

# Hochschuldidaktische Qualifizierung in der Medizin IV: Messung von Wirksamkeit und Erfolg medizindidaktischer Qualifizierungsangebote: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg

## Faculty development initiatives in Medical Education in German-Speaking Countries: IV. Measuring Effectiveness and Success

### Abstract

The demand to strengthen the evidence base of medical education makes it necessary to evaluate the success of faculty development initiatives. To this end, it is essential to consider multifaceted variables and methods, since processes of teaching and learning are complex. Changes might be identified with regard to the individual participants (e.g., changes in knowledge, skills, and attitudes) or the students (e.g., changes in academic success). However, it is important not to hastily interpret these changes in terms of causality.

Studies show that, overall, participants are quite content with faculty development initiatives. There is also data available showing that participants gain knowledge and skills through these programs. Whether or not these changes also foster students' learning remains to be proven.

**Keywords:** medical education, faculty development, evaluation, empirical evidence

### Zusammenfassung

Angesichts der Forderung nach mehr Evidenzbasierung in der Lehre müssen auch medizindidaktische Qualifizierungsangebote auf ihren Erfolg hin evaluiert werden. Aufgrund der Komplexität von Lehr-Lernprozessen sind dabei unterschiedliche Zielkriterien und Methoden zu berücksichtigen. Veränderungen können sowohl bei den Teilnehmern selbst (z.B. Veränderung von Wissen, Fertigkeiten und Einstellungen) als auch bei den Studierenden (z.B. Lernerfolg) gemessen werden, dürfen jedoch nicht vorschnell im Sinne kausaler Zusammenhänge interpretiert werden.

Studien zeigen generell eine große Zufriedenheit der Teilnehmer mit medizindidaktischen Maßnahmen auch ein Zuwachs an Wissen und Fertigkeiten lässt sich feststellen. Ob damit auch das Lernen der Studierenden positiv beeinflusst wird, muss noch überprüft werden.

**Schlüsselwörter:** Medizindidaktik, Personal- und Organisationsentwicklung, Evaluation, Evidenzbasierung

**Götz Fabry<sup>1</sup>**

**Maria**

**Lammerding-Köppel<sup>2</sup>**

**Matthias Hofer<sup>3</sup>**

**Falk Ochsendorf<sup>4</sup>**

**Christian Schirlo<sup>5</sup>**

**Jan Breckwoldt<sup>6</sup>**

1 Albert-Ludwigs-Universität Freiburg, Medizinische Fakultät, Abteilung für Medizinische Psychologie und Soziologie, Freiburg, Deutschland

2 Universität Tübingen, Medizinische Fakultät, Kompetenzzentrum für Hochschuldidaktik in Medizin Baden-Württemberg, Tübingen, Deutschland

3 Heinrich-Heine-Universität Düsseldorf, Medizinische Fakultät, Anatomisches Institut II, Arbeitsgruppe Medizindidaktik, Düsseldorf, Deutschland

4 Johann Wolfgang Goethe-Universität Frankfurt/Main, Klinikum, Zentrum Dermatologie und Venerologie, Frankfurt/Main, Deutschland

5 Universität Zürich, Medizinische Fakultät,

Studiendekanat, Zürich,  
Schweiz

6 Charité - Universitätsmedizin  
Berlin, Campus Benjamin  
Franklin, Klinik für  
Anästhesiologie und  
operative Intensivmedizin,  
Berlin, Deutschland

