



## MEDIZINISCHE FAKULTÄTEN

# Der Ausbildungserfolg im Vergleich (III)

## DÄ-ARTIKELSERIE



In zwei bereits in dieser Zeitschrift veröffentlichten Arbeiten wurde ein adjustiertes Ranking der medizinischen Fakultäten auf der Basis der Erfolgsraten (DÄ, Heft 25/2006) und des Notendurchschnitts (DÄ, Heft 34–35/2006) im schriftlichen Teil der Ärztlichen Vorprüfung vorgestellt. In dieser Arbeit werden nunmehr die Resultate der mündlichen Prüfungen in ihrem Verhältnis zu den schriftlichen analysiert. Die Methodik der Datenerhebung und -analyse wurden in einer gesonderten Dokumentation beschrieben ([www.aerzteblatt.de/plus3406](http://www.aerzteblatt.de/plus3406)).

Zwischen den Ergebnissen im schriftlichen und im mündlichen Teil der Ärztlichen Vorprüfung besteht an den Hochschulen nur ein geringer Zusammenhang.

Hendrik van den Bussche, Karl Wegscheider,  
Thomas Zimmermann

**D**ie Ärztliche Vorprüfung (ÄVP) wurde 1986 um eine mündliche Prüfung erweitert. Damit sollten weitere, für das ärztliche Handeln notwendige Qualifikationen getestet werden, die über das von den Multiple-Choice-Fragen in den schriftlichen Prüfungen in erster Linie erfasste Faktenwissen hinausgehen. In einem halbamtlichen Kommentar des Jahres 1989 zur Approbationsordnung wird die damals vorherrschende Meinung über das Verhältnis von schriftlichen und mündlichen Prüfungen wie folgt beschrieben:

*„Da schriftliche Prüfungen nach dem Antwort-Wahl-Verfahren geeignet sind, vor allen Dingen kognitives Wissen abzufragen, bedürfen sie einer Ergänzung durch mündliche und mündlich-praktische Prüfungen. In*

*der Ausbildung zum Arzt kommt es entscheidend auf Beobachtungen und zielsicheres Reagieren, auf das Denken in großen Zusammenhängen, die Fähigkeit zur Gesprächsführung, Fertigkeiten und Fähigkeiten zu ärztlichem Handeln an. Ein ausgewogenes Prüfungswesen verhindert somit die einseitige Ausrichtung bei Lehrenden und Lernenden und trägt dazu bei, dass sowohl dem theoretischen als auch dem praktischen Unterricht der notwendige Anteil zugesichert ist.“\**

Vor diesem Hintergrund erscheint es 20 Jahre nach Einführung der mündlichen Prüfung von Interesse, ob und in welchem Umfang die fa-

\* Approbationsordnung für Ärzte – Bundesärzteordnung. Bearbeitet von H.-P. Brauer und F. F. Stobrawa. Deutscher Ärzte-Verlag, Köln 1989, Seite 38

Institut für Allgemeinmedizin, Universitätsklinikum Hamburg-Eppendorf (Prof. Dr. med. van den Bussche, Dipl.-Psych. Dr. Zimmermann)

Institut für Statistik und Ökonometrie, Department Wirtschaftswissenschaften, Universität Hamburg (Prof. Dr. rer. pol. Wegscheider)

kulturellen Notendurchschnitte in den mündlichen Prüfungen in der Routine von den Notendurchschnitten in den schriftlichen Prüfungen abweichen und welche Faktoren gegebenenfalls mit diesen Abweichungen assoziiert sind.

In der mündlichen Prüfung werden die Studierenden nach einem Zufallsverfahren in zwei der Fächer Physiologie, Anatomie, Biochemie und Psychologie/Soziologie geprüft. In der Bewertung des Gesamtergebnisses hatte die schriftliche Prüfung doppeltes Gewicht. Dieses Verfahren galt bis Frühjahr 2004 und somit für den hier untersuchten Zeitraum von 1994 bis 2004.

