

Einflüsse auf die Notengebung an deutschen Hochschulen

Eine Analyse am Beispiel der Wirtschaftswissenschaften

Gerd Grözinger
Flensburg

Uns geht die Bedeutung des Tabu nach zwei entgegengesetzten Richtungen auseinander. Es heißt uns einerseits: heilig, geweiht, andererseits: unheimlich, gefährlich, verboten, unrein ... Unsere Zusammensetzung „heilige Scheu“ würde sich oft mit dem Sinn des Tabu decken.

Sigmund Freud, Totem und Tabu, 1913

Die Benotungspraxis an Hochschulen scheint in Deutschland tabubeladen zu sein. Obwohl mit Examensbewertungen entscheidende Weichenstellungen für Karriereoptionen eröffnet werden - man denke nur an festgelegte Schwellenwerte für Masterstudiengänge oder Promotionszulassungen, an Stipendien für diese oder natürlich auch generell an Chancen auf dem Arbeitsmarkt – findet eine kritische Auseinandersetzung dazu hierzulande kaum statt. Es herrscht ganz überwiegend eine ‚heilige Scheu‘ von Prüfenden, Hochschulleitungen und Bildungspolitikern, die Benotungspraxis intensiver zu thematisieren. Und dies trotz empirischer Anzeichen, dass bei erwarteten sehr guten Durchschnittsnoten der studentische Arbeitsaufwand erheblich absinkt, also die eigene Arbeit an der Hochschule darunter leidet (Babcock 2010).

Auch die seit nun schon mehreren Jahren vorliegenden kritischen Berichte des Wissenschaftsrates haben am weitgehenden Ignorieren des Themas nicht viel geändert (zuletzt: Wissenschaftsrat 2012). Ebenfalls recht wenig beachtet werden die vom Institut für Forschungsinformation und Qualitätssicherung zusammengestellten und kartografisch aufbereiteten Promotionsnoten,¹ trotz der damit verbundenen pessimistischen Schlussfolgerung: „Noten sagen kaum noch etwas über die Qualität von Doktorarbeiten aus“ (IFQ 2014).

Man kann diese sich kaum auflösende Nebelbildung bezüglich der Praxis der Notengebung an den deutschen Hochschulen nur für ein weite-

¹ <http://www.forschungsinfo.de/promotionsnoten/>

res der vielen Privilegien halten, die eine Konzentration der Organisation auf die Dimension der ‚akademischen Oligarchie‘ mit sich bringt (Clark 1983). Wenn weder ‚Markt‘ noch ‚Staat‘ hierin Korrektiv sein wollen oder können, sind weitreichende Nicht-Angriffspakte der Lehrenden untereinander erwartbar. Und die funktionieren am besten, je weniger man voneinander weiß.

Aber es könnte noch eine andere, tiefer gehende Begründung dafür geben. In Überlegungen zu ‚Tabus über den Lehrberuf‘ hat Adorno Mitte der 60er Jahre des letzten Jahrhunderts einer für ihn wahrnehmbaren gewissen gesellschaftlichen Verachtung für die (Schul-)Lehrer psychoanalytisch nachgespürt. Und es in dem Zusammenhang für beachtenswert gefunden, dass die an der Universität davon ausgenommen sind:

„Bezeichnend, dass die Lehrer, die in Deutschland am meisten Ansehen genießen, nämlich eben die akademischen, in praxi nur höchst selten disziplinaire Funktionen ausüben, und dass sie wenigstens der Idee und der öffentlichen Vorstellung nach produktiv forschen, also nicht in dem als sekundär, und wie ich sagte, als scheinhaft verdächtigen pädagogischen Bereich fixiert sind“ (Adorno 1973: 75).

Dann wäre die heilige Scheu darin begründet, sich auf keinen Fall mit schulischen Praktiken, wo es z.B. interne Transparenz durch Notenkonferenzen gibt, vergleichbar machen zu wollen. Es ist übrigens schon für Adorno ein auf längerer Sicht verlorener Kampf: „So wie längst in Amerika, wo derlei Prozesse viel krasser verlaufen als hierzulande, wird der Professor allmählich, aber ich würde denken: unaufhaltsam zum Verkäufer von Kenntnissen, ein wenig bemitleidet, weil er jene Kenntnisse nicht besser für sein eigenes materielles Interesse zu verwerten vermag“ (ebd.: 76).

In den USA jedenfalls ist das Tabu einer Auseinandersetzung zur Benotungspraxis und zur Noteninflation schon weitgehend verschwunden. Die Praxis mag sich häufig noch vergleichsweise wenig unterscheiden, aber eine kritische Aufmerksamkeit ist dort auf jeden Fall gegeben. Gab man etwa im April 2015 in Google Scholar *grade inflation* ein, erhielt man über 12.000 Nennungen. Versuchte man es dagegen mit „Noteninflation“, waren es kaum 800. Da ist sicher einiges der Weltsprache Englisch zu verdanken, aber es ist nicht bestreitbar, dass ein ganz überwiegender Teil der Literatur zur Benotungspraxis aus den Vereinigten Staaten stammt (Überblicke bei Hu 2005; Hunt 2008; Johnson 2003; Ogilvie/Jelavic 2013). Und es gibt hochreputierliche Einrichtungen wie die Princeton University oder das Wellesly College, die sich einer anti-inflationären und von der Hochschulleitung dann kontrollierten Praxis bei der Benotung verschrieben haben (Butcher/McEwan/Weerapana 2014).

Für eine Diskussion hiesiger Verhältnisse ist diese Literatursituation nicht unproblematisch. Denn sowohl die Vielfältigkeit des amerikanischen Hochschulsystems wie einige seiner Spezifika machen eine Übertragbarkeit dortiger Erklärungsansätze oft schwierig. So sind etwa die nordamerikanische inner-universitäre Wettbewerbssituation im Bachelor, wo die einzelnen Fächer in Colleges um Studierende für ihre Schwerpunktsetzung konkurrieren oder die dort aufzufindende stärkere Gewichtung studentischer Lehrevaluationen für Einkommens-, Berufungs- und Tenure-Entscheidungen in Deutschland weniger bedeutsam. Und auch der für die neuere Zeit erste große Verbesserungssprung in der Benotung Ende der 60er Jahre, wo Lehrende Studierende durch gute Bewertungen vor einer Einberufung in den Vietnam-Krieg bewahren konnten (Rojstaczer/Healy 2012), findet natürlich hierzulande keine politische Entscheidung.

Die folgende Ausarbeitung beruht auf einem DFG-Projekt, das, als ‚Mixed Methods‘ konstruiert, sowohl Längsschnittanalysen als auch Gruppendiskussionen und eine Auswertung der amtlichen Prüfungsstatistik beinhaltet (siehe auch den Beitrag von Müller-Benedict/Gaens in diesem Band).² Der Beitrag hier konzentriert sich methodisch auf letzteres und ist inhaltlich weniger längs- denn querschnittlich angelegt. Er enthält zunächst eine auf die Fragestellung übertragbare knappe Literaturdarstellung, gefolgt von einer Beschreibung des Datenmaterials, ausgewählten Ergebnissen sowie deren Diskussion und schließt mit einer Zusammenfassung ab.

