder Testatergebnis in einem Lernziel für einen Studierenden

(Quelle (Raupach et al. 2011))

Aus diesen individuellen Zuwächsen konnten die Mittelwerte aller Studierenden pro Lernziel gebildet werden. Der Mittelwert aller Lernzuwachsrate pro Lernziel wurde nun ermittelt. Es ergaben sich nun 46 Prozentzahlen, für jedes Lernziel ein durchschnittlicher Lernzuwachs.

Es konnten entweder die Selbsteinschätzungen zur Ermittlung der sILZ- oder die Testatergebnisse für die oILZ-Werte verwendet werden.

Wenn ein/e Student/in bereits vor dem Modul eine 1 in der Selbsteinschätzung angab, war ein Lernzuwachs nicht mehr möglich. Der Lernzuwachs betrug dann 0%. Auch negative Lernzuwachsrate waren möglich (siehe Kapitel 4.2.2, Seite 53).

2.7.2 Fokusgruppengespräche (Studienfrage 1)

Zur Beantwortung der ersten Studienfrage wurden Fokusgruppengespräche analysiert. Die Gespräche wurden per Tonband aufgenommen und anschließend schriftlich festgehalten. Haupt- und Unterthemen jeder Fokusgruppe wurden von zwei unabhängigen Prüfern in Mind-Maps zusammengetragen, die letztlich vier Hauptthemen ergaben. Hierbei wurde das System MaxQDA® verwendet. Durch dieses Vorgehen konnten die mehrfach von den Studenten angesprochenen Themen hervorgehoben, aber auch einzelne Anmerkungen der Studierenden festgehalten werden. Es ergab sich ein Meinungsbild der Studierenden zur Lehrevaluation.

2.7.3 Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (individual-level vs. group-level) (Studienfrage 2)

Zur Beantwortung der zweiten Studienfrage war es erforderlich die Lernzuwachsrate mit zwei verschiedenen Berechnungsarten zu ermitteln. Es wurden dabei die aggregierten (ALZ) von den individuellen (ILZ) Lernzuwachsrate unterschieden. Verwendet wurden die Ergebnisse der prospektiven Selbsteinschätzungen bezogen auf alle im Modul abgebildeten Lernziele sowie die Selbsteinschätzungen nach dem Modul. Pro Lernziel wurde ein Mittelwert der Lernzuwachsrate sowohl auf *group-* als auch auf *individual-level* gebildet und von diesen die Korrelation bestimmt (siehe Kapitel 2.7.1). Die ermittelten prozentualen Lernzuwachsrate sind in Tabellenform in Kapitel 3.4 (Seite 36) dargestellt.

Der Unterschied zwischen der *group-level*- und der *individual-level*-Berechnung liegt darin, dass zur *group-level*-Berechnung zuerst ein Mittelwert der Selbsteinschätzungen aller Studenten pro Lernziel gebildet und anschließend der Lernzuwachs berechnet wird. Bei der *individual-level*-Methode dagegen wird zunächst der Lernzuwachs pro Lernziel eines/r jeden Studenten/in berechnet. Anschließend kann der Mittelwert aus allen Zuwächsen gebildet werden.

Die Ergebnisse der beiden Berechnungsarten können direkt miteinander verglichen werden.

Die Abweichungen der beiden Messmethoden voneinander sind in einem Bland-Altman-Plot dargestellt (siehe Kapitel 3.4, Seite 36).

Der Bland-Altman-Plot ist ein Punktdiagramm. An den Koordinaten wird die Differenz der Werte der zu vergleichenden Methoden (hier sALZ- minus sILZ-Werte) gegen den Mittelwert derselben Werte $((sALZ + sILZ) / 2)$ aufgetragen ist. Die Punkte im Diagramm stellen jeweils ein Lernziel dar. Zur besseren Interpretation werden drei waagerechte Zusatzlinien eingezeichnet:

der Mittelwert aller eingetragenen Differenzen

dieser Mittelwert plus Standardabweichung mal 1,96 und

dieser Mittelwert minus Standardabweichung mal 1,96 (Bland and Altman 1999).

Je geringer die Differenz und die Standardabweichung ist, desto höher die Übereinstimmung der Ergebnisse der beiden Methoden.

2.7.4 Prospektiv vs. retrospektive Erhebung der initialen Selbsteinschätzungen (Studienfrage 3)

Zur Beantwortung der dritten Studienfrage wurden die Studierenden zusätzlich zur Selbsteinschätzung nach dem Modul um eine retrospektive Einschätzung ihres Wissens vor dem Modul gebeten und der verwendete Fragebogen entsprechend erweitert.

Die Studierenden schätzten wie bisher üblich ihre Leistungen vor und nach dem Modul ein. Sie wurden nun um eine zusätzliche Einschätzung ihres Leistungsstandes vor dem Modul gebeten, wobei der Zeitpunkt, zu dem diese Einschätzung stattfand, erst am Modulende lag. Somit konnten für jede/n Studierende/n für jedes Lernziel zwei Einschätzungen des initialen Leistungsstandes ermittelt werden.

Im nächsten Schritt wurden die Lernzuwachsrate berechnet, wobei hierfür in der *group-level*-Analyse pro Lernziel zwei Zuwächse relevant waren. Einbezogen wurden die prospektiv ($sALZ_{pro}$) und die retrospektiv ($sALZ_{retro}$) erhobenen Selbsteinschätzungen. Der Korrelationskoeffizient nach Pearson zwischen den Mittelwerten der Lernzuwachsrate nach beiden Methoden wurde ermittelt.

Zusätzlich wurden die Lernzuwachsrate $sALZ_{pro}$ und $sALZ_{retro}$ in einem Bland-Altman-Plot aufgetragen, um die beiden Methoden miteinander vergleichen zu können (siehe Kapitel 3.5, Seite 38).

Es wurden zur Beantwortung dieser sowie der folgenden Studienfragen die aggregierten Werte aus der *group-level*-Analyse verwendet, da diese, wie unten in Abschnitt 3 (Ergebnisteil) beschrieben, bei einer guten Korrelation mit den objektiven Werten eine geringere Schwankungsbreite als die *individual-level*-Werte aufweisen.

2.7.5 Einfluss der Anzahl der Lernziele auf die Reliabilität der Lernzuwachsrechnung aus vergleichenden studentischen Selbsteinschätzungen (Studienfrage 4)

Zur Beantwortung der vierten Studienfrage wurden 45 für das Modul relevanten Lernziele in der Evaluation erfasst und mittels einer Simulationsstudie die Schwankungsbreite der Lernzuwachsraten je nach einbezogener Zahl der Lernziele ermittelt.

Bisher wurden für jedes der 21 klinischen Module sALZ-Werte für 15 zufällig ausgewählte Lernziele erhoben (siehe Kapitel 1.5.1.1, Seite 14). In der vorliegenden Studie wurden alle 46 für das Modul relevanten Lernziele durch Selbsteinschätzungsfragen erfasst.

In einer Simulationsstudie wurde überprüft, wie stark der mittlere geschätzte Lernzuwachs von der Selektion der verwendeten Parameter abhängt und ob die Nutzung von 15 zufällig ausgewählten Lernzielen zu reliablen Ergebnissen führt.

Es wurden in jeweils 10.000 Simulationsschritten, jeweils 1 aus 45, 2 aus 45, ..., 45 aus 45 Lernzielen zufällig ohne zurücklegen ausgewählt und der subjektive Lernzuwachs aus Selbsteinschätzungen berechnet. Die daraus entstehenden Verteilungen aus jeweils 10.000 Mittelwerten wurden anschließend in einem Boxplot zusammengefasst.

2.7.6 Einfluss des Selbstwertgefühls auf studentische Selbsteinschätzungen (Studienfrage 5)

Zur Beantwortung der 5. Studienfrage wurde das Selbstwertgefühl der teilnehmenden Studierenden mittels der Rosenberg-Skala (Ferring and Filipp 1996) ermittelt.

Um nun überprüfen zu können, ob die Selbsteinschätzungen bei Personen mit unterschiedlichem Selbstwertgefühl variieren, wurden die Teilnehmer je nach Selbstwertgefühl in vier Quartile eingeteilt. Das erste Quartil umfasste die Teilnehmer mit dem geringsten Selbstwertgefühl, das zweite Quartil diejenigen mit dem nächsthöheren etc. Es wurden die sALZ_{pro}-Werte verwendet und ein Mittelwert aus allen Lernzuwachsen pro Lernziel in jedem Quartil gebildet. Es ergab sich für die einzelnen Quartile pro Lernziel ein gemittelter Lernzuwachs.

Auch wurde für jedes Quartil ein Mittelwert aller Lernzuwachsraten gebildet. Die Korrelation nach Pearson zwischen diesen Mittelwerten wurde berechnet, um das Maß des Zusammenhangs zwischen den Quartilen aufzuzeigen.

Eine Rangvarianzanalyse nach Kruskal-Wallis wurde für jedes Lernziel durchgeführt, um zu ermitteln, ob zwischen den Quartilen im Hinblick auf ihre Selbsteinschätzungen bei der Erreichung der Lernziele ein Unterschied besteht oder alle Studierenden unabhängig von ihrem Selbstwertgefühl ihre Leistungen einschätzen können (siehe Kapitel 3.7, 42).

2.7.7 Kriteriumsvalidität, Cut-Off für erfolgreich bzw. erfolglos gelehrte Inhalte (Studienfrage 6)

Zur Beantwortung der Studienfrage 6 wurden die Selbsteinschätzungs- mit den Testatdaten verglichen und eine ROC-Analyse durchgeführt.

Im Testat konnten in jeder Frage zwischen null (alles falsch) und fünf (alles richtig) Punkte erreicht werden. Die Angaben zur Selbsteinschätzung erfolgten auf einer sechsstufigen Likert-Skala, die sich von 1 (stimme voll zu) bis 6 (stimme überhaupt nicht zu) erstreckte.

Um hier eine Vergleichbarkeit zu schaffen, wurde der Bewertungsmaßstab von Selbsteinschätzungen und Testateleistungen angepasst. Die Skala der Selbsteinschätzungen wurde umgedreht (6 = gute Leistung, 1 = schlechte Leistung) und die Skala der Testateleistungen um jeweils einen Punkt gehoben (O = 1 Punkt, 1 = 2 Punkte...). Somit konnten beide Leistungserfassungen auf einer einheitlichen Skala abgebildet werden.

Um die Kriteriumsvalidität des neuen Evaluationsverfahrens zu untersuchen, wurden die Ergebnisse der Selbsteinschätzungen mit den Testatergebnissen als objektive Messwerte verglichen.

Es wurde zunächst mit den Ergebnissen der Testate (oALZ) ein Lernzuwachs pro Lernziel berechnet, ebenso ein Lernzuwachs mit sALZ-Werten. Diese auf *group-level*-Niveau berechneten Lernzuwachsraten wurden in einem Bland-Altman-Plot (Kapitel 3.9.1, Seite 49) der sALZ- und oALZ-Werte dargestellt.

Auch die Korrelation auf *individual-level* wurde mit sILZ- sowie oILZ-Werten berechnet. Es ergaben sich 46 Korrelationskoeffizienten (Pearson's r), pro Student/in eine über alle Lernziele gemittelte Korrelation, deren Verteilung in einem Balkendiagramm dargestellt wurde. Es wurden die ermittelten Korrelationskoeffizienten gegen die Anzahl der Studierenden aufgetragen, bei denen sich eben diese Korrelation zeigte (Kapitel 3.9.2, Seite 50).

Durch die Berechnung der Korrelationen sowohl auf *group*- als auch auf *individual-level* sollten mögliche Differenzen zwischen den Berechnungsformen aufgedeckt und diejenige Berechnungsform ermittelt werden, die die bessere Vergleichbarkeit mit objektiven Daten zeigt. Diese Vergleichbarkeit wird für die Beurteilung der Kriteriumsvalidität benötigt.

Nach dem Vergleich der subjektiven mit den objektiven Daten wurde eine ROC (Receiver Operating Characteristic)-Analyse durchgeführt, um den günstigsten Cut-Off-Wert (hohe Sensitivität bei niedriger Falsch-Positiv-Rate, hohe Trennschärfe) zu ermitteln.

Die ROC-Analyse wird in der Medizin verwendet, um die Trennschärfe eines Diagnoseverfahrens, insbesondere im Bereich der bildgebenden Verfahren, beurteilen zu können (Konietzko et al. 1995).

Die ROC-Kurve wird in einem Diagramm mit den Achsen Sensitivität und 1- Spezifität dargestellt (Cook 2007). Die Sensitivität beschreibt den Anteil der gut gelehnten Themen, die mittels der Evaluation als solche erkannt werden. Die Spezifität würde die schlecht gelehnten Themen, die als solche erkannt werden, umfassen. 1-Spezifität wird oft auch als „Falsch-Positiv-Rate“ bezeichnet und zeigt hier den Anteil der Lernziele, die zwar als gut gelehnt erfasst wurden, im objektiven Test aber ein eher schlechtes Ergebnis ergaben. Im Diagramm wurden Cut-Off-Werte mit ihrer jeweiligen Sensitivität und 1-Spezifität-Rate eingefügt.

In vorliegender Studie wurde die ROC-Analyse genutzt, um das Vermögen der Evaluationsmethode bezüglich der Unterscheidung zwischen erfolgreich und erfolglos gelehnten Themen zu erfassen. Im ersten Schritt wurde ein oAGP-Cut-Off-Wert definiert: eine Steigerung von zwei auf vier (aus sechs möglichen) richtigen Antworten im Test vor und nach dem Modul entsprach einem ALZ-Wert von 50%. Ein solcher Zuwachs kann als gut betrachtet werden, so dass er als Cut-Off für den objektiven aggregierten Lernzuwachs verwendet wurde. In der ROC-Analyse wurden verschiedene Cut-Off-Werte für den subjektiven aggregierten Lernzuwachs im Hinblick auf ihre Trennschärfe zwischen guten und schlechten objektiven Lernzuwachsraten getestet. Als Maßstab wurde hierbei der oALZ-

Wert > 50% zur Beurteilung gut gelehrter Lernziele genutzt. Der optimale sALZ-Cut-Off-Wert ergab sich aus der besten Sensitivität bei maximaler Spezifität.

2.8 Statistische Auswertung

Zur Auswertung der Selbsteinschätzungs- sowie Testatdaten wurde die Statistiksoftware SPSS® (14.0) verwendet.

In der Analyse der vorliegenden Daten wurden verschiedene Abkürzungen für die Angabe statistischer Merkmale verwendet. Der Mittelwert wurde als MW angegeben. Die Standardabweichung wurde immer als SD abgekürzt. Der Mittelwert wurde als MW +/- SD angegeben.

Cohen's d gibt die Effektstärke für Mittelwertunterschiede zwischen zwei Gruppen mit gleichen Gruppengrößen und –varianzen an. Er ist für die Beurteilung der praktischen Relevanz eines signifikanten Mittelwertunterschieds hilfreich. Nach Cohen indiziert $d = 0,2$ einen kleinen Effekt, $d = 0,5$ einen mittleren und $d = 0,8$ einen starken Effekt (Cohen 1992).

Die berechneten Korrelationen werden mit Hilfe des Korrelationskoeffizienten r angegeben. Ein $r > 0$ ist dabei Hinweis auf einen positiven, ein $r < 0$ auf einen negativen Zusammenhang.

Diese Ergebnisse sind nur dann aussagekräftig, wenn sie auch signifikant sind. $r < 0,3$ wird als kleine, $r < 0,5$ als gute und $r > 0,7$ als sehr hohe Korrelation interpretiert. Die Signifikanz der Korrelationen ist auf einem Niveau von $\alpha < 0,05$ (*), $\alpha < 0,01$ (**) oder $\alpha < 0,001$ (***) angegeben. $\alpha < 0,01$ bedeutet hier, dass in weniger als 1% der Fälle die Korrelation zufällig auftritt.

3 Darstellung der eigenen Untersuchungen („Ergebnisse“)

3.1 Fokusgruppengespräche (Studienfrage 1)

In den Fokusgruppengesprächen ergaben sich vier wichtige, an den Leitfragen orientierte Themen. Dies waren die Lehrqualität, die Vorstellung von guter Evaluation, Evaluationsinstrumente und Konsequenzen die sich aus der Evaluation ergeben sollten.

Die wichtigsten Aussagen sind in folgender Tabelle dargestellt (Schiekirka et al. 2012).

Tabelle 2: Kernaussagen der Fokusgruppenteilnehmer zu Kennzeichen guter Lehre

Teil A: Lehrqualität

INHALT	PROZESS	CHARAKTERISTIKA DER LEHRENDEN	LERNZUWACHS
Unterricht dem Niveau der Studenten angepasst	Interaktive Lehrformen	Ausreichendes Fachwissen	Keine Begrenzung auf einige wichtige Themen
Betonung wichtiger Themen	Freie Verfügbarkeit von Lehrmaterial	Klinische Erfahrung	Vorbereitung für lebenslanges Lernen
Keine zu häufigen Wiederholungen	Häufige Rückmeldungen bezüglich der praktischen Fertigkeiten	Gute Vorbereitung	
Widerspiegeln relevanter Themen in Prüfungen		Motivation	

Teil B: Vorstellungen von Evaluation

ZIELE	VERHINDERUNG DER TEILNAHME	STÖRFAKTOREN
Lernprozesse und Outcome verbessern	Zu häufige Evaluationen	Individuelle Definitionen von „guter“ Lehre
Individuelle Rückmeldungen an Lehrende	Mangel an Feedback	Motivation der Studenten
Erkennung, welche Lernziele erfüllt wurden	Mangel an Zeit für die Evaluation	Positive Erwartungen
		Halo-Effekt

Teil C: Evaluationsinstrumente und Datensammlung

ZIELE DER EVALUATION	BEVORZUGTE FORM	BEVORZUGTER ZEITPUNKT
Lernzuwachs	Offene Fragen	Nach der Klausur
Prozesse und Struktur der Lehre (Stundenpläne, Einrichtung etc.)	Nicht mehr als 15 Items	Kontinuierlich (Online-Plattform)
Angemessenheit der Prüfung	Online	
Individuelle Gestaltung durch den/die Lehrende(n)	freiwillig	

Teil D: Konsequenzen der Evaluation

VERÖFFENTLICHUNG DER ERGEBNISSE	BELOHNUNG POSITIVER ERGEBNISSE	KONSEQUENZEN NEGATIVER ERGEBNISSE	FEEDBACK
Sichtbar für alle Fakultäten	Geldmittel	Verbindlicher Lehr-Trainings-Kurs	Bekanntgabe der Konsequenzen als Antwort auf die Evaluation
Ranking der Module	Zusätzliche Zeit für Lehre	Ausschluss von der Lehrtätigkeit	
Sicherung der individuellen Daten			
Diskussion besonderer Ergebnisse in einem Evaluationsausschuss			

Die in der Tabelle dargestellten Kernaussagen beruhen auf verschiedenen Äußerungen der Fokusgruppenteilnehmer. In den vertikalen Spalten sind die zu einem Thema gehörenden Stichworte aufgeführt, die horizontale Richtung ist hier unwichtig.

Zu jedem Themengebiet ergaben sich charakteristische Meinungen der Studierenden, die mittels Zitat im Folgenden aufgeführt werden.

Zur Lehrqualität sagte ein Student des vierten Studienjahres: „Für mich ist gute Lehre etwas, das in klinische Praxis umgesetzt werden kann.“ Eine Studentin desselben Studienjahres äußerte sich folgendermaßen: „Ich denke, interaktive Lehre in Kleingruppen ist der beste Weg zu lehren und zu lernen.“ In Bezug auf die Lehrenden meinte eine Studentin des vierten Jahres, dass die Lehre völlig vom Lehrer abhängig sei.

Wie Studenten sich die Evaluation vorstellten, zeigte folgende Aussage: „Im Allgemeinen würde ich dem Kurs eine positive Bewertung geben, in dem ich das Gefühl habe, auf angenehme Weise viel gelernt zu haben.“

Zur Datensammlung äußerten sich die Studierenden eher kritisch. „Im Großen und Ganzen ist es leicht Statistiken zu analysieren, doch ich denke nicht, dass diese irgendetwas aussagen.“ (Studentin im dritten Studienjahr). Eher gewünscht sind konkrete Aussagen, wie eine weitere Studentin des 3. Studienjahres forderte: „Also gerade, wenn da halt konkrete Sachen stehen. Ich denke, auch wenn ich als Dozent sage, ich habe insgesamt von allen Studenten eine drei gekriegt, ja, also was soll ich damit anfangen als Dozent - wenn da aber steht ‚War prinzipiell gut, aber die PowerPoint-Folien waren irgendwie doof gestaltet oder so.‘, dann ist das ja wirklich ein konkreter Punkt, wo ich sage, da kann ich an mir arbeiten.“

Zu den Konsequenzen der Evaluation sagte ein Student des 4. Studienjahres: „ Wenn jemand einfach nicht an der Lehre interessiert ist, wird sich nichts ändern, da sein Job sicher ist – er ist einfach nicht

interessiert.“ Eine Studentin desselben Studienjahres äußerte sich folgendermaßen: „Ich denke, dass mehr Studenten bereit wären zu evaluieren, wenn sie wüssten, dass es etwas nützt.“

Die hier aufgeführten Aussagen der Studierenden stellen einen charakteristischen Querschnitt über die in den Fokusgruppengesprächen geäußerten Meinungen der Studenten/innen dar.

Zu betonen ist nochmals, dass der Lernzuwachs für die Beurteilung der Lehre als sehr wichtig eingeschätzt wurde: „Und wie würdet ihr selbst dann Lehre evaluieren?“ „Wie der Lernzuwachs ist.“ (Antwort einer Studentin des vierten Studienjahres auf die Frage der Gesprächsleiterin.)

3.2 Deskriptive Statistiken der Stichprobe und Rücklaufquote

Von den 145 Studierenden, die im Sommersemester 2011 zum Modul 3.1 angemeldet waren, nahmen 83 an allen Teilen der quantitativen Studie teil. Somit ergab sich eine Beteiligung von 57,2%. Von diesen 83 Personen waren 33 männlichen und 50 weiblichen Geschlechts. Die Altersspanne erstreckte sich von 21 bis 32 Jahre, der Mittelwert betrug $24,8 \pm 2,3$ Jahre.

