

Noteninflation an deutschen Hochschulen – Werden die Examensnoten überall immer besser?

Thomas Gaens

Die langfristige Entwicklung von Abschlussnoten an deutschen Hochschulen ist aufgrund des bisherigen Mangels an geeigneten Daten im Zeitreihenformat noch weitestgehend unerforscht. Dennoch wird in den Medien häufig das Bild einer umfassenden Noteninflation gezeichnet. Dieser Beitrag fasst zunächst die bislang verfügbaren empirischen Ergebnisse zur Notenentwicklung im Zeitverlauf an deutschen Hochschulen zusammen. Anschließend wird anhand erstmals verfügbarer Daten die langfristige Entwicklung der Notengebung an Hochschulen in Deutschland nachgezeichnet. Es zeigt sich, dass langfristige Verbesserungen nicht in allen Studiengängen und an allen Hochschulen auftreten.

1 **Grade Inflation – Ein Phänomen und seine Wahrnehmung**

Eine wesentliche Aufgabe von Examensnoten ist es, die Leistungen von Studierenden miteinander vergleichbar zu machen. Sie sollen nicht nur abbilden, in welchem Maße die im Studium erlernten Inhalte und Methoden vom Prüfling verstanden und internalisiert wurden, sie sollen auch verdeutlichen, welche Absolventen ihr Studium in welchem Maße besser oder schlechter als andere abgeschlossen haben. Diese Vergleichbarkeit ist jedoch nur dann gewährleistet, wenn, unabhängig von Zeitpunkt und Ort der Prüfung, für die gleiche Leistung auch die gleiche Beurteilung erfolgt. Nur wenn gute Noten auch auf Studierende mit guten Leistungen beschränkt bleiben und schlechtere Leistungen mit schlechteren Noten sanktioniert werden, sagen sie tatsächlich etwas über das Verhältnis der erbrachten Leistungen untereinander aus. Werden die vergebenen Noten immer besser, obwohl die Leistungen von Studierenden konstant bleiben, wird damit das Prinzip der Vergleichbarkeit von Noten sowohl im Querschnitt als auch im Zeitverlauf ausgehebelt. Absolventen mit guten Leistungen erhalten weiterhin gute Noten, solche mit schlechteren Leistungen dann aber ebenfalls, womit eine Differenzierung zwischen ihnen erschwert wird (*Hu, 2005; Johnson, 2003*). Die Entwertung von Noten, die sich im Zeitverlauf durch eine solche Dynamik vollzieht, wird als Noteninflation (englisch: grade inflation) bezeichnet. Dass eine Entwertung von Noten im Zeitverlauf möglicherweise auch den Vergleich von Leistungen im Querschnitt erschwert, lässt sich durch den begrenzten Charakter von Noten erklären. Im Gegensatz zur Entwertung einer Währung ist die Entwertung von Noten durch die Festsetzung von „sehr guten“ bzw. „ausgezeichneten“ Prädikaten als bestem Ergeb-

nis nur bis zu einem bestimmten Grad möglich. Eine Annäherung an einen Zustand, in dem sich immer mehr Noten der Bestnote nähern, wird als „grade compression“ (Hu, 2005:17) bezeichnet¹.

Während Verbesserungen von Notenniveaus im Zeitverlauf in den USA unter dem Stichwort *grade inflation* bereits seit den 1970er Jahren (Juola, 1976) in zahlreichen Studien untersucht werden (Jewell et al., 2013; Kolevzon, 1981; Kuh/Hu, 1999; Lowe et al., 2008; Prather et al., 1979; Sabot/Wakeman-Linn, 1991 u. v. m.), hat dieses Phänomen hierzulande erst Einzug in die öffentliche Diskussion erhalten, nachdem der Wissenschaftsrat 2003 feststellte, dass die Vergleichsfunktion von Noten durch eine Noteninflation gefährdet sein könnte. Die Forscher fanden mithilfe der vom Statistischen Bundesamt seit 1993 geführten Prüfungsstatistik heraus, dass im Zeitraum von 1996 bis 2000 „in einigen Fächern besonders gute Noten auffallend häufig vergeben“ wurden (Wissenschaftsrat, 2003:28). 2007 und 2012 bekräftigte der Wissenschaftsrat dieses Ergebnis und wies warnend auf die in den letzten zehn Jahren „fortgesetzte Tendenz zur Vergabe besserer Noten“ hin (Wissenschaftsrat, 2012:7). In den Medien wurde die Feststellung des Wissenschaftsrats, dass 1996, 1998 und 2000 in den meisten Fächern vorwiegend gute Noten vergeben wurden und entsprechend gute Notendurchschnitte vorliegen ohne jeden Zweifel als Folge eines inflationären Verbesserungsprozesses aufgefasst. Und das, obwohl zu dieser Einschätzung gar keine Vergleichszahlen zu früheren Zeitpunkten herangezogen wurden und im deskriptiven Arbeitsbericht auch keine Ursachen für die Vergabe guter Noten ermittelt wurden: „Viele Hochschulen vergeben fast nur noch Einsen und Zweien. Jetzt kritisiert der Wissenschaftsrat die Inflation der Spitzenzensuren“ (Spiewak, 2003). Das mediale Echo auf die folgenden Berichte des Wissenschaftsrats 2007 und 2012 fiel nach dem gleichen Muster aus: Aus der „fortgesetzten Tendenz zur Vergabe besserer Noten“ (Wissenschaftsrat, 2012:7) wurde die „Tendenz zu immer besseren Prüfungsnoten“ (Friedmann, 2012), die Rede war von „Kuschelnoten“ (Schlicht, 2012) und von: Noteninflation, und zwar von „schleichender Noteninflation“ einhergehend mit einer „Aufweichung der Bewertungsstandards“ (Preuss, 2012) – zitiert aus dem Vorwort des Arbeitsberichts, in dem der Vergleich der Prüfungsjahre 2000 und 2012 angekündigt wird (Wissenschaftsrat, 2012:10). Verstärkt wurde der öffentliche Eindruck der allgemeinen Entwertung von Noten von Kommentaren des Wissenschaftsrats selbst (vgl. ebd.).

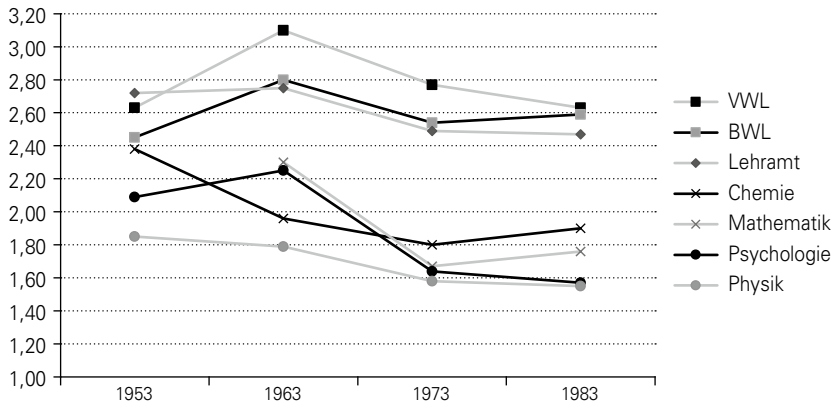
Doch ist das auf diese Weise vermittelte Bild einer umfassenden Noteninflation an deutschen Hochschulen zutreffend? Und betrifft dieses Phänomen die Studierenden aller Fächer und Hochschulen gleichermaßen? Sollten die Bewertungsstandards tat-

¹Theoretisch ist auch eine entgegengesetzte Entwicklung zu immer schlechteren Abschlussnoten denkbar, die eine Differenzierung der Leistungen einschränkt. Empirische Indizien für eine langfristige Entwicklung zu schlechteren Noten sind dem Autor jedoch nicht bekannt.

sächlich überall im Zeitverlauf gesunken sein, würde dies eine korrigierende Reaktion erfordern, die fach- und hochschulübergreifend auf die gesamte akademische Bewertungspraxis abzielen müsste, etwa die bundesweite Einführung konkreter Rahmenprüfungsvorschriften. Derartige Maßnahmen würden jedoch ins Leere laufen, sollten sich langfristige Verbesserungen als fach- und/oder hochschulspezifisches Phänomen erweisen. In diesem Falle müssten konkret abgestimmte Maßnahmen zur Aufrechterhaltung gewünschter, einheitlicher Bewertungsstandards entwickelt werden. Auch die zeitliche Dimension möglicher Verbesserungen muss berücksichtigt werden: Sollte sich das Notenniveau in der Vergangenheit zwar stark verändert haben, in jüngerer Zeit aber relativ stabil sein, würde dies den gesamten Handlungsbedarf einschränken. Vor dem Hintergrund dieser Überlegungen geht der Aufsatz den Fragen nach, welche langfristigen Veränderungen sich in der Notengebung an deutschen Hochschulen im Zeitverlauf nachweisen lassen und in welcher (fach- und hochschul-)spezifischen Form sie auftreten. Um diese Fragen zu beantworten, ist eine differenzierte Betrachtung der hiesigen Notengebung im Längsschnitt notwendig. Eine solche Analyse war aufgrund mangelnder Daten im Zeitreihenformat bisher nicht möglich, kann nun jedoch anhand neuer Daten vorgenommen werden.

2 **Grade Inflation an deutschen Hochschulen? Die bisherige Datenlage**

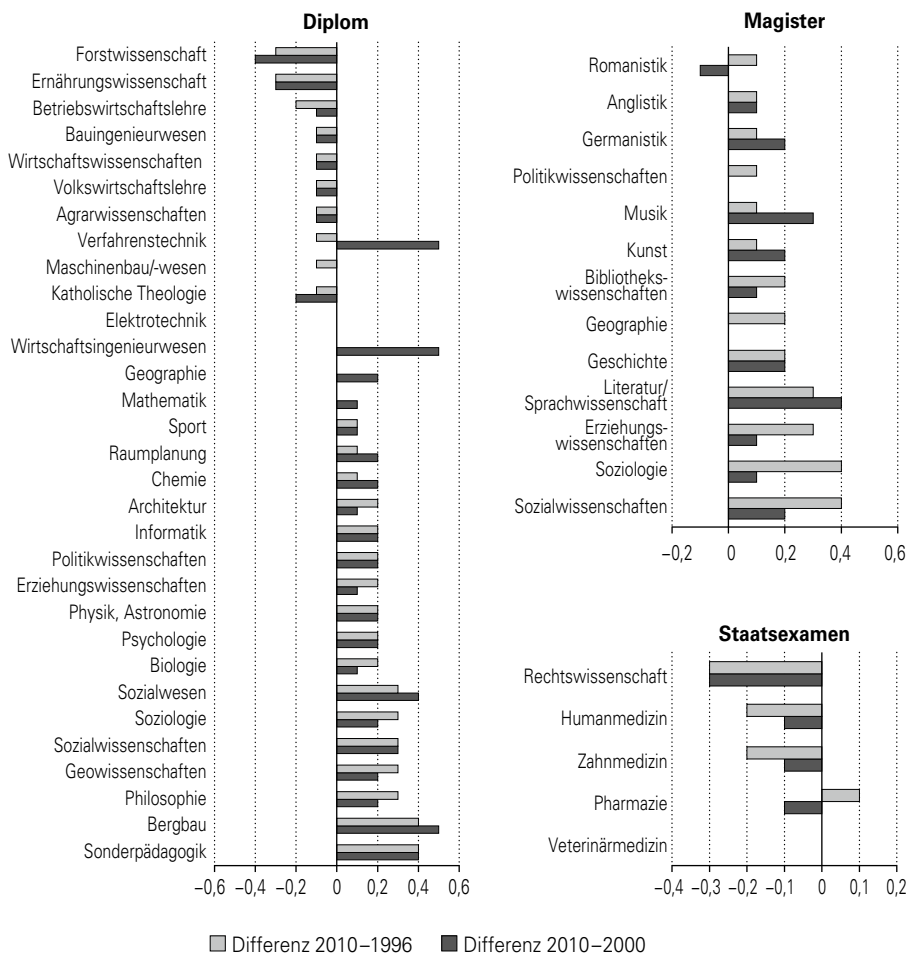
Die langfristige Entwicklung der Noten aus empirischer Sicht ist hierzulande bisher nur marginal behandelt worden. Hitpass und Trosien (1987) waren die ersten, die sich der Analyse der Notengebung an deutschen Hochschulen über einen längeren Zeitraum gewidmet haben. Die analysierten Durchschnittsnoten wurden von den beiden Autoren aus einer Stichprobe von Absolventen mehrerer Universitäten in verschiedenen Studiengängen zu vier Messzeitpunkten gewonnen (Abb.1). Die Daten zeigen, dass die Noten in vier der sechs detailliert dargestellten Diplomfächer (Physik, Chemie, Mathematik, Psychologie) sowie in fächerübergreifend zusammengefassten Lehramtsstudiengängen 1983 besser sind als zum ersten Messzeitpunkt 1953. Lediglich in Physik ist aber tatsächlich eine *kontinuierliche* Verbesserung über alle Erhebungszeitpunkte festzustellen. In Volkswirtschaftslehre (VWL), Psychologie und den Lehramtsstudiengängen sinkt das Notenniveau von 1963 an. In Betriebswirtschaftslehre (BWL), Mathematik und Chemie liegt das Notenniveau 1973 niedriger als 1963, steigt 1983 jedoch noch einmal an. Der lange Zeitraum zwischen den Messzeitpunkten, immerhin jeweils zehn Jahre, lässt dabei offen, wie sich das Notenniveau zwischen den Datenpunkten entwickelt hat. Zudem birgt die Betrachtung einzelner Messzeitpunkte das Risiko, dass einer (oder mehrere) dieser Datenpunkte einen Ausreißerwert darstellt, der den tatsächlichen Verlauf nicht widerspiegelt.