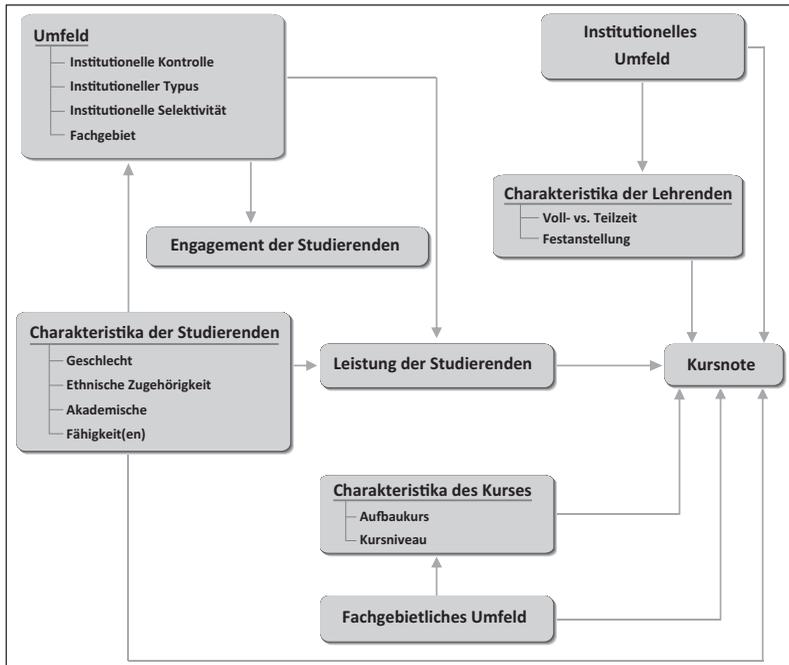
Welche Einflüsse werden thematisiert, wie wurden sie operationalisiert?

Hochschulnoten können als das Zusammenspiel zahlreicher Einflussgrößen angesehen werden. Auf der – bisher zumeist untersuchten – Ebene einer Lehrveranstaltung werden thematisiert: die Eigenschaften des/der

² Der Beitrag ist die stark überarbeitete Schriftvariante eines Vortrags, der – damals noch ohne Integration der Personalstatistik – von Florence Baillet und mir auf der GFHF-Tagung gehalten wurde. Ich bedanke mich besonders bei Frau Baillet, die bis Herbst 2014 im Projekt beschäftigt war. Zahlreiche wertvolle Anregungen kamen von den anderen Mitgliedern des Teams, Thomas Gaens, Volker Müller-Benedict, Margarita McGrory, Elena Tsarouha sowie den TeilnehmerInnen der Jahrestagung der GFHF 2014 und der Wissenschaftlichen Kommission Hochschulmanagement des Verbands der Hochschullehrer für Betriebswirtschaftslehre im Feb. 2014 in Duisburg. – Das DFG-Projekt „Die Notengebung an Hochschulen in Deutschland von den 1960er Jahren bis heute. Trends, Unterschiede, Ursachen“ hatte eine Laufzeit von drei Jahren und wurde zum September 2015 beendet.

Studierenden und das Engagement dieser. Dazu kommen institutionelle Rahmenbedingungen wie Fachcharakteristika, Kurs- und Prüfungsspezifika – z.B. die Veranstaltungsgröße oder ob Klausur oder nicht –, Besonderheiten der Lehrenden – z.B. überwiegend professoral oder nicht –, und der Institution, etwa ihre Selektivität bei der Auswahl der Studierenden oder inwieweit öffentlich bzw. privat (s. Abb. 1).

Abbildung 1: Einflüsse auf Kursnoten



Nach Hu (2005: 11)

Die deutsche amtliche Statistik, die Grundlage der Ausarbeitung hier ist, hat gegenüber einer solchen veranstaltungsbezogenen Untersuchungsebene etliche Beschränkungen. Vor allem (1) zeigt sie nur die Abschlussgesamtnote, sodass Einzelbewertungen und deren Umstände außen vor bleiben, (2) enthält sie bloß wenige Informationen über die akademischen Fähigkeiten der Studierenden. Vor allem die Abiturnote, die trotz aller Problematik (Müller-Benedict 2011) immer noch die beste Prognostik für einen Studienerfolg bietet ist hier leider nicht enthalten. Die deutsche Notenstatistik beinhaltet aber eine ganze Reihe nützlicher Informationen sowohl über die Eigenschaften der Examinierten als auch zu den die Noten

vergebenden Facheinheiten, dazu kommen einige von uns selbst erhobene Größen. Da das Gebiet der Notenerklärung in Deutschland noch weitgehend Terra Incognita ist, ist der Ansatz eher Strukturen entdeckend denn Zusammenhänge belegend.

Neben den unten dargestellten üblicheren Variablen wie z.B. das Alter der Examinierten oder die Personalrelation interessieren uns dabei besonders zwei in der Diskussion bisher eher weniger thematisierte Dimensionen:

A beeinflusst die wahrgenommene Bedeutungsposition der Lehrenden die Benotung und

B spielen die Arbeitsmarktbedingungen eine Rolle?

Beide können als besonders tabuisierte Dimensionen angesehen werden. Dass Bewertende ihre Notengebung vielleicht auch von ihrem eigenen Bedeutungshorizont abhängig machen könnten würde wohl in Diskussionen mit Lehrenden ebenso bestritten werden wie, dass man offen gewillt sein könnte, die Note nicht nur nach rein innerakademischen Standards zu vergeben sondern auch die Interessen späterer Arbeitgeber zu berücksichtigen.

Da wir weiter davon ausgehen, dass bei der Notengebung Fachcharakteristika eine ganz entscheidende Rolle spielen, die nicht nur die Stärke von Einflussgrößen sondern auch deren Wirkrichtung beeinflussen können, werden hier exemplarisch nur die Ergebnisse eines einzigen Faches dargestellt und diskutiert. Die Wirtschaftswissenschaften (in der Klassifikation der Hochschulstatistik: STB 26) haben dabei den Vorteil, ein von der AbsolventInnenanzahl her großes Fach mit vielen Standorten darzustellen, das sowohl an Fachhochschulen wie Universitäten gelehrt wird, sodass auch unterschiedliche institutionelle Einflüsse wohl repräsentiert sind.

Was man mit unserem Datensatz – amtliche Statistik sowie von uns aus anderen Quellen zugespielte bzw. eigens erhobene Daten – untersuchen kann, sind bei den Studenten die Eigenschaften: Geschlecht, Alter, ob man das Abitur im Ausland gemacht hat, ob man zwar das Abitur hierzulande abgelegt hat, aber eine ausländische Staatsbürgerschaft besitzt, ob man vorher eine Berufsausbildung abgeschlossen hatte und ob man aus der Nähe des Studienortes stammt. Daraus ergeben sich folgende Überlegungen. Im weitesten Sinne sind diese als leistungsbezogene Einflüsse zu kategorisieren:

■ Man kann erwarten, dass im Vergleich der Geschlechter die Einflussgröße *Männer* notenverschlechternd ist, da bei diesen durchschnittlich

schlechtere Abiturnoten beobachtet werden (Helbig 2012). Wie bei Regressionen üblich wurde dazu eine ‚Dummy‘-Variable konstruiert, bei der bei Vorliegen der Eigenschaft eine ‚1‘ oder im anderen Fall (weiblich) eine ‚0‘ steht.

- Eine *ausländische Hochschulzugangsberechtigung* lässt wegen oft geringerer Vorbildung, Akkulturations- und Sprachproblemen sowie verstärkten finanziellen Herausforderungen eine schlechtere Examensnote vermuten (Rech 2012).

- *Jüngerer Alter* beim Abschluss und

- eine *abgeschlossene Berufsbildung* sind vermutlich ebenfalls implizite Leistungsindikatoren und sollten deshalb notenverbessernd sein (Volkwein et al. 2000).

- Dagegen wird oft vermutet, dass Studierende aus dem *engeren regionalen Umfeld* weniger intrinsisch motiviert und deshalb leistungsschwächer sein könnten (Allmendinger 2003).

