

Bundesweite Erhebung zur Lehrevaluation und Leistungsorientierten Mittelvergabe in der Lehre an den medizinischen Fakultäten in Deutschland

Zusammenfassung

Hintergrund: Bislang fehlen umfassende Daten zu Evaluationspraktiken und Leistungsorientierter Mittelvergabe (LOM) in der Lehre für die deutschen medizinischen Fakultäten. Vor diesem Hintergrund haben sich die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) und der Medizinische Fakultätentag (MFT) das gemeinsame Ziel gesetzt, die Praxis der Evaluationen und LOM in der Lehre (LOM-Lehre) an den medizinischen Fakultäten in Deutschland zu analysieren.

Methoden: Die Datenerhebung erfolgte mittels Fragebogen, der an alle medizinischen Fakultäten in Deutschland gesandt wurde.

Ergebnisse: An der Befragung nahmen 30 Fakultäten mit insgesamt 33 Studiengängen teil (Rücklauf: 83%). Die an den Fakultäten eingesetzten Erhebungsinstrumente erfassen vorrangig strukturelle und prozedurale Aspekte sowie einen Gesamteindruck der Lehre. Zwischen den Fakultäten herrscht bezüglich der verwendeten Instrumente eine recht hohe Heterogenität. Teilweise bleibt unklar, inwiefern die Erhebungsinstrumente internationalen Qualitätsstandards genügen. Die finanzielle Honorierung der Lehre erfolgt überwiegend im Rahmen der Grundausstattung bzw. nach Kriterien der Lehr-Quantität. Qualitätsbasierte Mittelzuweisung spielt eine eher untergeordnete Rolle.

Schlussfolgerung: Eine möglichst bundesweite Konsentierung eines Leitbilds guter Lehre sowie die Identifikation bzw. Entwicklung valider und reliabler Erhebungsinstrumente in deutschlandweiter Zusammenarbeit scheint erstrebenswert und würde eine Weiterentwicklung der gültigen LOM-Lehre darstellen.

Schlüsselwörter: Evaluation, Evaluationspraxis, Leistungsorientierte Mittelvergabe