Abbildung 1: Durchschnittliche Abschlussnoten in Diplom und Lehramt (1.Staats-examen)

Quelle: Hitpass/Trosien, 1987, eigene Darstellung

Nach Hitpass und Trosien beschäftigte sich erst der Wissenschaftsrat (2003; 2007; 2012) wieder mit der Entwicklung von Examensnoten an deutschen Hochschulen, was zur geschilderten öffentlichen Wahrnehmung geführt hat. Eine genauere Betrachtung der vom Wissenschaftsrat veröffentlichten Daten zeigt jedoch, dass es verfrüht ist, vom endgültigen Ende der Vergleichbarkeit von Noten auszugehen. Betrachtet man die Veränderungen in den einzelnen Studiengängen, zeigt sich ein Bild, das dem einer umfassenden Noteninflation im Zeitraum zwischen 1996 bzw. 2000 und 2010 entgegensteht (Abb.2): Nur in zehn von 31 Diplomfächern an Universitäten, für die in allen drei Arbeitsberichten Zahlen vorliegen, hat sich das Notenniveau zwischen 1996 und 2010 verbessert, und das in sieben Fällen um gerade einmal eine Zehntelnote. Gegenüber dem Jahr 2000 sind 2010 sogar nur in acht Diplomfächern bessere Noten verzeichnet. In Studiengängen mit Abschluss Magister sinken die Noten zwischen 1996 und 2010 in keinem von 13 Studiengängen, zwischen 2000 und 2010 in einem einzigen. Im Staatsexamen gibt es immerhin in drei von fünf Fällen eine Verbesserung im Vergleich der Jahre 1996 und 2000, in vier von fünf beim Vergleich zwischen 2000 und 2010. Dabei ist zu beachten, dass in den Staatsprüfungen im Durchschnitt schlechtere Ergebnisse erzielt werden als in den übrigen Abschlussarten.

Abbildung 2: Veränderungen im Notenniveau in ausgewählten Studiengängen (nur Universitäten)

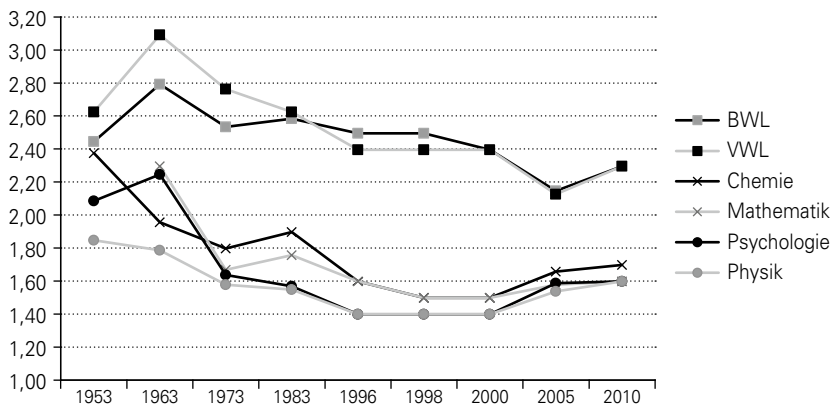


Quelle: Wissenschaftsrat 2003, 2007, 2012, eigene Berechnungen;
 Lesehilfe: Balken im positiven Bereich bedeuten ein schlechteres, Balken im negativen Bereich ein besseres Notenniveau 2010 im Vergleich zum Wert 1996 bzw. 2000.

Zwischen 1996 und 2010 ist eine Notenverbesserung auf Studiengangebene also nur vereinzelt nachweisbar. Doch wie sieht es mit dem Zeitraum davor aus? Bauer und Grave (2011) präsentieren Notendurchschnitte aus dem Konstanzer Studierendensurvey für die dort enthaltenen Hochschulen aus Nordrhein-Westfalen, Baden-Württemberg, Bayern, Hessen und Sachsen, aggregiert auf Bundeslandebene. Sie stellen für alle Länder außer Sachsen eine Verbesserung der Noten von 1983 bis 2007 fest. Die Autoren geben dabei an, „the average grade of student i at time t they have earned during their undergraduate study“ (Bauer/Grave, 2011:7, Herv.i.O.) zu nutzen – der

Konstanzer Studierendensurvey enthält laut Codebuch allerdings nur die Abitur- und die Zwischenprüfungsnote sowie alternativ eine Selbsteinschätzung der bisherigen Leistungen in Notenform, falls noch keine Zwischenprüfung abgelegt wurde. Diese Daten als aussagekräftige Durchschnittsnoten zu verwenden ist zumindest diskussionswürdig. Mehr Aufschluss bietet ein Vergleich der Noten aus den Arbeitsberichten des Wissenschaftsrates mit denen von Hitpass und Trosien (Abb.3). Hier zeigt sich für die sechs vergleichbaren Fächer mit Abschluss Diplom tatsächlich eine Verbesserung der Noten vom ersten Messzeitpunkt 1953 bis zum letzten 2010. Zwischen 1963 und 1996 sinken die Noten in allen Fächern tendenziell, bevor sie sich 1996 stabilisieren und dann wieder leicht ansteigen.

Abbildung 3: Abschlussnoten in sechs ausgewählten Fächern (Diplom)



Quelle: Hitpass/Trosien, 1987; Wissenschaftsrat 2003,2007,2012, eigene Darstellung

Eine langfristige Verbesserung der Notenniveaus könnte demnach, zumindest in einzelnen Studiengängen, in der Tat stattgefunden haben, und zwar bereits lange, bevor die öffentliche Diskussion darüber in Gang kam. Problematisch an der bisher verfügbaren Datenlage ist, dass nicht überprüft werden kann, ob einzelne Zeitpunkte möglicherweise Ausreißerwerte oder Extrempunkte von Zyklen darstellen. Um zuverlässige Aussagen über die langfristige Entwicklung treffen zu können, werden Zeitreihen mit kürzeren Abständen zwischen den einzelnen Messzeitpunkten benötigt. Solche Zeitreihen stehen inzwischen durch das Forschungsprojekt „Die Notengebung an Hochschulen in Deutschland von den 1960er Jahren bis heute. Trends, Unterschiede, Ursachen.“ erstmals zur Verfügung².

²Das an der Europa-Universität Flensburg angesiedelte und von der DFG finanzierte Projekt wurde von Prof. Dr. Volker Müller-Benedict (ZML) und Prof. Dr. Gerd Grözinger (IIM) beantragt. Neben dem Autor sind Florence Baillet (bis Ende 2014), Elena Tsarouha und Margarita McGrory in dem Projekt beschäftigt.

3 Die langfristige Entwicklung von Noten an deutschen Hochschulen

Im Rahmen des Projekts wurden in ausgewählten Studiengängen die Examensnoten der letzten Jahrzehnte (bis maximal 1943) an den Universitäten Berlin (FU), Braunschweig (TU), Göttingen, Heidelberg, Karlsruhe (KIT, ehemals TH) Münster, Saarbrücken (nur Germanistik) und Tübingen erhoben (Vollerhebung). Es liegen Daten vor für: BWL, VWL, Psychologie, Mathematik, Biologie, Chemie, Maschinenbau (alle mit Abschluss Diplom), Soziologie (Magister und Diplom), Germanistik (Magister) sowie Mathematik und Deutsch als Lehramtsstudium (1. Staatsexamen für Gymnasien) – jeweils soweit an den Hochschulen angeboten. Die Zahl der Prüflinge in den Studiengängen reicht in den verwendeten Daten der Stichprobe von $n=9$ (Mathematik Lehramt/2010) bis $n=1695$ (BWL/1993) pro Jahr (Tab. 3 im Anhang).

Die Auswahl der in der Erhebung berücksichtigten Hochschulen und Fächer erfolgte zunächst theoretisch, dann forschungspragmatisch, das heißt nach Zugangsmöglichkeiten. Es wurden Hochschulen ausgewählt, die die Überprüfung regionaler (z. B. Bundesland politisch/Stadtstaat vs. Flächenstaat) sowie hochschulspezifischer Unterschiede (z. B. Größe) zulassen. Die Auswahl der Studiengänge bildet zentrale Fächergruppen ab und ermöglicht die Überprüfung von Unterschieden nach Zugangsvoraussetzungen, der Heterogenität der Studieninhalte (z. B. standardisierter Studienverlauf vs. große Auswahl) und nach Abschlussart. Die Noten wurden an den Hochschulen in einer zweijährigen Erhebungsphase bis zum Abschlussjahrgang 1997 aufgenommen. Anschließend wurden sie mit den aggregierten Daten der amtlichen Hochschulprüfungsstatistik verknüpft, die seit 1992 personenbezogene Informationen zu den von den Hochschulen gemeldeten Prüfungsergebnissen jedes Prüflings erfasst. Der Zugang zur Hochschulprüfungsstatistik erfolgte über das Statistische Landesamt Schleswig-Holstein (2012). Außerdem stehen die Abschlussnoten in den Rechtswissenschaften als aggregierte Durchschnittsnoten auf Bundeslandebene von 1959 bis 2007 zur Verfügung. Diese konnten aus der Fachzeitschrift „Juristische Schulung“ erhoben werden. In der bundesweiten Totalerfassung der ersten juristischen Staatsprüfungen spannt sich die Fallzahl von $N=2698$ (1965) bis $N=11424$ (1996).

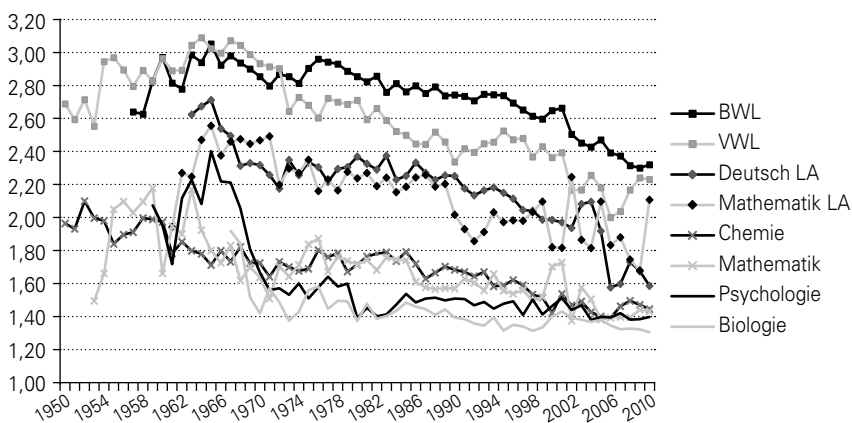
Die verwendeten Fächerdurchschnitte sind als arithmetisches Mittel der Examensnoten aller Absolventinnen und Absolventen (der erfassten Hochschulen) im gleichen Studiengang berechnet. Hierbei wird angenommen, dass die Prüfungsergebnisse die über die Hochschulen zufällig verteilten Leistungen der Studierenden abbilden und damit kein „Klumpeneffekt“, verursacht durch die jeweilige Hochschule, auftritt. Inwiefern diese Annahme empirisch haltbar ist, werden folgende Analysen der Unterschiede im Notenniveau zwischen Hochschulen zeigen. Es ist im Folgenden nur der Zeitraum

abgebildet, für den für mindestens zwei Hochschulen Werte vorliegen, auf Studiengänge gehen die Zeitreihen damit bis maximal 1950 zurück. Da in den Archiven keine verlässlichen Informationen über nicht bestandene Prüfungen erhoben werden konnten und die amtliche Statistik nur endgültig nicht bestandene Prüfungen erfasst und damit ebenfalls keine aussagekräftigen Informationen enthält (Gaens, 2013), geben die Notendurchschnitte das Mittel der bestandenen Prüfungen wieder. Die Abschlussnoten wurden als ganze Noten (Prädikate) mit den Werten 1 („sehr gut“) bis 4 („ausreichend“) erfasst, mit der Ausnahme des juristischen „vollbefriedigend“ als 2,5. Das an einigen Hochschulen zu unterschiedlichen Bedingungen vergebene Prädikat „mit Auszeichnung“ wurde als „sehr gut“ kodiert (vgl. Wissenschaftsrat, 2012).