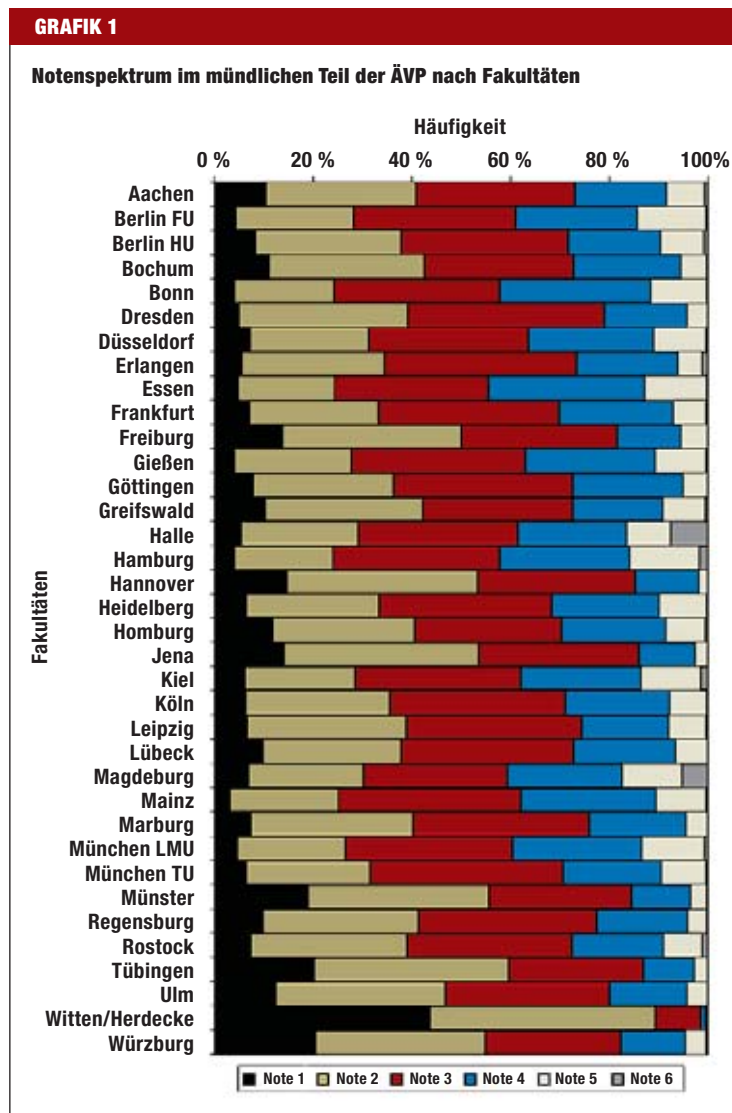
Würden die mündlichen Noten lediglich ein getreues Abbild der schriftlichen Noten darstellen, so wäre die mündliche Prüfung mögli-

cherweise überflüssig. Als Ausgangsmodell für die nachfolgenden Analysen wurde daher ein Metaregressionsmodell mit den schriftlichen Noten als Prädiktor gewählt. Dieses Modell wurde im Folgenden um die bereits bei der Analyse der schriftlichen Noten verwendeten Kontrollvariablen erweitert; so sollten Einflussfaktoren identifiziert werden können, die sich spezifisch in den mündlichen Noten abbilden.

Der Notendurchschnitt aller bestandenen mündlichen Prüfungen über zehn Jahre (1994–2004) liegt gemittelt über alle Fakultäten bei 2,70 und ist somit um mehr als einen halben Punkt besser als der Notendurchschnitt von 3,27 in der schriftlichen Prüfung. Der beste fakultätspezifische Notendurchschnitt liegt bei 1,68 (Witten/Herdecke), der

schlechteste bei 3,03 (Bonn und Essen). Das Notenspektrum der einzelnen Fakultäten mit den deutlichen Verteilungsunterschieden zeigt *Grafik 1*. Auffällig ist sicherlich die Universität Witten/Herdecke, die in fast 90 Prozent der mündlichen Prüfungen die Noten 1 und 2 vergibt.

Die mündlichen Notendurchschnitte unterscheiden sich von den schriftlichen Notendurchschnitten nicht nur im Niveau (Mittelwert). Tatsächlich sind auch die Unterschiede zwischen den medizinischen Fakultäten bei den mündlichen Noten erheblich größer als bei den schriftlichen Noten. Die mithilfe eines Metaregressionsmodells geschätzte Varianz bei den fakultätspezifischen Notendurchschnitten ist bei den mündlichen Noten mehr als fünfmal so groß wie bei den schriftlichen Noten. Die Standardabweichung beträgt 0,26 Notenstufen bei den mündlichen Noten gegenüber 0,11 Notenstufen bei den schriftlichen Noten und ist damit etwa 2,5-mal so groß.



**Auffällig** sind bei den mündlichen Prüfungen die Verteilungsunterschiede im Notenspektrum der einzelnen medizinischen Fakultäten.

**Frauen- und Ausländeranteil**

In Anbetracht der erheblichen Extravariabilität in den mündlichen Noten überrascht es nicht, dass diese sich nicht besonders gut aus den schriftlichen Noten vorhersagen lassen. Zwar ist die Assoziation signifikant ( $p = 0,0001$ ), die schriftlichen Noten erklären aber die Variation in den mündlichen Noten im Metaregressionsmodell dennoch nur zu 33,4 Prozent.