- Die Kombination deutsches Abitur und ausländische Staatsbürgerschaft ist schwer einzuschätzen. Man kann einerseits beobachten, dass die schulischen Bildungsaspirationen von *MigrantInnen* erstaunlich hoch sind (Relikowski/Yilmaz/Blossfeld 2012), andererseits wird dies nur sichtbar, wenn der sozio-ökonomische Status dabei rechnerisch neutralisiert wird. Da eine Herkunft aus eher nicht-akademischen Haushalten aber relativ häufiger sein dürfte, ist die Wahrscheinlichkeit hoch, dass bei dieser Gruppe wegen schlechterer Ausgangsbedingungen ein notenverschlechternder Einfluss sichtbar ist.

Einige der genannten individuellen Informationen lassen sich auch nutzen um noch einmal als Aggregat die spezifischen lokalen Charakteristika der Studienbedingungen näher zu bestimmen.

- Ein höherer Frauenanteil unter den Studierenden könnte auch den Männern helfen, da hochschulisches Lernen stark von den Peers abhängt (Pascarella/Terenzini 2005), ein stärkerer *Männeranteil* würde umgekehrt also als notenverschlechternd eingeschätzt.

- Ebenso ist ein starker *Anteil von Regionalstudierenden* eventuell negativ konnotiert, da hier ceteris paribus im Durchschnitt der StudentInnen weniger Interesse an den Inhalten des Fachs besteht.

- Der *Anteil von Studierenden mit ausländischer Studienberechtigung* könnte einerseits ebenfalls einen negativen Einfluss beinhalten, wenn die These der schlechteren Note bei dieser Gruppe stimmte, aber er könnte auch seitens der Lehrenden als Zeichen von erfolgreicher Internationalität

gelesen werden (möglicher Bestandteil der vorne benannten Dimension A), was dann eventuell durch allgemein bessere Noten honoriert würde.

■ Als wichtiges lokales Charakteristikum wurde weiter die *Anzahl der jährlichen AbsolventInnen* berechnet und zwar einmal *einfach* und dann als *quadratischer Term* um möglichen Nicht-Linearitäten auf die Spur zu kommen. Der Einfluss der Fachgröße wird als verschlechternd eingeschätzt (Kokkelenberg/Dillon/Christy 2008), vermutlich aber bei weiter steigenden AbsolventInnenzahlen in der Bedeutung abnehmend. Technisch gesprochen: der lineare Term sollte ein positives Vorzeichen, der quadratische ein negatives aufweisen und der quadratische Term klein genug sein um in der Summe der beiden Einflüsse nicht insgesamt ein negatives und damit notenverbesserndes Ergebnis zu zeitigen. Offen jedoch ist, ob bei einem solchen prognostizierten Bild damit eher objektiv verschlechterte Lernbedingungen erfasst oder eher den Prüfenden ‚Überfüllung‘ signalisiert wird.

Bei den Charakteristika der Lehrenden haben wir nur wenige sichere Informationen. Nebenberufliches Personal etwa kann je nach Bundesland unterschiedlich definiert werden (Destatis 2012) und wird deshalb hier nicht berücksichtigt, obwohl die unsichere Arbeitssituation vielleicht Einfluss auf die Notengebung aufweist (Sonner 2000).

■ Weiter darf nach amerikanischen Untersuchungen vermutet werden, dass Frauen besser bewerten (Jewell/McPherson 2012). Allerdings war es für Frauen auch schwieriger Feststellungen zu erhalten und Studienendenbewertungen entsprechend wertvoller. Wir nutzen hier den *weiblichen Anteil an allen ProfessorInnen* um diese Überlegung zu testen.

■ Zusätzlich enthält der Datensatz auch das *Durchschnittsalter der ProfessorInnen*. Hier ist die US-Literatur wenig hilfreich, da Alter dort fast immer in Relation zum Lehrenden-Status interpretiert wird, also festgestellt oder nicht und eine deutsche Literatur ist uns trotz der Selbstverständlichkeit in der Nutzung des Begriffs ‚Altersmilde‘ nicht bekannt. Die zu klärende Frage ist: gibt es einen solchen Effekt in der Bewertung?

■ Wenn ja, könnte sich das zusätzlich auch noch an der *Relation von Wissenschaftlichen MitarbeiterInnen auf Haushaltsstellen zu ProfessorInnen* zeigen. Denn diese Gruppe lehrt ja ebenfalls, ist erheblich jünger und der Indikator sollte dann notenverschlechternd wirken.

■ Anders ist es mit dem *Anteil der Wissenschaftlichen MitarbeiterInnen aus Drittmitteln zu ProfessorInnen*. Eine der hier vertretenen und zu untersuchenden Thesen ist (sie gehört zu der vorne angesprochenen Dimension A): je forschungsorientierter sich ein Fach in der Wahrnehmung der

Lehrenden darstellt, desto mehr strahlt diese Positionierung auch auf die Bewertung studentischer Leistungen aus. Der genannte Indikator sollte deshalb notenverbessernd wirken.

■ Das gleiche gilt für den *Anteil der höheren Abschlüsse Master und Promotionen* an allen Abschlüssen in dem betrachteten Zeitraum (hier: Jahr).

■ Auch auf individueller Ebene sollte diese relative Skala wirken: ein *Bachelor* würde dann – immer im Vergleich zum Diplom – als weniger wissenschaftlich interpretiert werden und deshalb relativ schlechtere Noten ergeben, *Master* und *Promotion* dagegen höhere.

Vom institutionellen Umfeld her haben wir nur wenige Entscheidungen getroffen bzw. Variablen definieren können. Einmal wurden Universitäten und Fachhochschulen getrennt ausgezählt. Sie sind einfach zu unterschiedlich – z.B. in der Größe oder der AbsolventInnen-ProfessorInnen-Relation – um sinnvoll in ein einheitliches Modell integrierbar zu sein. Die Berechnungen sind identisch, wobei bei den Universitäten zusätzlich noch die nur dort vorhandenen Variablen *Promotionen*, *Promotionsanteil* und die *Dummys* für die Sonderfächer *Volkswirtschaftslehre* und *Wirtschaftspädagogik* aufgenommen wurden.

■ Unterschieden wurde dagegen nach *privat-öffentlich*, wobei man bei privaten Einrichtungen bessere Noten erwarten darf (Rojstaczer/Healy 2012).

■ Dann wurde noch eine Relation gebildet aus den Absolventen dieses Fachs zu allen Absolventen der jeweiligen Hochschule („*Fachbedeutung*“). Ähnlich wie bei der Forschungsbedeutung erwarten wir, dass die Machtposition eines Fachs innerhalb ihrer Institution eine positive Wirkung auf die Selbstwahrnehmung der Lehrenden ausübt, was dann auf die Studierenden als Qualitätsmerkmal übertragen wird (ein weiterer möglicher Bestandteil von Dimension A).

■ Weiter wurde mit der *Arbeitslosenquote in der Raumordnungsregion* noch eine jährliche regionalbezogene Arbeitsmarktvariable aufgenommen, da davon ausgegangen werden kann, dass durchaus Beziehungen zwischen dem Verhalten von PrüferInnen und Studierenden und den Einkommens- und Karrierechancen bestehen (Freeman 1999). Zumeist sind damit Reaktionen auf Veränderungen in der Zeit gemeint (Paul/Chris/Ruhland 2013), aber hier kann dank der Regionalzeiger im Datensatz auch die räumliche Differenziertheit miteinfließen (die oben angesprochene Dimension B). Die Nutzung der knapp 100 Raumordnungsregionen statt der über 400 Kreise/Städte erklärt sich damit, dass dadurch

Pendlerverflechtungen stärker ausgeschlossen werden können, die sonst die regionale Arbeitslosenquote stark verzerren. Auch in vielen anderen sozialwissenschaftlichen Analysen wird die Ebene der Raumordnungsregionen bevorzugt (Grözinger/Matiaske 2005). Offen bleiben muss die vermutliche Richtung eines möglichen Einflusses, da diese vielleicht nach Fach verschieden sind. Man kann nach den Einflüssen über die Zeit vermuten, dass die Lehrenden in den Wirtschaftswissenschaften eher eine selektionsorientierte Rolle einnehmen und das Vorzeichen in einer Regressionsberechnung deshalb positiv sein dürfte (siehe den Beitrag von Müller-Benedict/Gaens in diesem Band).