3.1 Es gibt kontinuierliche Verbesserungen im durchschnittlichen Notenniveau

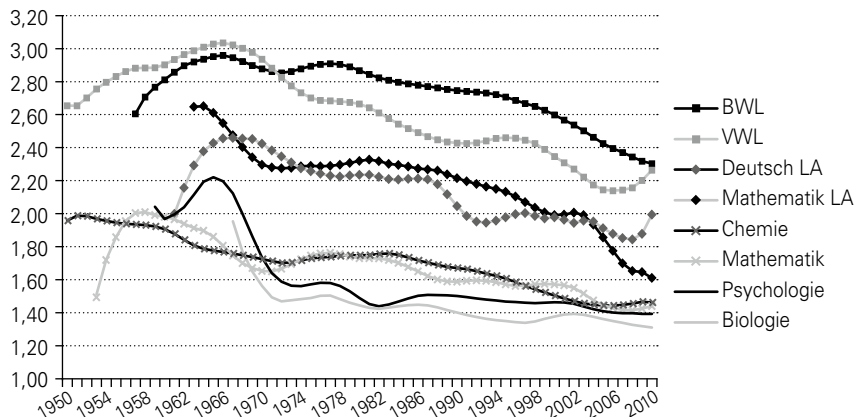
Tatsächlich zeigen die Daten in acht von zwölf berücksichtigten Studiengängen eine nennenswerte Verbesserung der Notendurchschnitte im Zeitverlauf (Abb.4 und 5). Abbildung 5 präsentiert die mittels der LOWESS-Technik³ geglätteten Daten.

Abbildung 4: Verlauf der Abschlussnoten in Studiengängen mit langfristiger Notenverbesserung



³ „**LO**cally **WE**ighted regression **S**catterplot **S**moothing“: Die Glättung der Daten wird durch eine polynomiale Regression erreicht. Vor Anwendung des nicht-parametrischen Regressionsmodells wird lediglich die Bandbreite der benachbarten Datenpunkte, die in die Glättung jedes einzelnen Werts eingehen, festgelegt. Je höher diese gewählt wird, umso glatter werden die Daten (Cleveland, 1979; Wolf/Best, 2010). Zur Glättung der Durchschnittsnoten wurden diese als abhängige, die Zeitvariable als unabhängige Variable in die Regression aufgenommen. Die jeweils vorgegebene Bandbreite der Glättung ist angegeben.

Abbildung 5: Verlauf der Abschlussnoten in Studiengängen mit langfristiger Notenverbesserung (LOWESS 0.2)



Die Graphiken zeigen, dass die Verbesserung jeweils zu Beginn/Mitte der 1960er Jahre einsetzt, was den Daten von Hitpass und Trosien entspricht. Die Verbesserungsprozesse vollziehen sich allerdings in unterschiedlichem Ausmaß (in Chemie z. B. Verbesserung um ca. eine halbe Note seit 1960, in Deutsch mit Abschluss Lehramt um mehr als eine ganze Note seit 1963) und auf unterschiedlich hohem Niveau. Die durchschnittliche Abschlussnote in BWL ist 2010 trotz langfristiger Verbesserung immer noch signifikant schlechter als das Notenniveau in Chemie 1960 ($p=0.000$) und in Biologie 1967 ($p=0.009$), also bevor dort die Verbesserung auftritt⁴. Die statistische Signifikanz ist hier in beiden Fällen bedeutsam, der Effekt in beiden Vergleichen als mittelstark einzustufen (Hedges' $g=0.548$ für Chemie vs. BWL bzw. 0.733 für Biologie vs. BWL). Die Abstufung der Studiengänge nach Notenniveau entspricht für die fünf der sechs Diplomfächer, für die zum Vergleich sowohl Daten von Hitpass und Trosien als auch vom Wissenschaftsrat vorliegen, der dortigen Abstufung. In Biologie und Psychologie kann man die Notenlage spätestens seit Beginn der 1970er Jahre als derart gut einstufen, dass die Leistungsdifferenzierung dort durch das Phänomen der grade compression gefährdet ist. Dafür sprechen sowohl die in Richtung der besten Noten verzerrten Notenverteilungen als auch die entsprechend geringen Standardabweichungen der Individualnoten in diesen beiden Fächern. In Psychologie wurden 54,6 Prozent der 11 467 zwischen 1971 und 1997 bestandenen Prüfungen mit einem „sehr gut“, 95,5 Prozent mit einem „sehr gut“ oder „gut“ bewertet (Schiefe: 0.713). In Biologie liegen diese Anteile im gleichen Zeitraum bei 62,8 bzw. 96,5 Prozent ($n=11\ 611$, Schiefe: 1.033). Der Wert der Standardabweichungen beträgt in Psychologie seit 1971

⁴Zur groben Einschätzung, bei welchem Ausmaß die Unterschiede im Notenniveau statistisch signifikant ($p \leq 0.05$) sind, kann dem Leser folgender Orientierungswert dienen: Bei einer Fallzahl von $n=50$ und einer Standardabweichung von $\sigma=0.5$ unterscheiden sich zwei Durchschnittsnoten \bar{x}_1 und \bar{x}_2 zwischen zwei Studiengängen ab einer Differenz $\bar{x}_1 - \bar{x}_2$ von ca. $0.14 (= 1.96 \cdot 0.5 / \sqrt{50}, \alpha = 0.05)$ signifikant voneinander.

in jedem Jahr zwischen $s=0.53$ und $s=0.61$, in Biologie seit 1977 zwischen $s=0.49$ und $s=0.63$. Die Abbildungen 6 und 7 verdeutlichen, dass die Streuung der Noten mit der Zeit ebenfalls sinkt – jedoch nicht nur dort, wo die Noten die untere Begrenzung der Skala zu erreichen drohen – auch in allen anderen Studiengängen mit sinkenden Notendurchschnitten nimmt die Streuung parallel ab, wie hier das Beispiel BWL zeigt. Dort sinkt die Standardabweichung im Zeitverlauf in etwa genauso stark wie in Chemie und Mathematik Diplom – trotz eines wesentlich schlechteren Notenniveaus. Die Verbesserungen gehen also unabhängig von der Begrenzung des Notenspektrums mit einer Kompression der Noten einher.

Abbildung 6: Abschlussnoten vs. Standardabweichungen*3 Biologie (LOWESS 0.3)

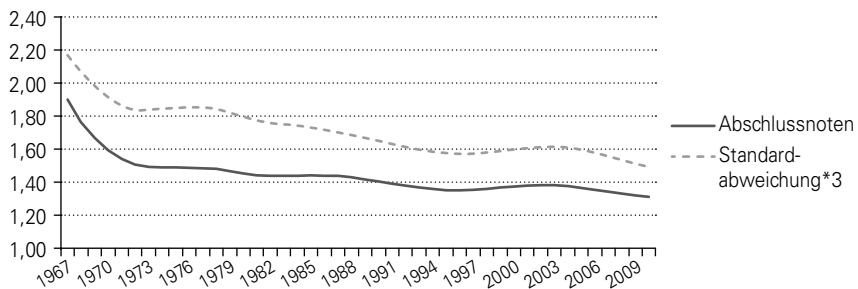
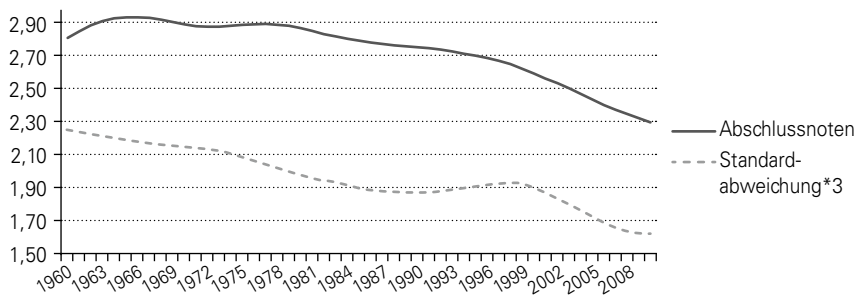


Abbildung 7: Abschlussnoten vs. Standardabweichungen*3 BWL (LOWESS 0.3)



Gemeinsam ist allen Studiengängen, dass die Verbesserung nicht perfekt linear verläuft. Die Abwärtsbewegung wird von zyklischen Schwankungen begleitet, deren Aufwärtsbewegungen Plateauphasen erzeugen (Abb. 5). Diese unterbrechen die Abwärtsdynamik und sorgen für zeitweilig stabile Notenniveaus. Die eigentliche Verbesserung im Zeitverlauf vollzieht sich damit in bestimmten Phasen von unterschiedlicher Länge: In Biologie ist der Großteil der Verbesserung, die sich 2010 im Vergleich zu 1967 feststellen lässt, bereits 1973 erreicht. Sechs Jahre nachdem die verfügbare Zeitreihe beginnt, ist das Notenniveau von 1,92 auf 1,38 gesunken. Es lässt sich anhand der vorliegenden Daten nicht genau sagen, wann dieser Prozess in Biologie eingesetzt

hat und wie lange er anhielt. Da der Diplomabschluss in Biologie an den meisten Hochschulen aber erst in den 1960er Jahren eingeführt wurde, lässt sich auch ohne weiter zurückgehende Daten folgern, dass sich der Verbesserungsprozess in wenigen Jahren vollzogen haben muss. In den anderen Studiengängen lassen sich mehrere maßgebliche Verbesserungsperioden ausmachen: In Chemie sinkt das Notenniveau zwischen 1958 und 1971 (-0,35) sowie zwischen 1987 und 2006 (-0,23). Die maximale in der Reihe enthaltene Verbesserungsspanne von -0,70, die über eine Dauer von 54 Jahren (1952–2006) festzustellen ist, beruht also größtenteils auf zwei Phasen der Verbesserung, die in zusammen 32 Jahren für ein Absinken des Notenniveaus um 0,58 Notenwerte verantwortlich sind. In Psychologie findet die größte Verbesserung zwischen 1965 und 1971 statt, eine zweite, sehr kurze Verbesserungsperiode lässt sich zwischen 1979 und 1982 beobachten. Zwischen 1971 und 1979 wie auch zwischen 1982 und 2010 bewegen sich die Durchschnittsnoten in einem relativ stabilen Rahmen. Auch die wirtschaftswissenschaftlichen und die Lehramtsstudiengänge weisen das Muster einer phasendominierten Verbesserung auf: In BWL sinkt das Notenniveau zwischen 1965 und 1971 und noch einmal zwischen 1984 und 2009. In VWL sind es mit den Jahren 1967–1973, 1982–1990 und 2001–2006 drei Phasen, die hauptsächlich für die langfristige Verbesserung im Studiengang verantwortlich zeichnen. In den beiden Lehramtsstudiengängen verlaufen die Entwicklungen beinahe parallel: In Deutsch verbessern sich die Noten wesentlich zwischen 1965–1970 und zwischen 1986–2006, in Mathematik zwischen 1965–1972 und zwischen 1989–2009, wobei in Mathematik am Ende der Zeitreihe einige Ausreißerwerte das Bild verzerren. Der Verlauf in Mathematik (Diplom) weicht von der in den anderen Studiengängen zu beobachtenden Verlaufsform der durch (eine) Plateauphase(n) unterbrochenen Verbesserung ab. Zwar lassen sich mit den Zeiträumen von 1963–1971 und von 1985–2002 ebenfalls zwei maßgebliche Verbesserungsperioden bestimmen, jedoch folgt auf die erste dieser Phasen zunächst wieder eine Verschlechterung bis 1976, bevor sich eine neunjährige Plateauphase anschließt. Hier überlagert die Aufwärtsbewegung der zyklischen Komponente den Abwärtstrend.