Welche anderen Faktoren könnten die Extravariabilität in den mündlichen Noten erklären? Von den in Betracht gezogenen Kontrollvariablen wiesen zwei Faktoren, der Frauenanteil ( $p = 0,0072$ ) und der Anteil ausländischer Studierender ( $p = 0,0166$ ), einen eigenständigen Erklärungswert für die Extravariabilität auf: Fakultäten mit einem höheren Frauenanteil und/oder einem höheren Ausländeranteil vergaben im Vergleich zu den schriftlichen Noten bessere Noten in der mündlichen Prüfung. Frauenanteil und Ausländeranteil erhöhen zusammen mit den schriftlichen Noten den Erklärungswert des Modells auf 57,5 Prozent. ▶

Grafik 2 stellt die schriftlichen und mündlichen Durchschnittsnoten der Fakultäten im Einzelnen dar. Die Fakultäten sind absteigend nach der Differenz zwischen mündlichen und schriftlichen Noten geordnet.

Die Grafik illustriert die größere Variabilität in den mündlichen Noten im Vergleich zu den schriftlichen Noten. Die eingezeichneten Trendlinien (lineare Regressionen) lassen erkennen, dass die Größenordnung der Differenzen im Wesentlichen von den mündlichen Noten bestimmt wird. Die Differenz ist am geringsten in Kiel, Heidelberg und Bonn (< 0,30), am größten in Witten/Herdecke (1,38). Ein Ranking nach mündlichen Durchschnittsnoten würde eine grundsätzlich andere Reihenfolge als ein Ranking nach schriftlichen Noten ergeben (Kendall- $\tau = 0,352$ ).

In der Untersuchung traten erhebliche Unterschiede zwischen mündlicher und schriftlicher Notengebung an den einzelnen Fakultäten zutage. Drei Erklärungsansätze scheinen möglich:

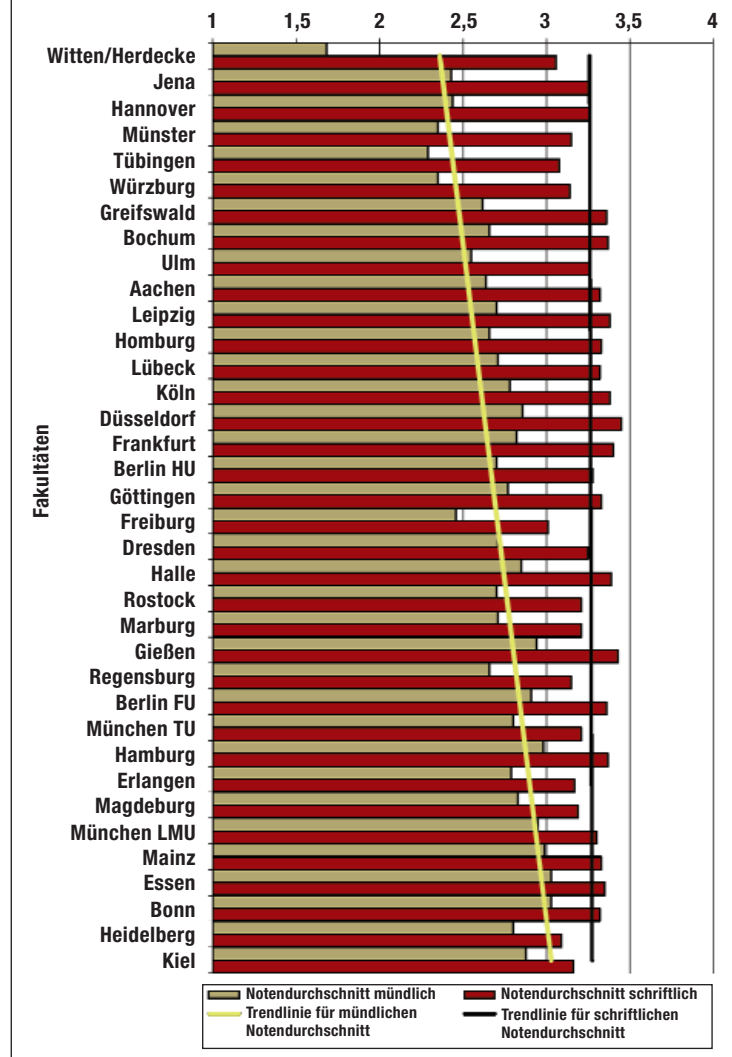
- Die mündliche Prüfung bildet tatsächlich – wie bei der Einführung postuliert – grundsätzlich andere Eigenschaften als die schriftliche Prüfung ab. Diese Eigenschaften sind nicht oder nur schwach mit dem in der schriftlichen Prüfung abgefragten Faktenwissen assoziiert.