## Hintergrund: Evidenzbasierte Lehre

Zunehmend wird auch für die Lehre gefordert, Methoden auf ihre Wirksamkeit hin zu überprüfen, um so in Analogie zur klinischen Medizin evidenzbasierte Entscheidungen treffen zu können [1]. Um diese Zielsetzung systematisch zu verfolgen, wurde Anfang des Jahrtausends ein Netzwerk (BEME-Collaboration – Best Evidence Medical Education, [www.bemecollaboration.org](http://www.bemecollaboration.org)) ins Leben gerufen, das aus zeitlich befristeten Arbeitsgruppen (Topic Review Groups) besteht, die Übersichtsarbeiten zu verschiedenen Themen erarbeiten [2]. Die Analogie zwischen evidenzbasierter Medizin und evidenzbasierter Lehre hat allerdings ihre Grenzen. Denn die durchaus umstrittene Bewertung der verschiedenen methodischen Vorgehensweisen kann nicht einfach von der klinischen Medizin auf empirische Studien zu Wirksamkeit und Erfolg von Lehre übertragen werden [3]. Eine randomisierte kontrollierte Studie, die in der klinischen Medizin als Goldstandard des Wirksamkeitsnachweises gilt, ist für viele medizindidaktische Fragestellungen nur schwer durchführbar, in vielen Fällen aber auch gar nicht sinnvoll [4]. Während eine Randomisierung der Gruppen häufig unproblematisch ist, bereitet die kontrollierte Darbietung der zu messenden Intervention erhebliche Schwierigkeiten. Sie ist nämlich nicht nur von der Person des Lehrenden, sondern auch von der Reaktion der Lernenden abhängig und diese wiederum unterliegt situativen, gruppenspezifischen und anderen Einflüssen. Was immer als Ergebnis gemessen wird, die Unsicherheit bleibt groß, ob sich etwaige Unterschiede wirklich auf die Methode selbst zurückführen lassen, zumal eine Verblindung ohnehin nicht möglich ist. Selbst wenn sich eine Methode zweifelsfrei als überlegen erweisen sollte, ist immer noch fraglich, in wieweit sich die Ergebnisse unter anderen Rahmenbedingungen, z.B. an einer anderen Fakultät mit einer anderen Studierenden- und Lehrendenpopulation wiederholen lassen [5].

Abbildung 1 veranschaulicht diesen Zusammenhang: Die Teilnehmer an der medizindidaktischen Fortbildung kommen aus unterschiedlichen Bereichen, in denen sie neben ihrer Lehrtätigkeit noch andere Funktionen haben, so dass die Gruppen jeweils unterschiedlich zusammengesetzt sind. Daraus entwickelt sich eine jeweils unterschiedliche Gruppendynamik, die den Lernprozess und dessen Ergebnisse beeinflusst. Zusätzlich unterliegt die Veranstaltung wechselnden Kontexteinflüssen (z.B. un-

terschiedliche Zeitpunkte, Dozenten, aktuelle Ereignisse, etc.). Auch der Transfer der erworbenen Kenntnisse und Fähigkeiten in den Lehralltag unterliegt verschiedenen intervenierenden Variablen, die sich aus der Interaktion mit den Studierenden und anderen Faktoren ergeben. Dieses komplexe Bedingungsgefüge ist beim Vergleich der tatsächlich festgestellten Ergebnisse mit den geplanten Ergebnissen zu berücksichtigen.

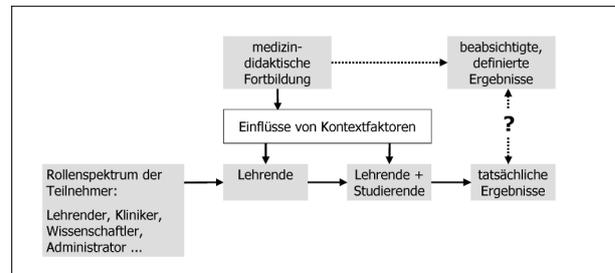


Abbildung 1: Einflussfaktoren auf das Ergebnis medizindidaktischer Qualifizierungsangebote (modifiziert n. [19])

## Zielkriterien für die Evaluation

Vor diesem Hintergrund wurde nach alternativen Kategorisierungsmöglichkeiten gesucht, um dennoch eine differenzierte Bewertung empirischer Studien im Bereich der Lehre vornehmen zu können. Die Arbeitsgruppen des BEME-Netzwerks benutzen dazu hauptsächlich die von Kirkpatrick [6], [7] vorgeschlagenen Zielkriterien zur Evaluation von Trainingsprogrammen, die in Tabelle 1 dargestellt sind. In einer systematischen Übersichtsarbeit zu den Effekten von Maßnahmen der Personalentwicklung in der medizinischen Lehre wurde eine erweiterte Form dieser Kategorien verwendet, die in Tabelle 2 wiedergegeben ist [8].