Schließlich soll noch erwähnt werden, dass wir ursprünglich vorhatten, auch die Hochschulfinanzstatistik in den Datensatz zu integrieren. Leider ist diese bis heute nicht in der nötigen Detailliertheit fertiggestellt und steht somit der Forschung nicht zur Verfügung.

Das Datenmaterial

Die Ausarbeitung basiert auf einer Auszählung der amtlichen Prüfungsstatistik, in der für jede bestandene Prüfung an einer deutschen Hochschule – im Prinzip jedenfalls – ab 1995 ein eigener Eintrag angelegt wird. Aus Datenschutzgründen sind diese Informationen nur an Forschungsdatenzentren der Statistischen Ämter der Länder (FDZ) zugänglich, wobei für uns das FDZ Kiel/Hamburg zuständig ist. Die Berechnungen wurden mit der dort vorhandenen Statistiksoftware IBM SPSS durchgeführt. Wir untersuchen die Periode 1995 – 2010, wobei 1995 den Beginn dieser Statistik darstellt und 2010 das zum Zeitpunkt der Antragstellung letztverfügbare Jahr.

Insgesamt beinhaltet der uns zur Verfügung stehende Datensatz etwas über 3,8 Mio. Zeilen. Für den Bereich Wirtschaftswissenschaften umfasst der Datensatz immerhin noch mehr als eine halbe Million Eintragungen. Davon sind allerdings noch knapp 10.000 jeweils in den für hier unwichtigen Unterkategorien ‚Endgültig nicht bestanden‘, ‚Bestanden, Gesamtnote unbekannt‘ eingeordnet. Von diesen Ausnahmen abgesehen, handelt es sich also um eine Vollerhebung mit sämtlichen Prüfungsarten.

Für die AbsolventInnen mit Noteninformation liegen (leider nur) ganzzahlige Noten vor. Für Promotionsabschlüsse, die ja keine Vier aber dafür ein ‚mit Auszeichnung‘ kennen, wurden diese folgendermaßen umkodiert (Summa Cum laude = 1, Magna Cum Laude = 2, Cum laude = 3, Rite = 4), einige wenige Summas bei den anderen Abschlüssen der 1 zu-

geschlagen. Im Ergebnis lautet dadurch die Verteilung: 1 = 10,1 %; 2 = 56,0 %; 3 = 29,0 %; 4 = 1,5 %.

Zu den Notendaten haben wir am FDZ Informationen der hochschulischen Personalstatistik zugespielt, die Vollzeitäquivalente ausgibt. Diese Informationen liegen erst ab 1997 vor, so dass dadurch leider weitere Fälle verloren gehen. Auch sind, besonders am Anfang der Personalstatistik, relativ viele fehlende Werte zu verzeichnen, selbst wenn man notorisch unzuverlässige Größen wie studentische MitarbeiterInnen oder Lehrbeauftragte wegen dieser Problematik in der Regression nicht berücksichtigt.

Schließlich wurden auch noch einige wenige unplausible Werte wg. vermutlicher Fehlkodierung herausgefiltert: ein Studienabschluss im Alter von unter 19 Jahren wurde als ebenso wenig aussagekräftig angesehen wie eine AbsolventInnen-ProfessorInnen-Relation von unter 3 oder über 100. Letztere Fälle betreffen etwa im Entstehen begriffene Studiengänge oder private Fachhochschulen, wo manchmal sehr wenig ProfessorInnen den Abschlüssen zugeordnet waren, auch wenn sie eventuell in anderen organisatorischen Einheiten dieser Hochschule tätig sind. So verbleiben 199.250 Eintragungen für die Universitätsabschlüsse und 153.025 für die der FHs.

Zu jeder Hochschule haben wir selbst noch die Information erhoben und zugespielt, ob öffentlich oder privat, wobei kirchliche Einrichtungen als öffentlich kodiert wurden. In den Datensatz integriert wurden weiter die regionalen Arbeitslosigkeitsdaten, die ebenfalls seit 1997 jährlich vorliegen, seit kurzem sogar online abrufbar.³ Die Werte werden nach offiziellen Quellen von der Bundesanstalt für Bau-, Stadt- und Regionalplanung zusammengestellt, von dieser haben wir auf Anfrage auch Distanzmatrizen für die durchschnittliche Entfernung zwischen dem Kreis der Hochschulzugangsberechtigung und dem Sitzkreis der Hochschule erhalten, die wir ebenfalls zugespielt haben. Bei einer Herkunft von bis zu 50 km Luftlinie wurde von einer Pendlermöglichkeit ausgegangen, diese Gruppe wurde als ‚*RegionalstudentIn*‘ gekennzeichnet.

Wirtschaftswissenschaften beschreiben ein Feld mit mehreren fachlichen Unterteilungen. Am weitaus stärksten besetzt ist für Universität und FHs zusammen und über alle Jahre seit 1997 mit 61,5 % die Kategorie Betriebswirtschaftslehre (Bwl), dazu kommen noch Internationale Bwl/Management mit 5,9 %. Es folgen mit 19,7 % Wirtschaftswissenschaften. Dieses ist ein Mischfach, in dem Studierende eigene Schwerpunkte wählen können, wobei davon auszugehen ist, dass ebenfalls überwiegend eine

³ <http://www.inkar.de/>

betriebswirtschaftliche Spezialisierung gesucht wird. Das Gleiche gilt für wirtschaftswissenschaftliche Sonderkategorien für Tourismus, Verkehr, Sport. Als inhaltlich etwas anders aufgestellt könnten nur die beiden universitären Kategorien Volkswirtschaftslehre (6,2 %) und Wirtschaftspädagogik (2,3 %) gelten. Sie wurden deshalb jeweils wieder mit einem Dummy berücksichtigt.

Von den Abschlüssen her gibt es ebenfalls eine klare Hierarchie. Die mit Abstand stärksten sind die Kategorien Dipl. (Uni) mit 38,8 % und Dipl. (FH) mit 38,1 % der Fälle. Einen Bachelor haben 11,7 %, einen Master 4,4 % und 3,0 % eine Promotion. Der Rest verteilt sich auf verschiedene Lehramtsstudiengänge und einige wenige Magister etc. Wir haben hier Diplom und Sonstige zur Basiskategorie definiert, Bachelor, Master und Promotion sind über Dummies einzeln gekennzeichnet. Da die verschiedenen Abschlüsse zu unterschiedlichen Zeiten im Lebenslauf durchschnittlich anfallen und auch unterschiedlich lang dauern, wurde das Alter der AbsolventInnen nicht absolut genommen sondern als Relation zum Durchschnitt ihrer jeweiligen Abschlussart gebildet.

