



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER



Diskussionspapier des
Instituts für Organisationsökonomik

7/2013

**Können sich Hochschuldozenten bessere
studentische Lehrevaluationen „erkaufen“?**

Laura Lütkenhöner

Discussion Paper of the
Institute for Organisational Economics

**Diskussionspapier des
Instituts für Organisationsökonomik
7/2013**

Juli 2013

ISSN 2191-2475

**Können sich Hochschuldozenten bessere studentische
Lehrevaluationen „erkaufen“?**

Laura Lütkenhöner

Zusammenfassung

In dieser Studie werden 128 Fälle betrachtet, in denen dieselben Studenten eine Vorlesung zu zwei verschiedenen Zeitpunkten evaluiert haben. Die Besonderheit dabei ist, dass ausschließlich Fälle berücksichtigt werden, in denen die zum Zeitpunkt der Evaluation erwartete oder bereits erhaltene Note zum späteren Zeitpunkt besser ist als zum früheren Zeitpunkt. Die Ergebnisse lassen darauf schließen, dass sich (vielfach beobachtete) Zusammenhänge zwischen Noten und Evaluationsergebnissen nicht allein mit dem ‚Effektive-Lehre-Ansatz‘ und/oder dem ‚Studenten-Charakteristika-Ansatz‘ erklären lassen. Vielmehr scheinen Dozenten durchaus die Möglichkeit zu haben, bessere Lehrbewertungen zu „erkaufen“. Binäre logistische Regressionen zeigen aber auch, dass die Wahrscheinlichkeit, mit der ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertungen reagiert, vom Geschlecht und vom Elternhaus sowie von einer bereits absolvierten Ausbildung beeinflusst wird. In Bezug auf die Wahrscheinlichkeit für eine solche Reaktion bestehen auch signifikante Unterschiede zwischen verschiedenen Dozenten.

JEL-Codes: A20, A22, C23, C83, I21, I23

Can University Lecturers “Buy” Themselves Good Results in Their Students’ Teaching Evaluations?

Abstract

This study comprises 128 cases in which the same students have evaluated one lecture at two different points of time. The special feature of the study is that only such cases are considered in which the grade, expected or already received, has improved at the second point of time in comparison to the first point of time. The results suggest that correlations between grades and students’ evaluations of teaching (which are often observed) are not fully explainable by the *grading leniency hypothesis* and/or the *students’ characteristics hypothesis*. Rather university lecturers seem to have the opportunity to “buy” themselves good results in their students’ teaching evaluations. However, binary logistic regressions also reveal the influence of other aspects on the likelihood that a student whose grade has improved gives a better teaching evaluation. These aspects are the student’s gender, his or her parental home and the fact whether or not he or she has already finished an apprenticeship. The likelihood for such a reaction also differs between individual lecturers.

Im Internet unter:

http://www.wiwi.uni-muenster.de/io/forschen/downloads/DP-IO_07_2013.pdf

Westfälische Wilhelms-Universität Münster
Institut für Organisationsökonomik
Scharnhorststraße 100
D-48151 Münster

Tel: +49-251/83-24303 (Sekretariat)
E-Mail: io@uni-muenster.de
Internet: www.wiwi.uni-muenster.de/io