Tabelle 1: Evaluationskriterien für didaktische Trainingsprogramme [7]

1	Reaktion	auf die Erfahrungen mit dem Kurs, dem Seminar, etc.
2	Lernen	Erwerb von Kenntnissen und Fertigkeiten
3	Verhaltensänderung	Haben die Teilnehmer ihr Verhalten verändert? Wird das Gelernte in der Praxis angewendet?
4	Ergebnisse	vor allem im Hinblick auf die beabsichtigten Ziele (z.B. bessere Examensergebnisse der Studierenden)

**Tabelle 2: Bewertungskriterien medizindidaktischer Fortbildungen [19]. In der letzten Spalte ist wiedergegeben, wie viele der dort aufgenommen 53 Studien diese Kategorie erfasst haben (Mehrfachnennungen waren möglich).**

1	Reaktion	Bewertung der Lern-Erfahrung, ihrer Organisation, Darbietung, Inhalte, Methoden und der Instruktionsqualität	39 (74%)
2	a) Lernen: Veränderung von Einstellungen	Veränderungen von Wahrnehmung und Einstellung zum Lehren & Lernen bei den Teilnehmern	41 (77%)
	b) Lernen: Veränderung von Wissen & Fertigkeiten	Wissen: Konzepte, Abläufen, Prinzipien Fertigkeiten: Denken, Problemlösen, psychomotorische & soziale Fertigkeiten	
3	Verhaltensänderung	Transfer in den Alltag, Bereitschaft neues Wissen & Fertigkeiten anzuwenden	38 (72%)
4	a) Ergebnisse: system- / organisationsbezogene Veränderungen	Veränderungen auf Organisationsebene, die auf die Maßnahme zurückgeführt werden können	7 (13%)
	b) Ergebnisse: Veränderungen bei den Studierenden und den Kollegen bzw. Assistenzärzten der Teilnehmer	Verbesserungen im Lernen/in der Leistung von Studierenden, Assistenten, Kollegen, die auf die Maßnahme zurückgeführt werden können	3 (6%)

Diese Kategorien sollen ausdrücklich nicht als eine Hierarchie verstanden werden: Eine Evaluation, die sich auf die Reaktionen der Teilnehmer beschränkt, ist bei adäquater methodischer Qualität, nicht grundsätzlich schlechter als eine andere, mit der Verhaltensänderungen oder Ergebnisse festgestellt werden. Diese Aussage mag vielleicht zunächst überraschen, sie ist aber insofern angemessen, als eine empirisch abgesicherte Theorie der „Wirkfaktoren“ von guter Lehre noch weitgehend fehlt. So wäre es durchaus denkbar, dass z.B. eine durch medizindidaktische Fortbildung gesteigerte Motivation der Dozenten, die sich positiv auf das Lernklima in ihren Veranstaltungen auswirkt, auch zu einer größeren Zufriedenheit der Studierenden führt.

Wird umgekehrt eine bessere Leistung der Studierenden im Gefolge medizindidaktischer Qualifizierung festgestellt, ist die Frage, was diese Verbesserung bewirkt hat, noch nicht geklärt. Auch hier könnten neben methodischen, strukturellen oder inhaltlichen Verbesserungen ebenfalls motivationale Aspekte eine große Rolle spielen.

Diese Überlegungen machen deutlich, dass es auf jeder Kategoriestufe schwieriger wird, Evaluationen durchzuführen, deren Daten sich eindeutig auf die durchgeführte Maßnahme beziehen lassen. Entsprechend sind Studien, die Parameter der Stufe 4 erfassen, eher selten, vgl. Tabelle 5. Gleichzeitig wären aber gerade solche Studien besonders wichtig, um besser zu verstehen, wie sich die Lehren und Lernen gezielt verbessern lassen.

## Methodische und Praktische Aspekte

Die verschiedenen Parameter, mit denen die Wirksamkeit medizindidaktischer Qualifikationsangebote evaluiert werden kann, legen unterschiedliche methodische Zugänge nahe, die unterschiedlich aufwändig sind.