Sarah

Schiekirka-Schwake¹

Janina Barth¹

Josef Pfeilschifter^{2,3}

Reinhard Hicel^{2,4}

Tobias Raupach^{1,5,6}

Christoph

Herrmann-Lingen^{7,8}

1 Studiendekanat,
Universitätsmedizin
Göttingen, Deutschland

2 Präsidium, Medizinischer
Fakultätentag, Berlin,
Deutschland

3 Dekanat des Fachbereichs
Medizin, Goethe-Universität
Frankfurt, Deutschland

4 Dekanat der Medizinischen
Fakultät, Ludwig-Maximilians-
Universität München,
Deutschland

5 Klinik für Kardiologie und
Pneumologie,
Universitätsmedizin
Göttingen, Deutschland

6 Health Behaviour Research
Centre, University College
London, Großbritannien

7 Klinik für Psychosomatische
Medizin und Psychotherapie,
Universitätsmedizin
Göttingen, Deutschland

8 Arbeitsgemeinschaft der
Wissenschaftlichen
Medizinischen
Fachgesellschaften e.V.,
Berlin, Deutschland

Einleitung

Im Studium der Humanmedizin werden an den 37 deutschen Fakultäten aktuell rund 92.000 Menschen zu künftigen Ärztinnen und Ärzten ausgebildet. Die Qualität des Studiums muss hohen Ansprüchen genügen: Neben der inhaltlichen Ebene [1] sind hierbei u.a. auch Aspekte der Interprofessionalität und der Wissenschaftlichkeit abzudecken [2]. Das „Ergebnis“ der Lehre an den medizinischen Fakultäten wird – auch zum Zweck eines nationalen Rankings – zuweilen an den studentischen Leistungen im schriftlichen Teil des Zweiten Staatsexamens abgeschätzt. Diese Prüfung bildet lediglich Faktenwissen ab und die aggregierten Examensleistungen bieten nur wenige Anhaltspunkte für konkrete Stärken und Schwächen einzelner Curricula. Somit benötigen die Fakultäten andere Datenquellen zur Beurteilung ihrer Lehrqualität. Hierzu werden in der Regel studentische Evaluationen durchgeführt. Hier können – je nach dem eingesetzten Erhebungsinstrument – bis zu vier Dimensionen der Lehrqualität abgebildet werden. Nach Gibson et al. werden diesbezüglich strukturelle und prozedurale Charakteristika der Lehre sowie die didaktischen Fertigkeiten der Lehrenden und der studentische Lernerfolg unterschieden [3]. Viele Evaluationsinstrumente legen ihren Fokus allerdings auf strukturelle und prozessbezogene Parameter [4] und erfragen in erster Linie die studentische Zufriedenheit mit der Lehre. Dabei ist bekannt, dass insbesondere Global-Bewertungen durch Studierende mannigfachen Störfaktoren (z.B. individuelle Charakteristika wie Geschlecht, Interesse am Thema, Leistungsniveau) unterliegen [5]. Eine solche Verzerrung der Evaluation ist problematisch, da die Ergebnisse an einigen Fakultäten als Grundlage einer Leistungsorientierten Mittelvergabe (LOM) für Lehre (LOM-Lehre) diskutiert werden. In der Forschung ist eine Leistungsorientierte Mittelzuweisung mittlerweile allgemein etabliert. Obwohl die Parameter und Algorithmen der Forschungsevaluation intensiv kritisiert werden [6], ist ein Ungleichgewicht zwischen Anreizen für gute Forschung und solchen für gute Lehre entstanden. Zwar fließen große Summen in die Grundausstattung der Lehre, die Lehrqualität wird bei der Ressourcen-Allokation jedoch häufig nicht ausreichend einbezogen. Dies hat zu Folge, dass sich in der Wahrnehmung vieler WissenschaftlerInnen Engagement in der Lehre weniger auszahlt als Engagement in der Forschung. Bislang fehlen umfassende Daten zu den Evaluationspraktiken und zur Gestaltung von LOM-Algorithmen in der Lehre für die deutschen medizinischen Fakultäten. Vor diesem Hintergrund haben sich die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF) und der Medizinische Fakultätentag (MFT) das gemeinsame Ziel gesetzt, die an den medizinischen Fakultäten in Deutschland gängigen Evaluationspraktiken zu analysieren. In dieser Arbeit werden die Ergebnisse einer an allen medizinischen Fakultäten Deutschlands durchgeführten Befragung vorgestellt.

Methoden

Es wurde ein standardisierter und im Vorfeld in einem mehrstufigen Verfahren pilotierter Fragebogen entwickelt. Der Fragebogen bildete in elf Abschnitten die folgenden Aspekte der Lehrevaluation und LOM-Vergabepaxis ab:

1. Grundlagen der Evaluation,
2. Gegenstände der Evaluation,
3. an der Evaluation beteiligte Personengruppen,
4. Regelmäßigkeit und Häufigkeit der Evaluation des Kerncurriculums,
5. Format der Evaluation durch Studierende,
6. Format der Evaluation durch Dozenten,
7. inhaltliche Aspekte der Evaluation durch Studierende,
8. Nutzung objektiver Daten zu Qualitätssicherung und Evaluation,
9. Aufbereitung und Verbreitung der Evaluationsergebnisse,
10. Konsequenzen der Evaluation,
11. Mittelvergabe für die Lehre.

Der Fragebogen enthielt vor allem Ja/Nein-Fragen, teilweise mit der Möglichkeit zu ergänzenden Freitextangaben. Vereinzelt wurden Zahlenangaben, etwa zum Umfang der Evaluationen, den Rücklaufquoten studentischer Evaluationen sowie zur Aufteilung der fakultären Landesmittel erbeten.

Der Fragebogen wurde im Juli 2013 über die Geschäftsstelle des Medizinischen Fakultätentages allen medizinischen Fakultäten in Deutschland zugesandt. Die Fakultäten wurden um schriftliche Beantwortung durch die zuständigen Mitarbeiter sowie die Zusendung relevanter Materialien gebeten. Bei Nichtbeantwortung erfolgte eine Erinnerung Ende August, weitere fehlende Daten wurden im November und Dezember telefonisch erfragt. Der Fragebogen kann bei den Autoren angefordert werden.

