

# Ständige Verbesserung

## Notenvergabe und -unterschiede an Universitäten

| GERD GRÖZINGER | VOLKER MÜLLER-BENEDICT |

**Studienabschlussnoten an Universitäten bewegen sich entweder seit Jahrzehnten bereits auf hohem Niveau oder werden immer besser. Zu diesem Ergebnis kommt eine aktuelle Untersuchung. Ein Beitrag über die Effekte und Hintergründe dieser Entwicklung. (Zur Entwicklung bei Promotionsnoten siehe Beitrag Hornbostel/Johann.)**

**S**ein Beginn der siebziger Jahre werden Examensnoten an deutschen Universitäten fast überall immer besser. Zu diesem ersten Schluss kommen wir aufgrund der Ergebnisse unseres von der DFG geförderten Forschungsprojektes zur Notengebung an deutschen Hochschulen. Dabei wurden in sieben Universitätsarchiven aus 138 000 Prüfungsakten ca. 700 000 Examensnoten für eine repräsentative Fächerauswahl zwischen 1960 (teilweise auch früher) und 1996 erhoben, dazu die vollständige elektronische Prüfungsdatenbank aller Fächer und Hochschulen an einem Forschungsdatenzentrum für die maximal mögliche Periode 1996 bis 2013 ausgewertet und Gruppeninterviews mit Prüfenden unternommen. Die folgenden Aussagen

beziehen sich deshalb zunächst auf die klassischen Abschlüsse Staatsexamen, Diplom, Magister.

Die bisher umfassendste Untersuchung deutscher Examensnoten bestätigt ein Bild breiter Noteninflation, die der deutsche Wissenschaftsrat für eine

### »Der überwiegende Trend zur Notenverbesserung verläuft nicht linear, sondern in Zyklen.«

kürzere Periode bereits 2012 moniert hatte. Am meisten verbessern sich in unserer Fächerauswahl auf Basis der Analyse der Prüfungsakten die Noten in Deutsch für Lehramt um mehr als eine ganze Note im Durchschnitt (auf der bekannten Skala zwischen 4,0 und 1,0), in Biologie am geringsten um immerhin noch 0,6. Allerdings gilt auch: in Biologie und Psychologie kann man die Notenlage spätestens seit Beginn der 1970er Jahre als derart gut einstufen, dass die Leistungsdifferenzierung dort zwangsläufig durch eine Häufung der Noten im Bestbereich gefährdet ist. In Psychologie wurden 55 Prozent der bestandenen Prüfungen zwischen 1971 und 1997 mit einem „sehr gut“, und 96 Prozent mit einem „sehr gut“ oder „gut“ bewertet. In Biologie liegen diese Anteile im gleichen Zeitraum bei 63 Prozent bzw. 97 Prozent. Abbildung 1 zeigt die zeitlichen Verläufe für die Fächer mit dem Trend zur Notenverbesserung für die untersuchten Hochschulen. Keinen

Trend konnten wir für die lange Periode von ca. 1960 bis 2010 für Maschinenbau, Soziologie und Germanistik Magister sowie Jura feststellen. Für die neuere Zeit von Mitte der 90er Jahre bis fast Mitte der 2010er dagegen und auf ganz Deutschland bezogen unterliegen nur noch die universitären Diplomstudienfächer Mathematik und Soziologie nicht diesem breiten Trend der Verbesserung. Der langfristige Trend zur Notenverbesserung geht einher mit einer Verbesserung der Abiturnoten (siehe Beitrag Grözinger/Baillet in *Forschung & Lehre* 1/17).

Der überwiegende Trend zur Notenverbesserung verläuft dabei nicht linear, sondern in Zyklen. Phasen von stärkerer Verbesserung wechseln mit Phasen von schwächerer Verschlechterung ab. Diese Phasen sind für viele Fächer mit den Arbeitsmarktaussichten oder den Studierendenzahlen gekoppelt. Bei schlechten Arbeitsmarktaussichten werden im Durchschnitt schlechtere Noten vergeben, d.h. die „Guten“ werden stärker hervorgehoben. Bei starkem Studienandrang werden ebenfalls schlechtere Noten vergeben. Weil jedoch in den Zeiten der Verbesserung der Noten – gute Arbeitsmarktlage oder weniger Studierende – die Höhe der Verbesserung immer größer ausfällt als in der anschließenden Verschlechterungsphase, ergibt sich langfristig der Trend zur ständigen Notenverbesserung. Das gilt wohl, weil die Vergabe schlechterer Noten auf sehr viel mehr Widerstand stößt. Dieses Aufdecken einer Asymmetrie in der Zyklizität ist unsere zweite relevante Erkenntnis. Sie überlagert die grüne Inflation.

