

## » Ergebnisse der IMPP-Prüfungen sind keine geeignete Grundlage für eine jährliche leistungsbezogene Mittelzuweisung an deutschen medizinischen Fakultäten

B.-P. Robra<sup>1,2</sup>, H. Schmitt<sup>2</sup>

<sup>1</sup> Studiendekanat

<sup>2</sup> Institut für Sozialmedizin und Gesundheitsökonomie der Medizinischen Fakultät der Otto-von-Guericke-Universität Magdeburg

**Zusammenfassung:** Die Evaluation der Lehre gewinnt Bedeutung als Output-Größe im Rahmen der leistungsorientierten Mittelvergabe an den deutschen medizinischen Fakultäten. Die bundeseinheitlich erhobenen Prüfungsergebnisse des IMPP sind mögliche Indikatoren des Lehrerfolgs, vorausgesetzt die Ergebnisse der Fächer sind voneinander unabhängig und die Ergebnisse der Fakultäten untereinander hinlänglich distinkt. Eine Überprüfung der Ergebnisse von sechs akademischen Jahren (Herbstprüfung 1994 bis Frühjahrprüfung 2000) mit Hilfe von Rangkorrelationen und fächerspezifischen Regressionsanalysen zeigt, dass beide Voraussetzungen nicht erfüllt sind. Die Fächer des Physikums korrelieren untereinander konsistent positiv, ebenso die Fächer des ersten Teils der Ärztlichen Prüfung. Zwar sind Ergebnisse einzelner Fakultäten signifikant unterschiedlich von denen anderer Fakultäten. Dies ist allerdings selbst dann wenig trennscharf, wenn Daten des ganzen Berichtszeitraums einbezogen werden. Bei den Physiologieergebnissen z. B. ist der Gradient der Fakultäten so flach, dass sich überhaupt keine Fakultät signifikant von irgendeiner anderen unterscheidet. Eine Prämierung jährlicher fachspezifischer IMPP-Ergebnisse im Regelwerk der fakultätsinternen leistungsorientierten Mittelvergabe kann nach den vorliegenden Daten nicht sachgerecht sein. Alternative zum externen Benchmarking durch IMPP-Ergebnisse könnten Elemente eines internen kontinuierlichen Verbesserungsprozesses sein.

**Schlüsselwörter:** Lehrevaluation – leistungsbezogene Mittelvergabe – IMPP-Prüfungsergebnisse

**Results of the IMPP Exams are not a Suitable Basis for an Annual Performance-orientated Allocation of Funds at German Medical Faculties:** Evaluation of teaching is becoming increasingly important as an output factor within the framework of performance-orientated allocation of funds at German medical schools. Results of the federal uniform IMPP exams are potential indicators of teaching success, provided that results of subjects are independent of each other and that results of different faculties are sufficiently distinct. A survey of the results from six academic years (autumn exams 1994 until spring exams 2000) using rank correlations and subject-specific regression analyses

shows that none of these requirements are fulfilled. The subjects of the pre-medical examinations are consistently positively correlated, as well as the subjects of the first section of the medical finals. Although results of some faculties differ significantly from those of others, these findings are hardly distinctive overall even when data from the whole reporting period are analysed. Concerning the results in physiology, for example, the gradient of the faculties is so flat that no faculty differs significantly from any other. Fixing a premium for annual subject-specific IMPP results within the framework of an internal performance-orientated allocation of funds in faculties is not justified on the basis of the available data. An alternative to external benchmarking procedures via IMPP results might be an internal process of ongoing quality improvement within the faculties.

**Key words:** Performance-oriented allocation of funds – IMPP exams – Evaluation of teaching

### Einführung

Die Evaluation der Lehre an den medizinischen Fakultäten soll erstens einen Anreiz schaffen, die als defizitär geltende akademische Ausbildung zu verbessern. Dazu brauchen wir Indikatoren des Lehrerfolgs und Kenntnis seiner Determinanten. Indikatoren des Lehrerfolgs reichen von subjektiven Einschätzungen der Studierenden und der Lehrenden über die Feststellung von Tatsachen-, Zusammenhangs- und Orientierungswissen bis zur erfolgreichen Lösung praxisnah gestellter Aufgaben.