Ergebnisse

An der Befragung nahmen 30 deutsche medizinische Fakultäten mit insgesamt 33 Studiengängen (23 Regelstudiengänge, 6 Modellstudiengänge, 3 reformierte Regelstudiengänge, und 1 Studiengang für Molekulare Medizin) teil (Rücklaufquote 83%).

Die Ergebnisse werden im Folgenden orientiert an den Abschnitten des Fragebogens dargestellt.

Grundlagen der Evaluation

Lediglich 21% der Studiengänge verfügten über eine fakultätseigene Evaluationsordnung. Die Mehrzahl berief sich auf die Evaluationsordnung der jeweiligen Universität bzw. auf entsprechende Fachbereichsratsbeschlüsse. Der Großteil (85%) gibt seine Evaluationspraxis als „durch Erfahrung gewachsen“ an, bei 75% ist diese auch wissenschaftlich begründet. Als Quelle verweisen 40% auf externe und 36% auf interne Erhebungsinstrumente. Des

Weiteren werden vereinzelt Punkte wie einschlägige Kenntnisse in der Methodenforschung oder Standards für Evaluation der Gesellschaft für Evaluation e.V. (DeGEVal), der AG Evaluation und der Ausbildungskommission genannt.

Gegenstände der Evaluation und beteiligte Personengruppen

Hauptsächlich werden Fächer (70%), Einzelveranstaltungen (67%) und Studienabschnitte (64%) evaluiert. Insbesondere das Praktische Jahr (PJ) ist für 85% der Studiengänge evaluationsrelevant. Auch Lehrende werden an drei Vierteln der Fakultäten evaluiert. Für Prüfungen (Qualität und Ergebnisse) und AbsolventInnen erfolgt in etwa der Hälfte der Fakultäten eine Evaluation. Alle Studiengänge nutzen studentische Evaluationen, 45% beziehen Lehrende mit ein und 20–25% interne Gremien jedoch auch externe GutachterInnen.

Regelmäßigkeit und Häufigkeit der Evaluation im Kerncurriculum

In weniger als 50% der Studiengänge erfolgt eine flächendeckende Evaluation aller Veranstaltungstypen und Personengruppen. Lediglich Fächer bzw. Querschnittsbereiche werden von etwa zwei Dritteln der Fakultäten durchgängig evaluiert, einzelne Studienabschnitte von 42%, und Lehrende (27% bis 36%) sowie Prüfungen (33%) von ca. einem Drittel.

Format der Evaluationen

Nahezu alle Studiengänge (97%) nutzen bei studentischen Evaluationen Online-Formate, etwa 50% evaluieren (zusätzlich) papierbasiert. Mehrheitlich (94%) wird außerhalb von Lehrveranstaltungen, vorwiegend vor oder nach der Abschlussklausur, evaluiert. Aber auch während laufender Veranstaltungen finden in ca. 46% der Studiengänge Evaluationen statt. Die Fragenanzahl im Fragebogen variiert stark (1 bis 140 Items). Am häufigsten kommen ordinal- bzw. intervallskalierte Items (z.B. Schulnoten, %-Angaben) sowie Freitextkommentare zum Einsatz. Auch werden dichotome, offene und Mehrfachantwort-Fragen verwendet.

Die Rücklaufquoten studentischer Evaluationen wurden im Mittel mit ungefähr 60% angegeben. Anreize für einen erhöhten Rücklauf wie Transparenz (18%) und Bonusysteme für Studierende (33%) werden ebenso wie Zwang und negative Konsequenzen nur vereinzelt als hilfreich beschrieben. Als weitere Evaluationsformen wurden Nachbesprechungen (54%) und Berichte von SemestersprecherInnen (48%) genannt. Interviews (15%) und Fokusgruppen (21%) kommen seltener zum Einsatz.