### Notenunterschiede zwischen Fächern und Universitäten

Über diese prinzipielle Verbesserung der Noten hinaus existieren erhebliche

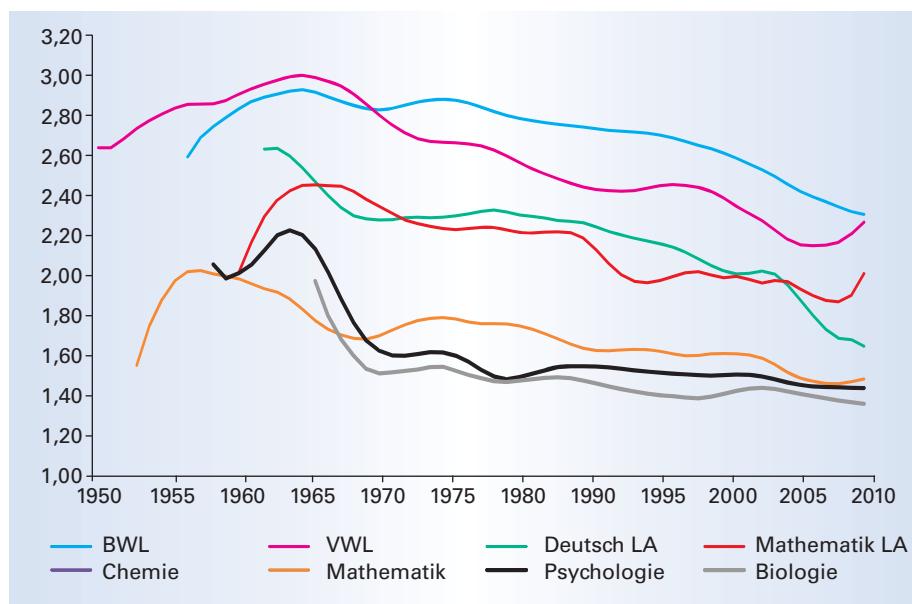
#### AUTOREN



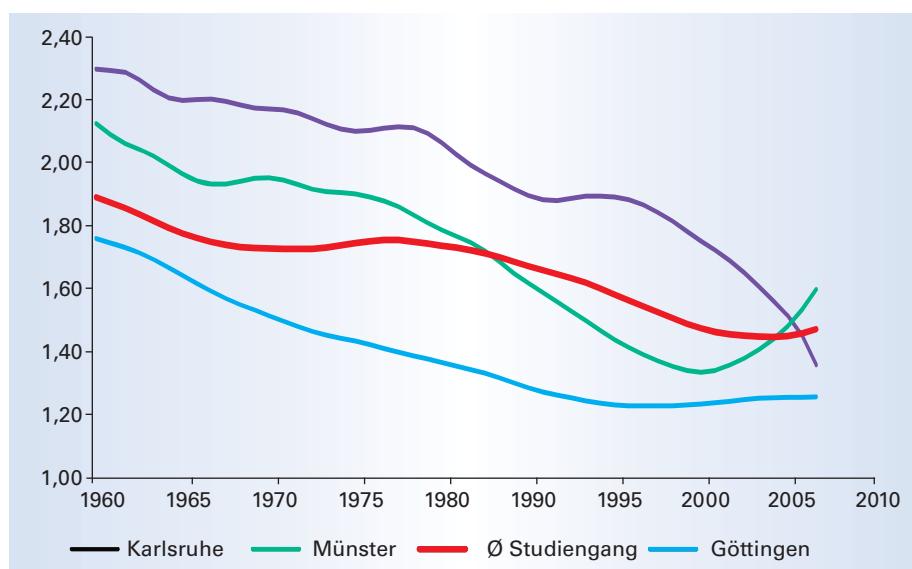
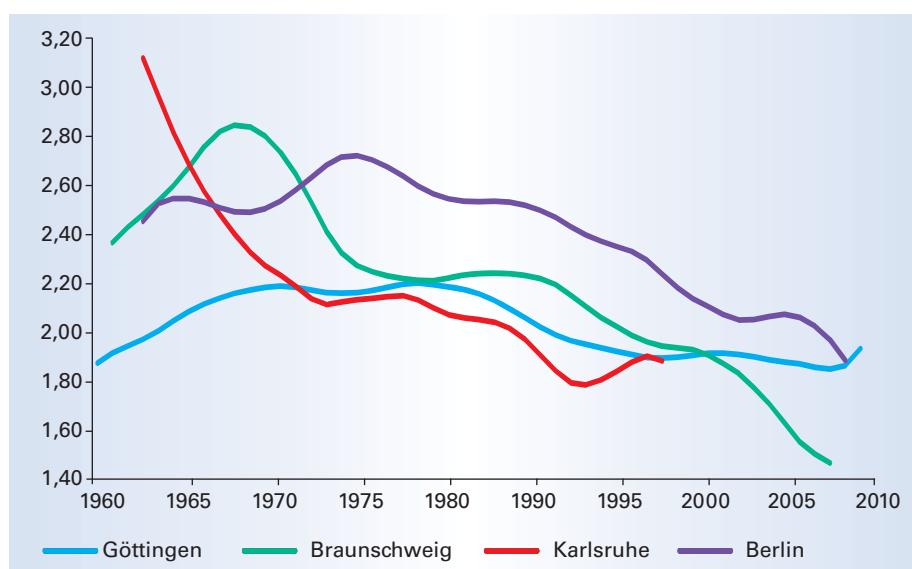
**Gerd Grözinger** ist Professor für Sozial- und Bildungswissenschaften an der Europa-Universität Flensburg.



**Volker Müller-Benedict** ist Professor für Methoden und Statistik an der Universität Flensburg und Leiter des Zentrums für Methodenlehre (ZML).



**Abbildung 1:** Verlauf der Abschlussnoten in Studiengängen mit langfristiger Notenverbesserung (geglättet)



**Abbildung 2:** Verlauf der Abschlussnoten in Mathematik Lehramt an vier Universitäten (oben) und in Chemie an drei Universitäten (unten) (geglättet)

und dauerhafte Notenunterschiede zwischen den Fächern. So erhalten die meisten Juristen im Durchschnitt seit jeher eine 4,0 und werden damit am schlechtesten benotet. Für die weiteren untersuchten Fächer lassen sich über den ganzen Zeitraum hinweg sieben Gruppen statistisch signifikant abgrenzen, die von der Spitzengruppe 1: Biologie/Psychologie abwärts im Durchschnitt immer schlechtere Noten bekommen: 2: Mathematik/Chemie; 3: Maschinenbau/Germanistik/Soziologie 4: Mathematik Lehramt/Deutsch Lehramt; 5: VWL; 6: BWL; 7: Jura. Die Notenunterschiede zwischen den Fächern lassen sich zu einem kleinen Teil durch Fachspezifika erklären, wie z.B. dass der Wissenskorpus in Mathematik oder Linguistik klarer strukturiert ist, und dass dort eine größere Passung zwischen Vorlesungsstoff und Prüfungsstoff besteht, was zu besseren Beurteilungsmöglichkeiten führt. Die stabilen und großen Unterschiede zwischen den Fächern müssen jedoch zum größten Teil als „Fachkulturen“ begriffen werden, die historisch bedingt sind.