- Bezüglich des mündlichen Prüfungsteils hat der Ordnungsgeber den Fakultäten keine nennenswerten Vorgaben im Hinblick auf die Zielsetzungen und die Durchführung der Prüfung gemacht. Es war und ist den einzelnen Fakultäten oder – sofern nicht explizite intrafakultäre Richtlinien existieren – gar den einzelnen Prüfern weitgehend überlassen, wie die Prüfungssituation ausgestaltet wird und welche Anforderungen gestellt werden. Die mündlichen Prüfungsergebnisse unterscheiden sich somit hauptsächlich, weil das Anforderungsniveau variiert und weil die Prüfer ein hohes Maß an individuellem Entscheidungsspielraum für die Benotung haben.

- Mündliche Prüfungen sind ein sensibleres „Messinstrument“ als schriftliche Prüfungen. Die Standardisierung der schriftlichen Prüfung

GRAFIK 2

Notendurchschnitte über den Zeitraum 1994–2004 im schriftlichen und im mündlichen Teil der ÄVP, geordnet nach abnehmendem Unterschied, illustriert durch Trendlinien



**Der Vergleich** zwischen den Ergebnissen der schriftlichen und der mündlichen Prüfungen an den einzelnen medizinischen Fakultäten zeigt deutliche Unterschiede.

und die dadurch mögliche gezielte Prüfungsvorbereitung nivellieren die unterschiedliche Befähigung der Studenten. In der mündlichen Prüfung hingegen treten die Qualitätsunterschiede zwischen den Studierenden offen zutage. Damit werden auch die Qualitätsunterschiede zwischen den Fakultäten besser sichtbar.

### Unterschiede bei der Prüfung?

Aufgrund der Analyseergebnisse erscheint der letzte Erklärungsansatz, obwohl theoretisch möglich, wenig plausibel, da sich die mündlichen Durchschnittsnoten der Fakultäten nur zu etwa einem Drittel aus den schriftlichen Noten vorhersagen lassen und die resultierenden Rankings

wenig Übereinstimmung aufweisen würden. Die Eigenständigkeit der mündlichen Notengebung erscheint offensichtlich. Es bleibt die Frage zu diskutieren, ob sich in den mündlichen Noten tatsächlich andere Eigenschaften oder eher unterschiedliche Prüfungskulturen abbilden.

Einen Hinweis geben die Regressionsanalysen. Offenkundig kommen in den mündlichen Noten Einflussfaktoren zum Tragen, die in den schriftlichen Noten nicht abgebildet sind und eigenen Gesetzen unterliegen. So erweisen sich die Abweichungen der mündlichen Noten von den schriftlichen Noten als geschlechtsspezifisch. Das ist insofern bemerkenswert, als weder bei der

Analyse der Erfolgsraten (DÄ, Heft 25/2006) noch bei der Analyse der schriftlichen Noten (DÄ, Heft 34–35/2006) ein Unterschied zwischen den Geschlechtern zu erkennen war. Es ist vorstellbar, dass Frauen in den nicht auf Fachkenntnisse begrenzten Eigenschaften bessere Leistungen bringen, aufgrund ihrer überlegenen verbalen Fähigkeiten (siehe die PISA-Ergebnisse) in mündlichen Prüfungen besser abschneiden oder aber von den noch überwiegend männlichen Prüfern besser bewertet werden. Zwischen diesen Deutungen kann man auf der Basis der Untersuchungsergebnisse nicht unterscheiden. Die Assoziation der mündlichen Noten mit dem Frauenanteil kann aber als gesichert angesehen werden.

Bemerkenswert ist auch die Relation der mündlichen Noten zum Ausländeranteil. Die schriftlichen Noten fielen mit steigendem Ausländeranteil schlechter aus (DÄ, Heft 34–35/2006). Die mündlichen Durchschnittsnoten hingegen fallen bei hohem Ausländeranteil besser aus, als aufgrund der schriftlichen Noten zu erwarten gewesen wäre. Zu vermuten ist, dass die Prüfer in den mündlichen Prüfungen die bei ausländischen Studierenden häufiger vorkommenden Sprachprobleme in ihrer Bewertung ausgleichen. Das ist in den schriftlichen Prüfungen nicht möglich.