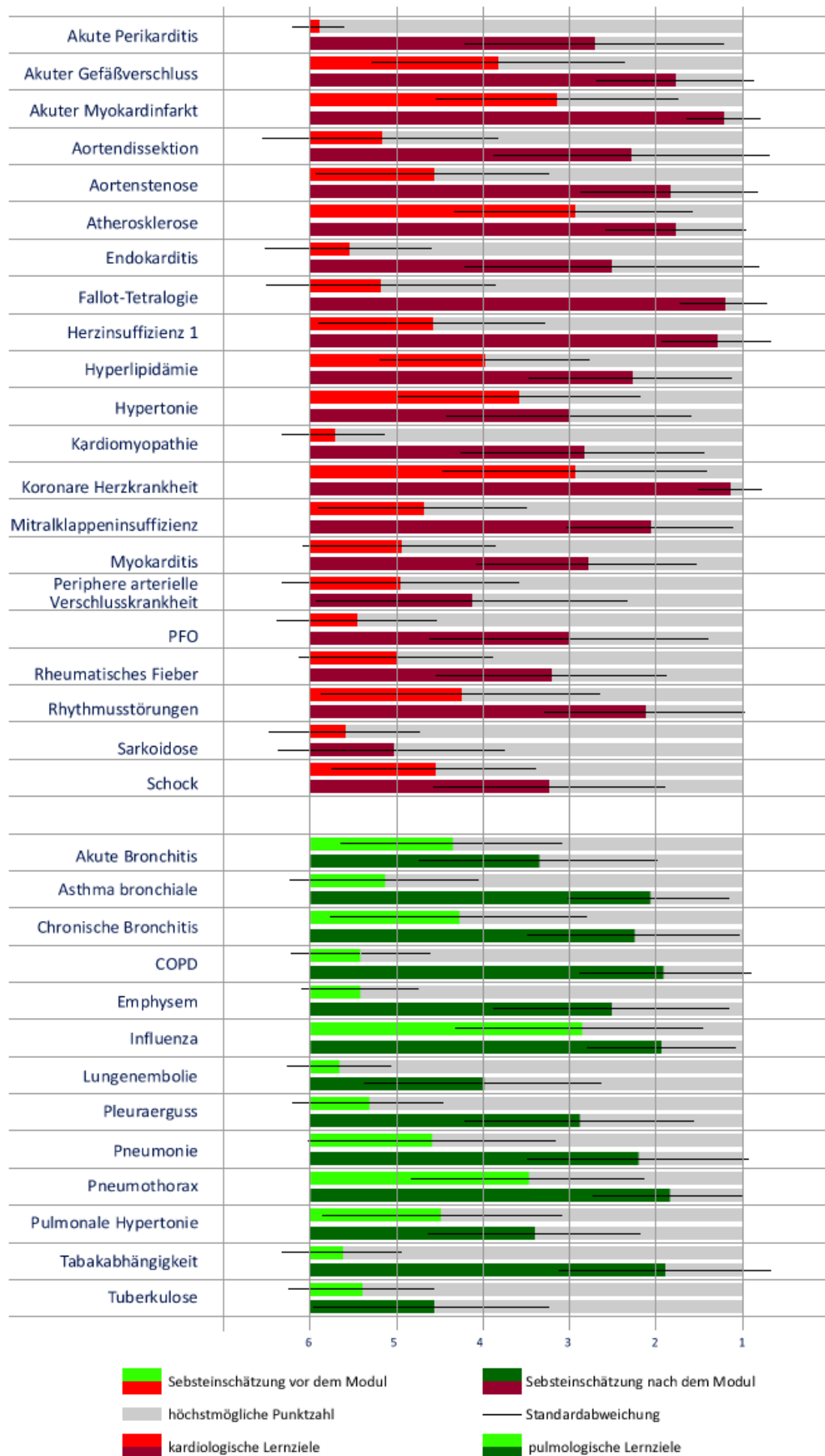
3.3 Darstellung der Lernzuwachsdaten

Zunächst werden die subjektiven aggregierten Lernzuwachsdaten (sALZ) dargestellt. Im Mittel wurde ein Zuwachs von $56,9 \pm 22,2\%$ erreicht (Bereich: 12,0% - 95,3%). Die Mittelwerte der Selbsteinschätzungen sind in Abbildung 1, die pro Lernziel berechneten mittleren Lernzuwachsraten in Abbildung 2 dargestellt.

Das Lernziel „Herzinsuffizienz“ ist in den dargestellten Ergebnissen zweimal aufgeführt. Dies ist darauf zurückzuführen, dass es sowohl im Rahmen von Wissensinhalten als auch bezüglich der praktischen Fertigkeiten in die Selbsteinschätzung einfließt, in der Studie also zwei Fragen zu diesem Thema gestellt wurden.

Herzinsuffizienz 1 stellt hierbei das kognitive, Herzinsuffizienz 2 das praktische Lernziel dar.

Abbildung 1: Mittelwerte der Selbsteinschätzungen aller betrachteten Lernziele mit Standardabweichung
 Teil A: kardiologische Lernziele



Teil B: praktische Lernziele

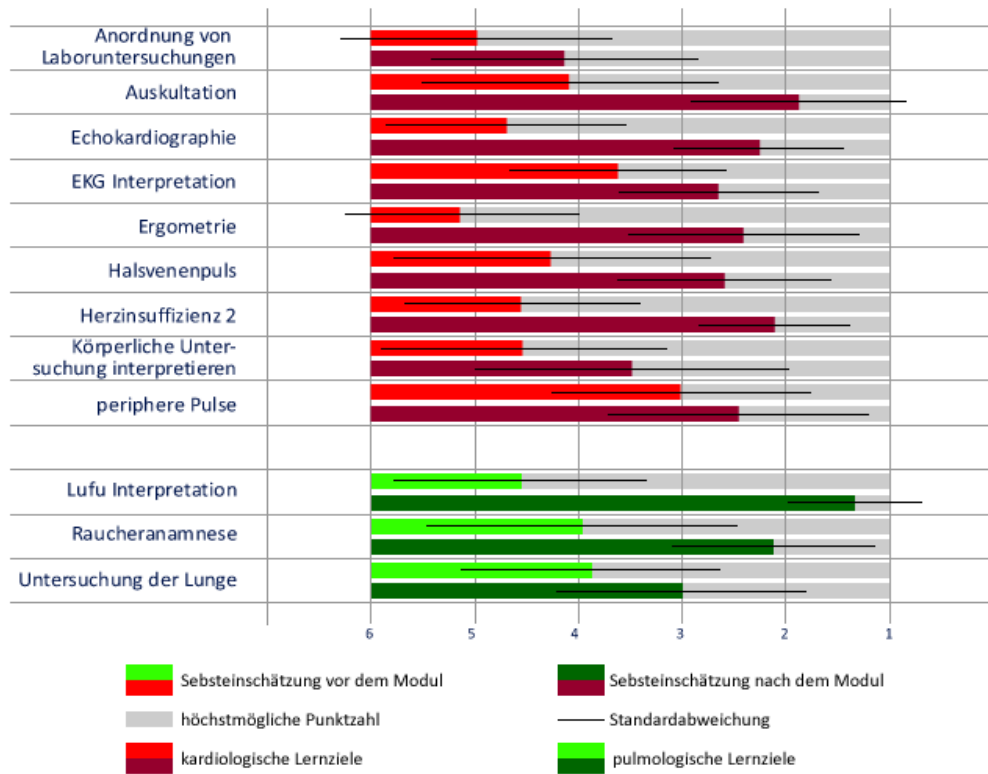
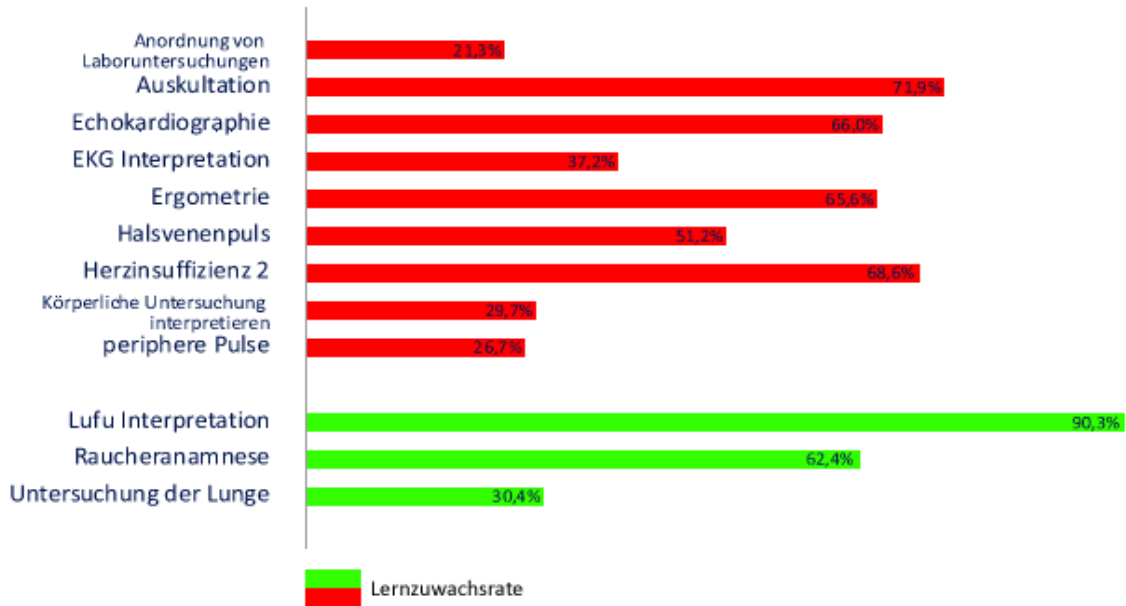


Abbildung 2: Lernzuwachsrate berechnet aus den Mittelwerten der pro Lernziel aggregierten studentischen Selbsteinschätzungen (sALZ) vor und nach dem Modul
 Teil A: kognitive Lernziele



Teil B: praktische Lernziele



Aus Abbildung 1 lässt sich erkennen, dass das Vorwissen zwar meist eher gering war, trotzdem aber eine Spannweite zwischen 2,87 und 5,90 aufwies. Somit lag hier kein Bodeneffekt vor, sodass die Selbsteinschätzungsfragen auch bei niedrigem Wissensstand noch gut differenzieren. Auch in den Selbsteinschätzungen am Modulende zeigte sich eine hohe Spannweite mit einem Bereich von 1,13 bis 5,05. Hinweise auf einen Deckeneffekt, die Selbsteinschätzungsfragen könnten dann bei hohem Wissensstand nicht mehr gut differenzieren, lagen nicht vor.

Abbildung 2 stellt die jeweiligen Lernzuwachsrate pro Lernziel dar.

Ein hohes oder geringes Vorwissen ist kein Anhaltspunkt für den Lernzuwachs. Für das Lernziel „KHK“ z.B. schien das Vorwissen mit einem Mittelwert von 2,93 relativ ausgeprägt zu sein. Trotzdem konnte ein Lernzuwachs von 93,1% erzielt werden. Die „Fallot-Tetralogie“ dagegen war vor Absolvierung des Moduls eher unbekannt (Selbsteinschätzung 5,18). Der Lernzuwachs war trotzdem mit 95,3% mit dem des Lernziels „KHK“ vergleichbar. Andererseits gab es Lernziele wie die „Sarkoidose“ mit geringer oder „Hypertonie“ mit eher hoher initialer Selbsteinschätzung, bei denen nur ein geringer Lernzuwachs, hier 12,0% bzw. 22,9%, verzeichnet werden konnte. Daraus lässt sich schließen, dass der Lernzuwachs trotz großen Vorwissens erfasst werden kann.

Die Lernzuwachsrate war bei den kognitiven Lernzielen mit durchschnittlich 59,3% höher als bei den praktischen, die einen Lernzuwachs von 46,6% aufwiesen.

Auch in der Betrachtung der kardiologischen und pulmologischen Lernziele war ein Unterschied erkennbar. Der Lernzuwachs der pulmologischen Themen betrug durchschnittlich nur 47,36%, während in den kardiologischen Lernzielen ein Zuwachs von 63,93% zu verzeichnen war.

3.4 Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (*individual-level vs. group-level*) (Studienfrage 2)

Zu Beantwortung der Studienfrage 2 wurden die Ergebnisse der prospektiven Selbsteinschätzungen und der Selbsteinschätzungen nach dem Modul verwendet. Die Mittelwerte pro Lernziel wurden auf *group-level*- wie auch auf *individual-level*-Niveau erhoben und die durchschnittlichen Lernzuwachsrate pro Lernziel miteinander korreliert. Die Mittelwerte der Lernzuwächse sind in Tabelle 3 dargestellt.

Zusätzlich sind die Lernzuwachsrate, berechnet aus den retrospektiv erhobenen Selbsteinschätzungen ($sALZ_{retro}$), in der Tabelle dargestellt. Diese werden in Kapitel 3.5 erläutert.

Die kardiologischen Themen sind rot, die pulmologischen grün hinterlegt. Die Tabelle zeigt in Teil A die kognitiven, in Teil B die praktischen Lernziele.

Tabelle 3: Lernzuwachsrate bezogen auf die in den Selbsteinschätzungsfragen verwendeten Lernziele – *group-level* ($sALZ$) und *individual-level* ($sILZ$); zusätzlich $sALZ_{retro}$

Teil A: kognitive Lernziele

LERNZIEL	sALZ	sILZ	sALZ _{retro}
		MW (95%-Konfidenzintervall)	
Akute Perikarditis	65,2%	52,1% (40,5-63,7%)	62,2%
Akuter Gefäßverschluss	72,7%	73,3% (60,8-85,9%)	75,5%
Akuter Myokardinfarkt	90,9%	89,4% (79,9-98,9%)	92,6%
Aortendissektion	69,5%	76,2% (65,1-87,2%)	67,3%
Aortenstenose	76,4%	71,4% (60,8-82,0%)	78,6%
Arteriosklerose	60,2%	65,6% (51,8-79,3%)	67,7%
Endokarditis	67,1%	67,9% (55,5-80,3%)	66,9%
Fallot-Tetralogie	95,3%	95,0% (90,2-99,8%)	95,1%
Herzinsuffizienz 1	91,6%	92,7% (86,0-99,5%)	92,6%
Hyperlipidämie	57,1%	57,3% (40,2-74,3%)	64,7%
Hypertonie	22,9%	15,2% (-5,3 - 35,6%)	44,2%
Kardiomyopathie	61,3%	51,1% (36,8-65,4%)	59,7%
Koronare Herzkrankheit	93,1%	95,0% (87,4-102,6%)	94,1%
Mitralklappeninsuffizienz	71,2%	67,8% (58,2-77,4%)	73,4%
Myokarditis	54,5%	49,4% (34,8-64,1%)	57,8%
Periphere arterielle Verschlusskrankheit	20,7%	18,1% (2,5-33,8%)	25,6%
PFO	54,7%	46,3% (32,1-60,5%)	54,8%
Rheumatisches Fieber	45,1%	34,3% (20,9-47,7%)	51,4%
Rhythmusstörungen	65,5%	60,7% (48,4-73,0%)	71,1%
Sarkoidose	12,0%	16,1% (1,5-30,6%)	14,7%
Schock	37,3%	33,7% (20,0-47,4%)	40,7%

Akute Bronchitis	29,9%	22,1% (4,6-39,5%)	37,4%
Asthma bronchiale	73,8%	71,3% (63,1-80,5%)	74,8%
Chronische Bronchitis	62,1%	62,7% (47,6-77,8%)	66,1%
COPD	79,8%	81,2% (72,8-89,6%)	79,5%
Emphysem	65,5%	68,8% (58,6-79,0%)	66,0%
Influenza	50,4%	49,4% (33,7-65,2%)	67,1%
Lungenembolie	36,3%	43,8% (33,6-54,0%)	36,4%
Pleuraerguss	56,5%	47,7% (36,2-59,1%)	56,6%
Pneumonie	66,4%	57,3% (42,4-72,3%)	68,7%
Pneumothorax	65,2%	56,7% (41,5-72,0%)	71,1%
Pulmonale Hypertonie	31,2%	25,5% (10,7-40,4%)	44,1%
Tabakabhängigkeit	81,0%	83,6% (74,4-92,9%)	79,9%
Tuberkulose	18,4%	18,4% (4,2-32,6%)	20,0%

Teil B: praktische Lernziele

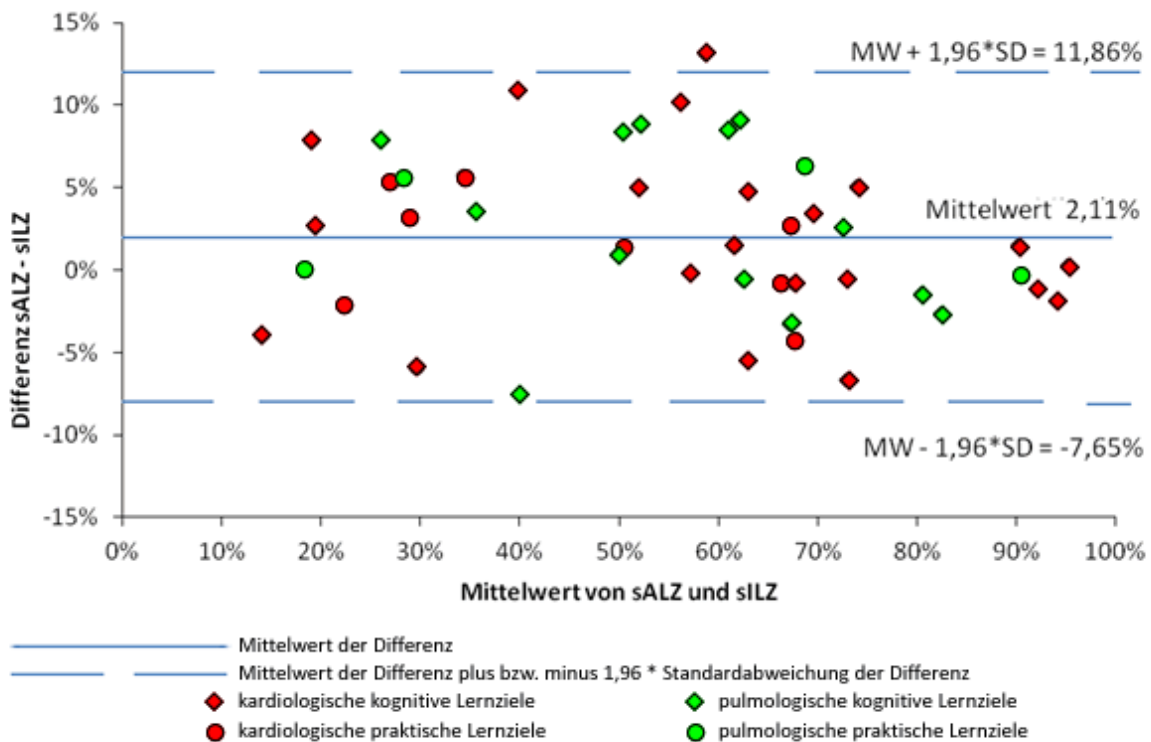
	sALZ	sILZ	sALZ _{retro}
LERNZIEL		MW (95%-Konfidenzintervall)	
Anordnung von Laboruntersuchung	21,3%	23,5% (10,5-36,5%)	25,9%
Auskultation	71,9%	65,5% (51,5-79,4%)	77,2%
Echokardiographie	66,0%	66,7% (58,5-74,9%)	68,3%
EKG-Interpretation	37,2%	31,5% (15,9-47,1%)	53,7%
Ergometrie	65,6%	69,8% (61,0-78,5%)	64,5%
Halsvenenpuls	51,2%	49,8% (36,3-63,3%)	61,0%
Herzinsuffizienz 2	68,6%	66,0% (60,7-71,3%)	69,3%
Körperliche Untersuchung interpretieren	29,7%	24,3% (13,2-35,4%)	35,7%
periphere Pulse	26,7%	32,6% (15,8-49,5%)	37,7%
Lungenfunktions-Interpretation	90,3%	90,6% (85,0-96,2%)	89,9%
Raucheranamnese	62,4%	60,9% (52,8-68,9%)	65,7%
Untersuchung der Lunge	30,4%	27,2% (16,4-38,1%)	44,4%

Die sILZ-Werte wiesen eine große Spannweite auf (7,7% bis 94,8%).

Die Korrelation der aggregierten (ALZ; Spalte 2) mit den gemittelten individuellen Werten (ILZ; Spalte 3) lag nach Pearson bei 0,996 ($p < 0,001$).

Ein Bland-Altman-Plot zeigt die Abweichungen der beiden Messmethoden. In Abbildung 3 ist der Mittelwert des Unterschieds zwischen beiden Methoden aufgetragen sowie die Unterschiede in den einzelnen Lernzielen. Die Schwankungsbreite lag zwischen 0,02% (Tuberkulose) und 10,81% (Rheumatisches Fieber). Die kognitiven Lernziele sind in Form von Rauten, die praktischen Lernziele als graue Kreise dargestellt.