Tabelle 1: Mittelwerte und das 10te/90te Perzentil der Variablen

| Variable | Universität | | | Fachhochschule | | |
|---|-------------|------------|-------|----------------|------------|-------|
| | 10 P. | Mittelwert | 90 P. | 10 P. | Mittelwert | 90 P. |
| Männer | ,48 | ,59 | ,70 | ,35 | ,48 | ,60 |
| Dt. Abitur, ausld. Staatsangehörigkeit | ,00 | ,03 | ,00 | ,00 | ,03 | ,07 |
| Ausländisches Abitur | ,02 | ,08 | ,17 | ,00 | ,06 | ,15 |
| RegionalstudentIn | ,05 | ,36 | ,62 | ,08 | ,45 | ,71 |
| Relatives Alter | ,87 | 1,00 | 1,12 | ,87 | 1,00 | 1,15 |
| Berufsausbildung | ,00 | ,12 | ,12 | ,00 | ,29 | ,53 |
| Volkswirtschaftslehre | ,00 | ,11 | ,09 | – | – | – |
| Wirtschaftspädagogik | ,00 | ,05 | ,00 | – | – | – |
| Bachelorabschluss | ,00 | ,09 | ,00 | ,00 | ,13 | ,39 |
| Masterabschluss | ,00 | ,04 | ,12 | ,00 | ,05 | ,15 |
| Promotionsabschluss | ,02 | ,06 | ,09 | – | – | – |
| Fachbedeutung | ,07 | ,21 | ,44 | ,19 | ,42 | ,74 |
| Absolventenanzahl/Jahr | 38 | 249 | 552 | 57 | 205 | 382 |
| Privat | ,00 | ,02 | ,00 | ,00 | ,03 | ,00 |
| AbsolventInnen pro ProfessorIn | 8,6 | 17,0 | 27,4 | 4,4 | 8,6 | 13,0 |
| Durchschnittsalter ProfessorInnen | 46,0 | 50,4 | 54,5 | 46,3 | 50,0 | 52,3 |
| Anteil Frauen an allen ProfessorInnen | ,00 | ,08 | ,17 | ,05 | ,14 | ,24 |
| MitarbeiterInnen (Haushaltsstellen)/ ProfessorIn | 1,5 | 2,7 | 4,0 | ,00 | 0,15 | ,34 |
| MitarbeiterInnen (Drittmittel)/ProfessorIn | 0,08 | ,6 | 1,2 | ,00 | ,03 | ,08 |
| Regionale Arbeitslosenquote | 4,1 | 7,1 | 11,4 | 3,8 | 7,0 | 12,2 |

Tabelle 1 stellt die Mittelwerte der erklärenden Variablen dar (die in der Regression zusätzlichen aggregierten Größen der studentischen Zusammensetzung wie etwa der Ausländeranteil sind identisch mit dem Mittelwert der Individualinformation und wurden deshalb weggelassen). Aus Datenschutzgründen ist es teilweise schwierig, auch die Minima angeben zu können. Ersatzweise wurden die Grenzen zum untersten und zum obersten Zehntel (10tes und 90stes Perzentil) pro Jahr und Hochschule berechnet. Das führt bei sehr ungleichen Verteilungen dazu, dass bei einzelnen Kategorien die obere Angabe dem Mittelwert ähnelt oder sogar darunter liegt. So ist z.B. ein Großteil der Volkswirtschaftsausbildung in sehr wenigen Einrichtungen konzentriert.

Universitäten und Fachhochschulen unterscheiden sich in der Mehrheit der Bereiche nur wenig, in manchen aber erheblich. So haben an FHs mehr als doppelt so viele AbsolventInnen einen vorherigen Berufsabschluss als an Unis und auch der RegionalstudentInnen-Anteil ist um einiges größer. Beides war erwartbar. Die Anzahl der Geprüften unterscheidet sich wieder nicht so stark, aber da Fachhochschulen durchschnittlich kleiner sind, ist die Fachbedeutung (wirtschaftswissenschaftliche Abschlüsse zu allen Abschlüssen) dort erheblich größer. Bedeutende Differenzen finden sich bei den personalbezogenen Informationen. FHs haben von ihrer Anlage her kleinere Klassen, und die ProfessorInnen sind die Hauptlehrenden. Daraus folgt ein niedriges AbsolventInnen-ProfessorInnen-Verhältnis. Dazu passend sind die beiden Relationen von MitarbeiterInnen zu ProfessorInnen viel geringer als an Universitäten.

Da alle quantitativen Informationen zumindest Intervallqualität besitzen oder in der Form von Dummy-Variablen vorliegen, ist eine klassische Kleinstquadrate-Regression (oder: OLS-Regression) die statistische Methode der Wahl. Sie berücksichtigt viele Kontrolldimensionen und die Kenntnisse darüber sind weitverbreitet, was eine Kommunikation der Ergebnisse erleichtert.

Da wir mit der Arbeitslosigkeit auch eine Variable auf der Regional-ebene einsetzen, wäre im Prinzip auch an ein Multi-Level-Modell zu denken. Aber das beträfe nur relativ wenige Fälle, wo in der gleichen Region mehrere Hochschulen wirtschaftswissenschaftliche Abschlüsse aufweisen und ist in den Ergebnissen viel schwieriger darstellbar, so dass davon Abstand genommen wurde. Tabelle 2 gibt die Regressionswerte wieder. Zusätzlich zu den erwähnten Variablen wurde noch die zeitliche Entwicklung in Form der Jahre berücksichtigt, um nach einem eventuellen Trend jenseits der strukturellen Einflüsse zu fragen.

Tabelle 2: Ergebnisse der Regression

| Name | Universitäten | | Fachhochschulen | |
|---|---------------|-------------|-----------------|-------------|
| | Beta | Signifikanz | Beta | Signifikanz |
| Männer | ,011 | *** | ,072 | *** |
| Dt. Abitur, ausld. Staatsangehörigkeit | ,061 | *** | ,071 | *** |
| Ausländisches Abitur | ,121 | *** | ,115 | *** |
| RegionalstudentIn | ,035 | *** | ,020 | *** |
| Relatives Alter | ,186 | *** | ,144 | *** |
| Berufsausbildung | -,023 | *** | -,029 | *** |
| Volkswirtschaftslehre | -,021 | *** | – | – |
| Wirtschaftspädagogik | -,005 | ** | – | – |
| Bachelorabschluss | ,056 | *** | -,028 | *** |
| Masterabschluss | -,056 | *** | -,110 | *** |
| Promotionsabschluss | -,060 | *** | – | – |
| Fachbedeutung | -,045 | *** | -,053 | *** |
| Absolventenanzahl | ,136 | *** | ,103 | *** |
| Absolventenanzahl² | -,035 | – | -,154 | *** |
| Privat | -,043 | *** | -,031 | *** |
| Männeranteil | ,073 | *** | ,030 | *** |
| AusländerInnenanteil | -,062 | *** | -,015 | *** |
| Masteranteil | -,052 | *** | -,049 | *** |
| Promovierendenanteil | -,043 | *** | – | – |
| RegionalStud.-Anteil | ,038 | *** | ,041 | *** |
| AbsolventInnen pro ProfessorIn | -,009 | *** | ,007 | ** |
| Durchschnittsalter ProfessorInnen | -,033 | *** | ,017 | ** |
| Anteil Frauen an allen ProfessorInnen | ,016 | *** | -,019 | *** |
| MitarbeiterInnen (Haushaltstellen)/Prof. | ,009 | *** | -,014 | *** |
| MitarbeiterInnen (Drittmittel)/Prof. | -,054 | *** | -,026 | *** |
| Regionale Arbeitslosenquote | ,028 | *** | ,085 | *** |
| Zeit | -,061 | *** | ,109 | *** |
| N | 199.250 | | 153.025 | |
| R² | 0,11 | | 0,08 | |

Diskussion der Ergebnisse

Als erstes fällt auf, dass sowohl für Universitäten wie für Fachhochschulen die erklärte Varianz nicht besonders hoch ist. Hier macht sich vor allem wohl die im Datensatz fehlende Abiturnote bemerkbar. Die Bedeutung dieser leider in der amtlichen Prüfungsstatistik nicht mitaufgenommenen Größe wird an einem Vergleich deutlich: in den USA hat eine gro-

ße Untersuchung mit 1,5 Mio. Fällen von Erststudiums-Abschlussnoten ergeben, dass der – einer Abiturnote in der Aussagekraft in etwa ähnliche und ebenfalls gegen Ende der Schulzeit abgelegte – Durchschnittswert des Scholarly Aptitude Test (SAT) als einzige erklärende Variable schon zu einem R^2 von 0,39 (öffentliche Hochschulen) bis 0,62 (private) für die Erklärung des addierten Durchschnittsanteils von Sehr gut/Gut führte (Rojstaczer/Healy 2012).