## Befragung der Teilnehmer

Grundsätzlich lassen sich alle Ebenen durch Befragung der Teilnehmer erfassen. Diese Form der Evaluation ist

weit verbreitet, in den meisten Studien wird sie entweder als einzige Methode oder in Kombination mit anderen Verfahren eingesetzt [8]. Verwendet werden können offene Fragen, aber auch geschlossene und standardisierte Formate. Wie bei der studentischen Evaluation von Lehrveranstaltungen auch, ist eine Kombination beider Verfahren am besten. So lassen sich Daten gewinnen, die relativ einfach auf ihre testtheoretische Qualität (Objektivität, Reliabilität, Validität) überprüft werden können. Gleichzeitig lassen sich so aber auch spezifische und konkrete Informationen erfassen, aus denen z.B. wichtige Anregungen für die weitere Gestaltung der Maßnahme entnommen werden können. In der Praxis zeigt sich allerdings, dass standardisierte Erhebungsinstrumente zur Evaluation zwar häufig verwendet werden, dass ihre testtheoretische Qualität aber nur selten überprüft wird. Das hängt sicherlich damit zusammen, dass medizindidaktische Maßnahmen meist für einen spezifischen Kontext geplant und durchgeführt werden – daher besteht das Bedürfnis, einen auf dieses Angebot zugeschnittenen Fragebogen einzusetzen, der selbst konstruiert und ohne weitere statistische Überprüfung verwendet wird. Erkauft wird damit jedoch, dass die Aussagekraft der so gewonnenen Ergebnisse nur schlecht überprüft werden kann. Auch der Vergleich verschiedener Programme wird dadurch erschwert, so dass sich kaum verallgemeinerbare Aussagen etwa zum Vergleich der Wirksamkeit bestimmter Elemente des Trainings gewinnen lassen. Vor diesem Hintergrund wäre es wünschenswert, wenn ein Kernbereich allgemeiner, testtheoretisch überprüfter Items zur Verfügung stünde, der an mehreren Standorten verwendet werden kann, um medizindidaktische Qualifizierungsangebote zu evaluieren. Für diesen Kernbereich sollte nach Möglichkeit, wie in der Lehrevaluation auch, mehrdimensional vorgegangen werden, damit überhaupt zwischen unterschiedlichen Faktoren differenziert werden kann, die den Erfolg des Angebots beeinflussen können (trainer-spezifischen, inhaltlichen, organisatorischen, usw.) (vgl.[9], [10]). Zusätzlich kann ein solcher Fragebogen um spezifische Fragen, die auch offen gestellt werden können, zum jeweiligen Angebot ergänzt werden, so dass sich die lokalen Besonderheiten ausreichend konkret abbilden lassen.

## Vorher-Nachher-Vergleiche

Ein wichtiger Aspekt der Wirksamkeitsüberprüfung medizindidaktischer Trainings sind Vorher-Nachher-Vergleiche z.B. von Einstellungen zu verschiedenen Aspekten der Lehre, Kenntnissen von Unterrichtsstrategien oder auch von lehrbezogenen Fertigkeiten. Werden dazu Befragungen der Teilnehmer eingesetzt, muss allerdings berücksichtigt werden, dass hier Verzerrungen auftreten können, weil die Einschätzung der eigenen Leistungsfähigkeit insbesondere dann unzuverlässig ist, wenn das entsprechende Kompetenzniveau insgesamt noch niedrig ist [11]. Umgekehrt wächst mit zunehmender Expertise auch die Fähigkeit, die eigene Kompetenz realistischer beurteilen zu können [12]. Das kann allerdings dazu führen,

dass Teilnehmer an medizindidaktischen Qualifizierungsmaßnahmen ihre Kenntnisse und Fähigkeiten nach der Maßnahme wegen (!) des Lernfortschritts schlechter einschätzen als zuvor. Möglicherweise lässt sich das Problem umgehen, indem die Teilnehmer erst retrospektiv nach einer Einschätzung ihres Kompetenzniveaus vor Kursbeginn gefragt werden. Jedenfalls zeigte sich in einer Studie von Skeff et al. [13], dass diese retrospektiv, also nach dem Kurs erhobene Selbsteinschätzung besser mit einer vor Kursbeginn durchgeführten studentischen Bewertung der Lehrkompetenz übereinstimmte als die vor dem Kurs abgegebene Selbsteinschätzung der Dozenten.