Aber nicht nur zwischen den Fächern gibt es Unterschiede im Notenniveau, auch bei den einzelnen Universitäten existieren Differenzen. Jede einzelne Hochschule praktiziert in jedem Fach eine eigene Prüfungskultur, die wesentlich die Unterschiede in der Noteinheit zwischen den Fächern und den Hochschulen beeinflusst, wie auch die qualitative Analyse der Gruppendiskussionen mit Prüfenden der Fächer Mathematik und Germanistik ergab. Zwei eklatante Beispiele sind in Abb. 2. zu sehen: Im gesamten Zeitraum erhalten die Geprüften im Studiengang Mathematik für Lehramt an Gymnasien an der FU Berlin im Durchschnitt um 0,7 schlechtere Noten als an der Georg-August-Universität Göttingen, im Studiengang Chemie entsprechend an der TU Karlsruhe um 0,5 schlechtere Noten als in Göttingen. Das bedeutet, Studienwillige können schon durch die Wahl der „richtigen“ (besser benotenden) Universität ihre erwartete Abschlussnote steigern, ohne das Studium auch nur angefangen zu haben. Die relative Persistenz der Unterschiede auf hochschulischer Ebene ist unsere *dritte* relevante Erkenntnis.

Solche Unterschiede müssen gar nicht direkt von den Lehrenden mit etwaig unterstellter differenter Motivation ausgehen. Aus der Analyse der Prüfungsakten konnte etwa entnommen werden, welche Hierarchie bei den Prü-

fungsformen herrscht. Die relativ besten Bewertungen gibt es bei schriftlichen Abschlussarbeiten, es folgen mündliche Prüfungen, dann schriftliche Klausuren. Also alleine durch deren Gewichtungen bei der Abschlussnote kann es schon zu Differenzen bei Hochschulen oder auch bei der Einführung neuer Prüfungsordnungen zu Veränderungen kommen, eine *vierte* Erkenntnis.

### **Leistungskonforme und leistungsfremde Einflüsse auf die Notengebung**

Generell lassen sich leistungskonforme und leistungsfremde Einflüsse auf die Notengebung unterscheiden. In der Querschnittsanalyse auf Basis der amtlichen Notenstatistik wurden eine Vielzahl davon getestet. Seit 1996 liegen Individualangaben dazu vor, insgesamt konnten wir ca. 5,3 Millionen Fälle regressionsanalytisch untersuchen. Allerdings hat die Hochschulstatistik hier ein entscheidendes Manko. Ausgerechnet die Abiturnote fehlt in der Absolventenstatistik. Nach allem, was wir aus Untersuchungen wissen, ist sie aber immer

noch der beste Prädiktor für die Examsnote. Es ist völlig unverständlich und wohl nur dem Druck der das Licht der Öffentlichkeit scheuenden Schulpolitik zu verdanken, dass diese zentrale Größe, die bei der Einschreibung den Hochschulen selbstverständlich vorliegt, nicht in den Datensatz mit aufgenommen wurde.

Unsere Fächerauswahl hier war bei Universitäten: Maschinenbau, Elektrotechnik, Chemie, Biologie, Mathematik, Jura, Wirtschaftswissenschaften, Soziologie, Psychologie, Erziehungswissenschaften, Germanistik, Geschichte. Und bei den Fachhochschulen: Maschinenbau, Elektrotechnik, Wirtschaftswissenschaften, Sozialwesen.

Vermutlich leistungskonform sind dabei statistische Relationen, die den notenverschlechternden Einfluss des Alters bei Studienabschluss belegen, einer der stärksten Einflüsse. Trotz des Alterseffekts ist aber eine abgeschlossene Berufsausbildung notenverbessernd, was auf eine hohe Motivation dieser Gruppe hinweist. Einen rechnerisch immer starken negativen Einfluss dagegen hat eine ausländische Hochschulzugangsberechtigung. Das gleiche gilt quantitativ zwar etwas abgeschwächt, aber immer noch deutlich sichtbar für die Kombination

deutsches Abitur und ausländische Staatsangehörigkeit. Bei beiden Studierendengruppen dürfte an Hochschulen also noch Spielraum für eine verbesserte Betreuung sein. Manchmal allerdings ist eher auf fehlende Motivation zu schließen. Wir konnten bei denen mit deutschem Abitur noch Entfernungsdaten zwischen dem Kreis der Hochschulzugangsberechtigung und dem Kreis der Hochschule zuspielen. Absolventen aus der direkten Nähe wiesen ebenfalls relativ schlechtere Noten auf. Wir vermuten, dass bei der Wahl einer speziellen Fach-/Hochschulort-Kombination hier die räumliche Nähe wichtiger als das Fachinteresse war. Die Bedeutung dieser individuellen Merkmale markieren eine *fünfte* relevante Erkenntnis. Wenn etwa eine Hochschule viele Regionalstudierende und ausländische Staatsangehörige unter ihren Absolventen hat, wird sie rechnerisch ein schlechteres Notenniveau

sieren. *Siebtens* gilt also, Noten sind auch auf Fachebene eine Mischung aus leistungskonform und leistungsfremd.