Zweitens erlangt die Lehre auch zunehmende Bedeutung als Output-Größe im Rahmen der leistungsorientierten Mittelvergabe an den und zwischen den Fakultäten. Das Hochschulrahmengesetz fordert eine Orientierung der staatlichen Hochschulfinanzierung u. a. an den in der Lehre erbrachten Leistungen<sup>1</sup>. Dafür brauchen wir weitergehende Informationen über

<sup>1</sup> HRG § 5 Staatliche Finanzierung. Die staatliche Finanzierung der Hochschulen orientiert sich an den in Forschung und Lehre sowie bei der Förderung des wissenschaftlichen Nachwuchses erbrachten Leistungen. Dabei sind auch Fortschritte bei der Erfüllung des Gleichstellungsauftrages zu berücksichtigen.

die Lehrleistung, das heißt eine Vorstellung darüber, mit welchen Mitteln oder Inputs welcher Lehrerfolg erreicht wurde.

Realisierte Evaluationsmethoden an deutschen medizinischen Fakultäten reichen von Befragungen der Studierenden oder der Lehrenden bis zur Nutzung der Ergebnisse des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) in Mainz [5,6].

Die schriftlichen Prüfungen des IMPP haben für eine vergleichende Evaluation erhebliche praktische Vorzüge. Sie werden mit hoher Kompetenz vorbereitet, bundeseinheitlich durchgeführt und fakultätsübergreifend ausgewertet. Ihre Ergebnisse werden in einigem Detail zur Verfügung gestellt. Soweit sie den im Medizinstudium angestrebten Lehrerfolg messen – wenn dies gar nicht der Fall wäre, sollten wir so nicht prüfen – können sie für eine Evaluation des Lehrerfolgs und der Lehrleistung der medizinischen Fakultäten kaum unberücksichtigt bleiben.

Als Grundlage einer leistungsorientierten Mittelzuweisung innerhalb der medizinischen Fakultäten sollten die IMPP-Ergebnisse der Fächer voneinander unabhängig und die Ergebnisse der Fakultäten untereinander hinlänglich distinkt sein. Mit Blick auf diese beiden Forderungen prüft dieser Beitrag die Eignung der IMPP-Ergebnisse als Maßzahlen für eine leistungsorientierte Mittelallokation.

### Material und Methoden

Das IMPP stellt die Ergebnisse der halbjährlichen Prüfungen zur Verfügung, disaggregiert nach Fakultäten, Art der Prüfung, Prüfungsfächern und – nur die Gesamtergebnisse – nach Zugehörigkeit zur plantreu studierenden so genannten Referenzgruppe. Wir konnten die Ergebnisse ab Herbst 1994 bis Frühjahr 2000 auswerten<sup>2</sup>.

Zur Referenzgruppe gehören die Studierenden, die ihre Prüfungen innerhalb der Mindeststudienzeit und zum ersten Mal ablegen. Wer einmal die Referenzgruppe aus irgendwelchen Gründen verlassen musste, wird nicht wieder dazugehört. Die Prüfungsergebnisse der Studierenden in der Referenzgruppe sind nach konstanter Erfahrung deutlich besser als die der Nichtreferenzgruppe, unabhängig davon ob man die erreichte Punktzahl oder die Misserfolgsquote betrachtet. Auch die Nichtreferenzgruppe wird aber unterrichtet. Ihr Anteil steigt in den von uns betrachteten Daten von 43,6% in der Vorprüfung auf 76,2% in der zweiten Ärztlichen Prüfung. Es wäre also falsch, evaluative Betrachtungen nur auf die Teilnehmer der Referenzgruppe zu beschränken.

Eine Mehrheit der Fakultäten lässt den Studienbeginn nur zum Herbst zu. Daher treten Studierende der Referenzgruppe an diesen Fakultäten (abgesehen von Quereinsteigern) de facto nur in den Herbstprüfungen an, während Frühjahrprüfungen vor allem Nachzügler betreffen. Auch von der Schule her sind die Ausbildungszyklen auf einen Studienbeginn im Herbst hin angelegt. Eine alleinige Berücksichtigung der Herbstprüfungen ist also verzerrt. Wir haben daher jeweils die Herbstprüfung

mit der nachfolgenden Frühjahrprüfung zum Ergebnis eines akademischen Jahres zusammengefasst. Diese Zusammenfassung der Prüfungstermine scheint uns dem betrachteten Zweck besser zu entsprechen als eine Zusammenfassung nach Kalenderjahren, die das IMPP mitteilt (im Fließgleichgewicht sollten die Unterschiede gering sein).