Tabelle 1 fasst die Verbesserungs- und Plateauphasen zusammen. Spalte 2 gibt die maßgeblichen Verbesserungsperioden an, also die Zeiträume, in denen die Noten den stärksten Abwärtstrends unterliegen, in den Klammern findet sich das Ausmaß der jeweiligen Verbesserung in dieser Zeit. Dass das Ausmaß der Verbesserung als absolute Angabe zwischen den Studiengängen sinnvoll vergleichbar ist, zeigt sich durch einen Abgleich der Werte mit der jeweiligen Effektstärke (hier nicht abgebildet): Die Korrelation zwischen den beiden Kennwerten beträgt $r=0.964$. In Spalte 3 sind die Plateauphasen und in Klammern die Spannweiten verzeichnet, innerhalb derer sich das Notenniveau in diesen Phasen bewegt. In Chemie etwa beträgt die Differenz zwischen höchster und niedrigster Durchschnittsnote zwischen 1971 und 1987 0,13 Notenwerte. Die Plateauphasen zeichnen sich dadurch aus, dass sie eine Spannweite

von $R=0.20$ nicht überschreiten. Die Spalten 4 und 5 bieten einen Vergleich der maximal in der Zeitreihe zu beobachtenden Verbesserung, also der Differenz zwischen höchstem und niedrigstem Notendurchschnitt, und der Zeitspanne in der dieses Maximum erreicht wird (Spalte 4) mit dem Ausmaß, das nur in den beschriebenen Verbesserungsperioden (Spalte 5) zustande kommt⁵. Hier zeigt sich, dass die größte Verbesserung über den gesamten Zeitraum in Deutsch (Lehramt), die geringste in Biologie stattgefunden hat (Spalte 4). Die letzte Spalte enthält zwei Werte, die die durchschnittliche Verbesserung im Notenniveau pro Jahr abbilden: Der erste Wert über den gesamten Zeitraum seit Einsetzen der ersten Verbesserung bis zum letzten Messzeitpunkt 2010, der zweite Wert nur für die Jahre, die in die Verbesserungsphase(n) fallen (siehe Spalte 1). Berechnet sind diese Angaben als arithmetisches Mittel der ersten Differenzen der entsprechenden Datenpunkte, also als Durchschnittswert der jährlichen Veränderungen gegenüber den Vorjahreswerten⁶. Der über den gesamten Zeitraum gemittelte Wert gibt ähnlich wie der absolute Wert in Spalte 4 das Ausmaß dieser Verbesserung seit dem ersten Einsetzen der Verbesserung an, nur eben nicht begrenzt auf die Zeitspanne der maximalen Verbesserung. Am zweiten Wert, dem Durchschnitt innerhalb der Verbesserungsphase(n), lässt sich ablesen, wie stark die Noten in den Jahren sinken, in denen die Verbesserung maßgeblich stattfindet. Spalte 6 lässt erkennen, dass sich die Notenveränderung in Psychologie als besonders dynamisch erweist: Dort verbessert sich das Niveau während der Verbesserungsphasen mit Abstand am stärksten. Es wird außerdem deutlich, dass die einfache Differenz eines Anfangs- und Endwerts die Dynamik zwischen diesen Messpunkten in der Regel verschweigt: Die jährlichen Verbesserungen in den entsprechenden Phasen liegen in allen Fällen über den Durchschnittswerten für den gesamten Zeitraum. Vergleicht man den Zeitraum der maximalen Verbesserung (Spalte 4) mit der summierten Dauer der Verbesserungsphasen (Spalte 5), veranschaulicht dies die Dauer, über die tatsächlich dynamische Bewegungen in den Noten zu beobachten sind. Es fällt auf, dass in allen Studiengängen, in denen sich eine langfristige Verbesserung zeigt, die erste dafür maßgeblich verantwortliche Phase bereits Anfang der 1970er Jahre abgeschlossen ist. Das folgende Plateau besteht je nach Studiengang zwischen acht und 17 Jahren, die zweite Verbesserungsphase setzt entsprechend versetzt Mitte/Ende der 1980er Jahre ein (in Psychologie bereits 1979).

⁵Größere Werte für letztere als für die maximale Gesamtverbesserung kommen dadurch zustande, dass die zweite Periode der Verbesserung leicht über dem Wert wieder einsetzt, mit dem die erste Periode abschließt. In Mathematik (Diplom) liegt dieser Wert deshalb deutlich höher, weil der Ausgangswert der zweiten Verbesserungsperiode dem Peak des zwischenliegenden Zyklus entspricht, der weit über dem Plateauniveau liegt.

⁶Auf diese Weise erhält man einen Durchschnittswert, der eine genauere Einschätzung der durchschnittlichen jährlichen Entwicklung bietet als bspw. eine Regression der (nicht perfekt linear verlaufenden) Noten auf eine Zeitvariable.

Tabelle 1: Verlaufsphasen und Verbesserungsausmaß in den Studiengängen mit langfristiger Notenverbesserung

| Studien- gang | Phasen der Verbesserung (Ausmaß) | Plateauphasen (Spannweite der Schwankungen) | Maximale Verbesserung (Wert/Jahre) | Verbesserung in den Phasen (Wert/Jahre) | $\bar{\sigma}$ Verbesserung pro Jahr (Gesamt/ Verbesserungs- phasen) |
|-----------------------|--|--|--|---|---|
| Biologie Diplom | 1967–1973 (–0,54**) | 1973–2010 (0,19) ^a | –0,61***/43 | –0,54/6 | –0,014/–0,091 |
| Psychologie Diplom | 1965–1971 (–0,83***) 1979–1982 (–0,19***) ^b | 1971–1979 (0,13) 1982–2010 (0,15) | –1,01***/39 | –1,02/9 | –0,022/–0,115 |
| VWL Diplom | 1967–1973 (–0,43***) 1982–1990 (–0,32***) 2001–2006 (–0,40***) | 1973–1982 (0,13) 1990–2001 (0,19) | –1,09***/42 | –1,15/19 | –0,020/–0,060 |
| Mathematik Diplom | 1963–1971 (–0,66***) 1985–2002 (–0,37***) | 1971–1976 (+0,37) ^c 1976–1985 (0,20) | –0,79***/39 | –1,03/25 | –0,016/–0,041 |
| Deutsch Lehramt | 1965–1970 (–0,39***) 1986–2006 (–0,76***) | 1970–1986 (0,20) | –1,13***/41 | –1,15/25 | –0,025/–0,046 |
| Mathematik Lehramt | 1965–1972 (–0,36***) 1989–2009 (–0,53***) | 1972–1989 (0,19) | –0,88***/44 | –0,89/27 | –0,010/–0,033 |
| BWL Diplom | 1965–1971 (–0,25***) 1984–2009 (–0,51***) | 1971–1984 (0,20) | –0,75***/44 | –0,76/31 | –0,016/–0,025 |
| Chemie Diplom | 1958–1971 (–0,35***) 1987–2006 (–0,23***) | 1971–1987 (0,13) | –0,70***/54 | –0,58/32 | –0,011/–0,019 |

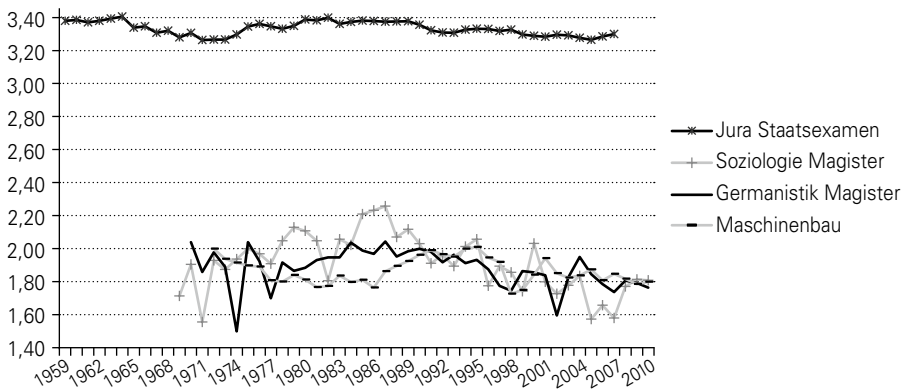
** $p \leq 0,01$; *** $p \leq 0,001$; ^a ohne Werte für 1975 und 1976; ^b ohne Wert für 1980; ^c Verschlechterungsperiode

3.2 Nicht überall verbessern sich die Noten im Zeitverlauf

Im Diplomstudiengang Maschinenbau, in den Magisterstudiengängen Soziologie und Germanistik sowie im ersten Staatsexamen der Rechtswissenschaften kann keine langfristige Verbesserung des Notenniveaus festgestellt werden (Abb. 8 und 9). Während die Noten der ersten drei Studiengänge offensichtlich zyklisch verlaufen, scheinen die Noten in den juristischen Staatsexamen sich auf den ersten Blick konstant auf demselben Niveau zu bewegen. Der Eindruck eines linearen Verlaufs ist jedoch auf die graphische Skalierung zurückzuführen. Wird die Skalierung an die einzelne Reihe angepasst, offenbart sich auch für die scheinbar konstanten Noten der rechtswissenschaftlichen Prüfungen ein in wesentlich geringeren Spannweiten eingegrenzter, aber dennoch zyklischer Verlauf (Abb. 10). Die durchschnittlichen Abschlussnoten bewegen sich in Jura über den gesamten Zeitverlauf im Rahmen einer maximalen Spannweite von $R=0,14$ bei einer Standardabweichung von $s=0,04$. In Maschinenbau ($R=0,28/ s=0,08$), Germanistik ($R=0,54/ s=0,12$) und Soziologie ($R=0,70/ s=0,17$) fallen diese Streuungswerte wesentlich höher aus. Die Spannweite der Durchschnitts-

noten in Jura liegt damit innerhalb des für die Studiengänge mit Verbesserung beobachteten Toleranzbereiches für die Plateauphasen. Zum letzten Messzeitpunkt 2007 liegt die Durchschnittsnote hier bei 3,30 und damit 0,08 Notenwerte niedriger als zum ersten Messzeitpunkt 1959. Diese, auch aufgrund der hohen Fallzahlen hochsignifikante Veränderung ($p=0.000$, Hedges' $g=0.121$) kommt dadurch zustande, dass der Wert 1959 den oberen, der Wert 2007 den unteren Wendepunkt eines Zyklus darstellt – ein Beleg dafür, wie leicht der Vergleich von zwei Zeitpunkten ohne Kenntnis der zwischenliegenden Entwicklung ein falsches Bild erzeugen kann. Auch in Maschinenbau ist die Veränderung zwischen erstem und letztem Datenpunkt ($-0,20$) aus diesem Grund hochsignifikant ($p=0.000$, Hedges' $g=0.341$). Im Vergleich zu den Verläufen in den anderen Fächern muss das *Notenniveau* in Jura dennoch – und auch trotz der zyklischen Verlaufsform – als über den Zeitverlauf konstant eingestuft werden. In Germanistik ($-0,28$, $p=0.077$) und Soziologie ($+0,10$, $p=0.639$) ist die Differenz zwischen Beginn und Ende der Zeitreihen nicht signifikant⁷. Die Streuung der Noten nimmt in den vier Studiengängen ohne Verbesserung im Zeitverlauf nicht ab (Abb. 11). Dies belegt, dass die sinkende Streuung tatsächlich in Verbindung mit der Verbesserung im Zeitverlauf zu sehen ist und keine generelle Tendenz der Notengebung darstellt.