## Direkte Beobachtung/Messung von Lehrkompetenzen

Für die direkte Beobachtung bzw. Messung von Lehrkompetenzen können ebenfalls verschiedene Methoden eingesetzt werden. Es läge z.B. nahe, die Ergebnisse studentischer Evaluation heranzuziehen, auch wenn immer wieder Zweifel daran geäußert werden, ob Studierende in der Lage sind, die Qualität der Lehre beurteilen zu können. Unter der Voraussetzung einer adäquaten methodischen Qualität haben empirische Untersuchungen jedoch wiederholt gezeigt, dass die Übereinstimmung zwischen studentischen Urteilen und solchen von geschulten Fremd ratern recht hoch ist [14]. Die Korrelation liegt in der Größenordnung von  $r=.54$  [9]. Für die Beurteilung durch Studierende spricht zudem, dass sie die eigentliche Zielgruppe der Lehre sind, deren Bedürfnisse im Zentrum der Anstrengungen der Dozenten stehen sollten. Nicht zuletzt sind auch die organisatorischen Voraussetzungen günstig, weil die studentische Evaluation ohnehin durchgeführt wird und daher den Unterrichtsablauf nicht zusätzlich stört. In 15 (28%) der 53 Studien, die Steinert et al. [8] für ihre Übersichtsarbeit ausgewertet haben kommen studentische Bewertungen zum Einsatz, wobei in 11 Fällen eine Verbesserung im Vorher-Nachher-Vergleich festgestellt wurde. Bei den anderen Studien wurden keine statistisch signifikanten Unterschiede gefunden.

Eine weitere Möglichkeit der direkten Beobachtung ist die Beurteilung durch didaktisch geschulte und spezifisch trainierte Fremdbeobachter. Die didaktische Expertise ist hierbei besonders wichtig, weil bei einer Beurteilung durch Fachkollegen ansonsten leicht die Gefahr besteht, wissenschaftliche Kompetenz und Lehrkompetenz miteinander zu verwechseln. Ein weiteres Problem besteht darin, dass der Unterricht durch die Anwesenheit eines Beobachters beeinflusst oder gar gestört werden kann (manche klinische Unterrichtsformate lassen sich aus diesem Grund möglicherweise gar nicht durch Fremdbeurteiler evaluieren). Um solche Störungen zu minimieren, kann eine Videoaufzeichnung angefertigt und im Anschluss bewertet werden.

In den letzten Jahren wurden verschiedene Versuche publiziert, die Lehrkompetenz mit Hilfe von Simulationen zu prüfen, die sich am Modell der OSCE (Objective Structured Clinical Examination) orientieren (z.B. [8], [15],

[16], [17], [18]). Bei diesen OSTE (Objective Structured Teaching Examination/Evaluation) kommen meist „standardisierte Studierende“ zum Einsatz, die typische Lehrsituationen spielen (z.B. Interaktion mit schwierigen Studierenden, in Prüfungssituationen, am Krankenbett), auf welche die Lehrenden reagieren müssen: Bisher vorgestellte OSTEs umfassen zwischen fünf und acht Stationen, mit einer Dauer von 5 bis 15 min pro Station; die Gesamtdauer liegt bei bis zu 3,5 Stunden. Die Leistung wird analog zur OSCE mittels Checklisten oder globalen Schätzskaalen beurteilt. Eine fundierte Bewertung dieser Methode ist derzeit noch kaum möglich, da hierzu bislang nur wenige Studien mit sehr heterogener methodischer Qualität publiziert wurden.