Eine neutrale Beurteilung aller Prüfungen trotz unterschiedlich großer Referenzgruppenanteile wird durch einen adjustierten Jahresindex (AJAX) erreicht. Er berücksichtigt den Anteil der Studierenden in der Referenzgruppe  $\rho_{i,t}$  und in ihrer Komplementärgruppe  $(1 - \rho_{i,t})$  explizit, und zwar mittels indirekter Standardisierung der jährlichen Ergebnisse der 36 Fakultäten ( $\mu_{i,t}$ , d. h. Anteil gelöster Aufgaben) auf das gleichzeitige Bundesergebnis, das man erwarten würde, wenn bei gegebenem Anteil der Referenzgruppe in der jeweiligen Fakultät der Bundesdurchschnitt der jeweiligen Gruppe erreicht worden wäre:

$$AJAX(i,t) := \frac{\mu_{i,t}}{E\mu_{i,t}} = \frac{\rho_{i,t} \cdot \tilde{\mu}_{i,t} + (1 - \rho_{i,t}) \cdot \bar{\mu}_{i,t}}{\rho_{i,t} \cdot \tilde{\mu}_{0,t} + (1 - \rho_{i,t}) \cdot \bar{\mu}_{0,t}}$$

wobei Index  $t$  das jeweilige Jahr kennzeichnet, Index  $i$  die Fakultäten,  $0$  das Ergebnis auf Bundesebene, die Tilde die Referenzgruppe und der Überstrich die Nichtreferenzgruppe. Der von der Fakultät  $i$  im Jahr  $t$  erreichte Anteil der gelösten Aufgaben ergibt sich als gewichteter Mittelwert der Ergebnisse der Teilnehmer in der Referenz- und in der Nicht-Referenzgruppe im Zähler des Bruches.  $E\mu_{i,t}$  ist der nach den Gruppenergebnissen auf Bundesebene erwartete Wert in der Fakultät  $i$  im Jahr  $t$  (für eine Ableitung aus den IMPP-Daten, s. [4]). Der frei verfügbaren Datenlage nach kann der gruppenadjustierte Wert nur für die Gesamtergebnisse errechnet werden. Fächerspezifische Jahresergebnisse müssen unkorrigiert betrachtet werden.

Die Ergebnisse der Fakultäten im Berichtszeitraum werden mit Rangkorrelationen und mit Hilfe einer Regressionsgleichung verglichen, in die Fakultätsdummies (0,1) und eine Gewichtung mit der Zahl der Prüfungsteilnehmer eingehen. Für die Gesamtergebnisse und die Ergebnisse der einzelnen Fächer werden je spezifische Gleichungen geschätzt (SPSS für Windows 10.0.7).

### Ergebnisse

Abb. 1 zeigt die Physikumsergebnisse der Fakultäten (alle Fächer) für die zusammengefasste Berichtsperiode 1994–2000. Die Anteile richtig beantworteter Fragen reichen im Mittel der Beobachtungsjahre von 66,4% in Witten (wo allerdings nur für vier Jahre Werte für die Herbst- und die Frühjahrprüfungen vorliegen) bis 58,3% in Frankfurt. Der Gradient ist flach. Die 95%-Konfidenzintervalle der Fakultätsparameter signalisieren, dass die einzelnen Fakultäten von einer breiten Gruppe ihrer jeweiligen Nachbarn nicht signifikant zu unterscheiden sind.

Die durch den AJAX definierten Erfolgsquoten liegen während der Jahre 1994–2000 für die Fakultäten mit wenigen Ausnahmen im Intervall zwischen 0,95 und 1,05. Die geringe Streubreite spricht für eine hohe Standardisierung der medizinischen akademischen Lehre, soweit sie über schriftliche Prü-

<sup>2</sup> Wir danken dem IMPP, Mainz, für die Überlassung der Daten, soweit sie nicht im Internet verfügbar waren ([www.impp.de](http://www.impp.de)).