Abbildung 8: Verlauf der Abschlussnoten in Studiengängen ohne langfristige Notenverbesserung



⁷Im Studiengang Maschinenbau liegen nur Noten von zwei Hochschulen vor. In Soziologie und Germanistik sind zu Beginn der Zeitreihen zwei bzw. sechs Datenpunkte mit geringen Fallzahlen ($n \leq 13$) entfernt worden.

Abbildung 9: Verlauf der Abschlussnoten in Studiengängen ohne langfristige Notenverbesserung (LOWESS 0.3)

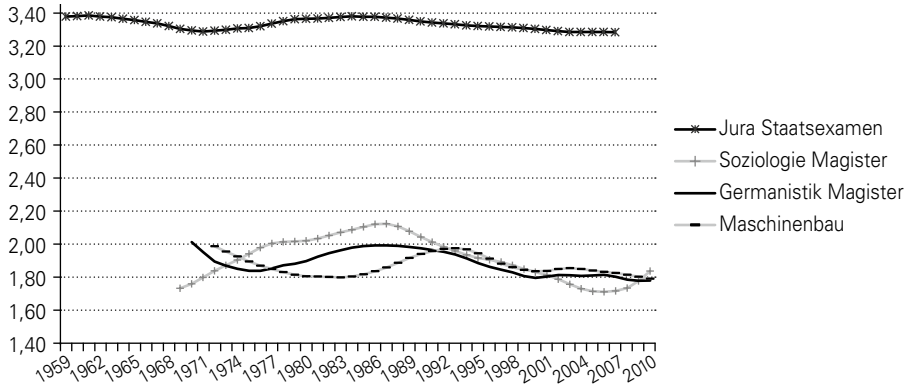


Abbildung 10: Verlauf der Abschlussnoten im ersten juristischen Staatsexamen enger skaliert

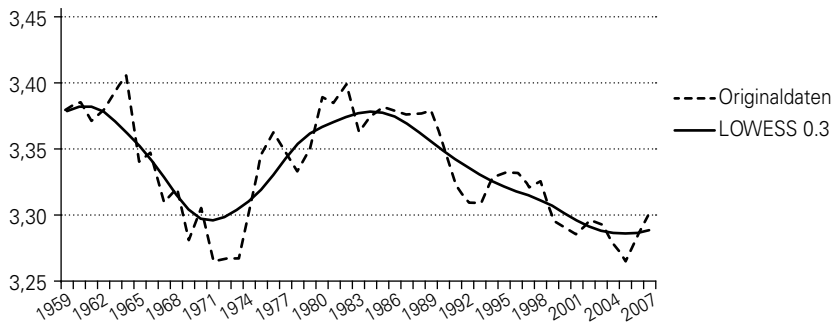
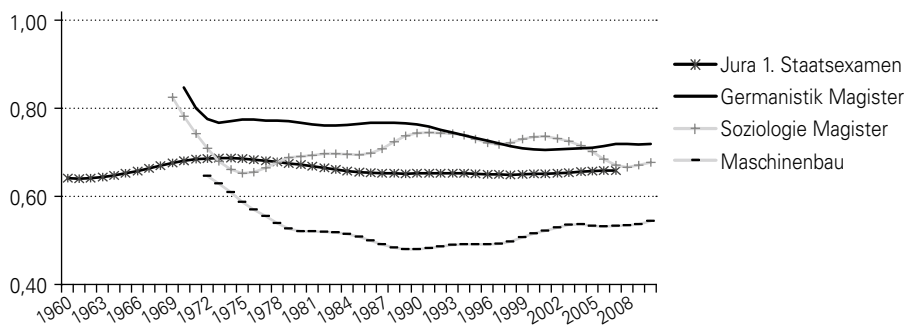


Abbildung 11: Standardabweichungen in den vier Studiengängen ohne Verbesserung (LOWESS 0.3)



Da die Noten auf Studiengangebene als Mittel aller Prüflinge berechnet wurden, ist der jeweilige Verlauf auch vom Verlauf der Noten an den Hochschulen mit den größeren Anteilen am gesamten Prüfungsvolumen abhängig. Betrachtet man die nach Studiengängen differenzierten Noten an den einzelnen Hochschulen, wird deutlich, dass die langfristige Notenentwicklung nicht nur studiengang-, sondern auch hochschulspezifisch verläuft. Am Beispiel des Diplomstudiengangs Chemie lässt sich dies exemplarisch darstellen. Der kontinuierliche Trend zu besseren Noten auf Studiengangebene wird vor allem durch die drei Universitäten Karlsruhe, Münster und Göttingen bestimmt (Abb.12). Die Noten verbessern sich an allen drei Hochschulen nahezu durchgängig, lediglich unterbrochen von den bereits auf Studiengangebene identifizierbaren Plateauphasen: Zwischen 1960 und 2010 in Karlsruhe um 1,16 ($p=0.000$, Hedges' $g=1.692$), in Münster um 0,64 ($p=0.000$, Hedges' $g=0.898$) und in Göttingen um 0,54 ($p=0.001$, Hedges' $g=0.992$) Notenwerte. In Braunschweig liegt die Durchschnittsnote 2010 um 0,07 Notenwerte höher als bei der ersten Messung 1972 ($p=0.662$). An den geglätteten Daten lässt sich jedoch eine geringe Verbesserungstendenz im Zeitverlauf erkennen, wobei sich die Noten stärker zyklisch entwickeln. (Abb.13). An den drei übrigen Hochschulen ist die Abweichung vom allgemeinen Verlauf dann nicht mehr zu übersehen. In Heidelberg ist über den gesamten Zeitraum zwar ebenfalls eine deutliche Verbesserung und ein langfristiger Abwärtstrend festzustellen ($-0,63$, $p=0.001$, Hedges' $g=0.930$), die Verbesserung vollzieht sich allerdings in einer großen zyklischen Bewegung und damit deutlich stärker schwankend als auf Studiengangebene. In Tübingen ($+0,17$ seit 1970, $p=0.489$) und Berlin ($+0,36$, $p=0.291$) sind die Noten 2010 schlechter als zu Beginn der Reihen und auch hier verlaufen die Notendurchschnitte deutlich abweichend vom Trend der gemittelten Zeitreihe.

Abbildung 12: Verlauf der Abschlussnoten in Chemie (LOWESS 0.3) – entsprechend Verlauf im Studiengang

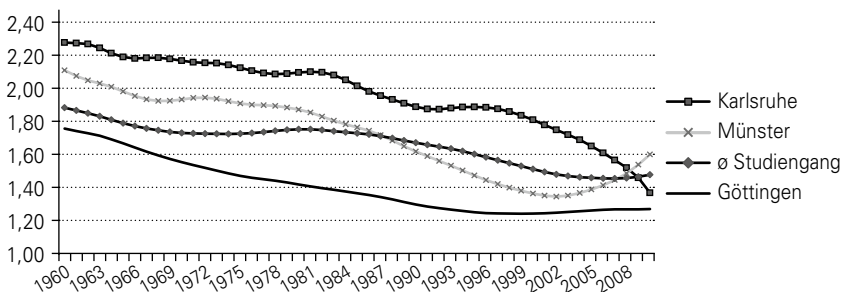
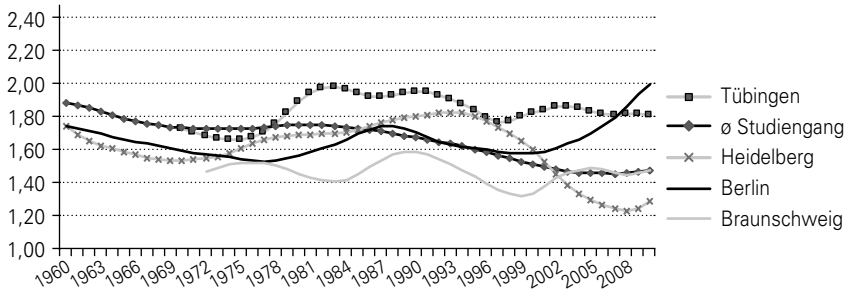


Abbildung 13: Verlauf der Abschlussnoten in Chemie (LOWESS 0.3) – gegenläufig zum Verlauf im Studiengang



Aus Tabelle 2 wird ersichtlich, dass Chemie zwar ein Extrembeispiel ist, aber auch keine echte Ausnahme darstellt. Lediglich in zwei der acht Studiengänge, in denen die Gesamtheit der Prüflinge im Durchschnitt immer bessere Noten bekommt, in VWL und Mathematik Lehramt, findet sich diese Entwicklung an allen Hochschulen wieder. Die anderen sechs weisen immer mindestens eine Hochschule auf, die den Trend zu besseren Noten nicht teilt. Umgekehrt ist eine langfristige Verbesserung der Noten im Zeitverlauf an einzelnen Hochschulen auch in den Studiengängen zu finden, in denen über alle Absolventinnen und Absolventen gemittelt keine Tendenz zu sinkenden Noten besteht (Jura, Germanistik, Soziologie). Schließlich wird an den Hochschuldaten sichtbar, dass eine zu kleine Stichprobe (wie zu erwarten) zu Verzerrungen führt: So ist der zyklische Verlauf in Maschinenbau nur in Braunschweig zu finden – seine Entsprechung auf der höheren Aggregatebene beruht auf der Dominanz des dortigen Absolventenanteils in genau dem Zeitraum, in dem in Karlsruhe eine Plateauphase die langfristige Verbesserung unterbricht. Ob sich die Noten in Maschinenbau auf Studiengangebene wirklich nicht verbessert haben, kann anhand der Daten demnach nicht geklärt werden.

Tabelle 2: Die Notenentwicklung an den Hochschulen im Vergleich zum Studiengangstrend

| Studiengang | Trend zur Verbesserung auf Studiengangebene | Hochschulen im sample | Hochschulen mit Trend zur Verbesserung | Hochschulen ohne Trend zur Verbesserung |
|----------------------|---|-----------------------|--|---|
| VWL Diplom | ja | 6 | 6 | 0 |
| Mathematik Lehramt | ja | 5* | 5 | 0 |
| Mathematik Diplom | ja | 7 | 6 | 1 |
| Psychologie Diplom | ja | 6 | 5 | 1 |
| Deutsch Lehramt | ja | 5* | 4 | 1 |
| BWL Diplom | ja | 5 | 4 | 1 |
| Biologie Diplom | ja | 7 | 4 | 3 |
| Chemie Diplom | ja | 7 | 4 | 3 |
| Jura 1. Staatsexamen | nein | 10/11** | 2 | 8 |
| Germanistik Magister | nein | 6 | 2 | 4 |
| Soziologie Magister | nein | 4 | 1 | 3 |
| Maschinenbau Diplom | nein | 2 | 1 | 1 |

*Die Lehramtsnoten sind in drei von fünf Fällen nur zu ganzen Landesprüfungsämtern zuzuordnen, umfassen dann mehrere Hochschulen.

**Bundesländer statt Hochschulen, für Bremen liegt nur eine sehr kurze Zeitreihe vor (1990–2007).

4 Diskussion der deskriptiven Ergebnisse und möglicher Ursachen der Verbesserung

Notenverbesserungen im Zeitverlauf existieren also auch an deutschen Hochschulen. Mit dieser Erkenntnis allein ist jedoch noch keine *Noteninflation*, also eine Entwertung von Noten belegt. Denn es gibt auch *leistungskonforme Ursachen*, anhand welcher Verbesserungen im Notendurchschnitt theoretisch erklärt werden können, ohne dass sich das Verhältnis zwischen individueller Leistung und dafür vergebener Note im Zeitverlauf verändert haben muss: steigende *individuelle* Leistungen der Studierenden (durch verbesserte Eignung oder verbesserte Lehre) sowie steigende *durchschnittliche* Leistungsniveaus, etwa bedingt durch eine veränderte Zusammensetzung der Studierenden (Geschlecht, soziale Herkunft, Alter - siehe etwa *Birnbaum, 1977; Potter et al., 2001; Winzer, 2002*). Auch eine zunehmende Selbstselektion Studierender in Kurse oder Hochschulen mit bekanntermaßen guten Noten würde sinkende Durchschnittsnoten begünstigen (*Bar et al., 2009; Sabot/Wakeman-Linn, 1991*). Erst wenn diese leistungskonformen Faktoren ausgeschlossen werden können, sollte in Betracht gezogen werden, dass *leistungsunabhängige Ursachen*, also sinkende Bewertungsstandards, den Grund für veränderte Notenniveaus darstellen.