Abbildung 3: Bland-Altman-Plot der Lernzuwachsrate auf sALZ- und sILZ-Niveau



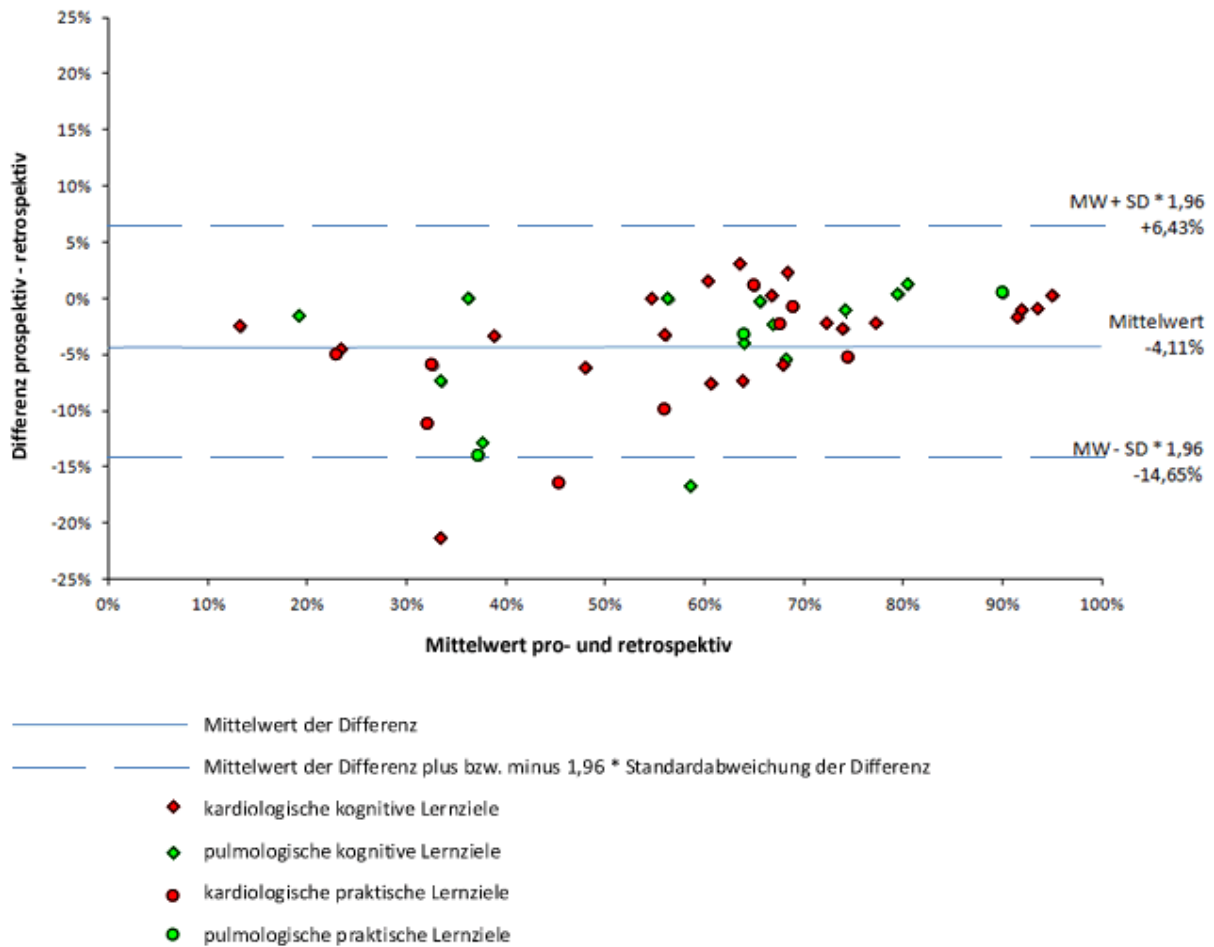
In Tabelle 3 wurden die großen Spannweiten der Lernzuwachsrate auf *individual-level* gezeigt. Zusammen mit der Betrachtung des Bland-Altman-Plots ergibt sich nun, dass eine Verwendung der aggregierten Daten der *group-level*-Analyse die bessere Alternative darstellt. In den nächsten Kapiteln werden deshalb nur die ALZ, nicht mehr die ILZ-Werte, verwendet.

3.5 Prospektive versus retrospektive Selbsteinschätzung des initialen Lernzuwachses (Studienfrage 3)

Zur Beantwortung der Studienfrage 3 wurden die Studierenden aufgefordert, ihren initialen Leistungsstand zusätzlich retrospektiv nach dem Modul einzuschätzen. Die Lernzuwachsrate, die aus den retrospektiv erhobenen Selbsteinschätzungen errechnet wurden, sind in Tabelle 3 (Kapitel 3.4) dargestellt. Es ergab sich eine Korrelation des Mittelwerts der prospektiven mit den retrospektiven Lernzuwachsrate nach Pearson von 0,974 ($p < 0,01$).

Ein Bland-Altman-Plot der pro- und retrospektiv errechneten Lernzuwachsraten (Abbildung 4) zeigt, dass die beiden Methoden durchaus vergleichbar sind. Die Schwankungsbreite der Differenz zwischen pro- und retrospektiver Einschätzung liegt zwischen 0,05% (PFO, Lungenembolie) und 21,28% (Hypertonie).

Abbildung 4: Bland-Altman-Plot der Lernzuwachsrate berechnet mit prospektiv und retrospektiv erfassten Selbsteinschätzungen zu Modulbeginn sowie Selbsteinschätzungen zu Modulende

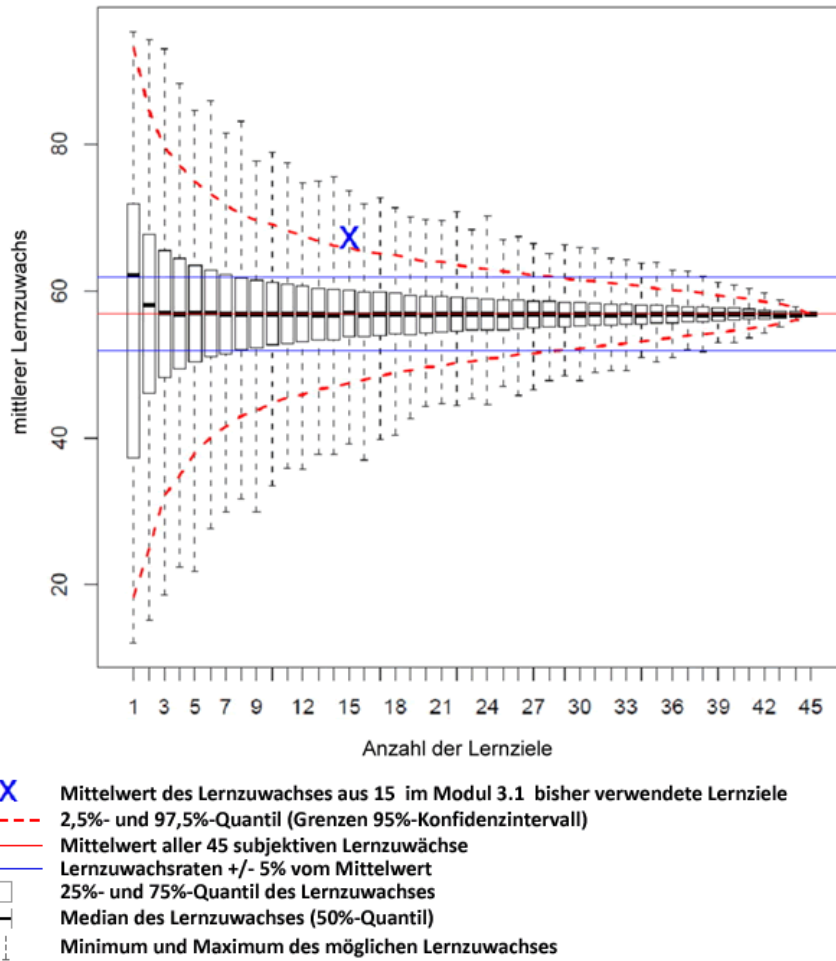


Wie in Kapitel 3.4 sind auch hier die kognitiven Lernziele als Raute, die praktischen als Kreis dargestellt. Die kardiologischen Lernziele sind rot, die pulmologischen grün gekennzeichnet. Der Mittelwert der Differenz ist mit 4,11% gering, ebenso die Standardabweichung mit 5,38%.

3.6 Einfluss der Anzahl der Lernziele auf die Reliabilität der Lernzuwachsrechnung aus vergleichenden studentischen Selbsteinschätzungen (Studienfrage 4)

Zur Beantwortung der Studienfrage 4 wurden 45 für das Modul 3.1 relevante Lernziele einbezogen. Es wurde in einer Simulationsstudie überprüft, wie stark der mittlere Lernzuwachs von der Auswahl der verwendeten Lernziele abhängt. Ein Lernziel wurde aus der Berechnung genommen (siehe Kapitel 2.5), deshalb sind nur 45 statt 46 Lernziele in die Berechnung einbezogen. Abbildung 5 zeigt einen Boxplot mit der errechneten Schwankungsbreite des Lernzuwachses abhängig von der Anzahl der Lernziele.

Abbildung 5: Darstellung der Schwankungsbreite je nach Anzahl der erfassten Lernziele mittels der sALZ-Werte



Die Abbildung zeigt die Bandbreite der Lernzuwachsrate in Abhängigkeit von der Anzahl der einbezogenen Ziele. Bei der Auswahl eines einzigen Lernziels zeigen sich alle möglichen zu erreichenden Zuwächse (12,0% bis 95,3%; MW 56,9% ± 22,2%; siehe Kapitel 3.3). Bei einer zufälligen Auswahl von 15 Lernzielen (aktuelle Praxis an der UMG) konnte der errechnete Mittelwert zwischen 36,6% und 73,2% liegen. Das 95%-Konfidenzintervall reichte dabei von 47,4% bis 66,0%. Die Breite des 95%-Konfidenzintervalls lag erst bei der Mittelung von 30 Lernzielen unter 10% und die gesamte Schwankungsbreite unterschritt den Wert von 10% erst bei der Verwendung von 39 Lernzielen.

3.7 Einfluss des Selbstwertgefühls auf die studentischen Selbsteinschätzungen (Studienfrage 5)

Um die 5. Studienfrage zu beantworten wurden die Studierenden gebeten 10 Fragen zum Selbstwertgefühl zu beantworten. Das Selbstwertgefühl nach Rosenberg wurde für jede/n Studierende/n errechnet und die gesamte Kohorte in vier Quartile eingeteilt. Anschließend wurde der Vergleich der Mittelwerte aller Selbsteinschätzungen sowie der Mittelwerte pro Lernziel zwischen den Quartilen tabellarisch dargestellt. In Tabelle 5 und 6 ist das Ergebnis des Kruskal-Wallis-Tests für jedes Lernziel aufgeführt.

Das Cronbach's α der Rosenberg-Skala betrug in dem untersuchten Kollektiv 0,89.

Die Studierenden erreichten einen mittleren Wert von $24,7 \pm 5,0$ Punkten. Zwischen männlichen und weiblichen Studierenden fand sich kein signifikanter Unterschied (MW $26,0 \pm 3,5$ vs. MW $24,0 \pm 5,7$ ($p = 0,083$)).

Tabelle 4 zeigt die Quartilengrenzen, die sich nach der Punkteverteilung der Rosenbergskala ergaben, sowie die Mittelwerte der Antworten auf die Fragen nach dem Selbstwertgefühl.

Tabelle 4: Quartile nach Selbstwertgefühl

	Quartil 1	Quartil 2	Quartil 3	Quartil 4
N (Anzahl)	21	23	19	19
Quartilengrenzen (Punkte Rosenbergskala)	< 22	22-25	26-28	≥ 28
Mittelwert des Selbstwertgefühls nach Rosenberg (\pm SD)	17,8	24,8	27,7	29,4

Tabelle 5 gibt die nach Quartilen unterteilten studentischen Selbsteinschätzungen zu Modulbeginn wieder, Tabelle 6 die Selbsteinschätzungen zu Modulende.

Für jedes Lernziel wurde ein Kruskal-Wallis-Test durchgeführt. Die p-Werte lagen zu Modulbeginn für die kognitiven Lernziele zwischen 0,007 (akute Bronchitis) und 0,931 (PFO), für die praktischen zwischen 0,178 (Anordnung von Laboruntersuchungen) und 0,968 (Untersuchung der Lunge). Am Modulende lagen die p-Werte für die kognitiven Lernziel zwischen 0,001 (Emphysem) und 0,839 (akuter Gefäßverschluss), für die praktischen zwischen 0,004 (EKG-Interpretation) und 0,721 (Ergometrie). Die Kruskal-Wallis-Tests konnten zeigen, dass die ermittelten Werte in den vier Quartilen der gleichen Grundgesamtheit entstammen.

Die kardiologischen Themen sind rot, die pulmologischen Themen grün hinterlegt. Teil A der Tabelle zeigt die kognitiven, Teil B die praktischen Lernziele.

Tabelle 5: Selbsteinschätzungen gemittelt über die Selbstwert-Quartile pro Lernziel; Einschätzung zu Modulbeginn

Teil A: kognitive Lernziele

LERNZIEL	Quartil 1	Quartil 2	Quartil 3	Quartil 4	p-Wert (Kruskal-Wallis-Test)
	MW ± SD	MW ± SD	MW ± SD	MW ± SD	
Akute Perikarditis	5,90 ± 0,30	5,78 ± 0,42	5,95 ± 0,23	6,00 ± 0,00	0,147
Akuter Gefäßverschluss	3,86 ± 1,59	3,91 ± 1,28	3,63 ± 1,61	3,74 ± 1,41	0,308
Akuter Myokardinfarkt	3,05 ± 1,36	3,22 ± 1,38	2,89 ± 1,56	3,37 ± 1,42	0,301
Aortendissektion	5,29 ± 1,35	4,91 ± 1,50	5,47 ± 0,90	5,11 ± 1,59	0,140
Aortenstenose	4,52 ± 1,17	4,70 ± 1,22	4,11 ± 1,76	4,95 ± 1,22	0,350
Atherosklerose	3,10 ± 1,22	2,78 ± 1,09	2,58 ± 1,46	3,32 ± 1,73	0,512
Endokarditis	5,48 ± 1,25	5,39 ± 1,08	5,53 ± 0,90	5,84 ± 0,37	0,334
Fallot-Tetralogie	4,90 ± 1,51	5,17 ± 1,37	5,05 ± 1,47	5,58 ± 0,84	0,040
Herzinsuffizienz 1	4,90 ± 0,83	4,48 ± 1,16	4,00 ± 1,56	4,89 ± 1,49	0,342
Hyperlipidämie	4,10 ± 1,18	4,04 ± 0,98	3,84 ± 1,46	3,79 ± 1,27	0,203
Hypertonie	3,52 ± 1,50	3,61 ± 1,27	3,84 ± 1,50	3,21 ± 1,36	0,256
Kardiomyopathie	5,52 ± 0,81	5,87 ± 0,34	5,72 ± 0,67	5,79 ± 0,42	0,903
Koronare Herzkrankheit	2,95 ± 1,28	3,00 ± 1,54	2,68 ± 1,70	2,89 ± 1,56	0,062
Mitralklappeninsuff.	4,62 ± 1,02	4,74 ± 1,14	4,47 ± 1,54	4,95 ± 1,22	0,704
Myokarditis	5,00 ± 1,18	5,09 ± 1,06	4,95 ± 0,91	4,79 ± 1,40	0,631
Periph. art. Verschl.kr.	5,00 ± 1,14	4,74 ± 1,71	5,32 ± 1,06	4,95 ± 1,35	0,479
PFO	5,38 ± 0,86	5,65 ± 0,57	5,16 ± 1,42	5,58 ± 0,77	0,931
Rheumatisches Fieber	4,71 ± 1,15	5,17 ± 1,03	5,16 ± 1,21	4,95 ± 1,18	0,748
Rhythmusstörungen	4,33 ± 1,39	4,39 ± 1,67	3,95 ± 1,65	4,26 ± 1,82	0,074
Sarkoidose	5,52 ± 1,08	5,61 ± 0,72	5,74 ± 0,73	5,53 ± 0,96	0,186
Schock	4,67 ± 1,15	4,57 ± 1,27	4,26 ± 1,33	4,68 ± 1,11	0,030
Akute Bronchitis	4,19 ± 1,25	4,26 ± 1,25	4,68 ± 1,49	4,26 ± 1,33	0,007
Asthma bronchiale	4,95 ± 0,97	5,26 ± 1,05	5,32 ± 1,20	5,05 ± 1,22	0,396
Chronische Bronchitis	4,19 ± 1,44	4,61 ± 1,34	4,00 ± 1,63	4,16 ± 1,54	0,743
COPD	5,29 ± 0,85	5,57 ± 0,59	5,32 ± 1,16	5,53 ± 0,51	0,457
Emphysem	5,38 ± 0,67	5,48 ± 0,79	5,32 ± 0,75	5,53 ± 0,51	0,328
Influenza	2,67 ± 1,35	3,13 ± 1,52	2,63 ± 1,30	2,89 ± 1,56	0,752
Lungenembolie	5,57 ± 0,75	5,70 ± 0,56	5,79 ± 0,54	5,63 ± 0,60	0,131
Pleuraerguss	5,19 ± 1,12	5,22 ± 0,80	5,33 ± 0,97	5,58 ± 0,61	0,522
Pneumonie	4,62 ± 1,50	4,78 ± 1,24	4,56 ± 1,54	4,32 ± 1,57	0,437
Pneumothorax	3,43 ± 1,54	3,48 ± 1,20	3,42 ± 1,26	3,53 ± 1,50	0,610
Pulmonale Hypertonie	4,38 ± 1,36	4,61 ± 1,20	4,26 ± 1,48	4,68 ± 1,53	0,318
Tabakabhängigkeit	5,57 ± 0,60	5,61 ± 0,89	5,63 ± 0,83	5,74 ± 0,45	0,151
Tuberkulose	5,48 ± 0,81	5,13 ± 1,01	5,53 ± 0,70	5,53 ± 0,77	0,130

Teil B: praktische Lernziele

LERNZIEL	Quartil 1	Quartil 2	Quartil 3	Quartil 4	p-Wert (Kruskal-Wallis-Test)
	MW ± SD	MW ± SD	MW ± SD	MW ± SD	
Anordn. von Laborunters.	4,71 ± 1,42	4,87 ± 1,46	5,32 ± 1,00	4,89 ± 1,24	0,178
Auskultation	3,62 ± 1,50	4,13 ± 1,32	4,16 ± 1,46	4,26 ± 1,37	0,502
Echokardiographie	4,43 ± 1,29	4,96 ± 1,19	4,58 ± 1,12	4,63 ± 1,01	0,553
EKG-Interpretation	3,62 ± 0,97	3,78 ± 1,20	3,61 ± 0,85	3,32 ± 0,95	0,233
Ergometrie	5,24 ± 1,09	4,83 ± 1,03	5,37 ± 1,07	5,00 ± 1,33	0,208
Halsvenenpuls	3,95 ± 1,47	4,32 ± 1,67	4,16 ± 1,57	4,53 ± 1,43	0,493
Herzinsuffizienz 2	4,38 ± 1,07	4,61 ± 1,23	4,32 ± 1,06	4,79 ± 1,08	0,276
Körperl. Unters. interpr.	4,15 ± 1,79	4,48 ± 1,04	4,79 ± 1,40	4,63 ± 1,16	0,633
Periphere Pulse	2,95 ± 1,19	3,22 ± 1,28	3,00 ± 1,25	2,74 ± 1,37	0,237
Lungenfkt.-Interpr.	4,71 ± 1,06	4,48 ± 1,31	4,21 ± 1,23	4,68 ± 1,34	0,226
Raucheranamnese	3,38 ± 1,28	4,30 ± 1,52	3,95 ± 1,43	4,05 ± 1,61	0,221
Untersuchung der Lunge	3,76 ± 1,45	4,22 ± 1,09	3,53 ± 1,17	3,89 ± 1,24	0,968

Tabelle 6: Selbsteinschätzungen gemittelt über die Selbstwert-Quartile pro Lernziel; Einschätzung zu Modulende

Teil A: kognitive Lernziele

LERNZIEL	Quartil 1	Quartil 2	Quartil 3	Quartil 4	p-Wert (Kruskal -Wallis- Test)
	MW ± SD	MW ± SD	MW ± SD	MW ± SD	
Akute Perikarditis	2,86 ± 1,65	2,65 ± 1,30	2,47 ± 1,31	2,89 ± 1,84	0,101
Akuter Gefäßverschluss	1,81 ± 0,93	1,82 ± 0,85	1,68 ± 1,00	1,74 ± 0,99	0,839
Akuter Myokardinfarkt	1,14 ± 0,48	1,22 ± 0,52	1,28 ± 0,46	1,11 ± 0,32	0,443
Aortendissektion	2,05 ± 1,53	2,65 ± 1,72	2,53 ± 1,68	1,83 ± 1,47	0,365
Aortenstenose	1,62 ± 0,86	1,91 ± 1,12	1,74 ± 0,87	2,00 ± 1,20	0,068
Atherosklerose	1,86 ± 0,96	1,65 ± 0,71	1,79 ± 0,79	1,74 ± 0,73	0,497
Endokarditis	2,62 ± 1,75	2,61 ± 1,64	2,32 ± 1,92	2,44 ± 1,72	0,130
Fallot-Tetralogie	1,29 ± 0,78	1,22 ± 0,42	1,21 ± 0,54	1,05 ± 0,23	0,273
Herzinsuffizienz 1	1,48 ± 0,81	1,22 ± 0,52	1,26 ± 0,56	1,16 ± 0,50	0,350
Hyperlipidämie	2,29 ± 1,45	2,09 ± 1,04	2,53 ± 1,17	2,16 ± 1,12	0,034
Hypertonie	3,14 ± 1,35	2,96 ± 1,40	2,63 ± 1,38	3,16 ± 1,64	0,505
Kardiomyopathie	2,76 ± 1,37	2,57 ± 1,20	2,84 ± 1,68	3,21 ± 1,55	0,060
Koronare Herzkrankheit	1,10 ± 0,30	1,04 ± 0,21	1,11 ± 0,32	1,26 ± 0,56	0,379
Mitralklappeninsuff.	2,10 ± 0,94	1,87 ± 0,97	2,00 ± 0,82	2,26 ± 1,19	0,118
Myokarditis	3,19 ± 1,29	2,50 ± 1,19	2,89 ± 1,24	2,63 ± 1,42	0,144
Periph. art. Verschl.kr.	4,19 ± 1,83	3,87 ± 1,91	4,21 ± 1,90	4,26 ± 1,76	0,002
PFO	3,52 ± 1,75	2,87 ± 1,71	2,74 ± 1,37	2,95 ± 1,54	0,078
Sarkoidose	5,24 ± 1,22	4,83 ± 1,53	5,11 ± 1,29	5,16 ± 1,26	0,643
Schock	3,62 ± 1,40	3,04 ± 1,33	2,95 ± 1,27	3,37 ± 1,46	0,034
Rheumatisches Fieber	3,29 ± 1,59	2,96 ± 1,30	3,05 ± 1,31	3,56 ± 1,25	0,036
Rhythmusstörungen	2,43 ± 1,29	1,86 ± 0,83	2,11 ± 1,24	2,05 ± 1,31	0,055
Akute Bronchitis	3,29 ± 1,27	3,13 ± 1,32	3,21 ± 1,44	3,89 ± 1,52	0,277
Asthma bronchiale	2,05 ± 0,86	1,91 ± 0,67	2,21 ± 1,03	2,21 ± 1,23	0,125
Chronische Bronchitis	2,24 ± 1,45	2,35 ± 1,19	2,16 ± 1,26	2,11 ± 1,10	0,069
COPD	1,76 ± 0,89	1,91 ± 0,85	1,89 ± 1,10	1,95 ± 1,27	0,036
Emphysem	2,40 ± 1,43	2,43 ± 1,24	2,58 ± 1,22	2,68 ± 1,63	0,001
Influenza	2,19 ± 1,12	2,09 ± 0,87	1,47 ± 0,70	1,89 ± 0,66	0,640
Lungenembolie	3,67 ± 1,46	3,91 ± 1,28	4,32 ± 1,45	4,16 ± 1,42	0,040
Pleuraerguss	3,19 ± 1,25	2,52 ± 1,31	2,63 ± 1,26	3,21 ± 1,47	0,066
Pneumonie	2,10 ± 1,51	2,17 ± 1,23	2,16 ± 1,38	2,37 ± 1,12	0,146
Pneumothorax	1,95 ± 1,32	1,91 ± 0,61	1,67 ± 0,69	1,84 ± 0,76	0,009
Pulmonale Hypertonie	3,52 ± 1,29	3,64 ± 1,00	3,16 ± 1,34	3,26 ± 1,41	0,002
Tabakabhängigkeit	2,10 ± 1,48	1,74 ± 1,18	1,79 ± 1,18	1,84 ± 1,21	0,005
Tuberkulose	5,00 ± 1,18	4,00 ± 1,31	4,68 ± 1,42	4,89 ± 1,33	0,798

Teil B: praktische Lernziele

LERNZIEL	Quartil 1	Quartil 2	Quartil 3	Quartil 4	p-Wert (Kruskal -Wallis- Test)
	MW ± SD	MW ± SD	MW ± SD	MW ± SD	
Anordn. von Laborunters.	3,86 ± 1,28	4,04 ± 1,40	4,37 ± 0,90	4,26 ± 1,48	0,187
Auskultation	1,90 ± 1,22	1,70 ± 0,76	2,11 ± 1,29	1,74 ± 0,87	0,017
Echokardiographie	2,10 ± 0,89	2,00 ± 0,60	2,32 ± 0,75	2,56 ± 0,92	0,070
EKG-Interpretation	2,90 ± 1,09	2,65 ± 0,78	2,53 ± 0,90	2,37 ± 1,07	0,004
Ergometrie	2,67 ± 0,91	2,04 ± 0,98	2,58 ± 1,07	2,32 ± 1,38	0,720
Halsvenenpuls	2,62 ± 1,12	2,65 ± 1,19	2,63 ± 0,90	2,37 ± 0,96	0,372
Herzinsuffizienz 2	2,24 ± 0,77	2,09 ± 0,67	2,05 ± 0,85	2,00 ± 0,59	0,310
Körperl. Unters. Interpr.	3,05 ± 1,47	3,91 ± 1,41	3,32 ± 1,63	3,63 ± 1,46	0,436
Periphere Pulse	2,71 ± 1,35	2,61 ± 1,37	2,21 ± 1,27	2,21 ± 0,98	0,428
Lungenfkt.-Interpr.	1,35 ± 0,49	1,22 ± 0,52	1,32 ± 0,75	1,53 ± 0,84	0,360
Raucheranamnese	2,19 ± 1,25	1,96 ± 0,71	2,37 ± 1,07	1,89 ± 0,74	0,401
Untersuchung der Lunge	3,24 ± 1,37	2,74 ± 1,01	3,00 ± 1,25	2,95 ± 1,22	0,036

Die in Tabelle 5 und 6 gezeigten Selbsteinschätzungen ergaben im Mittel sALZ-Werte, die in den vier Quartilen hohe Korrelationen zeigten (siehe Tabelle 7).