### Präzisierung wäre sinnvoll

Interessant ist ferner, welche Einflussgrößen keine spezifische Assoziation mit den mündlichen Noten aufweisen. Während die schriftlichen Durchschnittsnoten der Fakultäten mit dem schulischen Leistungsniveau und mit der Personalausstattung der Fakultäten assoziiert sind (DÄ, Heft 34–35/2006), können diese Faktoren zur Erklärung der Besonderheit der mündlichen Prüfungsergebnisse nicht herangezogen werden ( $p = 0,10$  beziehungsweise  $0,20$  im univariaten,  $p = 0,59$  beziehungsweise  $0,71$  im multivariaten Regressionsmodell).

Damit ist belegt, dass die mündlichen Noten tatsächlich etwas anderes abbilden als die schriftlichen. Und es ist somit denkbar, dass die



Foto: mauritius images

### Ein Ranking

wie nach Auswertung des Erfolgs bei den schriftlichen Prüfungen scheint bei den mündlichen Prüfungen nicht sinnvoll.

ses bei der Einführung der mündlichen Noten formulierte Ziel tatsächlich erreicht wurde. Allerdings stimmt die Größenordnung der interfakultären Unterschiede in der Notengebung nachdenklich. Nur schwer vorstellbar ist nämlich, dass die Leistungsdifferenzen im mündlichen Teil um ein Vielfaches größer sind als im schriftlichen. Viel plausibler erscheint, dass bei der mündlichen Notengebung unterschiedliche Prüfermaßstäbe und unterschiedliche Fähigkeiten der Studierenden aufeinandertreffen und mehr oder minder zufällig interagieren. Unterschiedliche Einstellungen zur Notengebung im Allgemeinen mögen ein Übriges tun.

Weitere Erklärungen bieten sich an: Die Fächerzulassung könnte das Ergebnis der mündlichen Prüfung beeinflussen, da nicht überall die gleichen Prüferkapazitäten vorhanden sind. Ein weiterer Einflussfaktor dürften die unterschiedlichen Zugangswege zur mündlichen Prüfung sein, da diese sowohl vor als auch nach der schriftlichen Prüfung – im letzteren Fall in Kenntnis der schriftlichen Ergebnisse – stattfinden kann.

Jedenfalls schwankt die Anzahl der Prüfungsteilnehmer in den mündlichen Prüfungen im Verhältnis zu der bei den schriftlichen Prüfungen über den gesamten Zeitraum von minus 18 Prozent bis plus drei Prozent zwischen den einzelnen Fakultäten.

Letztlich ist eine eindeutige inhaltliche Interpretation der mündlichen Notendurchschnitte (im Gegensatz zu den schriftlichen Notendurchschnitten) nicht möglich. Der Prüfungserfolg in der mündlichen Prüfung besagt lediglich, wie schwer oder leicht es für die Studierenden an der jeweiligen Fakultät ist, diese Prüfung erfolgreich zu bestehen. Ob die Ursache im Ausbildungsniveau der Studierenden oder in den lokalen Ausformungen des Prüfungswesens liegt, muss offen bleiben. Auf der Basis der mündlichen Prüfungsergebnisse ist unter diesen Umständen ein fakultäres Ranking nicht sinnvoll. Das gilt auch für die Gesamtnoten, da die Resultate der mündlichen Prüfung in das Gesamtergebnis eingehen und aufgrund ihrer größeren Streuung die Gesamtnote zu dominieren drohen. Deshalb werden die Fakultäten auch nicht nach mündlichen Notendurchschnitten oder Gesamtnotendurchschnitten in einer Ligatabelle dargestellt.

Die Ergebnisse dieser Analyse sollten Anlass sein, die Zielsetzung und den Durchführungsmodus der mündlichen Prüfung kritisch zu reflektieren. Zu diskutieren wäre, ob nicht eine Präzisierung der Prüfungsziele und Prüfungsanforderungen sowie eine Normierung der Prüfungsdurchführung angebracht wären. Nur so kann ein Übermaß an Vielfalt abgebaut werden – zugunsten eines höheren Maßes an Prüfungsvalidität und -reliabilität. Die Ergebnisse untermauern die Forderung nach Prüfungsforschung, die einen Beitrag zur Klärung der hier aufgeworfenen Fragen leisten könnte.

■ Zitierweise dieses Beitrags:  
Dtsch Arztebl 2006; 103(47):A 3170–6

### Anschrift für die Verfasser

Prof. Dr. med. Hendrik van den Bussche  
Institut für Allgemeinmedizin  
Universität Hamburg-Eppendorf  
Martinistraße 52  
20246 Hamburg  
E-Mail: bussche@uke.uni-hamburg.de