## Lernerfolg als Maß für die Lehrqualität?

Die zunehmende Forderung nach Ergebnisorientierung in der Medizin könnte zu der Überlegung führen, den Lernerfolg der Studierenden, z.B. in Form von Prüfungsergebnissen, als Maß für die Verbesserung der Lehrqualität heranzuziehen. Dem liegt die zunächst plausible Annahme zugrunde, dass gute Lehre letztendlich auch zu guten Leistungen führen müsse. Wie eingangs bereits angedeutet wurde, sind die realen Umstände allerdings wesentlich komplexer: Oft stimmen Unterrichts- und Prüfungsinhalt nicht überein, oder es wird in Prüfungen vielfach nur Wissen geringen Elaborationsgrades (z.B. Reproduktion oder Wiedererkennen bei einfachen Multiple-Choice-Fragen) abgefragt. Dieser Mangelzustand besteht vor allem deshalb, weil solche Aufgaben mit geringerem Aufwand zu konstruieren sind. Viele Lehrende sehen aber gerade nach einer medizindidaktischen Sensibilisierung höherrangige kognitive Lernziele (Verständnis, Anwendung, etc.) oder gar affektive Lernziele (etwa die Veränderung von Einstellungen) als wichtige Ziele ihres Unterrichts an. Gute Lehre kann damit erheblich über gute Prüfungsergebnisse hinausgehen. Dabei kann schlimmstenfalls das durch die Lehrveranstaltung induzierte Lernen mit der zur Prüfungsvorbereitung erforderliche Lernstrategie auch im Widerspruch stehen.

Zudem kann in der Regel nicht kontrolliert werden, inwieweit das individuelle, selbständige Lernen der Studierenden, das zuhause oder in der Bibliothek stattfindet und für den Wissenserwerb an der Universität eine zentrale Rolle spielt, überhaupt durch die Lehre und die Dozenten beeinflusst wird. Vor dem Hintergrund dieser Überlegungen ist es umso erstaunlicher, dass dennoch in einigen Studien Zusammenhänge zwischen der individuellen Lehrqualität und Leistungsparametern gefunden wurden: So konnten Griffith et al. [19] zeigen, dass Studierende im dritten klinischen Jahr ( $N=502$  aus sechs aufeinanderfolgenden Kohorten) signifikant bessere Ergebnisse sowohl im fachbezogenen Abschlusstest (NBME)<sup>1</sup> als auch im späteren Examen (USMLE II)<sup>2</sup> hatten, wenn sie während ihrer klinischen Praxisphase mindestens vier Wochen von den besten Dozenten (bestimmt durch ein von den Untersuchten unabhängiges studentisches Expertenpanel sowie mittels studentischer Evaluation) unterrichtet wor-

den waren. Die Autoren vermuten, dass für diese Effekte weniger direkte Einflüsse, etwa durch eine bessere Stoffvermittlung, als vielmehr indirekte, vor allem motivationale Faktoren verantwortlich sind. Ganz ähnlich fanden Stern et al. [20], dass die Evaluationsergebnisse der Dozenten signifikante Prädiktoren für die Ergebnisse waren, welche die von ihnen unterrichteten Studierenden des dritten klinischen Jahres im fachbezogenen Abschlusstest (NBME) erreichten (a priori bestehende Leistungsunterschiede der Studierenden wurden dabei kontrolliert). Solche Ergebnisse dürfen allerdings nicht dazu verleiten, Prüfungsergebnisse unbesehen, ohne Berücksichtigung intervenierender Variablen (z.B. der jeweiligen Rahmenbedingungen oder dem zeitlichen Abstand zwischen der Veranstaltung und der entsprechenden Prüfung), einfach als einen Indikator der dozenten-spezifischen Lehrqualität zu verwenden.