Auf der individuellen Ebene wird der Zusammenhang von Noten und SAT etwas schwächer, ist aber immer noch deutlich sichtbar (Camara/Echternacht 2000). Vergleichbare Vorhersagewerte erzielt man natürlich auch für das Abitur in Deutschland (Köller 2013). Dies gegeben, stellt ein R^2 für eine Berechnung ohne Abiturnote von um die 0.1 zwar alles andere als ein Bestergebnis dar, bedeutet aber noch keinen Hinderungsgrund, Richtung und Stärken von Einflussfaktoren zu diskutieren. Auch die Regressionsergebnisse der boomenden Glücksforschung etwa liegen in der Regel in einem ähnlichen Wertebereich (z.B. Grözinger/Matiaske 2014).

Von den Individualmerkmalen beschreibt das relative Alter und ob man aus dem Ausland kommt einen besonders starken Einfluss. Alle anderen, also das Geschlecht, ob man eine Berufsausbildung vorher abschloss, ob man das Abitur in Deutschland mit ausländischer Staatsangehörigkeit ablegte oder ob man RegionalstudentIn ist, sind ebenfalls hochsignifikant und zeigen auch das erwartete Vorzeichen. Damit verlässt man aber schon das Feld, in dem Notenunterschiede recht sicher durch unterstellte Leistungsunterschiede erklärt werden.

In einem etwas weiteren Sinne kann noch der Einfluss der Zusammensetzung der Studierenden dazu gezählt werden. Hier ergibt sich jedoch ein gemischtes Bild: ein größerer Männeranteil wirkt – wie angenommen – notenverschlechternd, ein höherer RegionalstudentInnenanteil dito. Bei dem AusländerInnenanteil scheint dagegen der Einfluss durch den dadurch erzeugten Ausweis von Internationalität gegenüber einem eventuell zu vermutenden verschlechternden Einfluss auf die Leistungsbereitschaft der anderen Studierenden zu überwiegen.

Schwer mit Leistungsunterschieden erklärbar ist die Wirkung von Fachbedeutung: Wenn eine Fakultät an ihrer Einrichtung relativ viele AbsolventInnen hat, wird dadurch vermutlich ein positives Gefühl hoher Bedeutung unter den Prüfenden erzeugt, von dem auch die Geprüften profitieren. Die Alternativhypothese, man habe sich dadurch vielleicht mehr Mittel zur Verbesserung der Lernbedingungen sichern können, ist mangels der Finanzdaten zwar nicht direkt widerlegbar, das sollte sich aber auch in der Personalausstattung widerspiegeln und für diese ist in der

wichtigsten Dimension der Mitarbeiter-/ProfessorInnenrelationen kontrolliert.

Weiter wird deutlich, welche große Bedeutung die Anzahl der Geprüften pro Einrichtung hat. Der sich abschwächende negative Einfluss der ‚großen Zahl‘ (signifikant nur bei den Fachhochschulen), und sichtbar beim Minusvorzeichen des Quadrats der AbsolventInnenzahl ist auch nur eine Abschwächung und darf nicht bei größeren Einrichtungen als Umkippen ins Positive interpretiert werden. Es liegt dieser hypothetische Punkt erheblich über den empirisch beobachteten Maxima. Nur bei diesen beiden Größen – linear und Quadrat einer Variable – wird eine Multikollinearität sichtbar, also eine enge Korrelation zweier erklärender Variablen nicht nur mit der zu erklärenden Größe sondern auch miteinander. Diese ist in dem angesprochenen Fall natürlich modellimmanent und in Kauf zu nehmen. Bei allen anderen Variablen wurde ebenfalls auf Multikollinearität getestet. Die dafür berechneten Varianzinflationsfaktoren (‚VIF‘-Werte) lagen jedoch immer im unproblematischen Bereich und sind deshalb hier auch nicht im Detail berichtet.

Der verschlechternde Einfluss hoher AbsolventInnenzahlen auf die Note lässt sich mehrfach deuten und der tatsächliche Zusammenhang mit unseren Daten alleine auch nicht wirklich klären:

- Es könnte erstens so sein, dass die Lernbedingungen an ‚Massenhochschulen‘ schlechter sind, was sich nicht zwingend in der hier kontrollierten Personalausstattung alleine widerspiegeln muss, sondern andere Bereiche wie etwa mangelnde Tutorien beinhalten könnte.
- Zweitens könnten wegen dadurch größerer Teilnehmerzahlen bei Veranstaltungen der Anteil formalerer Prüfungen (vor allem Klausuren mit hohem Multiple-Choice-Anteil) steigen. Bei diesen wird erfahrungsgemäß stärker die Notenskala ausgereizt, zum einen weil mit klaren Punktvorgaben/Schwellenwerten gearbeitet wird und zum anderen die Schwierigkeit eine Note gegenüber sich beschwerenden Studierenden zu verteidigen geringer ist.
- Drittens könnte sich auch die Zusammensetzung von Studierenden jenseits der hier geprüften individuellen Dimensionen in ihrer Leistungsfähigkeit unterscheiden.
- Viertens könnten größere Absolventenzahlen eines Fachs vor allem an größeren Hochschulen zu finden sein, die den Studierenden mehr Ablenkungspotential bieten.
- Fünftens schließlich könnte die Größe alleine auch einen psychologischen Effekt in dem Sinne ausüben, dass bei umfangreichen Kohorten die

Selektionsaufgabe der Bewertung mehr in das Bewusstsein der Prüfenden rückt.

Die letztere Erklärung passt jedenfalls gut zu dem ebenfalls notenverschlechternden Einfluss der regionalen Arbeitslosigkeit, der bei beiden Berechnungen ebenfalls deutlich wird. Die vorne angesprochene Einflussdimension B, dass sich bei den Wirtschaftswissenschaften die Lehrenden in einer Selektionsrolle sehen, hat also einige empirische Plausibilität.

Ökonomisch nachvollziehbar ist auch der notenverbessernde Einfluss des Studiums an einer privaten Einrichtung. Ist im universitären Kontext hier noch die Möglichkeit von ‚Elite‘-Einrichtungen mit einer Bestauswahl der Studierenden denkbar, dürfte bei den Fachhochschulen dieses Argument kaum als valide gelten sondern eher ein Tausch Studiengebühren gegen gute Noten anzunehmen sein.

Bei den Personalrelationen ist erst einmal unerwartet, wie wenig die AbsolventInnen-ProfessorInnen-Relation zur Erklärung der Notendifferenzen beiträgt. Sie ist nicht nur im Vergleich zur Wirkung der AbsolventInnenanzahl sehr gering sondern auch zu vielen anderen hier aufgenommenen Dimensionen.

Ob es eine Altersmilde bei den Prüfenden gibt, ist nicht eindeutig, aber weiter denkbar. Dazu passt, dass diese Größe zumindest bei der Universität mit einem gewissen Einfluss notenverbessernd wirkt (an den FHs mit einem kleineren Signifikanzwert dagegen umgekehrt), viele Haushaltsstellen, und damit jüngere Bewertende, dagegen zumindest im universitären Kontext notenverschlechternd wirken und das Gleiche auch für den Frauenanteil gilt. Zwar war hier der gegenteilige Effekt erwartet worden, aber wenn Professorinnen eher in den letzten Jahren berufen wurden und damit jünger sind, könnte dieser Alterseffekt einen eventuell gegensätzlichen der Geschlechtswirkung überstrahlen.

An den FHs jedenfalls, wo sich schon erheblich mehr weibliche Professoren finden, wechselt das Vorzeichen beim Anteil der Professorinnen. Und der positive Wert bei den (wenigen) Haushaltsstellen auf Fachhochschulseite könnte dort eher als Beleg für Fachbedeutung interpretiert werden als dass das jüngere Alter der hier eventuell in die Lehre Eingebundenen eine Wirkung der Notenverschlechterung hätte.