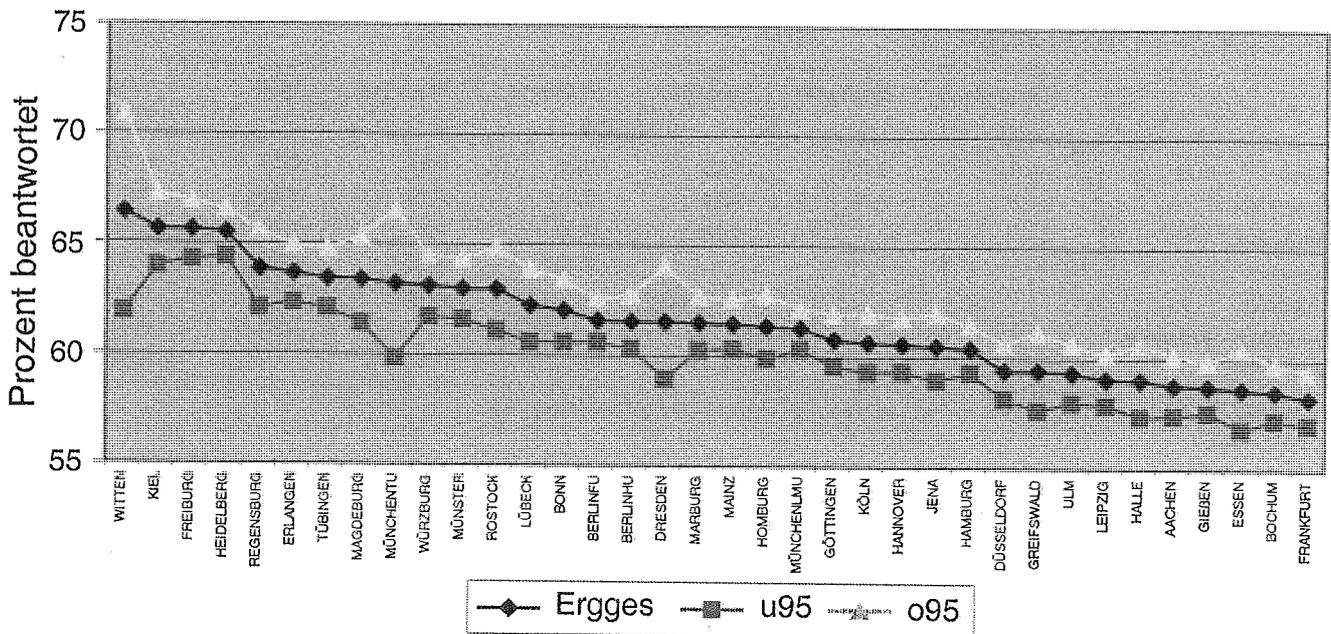


Abb. 1 Ärztliche Vorprüfung 1994 - 2000, Gesamtergebnis nach Fakultäten.

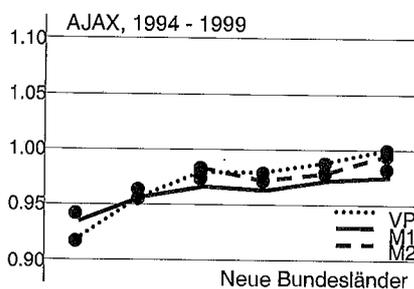
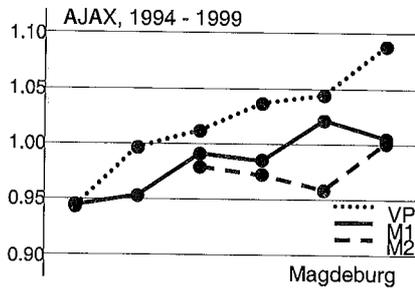


Abb. 2 Entwicklung des AJAX in Magdeburg und in den Fakultäten der neuen Bundesländer aggregiert, alle Fächer, 1994 - 1999; AJAX = 1 entspricht Bundesdurchschnitt.

fungen sichtbar gemacht werden kann. Dabei streuen die Ergebnisse der Ärztlichen Vorprüfung am stärksten. Im Längsschnitt gibt es nur selten starke Schwankungen.

Die Fakultäten der neuen Bundesländer zeigen im betrachteten Zeitraum einen deutlichen Aufwärtstrend (Abb. 2). Als Gruppe haben sie den Bundesdurchschnitt in allen Prüfungsarten erreicht. Das spricht für eine überlegene Lehrleistung der bekanntlich unterdurchschnittlich ausgestatteten Ostfakultäten.

Für die fakultätsinterne leistungsorientierte Mittelzuweisung sind nicht die fächerübergreifenden Ergebnisse entscheidend, sondern das Ergebnis einzelner Fächer relativ zu den fächerspezifischen Ergebnissen im Bund und in den anderen Fakultäten. Diese fächerspezifischen Ergebnisse werden vom IMPP nicht getrennt für die Referenz- und die Nichtreferenzgruppe ausgewiesen.