Tabelle 7: Korrelationen der Lernzuwächse (sALZ) der 4 Quartile

	QUARTIL 1	QUARTIL 2	QUARTIL 3	QUARTIL 4
QUARTIL 1	-	0,927**	0,911**	0,924**
QUARTIL 2	0,927**	-	0,905**	0,927**
QUARTIL 3	0,911**	0,905**	-	0,896**
QUARTIL 4	0,924**	0,927**	0,896**	-

Die Korrelationen sind auf dem Niveau $p < 0,01$ signifikant.

3.8 Darstellung der Testleistungen

Im folgenden Kapitel sind die Lernzuwachsrate, errechnet aus den Leistungen im formativen Eingangs- und Abschlusstestat, dargestellt.

Die interne Konsistenz als unteres Grenzmaß der Reliabilität (Cronbach's α) betrug im Eingangstestat 0,890 und im Abschlusstestat 0,871. Die mittlere Item-Schwierigkeit P zeigte sich zu den beiden Prüfungszeitpunkten mit $0,18 \pm 0,18$ bzw. $0,57 \pm 0,26$. Die mittlere Trennschärfe betrug $0,23 \pm 0,14$ bzw. $0,19 \pm 0,13$. Im Eingangstestat wiesen neun, im Abschlusstestat acht Items eine negative Trennschärfe auf.

Von den maximal erreichbaren 165 Punkten erreichten die Studienteilnehmer/innen im Eingangstestat im Mittel $29,0 \pm 12,3$ Punkte (Bereich 9 bis 68 Punkte), im Abschlusstestat $93,2 \pm 14,6$

Punkte (Bereich 59 bis 120 Punkte). In den Lernzielen „Tabakabhängigkeit“ und „Tuberkulose“ haben die Studierenden im Eingangstestat keine Punkte erreicht.

Abbildung 6: Darstellung der Mittelwerte der Testatergebnisse sowie der Standardabweichung pro Lernziel vor und nach dem Modul basierend auf den Testatergebnissen

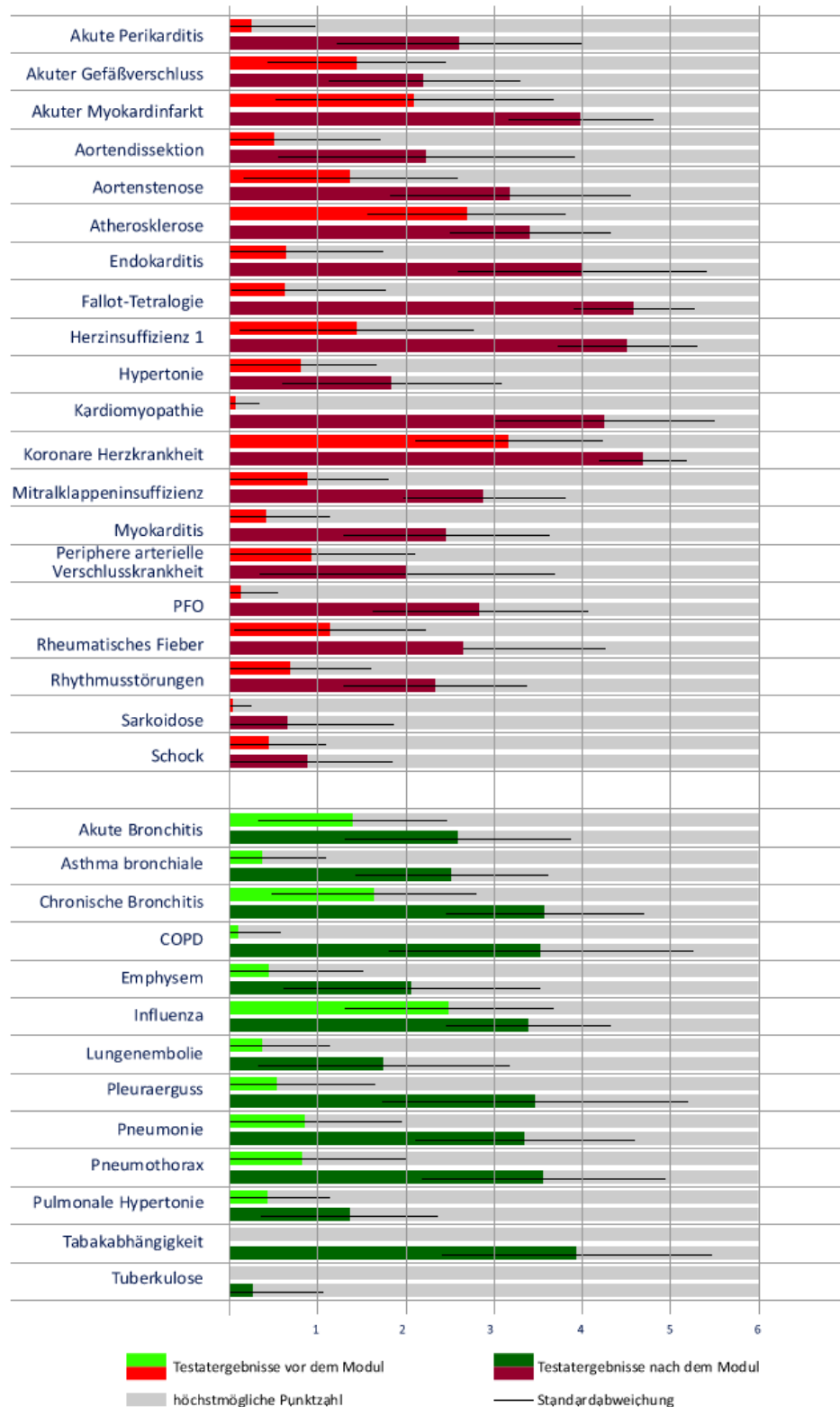
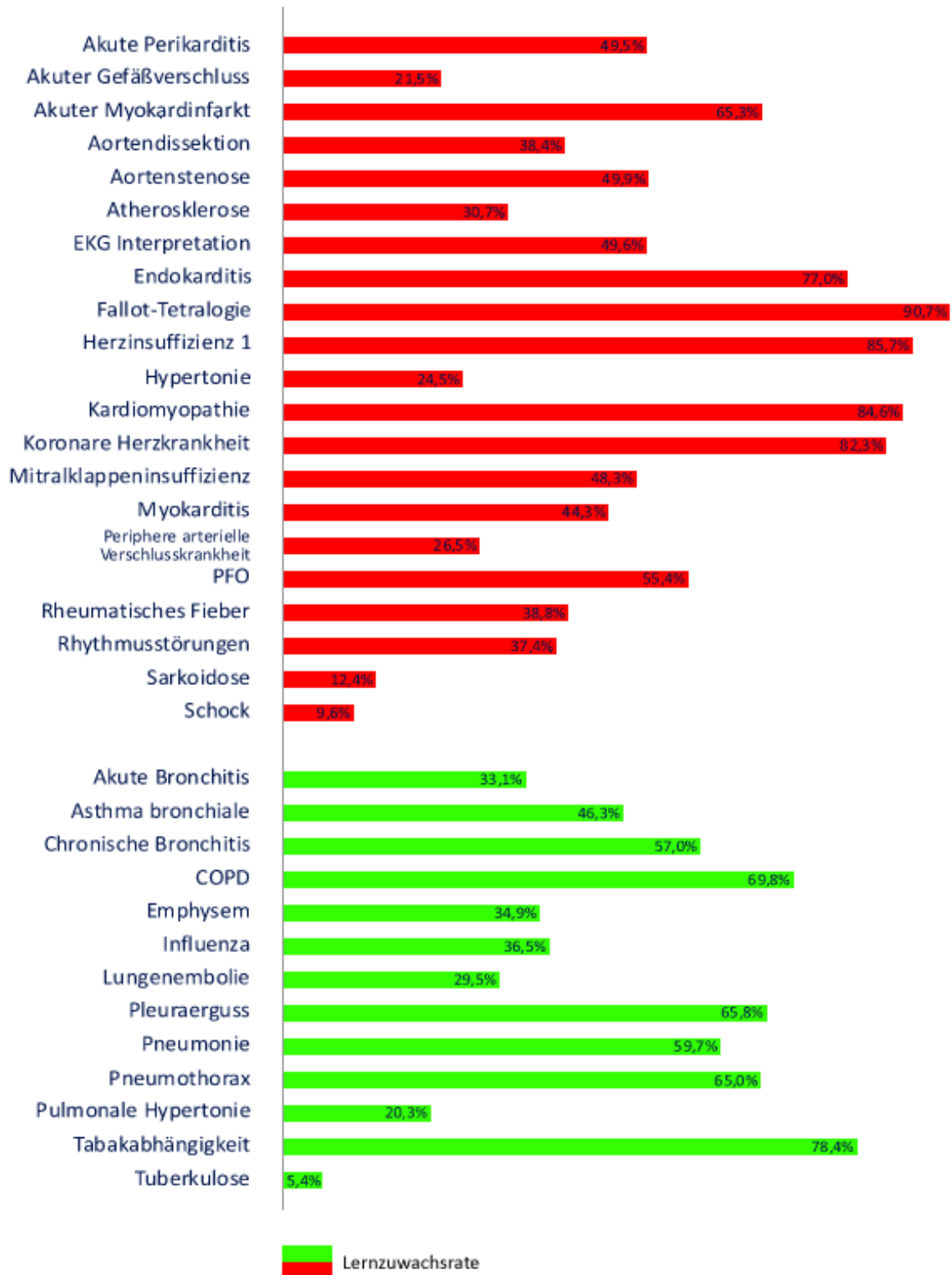


Abbildung 7: Darstellung der Lernzuwachsrate pro Lernziel vor und nach dem Modul basierend auf den Testatsergebnissen (oALZ)



Im Abschlusstestat zeigte sich eine breite Verteilung der Anzahl der richtigen Antworten. Einen Hinweis auf einen Boden- oder Deckeneffekt gab es nicht.

Allein aus der initialen Leistung ließ sich nicht auf den später resultierenden Lernzuwachs schließen. Bei einer hohen initialen Leistung konnte sowohl ein großer, z.B. „Koronare Herzkrankheit“ (82,3%), wie auch ein geringer, z.B. „Atherosklerose“ (30,7%), Lernzuwachs erreicht werden. Gleiches galt für eine eher schwache initiale Leistung (vergleiche „Tabakabhängigkeit“ (78,4%) und „Tuberkulose“ (5,4%)).

Die initiale Testatleistung war in den kardiologischen Themen etwas höher (MW 0,98) als in den pulmologischen Themen (MW 0,72), ebenso die Testatleistung nach dem Modul (mittlere Leistung in kardiologischen Themen 2,90, in pulmologischen Themen 2,71). Der Lernzuwachs in beiden Themengebieten war fast identisch (kardiologisch 49,3%, pulmologisch 50,6%).

3.9 Kriteriumsvalidität (Studienfrage 6)

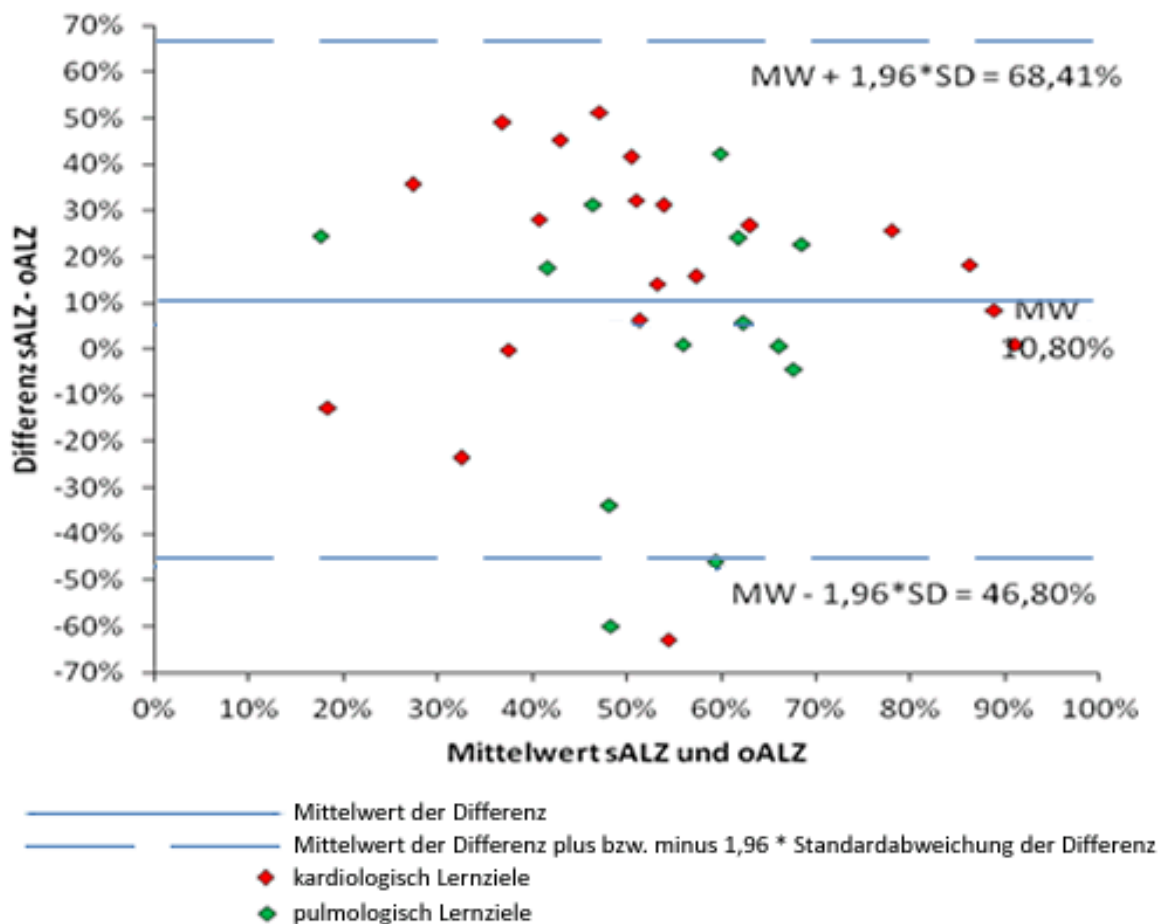
Zur Beantwortung der Studienfrage 6 wurden die Lernzuwachsraten sowohl auf *group-* als auch auf *individual-level* mittels der prospektiven subjektiven Selbsteinschätzungen errechnet. Es erfolgte eine zusätzliche Berechnung der Lernzuwachsraten beruhend auf den Testatleistungen. Die Korrelation der *group-level*-Analyse mit dem objektiven Lernzuwachs wurde in Abbildung 8 mittels eines Bland-Altman-Plots dargestellt. Die Korrelation der *individual-level*-Analyse mit dem objektiven Lernzuwachs ist in Abbildung 9, aufgrund der breiten Verteilung der Korrelationskoeffizienten, in einem Balkendiagramm zu sehen. Außerdem wurde eine ROC-Analyse durchgeführt.

3.9.1 Korrelation des Lernzuwachses errechnet aus Selbsteinschätzungen und Testatergebnissen auf *group-level*

Betrachtet man die 33 Lernziele zusammen, so fand sich zwischen den subjektiven und den objektiven aggregierten Lernzuwachsdaten eine Korrelation von $r = 0,78$ ($p > 0,0001$).

Im Bland-Altman-Plot ist die Vergleichbarkeit der beiden Methoden (sALZ- bzw. oALZ-Berechnung) gezeigt. Die kardiologischen Lernziele sind rot, die pulmologischen grün dargestellt.

Abbildung 8: Bland-Altman-Plot der sALZ- und oALZ-Werte (Korrelation zwischen dem prospektivem Lernzuwachs, errechnet aus Selbsteinschätzungen (sALZ), und dem prospektiven Lernzuwachs, errechnet aus Testatergebnissen (oALZ))



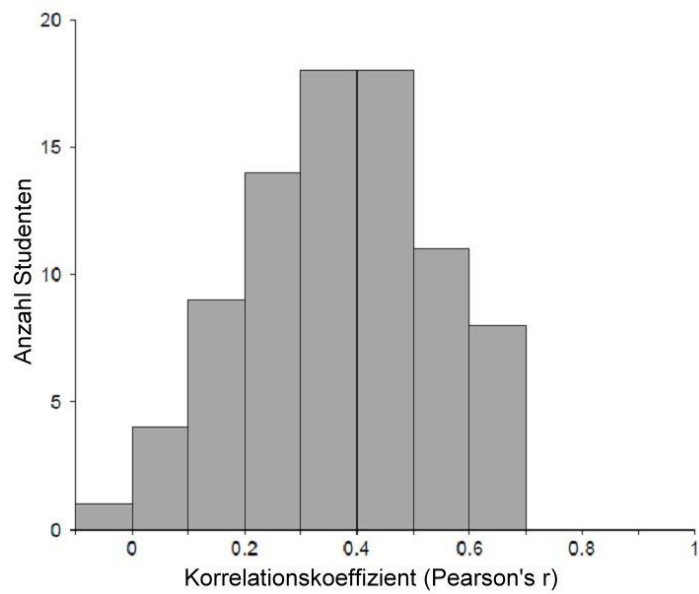
Es zeigt sich in der Differenz der beiden Berechnungsmethoden ein Mittelwert von 10,80%. Die Standardabweichung liegt bei 29,39%.

3.9.2 Korrelation des Lernzuwachses errechnet aus Selbsteinschätzungen und Testatergebnissen auf *individual-level*

Unter Verwendung der sILZ-Werte wurden die mittleren Lernzuwachsraten der Studierenden errechnet und diese mit der objektiven Lernzuwachsrate korreliert.

In Abbildung 9 ist die Korrelation zwischen subjektiv und objektiv eingeschätztem Lernzuwachs nach der *individual-level*-Berechnung in einem Balkendiagramm dargestellt. Es ergaben sich Korrelationskoeffizienten zwischen -0,09 und 0,69 mit einem mittleren $r=0,37$.

Abbildung 9: Darstellung der Korrelation zwischen subjektiv und objektiv eingeschätztem Lernzuwachs berechnet auf individual-level-Niveau

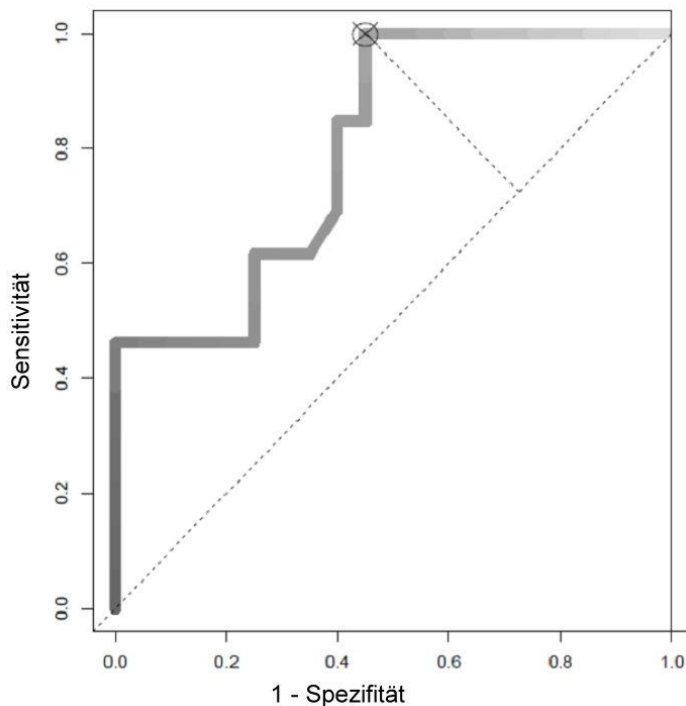


Im Unterschied zur *group-level*-Analyse wurden hier die mittleren Lernzuwachsrate pro Student/in, nicht pro Lernziel, errechnet.