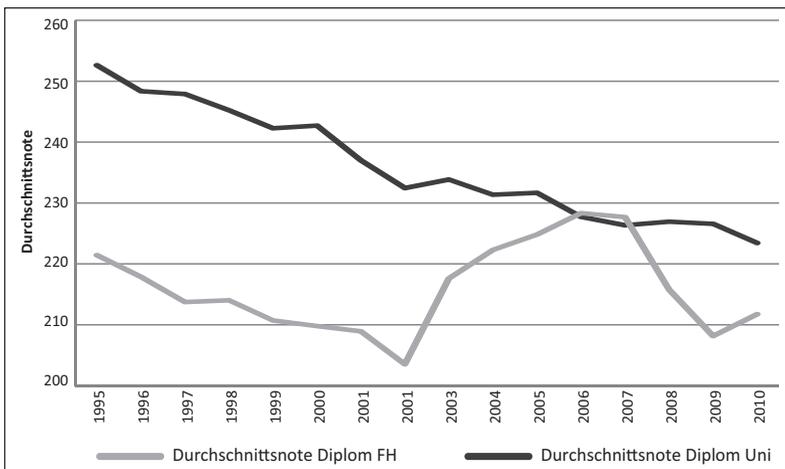
Bei den Abschlüssen gibt es wieder mehr Gemeinsamkeiten. Immer sind die Master in der Note im Schnitt besser als Diplom/Sonstige und das Gleiche gilt auch für die Promotionen an den Universitäten. Das korrespondiert mit der Beobachtung des Wissenschaftsrats:

„Stellt man die Prüfungsnoten der einzelnen Prüfungsgruppen einander gegenüber, so fällt auf, dass die Noten in den Prüfungsgruppen mit vergleichsweise höherem Qualifikationsniveau tendenziell besser ausfallen als die Noten in den Prüfungsgruppen mit vergleichsweise niedrigerem Qualifikationsniveau. Es ist davon auszugehen, dass dies teilweise, z. B. beim Master, auf Selektionsprozesse zurückzuführen ist.“ (Wissenschaftsrat 2012: 41).

Diese letzte Schlussfolgerung ist aber einigermaßen eigenartig: Warum sollte für aufeinander folgende Abschlüsse nicht gelten, dass immer wieder ein neuer Durchschnittswert bestimmt wird, der idealerweise auf allen Ebenen gleich ist? Bei den Schulen würde man ja auch nicht annehmen, dass z.B. in der Oberstufe schon deshalb bessere Noten gegeben werden sollten, weil einige SchülerInnen nach der Mittelstufe abgegangen sind.

Der Bachelorabschluss ist an den Universitäten schlechter als das Diplom bewertet, an den Fachhochschulen dagegen ein wenig besser. Aber hier sind auch die unterschiedlichen Studienzeiten zu beachten: universitär liegt ein 3jähriger BA-Abschluss zwischen einem Vordiplom etwa nach 2 Jahren und einem Hauptdiplom nach 5 Jahren. An der FH ist er nahe an dem alten Diplom von 4 Jahren. In beiden Fällen strahlt ein hoher Anteil von Masterstudierenden (an Universitäten auch von Promovierenden) ebenso positiv auf das allgemeine Notenniveau aus wie viele Drittmittelbeschäftigte. Die vorne benannte Einflussdimension A – Prüfende schätzen auch in der Lehrsituation eine Forschungsorientiertheit und eine hohe Bedeutung ihres Fachs im hochschulischen Kontext – bestätigt sich: Alle dafür stehenden Indikatoren sind hier notenverbessernd.

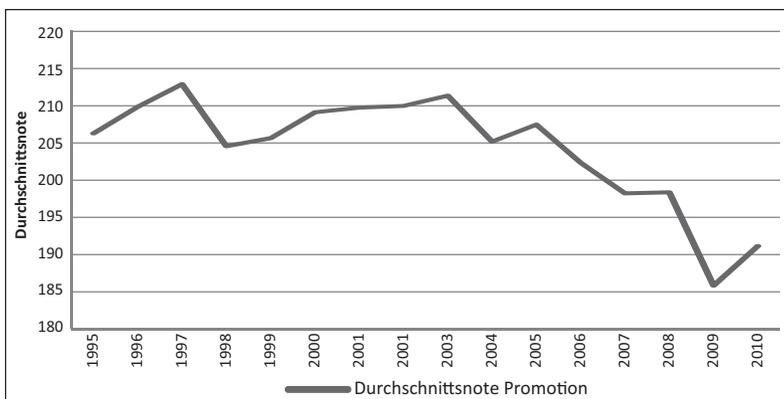
Abbildung 2: Notenentwicklung beim Diplom



Ein großer Unterschied zwischen Universität und Fachhochschule scheint bei der Zeitvariable zu herrschen. Aber von den beiden hochsignifikanten Ergebnissen mit unterschiedlichem Vorzeichen darf man sich nicht zu der Einschätzung verleiten lassen, dass dies für alle Abschlüsse gilt. Die Werte kommen überwiegend durch die Entwicklung bei den jetzt auslaufenden Diplom-Abschlüssen zustande, die ja in dem Datensatz den allergrößten Anteil darstellen. In Abbildung 2 ist der zeitliche Verlauf für das Diplom bei Universitäten und Fachhochschulen für die gesamte zeitliche Periode 1995 – 2010 dargestellt.

Es wird zweierlei an der Abbildung deutlich: An den Universitäten herrscht tatsächlich im Diplom eine strikte Notenverbesserung vor. An den Fachhochschulen überwiegt im Diplom dagegen ein Schwankungseffekt mit leichtem Trendanstieg hin zu schlechteren Noten. Allerdings sind die Ausgangsniveaus unterschiedlich, sodass man statt von einer Noten-inflation bei den Universitäten auch von einem von beiden Seiten betriebenen Angleichungsprozess sprechen kann.

Abbildung 3: Notenentwicklung bei der Promotion



Eine insgesamt eindeutig sich verbessernde Notensituation ist bei den Promotionsabschlüssen beobachtbar (Abb. 3). Allerdings ist der Trend weniger konstant als beim universitären Diplom, sondern setzt erst Mitte der 2000er Jahre ein. Ob dies eventuell mit der Umstellung auf den Bachelor zu tun hat, der der Promotion eine neue Distinktionsbedeutung verlieh, ist denkbar, bleibt aber spekulativ. Wir werden für andere Fächer überprüfen, ob die Promotionsnoten dort ebenfalls einen solchen zeitlichen Verlauf aufweisen.

Die Bologna-Erklärung datiert von 1999, die ersten darauf basierenden Studienprogramme starteten in den Jahren darauf. Die Bologna-Abschlüsse Bachelor und Master selbst beginnen in der Experimentalphase mit oft wenigen Fällen zwar mit sehr guten Noten, stabilisieren sich aber dann schnell auf einem wieder schlechteren Niveau. Berechnet man z.B. für Universitäten eine Regression mit nur der Zeit als erklärender Variable, so erhält man ein leicht positives Beta von $0,07^{***}$ für den BA, von $0,02^{**}$ für den MA. Bei den Fachhochschulen lauten die Werte $0,05^{***}$ im BA-Fall und $-0,08^{***}$ beim MA. Von den vier Fällen der Bologna-Abschlüsse sind in der Entwicklung also drei bisher notenverschlechternd und einer verbessernd. Das kann man sicher nicht einen inflationären Trend nennen.