Bei der Lehrveranstaltungsevaluation durch Dozierende wird in der Mehrzahl (80%) ein strukturiertes Vorgehen gewählt. In 40% der Fälle evaluieren alle Lehrenden, in 20% eine Auswahl von Lehrenden in Kommissionen.

Inhaltliche Aspekte der Evaluation durch Studierende

Mittels studentischer Evaluationen werden vor allem Struktur- und Prozessparameter sowie der Gesamteindruck der Lehre erfasst. Besonders hoch (>80%) schätzen Studierende subjektiv die Prüfungs- und Praxisrelevanz der Inhalte ein, bewerten den inhaltlichen Aufbau und geben ihre globale Zufriedenheit mit der Veranstaltung an bzw. eine Gesamtnote ab. Für weitere Details siehe Tabelle 1.

Nutzung objektiver Daten zur Qualitätssicherung und Evaluation

Von 94% der Fakultäten werden objektive Daten zur internen oder externen Qualitätssicherung der Lehre genutzt. Besonders häufig (>80%) wurde das Zweite Staatsexamen (Bestehensquote in der Referenzkohorte, Durchfallquote) genannt, gefolgt von durchschnittlicher Studiendauer, durchschnittlicher Punktzahl im Zweiten Staatsexamen und Absolventenzahlen (je 79%), Bestehen des Ersten Staatsexamens und Weiterqualifikationen der Lehrenden (je 76%).

Betreuungsrelation, Zahl und Qualität von Promotionen sowie Forschungsprojekte bzw. Publikationen zur Lehre, aber auch durchschnittliche Punktzahlen und fächer-spezifische Auswertungen im Ersten Staatsexamen sind weniger relevante Parameter für die Qualitätssicherung (50–70%)

Aufbereitung und Verbreitung der Evaluationsergebnisse

Die Evaluationsergebnisse werden überwiegend als schriftliche Berichte verfasst und meist elektronisch in einem geschützten Bereich abgelegt. In der Regel werden Dekanat (79%) und Scheinverantwortliche (76%) aktiv über die Ergebnisse informiert, häufig auch LehrkoordinatorInnen (67%), einzelne Dozierende (61%) und Studierende (64%). Eine aktive Ergebnis-Kommunikation erfolgt in gut 40%, und in knapp 10% der Fälle wird aktiv die (externe) Öffentlichkeit informiert.

Konsequenzen der Evaluation

Die Bewertung der Lehrqualität erfolgt in zwei von drei Fakultäten anhand fester Kriterien/Kategorien, bei 82% werden Veranstaltungen und Dozierende auch relativ zueinander bewertet. Mehrheitlich (82%) werden Feedbackgespräche mit Dozierenden geführt, seltener wird Feedback zu Lehrinhalten an Querschnittsbereiche (73%), Fächer (61%) und Module (55%) gegeben. Zudem haben knapp 80% (79%) der Studiengänge Konsequenzen für besonders positive aber auch für besonders negative Evaluationsergebnisse vorgesehen. Für positive Ergebnisse wird vorwiegend (54%) die Berücksichtigung in der LOM-Lehre genannt und nur vereinzelt

Tabelle 1: Aspekte, die mittels Lehrveranstaltungsevaluation durch Studierenden erfasst werden

Dimension	Inhaltliche Aspekte	Angaben in %
Struktur	Äußere Rahmenbedingungen	76
	Zeitliche Rahmenbedingung	73
	Eingliederung ins Curriculum	58
	Verfügbarkeit der Stundenpläne	27
	Prüfungs- und Praxisrelevanz der Veranstaltungsinhalte	82
	Wissenschaftliche Relevanz	30
	Prüfungen (Format, Ergebniskommunikation)	39
	Zufriedenheit der Studierenden mit der Veranstaltung insgesamt	94
Prozess	Inhaltlicher Aufbau des Unterrichts	94
	Prüfungsqualität	27
	Feedbackschleifen	30
	Zeitaufwand	46
	Schwierigkeitsgrad	49
Dozent	Globale Aussagen/Freitext	85
	Sprache	30
	Didaktisches Geschick, Lehrkompetenz	70
	Unterrichtsvorbereitung	30
	Begeisterungs- und Motivationsfähigkeit	49
	Interaktion mit Studierenden	49
	Erreichbarkeit	21
	Sprechstunden	9
Ergebnis	Subjektiver Lernerfolg	76
	Notwendigkeit im zukünftigen Beruf	58
Gesamteindruck	Zufriedenheit mit der Veranstaltung insgesamt	94
	Globalnoten für Veranstaltungen	88