Wir haben daher zunächst geprüft, wie das nicht adjustierte Jahresergebnis der Fakultäten mit dem AJAX korreliert. Dieser Zusammenhang ist in allen überprüften Jahren sehr eng. Offensichtlich erreicht man durch das Zusammenfassen der Herbst- und der Frühjahrsergebnisse bereits den größten Teil der Referenzgruppenadjustierung, die der AJAX leistet. Es er-

scheint daher zulässig, die Jahresergebnisse der Fächer orientierend auch unstandardisiert miteinander in Beziehung zu setzen.

Alle Fächer korrelieren untereinander positiv und sind auch hoch mit dem AJAX und dem nicht adjustierten Jahresergebnis der Fakultät korreliert, zu denen die fächerspezifischen Ergebnisse natürlich entsprechend ihrem Anteil an den Prüfungsfragen beitragen. Tab. 1 a u. b enthalten Ergebnisse einzelner Jahre für die Gruppe der Fächer im Physikum und im ersten Teil der Ärztlichen Prüfung. Das erhobene Befundmuster ist in beiden Fächergruppen in jedem Jahr des Berichtszeitraums zu finden (die Fächergruppe des zweiten Teils der Ärztlichen Prüfung ist noch nicht überprüft). Für eine leistungsorientierte Differenzierung der Fächer innerhalb der Fakultäten gibt dieses positive Assoziationsmuster daher keinen guten Anlass. Lediglich die Biometrie und weniger deutlich die Immunologie sind relativ unabhängig von den anderen Fächern.

Das Physikumsergebnis der eigenen Fakultät (Magdeburg) ist unter Berücksichtigung der ganzen Beobachtungszeit, innerhalb derer sie sich noch deutlich verbessern konnte, um durchschnittlich 2,1 Punkte besser als das der zusammengefassten übrigen Fakultäten. Sie hat damit das achtbeste Gesamtergebnis. Nimmt man die eigene Fakultät als Bezugsgröße einer Re-

**Tab. 1a** Korrelationen (Spearman-Koeffizienten) der fächerspezifischen Ergebnisse in der Ärztlichen Vorprüfung. 36 Medizinische Fakultäten, Jahresergebnis Herbst 1994 plus Frühjahr 1995.

	Physiologie	Biochemie	Biologie	Anatomie	Psych-Soz	AJAX	Gesamtergebnis (roh)
Physik	0,52	0,39	0,46	0,26	0,47	0,52	0,52
Physiologie		0,76	0,79	0,61	0,70	0,86	0,91
Biochemie			0,70	0,39	0,67	0,83	0,84
Biologie				0,46	0,90	0,80	0,89
Anatomie					0,38	0,53	0,69
Psychologie-Soziologie						0,81	0,84
AJAX							0,92

Koeffizienten ab 0,33 sind auf dem Niveau von 0,05 signifikant (2-seitig), ab 0,42 auf dem Niveau 0,01

**Tab. 1b** Korrelationen (Spearman-Koeffizienten) der fächerspezifischen Ergebnisse im ersten Teil der Ärztlichen Prüfung, 36 Medizinische Fakultäten, Jahresergebnis Herbst 1999 plus Frühjahr 2000.

	Pathologie	PaPa-KlCh	Human-genetik	Mikro-biologie	Biometrie	Anamnese	Radiologie	Pharmakologie	Akut-med.	Immunologie	AJAX	Gesamtergebnis (roh)
Geschichte	0,67	0,58	0,48	0,35	0,11	0,54	0,42	0,56	0,56	0,21	0,54	0,68
Pathologie		0,83	0,48	0,54	0,15	0,77	0,65	0,54	0,68	0,22	0,81	0,87
PaPaKlCh			0,60	0,46	0,25	0,79	0,63	0,60	0,73	0,26	0,85	0,90
Humangenetik				0,34	0,42	0,50	0,24	0,61	0,58	0,59	0,50	0,69
Mikrobiologie					-0,25	0,36	0,58	0,40	0,40	0,22	0,57	0,65
Biometrie						0,38	0,15	0,25	0,10	0,25	0,26	0,24
Anamnese							0,59	0,47	0,75	0,15	0,83	0,81
Radiologie								0,42	0,50	0,06	0,74	0,73
Pharmakologie									0,47	0,48	0,55	0,74
Akutmedizin										0,26	0,76	0,78
Immunologie											0,17	0,41
AJAX												0,89

gressionsgleichung (mit Fakultätsdummies und Teilnehmergewichtung; Abb. 3), unterscheidet sie sich signifikant positiv von 15 Fakultäten am unteren Ende der Verteilung. Signifikant besser ist keine andere Fakultät.