Zusammenfassung

Die große Frage nach den Einflüssen des sozialen Kontexts auf die Notengebung an deutschen Hochschulen (Müller-Benedict/Tsarouha 2010) kann so beantwortet werden: Sie sind zweifellos vorhanden und sie sind recht mannigfaltig. Am Beispiel der Auswertung der Prüfungsstatistik für die Wirtschaftswissenschaften unter Einschluss zentraler Größen der Personalstatistik und weiteren Informationen ergab sich vor allem, dass das Selektionsklima von Bedeutung ist (Absolventenzahl, regionale Arbeitslosigkeit). Aber auch die wahrgenommene eigene Bedeutung in Punkto Forschung und Durchsetzungsmacht (Anteil der Master- und Promotionsabschlüsse sowie Drittmittelpersonal und relative Bedeutung des Fachs im Hochschulkontext) verbessert die Noten. Dazu passt, dass sich die Durchschnittsnote je Abschlussart parallel zu dem unterstellten Forschungsgehalt nach oben bewegt.

Der Hoffnung, dass Examensnoten eine reine Reflexion studentischer Leistung darstellen, die wenigstens innerhalb eines Fachs leicht zwischen Hochschulen oder Abschlussarten vergleichbar sind, darf man deshalb mit gutem Grund skeptisch gegenüber sein. Dies erklärt, warum an Hochschulen eine ‚heilige Scheu‘ herrscht, bei dem Thema Transparenz herzustellen. Künftige Arbeiten werden analysieren müssen, inwieweit die hier berichteten Ergebnisse sich wie bei anderen Fächern darstellen und auch, inwieweit mit anderen Methoden die konkrete Wirkungsweise der hier mehr beschriebenen denn erklärten Einflüsse gefunden werden kann.

Literatur

- Adorno, Theodor W. (1973): *Erziehung zur Mündigkeit*. Frankfurt am Main: Suhrkamp.
- Allmendinger, Jutta (2003): Soziologie in Deutschland. Standortbestimmung und Perspektiven, in: dies. (Hg.): *Entstaatlichung und soziale Sicherheit*. Verhandlungen des 31. Kongresses der Deutschen Gesellschaft für Soziologie in Leipzig 2002. Teil 1. Opladen: Leske + Budrich, S. 33–50.
- Babcock, Philip (2010): Real costs of nominal grade inflation? New evidence from student course evaluations, *Economic Inquiry* 48 (4), S. 983–996.
- Butcher, Kristin F/McEwan, Patrick J/Weerapana, Akila (2014): The Effects of an Anti-Grade-Inflation Policy at Wellesley College, *Journal of Economic Perspectives* 28 (3), S. 189–204.
- Camara, Wayne J./Gary Echternacht (2000): The SAT® I and High School Grades: Utility in Predicting Success in College. *Research Notes* 10, July 2000; URL <http://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2000-10-sat-high-school-grades-predicting-success.pdf> (3.10.2015)
- Clark, Burton R. (1983): *The higher education system: academic organization in cross-national perspective*. Berkeley/Los Angeles/London: University of California Press.
- Destatis (2012): *Personal an Hochschulen 2011 – Vorläufige Ergebnisse*. Wiesbaden; URL https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/PersonalVorbericht5213402118004.pdf;jsessionid=678ED5D1C4BDE69B271A946F8AFD2629.cae1?__blob=publicationFile (17.10.2015)
- Freeman, Donald G. (1999): Grade Divergence as a Market Outcome, *Journal of Economic Education* (Fall), S. 344–351.
- Grözinger, Gerd/Wenzel Matiaske (2005): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund. München/Mering: Hampp.
- (2014): The direct and indirect impact of religion on well-being in Germany, *Social indicators research* 116 (2), S. 373–387.
- Helbig, Marcel (2012): Die Umkehrung – Geschlechterungleichheiten beim Erwerb des Abiturs im Wandel, in: Rolf Becker/Heike Solga (Hg.), *Soziologische Bildungsforschung* (=Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderheft 52/2012). Wiesbaden: Springer, S. 393–414.
- Hu, Shouping (2005): *Beyond Grade Inflation – Grading Problems in Higher Education*. San Francisco: Jossey-Bass.
- Hunt, Lester H. (Hg.) (2008): *Grade inflation: Academic standards in higher education*. New York: SUNY Press.
- IFQ (2014): Noten sagen kaum noch etwas über die Qualität von Doktorarbeiten aus (Pressemitteilung vom 26. Juni 2014). Berlin; URL http://www.forschungsinfo.de/Presse/dokumente/PM%2006-2014-01_Noten%20sagen%20kaum%20noch%20etwas%20ueber%20die%20Qualitaet%20von%20Doktorarbeiten%20aus.pdf (10.9.2015)
- Jewell, R Todd/Michael A. McPherson (2012): Instructor-Specific Grade Inflation: Incentives, Gender, and Ethnicity, *Social Science Quarterly* 93 (1), S. 95–109
- Johnson, Valen E. (2003): *Grade Inflation. A Crisis in College Education*. New York: Springer.
- Kokkelenberg, Edward C./Michael Dillon/Sean M. Christy (2008): The effects of class size on student grades at a public university, *Economics of Education Review* 27, S. 221–233.

- Köller, Olaf (2013): Abitur und Studierfähigkeit, in: Jupp Asdonk/Sebastian U. Kuhnen/Philipp Bornkessel (Hg.): Von der Schule zur Hochschule. Analysen, Konzeptionen und Gestaltungsperspektiven des Übergangs. Münster: Waxmann, S. 25–49
- Müller-Benedict, Volker/Elena Tsarouha (2011): Können Examensnoten verglichen werden? Eine Analyse zur Stabilität und Vergleichbarkeit von Hochschulnoten. *Zeitschrift für Soziologie* 5/2011, S. 388–409.
- Ogilvie, Kristie/Matthew Jelavic (2013) Grade inflation in the US higher educational environment: a faculty perception study, *International Journal of Management in Education* 7 (4): S. 406–416.
- Pascarella, Ernest T./Patrick T. Terenzini (2005): How College Affects Students. Volume 2. A Third Decade of Research. San Francisco: Jossey-Bass.
- Paul, Chris W./Joseph S. Ruhland (2013): A Note On Job Market Conditions And Students' Academic Performance, *Journal of Business & Economics Research (JBER)* 11 (5), S. 223–228.
- Rech, Jörg (2012): Studienerfolg ausländischer Studierender. Eine empirische Analyse im Kontext der Internationalisierung der deutschen Hochschulen. Münster: Waxmann.
- Relikowski, Ilona/Erbil Yilmaz/Hans-Peter Blossfeld (2012): Wie lassen sich die hohen Bildungsaspirationen von Migranten erklären? Eine Mixed-Methods-Studie zur Rolle von strukturellen Aufstiegschancen und individueller Bildungserfahrung, in: Rolf Becker/Heike Solga (Hg.), *Soziologische Bildungsforschung* (=Kölner Zeitschrift für Soziologie und Sozialpsychologie Sonderheft 52/2012). Wiesbaden: Springer, S. 111–136.
- Rojstaczer, Stuart/Christopher Healy (2012): Where A is ordinary: The evolution of American college and university grading, 1940–2009. *Teachers College Record* 114 (7), S. 1–23.
- Sonner, Brenda S. (2000): A is for “adjunct”: Examining grade inflation in higher education, *Journal of Education for Business* 76 (1), S. 5–8.
- Volkwein, J. Fredericks/Sondra Valle/Kelli Parmley/Gary Blose/Ying Zhou (2000): A Multi-Campus Study of Academic Performance and Cognitive Growth among Native Freshman, Two-Year Transfers, and Four-Year Transfers. AIR 2000 Annual Forum Paper. Washington: ERIC Clearinghouse.
- Wissenschaftsrat (2012): Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar des Wissenschaftsrates. Hamburg; URL <http://www.wissenschaftsrat.de/download/archiv/2627-12.pdf> (17.10.2015)