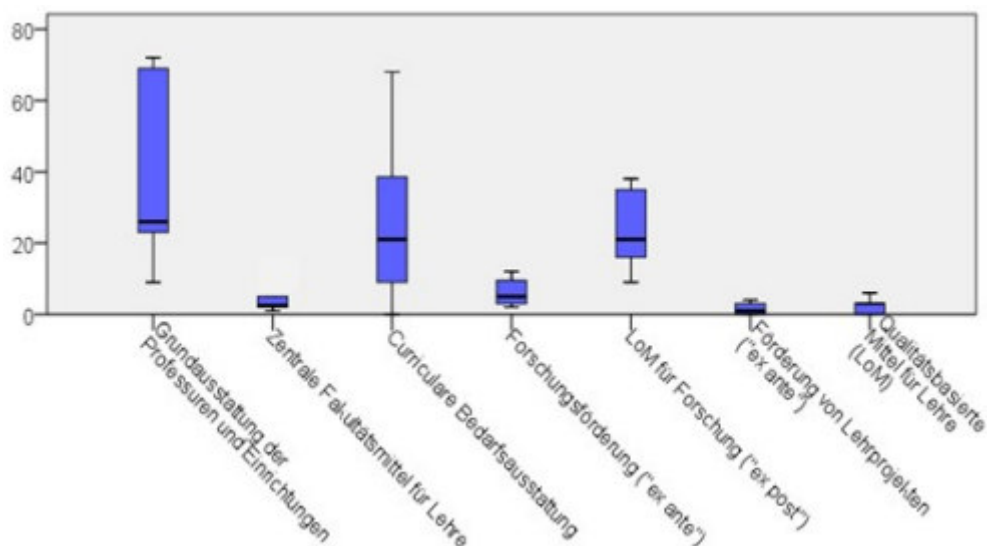


Abbildung 1: Prozentuale Verteilung der Landeszuführungsbeiträge an den Fakultäten

Auszeichnungen als beste Dozentin bzw. bester Dozent oder ein Bonus bei Beförderungen erwähnt. Dagegen stehen bei 73% der Studiengänge Schulungs- und Unterstützungsangebote für Lehrende mit besonders negativen Ergebnissen zur Verfügung, gefolgt von persönlichen Gesprächen (45%) und Effekten auf die LOM-Lehre (21%).

Mittelvergabe für Lehre

Für die Verteilung der Landes-Zuführungsbeiträge liegen Antworten von 17 Fakultäten vor. Diese Daten sind in Abbildung 1 dargestellt. Deutlich wird hier, dass die drei wichtigsten Posten die allgemeine Grundausstattung der Lehrstühle sowie die curriculare Basisausstattung und die evaluationsbasierte Forschungs-LOM darstellen. Demgegenüber spielen diese Mittel für die Lehre ebenso eine untergeordnete Rolle wie antragsbasierte Forschungs-

und Lehrförderung. Die evaluationsbasierte LOM-Lehre bildet mit durchschnittlich 1,6% (Minimum 0%, Median 3%, Maximum 6%) eher einen kleinen Teil der Landesmittel ab.

Die LOM-Lehre folgt in gut 60% einem festen Algorithmus. Sie kommt mit knapp 70% insbesondere den Kliniken bzw. Instituten zugute, deutlich seltener kompletten Modulen (15%) oder Einzelpersonen (21%).

Diskussion

Die Ergebnisse der Fakultätenbefragung decken sich insofern mit der aktuellen Literatur [4], als die in Deutschland eingesetzten Erhebungsinstrumente vorrangig strukturelle und prozedurale Aspekte sowie einen Gesamteindruck der Lehre erfassen.