In den auf gleiche Weise durchgeführten fachspezifischen Analysen übertreffen die Ergebnisse der Magdeburger Studierenden in der Anatomie die von 22 anderen Fakultäten signifikant, in der Biochemie und der Physik von acht, in der Biologie nur von einer. In der Physiologie gibt es keinerlei signifikante Unterschiede. In den psychosozialen Fächern wird Magdeburg von drei Fakultäten geschlagen, von allen anderen ist es nicht zu unterscheiden. Das ist der Befund, wenn sechs Jahre kumulieren. Die Ergebnisse einzelner Jahre sind fächerspezifisch kaum belastbar<sup>3</sup> (Tab. 2).

Diese Analyse lässt über einen längeren Beobachtungszeitraum durchaus Schwerpunkte der Fakultät erkennen. In den meisten Fächern gibt es aber wenig (oder keinen) gesicherten Anlass, ihre relative Position gegenüber anderen Fakultäten zu prämiieren oder zu sanktionieren. Das IMPP-Ergebnis – absolut

**Tab. 2** Ärztliche Vorprüfung 1994–2000, fächerspezifische Vergleiche Magdeburg gegen übrige 35 Fakultäten.

	sign. bessere Fakultäten	sign. schlechtere Fakultäten
Gesamtergebnis	0	15
Anatomie	0	22
Physik	0	8
Biochemie	1	8
Biologie	0	1
Physiologie	0	0
Psychologie-Soziologie	3	0

oder als Ranking – bietet daher für eine fakultätsinterne leistungsorientierte Mittelvergabe keine angemessene Grundlage, insbesondere keine kurzfristig nutzbare.

Entsprechende Analysen können für jede Fakultät besonders durchgeführt werden, doch sollte dann die statistische Auffälligkeitsgrenze wegen des multiplen Testens adjustiert werden.

<sup>3</sup> Das IMPP stellt den 95%-Vertrauensbereich der fächerspezifischen Bundesergebnisse grafisch dar; außerhalb liegende Fakultätsergebnisse werden erkennbar (ungewichtet; IMPP 2001, S. 38 ff).

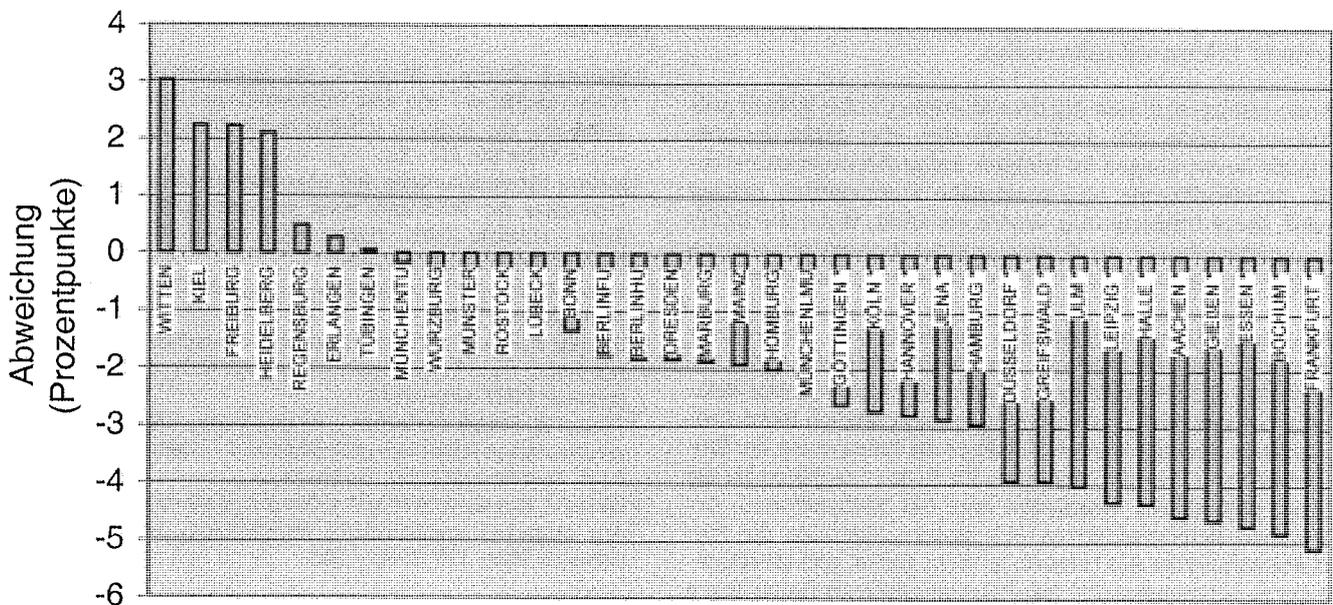


Abb. 3 Ärztliche Vorprüfung, Gesamtergebnis Herbst 1994 bis Frühjahr 2000, Referenz Magdeburg (\* = sign. Unterschied zur Referenz,  $p < 0,05$ ).