### **Übergang BA/MA-System**

Interessant ist die Situation beim Übergang zum BA/MA-System. Hier haben die Universitäten die Gelegenheit genutzt, ihre Inflationstendenz bei den Altabschlüssen zu stoppen und beim Bachelor überall ein neues schlechteres Notenniveau zu etablieren. Die Fachhochschulen jedoch, die ja jetzt gleichwertige Abschlüsse vergeben, haben dagegen die Noten beim Bachelor verbessert. In den überlappenden Fächern Maschinenbau, Elektrotechnik, Wirtschaftswissenschaften haben FH-Absolventen nun einen beachtlichen Vorteil, wenn sie sich auf einen Masterstudienangang bewerben. Unsere *achte* Erkenntnis ist somit, Universitäten und Fachhochschulen haben unterschiedliche Notenniveaus- und auch Trends.

Während der Bachelor die Inflationstendenz an Universitäten unterbrach, zeigt der Masterabschluss auf breiter Linie sowohl von Anfang

an ein erheblich besseres Niveau und eine weitere Verbesserung. In der Kombination bedeutet das etwas für das quantitativ bedeutende Lehramt, dass die nun für den Eintritt in den Schuldienst relevanten Masternoten erheblich besser als die des alten Staatsexamens ausfallen. *Neunte* Erkenntnis: bei Masterabschlüssen zeigt sich wieder der Trend zur Noteninflation.

### **Mehr Transparenz**

Was folgt daraus? Unsere *zehnte* Erkenntnis ist, dass dringend mehr Transparenz nötig ist. So sollten Notenniveaus und Notenverteilung der einzelnen Lehrveranstaltungen innerhalb der Kollegialorgane besprochen werden. Vor allem aber sollten die Durchschnittsnoten der Fächer regelmäßig auf Bundesebene veröffentlicht werden, und bei Abschlussnoten von Absolventen sollte die ungefähre Relation zur Fachkohorte an der jeweiligen Einrichtung auf den Zeugnissen mit angegeben sein (z.B. „gehört zum besten Viertel“ oder „liegt über dem Durchschnitt“) oder alternativ die Verteilung der Noten 1 bis 4 über die letzten fünf Jahre als kleines Balkendiagramm mit vier Säulen dahinter abgebildet sein, wie es z.B. in Norwegen üblich ist.

## »Die Universitäten haben beim Bachelor überall ein neues niedrigeres Notenniveau etabliert, Fachhochschulen dagegen die Noten beim Bachelor verbessert.«

veau aufweisen als eine andere mit niedrigeren Anteilen darin. Das macht Vergleiche ohne die Einbeziehung solcher Informationen problematisch.

In Übereinstimmung mit den Ergebnissen der Längsschnittanalyse zeigte sich weiter, dass in fast allen Fächern die ebenfalls zugespielte regionale Arbeitslosigkeit einen Einfluss auf die Notengebung hatte. Wo es einen definierten Arbeitsmarkt für die Absolventen gab, bedeutete höhere Arbeitslosigkeit schärfere Selektion. Diese breite Bedeutung der Arbeitsmarktsituation für die Notenvergabe – so wenig bewusst sie den Lehrenden vielleicht sein mag –, die aber räumlich wie zeitlich deutlich wird, ist somit eine *sechste* neue Erkenntnis.

Einige getestete Variablen blieben erstaunlicherweise uneindeutig. Dies gilt vor allem für das Verhältnis Professoren zu Studierenden. Wir konnten seit 1998 auch die Personalstatistik noch zuspielen und deshalb eine solche Relation kalkulieren. Dagegen zeigten andere Dimensionen einen Einfluss, den wir nur schlecht als leistungskonform interpretieren können. So ist ein hoher Anteil an Promotionen bei den Studienabschlussnoten verbessernd, dazu ein hoher Frauenanteil bei den Professoren und das Durchschnittsalter der Profes-