In den USA werden als ursprüngliche leistungsunabhängige Ursachen für Notenverbesserungen häufig der Vietnamkrieg, im Laufe dessen Studierende mit guten Noten vor dem Einzug in die Armee bewahrt werden sollten, sowie die Gleichstellung von ethnischen Minderheiten angeführt, die zu guten Noten als Fördermittel geführt habe (*Birnbaum, 1977*). Diese Entwicklungen können theoretisch das Einsetzen eines Verbesserungsprozesses erklären, aber nicht dessen weiteren Verlauf. Geeigneter sind dazu etwa Annahmen, die auf kontinuierlichen Arbeitsmarktentwicklungen aufbauen. Wie die, dass eine zunehmende Polarisierung zwischen Arbeitsplätzen mit hohem und niedrigem Anforderungsprofil die Vergabe guter Noten fördert, da Lehrende ihren Studierenden nicht die Chance auf gute Jobs nehmen möchten (*Yang/Yip, 2003*). Oder die Vermutung, dass ein zunehmender Konkurrenzkampf um wenige gute Stellen bei immer mehr geeigneten Absolventen zu besseren Noten beiträgt (*Pressman, 2007*). *Gesellschaftliche Entwicklungen* dieser Art beeinflussen womöglich die Selektionsneigung der Prüfer. Für die hiesige Notengebung gilt es zu klären, ob hierzulande ähnliche Entwicklungen mit milderndem Einfluss auf Prüfer identifiziert werden können. Ein Beispiel könnte hier die Studentenbewegung der 1960er Jahre sein, die möglicherweise Zweifel an der Aussagekraft von Noten geweckt hat (*Ziegenspeck, 1999*). Dieser Erklärungsansatz für eine erste, fachübergreifende Phase der Aufweichung von Bewertungsstandards ist konform mit der aufgezeigten Notenentwicklung: Da die Magisterprüfung erst im Laufe der 1960er Jahre wieder flächendeckend eingeführt wurde (*Grüner, 1971*), ist die Notengebung in Soziologie und Germanistik Magister – zwei der drei Studiengänge, in denen sicher keine Verbesserung gegeben ist – nicht betroffen. Dass das Notenniveau in den Rechtswissenschaften nicht gesunken ist, könnte auf die große Distanz zwischen der politisch links stehenden Bewegung mit dem vergleichsweise rechts einzuordnenden juristischen Lehrpersonal (vgl. *Maier-Leibnitz/Schneider, 1991*) zurückzuführen sein.

Auch Veränderungen in den *institutionellen Lehr-, Lern- und Prüfungsbedingungen* werden für sinkende Bewertungsstandards verantwortlich gemacht – am häufigsten die studentische Evaluation der Lehrleistung. Es wird argumentiert, dass Lehrende sich gute Bewertungen im Austausch gegen gute Noten quasi erkaufen (*Correa, 2001; McKenzie, 1975*). Vor allem bei Aufwertungen von Evaluationsergebnissen, etwa bei der Einführung von Bonuszahlungen (*Mangan, 2009*), ist es plausibel, kurzfristige Auswirkungen dieser Eingriffe auf die Notenhöhe zu erwarten, während ein kontinuierlicher Einfluss kaum begründbar ist. Da Lehrevaluationen an deutschen Hochschulen im 20. Jahrhundert noch keine große Bedeutung zukam (an keiner der Hochschulen in der Stichprobe wurden vor 2004 zentral verwaltete Evaluationen zur Qualitätssicherung durchgeführt), kann die langfristige Verbesserung der Noten hierzulande nicht mit einem unintendierten Effekt dieses Instruments erklärt werden.

Zunehmend schlechtere Rahmenbedingungen in der Lehre (gestiegene Studierendenzahlen und damit verschlechterte Betreuungsrelationen) könnten die Lehrenden dazu motivieren, den Studierenden im Ausgleich bessere Noten zu geben (*Müller-Benedict/Tsarouha, 2011*). Ein steigendes Prüfungsaufkommen könnte Prüfer zudem dazu bringen, milder zu bewerten, um die Zeit, die sie für Forschung und Verwaltung benötigen, nicht für zunehmende Verhandlungen von Beschwerden über schlechte Noten nutzen zu müssen (*Franz, 2010*). Eine ähnliche Argumentation verfolgt, wer behauptet, dass Studierende zunehmend prüfungsorientiert lernen und Lehrende ihnen dies ermöglichen, indem sie Prüfungsinhalte im Voraus immer genauer bekanntgeben (vgl. *Winzer, 2002*). Ein derartiger Mechanismus könnte Teil eines umfassenderen Handels sein, den Kuh als „disengagement compact“ (*Kuh, 2003:28*) bezeichnet: Lehrende erwarten weniger von ihren Studierenden und vergeben bessere Noten, die Studierenden erwarten weniger Lehrleistung und damit Zeitaufwand von den Lehrenden, den diese in Forschung und Verwaltung umlenken.

In der Stichprobe sind zunehmende Prüfungszahlen bis mindestens in die 1990er Jahre zu beobachten⁸, teils in zyklischen Bewegungen, bevor sie in einigen Studiengängen wieder absinken, in den 2000er Jahren aber immer noch über dem Niveau der 1960er bzw. 1970er liegen. Nur im Lehramt erreichen die Prüfungszahlen bereits um 1980 ihren Höhepunkt und liegen in den 2000er Jahren niedriger als in den 1960ern, was aber dadurch zustande kommt, dass für zwei der fünf Hochschulen/Prüfungsämter ab 1998 keine Daten mehr vorliegen (Tab. 3). Ein negativer Zusammenhang zwischen Studierendenzahlen und Notenhöhe ist also möglich.

Schließlich werden formale Regelungen, wie neue Optionen zur Vermeidung schlechter Noten, etwa durch Freiversuche (*Birnbaum, 1977*), gestiegene Wahlmöglichkeiten (*Falkenberg, 1996*) und wechselnde Bewertungsverfahren mit Notenverbesserungen in Verbindung gebracht. Für Letztere wird ein Wechsel von der sozialen zur sachlichen Bezugsnorm postuliert, der mehr gute Noten trotz unveränderter Leistungen ermöglicht (vgl. *Rojstaczer/Healy, 2012*). Hierzulande wurde der Einfluss derartiger Entwicklungen bisher nicht überprüft.

Auch *monetäre Anreize* werden häufig als Grund für die Verbesserung von Noten angeführt. Sowohl ganzen Hochschulen als auch einzelnen Instituten wird unterstellt, die Notenverbesserung bewusst zu fördern, um damit Outputindikatoren für Fördermittel, etwa Absolventenzahlen, positiv zu beeinflussen (*De Paola, 2008; Warning/Welzel, 2005*). Ein Interesse an möglichst hoher öffentlicher Reputation, über die

⁸Die Zahl der Hochschulen, die in die Berechnung der Durchschnittsnoten eingehen, umfasst zu Beginn der Zeitreihen i. d. R. noch nicht das ganze Sample. Der Anstieg der Fallzahlen wird damit in den ersten Jahren auch durch die Erweiterung des Samples begünstigt und stellt zu einem gewissen Anteil ein Artefakt dar (Tab. 3).

Leistungen der Studierenden in Form ihrer Noten angezeigt, könnte hier ebenfalls einen Anreiz leisten (Yang/Yip, 2003). Wie die empirischen Befunde zeigen, entwickeln sich die Noten allerdings nicht nur hierzulande studiengangspezifisch in unterschiedlichem Tempo und auch nicht an jeder Einrichtung in jedem Studiengang gleichermaßen zum Besseren (etwa Kuh/Hu, 1999; Sabot/Wakeman-Linn, 1991). Die hochschulinternen Unterschiede im Notenverlauf sprechen gegen die Auffassung, Hochschulen würden eine Noteninflation systematisch fördern. Dass der Wettbewerb um finanzielle Mittel innerhalb der Hochschulen Notenverbesserungen produziert, ist eher nachvollziehbar. So werden in kleinen Instituten und solchen mit niedrigen Betreuungsrelationen die besseren Noten vergeben (Baird, 2009; Dickson, 1984). Ihnen wird ein Interesse an der Vergabe immer besserer Noten nachgesagt, um die Teilnehmerzahlen zu halten oder zu erhöhen und im akademischen Betrieb konkurrenzfähig zu bleiben (Staples, 1998). Auf das konkrete Interesse der individuellen Lehrperson, ihren Arbeitsplatz nicht zu verlieren, wird ebenfalls hingewiesen (Dickson, 1984). Diese Annahme wird durch den Befund gestützt, dass befristete Lehrkräfte bessere Noten vergeben als unbefristet Angestellte (Sonner, 2000).

Auch das wechselnde Selbstverständnis von Hochschulen hin zu Dienstleistern, die Studierende als Kunden betrachten, die für ihre Ausbildung zahlen und eine entsprechende Leistung dafür erhalten sollen, wird in den Fokus genommen. Die gewünschte Leistung stellt dabei nicht mehr primär Bildung dar, sondern ein auf dem Arbeitsmarkt optimal verwertbares Zertifikat (vgl. Kirp, 2003). Dass Noten genutzt werden, um Studierende anzuziehen und dadurch finanzielle Mittel aufzustocken, ist für deutsche Hochschulen aufgrund der selbst in Zeiten des New Public Management immer noch geringen Kopplung von Fördermitteln an derartige Outputfaktoren nicht anzunehmen (Bauer/Grave, 2011). Auch der Konsumentenstatus von Studierenden ist nicht ohne weiteres aus den US-amerikanischen auf deutsche Verhältnisse übertragbar, wie ein Vergleich der Bedeutung von Studiengebühren zwischen den beiden Hochschulsystemen schnell offenbart.

Differenzen in den Notenniveaus zwischen Studiengängen oder Hochschulen könnten, bei entsprechenden *Einstellungen von Lehrenden*, einen Anreiz für streng bewertende Prüfer bieten, ihre Noten an bessere Niveaus anzupassen, um ihre Studierenden nicht zu benachteiligen (Achen/Courant, 2009). Gelegentlich wird eine Wende zu einem egalitären Lehr- und Lehrklima postuliert, als dessen Folge eine verringerte Orientierung der Bewertung an fachlichen Kompetenzen zugunsten von Sympathien vermutet wird (vgl. Winzer, 2002). Auch zunehmende Unterstützung bei der Stipendienbewerbung (vgl. Hu, 2005) und ein Wandel des wahrgenommenen Bildungsauftrags weg von Selektion hin zum Aufbau eines gesunden Selbstvertrauens von Studierenden (Mansfield, 2001) werden für den Abwärtstrend verantwortlich gemacht. Dass Lehrkräfte im Zeitverlauf zunehmend selbst gute Noten im Studium erhalten haben und

diese deswegen auch als Prüfer zur Bewertung heranziehen (*Winzer, 2002*), klingt wenig plausibel, wenn man bedenkt, dass zukünftiges Lehrpersonal während seines Studiums vermutlich unabhängig vom durchschnittlichen Notenniveau zu den besten Studierenden gehörte und vorwiegend beste Noten erzielt haben dürfte. Anpassungsdruck innerhalb der Institute könnte aber ein Grund sein, weshalb die Noten nicht schlechter werden (*Koedel, 2011*). Schließlich wird die Vergabe zunehmend guter Noten sogar als Ausgleichshandlung zu geringer Forschungsleistung betrachtet, um das Selbstwertgefühl aufzubessern (*Pressman, 2007*). Für einen merkbaren Effekt müsste jedoch die Zahl derer, deren Forschungsleistung vergleichsweise gering ausfällt, mit der Zeit deutlich ansteigen.