## Diskussion

Die medizinischen Fakultäten haben ein genuines Interesse, Studierende der Medizin so wirksam und wirtschaftlich wie möglich auszubilden. Es gibt dennoch keinen Zweifel, dass für Universitäten und Fakultäten eine stärkere Leistungsorientierung und Ausrichtung am Erfolg notwendig sind [3]. Der Gesetzgeber hat dies im § 5 des Hochschulrahmengesetzes zum Ausdruck gebracht.

Es liegt auf den ersten Blick nahe, die Ergebnisse der drei vom IMPP betreuten bundeseinheitlichen schriftlichen Prüfungen als Kriterium des Lernerfolgs und der Lehrleistung in den medizinischen Fakultäten anzusehen. Ihre Eignung als Parameter einer fakultätsinternen leistungsorientierten Mittelvergabe wird jedoch fraglich, weil die Ergebnisse der Fächer voneinander nicht unabhängig und die Ergebnisse der Fakultäten untereinander nicht hinlänglich distinkt sind.

Es lassen sich zwar durchaus Ergebnisse einzelner Fakultäten als signifikant unterschiedlich von denen anderer Fakultäten erkennen. Dies ist allerdings selbst dann wenig trennscharf, wenn Daten über einen längeren Zeitraum einbezogen werden (hier sechs Jahre). Bei den Physiologieergebnissen z.B. ist der Gradient der Fakultäten so flach, dass sich überhaupt keine Fakultät signifikant von irgendeiner anderen unterscheidet.

Eine fächerspezifische leistungsorientierte Mittelvergabe, vor allem eine jährlich neu taxierende, würde auf dieser Basis überwiegend nur statistisches Rauschen evaluieren und präzisieren.

Die eigene Fakultät hat beschlossen, die jährlichen IMPP-Ergebnisse in der Form gruppierter Rangplätze in ihrem Regelwerk der leistungsorientierten Mittelvergabe zu berücksichtigen<sup>4</sup>. Diese Beschlusslage ist nach den vorliegenden Daten nicht belastbar.

Die vorgelegte Auswertung geht von öffentlich verfügbaren Daten des IMPP aus. Dabei handelt es sich um Aggregatdaten, nicht um Individualdaten. Trotz einer immer noch erheblichen Disaggregation ergeben sich dadurch Grenzen der Auswertung. Wir haben uns entschieden, die Daten für ein akademisches Jahr (das heißt Herbst- und folgende Frühjahrsprüfung) zusammenzufassen. Den Anteil der Prüfungsteilnehmer in der Referenzgruppe berücksichtigen wir durch indirekte Standardisierung (AJAX). Die Vorteile des AJAX liegen in seiner Robustheit und Neutralität, die aufgrund der Jahresaggregation und der Referenzgruppenbereinigung erzielt werden. Das über ein akademisches Jahr zusammengefasste Gesamtergebnis ohne Berücksichtigung der Referenzgruppen korreliert mit dem AJAX sehr hoch. Die fächerspezifischen Ergebnisse, die nicht nach Gruppenzugehörigkeit stratifiziert vorliegen, haben wir daher als Jahresdaten analysiert. Das IMPP selbst könnte analoge Auswertungen auf Individualebene durchführen und damit die Gruppenzugehörigkeit (und weitere Merkmale) explizit berücksichtigen.

Die Daten des zweiten Teils der Ärztlichen Prüfung sind in die vorgelegten Analysen aus Kapazitätsgründen noch nicht einbezogen worden. Für die im Titel genannte Schlussfolgerung ist es aber schon hinreichend, wenn die Fächergruppe des Physikums nicht in eine leistungsorientierte Mittelvergabe einbezogen werden kann.

Die IMPP-Daten gelten bisher als einziges hartes Resultatmaß zur Bewertung der Lehre, ähnlich wie – trotz Kritik – die Impaktfaktoren für die Publikationsleistungen der Forschung. Es könnte für den Stellenwert der Lehre in den Fakultäten durch-

<sup>4</sup> Eine Präsidialkommission des Medizinischen Fakultätentages hatte 1999 empfohlen, die Prüfungsergebnisse des IMPP und die Bewertung der Fächer durch die Studierenden jeweils in Noten von 1–5 zu transformieren und zu mitteln (Prof. Dr. R. Greger, Protokoll des 60. Ordentl. Med. Fakultätentages 1999 in Dresden).

aus nachteilig sein, wenn man für ihre Evaluation über ein „objektives“ Erfolgsmaß nicht verfügt.