Das Ergebnis der Lehre – i.d.R. als Lernerfolg der Studierenden definiert – wird entweder anhand punktueller studentischer Angaben zum wahrgenommenen eigenen Lernerfolg oder anhand der Leistungen in den Staatsexamina beurteilt. Eine systematische Bewertung der Lehrenden erfolgt kaum.

Für die Dimensionen „Prozess“ und „Lernende“ wurden in der Literaturrecherche von Schiekirka et al. [4] bereits zahlreiche Erhebungsinstrumente für die medizinische Lehre mit guter bis sehr guter Reliabilität identifiziert. Etwas mehr als ein Drittel der Fakultäten geben an, externe Instrumente für die Evaluation heranzuziehen bzw. diese als wissenschaftliche Grundlage für fakultätseigene Instrumente zu nutzen. Es bleibt jedoch unklar, ob die genutzten Instrumente sich mit den bereits identifizierten Instrumenten decken. Auch bleibt offen, ob die Instrumente spezifisch zur Erfassung der Lehre in der Medizin entwickelt wurden, ob ihnen psychometrische Testungen zugrunde liegen und inwiefern sie den gängigen Gütekriterien genügen. Bei 88% der Fakultäten werden Globalnoten zur Bewertung der Lehrqualität herangezogen. Wenngleich diese Bewertungen einen groben Eindruck von der studentischen Zufriedenheit mit der Lehre vermitteln, sind sie aufgrund der starken Verzerrung durch verschiedene konstrukt-irrelevante Störfaktoren nicht als valide Messgrößen der Lehrqualität anzusehen [5]. So zeigen Daten, dass Studierende mit einem großen initialen Interesse an einer Lehrveranstaltung in der Evaluation positivere Einschätzungen angeben als Personen mit einem geringen Interesse [7], [8]. Weitere Studien fanden eine positive Korrelation zwischen Leistungen in der Abschlussprüfung und der Bewertung eines Anatomiekurses [9], [10]. In diesem Zusammenhang sei auf die Bedeutung einer klaren Definition des in der Evaluation abgebildeten Konstrukts von guter Lehre hingewiesen: Nur wenn eindeutig definiert ist, was unter „guter Lehre“ verstanden wird, kann ein Instrument identifiziert und eingesetzt werden, das genau dieses Konstrukt misst. Umgekehrt darf sich die Interpretation vorliegender Evaluationsdaten nur auf genau dieses Konstrukt (z.B. strukturelle Gegebenheiten) beziehen, und nicht auf andere Aspekte der

Lehrqualität (z.B. didaktisches Geschick der Dozenten) ausgeweitet werden.

Die psychometrische Untersuchung eingesetzter Evaluationsinstrumente, Minimierung verzerrender Effekte sowie die Erfassung aller vier Dimensionen nach Gibson [3] sind jedoch dringend erforderlich, um die Lehrqualität valide erfassen und entsprechend optimieren zu können. Auch die Erfassung des Lehrergebnisses durch scheinbar objektive Daten wie Prüfungen unterliegt Schwierigkeiten. Um valide Daten liefern zu können, müssen Prüfungen kongruent mit dem Lehrinhalt und der Lehrform [11] und somit frei von konstrukt-irrelevanter Varianz sein [12], [13] sowie internationalen Qualitätsstandards genügen [14]. Nicht immer ist dies gegeben [15].

Noch mehr Gewicht bekommen diese Voraussetzungen, wenn die Verteilung von LOM-Lehre auf Basis von Evaluationsergebnissen und Prüfungen erfolgen soll. Die finanzielle Honorierung der Lehre erfolgt bislang überwiegend im Rahmen der Grundausrüstung bzw. nach Kriterien der Lehr-Quantität. Qualitätsbasierte Mittelzuweisung für die Lehre spielt anders als im Bereich der Forschung nur eine untergeordnete Rolle. Auch weitere Gratifikationen für hohe Lehrqualität kommen lediglich vereinzelt zum Einsatz.