## Ergebnisse der Evaluation

Insgesamt besteht eine hohe Zufriedenheit mit medizinisch-didaktischen Qualifizierungsangeboten [19]. Die Teilnehmer fühlen sich hinsichtlich lehrbezogener Aspekte sensibilisiert, motiviert und geben an, eine positivere Einstellung zur Lehre, aber auch zur lehrbezogenen Personalentwicklung gewonnen zu haben. Sowohl nach subjektiven als auch nach objektiven Maßstäben lässt sich darüber hinaus ein Zuwachs an medizindidaktischen Kenntnissen und unterrichtsbezogenen Fertigkeiten feststellen. Dem entsprechend stellen die Teilnehmer ein verändertes Verhalten im Unterricht bei sich selbst fest, das, wenn auch nicht in gleichem Ausmaß, auch von den Studierenden wahrgenommen wird. Schließlich haben medizindidaktische Qualifizierungsangebote auch für die Fakultätsentwicklung positive Konsequenzen: Die Teilnehmer berichten über ein verstärktes Engagement in der Lehre (z.B. durch innovative Lehrprojekte) und eine verbesserte lehrbezogene Zusammenarbeit mit Kollegen.

## Zusammenfassung

Die Evaluation medizindidaktischer Qualifizierungsangebote ist nicht nur im Hinblick auf die Erfolgskontrolle der durchgeführten Maßnahmen relevant, sondern kann darüber hinaus auch dazu beitragen, das Verständnis komplexer Lernprozesse zu vertiefen. Voraussetzung dafür ist allerdings eine hohe methodische Qualität. Insgesamt zeigt sich, dass das übergeordnete Ziel, die Lehrenden für medizindidaktische Fragen zu sensibilisieren und ihre Lehrkompetenz zu vergrößern, nicht nur erreicht wird, sondern dass sich darüber hinaus tatsächlich auch eine Verbesserung der Lehrqualität sowohl auf individueller als auch auf institutioneller Ebene feststellen lässt.

## Anmerkung

<sup>1</sup>NBME = National Board of Medical Examiners: Eine Organisation, die als Dienstleister medizinische Prüfungen organisiert und damit in etwa dem Institut für Medizinische und Pharmazeutische Prüfungsfragen (IMPP) in Deutschland vergleichbar ist.

<sup>2</sup>USMLE = United States Medical Licensing Examination: In etwa den deutschen Staatsexamina vergleichbar, Voraussetzung für die Zulassung zur ärztlichen Tätigkeit, insgesamt werden drei Prüfungen durchgeführt.

## Literatur

1. van der Vleuten CP, Dolmans DH, Scherpbier AJ. The need for evidence in education. *Med Teach.* 2000;22(3):246-250. DOI: 10.1080/01421590050006205
2. Harden RM, Grant J, Buckley G, Hart IR. BEME Guide No. 1: Best evidence medical education. *Med Teach.* 1999;21(6):553-562. DOI: 10.1080/01421599978960
3. Norman G. The effectiveness and the effects of effect sizes. *Adv Health Sci Educ Theory Pract.* 2003;8(3):183-187. DOI: 10.1023/A:1026090406201
4. Norman G. RCT= results confounded an trivial: the perils of grand educational experiments. *Med Educ.* 2003;37(7): 582-584. DOI: 10.1046/j.1365-2923.2003.01586.x
5. Norman G. Reflections on BEME. *Med Teach.* 2000;22(2):141-144. DOI: 10.1080/01421590078553
6. Kirkpatrick DL. Evaluation of training. In: Craig R, Mittel I (Hrsg). *Training and development handbook.* New York: McGraw-Hill; 1967. S.87-112.
7. Belfield C, Thomas H, Bullock A, Eynon R, Wall D. Measuring effectiveness for best evidence medical education: a disussion. *Med Teach.* 2001;23(2):164-170. DOI: 10.1080/0142150020031084
8. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach.* 2006;28(6):497-526. DOI: 10.1080/01421590600902976
9. Rindermann H. *Lehrevaluation. Einführung und Überblick zu Forschung und Praxis der Lehrveranstaltungsevaluation an Hochschulen mit einem Beitrag zu Evaluation computerbasierter Unterrichts.* Landau: Verlag Empirische Pädagogik; 2001.
10. Marsh HW. Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *Int J Educ Res.* 1987;11(3):253-388. DOI: 10.1016/0883-0355(87)90001-2
11. Kruger J, Dunning D. Unskilled an unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999;77(6):1121-1134. DOI: 10.1037/0022-3514.77.6.1121
12. Hodges B, Regehr G, Martin D. Difficulties in recognizing one's own incompetence: novice physicians who are unskilled and unaware of it. *Acad Med* 2001;76(10 Suppl):S87-S89. DOI: 10.1097/00001888-200110001-00029
13. Skeff KM, Stratos GA, Bergen MR. Evaluation of a medical faculty development program: a comparison of traditional pre/post and retrospective self-assessment ratings. *Eval Health Prof.* 1992;15(3):350-366. DOI: 10.1177/016327879201500307