Die differenzierte Analyse der langfristigen Zeitreihen hat gezeigt, dass das Argument einer *generellen* Aufweichung von Bewertungsstandards zur Begründung langfristiger Notenverbesserungen nicht ausreicht. Die Daten sind kompatibel mit einem gemeinsamen Ursprung der Verbesserung in den 1960er Jahren – in diesem Zeitraum bleiben nur die Noten im juristischen Staatsexamen konstant. Im weiteren Verlauf zeigen sich jedoch studien- und hochschulspezifische Entwicklungen. Sollten diese Unterschiede leistungsunabhängige Differenzen darstellen, liegt die Annahme nahe, dass auch studien-, hochschul- und zeitraumspezifische Ursachen vorliegen. Bevor sich also die Vorstellung einer umfassenden Entwertung von Noten durchsetzt, die den Prüflingen aller Fächer über alle Abschlüsse hinweg beste Noten für durchschnittliche Leistungen beschert, muss in kommenden Arbeiten erst einmal untersucht werden, welche Faktoren tatsächlich für die festgestellten Notenverbesserungen ursächlich sind und zu welchem Teil diese leistungskonformer oder -unabhängiger Art sind. Nur so kann verhindert werden, dass hochschulpolitische Schnellschüsse mit dem Ziel der Anhebung von Notenniveaus die jetzt schon breite Palette potentieller Einflussfaktoren auf die Notenvergabe noch erweitern und die Vergleichbarkeit von Noten damit erst recht einschränken.

Literatur

Achen, Alexandra C./ Courant, Paul N. (2009): What Are Grades Made Off? In: The Journal of Economic Perspectives. Vol. 23 (3), S. 77–92.

Bar, Talia/Kadiyali, Vrinda/Zussman, Asaf (2009): Grade Information and Grade Inflation: The Cornell Experiment. In: The Journal of Economic Perspectives. Vol. 23 (3), S. 93–108.

Baird, Matthew (2009): Dynamic Estimation of the Incentive Schemes and Signalling Costs of Grade Inflation. <http://papers.ccpr.ucla.edu/papers/PWP-CCPR-2009-015/PWP-CCPR-2009-015.pdf> (zuletzt überprüft: 24.11.2015)

Bauer, Thomas K./Grave, Barbara S. (2011): Performance-related Funding of Universities: Does More Competition Lead to Grade Inflation? IZA Discussion Paper No.6073.

Birnbaum, Robert (1977): Factors Related to University Grade Inflation. In: The Journal of Higher Education. Vol. 48 (5), S. 519–539.

Cleveland, William S. (1979): Robust Locally Weighted Regression and Smoothing Scatterplots In: Journal of the American Statistical Association. Vol. 74 (368), S. 829–836.

Correa, Hector (2001): A game theoretic analysis of faculty competition and academic standards. In: Higher Education Policy. Vol. 14, S. 175–182.

De Paola, Maria (2008): Are easy grading practices induced by low demand? Evidence from Italy. MPRA Paper No. 14425.

Dickson, Vaughan A. (1984): An Economic Model of Faculty Grading Practices. In: Journal of Economic Education. Vol. 15 (3), S. 197–203.

Falkenberg, Steve (1996): Grade Inflation. <http://people.eku.edu/falkenbergs/grdinfla.htm> (zuletzt überprüft: 24.11.2015)

Franz, Wan-Ju I. (2010): Grade inflation under the threat of students' nuisance: Theory and evidence. In: Economics of Education Review. Vol. 29, S. 411–422.

Friedmann, Jan (2012): Wissenschaftsrat beklagt zu gute Noten an Unis. In: Unispiegel vom 09.11.2012.

Gaens, Thomas (2013): Von einem, der auszog, einen Leistungsindikator zu erheben. Durchfallquoten und die Problematik ihrer Bildung. In: Das Hochschulwesen. Vol. 6/2013, S. 200–206.

Grüner, Gustav (1971): Die Magisterprüfung in der Bundesrepublik Deutschland. Weinheim.

Hitpass, Josef/Trosien, Jürgen (1987): Leistungsbeurteilung in Hochschulabschlussprüfungen innerhalb von drei Jahrzehnten - Wandel von Prüfungsergebnis und Prüfungserlebnis an deutschen Universitäten. Bad Honnef.

Hu, Shouping (2005): Beyond Grade Inflation. Grading Problems in Higher Education. San Francisco.

Jewell, R. Todd/McPherson, Michael A./Tieslau, Margie A. (2013): Whose fault is it? Assigning blame for grade inflation in higher education. In: Applied Economics. Vol. 45, S. 1185–1200.

Johnson, Valen E. (2003): Grade Inflation: A Crisis in College Education. New York.

Juola, Arvo E. (1976): Grade Inflation in Higher Education: What Can Or Should We Do? Paper presented at the Annual Meeting of National Council on Measurement in Education in San Francisco, California.

Juristische Schulung (1959–2007): JuS - Zeitschrift für Studium und Referendariat. München/Frankfurt a. M.

Kirp, David L. (2003): Shakespeare, Einstein, and the Bottom Line: The Marketing of Higher Education. Cambridge.

Koedel, Cory (2011): Grading Standards in Education Departments at Universities. In: Education Policy Analysis Archives 19(23). S. 1–20.

Kolevzon, Michael S. (1981): Grade inflation in higher education: A comparative study. In: Research in Higher Education. Vol. 15 (3), S. 195–212.

Kuh, George D./Hu, Shouping (1999): Unraveling the Complexity of the Increase in College Grades From the Mid-1980s to the Mid-1990s. In: Educational Evaluation and Policy Analysis. Vol. 21 (3), S.297–321.

Kuh, George D. (2003): What We're Learning about Student Engagement From NSSE. In: Change. Vol. 35 (2), S. 24–32.

Lowe, S. Keith/Borstorff, Patricia C./Landry III, Robert J. (2008): An empirical examination of the phenomenon of grade inflation in higher education: a focus of grade divergence between business and other fields of study. In: Academy of Educational Leadership Journal. Vol. 12 (1), S. 15–33.

Maier-Leibnitz, Heinz/Schneider, Christoph (1991): The Status of Academic Research in the Federal German Republic: A Report on Two Surveys and the Testimony of Individual Scientists. In: Minerva. Vol. 29 (1), S. 27–60.

Mangan, Katherine (2009): Professors Compete for Bonuses Based on Student Evaluations. In: The Chronicle of Higher Education vom 30.01.2009.

Mansfield, Harvey C. (2001): Grade Inflation. It's Time to Face the Facts. In: The Chronicle of Higher Education vom 06.04.2001.

McKenzie, Richard B. (1975): The Economic Effects of Grade Inflation on Instructor Evaluations: A Theoretical Approach. In: The Journal of Economic Education. Vol. 6 (2), S. 99–105.

Müller-Benedict, Volker/Tsarouha, Elena (2011): Können Examensnoten verglichen werden? Eine Analyse von Einflüssen des sozialen Kontextes auf Hochschulprüfungen. In: Zeitschrift für Soziologie. Vol.40, S. 288–309.

Potter, William/Nyman, Melvin A./Klumpp, Karen S. (2001): Be careful what you wish for: Analysis of grading trends at a small liberal arts college. In: College and University: The Journal of the American Association of Collegiate Registrars. Vol. 76 (4), S. 9–14.

Prather, James E./Smith, Glynton/Kodras, Janet E. (1979): A Longitudinal Study of Grades in 144 Undergraduate Courses. In: Research in Higher Education. Vol. 10 (1), S.11–24.

Pressman, Steven (2007): The Economics of Grade Inflation. In: Challenge. Vol. 50 (5), S. 93–102.

Preuss, Roland (2012): Zu gute Noten an deutschen Hochschulen. In: Süddeutsche Zeitung vom 10./11.11.2012.

Rojstaczer, Stuart/Healy, Christopher (2012): Where A is ordinary: The evolution of American college and university grading, 1940–2009. In: Teachers College Record. Vol. 114 (7), S. 1–23.

Sabot, Richard/Wakeman-Linn, John (1991): Grade Inflation and Course Choice. In: Journal of Economic Perspectives. Vol. 5 (1), S. 159–170.

Schlicht, Uwe (2012): Kuschelnoten: Eine Eins für alle. In: Tagesspiegel vom 13.11.2012.

Sonner, Brenda S. (2000): A is for "Adjunct": Examining Grade Inflation in Higher Education. In: Journal of Education for Business. Vol. 76 (1), S.5–8.

Spiewack, Matthias (2003): Noten ohne Wert. In: Zeit vom 20.02.2003.

Staples, Brent (1998): Why Colleges Shower Their Students With A's. In: New York Times vom 08.03.1998.

Statistisches Landesamt Schleswig-Holstein, Forschungsdatenzentrum (2012): Hochschulprüfungsstatistik 1995–2010. Zugang über Arbeitsplatz für Gastwissenschaftler.

Warning, Susanne/Welzel, Peter (2005): A Note on Grade Inflation and University Competition. Paper based on presentations given at the 2005 meeting of the Allied Social Sciences Association in Philadelphia and at the 3rd Workshop on Business and Economic Policy in Sion. http://www.fep.up.pt/conferences/earie2005/cd_rom/session%20iii/iii.l/warning.pdf (zuletzt überprüft: 24.11.2015)

Winzer, Margret (2002): Grade Inflation. An Appraisal of the Research. Technical Report, University of Lethbridge Faculty of Education. <http://www.uleth.ca/edu/runte/inflation/Inflationb.pdf> (zuletzt überprüft: 24.11.2015)

Wissenschaftsrat (2003): Prüfungsnoten an Hochschulen 1996, 1998 und 2000 nach ausgewählten Studienbereichen und Studienfächern – Arbeitsbericht. Hrsg.v.d. Geschäftsstelle des Wissenschaftsrats, Drucksache 5536–03.

Wissenschaftsrat (2007): Prüfungsnoten im Prüfungsjahr 2005 an Universitäten (einschließlich KH, PH, TH) sowie an Fachhochschulen (einschließlich Verwaltungsfachhochschulen) nach ausgewählten Studienbereichen und Studienfächern – Arbeitsbericht. Hrsg. v. d. Geschäftsstelle des Wissenschaftsrats, Drucksache 7769–07.

Wissenschaftsrat (2012): Prüfungsnoten an Hochschulen im Prüfungsjahr 2010 - Arbeitsbericht mit einem wissenschaftspolitischen Kommentar des Wissenschaftsrates. Hrsg. v. d. Geschäftsstelle des Wissenschaftsrats, Drucksache 2627-12.

Wolf, Christof/Best, Henning (2010): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden : VS Verlag für Sozialwissenschaften.

Yang, Huanxing/Yip, Chun Seng (2003): An Economic Theory of Grade Inflation. Working Paper. <http://econ.ohio-state.edu/hyang/grade-inflation.pdf> (zuletzt überprüft: 24.11.2015)

Ziegenspeck, Jörg (1999): Handbuch Zensur und Zeugnis in der Schule: Historischer Rückblick, allgemeine Problematik, empirische Befunde und bildungspolitische Implikationen. Bad Heilbrunn.

Anhang

Tabelle 3: Fallzahlen der Stichprobe in den einzelnen Studiengängen seit 1950

| Jahr | Biologie | Chemie | Mathematik | Psychologie | BWL | VWL | Maschinenbau | Soziologie Magister | Germanistik Magister | Mathematik Lehramt | Deutsch Lehramt | Jura |
|------|----------|---------|------------|-------------|---------|---------|--------------|------------------------|-------------------------|-----------------------|--------------------|-----------|
| 1950 | | 61 (3) | | | | 94 (2) | | | | | | |
| 1951 | | 46 (3) | | | | 124 (2) | | | | | | |
| 1952 | | 58 (3) | | | | 113 (2) | | | | | | |
| 1953 | | 43 (3) | 18 (2) | | | 213 (3) | | | | | | |
| 1954 | | 64 (3) | 21 (2) | | | 185 (3) | | | | | | |
| 1955 | | 84 (3) | 19 (2) | | | 205 (4) | | | | | | |
| 1956 | | 70 (3) | 20 (2) | | | 233 (4) | | | | | | |
| 1957 | | 83 (3) | 20 (2) | | 67 (2) | 246 (4) | | | | | | |
| 1958 | | 76 (3) | 10 (2) | | 70 (2) | 226 (4) | | | | | | |
| 1959 | | 128 (4) | 11 (2) | 13 (2) | 135 (2) | 146 (4) | | | | | | 3028 (10) |
| 1960 | | 180 (5) | 24 (3) | 23 (3) | 146 (2) | 202 (5) | | | | | | 3400 (10) |
| 1961 | | 145 (5) | 20 (3) | 18 (3) | 131 (2) | 166 (5) | | | | 38 (2) | | 3283 (10) |
| 1962 | | 167 (5) | 21 (3) | 25 (3) | 164 (2) | 178 (5) | | | | 44 (2) | | 3305 (10) |
| 1963 | | 183 (5) | 35 (4) | 26 (3) | 224 (3) | 176 (6) | | | | 64 (4) | 166 (3) | 3150 (10) |
| 1964 | | 194 (5) | 41 (5) | 47 (3) | 257 (4) | 221 (6) | | | 2 (2) | 93 (4) | 188 (3) | 2792 (10) |
| 1965 | | 197 (5) | 46 (5) | 47 (3) | 262 (4) | 277 (6) | | | 4 (2) | 111 (4) | 231 (3) | 2698 (10) |