3.9.3 ROC-Analyse, Cut-Off-Wert

Zur Beurteilung der Kriteriumsvalidität ist es notwendig, das Unterscheidungsvermögen des neuen Evaluationsinstruments zwischen erfolgreich und erfolglos gelehrt Themen darzustellen. In Abbildung 10 ist die ROC-Analyse dargestellt. Die Koordinatenachsen zeigen die Sensitivität und 1-Spezifität. Der optimale sALZ-Cut-Off-Wert ergab sich aus der besten Sensitivität bei einer maximalen Spezifität.

Abbildung 10: ROC-Analyse



Die durchgezogenen Linien in Abbildung 10 stellen die verschiedenen Cut-Off-Werte dar.

Als Cut-Off für oALZ-Werte wird ein Lernzuwachs von 50% angenommen. Dies bedeutet, dass Lernziele, die weniger als 50% Lernzuwachs erzielten, als nicht gut gelehrt betrachtet wurden.

In der ROC-Analyse wurden verschiedene Cut-Off-Werte für die sALZ-Daten getestet und deren Diskriminationsfähigkeit mittels Angabe von Sensitivität und Falsch-Positiv-Wert aufgezeigt.

Unter Verwendung des Cut-Offs von 50% für oALZ ergab sich für die sALZ-Werte ein Cut-Off von 54,7% Lernzuwachs.

Wenn dieser Cut-Off genutzt wurde, entsprach die Sensitivität der Evaluation mittels vergleichender studentischer Selbsteinschätzung 100%. Dies bedeutet, dass alle Lernziele mit einem oALZ von $\geq 50\%$ mittels der sALZ-Daten korrekt als „gut“ identifiziert wurden. Die Spezifität lag bei 57%, was bedeutet, dass 57% der Lernziele mit einem oALZ von $< 50\%$ mittels der sALZ-Daten korrekt als „schlecht“ identifiziert wurden.

Der positive prädiktive Wert der sALZ-Werte lag bei 59%, was bedeutet, dass 59% der als gut gelehrt identifizierten Lernziele auch wirklich gut gelehrt wurden.

Der negative prädiktive Wert lag bei 100%. Somit waren alle der nach sALZ $< 54,7\%$ als schlecht eingestuften Lernziele auch in der oALZ-Überprüfung schlecht.

4 Besprechung der Ergebnisse im Zusammenhang mit den Angaben in der Literatur („Diskussion“)

Die Interpretation von Lehrevaluationen umfasst viele verschiedene Aspekte.

Für Studierende wichtige Gedanken zur Evaluation wurden im qualitativen Teil der hier vorgestellten Studie identifiziert.

Um eine Verbesserung der Lehre durch Lehrevaluationen zu erreichen, wäre es wünschenswert, dass beide Gruppen, sowohl Studierende als auch Lehrende, Bestandteile der Evaluation besser kennen und verstehen, um möglichst wenige Missverständnisse bei der Ergebnisinterpretation aufkommen zu lassen. Die vorliegende Arbeit soll einen Beitrag dazu leisten, das Verständnis für dieses Konstrukt zu erleichtern.

4.1 Wesentliche Ergebnisse der hier vorgestellten Studie

Die Studie hat gezeigt, dass der Lernzuwachs in den Fokusgruppengesprächen als ein für die Studierenden wichtiger Parameter der Lehrqualität anzusehen ist (Studienfrage 1).

Die Beantwortung von Studienfrage 2 ergab, dass die *group-level-* gegenüber der *individual-level-*Analyse deutliche Vorteile bietet.

In Bezug auf Studienfrage 3 ergab sich, dass die Korrelation zwischen pro- und retrospektiver Einschätzung des initialen Lernzuwachses $r=0,97$ beträgt.

Berechnungen zu Studienfrage 4 zeigten, dass die Bildung des Mittelwertes aus den Lernzuwächsen für 15 spezifische Lernziele eine hohe Schwankungsbreite zwischen 38% und 73% aufweist. Die Schwankungsbreite lag erst bei Einbeziehung von 39 Lernzielen unter 10%.

Studienfrage 5 konnte dahingehend beantwortet werden, dass ein Unterschied in der Selbsteinschätzungsfähigkeit in Gruppen mit unterschiedlich ausgeprägtem Selbstwertgefühl nicht festgestellt werden konnte.

Zu Studienfrage 6 ergab sich ein $r=0,78$ zwischen sALZ- und oALZ-Werten. Der Cut-Off für sALZ-Werte lag bei 54,7%.

4.2 Wesentliche Limitationen der hier vorgestellten Studie

4.2.1 Limitationen der qualitativen Studie

Limitationen in Fokusgruppengesprächen sind anzunehmen. Es ist möglich, dass ein Selektionsbias einfluss, da die Teilnahme an den Gesprächen freiwillig erfolgte. Die Meinung der Fokusgruppenteilnehmer/innen kann nicht als „Spiegelbild“ der allgemeinen Studierendenmeinung verstanden werden. Eher waren die an der Fokusgruppe teilnehmenden Studierenden besonders engagiert, da sie an einer zusätzlichen Veranstaltung teilnahmen. Dies muss nicht zwangsläufig zu

einer Verzerrung führen. Auch könnten hier Studierende teilgenommen haben, die Evaluation besonders gut finden oder so schlecht, dass sie daran gern etwas ändern würden. Die Geschlechterverteilung kam mit 76% weiblichen Teilnehmern der Rate an weiblichen Medizinstudierenden mit 65% nah. Es nahmen Studierende aus unterschiedlichen Semestern teil.

4.2.2 Limitationen der quantitativen Studie

Die Selbsteinschätzungsfähigkeit ist in der Literatur nicht immer einheitlich definiert, was Studien zu diesem Thema nur bedingt vergleichbar macht. Allerdings wird eine gute Selbsteinschätzungsfähigkeit beim Ausüben selbstregulierter Berufe als notwendig erachtet (Eva and Regehr 2007) und ist somit Thema aktueller Forschung. Davis zum Beispiel stellte in einer Metastudie fest, dass die Fähigkeit von Medizinern, sich selbst einzuschätzen, begrenzt ist (Davis 2006). Die Selbsteinschätzungsfähigkeit wird dagegen von anderen Autoren als stabiles Kriterium betrachtet (Fitzgerald et al. 2003; Raupach et al. 2011).

An der vorliegenden Studie haben sich 58% der möglichen Probanden/innen beteiligt. Diese Teilnehmerquote zeigt im Vergleich zu Online-Evaluationsstudien nur geringe Abweichungen. Johnson z.B. erfasste in einer 2003 durchgeführten Studie Teilnehmerquoten zwischen 40% und 62% (Johnson 2003). Eine Studie der Universität Michigan zum Vergleich Online- und papiergestützte Evaluation ergab, dass 60% der möglichen Probanden an der Online-Evaluation teilnahmen, wenn ohne Sanktionen bzw. Belohnung gearbeitet wurde (University-of-Michigan 2007).

Evtl. ist die eher moderate Teilnahmequote in der hier vorgestellten Studie durch den recht hohen Zeitaufwand zu erklären, da die Studierenden zwei Evaluationsfragebögen ausfüllen, sowie an zwei Testaten teilnehmen mussten. Das zweite Testat und die zweite Selbsteinschätzung wurden zudem in zeitlicher Nähe zur Modulabschlussklausur durchgeführt.

Ein Vergleich der Klausurergebnisse zeigte, dass die Studienteilnehmer/innen im Vergleich mit den Studierenden die zwar das Modul absolviert, jedoch nicht an der Studie teilgenommen haben, bessere Ergebnisse erzielten. Die Studienteilnehmer/innen konnten im Durchschnitt 80,2% der Fragen (SD 7,3%) beantworten, während die Vergleichsgruppe nur ein Ergebnis von 74,7% (SD 10,0%) erzielte. Cohen's d für den Unterschied zwischen den beiden Gruppen lag bei 0,72, was für einen mittleren bis großen Effekt bezüglich der Leistung spricht. Die besseren Klausurleistungen der Studienteilnehmer/innen lassen vermuten, dass diese motivierter waren als die Nicht-Teilnehmer. Diese höhere Motivation kann als potentieller Selektionsbias aufgefasst werden. Es ist auch denkbar, dass die Studierenden die in der Studie verwendeten Lernziele als Orientierung zum Lernen nutzten und somit die wichtigen und in der Klausur abgeprüften Themen besser als die anderen Studierenden erfassen konnten.

Die Berechnung des objektiven Lernzuwachses (oALZ) erfolgte mittels *neutral-marking*. Es wäre auch eine *negative-marking*-Berechnung möglich gewesen.

Die *neutral-marking*-Methode sollte eine Vergleichbarkeit der Testatergebnisse mit den Selbsteinschätzungen gewährleisten, da sich herausstellte, dass die Beantwortung der Fragen zunächst schwieriger schien als bei regulär verwendeten Klausurfragen. Die Abschlussklausur wies mit 62% aller Fragen im Vergleich mit dem formativen Testat (27%) eine höhere Rate an Fragen mit hoher Itemschwierigkeit (> 0,8) auf. Die Itemschwierigkeit ist durch einen Index dargestellt, der den

Anteil der Personen, die diese Aufgabe richtig lösen können, beinhaltet. Je höher die Itemschwierigkeit, desto weniger Teilnehmer/innen konnten die richtige Antwort geben. Aber auch die Anzahl der Fragen mit einem geringen Unterscheidungsindex war in der summativen Abschlussprüfung höher (62% in summativer versus 48% in formativer Prüfung). Cronbach's α ergab in der Abschlussklausur nur 0,69, während das formative Testat ein Cronbach's α von 0,87 aufwies. Die schlechteren Ergebnisse sind eher auf den geringeren Leistungsdruck zurückzuführen, der sich in einer formativen im Vergleich zu einer summativen Prüfung ergibt (Raupach et al. 2010).

In der Berechnung des sALZ ergaben sich auch negative Werte. Dies könnte zwei Gründe haben. Einerseits ist es möglich, dass Studierende im Verlauf eines Moduls Lehrinhalte vergessen, die sie früher beherrschten. Andererseits ist ein *response shift* denkbar, da Studierende im Laufe eines Moduls lernen, Fragen zu Lehrinhalten besser einzuschätzen und somit ihr Wissen nach dem Modul adäquater einschätzen können als zuvor. Möglicherweise lernten die Studierenden erst im Modul, welches umfangreiches Wissen zur Beantwortung einer Frage nötig sein würde, wobei diese Frage vor dem Modul ggf. als eher trivial eingeschätzt wurde. Dagegen sprechen die hohen Korrelationen zwischen retrospektiv und prospektiv erhobenen initialen Selbsteinschätzungen, da ein *response shift bias* hier zu einem Unterschied geführt hätte.

4.3 Paarweiser Vergleich versus Kalkulation des Lernzuwachses aus den Mittelwerten (Studienfrage 2)

Die Betrachtung der Lernzuwachsrate ergab sowohl in der prospektiven sALZ- als auch der sILZ-Berechnung (Lernzuwachs berechnet aus prospektiven Selbsteinschätzungen und Einschätzungen nach dem Modul) vergleichbare Ergebnisse. Dies deutete zunächst darauf hin, dass beide Berechnungsarten sinnvoll sein könnten. Wurden jedoch die subjektiven mit den objektiven Lernzuwachsrate verglichen, so ergab sich eine gute Korrelation nur bei der Verwendung der aggregierten Daten. Die Konfidenzintervalle in der sILZ-Berechnung zeigten eine breite Streuung. Ebenso die Korrelationen der Lernzuwächse auf *individual-level*-Niveau, welche bei jedem/ jeder einzelnen Studierenden sehr unterschiedlich waren. Der Aussage von Lam folgend, der feststellte, dass gemittelte Daten die gleich viele Über- wie Unterschätzungen beinhalten, nicht valide sind (Lam 2009), ist eine Verwendung der *individual-level*-Daten hier als nicht sinnvoll zu betrachten.

Somit kann die zeit- und ressourcensparende Berechnung aus Mittelwerten gewählt werden, die keine individuelle Kennzeichnung notwendig macht. Die erfolgreiche Verwendung aggregierter Daten wurde von D'Eon 2008 in einer Studie zur Evaluation von Bildungsmaßnahmen bereits gezeigt (D'Eon et al. 2008).

Trotzdem ist es sinnvoll durch paarweise Zuordnung sicherzustellen, dass an der Prä- und der Post-Evaluation die gleichen Studierenden beteiligt sind, um Verzerrungen des Ergebnisses zu vermeiden. Hierfür wäre zum Beispiel die Einführung eines Codes denkbar, der zwar die Zuordnung der Evaluationen zueinander, aber keine Zuordnung zu bestimmten Teilnehmern/innen ermöglicht. Ein solcher Code wurde auch in vorliegender Studie verwendet. Eine andere, allerdings eher unsichere Möglichkeit besteht darin, die Teilnehmer/innen zu befragen, ob sie an der Prä-Evaluation bereits

teilgenommen haben. Letztlich wäre auch eine retrospektive Erhebung des Wissensstandes denkbar (siehe Kapitel 4.4).

4.4 Pro-/Retrospektive Einschätzung (Studienfrage 3)

Der Literatur zur vergleichenden Selbsteinschätzung zufolge wäre ein Unterschied in der pro- und retrospektiven Selbsteinschätzung zu erwarten gewesen.

Eine Studie von Skeff erfasste zum Beispiel die mögliche Veränderung der Selbsteinschätzungen durch *response shift*. Es wurde der Einfluss eines Trainingsprogramms auf Selbsteinschätzungen untersucht. Klinische Dozenten sind gebeten worden, ihre Fähigkeiten vor einem Training einzuschätzen, das Training zu absolvieren und nach dem Training noch einmal anzugeben, wie sie ihre eigenen Fähigkeiten vor dem Training einschätzten, sodass für jeden Dozenten zwei Datensätze des selbst eingeschätzten Leistungsstandes vor dem Training vorlagen. Zusätzlich hat eine Einschätzung der eigenen Fähigkeiten nach dem Seminar stattgefunden. Außerdem wurde eine Einschätzung vor und nach jedem Training durch externe Mitarbeiter und Studierende durchgeführt. Es ergaben sich also pro Teilnehmer drei Datensätze in der Selbsteinschätzung und zwei in der externen Bewertung. Die subjektiven (Dozenten-) und objektiven (Mitarbeiter-/Studierenden-) Einschätzungen wurden nach dem Training auf *individual-level* miteinander verglichen. Die Einschätzungen des Leistungsstandes wichen hierbei voneinander ab. Außerdem wurde ein Vergleich der pro- und retrospektiv erhobenen Daten angestrebt. Die Korrelation der objektiven Einschätzungen vor dem Training mit der retrospektiven Selbsteinschätzung ist besser als mit der prospektiven Selbsteinschätzung. In der Bewertung der retrospektiven Einschätzung zeigten sich positive Zusammenhänge mit einem $p < 0,001$ mit den objektiven Daten, während die Zusammenhänge bei den traditionell erhobenen Daten (prospektiv) negativ bis positiv mit einem $p < 0,01$ waren (Skeff et al. 1992). In dieser von Skeff beschriebenen Studie scheint die retrospektive Einschätzung eher reliabel zu sein und mit einer externen Einschätzung übereinzustimmen als die prospektive. Allerdings könnte dies auch auf ein Bias wegen „sozialer Erwünschtheit“ zurückzuführen sein, indem die Probanden sich retrospektiv niedriger einschätzen, um den Studienleiter zufriedenzustellen (Skeff et al. 1992).

In der vorliegenden Studie wurde in den Vergleich von pro- und retrospektiver Einschätzung kein zusätzlicher externer Bewerter miteinbezogen, sondern die Leistungsfähigkeit der Studierenden selbst mittels eines objektiven Testats als externes Kriterium genutzt.

Ein signifikanter Unterschied in der pro- und retrospektiven Einschätzung konnte in der vorliegenden Studie nicht bestätigt werden. Dieses Ergebnis deckt sich mit der in Kapitel 1.5.1 beschriebenen Studie von D`Eon 2008. Dies könnte, in Anlehnung an Skeff, darauf zurückzuführen sein, dass sich der Bezugsstandard der Teilnehmer während des Kurses nicht ändert. Sie lernen zwar inhaltlich dazu, verfügen aber über eine Vorbildung, die es ihnen hier ermöglicht, ihr Wissen realistisch einzuschätzen. Dies ist auch in unserer Studie denkbar. Im Unterschied wurden in der Studie von D`Eon ausschließlich aggregierte Gruppendaten verwendet. Zudem liegt D`Eons Schwerpunkt auf punktuellen Selbsteinschätzungen, während in vorliegender Studie die vergleichende Selbsteinschätzung und die daraus resultierenden Lernzuwächse zur Evaluation herangezogen wurden.

Zudem wäre der geringe Unterschied zwischen pro- und retrospektiver Datenerhebung mit der Annahme vereinbar, dass die Selbsteinschätzungsfähigkeit ein stabiles Kriterium darstellt.

Die eingesetzten Lernziele und entsprechende Testatfragen waren in vorliegender Studie sehr spezifisch formuliert und operationalisiert, was zu einer deutlich besseren Vergleichbarkeit der Ergebnisse vorher und nachher führte. Die Operationalisierung beinhaltet eine spezifische Formulierung des Lernziels, eine klare Messbarkeit und eine „Aktionsorientierung“, die besagt, dass der/ die Studierende eine Erreichung des Lernziels weitgehend selbst beeinflussen kann. Zudem muss das Lernziel realistisch sein und ein Zeitraum, in dem es erreicht werden soll, festgelegt werden. All diese Forderungen können die betrachteten Lernziele erfüllen. Der Aspekt der spezifischen Zielformulierung in der Praxis der Selbsteinschätzungen ist bereits von Colthart in einer Metaanalyse erwähnt worden. Hier konnten Hinweise darauf gefunden werden, dass explizite Bewertungskriterien die Selbsteinschätzungen verbessern können (Colthart et al. 2008).

Welcher der aufgeführten Gründe für die hohe Korrelation der pro- und retrospektiven Erhebung der initialen Selbsteinschätzung letztlich ausschlaggebend ist, kann hier nicht erschlossen werden.

Die Konsequenz ist ein größerer Spielraum in der Durchführung der Evaluation, indem sowohl die pro- als auch die retrospektive Datenerhebung verwendet werden können. Der Vorteil einer retrospektiven Erhebung besteht darin, dass keine Differenz der Anzahl von Prä- und Post-erhebungen besteht. Nachteilig könnte sich die hohe Fragenanzahl auf die Motivation der Studierenden auswirken, sodass ggf. die Teilnahmequote sinkt. In den Fokusgruppengesprächen äußerten sich die Studierenden dahingehend, dass meist eine retrospektive Einschätzung leichter ist als eine prospektive. Andererseits wurde eine zu große Fragenanzahl als schlecht eingeschätzt.

Um zu wissen welche in der praktischen Umsetzung die bessere Alternative ist, wäre eine weitere Studie wünschenswert, in der beide Möglichkeiten getestet und verglichen werden.

Zu beachten ist, in welchem Format die Evaluation durchgeführt wird. Die Bearbeitung von zwei kompletten Bögen nacheinander (erst die Prä-Einschätzung aller Lernziele, dann die Post-Einschätzung aller Lernziele) wird ggf. andere Ergebnisse liefern, als ein direkter Vergleich der einzelnen Lernziele vor und nach dem Modul, wenn beide Einschätzungen pro Lernziel gleich nacheinander gegeben werden sollen. Die direkte Vergleichbarkeit der Leistungsstände vor (retrospektiv) und nach (aktuell) der Lehrexposition könnte hier eine Veränderung in der Selbsteinschätzung zur Folge haben. In der vorliegenden Studie wurde zunächst die retrospektive Einschätzung aller Lernziele angegeben, im Anschluss die Post-Einschätzung.

4.5 Anzahl der Lernziele (Studienfrage 4)

Die Bewertung eines Moduls kann bei Verwendung des neuen Evaluationsinstruments aus den Selbsteinschätzungen der Studierenden (sALZ) erfolgen. Die Studie erfasste nun alle in einem Modul unterrichteten Lernziele, um die Anzahl der zu evaluierenden Lernziele zu optimieren.

Die Anzahl der Lernziele muss hier als Kompromiss angesehen werden. Einerseits ist zu bedenken, dass, je mehr Fragen gestellt werden, eventuell umso weniger Studierende freiwillig teilnehmen

werden, da der Arbeitsaufwand umso größer wird. Andererseits sind die Ergebnisse genauer, je mehr Lernziele erfasst werden können.

Da die Breite des 95%-Konfidenzintervalls erst ab 39 gestellten Fragen unter 10% liegt, ergibt es sich, in einer Evaluation die Erfassung von mindestens 39 Lernzielen anzustreben. Dies entspricht 85% der in der Studie untersuchten Lernziele.

Nach Minimierung der Schwankungsbreite durch die Erfassung von mehr Lernzielen, könnte unter anderem der Lernzuwachs für die Verteilung von LOM-Geldern herangezogen werden (siehe Kapitel 4.8, Seite 61).

Ein Vergleich mit anderen Untersuchern des neuen Evaluationsinstruments ist in dieser Fragestellung leider nicht möglich, da die Studie in dieser Form bisher nicht durchgeführt und hier pilotiert wurde.

Um herauszufinden, inwiefern die Einführung von mehr Evaluationsfragen die Beteiligung der Studierenden wirklich beeinflusst, wäre es denkbar, in den nächsten Evaluationen die Anzahl der Fragen um z.B. jeweils fünf zu erweitern und somit zu erfassen, ob und in welchem Umfang die Anzahl der Teilnehmer/innen im Vergleich mit den Vorjahren abnimmt. Wenn sich ein prozentualer Abfall der Teilnehmerzahl ergeben sollte, der außerhalb der regulären Schwankungsbreite zwischen den Semestern liegt, so müsste über andere Möglichkeiten, die Studierenden zur Teilnahme zu motivieren, nachgedacht werden.

4.6 Selbstwertgefühl (Studienfrage 5)

Das Selbstwertgefühl wurde in der vorliegenden Studie mittels der Rosenberg-Skala erhoben. Eine Vergleichbarkeit ergibt sich aus der Betrachtung anderer Studien zum Thema Selbstwertgefühl.