Durch die derzeitige Heterogenität der Evaluation ist die Vergleichbarkeit der Ergebnisse zwischen Fakultäten nicht oder nur sehr eingeschränkt möglich. Durch eine Betrachtung schriftlicher Examensergebnisse kann zwar eine gewisse Vergleichbarkeit hergestellt werden – hier bleibt jedoch jeweils die Validität der Daten vor dem Hintergrund des jeweils gewählten Konstruktes guter Lehre zu diskutieren. Hinsichtlich einer Honorierung guter Lehre zeigen erste Daten, dass Lehrende derzeit ihre Motivation vorrangig aus studentischem Feedback ziehen, jedoch möglichen finanziellen Anreizen positiv gegenüber stehen [16], [17]. So scheint es nicht zielführend zu sein, ausschließlich finanzielle Anreize in den Vordergrund zu stellen, vielmehr können auch weitere Aspekte der Wertschätzung und konkrete Unterstützung (durch z.B. Verbesserung von Rahmenbedingungen und Karrierechancen) motivationsfördernd und -erhaltend wirken [18]. Ebenso wie gute Forschungsleistung sowohl immaterielle und karrierebezogene als auch direkte materielle Wertschätzung erfährt, sollte ein Anreizsystem für gute Lehre auf breiter Basis ansetzen, um das bislang bestehende Ungleichgewicht auszugleichen.

Fazit

Diese erste systematische Erhebung zu den Evaluationspraktiken in der Lehre an deutschen medizinischen Fakultäten hat gezeigt, dass die Evaluationsinstrumente inhaltlich und methodisch ähnlich ausgelegt sind und vor allem strukturelle und prozedurale Aspekte sowie einen Gesamteindruck der Lehre erfassen. Bezüglich der tatsächlich eingesetzten Instrumente herrscht eine erhebliche Heterogenität. Eine möglichst bundesweite Konsentierung eines Leitbilds guter medizinischer Lehre sowie

die Identifikation bzw. Entwicklung valider und reliabler Erhebungsinstrumente in deutschlandweiter Zusammenarbeit scheint sinnvoll.

Anmerkungen

Interessenkonflikte

Die Autoren erklären, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel haben.

Autorenschaft

Die Autoren Raupach T und Herrmann-Lingen C haben gleichermaßen zu der Arbeit beigetragen.

Danksagungen

Wir danken der AWMF für finanzielle Unterstützung, der MFT-Geschäftsstelle (Dr. Corinne Dölling) für logistische Unterstützung, Frau Sabine Gluth für die Komplettierung und Auswertung der Befragungsdaten, den Mitarbeiterinnen und Mitarbeitern der Studiendekanate für die Beantwortung der Fragebögen sowie den Mitgliedern der AWMF-Kommission für Leistungsevaluation in Forschung und Lehre und der MFT-AG Lehre für ihre Unterstützung bei der Durchführung und Interpretation der Befragung.