Andererseits setzt man mit dem Multiple-Choice-System prüfungsorientierte Lehr- und Lernanreize, über deren dysfunktionale Auswirkungen im Hinblick auf problemstrukturierende, problemlösende und praxisnahe Kompetenzen geklagt wird. Diese Fehlsteuerung würde verstärkt, wenn die IMPP-Prüfungsergebnisse als Parameter einer leistungsorientierten Mittelvergabe auf die Institute und Kliniken zurückwirken.

AJAX und fächerübergreifendes Jahresergebnis messen nur den Output des Lehrbetriebs. Zur fakultätsvergleichenden Beurteilung der Lehrleistung im Sinne einer Input-Output-Relation, wie das HRG sie anstrebt, müssen zusätzlich Input-Größen berücksichtigt werden, wie etwa die Abiturnoten der Studienanfänger (und der Wechsler), die in den Fakultäten für die Lehre verfügbaren Ressourcen oder das Vorhandensein bestimmter Reforminitiativen. Wir kennen allerdings die Produktionsfunktion des Lehrerfolgs nicht genau, und theoriegeleitet zusammengestellte fakultätsübergreifende Daten liegen nicht vor. Die zur Vorbereitung des Kostennormwertes gesammelten Vergleichsdaten werden zu prüfen sein.

Darüber hinaus messen die verwendeten Prüfungsergebnisse nur eine Output-Dimension. Die Studiengeschwindigkeit z. B. könnte explizit zusätzlich berücksichtigt werden [1].

Insgesamt kann man den externen IMPP-Ergebnissen daher zur Zeit keinen überzeugenden oder gar dominanten Stellenwert in der Evaluation der Lehre zusprechen. Immerhin arbeitet das IMPP an einer Weiterentwicklung der Prüfungsinhalte in Richtung auf eine durchgehende Fallorientierung und eine stärkere Berücksichtigung der medizinischen Praxis.

Als Alternative könnten die Fakultäten innovationsfördernde (nicht allein leistungsorientierte) Elemente in der Lehre zu honorieren versuchen, die bevorzugt auf interne Autoevaluation und einen kontinuierlichen Verbesserungsprozess zielen, wie z. B. das Vorhandensein von expliziten Scheinkriterien, Skripten, stoffbegleitenden Prüfungen ohne Sanktionscharakter, Studentenbefragungen und Peer-Reviews.

## Literatur

- <sup>1</sup> Brähler E, Wittig U, Beckert C. Der Studienerfolg an Medizinischen Fakultäten – wie viele Studienanfänger bestehen nach vier Semestern die ärztliche schriftliche Vorprüfung? Gesundheitswesen 1998; 60: 317–321
- <sup>2</sup> Institut für medizinische und pharmazeutische Prüfungsfragen. Ergebnisse der Prüfungen nach der Approbationsordnung für Ärzte – Frühjahr 2000 – Herbst 2000. Mainz: IMPP, 2001 (IMPP Ergebnisbericht Medizin 2000)
- <sup>3</sup> Müller-Böling D. Die entfesselte Hochschule. Gütersloh: Bertelsmann Stiftung, 2000
- <sup>4</sup> Schmitt H, Robra BP. Lehrevaluation an deutschen medizinischen Fakultäten – ein Verfahrensvorschlag zur Nutzung der IMPP-Ergebnisse. 2001, eingereicht
- <sup>5</sup> Theisel N, Stosch C, Koebke J. Evaluationsbemühungen an den medizinischen Fakultäten in Deutschland – Ergebnisse einer Umfrage. Med Ausbild 2000; 17: 18–21
- <sup>6</sup> Weber A, Wacker A, Weltle D, Lehnert G. Stellenwert der Lehre an den deutschen medizinischen Universitäten. Dtsch Med Wochenschr 2000; 125: 1560–1564

Prof. Dr. med. Bernt-Peter Robra, M.P.H.

Institut für Sozialmedizin und Gesundheitsökonomie  
der Medizinischen Fakultät der Otto-von-Guericke-Universität  
Magdeburg  
Leipziger Straße 44  
39120 Magdeburg