Einen Vergleich stellt eine Studie von Sinclair (Sinclair et al. 2010) dar. Eingesetzt wurde die englische Version der Rosenberg-Skala. Die Studie umfasste 503 Personen mit verschiedenen Bildungsabschlüssen.

Tabelle 8: Mittelwerte des Selbstwertgefühls

Altersgruppe/ Bildungsabschluss/ Geschlecht	Mittelwert Selbstwert	Standardfehler
Alter 18-25 Jahre	19,67	6,63
Alter 26-35	22,28	5,66
4 Jahre College	24,24	5,53
Männer	22,43	6,21
Frauen	22,79	5,41

Die Ergebnisse der Fragen der Rosenberg-Skala zeigen in vorliegender Studie mit einem Mittelwert von 24,73 ein eher hohes Selbstwertgefühl in der Kohorte auf, das aber mit den Werten der Collegeabsolventen von Sinclair vergleichbar ist.

Collani und Herzberg (Collani and Herzberg 2003) ermittelten das Selbstwertgefühl von Psychologiestudenten. Hier ergaben sich etwas niedrigere Werte als in vorliegender Studie. Der Mittelwert betrug 22,67 +/- 4,81.

Einen möglichen Einfluss des Selbstwertgefühls auf die Selbsteinschätzungsfähigkeit konnten Kruger und Dunning in Studien ermitteln. Es zeigte sich hier, dass die Teilnehmer, deren Leistung in intellektuellen Bereichen im untersten Quartil in Bezug auf die Studiengruppe lag, sich am meisten überschätzen. Je höher die Leistung in den Studientests ausfiel, desto besser war die Selbsteinschätzungsfähigkeit. Kruger und Dunning führten dies unter anderem darauf zurück, dass die besseren Teilnehmer über ausgereifere metakognitive Fähigkeiten verfügten (Kruger and Dunning 1999).

Die Ermittlung des Selbstwertgefühls in Bezug auf die Selbsteinschätzungen in der vorliegenden Studie ergab dagegen kein bestimmtes Muster. Es ließ sich nicht erkennen, dass das Selbstwertgefühl einen Einfluss auf die Selbsteinschätzungsfähigkeit hätte. Es war, im Unterschied zu Kruger und Dunning, keine systematische Über- oder Unterschätzung zu erkennen.

Somit konnte die Selbsteinschätzungsfähigkeit der Studierenden bezüglich ihres Wissenstandes als hinreichend stabil und unabhängig vom Selbstwertgefühl gewertet werden. Eine Erhebung des Selbstwertgefühls bei einer Evaluation mittels vergleichender Selbsteinschätzung ist dementsprechend in Zukunft nicht notwendig.

4.7 Kriteriumsvalidität (Studienfrage 6)

Im qualitativen Teil der hier vorgestellten Studie wurde als wichtige Zielgröße der Evaluation der Lernzuwachs identifiziert. Der quantitative Teil der Studie untersuchte ein Evaluationsinstrument, das genau diesen Aspekt messen kann. Die Studie ließ Hinweise auf eine gute Konstruktvalidität für das vorgestellte Messinstrument der vergleichenden Selbsteinschätzung erkennen. Z.B. scheint das Instrument gegen mögliche Einflussfaktoren wie das Selbstwertgefühl robust zu sein. Auch ein *response shift bias* ist nicht anzunehmen. Andere Biasfaktoren wie das Interesse am unterrichteten Fach (Marsh and Roche 2000) und die Erwartungen der Studierenden zu Beginn des Moduls (Raupach et al. 2012) sollten dagegen weiterhin mit erfasst werden. Außerdem ergab sich, dass die objektive Messung des Wissenszuwachses (oALZ) signifikant mit der subjektiven Erfassung (sALZ) korrelierte.

Dies entspricht einer Studie von Frank, in der Studierende auf einem Fragebogen angeben sollten, wie oft sie ihren Patienten gegenüber bestimmte Themen, z.B. rauchen, ansprechen und dieses Verhalten mit SPs (standardisierte Patienten) praktisch geprüft wurde. Die Auswertung zeigte, dass die Studenten, die angaben ihre Patienten häufiger auf diese Themen anzusprechen, dies auch in der Praxis taten (Frank 2005). Die Selbsteinschätzung der Studierenden scheint hier mit dem Ergebnis der objektiven Überprüfung zu korrelieren.

In der ROC-Analyse sind Hinweise für eine praktische Umsetzung des Evaluationsinstruments im Alltag beschrieben. Es ergab sich ein Cut-Off-Wert von 54,7% für die sALZ-Daten. Unter Einbeziehung dieses Cut-Offs werden bei einer Spezifität von 57% über die Hälfte der schlecht gelehrten Lernziele als solche identifiziert. Ist ein Lernziel in der vergleichenden studentischen Selbsteinschätzung als

schlecht gelehrt erkannt worden, so entspricht dies bei einem negativ prädiktiven Wert von 100% immer dem objektiven Ergebnis. Der positive prädiktive Wert von 59% zeigt dagegen, dass ein großer Teil der als gut gelehrt identifizierten Lernziele auch in der objektiven Überprüfung gute Ergebnisse erzielte.

Bei einer Sensitivität von 100% ist sicher, dass die Lernziele die in der objektiven Kontrolle als gut erkannt werden, auch in der vergleichenden studentischen Selbsteinschätzung gute Ergebnisse erzielen. Auf dieser Grundlage ist ein Ansatz geschaffen, um die Lehre zu verbessern, da die als schlecht identifizierten Lernziele mit sehr hoher Wahrscheinlichkeit von einem großen Teil der Studierenden tatsächlich nicht erreicht wurden. Dies eröffnet Möglichkeiten zur Verbesserung der Lehre.

Dass Schwächen in der Lehre durch die vergleichende Selbsteinschätzung abgebildet werden, zeigt der Vergleich der kardiologischen und pulmonologischen Themen, deren bisher unterschiedlicher Stellenwert in der Lehre sich in den subjektiven Lernzuwachs-Daten widerspiegelt (siehe Kapitel 3.3, Seite 31).

Die Berechnung auf *individual-level* zeigte im Vergleich mit dem objektiven Lernzuwachs eine sehr hohe Schwankungsbreite, die auf wenig stabile Daten schließen ließ. Aggregierte Daten erbrachten dagegen stabile Ergebnisse.

Die vorliegende Studie unterstützt die Annahme, dass das Konzept der vergleichenden Selbsteinschätzung als Evaluationsinstrument bei Verwendung der aggregierten Daten eine gute Validität besitzt.

Die vergleichende Selbsteinschätzung bietet hier eine kostensparende Möglichkeit der Lehrevaluation. Es werden keine kostenintensiven Tests benötigt. Zudem wird der anfängliche Leistungsstand miteinbezogen und lernzielbezogen statt nur global evaluiert. Dies ermöglicht es, spezifische Verbesserungen vorzunehmen.

4.8 Ausblick auf künftige Forschungsprojekte

Die vorliegende Studie konnte Fragen in Bezug auf das neue Evaluationsinstrument klären, warf aber zugleich weitere Fragen als Gegenstand künftiger Forschungsprojekte auf.

Zusätzlich zu den Fokusgruppengesprächen mit Studierenden, die einen Einblick in deren Meinung über die Evaluation bieten, könnten auch die Lehrenden zu ihren Einschätzungen befragt werden. Die dann sowohl bei Studierenden als auch bei Lehrenden erhobenen Daten würden einen direkten Vergleich zwischen den Erwartungen an eine Evaluation beider Gruppen ermöglichen. Das Evaluationsinstrument könnte noch besser an die Wünsche und Bedürfnisse beider Gruppen angepasst werden.

Eine Steigerung der Teilnehmerzahl bei Lehrevaluationen ist wünschenswert. Diese sollte jedoch nicht durch Zwang erfolgen, da die Ergebnisse nur als aussagekräftig zu verstehen sind, wenn die Studierenden die Fragebögen ernsthaft ausfüllen. Wird die Zahl der auszuwertenden Bögen nur um solche ergänzt, auf denen ohne nachzudenken angekreuzt wurde, würde dies das Ergebnis sicher verzerren. Besser sollte die Motivation der Studierenden gesteigert werden, freiwillig an Lehrevaluationen teilzunehmen. In den Fokusgruppengesprächen wurde von Studierenden

angemerkt, dass sie es wünschenswert fänden, wenn die aus den Evaluationen resultierenden Konsequenzen durch die Fakultät sichtbar gemacht würden. Die Studierenden möchten erkennen, welche Auswirkungen die Evaluation hat und ob Kritikpunkte wirklich verbessert worden sind. Eine Darstellung der durch die Evaluation bewirkten Veränderungen könnte die Motivation für weitere Evaluationen steigern.

Wie viel Prozent der möglichen Teilnehmer/innen die Fragebögen ausfüllen müssen, um ein reliables Ergebnis zu erhalten, ist unklar. Ein weiteres Problem stellt die Notwendigkeit dar, dass zur vergleichenden Selbsteinschätzung Daten vor und nach dem Modul vorliegen müssen.

Da die beschriebene Evaluationsmethode einen neuen Qualitätsaspekt erfasst, nämlich den der Ergebnisqualität, kann über eine Integration in z.B. die Verteilung der LOM-Gelder nachgedacht werden, wie es von den Studierenden im Fokusgruppengespräch vorgeschlagen wurde.

Die Betrachtung des Lernerfolgs für einzelne Lernziele würde bei dieser Verteilung einen Kompromiss darstellen, da die Erreichung vieler Ziele erst durch das Zusammenspiel verschiedener Lehrveranstaltungen möglich ist. Denkbar wäre die Auflage, dass ein bestimmter Anteil der Lernziele in einem Modul erfolgreich gelehrt werden muss, um die Förderung mit LOM-Geldern in Erwägung zu ziehen. Die Bestimmung des Lernzuwachses auf der Grundlage der vergleichenden Selbsteinschätzung bietet den Vorteil, dass auch die Ergebnisqualität in die Verteilung einfließen kann.

Die Ergebnisse der Evaluation werden dazu beitragen, gute Lehre anhand des Lernzuwachses erkennen zu können, aber auch Defizite in der Lehre aufzudecken. Die Gründe für diese Defizite zu suchen, diese zu beheben und somit die Lehre noch weiter zu optimieren, werden hoffentlich die Konsequenzen einer erfolgreich durchgeführten Lehrevaluation sein.

Inwiefern das neue Evaluationsinstrument auf andere medizinische und nicht-medizinische Curricula übertragbar ist, müsste Gegenstand weiterer Forschungsprojekte sein.

5 Zusammenfassung

Ziel der Lehrevaluation ist eine Einschätzung der Lehrqualität, die sich besonders im Studium der Humanmedizin in der Ergebnisqualität widerspiegelt. Ein Instrument zur Messung dieser Ergebnisqualität ist in vorliegender Studie zur „Störgrößenanalyse der vergleichenden studentischen Selbsteinschätzung zur Evaluation von Lehrveranstaltungen im Studium der Humanmedizin“ untersucht worden. Die vergleichende studentische Selbsteinschätzung wird hier als Evaluationsinstrument aufgefasst, das in Zukunft noch an Bedeutung für die Lehrevaluation gewinnen wird.

Die Studie untersuchte folgende Fragestellungen, um zu einer besseren Einschätzung der vergleichenden Selbsteinschätzung beizutragen: Welchen Stellenwert hat die Evaluation für die Studierenden? Liefert ein paarweiser Vergleich studentischer Selbsteinschätzungen andere Ergebnisse als die Kalkulation des Lernzuwachses aus den Mittelwerten einer gesamten Studentenkohorte? Liefert eine prospektive Datenerhebung bei der Lernzuwachevaluation andere Ergebnisse als eine retrospektive Erhebung des initialen Leistungsstands? Wie viele Lernziele pro Modul müssen abgefragt werden, um reliable Daten zu erhalten? Wird das Ergebnis der vergleichenden Selbsteinschätzung durch psychometrische Größen wie das Selbstwertgefühl beeinflusst? Besitzt das Evaluationsverfahren eine akzeptable Kriteriumsvalidität? Lässt sich für die Lernzuwachs-Evaluation ein Cut-Off für die Unterscheidung erfolgreich und erfolglos gelehrter Inhalte definieren?

Die Beantwortung dieser Fragen beruht auf einer 2011 an der Universitätsmedizin Göttingen durchgeführten Studie, an der Studierende des 3. und 4. Studienjahres teilnahmen. Die Studie bestand aus zwei Anteilen – den Fokusgruppengesprächen als qualitativem und der in einem Modul durchgeführten Befragungen der Studenten als quantitativem Anteil. Die Befragungen beinhalteten sowohl die Selbsteinschätzungen als auch formative Testate vor und nach Absolvierung der Lehrveranstaltungen.

Der Stellenwert der Evaluation wurde von den Studierenden, die an den Fokusgruppengesprächen teilgenommen haben, als sehr hoch eingestuft. Zudem wurde der Lernzuwachs als wichtiges Kriterium in der Lehrevaluation identifiziert.

Die *group-level*-Analyse bot gegenüber der *individual-level*-Analyse deutliche Vorteile. Die breite Streuung der Korrelationskoeffizienten zwischen subjektiven und objektiven individuellen Lernzuwachsrechnungen (sILZ vs. oILZ) wies auf die Nachteile der individuellen Berechnung hin. Der Vergleich der prospektiven mit der retrospektiven Erhebung des Leistungsstandes zu Modulbeginn zeigte keine relevanten Unterschiede. Bei Einschluss von 39 Lernzielen lag die Schwankungsbreite des 95%-Konfidenzintervalls bei unter 10%. Ein Einfluss des Selbstwertgefühls auf die vergleichende Selbsteinschätzung konnte nicht bestätigt werden. Die Studie gab Hinweise auf eine akzeptable Kriteriumsvalidität des Evaluationsinstruments. Der Cut-Off-Wert für die Unterscheidung von erfolglos und erfolgreich gelehrten Inhalten konnte für die aggregierten Selbsteinschätzungen (sALZ) bei 54,7% ermittelt werden.

Aus diesen Ergebnissen lässt sich schlussfolgern, dass die vergleichende Selbsteinschätzung als Evaluationsinstrument zukünftig an Bedeutung gewinnen sollte.

Die Verwendung der *group-level-* ist der *individual-level-*Analyse vorzuziehen. Die Erhebung der initialen Selbsteinschätzung kann laut vorliegender Studie pro- oder retrospektiv durchgeführt werden, ohne deutlichen Vorteil der einen gegenüber der anderen Methode. In Bezug auf die Reliabilität ist eine Erfassung von mindestens 39 Lernzielen pro Evaluation anzustreben. Eine Erhebung des Selbstwertgefühls ist nicht notwendig. Bei einem subjektiven nach *group-level-*Niveau berechneten Lernzuwachs (sALZ) von mindestens 54,7% kann das Lernziel als gut gelehrt betrachtet werden.

Die Ergebnisse der vorliegenden Studie bestätigen die vergleichende Selbsteinschätzung als gutes Evaluationsinstrument im Studium der Humanmedizin.

6 Anhang

6.1 Evaluationsbogen der UMG aus dem Wintersemester 06/07

GB Informationstechnologie	Modul 3.1	Seite 1/3
Modul 3.1 WS06/07 Humanmedizin - Modullehre Erkrankungen des Herz-Kreislauf-Systems und der Lunge		

Markieren Sie so: Bitte beurteilen Sie bei jeder Frage, inwieweit die Aussage aus Ihrer Sicht zutrifft. Dieser Fragebogen wird korrektur: maschinell erfasst. Bitte beachten Sie die links gegebenen Hinweise beim Ausfüllen.

Fragen zum Gesamtmodul (Bitte beantworten Sie alle Fragen)	trifft voll zu			trifft nicht zu		
	1	2	3	4	5	6
1) Die Ziele und Anforderungen des gesamten Moduls wurden klar formuliert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Die Zusammenstellung der Lehrinhalte war über das gesamte Modul thematisch ausgewogen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Die Darbietung und Bearbeitung der Lehrinhalte waren über das gesamte Modul gut und sinnvoll strukturiert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Bezogen auf meine inhaltlichen theoretischen Kenntnisse schätze ich meinen Lernzuwachs in diesem Modul als hoch ein	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Bezogen auf meine praktischen Fertigkeiten schätze ich meinen Lernzuwachs in diesem Modul als hoch ein	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Während des Moduls herrschte eine angenehme und anregende Lehr- und Lernatmosphäre	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7) Art und Umfang der Modulprüfung(en) und die dargebotenen und bearbeiteten Lehrinhalte waren gut aufeinander abgestimmt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8) Mit der organisatorischen Durchführung des Gesamtmoduls war ich zufrieden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9) Es gab genügend Beratungsangebote für mich (Sprechstunden, E-mailkontaktmöglichkeiten etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Folgende Fragen beziehen sich auf die einzelnen Unterrichtsformen des Moduls

Bitte beantworten Sie nur Fragen zu den Unterrichtsformen, die Sie auch tatsächlich in diesem Modul besucht haben.

Vorlesung	trifft voll zu			trifft nicht zu		
	1	2	3	4	5	6
1) Die inhaltlichen Ziele und der Aufbau der Vorlesung waren klar zu erkennen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Die dargebotenen und bearbeiteten Inhalte in der Lehrveranstaltung waren gut auf die Ziele und Anforderungen abgestimmt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Die Vorlesung war motivierend, anregend und lebendig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) In dieser Vorlesung habe ich viel gelernt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Nach- und Vorbereitung des Stoffes wurden durch angebotene Unterlagen, wie Skripte, Literatur oder Software, erleichtert	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Mit der organisatorischen Durchführung der Vorlesung war ich zufrieden	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Praktikum	trifft voll zu			trifft nicht zu		
	1	2	3	4	5	6
1) Die inhaltlichen Ziele der Lehrveranstaltung waren klar zu erkennen	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Die Lehrveranstaltung war motivierend, anregend und lebendig	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) In dieser Lehrveranstaltung habe ich viel gelernt	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Nach- und Vorbereitung des Stoffes wurden durch angebotene Unterlagen, wie Skripte, Literatur oder Software, erleichtert	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Mit der organisatorischen Durchführung der Lehrveranstaltung war ich zufrieden	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Ein Praktikum erscheint mir angemessen zur Vermittlung der dargebotenen und bearbeiteten Lehrinhalte	Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Gerätekunde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

m3102

4667075

4667075

bitte wenden!

7906620785

GB Informationstechnologie	Modul 3.1	Seite 2/3					
Seminar			trifft voll zu			trifft nicht zu	
		1	2	3	4	5	6
1) Die inhaltlichen Ziele und der Aufbau des Seminars waren klar zu erkennen	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Das Seminar war motivierend, anregend und lebendig	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) In diesem Seminar habe ich viel gelernt	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Nach- und Vorbereitung des Stoffes wurden durch angebotene Unterlagen, wie Skripte, Literatur oder Software, erleichtert	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Die Erarbeitung des Stoffes wurde von den Dozenten sinnvoll begleitet	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Mit der organisatorischen Durchführung des Seminars war ich zufrieden	Tutorial Harvey Humangenetik Freiw. Seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
UaK			trifft voll zu			trifft nicht zu	
		1	2	3	4	5	6
1) Die Lehrveranstaltung war motivierend, anregend und lebendig	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Bezogen auf meine theoretischen Kenntnisse habe ich viel gelernt	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Bezogen auf meine praktischen Fertigkeiten habe ich viel gelernt	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Der Patientenkontakt war in seinem Umfang ausreichend zur Vermittlung der Lehrinhalte	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Der Aufbau der Lehrveranstaltung war strukturiert und fügte sich gut in das Gesamtmodul ein	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Die Betreuung durch die Dozenten war gut	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7) Mit der organisatorischen Durchführung der Lehrveranstaltung war ich zufrieden	Kardiologie Pädiatrie Pulmonologie THG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

0482620787

Personenbezogene Kritik, Verbesserungsvorschläge, Kommentare:

Diese Lehrveranstaltungsform war für meine pers. Lernerfolg sehr wichtig:

	trifft zu 1	2	trifft 3	nicht zu 4	5
1) Vorlesung (Vormittags)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Repetitorium (Fr. 11-12h)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Seminar (16-17h)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Praktikum auf Station (13-16h)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Kursus Pathologie (14-15h)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Harvey (14-15h bzw. 15-16h)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7) Gerätekunde (1 Termin)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8) Tutorial (4 Fälle)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Der Umfang folgender Fachgebiete war im Hinblick auf die kardiopulmonale Thematik meiner Ansicht nach in folgender Weise repräsentiert:

	zu hoch	Umfang angemessen	zu niedrig
1) Kardiologie (Innere Medizin)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Pneumologie (Innere Medizin)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Thorax-Herz-Gefäß-Chirurgie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4) Kinderkardiologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5) Pharmakologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6) Radiologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7) Laboratoriumsmedizin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8) Herz- und Kreislaufphysiologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9) Pathologie	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10) Psychosomatische Aspekte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11) Mikrobiologische Aspekte	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wie beurteilen sie das EKG-Skript:

	1	2	3	4	5	6
1) Welche Gesamtnote geben sie dem Skript?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Fühlen sie sich nach dem EKG-Seminar in der Lage, EKGs zu befunden?	<input type="checkbox"/>	ja	<input type="checkbox"/>	unsicher	<input type="checkbox"/>	nein

Modulergebnis

	1	2	3	4	5	6
1) Wie würden Sie rückblickend Ihren Kenntnisstand im kardiopulmonalen Bereich vor Semesterbeginn beurteilen (als Schulnote).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2) Wie würden Sie heute ihren aktuellen Kenntnisstand im kardiopulmonalen Bereich vor Semesterbeginn beurteilen (als Schulnote).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3) Gesamtnote Modul 3.1 (Curriculum Kardiopulmunale Lehre)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6.2 Probandeninformation

Georg-August-Universität Göttingen • Bereich Humanmedizin
Universitätsklinikum • Medizinische Fakultät

Probanden/innen-Information zur Studie

„Störgrößenanalyse der vergleichenden studentischen Selbsteinschätzung zur Evaluation von Lehrveranstaltungen im Studium der Humanmedizin“

Sehr geehrte Probandin,
sehr geehrter Proband,

im Sommersemester 2011 nehmen Sie am sechswöchigen Lehrmodul „Erkrankungen des Herz-Kreislauf-Systems und der Lunge“ (Modul 3.1) teil. Wir führen in diesem Semester eine Studie zur Lehrforschung durch, in der mögliche Störgrößen der Lehrveranstaltungsevaluation mittels studentischer Selbsteinschätzungen näher charakterisiert werden sollen. Im Folgenden möchten wir Ihnen die Studie kurz vorstellen.