Literatur

- Hahn EG, Fischer MR. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). *GMS Z Med Ausbild.* 2009;26(3):Doc35. DOI: 10.3205/zma000627
- Wissenschaftsrat. Empfehlungen zur Weiterentwicklung des Medizinstudiums in Deutschland auf Grundlage einer Bestandsaufnahme der humanmedizinischen Modellstudiengänge. Dresden: Wissenschaftsrat; 2014.
- Gibson KA, Boyle P, Black DA, Cunningham M, Grimm MC, McNeil HP. Enhancing Evaluation in an Undergraduate Medical Education Program. *Acad Med.* 2008;83(8):787-93. DOI: 10.1097/ACM.0b013e31817eb8ab
- Schiekirka S, Feufel MA, Herrmann-Lingen C, Raupach T. Evaluation in medical education: A topical review of target parameters, data collection tools and confounding factors. *GMS Ger Med Sci.* 2015 Sep 16;13:Doc15. DOI: 10.3205/000219
- Herrmann-Lingen C, Brunner E, Hildenbrand S, Loew TH, Raupach T, Spies C, Treede RD, Vahl CF, Wenz HJ. Evaluation of medical research performance - position paper of the Association of the Scientific Medical Societies in Germany (AWMF). *GMS Ger Med Sci.* 2014;12:Doc11. DOI: 10.3205/000196
- Schiekirka S, Raupach T. A systematic review of factors influencing student ratings in undergraduate medical education course evaluations. *BMC Med Educ.* 2015 Mar 5;15:30. DOI: 10.1186/s12909-015-0311-8
- Berger U, Schleussner C, Strauss B. [Comprehensive evaluation of medical teaching – a task for the psychosocial disciplines?]. *Psychother Psychosom Med Psychol.* 2003 Feb;53(2):71-8. DOI: 10.1055/s-2003-36966
- Raupach T, Schiekirka S, Münscher C, Beißbarth T, Himmel W, Burckhardt G, Pukrop T. Piloting an outcome-based programme evaluation tool in undergraduate medical education. *GMS Z Med Ausbild.* 2012;29(3):Doc44. DOI: 10.3205/zma000814
- Canaday SD, Mendelson MA, Hardin JH. The effect of timing on the validity of student ratings. *J Med Educ.* 1978 Dec;53(12):958-64. DOI: 10.1097/00001888-197812000-00002
- Mendelson MA, Canaday SD, Hardin JH. The relationship between student ratings of course effectiveness and student achievement. *Med Educ.* 1978 May;12(3):199-204. DOI: 10.1111/j.1365-2923.1978.tb00337.x
- Kern DE, Thomas PA, Howard DM, Bass EB. Curriculum development for medical education – A six-step approach. Baltimore, London: The John Hopkins University Press; 1998.
- Downing SM. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Acad Med.* 2002 Oct;77(10 Suppl):S103-4. DOI: 10.1097/00001888-200210001-00032
- Haladyna TM, Downing SM. Construct-Irrelevant Variance in High-Stakes Testing. *Educ Meas.* 2005;23(1):17-27. DOI: 10.1111/j.1745-3992.2004.tb00149.x
- Möltner A, Schellberg D, Jünger J. Basic quantitative analyses of medical examinations. *GMS Z Med Ausbild.* 2006;23(3):Doc53.
- Möltner A, Duelli R, Resch F, Schultz J, Jünger J. School-specific assessment in German medical schools. *GMS Z Med Ausbild.* 2010;27(3):Doc44. DOI: 10.3205/zma000681
- Müller-Hilke B. A qualitative study to investigate the acceptance of performance-based allocation of resources for the improvement of education at German medical schools. *GMS Z Med Ausbild.* 2010;27(3):Doc43. DOI: 10.3205/zma000680
- Hofer M, Pieper M, Sadlo M, Reipen J, Heussen N. [Performance-related middle management in medical teaching. Attractiveness of incentive tools from the perspective of the teachers]. *Dtsch Med Wochenschr.* 2008 Aug;133(31-32):1615-20. DOI: 10.1055/s-2008-1081141
- Müller-Hilke B. Leistungsorientierte Mittelvergabe: Mehr Geld zur Lehre lenken. *Dtsch Arztebl.* 2013;110(50):A 2418-20.

Korrespondenzadresse:

Sarah Schiekirka-Schwake
Studiendekanat, Universitätsmedizin Göttingen,
Humboldtallee 38, 37073 Göttingen, Deutschland
sarah.schiekirka-schwake@med.uni-goettingen.de

Bitte zitieren als

Schiekirka-Schwake S, Barth J, Pfeilschifter J, Hickel R, Raupach T, Herrmann-Lingen C. National survey of evaluation practices and performance-guided resource allocation at German medical schools. *GMS Ger Med Sci.* 2019;17:Doc04. DOI: 10.3205/000270, URN: urn:nbn:de:0183-0002708

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/gms/2019-17/000270.shtml>

Eingereicht: 02.08.2018

Überarbeitet: 24.01.2019

Veröffentlicht: 18.04.2019

Copyright

©2019 Schiekirka-Schwake et al. Dieser Artikel ist ein
Open-Access-Artikel und steht unter den Lizenzbedingungen der
Creative Commons Attribution 4.0 License (Namensnennung).
Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.