Fortsetzung Tabelle 3 siehe nächste Seite

Fortsetzung **Tabelle 3**

| Jahr | Biologie | Chemie | Mathematik | Psychologie | BWL | VWL | Maschinen- bau | Soziologie Magister | Germanistik Magister | Mathematik Lehramt | Deutsch Lehramt | Jura |
|------|----------|---------|------------|-------------|----------|---------|-------------------|------------------------|-------------------------|-----------------------|--------------------|------------|
| 1966 | | 173 (5) | 74 (6) | 45 (3) | 425 (4) | 367 (6) | | | 6 (2) | 103 (4) | 211 (3) | 2850 (10) |
| 1967 | 13 (2) | 216 (5) | 114 (6) | 79 (4) | 433 (4) | 370 (6) | | 10 (2) | 6 (2) | 106 (4) | 227 (3) | 3089 (10) |
| 1968 | 13 (2) | 196 (5) | 106 (6) | 137 (4) | 538 (4) | 446 (6) | | 8 (2) | 8 (2) | 153 (4) | 243 (3) | 3465 (10) |
| 1969 | 23 (2) | 184 (5) | 135 (6) | 128 (4) | 512 (4) | 460 (6) | | 14 (3) | 13 (3) | 107 (4) | 282 (3) | 4284 (10) |
| 1970 | 33 (3) | 217 (7) | 109 (6) | 116 (4) | 500 (4) | 384 (6) | | 21 (4) | 25 (6) | 87 (4) | 218 (3) | 3712 (10) |
| 1971 | 45 (4) | 285 (7) | 140 (7) | 145 (5) | 442 (4) | 387 (6) | | 18 (4) | 57 (6) | 101 (4) | 256 (3) | 3532 (10) |
| 1972 | 104 (4) | 299 (7) | 157 (7) | 172 (6) | 451 (4) | 305 (6) | 173 (2) | 28 (4) | 48 (6) | 124 (4) | 290 (4) | 4359 (10) |
| 1973 | 100 (4) | 311 (7) | 196 (7) | 205 (6) | 519 (4) | 256 (6) | 181 (2) | 40 (4) | 55 (6) | 213 (5) | 440 (5) | 5132 (10) |
| 1974 | 136 (4) | 308 (7) | 269 (7) | 269 (6) | 500 (4) | 319 (6) | 179 (2) | 16 (4) | 47 (6) | 238 (5) | 699 (5) | 4887 (10) |
| 1975 | 181 (4) | 317 (7) | 272 (7) | 330 (6) | 534 (4) | 336 (6) | 206 (2) | 44 (4) | 49 (6) | 232 (5) | 705 (5) | 4326 (10) |
| 1976 | 196 (5) | 296 (7) | 233 (7) | 368 (6) | 523 (4) | 351 (6) | 232 (2) | 32 (4) | 79 (6) | 207 (5) | 623 (5) | 3496 (10) |
| 1977 | 173 (5) | 247 (7) | 252 (7) | 416 (6) | 610 (4) | 370 (6) | 214 (2) | 44 (4) | 87 (6) | 303 (5) | 808 (5) | 3857 (10) |
| 1978 | 174 (5) | 306 (7) | 221 (7) | 346 (6) | 583 (4) | 460 (6) | 271 (2) | 63 (4) | 108 (6) | 406 (5) | 819 (5) | 4324 (10) |
| 1979 | 187 (5) | 266 (7) | 237 (7) | 359 (6) | 571 (4) | 434 (6) | 277 (2) | 46 (4) | 120 (6) | 367 (5) | 974 (5) | 4930 (10) |
| 1980 | 180 (5) | 310 (7) | 230 (7) | 583 (6) | 491 (4) | 316 (6) | 325 (2) | 46 (4) | 174 (6) | 419 (5) | 918 (5) | 5580 (10) |
| 1981 | 205 (6) | 335 (7) | 202 (7) | 519 (6) | 411 (4) | 253 (6) | 356 (2) | 41 (4) | 191 (6) | 413 (5) | 879 (5) | 6158 (10) |
| 1982 | 291 (7) | 330 (7) | 195 (7) | 465 (6) | 571 (3) | 307 (6) | 297 (2) | 31 (4) | 172 (6) | 362 (5) | 854 (5) | 5592 (10) |
| 1983 | 321 (7) | 348 (7) | 210 (7) | 471 (6) | 536 (3) | 281 (6) | 313 (2) | 35 (4) | 214 (6) | 224 (5) | 504 (5) | 5535 (10) |
| 1984 | 412 (7) | 317 (7) | 164 (7) | 501 (6) | 515 (4) | 297 (6) | 346 (2) | 45 (4) | 242 (6) | 159 (5) | 524 (5) | 5854 (10) |
| 1985 | 444 (7) | 372 (7) | 161 (7) | 494 (6) | 678 (4) | 219 (6) | 372 (2) | 38 (4) | 267 (6) | 164 (5) | 552 (5) | 6015 (10) |
| 1986 | 447 (7) | 315 (7) | 171 (7) | 460 (6) | 771 (4) | 227 (6) | 413 (2) | 56 (4) | 300 (6) | 146 (5) | 507 (5) | 7082 (10) |
| 1987 | 587 (7) | 382 (7) | 215 (7) | 538 (6) | 869 (4) | 249 (6) | 508 (2) | 66 (4) | 342 (6) | 92 (5) | 412 (5) | 6951 (10) |
| 1988 | 619 (7) | 483 (7) | 235 (7) | 435 (6) | 961 (4) | 265 (6) | 498 (2) | 84 (4) | 391 (6) | 68 (5) | 342 (5) | 7927 (10) |
| 1989 | 658 (7) | 475 (7) | 258 (7) | 438 (6) | 1107 (4) | 280 (6) | 578 (2) | 76 (4) | 398 (6) | 58 (5) | 298 (5) | 8020 (10) |
| 1990 | 725 (7) | 484 (7) | 254 (7) | 450 (6) | 1297 (4) | 271 (6) | 668 (2) | 65 (4) | 411 (6) | 50 (5) | 252 (5) | 8175 (11) |
| 1991 | 792 (7) | 513 (7) | 276 (7) | 444 (6) | 1455 (4) | 288 (6) | 643 (2) | 68 (4) | 365 (6) | 58 (5) | 250 (5) | 7508 (11) |
| 1992 | 810 (7) | 498 (7) | 238 (7) | 474 (6) | 1532 (4) | 359 (6) | 657 (2) | 53 (4) | 390 (6) | 65 (5) | 209 (5) | 8411 (11) |
| 1993 | 807 (7) | 529 (7) | 289 (7) | 563 (6) | 1695 (4) | 408 (6) | 711 (2) | 66 (4) | 389 (6) | 72 (5) | 286 (5) | 9752 (11) |
| 1994 | 844 (7) | 556 (7) | 323 (7) | 577 (6) | 1644 (4) | 496 (6) | 727 (2) | 76 (4) | 417 (6) | 141 (5) | 350 (5) | 10017 (11) |
| 1995 | 826 (7) | 527 (7) | 296 (7) | 479 (6) | 1515 (4) | 558 (6) | 730 (2) | 70 (4) | 390 (6) | 206 (5) | 424 (5) | 10812 (11) |
| 1996 | 677 (7) | 510 (7) | 318 (7) | 450 (6) | 1369 (4) | 533 (6) | 751 (2) | 58 (4) | 411 (6) | 290 (5) | 466 (5) | 11424 (11) |
| 1997 | 670 (7) | 432 (7) | 310 (7) | 516 (6) | 1338 (4) | 540 (6) | 615 (2) | 66 (4) | 383 (6) | 308 (5) | 550 (5) | 11124 (11) |
| 1998 | 550 (7) | 330 (7) | 207 (7) | 603 (6) | 1134 (4) | 428 (6) | 560 (2) | 57 (4) | 368 (6) | 27 (3) | 88 (3) | 10709 (11) |
| 1999 | 543 (7) | 215 (7) | 187 (7) | 448 (6) | 1001 (4) | 323 (6) | 420 (2) | 53 (4) | 240 (6) | 30 (3) | 103 (3) | 10605 (11) |

Fortsetzung Tabelle 3 siehe nächste Seite

Fortsetzung **Tabelle 3**

| Jahr | Biologie | Chemie | Mathematik | Psychologie | BWL | VWL | Maschinen- bau | Soziologie Magister | Germanistik Magister | Mathematik Lehramt | Deutsch Lehramt | Jura |
|--------------|---------------|---------------|--------------|---------------|---------------|---------------|-------------------|------------------------|-------------------------|-----------------------|--------------------|----------------|
| 2000 | 558 (7) | 214 (7) | 130 (7) | 413 (6) | 902 (4) | 294 (6) | 317 (2) | 62 (4) | 289 (6) | 24 (3) | 90 (3) | 10371 (11) |
| 2001 | 545 (7) | 198 (7) | 144 (7) | 417 (6) | 996 (4) | 258 (6) | 278 (2) | 49 (4) | 300 (6) | 21 (3) | 78 (3) | 9624 (11) |
| 2002 | 564 (7) | 194 (7) | 142 (7) | 451 (6) | 969 (4) | 231 (6) | 263 (2) | 58 (4) | 298 (6) | 21 (3) | 102 (3) | 9466 (11) |
| 2003 | 608 (7) | 189 (7) | 156 (7) | 693 (6) | 878 (4) | 232 (6) | 291 (2) | 116 (4) | 364 (6) | 23 (3) | 93 (3) | 8357 (11) |
| 2004 | 623 (7) | 201 (7) | 132 (7) | 591 (6) | 981 (4) | 285 (6) | 316 (2) | 69 (4) | 321 (6) | 27 (3) | 71 (3) | 8514 (11) |
| 2005 | 621 (7) | 223 (7) | 127 (7) | 539 (6) | 964 (4) | 308 (6) | 312 (2) | 96 (4) | 339 (6) | 20 (3) | 115 (4) | 8053 (11) |
| 2006 | 617 (7) | 270 (7) | 157 (7) | 592 (6) | 964 (4) | 354 (6) | 383 (2) | 103 (4) | 380 (6) | 18 (3) | 222 (4) | 8999 (11) |
| 2007 | 705 (7) | 246 (7) | 233 (7) | 572 (6) | 988 (4) | 499 (6) | 401 (2) | 129 (4) | 428 (6) | 26 (3) | 254 (4) | 8887 (11) |
| 2008 | 678 (7) | 295 (7) | 243 (7) | 504 (6) | 881 (4) | 482 (5) | 452 (2) | 119 (4) | 404 (6) | 36 (4) | 230 (4) | |
| 2009 | 542 (6) | 322 (7) | 210 (7) | 496 (6) | 820 (4) | 472 (5) | 472 (2) | 129 (4) | 363 (6) | 41 (3) | 228 (4) | |
| 2010 | 306 (5) | 201 (7) | 182 (7) | 338 (6) | 498 (4) | 311 (5) | 358 (2) | 60 (4) | 257 (6) | 9 (2) | 83 (4) | |
| Total | 19 153 | 16 214 | 9 416 | 18 828 | 39 328 | 18 874 | 16 064 | 2 524 | 10 912 | 6 715 | 17 714 | 313 421 |

Manuskript eingereicht: 25.02.2015
 Manuskript angenommen: 04.08.2015

Anschrift des Autors:

Thomas Gaens, M.A. Soziologie
 Europa-Universität Flensburg
 Zentrum für Methodenlehre
 Auf dem Campus 1
 24943 Flensburg
 E-Mail: Thomas.Gaens@uni-flensburg.de

Thomas Gaens ist wissenschaftlicher Mitarbeiter am Zentrum für Methodenlehre an der Europa-Universität Flensburg und im DFG-Forschungsprojekt „Die Notengebung an Hochschulen in Deutschland seit den 1960er Jahren. Trends, Unterschiede, Ursachen.“ beschäftigt.