Die Evaluation einzelner Lehrveranstaltungen und ganzer Studienabschnitte ist integraler Bestandteil eines jeden Curriculums und ist in den Evaluationsordnungen der Universitäten, so auch der Georg-August-Universität Göttingen, vorgeschrieben.

In der Regel werden recht allgemeine studentische Bewertungen auf einer Schulnotenskala zur Evaluation herangezogen, die in erster Linie jedoch die studentische Zufriedenheit mit strukturellen und organisatorischen Aspekten der Lehre, nicht aber den tatsächlichen Lernerfolg, messen und noch dazu extrem störanfällig sind. Aufgrund dessen werden seit dem Wintersemester 08/09 alle Studierenden vor Beginn und am Ende jedes Moduls im Kerncurriculum der UMG darum gebeten, Selbsteinschätzungen ihres Leistungsstandes hinsichtlich der auf das jeweilige Modul bezogenen Lernziele vorzunehmen. Das studentische Feedback hat sich als extrem hilfreich für die Planung des Moduls erwiesen – in der Tat wurden einige Änderungen nur deshalb vorgenommen, weil die vergleichende Selbsteinschätzung Schwachstellen im Modul aufgedeckt hatte.

Das Evaluationsinstrument der vergleichenden Selbsteinschätzung erscheint daher sehr nützlich – allerdings sind wissenschaftliche Studien nötig, um zu untersuchen wie verlässlich die Daten sind, die mit dem neuen Instrument erhoben werden. Folgende Fragen konnten bisher nicht beantwortet werden:

1. Genügt es, den Lernerfolg aus den Mittelwerten der Angaben aller Studierenden vor und nach der Modulteilnahme zu berechnen, oder müssen die Selbsteinschätzungen auf individueller Ebene miteinander verglichen werden?
2. Schätzen Studierende sich anders ein, wenn sie erst am Ende des Moduls gebeten werden, ihren Wissensstand vor Modulbeginn *retrospektiv* einzuschätzen?
3. Besteht ein Zusammenhang zwischen subjektiven Selbsteinschätzungen und objektiven Prüfungsleistungen?
4. Inwiefern werden studentische Selbsteinschätzungen durch das Selbstwertgefühl beeinflusst?

Insbesondere die letzte Frage muss dringend untersucht werden, da eine Über- oder Unterschätzung des eigenen Wissens nicht ohne Einfluss auf die Evaluationsergebnisse bleiben kann. Auf diese Weise entsteht aber eine Störung der Ergebnisse, denn der Lernzuwachs soll möglichst „objektiv“ gemessen werden und von individuellen Merkmalen unabhängig sein.

Datenerhebung:

Wir möchten Sie zur Teilnahme an einer Studie einladen, in der die oben genannten Fragen beantwortet werden sollen. Die Teilnahme an der Studie setzt voraus, dass Sie bereit sind, uns die

Erlaubnis zu erteilen, Ihre Daten in pseudonymisierter Form auszuwerten. Im Rahmen der Studie werden Sie gebeten Ihr Wissen vor Modulbeginn einzuschätzen – gleichzeitig werden einige Daten zum Selbstwertgefühl erhoben, die absolut vertraulich behandelt und ohne Bezug zu Ihrer Person ausgewertet werden. Die Fragebögen zur Selbsteinschätzungen und zum Selbstkonzept werden sechs Tage lang bis zum Abend vor Beginn des Moduls online zur Verfügung stehen.

Die Post-Evaluation findet wie bisher am Ende des Moduls statt. Hier wird zusätzlich zur Einschätzung des derzeitigen Wissenstandes auch eine retrospektive Einschätzung des Wissenstandes vor Modulbeginn erbeten.

Zur Bestimmung des objektiven Lernerfolgs im Modul werden außerdem zwei Testate - abgehalten, die nicht benotet werden, Ihnen aber zur Einordnung Ihres eigenen Wissensstands dienen können. Außerdem werden unter allen Studienteilnehmern insgesamt 30 Büchergutscheine à 25 Euro verlost.

- 1) Eingangstestat: Am ersten Modultag (11. April 2011) wird der Leistungsstand vor Beginn der Lehre bestimmt. Es werden keine Leistungspunkte oder Geldpreise vergeben. **Die Teilnahme an der Eingangsprüfung ist aber Voraussetzung für die Teilnahme an der Studie.** Eine Vorbereitung auf das Eingangstestat ist **nicht** notwendig.
- 2) Abschlusstestat: Am Dienstag der letzten Modulwoche (17. Mai 2011) wird der Leistungsstand zu Modulende bestimmt. Es werden keine Leistungspunkte oder Geldpreise vergeben, aber Sie erhalten noch vor der Klausur eine Rückmeldung über Ihre Leistung, so dass Sie gegebenenfalls Wissenslücken vor der Klausur schließen können.

Datenschutz

Im Rahmen der Studie werden folgende personenbezogene Daten erhoben:

- Matrikelnummer
- Alter und Geschlecht
- Fachsemester
- Prüfungsleistungen (Eingangs- und Abschlusstestat, Modulabschlussklausur)

Zur Zusammenführung der Evaluationsdaten mit den Resultaten der bewerteten Klausur ist eine individuelle Kennzeichnung notwendig. Dies geschieht durch die Matrikelnummer und einen speziellen Code, den Sie selbst erstellen werden. Auch für die Verlosung der Büchergutscheine muss diese bekannt sein. Nach der Zusammenführung der Prüfungsergebnissen und Befragungen werden die Nummern komplett aus dem Datensatz gelöscht, so dass die Auswertung vollkommen anonymisiert erfolgt.

Vermeidung einer Benachteiligung durch Teilnahme an der Studie

Allen Studierenden stehen unabhängig von einer Studienteilnahme die gleichen Lehr-Ressourcen zur Verfügung, sodass es in keinem Fall zu einer Benachteiligung der Studenten bei Nichtteilnahme kommt, allerdings besteht dann keine Chance auf einen der Büchergutscheine.

Die Studie wird vom Studiendekanat der Fakultät unterstützt. Die Teilnahme an der Studie ist freiwillig.

Prof. Dr. med. G. Burckhardt
(Studiendekan)

Dr. med T. Raupach, MME
(Projektleiter)

6.3 Einverständniserklärung

Einverständniserklärung für Probanden/innen zur Studie

*„„Störgrößenanalyse der vergleichenden studentischen
Selbsteinschätzung zur Evaluation von Lehrveranstaltungen
im Studium der Humanmedizin““*

Proband/in

Nachname:

Vorname:

Matrikelnummer:

Ich wurde vom Studienleiter vollständig über Wesen, Bedeutung und Tragweite der Studie aufgeklärt. Ich hatte die Möglichkeit, per Mail Fragen zu stellen, habe falls erforderlich Antworten erhalten und diese verstanden.

Ich werde die Ergebnisse der Prüfungen sowie eine transparente Darstellung der Bewertung und der Büchergutscheinverlosung nach Modulende erhalten. Mir ist bekannt, dass meine persönlichen Daten in pseudonymisierter Form (Matrikelnummer und Code) gespeichert werden. Sobald die Datensammlung abgeschlossen ist, werden die Daten vollständig anonymisiert, so dass dann keine persönlichen Daten mehr gespeichert werden. Dokumente, die identifizierende Daten enthalten, werden vor Beginn der studienbezogenen Datenanalyse vernichtet.

Die Verwendung der Daten aus den Prüfungen und Befragungen setzt vor der Teilnahme an der Studie folgende freiwillig abgegebene Einwilligungserklärung voraus, d.h. ohne die nachfolgende Einwilligung kann ich nicht an der Studie teilnehmen.

Ich erkläre mich damit einverstanden, dass im Rahmen dieser Studie erhobene Daten, insbesondere die Matrikelnummer und fachspezifische Prüfungsergebnisse, in Papierform und auf elektronischen Datenträgern im Referat Lehre aufgezeichnet werden. Die Speicherung erfolgt in pseudonymisierter Form. Die Daten sind vor unberechtigtem Zugriff geschützt. Die Einwilligung kann jederzeit ohne Angabe von Gründen widerrufen werden.

Ich habe eine Kopie der Probanden/innen-Information erhalten. Hiermit erkläre ich meine freiwillige Teilnahme an der Studie.

Ort und Datum

Unterschrift des/der Probanden/in

Göttingen, 11. April 2011

Ort und Datum

Unterschrift des Studienleiters

6.4 Rosenbergskala

1. Alles in allem bin ich mit mir selbst zufrieden.	0	1	2	3
2. Hin und wieder denke ich, dass ich gar nichts tauge.	0	1	2	3
3. Ich besitze eine Reihe guter Eigenschaften.	0	1	2	3
4. Ich kann vieles genauso gut wie die meisten anderen Menschen auch.	0	1	2	3
5. Ich fürchte, es gibt nicht viel, worauf ich stolz sein kann.	0	1	2	3
6. Ich fühle mich von Zeit zu Zeit richtig nutzlos.	0	1	2	3
7. Ich halte mich für einen wertvollen Menschen, jedenfalls bin ich nicht weniger wertvoll als andere auch.	0	1	2	3
8. Ich wünschte, ich könnte vor mir selbst mehr Achtung haben.	0	1	2	3
9. Alles in allem neige ich dazu, mich für einen Versager zu halten.	0	1	2	3
10. Ich habe eine positive Einstellung zu mir selbst gefunden.	0	1	2	3

0 trifft überhaupt nicht zu

1 trifft eher nicht zu

2 trifft eher zu

3 trifft voll und ganz zu

(Quelle (Ferring and Filipp 1996))

6.5 Prä-Evaluationsbogen

EvaSys	Prä-Evaluation Modul 3.1	Electric Paper
Universitätsmedizin Göttingen Interdisziplinär	Deborah Reinhardt Prä 3.1 Papierformat	UMG

Markieren Sie so: Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
 Korrektur: Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

Dieser Evaluationsbogen wird im Rahmen der angekündigten Studie zur Qualitätskontrolle des Lernerfolgs, der Organisation und Struktur in diesem Modul eingesetzt. Das vollständige Ausfüllen des Bogens ist für die Auswertung der Studienergebnisse hilfreich.

VIELEN DANK!

1. Allgemeines

Zur anonymen Zusammenführung der Evaluationsdaten vor und nach dem Modul ist eine individuelle Kennzeichnung notwendig. Auf keinen Fall werden wir jedoch Ihre Angaben und Selbsteinschätzungen für andere Zwecke als die Studie zur Evaluation nutzen. Alle Daten werden höchst vertraulich behandelt. Ihre Angaben werden absolut keinen Einfluss auf Ihr Abschneiden in den Modulprüfungen haben! Zur Identifikation bitten wir Sie, an dieser Stelle einen individuellen, gut merkbaren Code zu erstellen, diesen sich ggf. zu notieren und bei allen weiteren Datensammlungen im Rahmen der Studie anzugeben. Der Code muss aus 4 Buchstaben und zwei darauffolgenden Zahlen zusammengesetzt sein.

1.1 Bitte geben Sie Ihren selbst generierten Code an.

1.2 Geschlecht männlich weiblich

1.3 Alter

- 1.4 Ich glaube dieses Modul ist wichtig für meine spätere berufliche Zukunft. trifft voll zu trifft überhaupt nicht zu
- 1.5 Ich freue mich schon auf dieses Modul. trifft voll zu trifft überhaupt nicht zu
- 1.6 Ich habe von Kommilitoninnen und Kommilitonen vor allem Gutes über dieses Modul gehört. trifft voll zu trifft überhaupt nicht zu

2. Lernzielfragen 1

- 2.1 Ich kenne den klinischen Unterschied zwischen der Pericarditis epistenocardica und dem Dressler-Syndrom. trifft voll zu trifft überhaupt nicht zu
- 2.2 Ich kenne die Sofortmaßnahmen, die beim Verdacht auf einen akuten Gefäßverschluss ergriffen werden müssen. trifft voll zu trifft überhaupt nicht zu
- 2.3 Ich weiß, welche therapeutischen Maßnahmen bei einem Patienten mit akutem Myokardinfarkt ergriffen werden müssen. trifft voll zu trifft überhaupt nicht zu
- 2.4 Ich weiß, wie die Aortendissektion „Typ A“ definiert ist. trifft voll zu trifft überhaupt nicht zu



2. Lernzielfragen 1 [Fortsetzung]

- | | | | | | | | | | |
|------|--|----------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------------------|
| 2.5 | Ich kann ein EKG interpretieren. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.6 | Ich weiß, welcher Wert auf einem Lungenfunktions-Befund anzeigt, ob bei dem betreffenden Patienten eine bronchiale Obstruktion vorliegt. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.7 | Ich bin in der Lage, anhand einer Ergometrie eine myokardiale Belastungsischämie zu erkennen. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.8 | Ich bin in der Lage, den Halsvenenpuls zu beurteilen. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.9 | Ich kann alle Pulse der unteren Extremität sicher tasten. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.10 | Ich kann eine vollständige Raucheranamnese erheben. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.11 | Ich bin in der Lage, bei der körperlichen Untersuchung eines Patienten mit Herzinsuffizienz die entscheidenden Befunde zu erheben. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.12 | Ich kenne die Indikationen zur Echokardiographie. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.13 | Ich bin in der Lage, eine Blutentnahme für die Bestimmung des Lipid-Profiles so anzuordnen, daß das Risiko für eine Verfälschung der Werte minimiert wird. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.14 | Ich kann bei der Lungenauscultation trockene von feuchten Rasselgeräuschen unterscheiden. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.15 | Ich kann in der körperlichen Untersuchung Unterschenkelödeme von einem Lymphödem unterscheiden. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 2.16 | Ich bin in der Lage, anhand des Auscultationsbefundes eine Mitralinsuffizienz von einer Aortenstenose zu unterscheiden. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |

3. Erfassung des Selbstkonzepts

Es gibt Hinweise darauf, dass Selbsteinschätzungen durch das individuelle Selbstkonzept beeinflusst werden bzw. mit diesem zusammenhängen. Das führt möglicherweise zu einer Verzerrung des selbsteingeschätzten Wissenstandes. Um diesen Zusammenhang zu untersuchen, werden Ihnen im Folgenden Fragen des meist verwendeten und weltweit anerkannten psychologischen Fragebogens zum Selbstwertgefühl präsentiert. Möglicherweise erscheinen Ihnen diese Fragen sehr persönlich, wir bitten Sie aber trotzdem inständig darum sie ehrlich und vollständig zu beantworten, da dies für den Erfolg der Studie notwendig ist. Ihre Daten werden selbstverständlich absolut vertraulich behandelt und ohne jeglichen Bezug zu Ihrer Person ausgewertet.

- 3.1 Wie sehr interessieren Sie die Inhalte dieses Moduls? sehr überhaupt nicht



3. Erfassung des Selbstkonzepts [Fortsetzung]

3.2	Alles in allem bin ich mit mir selbst zufrieden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.3	Hin und wieder denke ich, dass ich gar nichts taue.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.4	Ich besitze eine Reihe guter Eigenschaften.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.5	Ich kann vieles genauso gut wie die meisten anderen Menschen auch.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.6	Ich fürchte, es gibt nicht viel, worauf ich stolz sein kann.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.7	Ich fühle mich von Zeit zu Zeit richtig nutzlos.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.8	Ich halte mich für einen wertvollen Menschen, jedenfalls bin ich nicht weniger wertvoll als andere auch.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.9	Ich wünschte, ich könnte vor mir selbst mehr Achtung haben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.10	Alles in allem neige ich dazu, mich für einen Versager zu halten.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.11	Ich habe eine positive Einstellung zu mir selbst gefunden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.12	Ich lüge nie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
3.13	Halten Sie stets ein Versprechen, gleichgültig wie schwierig es auch sein könnte, das zu tun, was Sie gesagt haben?	immer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	nie

4. Lernzielfragen 2

4.1	Ich kann die drei Kardinalsymptome der Aortenklappenstenose aufzählen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.2	Ich kann das Stufenschema zur Behandlung des Asthma bronchiale anwenden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.3	Ich kenne die pathophysiologischen Mechanismen, die an der Entstehung der Atherosklerose beteiligt sind.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.4	Ich weiß, wie die Diagnose 'chronische Bronchitis' definiert ist.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu



4. Lernzielfragen 2 [Fortsetzung]

4.5	Ich kann den Schweregrad einer chronisch-obstruktiven Lungenerkrankung nach dem Stufenschema angeben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.6	Ich weiß, anhand welcher lungenfunktioneller Parameter eine Überblähung beim Lungenemphysem erkannt werden kann.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.7	Ich kenne die Hauptkriterien nach Duke für die Diagnose der Endokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.8	Ich kenne die vier Fehlbildungen, die zur Fallot'schen Tetralogie gehören.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.9	Ich weiß, welche Medikamente zur Therapie der Herzinsuffizienz in den einzelnen NYHA-Stadien zum Einsatz kommen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.10	Ich kenne die risiko-adaptierten LDL-Zielwerte, die bei der Behandlung der Hyperlipidämie angestrebt werden sollen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.11	Ich weiß, welcher Zielblutdruck bei spezifischen Patientengruppen mit arterieller Hypertonie angestrebt werden soll.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.12	Ich kann die einzelnen Stadien der Sarkoidose definieren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.13	Ich verstehe die genetischen Ursachen der hypertroph-obstruktiven Kardiomyopathie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.14	Ich kenne die fünf Risikofaktoren erster Ordnung für die Entstehung der Koronaren Herzkrankheit.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.15	Ich weiß, welche Risikofaktoren für eine Lungenembolie in den Wells-Score eingehen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.16	Ich kenne die Ursachen der Mitralinsuffizienz.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.17	Ich kenne die typischen Symptome und Befunde bei Myokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.18	Ich kenne die Stadieneinteilung der peripheren arteriellen Verschlusskrankheit nach Fontane.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.19	Ich kenne den Auskultationsbefund beim PFO (persistierendes Foramen ovale).	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.20	Ich weiß, welche Erkrankungen mit einem exsudativen bzw. transsudativen Pleuraerguß einhergehen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.21	Ich kenne die Empfehlungen für die ambulante antibiotische Therapie einer unkomplizierten Pneumonie bei Patienten ohne weitere Risikofaktoren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu



4. Lernzielfragen 2 [Fortsetzung]

4.22 Ich kenne die typischen Symptome und Befunde beim Pneumothorax.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.23 Ich kenne die therapeutischen Maßnahmen, die bei der pulmonalen Hypertonie zum Einsatz kommen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.24 Ich kenne die EKG-Kriterien für die Unterscheidung zwischen einer supraventrikulären und einer ventrikulären Tachykardie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.25 Ich kenne die fünf wichtigsten pathophysiologischen Mechanismen, die zum kardiogenen Schock führen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.26 Ich kenne die Hauptkriterien für die Diagnose des Rheumatischen Fiebers.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.27 Ich kann die fünf in der Tabakentwöhnung üblichen Maßnahmen benennen, die als die "5 A's" bekannt sind.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.28 Ich kenne die Nebenwirkungen der fünf wichtigsten Medikamente, die zur Therapie der Tuberkulose eingesetzt werden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.29 Ich kenne die typischen Erreger der akuten Bronchitis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.30 Ich weiß, für welche Personengruppen eine Impfung gegen Influenza empfohlen wird.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

6.6 Post-Evaluationsbogen

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
		UMG

Markieren Sie so: Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
Korrektur: Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

Dieser Evaluationsbogen wird im Rahmen der angekündigten Studie zur Qualitätskontrolle des Lernerfolgs, der Organisation und Struktur in diesem Modul eingesetzt. Das vollständige Ausfüllen des Bogens ist für die Auswertung der Studienergebnisse hilfreich und **VORAUSSETZUNG FÜR DIE TEILNAHME AN DER BÜCHERGUTSCHEIN-VERLOSUNG**.

VIELEN DANK!

1. Gesamtbewertung

1.1 Bitte bewerten Sie das Modul mit einer Gesamtnote. Note 1 Note 6 Enthaltung

2. Wochenarbeitszeit / -Lernzeit

2.1 Wie hoch ist ihr wöchentlicher Zeitaufwand für dieses Modul außerhalb der Lehrveranstaltungen (Angabe in Stunden/Woche)?

5 10 15 20 25 30 35 40

2.2 Wie hoch schätzen Sie Ihre durchschnittliche wöchentliche Mehrarbeit ein, die Sie aufwenden mussten, um Mängel in der Modulaufbereitung (Organisation, Struktur, Stoffpräsentation, Verständlichkeit) auszugleichen (Angabe in Stunden/Woche)?