14. Hofer M, Abanador N, Mödder U. Effektive Didaktiktrainings für Dozenten von CME-Fortbildungen (Continuing Medical Education). *Fortschr Röntgenstrahl bildgeb Verfahr.* 2005;177:1290-1296.
15. Morrison EH, Boker JR, Hollingshead J, Prislin MD, Hitchcock MA, Litzelman DK. Reliability and Validity of an objective structured teaching examination for generalist resident teachers. *Acad Med.* 2002;77(10):29-32. DOI: 10.1097/00001888-200210001-00010
16. Morrison EH, Rucker L, Boker JR, Hollingshead J, Hitchcock MA, Prislin MD, Hubbell FA. A pilot randomized, controlled trial of a longitudinal residents-as-teachers curriculum. *Acad Med.* 2003;78(7):722-729. DOI: 10.1097/00001888-200307000-00016
17. Dunnington GL, DaRosa D. A prospective randomized trial of a residents-as-teachers training program. *Acad Med.* 1998;73(6):696-700. DOI: 10.1097/00001888-199806000-00017
18. Prislin MD, Fitzpatrick C, Giglio M, Lie D, Radecki S. Initial experience with a multi-station objective structured teaching skills evaluation. *Acad Med.* 1998;73(10):1116-1118. DOI: 10.1097/00001888-199810000-00024
19. Griffith III CH, Georgesen JC, Wilson JF. Six-year documentation of the association between excellent clinical teaching and improved students' examination performances. *Acad Med.* 2000;75(10 Suppl):S62-S64. DOI: 10.1097/00001888-200010001-00020
20. Stern DT, Williams BC, Gill A, Gruppen LD, Wooliscroft JO, Grum CM. Is there a relationship between attending physicians' and residents' teaching skills and students' examination scores? *Acad Med.* 2000;75(11):1144-1146. DOI: 10.1097/00001888-200011000-00024
21. Srinivasan M, Litzelman D, Seshadri R, Lane K, Zhou W, Bogdevic S, Gaffney M, Galvin M, Mitchell G, Treadwell P, Willis L. Developing an OSTE to address lapses in learners professional behavior and an instrument to code educators responses. *Acad Med.* 2004;79(9):888-896. DOI: 10.1097/00001888-200409000-00017

**Korrespondenzadresse:**

Dr. med. Götz Fabry  
 Albert-Ludwigs-Universität Freiburg, Medizinische Fakultät,  
 Abteilung für Medizinische Psychologie und Soziologie,  
 Rheinstraße 12, 79104 Freiburg, Deutschland, Tel.:  
 0761/203-5512, Fax: 0761/203-5514  
 fabry@uni-freiburg.de

**Bitte zitieren als**

*Fabry G, Lammerding-Köppel M, Hofer M, Ochsendorf F, Schirlo C, Breckwoldt J. Hochschuldidaktische Qualifizierung in der Medizin IV: Messung von Wirksamkeit und Erfolg medizinischer Qualifizierungsangebote: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg. GMS Z Med Ausbild. 2010;27(4):Doc62. DOI: 10.3205/zma000699, URN: urn:nbn:de:0183-zma0006998*

**Artikel online frei zugänglich unter**

<http://www.egms.de/en/journals/zma/2010-27/zma000699.shtml>

**Eingereicht:** 13.01.2010

**Überarbeitet:** 26.03.2010

**Angenommen:** 26.03.2010

**Veröffentlicht:** 16.08.2010

**Copyright**

©2010 Fabry et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.