0 5 10 15 20 25 30 35

2.3 Nehmen Sie an den freiwilligen Lehrangeboten (z.B. Vorlesungen) teil? immer nie

2.4 Wie sehr interessieren Sie die Inhalte des Moduls? sehr überhaupt nicht

3. Allgemeines

3.1 Bitte geben Sie Ihren selbst generierten Code an.

3.2 Geschlecht männlich weiblich

3.3 Alter

4. Lernzielfragen - Kenntnisse nach dem Modul

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
--------	---------------------------	----------------

4. Lernzielfragen - Kenntnisse nach dem Modul [Fortsetzung]

4.1	Ich kenne den klinischen Unterschied zwischen der Pericarditis epistenocardica und dem Dressler-Syndrom.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.2	Ich kenne die Sofortmaßnahmen, die beim Verdacht auf einen akuten Gefäßverschuß ergriffen werden müssen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.3	Ich weiß, welche therapeutischen Maßnahmen bei einem Patienten mit akutem Myokardinfarkt ergriffen werden müssen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.4	Ich weiß, wie die Aortendissektion „Typ A“ definiert ist.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.5	Ich kann ein EKG interpretieren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.6	Ich weiß, welcher Wert auf einem Lungenfunktions-Befund anzeigt, ob bei dem betreffenden Patienten eine bronchiale Obstruktion vorliegt.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.7	Ich bin in der Lage, anhand einer Ergometrie eine myokardiale Belastungsschämie zu erkennen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.8	Ich bin in der Lage, den Halsvenenpuls zu beurteilen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.9	Ich kann alle Pulse der unteren Extremität sicher tasten.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.10	Ich kann eine vollständige Raucheranamnese erheben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.11	Ich bin in der Lage, bei der körperlichen Untersuchung eines Patienten mit Herzinsuffizienz die entscheidenden Befunde zu erheben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.12	Ich kenne die Indikationen zur Echokardiographie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.13	Ich bin in der Lage, eine Blutentnahme für die Bestimmung des Lipid-Profiles so anzuordnen, daß das Risiko für eine Verfälschung der Werte minimiert wird.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.14	Ich kann bei der Lungenauscultation trockene von feuchten Rasselgeräuschen unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.15	Ich kann in der körperlichen Untersuchung Unterschenkelödeme von einem Lymphödem unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.16	Ich bin in der Lage, anhand des Auscultationsbefundes eine Mitralinsuffizienz von einer Aortenstenose zu unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.17	Ich kann die drei Kardinalsymptome der Aortenklappenstenose aufzählen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
--------	---------------------------	----------------

4. Lernzielfragen - Kenntnisse nach dem Modul [Fortsetzung]

4.18 Ich bin in der Lage, anhand des Auskultationsbefundes eine Mitralinsuffizienz von einer Aortenstenose zu unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.19 Ich kann das Stufenschema zur Behandlung des Asthma bronchiale anwenden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.20 Ich kenne die pathophysiologischen Mechanismen, die an der Entstehung der Atherosklerose beteiligt sind.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.21 Ich weiß, wie die Diagnose 'chronische Bronchitis' definiert ist.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.22 Ich kann den Schweregrad einer chronisch-obstruktiven Lungenerkrankung nach dem Stufenschema angeben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.23 Ich weiß, anhand welcher lungenfunktioneller Parameter eine Überblähung beim Lungenemphysem erkannt werden kann.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.24 Ich kenne die Hauptkriterien nach Duke für die Diagnose der Endokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.25 Ich kenne die vier Fehlbildungen, die zur Fallot'schen Tetralogie gehören.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.26 Ich weiß, welche Medikamente zur Therapie der Herzinsuffizienz in den einzelnen NYHA-Stadien zum Einsatz kommen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.27 Ich kenne die risiko-adaptierten LDL-Zielwerte, die bei der Behandlung der Hyperlipidämie angestrebt werden sollen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.28 Ich weiß, welcher Zielblutdruck bei spezifischen Patientengruppen mit arterieller Hypertonie angestrebt werden soll.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.29 Ich kann die einzelnen Stadien der Sarkoidose definieren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.30 Ich verstehe die genetischen Ursachen der hypertroph-obstruktiven Kardiomyopathie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.31 Ich kenne die fünf Risikofaktoren erster Ordnung für die Entstehung der Koronaren Herzkrankheit.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.32 Ich weiß, welche Risikofaktoren für eine Lungenembolie in den Wells-Score eingehen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.33 Ich kenne die Ursachen der Mitralinsuffizienz.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.34 Ich kenne die typischen Symptome und Befunde bei Myokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
--------	---------------------------	----------------

4. Lernzielfragen - Kenntnisse nach dem Modul [Fortsetzung]

4.35 Ich kenne die Stadieneinteilung der peripheren arteriellen Verschlusskrankheit nach Fontane.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.36 Ich kenne den Auskultationsbefund beim PFO (persistierendes Foramen ovale).	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.37 Ich weiß, welche Erkrankungen mit einem exsudativen bzw. transsudativen Pleuraerguß einhergehen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.38 Ich kenne die Empfehlungen für die ambulante antibiotische Therapie einer unkomplizierten Pneumonie bei Patienten ohne weitere Risikofaktoren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.39 Ich kenne die typischen Symptome und Befunde beim Pneumothorax.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.40 Ich kenne die therapeutischen Maßnahmen, die bei der pulmonalen Hypertonie zum Einsatz kommen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.41 Ich kenne die EKG-Kriterien für die Unterscheidung zwischen einer supraventrikulären und einer ventrikulären Tachykardie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.42 Ich kenne die fünf wichtigsten pathophysiologischen Mechanismen, die zum kardiogenen Schock führen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.43 Ich kenne die Hauptkriterien für die Diagnose des Rheumatischen Fiebers.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.44 Ich kann die fünf in der Tabakentwöhnung üblichen Maßnahmen benennen, die als die '5 A's' bekannt sind.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.45 Ich kenne die Nebenwirkungen der fünf wichtigsten Medikamente, die zur Therapie der Tuberkulose eingesetzt werden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.46 Ich kenne die typischen Erreger der akuten Bronchitis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
4.47 Ich weiß, für welche Personengruppen eine Impfung gegen Influenza empfohlen wird.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

5. retrospektive Einschätzung der Vorkenntnisse

5.1 Ich kenne den klinischen Unterschied zwischen der Pericarditis epistenocardica und dem Dressler-Syndrom.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.2 Ich kenne die Sofortmaßnahmen, die beim Verdacht auf einen akuten Gefäßverschluss ergriffen werden müssen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.3 Ich weiß, welche therapeutischen Maßnahmen bei einem Patienten mit akutem Myokardinfarkt ergriffen werden müssen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
--------	---------------------------	----------------

5. retrospektive Einschätzung der Vorkenntnisse [Fortsetzung]

5.4	Ich weiß, wie die Aortendissektion „Typ A“ definiert ist.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.5	Ich kann ein EKG interpretieren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.6	Ich weiß, welcher Wert auf einem Lungenfunktions-Befund anzeigt, ob bei dem betreffenden Patienten eine bronchiale Obstruktion vorliegt.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.7	Ich bin in der Lage, anhand einer Ergometrie eine myokardiale Belastungsischämie zu erkennen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.8	Ich bin in der Lage, den Halsvenenpuls zu beurteilen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.9	Ich kann alle Pulse der unteren Extremität sicher tasten.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.10	Ich kann eine vollständige Raucheranamnese erheben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.11	Ich bin in der Lage, bei der körperlichen Untersuchung eines Patienten mit Herzinsuffizienz die entscheidenden Befunde zu erheben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.12	Ich kenne die Indikationen zur Echokardiographie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.13	Ich bin in der Lage, eine Blutentnahme für die Bestimmung des Lipid-Profiles so anzuordnen, daß das Risiko für eine Verfälschung der Werte minimiert wird.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.14	Ich kann bei der Lungenauscultation trockene von feuchten Rasselgeräuschen unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.15	Ich kann in der körperlichen Untersuchung Unterschenkelödeme von einem Lymphödem unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.16	Ich bin in der Lage, anhand des Auscultationsbefundes eine Mitralinsuffizienz von einer Aortenstenose zu unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.17	Ich kann die drei Kardinalsymptome der Aortenklappenstenose aufzählen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.18	Ich bin in der Lage, anhand des Auscultationsbefundes eine Mitralinsuffizienz von einer Aortenstenose zu unterscheiden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.19	Ich kann das Stufenschema zur Behandlung des Asthma bronchiale anwenden.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.20	Ich kenne die pathophysiologischen Mechanismen, die an der Entstehung der Atherosklerose beteiligt sind.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

MUSTER

EvaSys	Post-Evaluation Modul 3.1	Electric Paper
--------	---------------------------	----------------

5. retrospektive Einschätzung der Vorkenntnisse [Fortsetzung]

5.21 Ich weiß, wie die Diagnose ,chronische Bronchitis ' definiert ist.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.22 Ich kann den Schweregrad einer chronisch-obstruktiven Lungenerkrankung nach dem Stufenschema angeben.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.23 Ich weiß, anhand welcher lungenfunktioneller Parameter eine Überblähung beim Lungenemphysem erkannt werden kann.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.24 Ich kenne die Hauptkriterien nach Duke für die Diagnose der Endokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.25 Ich kenne die vier Fehlbildungen, die zurFallot'schen Tetralogie gehören.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.26 Ich weiß, welche Medikamente zur Therapie der Herzinsuffizienz in den einzelnen NYHA-Stadien zum Einsatz kommen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.27 Ich kenne die risiko-adaptierten LDL-Zielwerte, die bei der Behandlung der Hyperlipidämie angestrebt werden sollen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.28 Ich weiß, welcher Zielblutdruck bei spezifischen Patientengruppen mit arterieller Hypertonie angestrebt werden soll.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.29 Ich kann die einzelnen Stadien der Sarkoidose definieren.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.30 Ich verstehe die genetischen Ursachen der hypertroph-obstruktiven Kardiomyopathie.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.31 Ich kenne die fünf Risikofaktoren erster Ordnung für die Entstehung der Koronaren Herzkrankheit.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.32 Ich weiß, welche Risikofaktoren für eine Lungenembolie in den Wells-Score eingehen.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.33 Ich kenne die Ursachen der Mitralinsuffizienz.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.34 Ich kenne die typischen Symptome und Befunde bei Myokarditis.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.35 Ich kenne die Stadieneinteilung der peripheren arteriellen Verschlusskrankheit nach Fontane.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.36 Ich kenne den Auskultationsbefund beim PFO (persistierendes Foramen ovale).	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu
5.37 Ich weiß, welche Erkrankungen mit einem exsudativen bzw. transsudativen Pleuraerguß einhergehen können.	trifft voll zu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	trifft überhaupt nicht zu

MUSTER

EvaSys

Post-Evaluation Modul 3.1

Electric Paper

5. retrospektive Einschätzung der Vorkenntnisse [Fortsetzung]

- | | | | | | | | | |
|---|----------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|---------------------------|
| 5.38 Ich kenne die Empfehlungen für die ambulante antibiotische Therapie einer unkomplizierten Pneumonie bei Patienten ohne weitere Risikofaktoren. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.39 Ich kenne die typischen Symptome und Befunde beim Pneumothorax. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.40 Ich kenne die therapeutischen Maßnahmen, die bei der pulmonalen Hypertonie zum Einsatz kommen können. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.41 Ich kenne die EKG-Kriterien für die Unterscheidung zwischen einer supraventrikulären und einer ventrikulären Tachykardie. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.42 Ich kenne die fünf wichtigsten pathophysiologischen Mechanismen, die zum kardiogenen Schock führen können. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.43 Ich kenne die Hauptkriterien für die Diagnose des Rheumatischen Fiebers. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.44 Ich kann die fünf in der Tabakentwöhnung üblichen Maßnahmen benennen, die als die "5 A's" bekannt sind. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.45 Ich kenne die Nebenwirkungen der fünf wichtigsten Medikamente, die zur Therapie der Tuberkulose eingesetzt werden. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.46 Ich kenne die typischen Erreger der akuten Bronchitis. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |
| 5.47 Ich weiß, für welche Personengruppen eine Impfung gegen Influenza empfohlen wird. | trifft voll zu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | trifft überhaupt nicht zu |

Bitte denken Sie jetzt an den Anfang des Moduls zurück und überlegen sich, wie Sie - rückblickend - Ihr eigenes Wissen und Ihre eigenen Fertigkeiten vor Beginn des Moduls einschätzen würden. Bitte versuchen Sie nicht, sich an Ihre Angaben aus der Vorher-Evaluation zu erinnern, und stellen Sie bitte auch keinen Vergleich mit Ihren soeben gemachten Angaben zum aktuellen Wissensstand an. Im folgenden geht es wirklich nur darum, wie Sie aus heutiger Sicht Ihren damaligen Leistungsstand beurteilen. Oder anders gesagt: Wie gut waren Sie (hinsichtlich der folgenden Lernziele) VOR Beginn des Moduls?

7 Literaturverzeichnis

- Anderson, J. (2004). "Multiple-choice questions revisited." Med Teach **26**(2): 110-113.
- Baumeister, R. F., J. D. Campbell, J. I. Krueger and K. D. Vohs (2003). "Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles?" Psychol Sci Public Interest **4**: 1-44.
- Billings-Gagliardi, S., S. V. Barrett and K. M. Mazor (2004). "Interpreting course evaluation results: insights from thinkaloud interviews with medical students." Med Educ **38**(10): 1061-1070.
- Bland, J. M. and D. G. Altman (1999). "Measuring agreement in method comparison studies." Stat Methods Med Res **8**(2): 135-160.
- Braun, E. and B. Leidner (2009). "Academic course evaluation: Theoretical and empirical distinctions between self-rated gain in competences and satisfaction with teaching behavior." Eur Psychol **14**(4): 297-306.
- Cashin, W. E., Defining and evaluating college teaching (IDEA Paper No. 20), Kansas State University Manhattan (1989).
- Cohen, J. (1992). "A power primer." Psychol Bull **112**: 155-159.
- Collani, G. and P. Y. Herzberg (2003). "Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg." Zeitschrift für Differentielle und Diagnostische Psychologie, **24**(1): 3-7.
- Collani, G. and P. Y. Herzberg (2003). "[A Revised Version of the German Adaptation of Rosenberg's Self-Esteem Scale]." Z Diff Diagn Psychol **24**(1): 3-7.
- Colthart, I., G. Bagnall, A. Evans, H. Allbutt, A. Haig, J. Illing and B. McKinstry (2008). "The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10." Med Teach **30**(2): 124-145.
- Cook, N. (2007). "Use and misuse of the receiver operating characteristic curve in risk prediction." Circulation **115**: 928-935.
- d'Apollonia, S. and P. C. Abrami (1997). "Navigating student ratings of instruction." Am Psychol **52**(11): 1198-1208.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., Perrier, L. (2006). "Accuracy of physician self-assessment compared with observed measures of competence: a systematic review." JAMA **296**(9): 1094-1102.
- D'Eon, M., L. Sadownik, A. Harrison and J. Nation (2008). "Using Self-Assessment to Detect Workshop Success." Am J Eval **29**: 92-98.
- Epstein, R. M. (2007). "Assessment in medical education." N Engl J Med **356**(4): 387-396.
- Eva, K. W. and G. Regehr (2007). "Knowing when to look it up: a new conception of self-assessment ability." Acad Med **82**(10 Suppl): S81-84.
- Ferring, D. and S.-H. Filipp (1996). "Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala. / Measurement of self-esteem: Findings on reliability, validity, and stability of the Rosenberg Scale." Diagnostica **42**(3): 284-292.
- Ferring, D. and S. H. Filipp (1996). "Messung des Selbstwertgefühls: Befund zur Reliabilität, Validität und Stabilität der Rosenberg-Skala." Diagnostica **42**: 284-292.
- Fitzgerald, J. T., C. B. White and L. D. Gruppen (2003). "A longitudinal study of self-assessment accuracy." Med Educ **37**(7): 645-649.
- Frank, E., McLendon, L., Denniston, M., Fitzmaurice, D., Hertzberg, V., Elon, L. (2005). "Medical students' self-reported typical counseling practices are similar to those assessed with standardized patients." Medscape General Medicine **7**(1): 2.
- Germain, M. and T. A. Scandura (2005). "Grade inflation and Student Individual Differences as Systematic Bias in Faculty Evaluation." J Instruc Psychol **32**(1): 58-67.
- Gibson, B. P., D. A. Black, M. Cunningham, M. C. Grimm and M. H.P. (2008). "Enhancing evaluation in an undergraduate medical education program." Acad Med **83**: 787-793.

- Gordon, M. J. (1991). "A review of the validity and accuracy of self-assessments in health professions training." Acad Med **66**(12): 762-769.
- Göttinger-Lernzielkatalog. (2008). "Der Göttinger Lernzielkatalog für den klinischen Studienabschnitt." from http://www.med.uni-goettingen.de/de/media/G1-2_lehre/lernzielkatalog.pdf.
- Greenwald, A. G. (1997). "Validity concerns and usefulness of student ratings of instruction." Am Psychol **52**(11): 1182-1186.
- Griffin, B. W. (2001). "Instructor Reputation and Student Ratings of Instruction." Contemp Educ Psychol **26**(4): 534-552.
- Harden, R. M. (1999). "What is a spiral curriculum?" Med Teach **21**(2): 141-143.
- Harden, R. M., M. Stevenson, W. W. Downie and G. M. Wilson (1975). "Assessment of clinical competence using objective structured examination." Br Med J **1975**, **1**: 447-451.
- Herzig, S., B. Marschall, D. Nast-Kolb, S. Soboll, L. C. Rump and R. D. Hilgers (2007). "Positionspapier der nordrhein-westfälischen Studiendekane zur hochschulvergleichenden leistungsorientierten Mittelvergabe für die Lehre." GMS Z Med Ausbild **24**(2): Doc109.
- Howard, G. S. and S. E. Maxwell (1980). "Correlation Between Student Satisfaction and Grades: A Case of Mistaken Causation?" J Educ Psychol **72**(6): 810-820.
- Johnson (2003). "Online Student Ratings: Will Students Respond?" New Direct Teach Learn **96**: 49-59.
- Kern, D. E., P. A. Thomas, D. M. Howard and E. B. Bass, Curriculum development for medical education - A six-step approach., The Johns Hopkins University Press Baltimore and London (1998).
- Konietzko, N., H. Wendel and B. Wiesner, Erkrankungen der Lunge, de Gruyter Berlin (1995).
- Krathwohl, D. R. (2002). "A revision of bloom`s taxonomy: An overview. ." Theory Pract **41**(4): 212-260.
- Kruger, J. and D. Dunning (1999). "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." J Pers Soc Psychol **77**(6): 1121-1134.
- Lam, T. (2009). "Do Self-Assessments Work to Detect Workshop Success? An Analysis of Argument and Recommendation by D'Eon et al." Am J Eval **30**(1).
- Leopold, S. S., H. D. Morgan, N. J. Kadel, G. C. Gardner, D. C. Schaad and F. M. Wolf (2005). "Impact of educational intervention on confidence and competence in the performance of a simple surgical task." J Bone Joint Surg Am **87**: 1031-1037.
- Marsh, H. W. (1984). "Students evaluations of university teaching - dimensionality, reliability, validity, potential biases, and utility." J Educ Psychol **76**(5): 707-754.
- Marsh, H. W. and L. A. Roche (1997). "Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility." Am Psychol **52**(11): 1187-1197.
- Marsh, H. W. and L. A. Roche (2000). "Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?" J Educ Psychol **92**(1): 202-228.
- McKeachie, W. (1997). "Student ratings; the validity of use." Am Psychol **52**(11): 1218-1225.
- MFT. (2012). "Organisation der Lehrevaluation an der Medizinischen Fakultät." from http://www.landkarte-hochschulmedizin.de/TextEvaluationEvaluationsheet_ValidForAllYears.aspx?id=60.
- Miller, T. M. and L. Geraci (2011). "Unskilled but aware: Reinterpreting overconfidence in low-performing students." J Exp Psychol Learn Mem Cogn **37**: 502-506.
- Möltner (2006). "Grundlegende quantitative Analysen medizinischer Prüfungen." GMS Z Med Ausbild **23** (3): Doc53.
- Möltner (2010). "Fakultätsinterne Prüfungen an den deutschen medizinischen Fakultäten." GMS Z Med Ausbild **27** (3): Doc44.
- Naftulin, D. H., J. E. Ware, Jr. and F. A. Donnelly (1973). "The Doctor Fox Lecture: a paradigm of educational seduction." J Med Educ **48**(7): 630-635.
- Osterberg, K., S. Kölbl and K. Brauns (2006). "Der Progress Test Medizin: Erfahrungen an der Charité Berlin." GMS Z Med Ausbild. **23**(3): 62-66.

- Raupach, T., N. Hanneforth, S. Anders, T. Pukrop, O. Th J ten Cate and S. Harendza (2010). "Impact of teaching and assessment format on electrocardiogram interpretation skills." *Med Educ* **44**(7): 731-740.
- Raupach, T., C. Munscher, T. Beissbarth, G. Burckhardt and T. Pukrop (2011). "Towards outcome-based programme evaluation: Using student comparative self-assessments to determine teaching effectiveness." *Med Teach* **33**(8): 446-453.
- Raupach, T., S. Schiekirka, C. Munscher, T. Beissbarth, W. Himmel, G. Burckhardt and T. Pukrop (2012). "Piloting an outcome-based programme evaluation tool in undergraduate medical education." *GMS Z Med Ausbild* **29**(3): Doc44.
- Rogers, W. T. Y., P. (1996). "Test-Wiseness: Its nature and application." *Eur J Psychol Assess* **12**(3): 247-259.
- Schaub, H. and K. G. Zenke, Wörterbuch Pädagogik, Deutscher Taschenbuch Verlag München (1997). **2**: 131.
- Schiekirka, S., D. Reinhardt, S. Heim, G. Fabry, T. Pukrop, S. Anders and T. Raupach (2012). "Student perceptions of evaluation in undergraduate medical education: A qualitative study from one medical school." *BMC Med Educ* **12**: 45.
- Schmidt, H. G. (1983). "Problem-based learning: rationale and description." *Med Educ* **17**(1): 11-16.
- Schuwirth, L. W. and C. P. van der Vleuten (2004). "Different written assessment methods: what can be said about their strengths and weaknesses?" *Med Educ* **38**(9): 974-979.
- Shumway, J. M. and R. M. Harden (2003). "AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician." *Med Teach* **25**(6): 569-584.
- Sinclair, S. J., M. A. Blais, D. A. Gansler, E. Sandberg, K. Bistis and A. LoCicero (2010). "Psychometric properties of the Rosenberg Self-Esteem Scale: overall and across demographic groups living within the United States." *Eval Health Prof* **33**(1): 56-80.
- Skeff, K. M., G. A. Stratos and M. R. Bergen (1992). "Evaluation of a Medical Faculty Development Program: A Comparison of Traditional Pre/Post and Retrospective Pre/Post Self-Assessment Ratings." *Eval Health Prof* **15**: 350-366.
- Skeff, K. M., G. A. Stratos and M. R. Bergen (1992). "Evaluation of a Medical Faculty Development Program; A Comparison of Traditional Pre/Post and Retrospective Pre/Post Self-Assessment Ratings." *Eval Health Prof* **15**: 359-366.
- Spiel, C. and P. M. Gössler (2000). "Zum Einfluss von Biasvariablen auf die Bewertung universitärer Lehre durch Studierende." *Z Päd Psychol* **14**(1): 38 - 47.
- Spiel, C., B. Schober and R. Reimann (2006). "Evaluation of curricula in higher education: challenges for evaluators." *Evaluation review* **30**(4): 430-450.
- Theall, M. and J. Franklin (2001). "Looking for bias in all the wrong places: a search of truth or a witch hunt in student ratings of instruction?" *New Direct Inst Res* **109**: 45-56.
- Thompson, B. M. and C. Rogers (2008). "Exploring the Learning Curve in Medical Education: Using Self-Assessment as a Measure of Learning." *Acad Med* **83**(10): 568 - 588.
- University-of-Michigan. (2007). "Online Course Evaluations." from www.provost.umich.edu/reports/online_course_evaluations.pdf.
- Wikipedia. (1. März 2011). "<http://de.wikipedia.org/wiki/Evaluation>."
- Wissenschaftsrat, *Empfehlungen zur Qualitätsverbesserung von Lehre und Studium*, German Council of Sciences and Humanities Berlin (2008).

Danksagung

Ich bedanke mich für die Unterstützung und wertvollen Ratschläge bei meinem Doktorvater PD Dr. Tobias Raupach sowie bei Sarah Schiekirka, die an der Studie mitwirkte und mir bei Fragen immer zur Seite stand.