Können sich Hochschuldozenten bessere studentische Lehrevaluationen „erkaufen“?*

1. Einleitung

Eine Reihe von empirischen Studien, wie beispielsweise die von McPherson/Jewell/Kim (2009) und Brockx/Spooren/Mortelmans (2011) zeigen, dass bessere Noten mit besseren Lehrbewertungen einhergehen. Diese Erkenntnis ist in der Literatur weitgehend anerkannt. Unstimmigkeit besteht aber darüber, wie dieser Zusammenhang zu interpretieren ist (vgl. Brockx/Spooren/Mortelmans 2011, S. 292). Marsh (1984) nennt hierfür drei – mittlerweile von vielen Autoren übernommene – Interpretationsansätze (vgl. Marsh 1984, S. 722; Brockx/Spooren/Mortelmans 2011, S. 291 f.). Nach dem ersten Ansatz – hier als *„Effektive-Lehre-Ansatz“* bezeichnet – sind Noten und studentische Lehrbewertungen positiv korreliert, weil effektivere Lehre zu höheren Lernerfolgen und damit auch zu besseren Noten führt (vgl. Marsh 1984, S. 722).¹ Dem zweiten Ansatz zufolge – hier als *„Käuflichkeitsansatz“* bezeichnet – rührt der Zusammenhang zwischen Noten und Evaluationsergebnissen daher, dass Studenten (hier und im Folgenden als generisches Maskulinum zu verstehen) im Austausch für bessere Noten Dozenten mit besseren Lehrbewertungen belohnen (vgl. Marsh 1984, S. 722).² Der dritte Ansatz – hier als *„Studenten-Charakteristika-Ansatz“* bezeichnet – argumentiert hingegen, dass von vornherein bestehende Unterschiede in den Charakteristika der Studenten sowohl ihre Noten als auch ihre Lehrbewertungen beeinflussen (vgl. Marsh 1984, S. 722).³

Zusammenhänge zwischen Noten und Lehrbewertungen sind nur dann unproblematisch, wenn sie vollständig mit dem *„Effektive-Lehre-Ansatz“* erklärbar wären. In diesem Fall wären

* Herrn Prof. Dr. Alexander Dilger danke ich für wertvolle Ideen, Anregungen und Hinweise. Ihm und den anderen Professoren des Centrums für Interdisziplinäre Wirtschaftsforschung der Westfälischen Wilhelms-Universität Münster, namentlich Herrn Prof. Dr. Thomas Apolte, Herrn Prof. Dr. Dietmar Krafft sowie Herrn Prof. Dr. Christian Müller, danke ich außerdem für ihr Einverständnis, dass ich unter den Teilnehmern ihrer Lehrveranstaltungen Befragungen durchführen durfte. Schließlich danke ich allen Studierenden, die an den Befragungen teilgenommen haben. Für alle Inhalte und möglicherweise verbliebene Fehler bin selbstverständlich ausschließlich ich selbst verantwortlich.

¹ In englischsprachigen Veröffentlichungen wird dieser Ansatz oftmals als „validity hypothesis“ (Marsh 1984, S. 737; Brockx/Spooren/Mortelmans 2011, S. 291) bezeichnet.

² Dieser Ansatz wird in englischsprachigen Veröffentlichungen oftmals als „grading leniency hypothesis“ (Marsh 1984, S. 737; Griffin 2004, S. 411; Ewing 2012, S. 146) oder einfach nur als „leniency hypothesis“ (Brockx/Spooren/Mortelmans 2011, S. 291) bezeichnet.

³ Marsh (1984, S. 737) bezeichnet diesen Ansatz als „student characteristics hypothesis“; Brockx/Spooren/Mortelmans (2011, S. 291) als „student characteristic hypothesis“.

studentische Evaluationen jedoch insofern entbehrlich, als Lehrleistungen vollständig über studentische Noten abgebildet würden. Sollte der beobachtbare Zusammenhang zwischen Noten und Evaluationsergebnissen auf dem ‚*Studenten-Charakteristika-Ansatz*‘ beruhen, so kann der Dozent die Evaluationsergebnisse zwar nicht beeinflussen, sie können aber verzerrt sein. Besonders problematisch ist der ‚*Käuflichkeitsansatz*‘, zumal Evaluationsergebnisse auch für Leistungsvergleiche und zur Auszeichnung von Lehrenden herangezogen werden. Prinzipiell könnte natürlich auch eine Kombination der drei Interpretationsansätze erforderlich sein.

In der vorliegenden Studie wird nach Indizien für den ‚*Käuflichkeitsansatz*‘ gesucht. Dabei wird nach Anhaltspunkten dafür gesucht, dass Studenten allein aufgrund einer Notenverbesserung bessere Lehrbewertungen vergeben. In diesem Fall hätten Hochschuldozenten die Möglichkeit, bessere Lehrbewertungen zu „erkaufen“. Außerdem wird, was in bisherigen Studien offenbar noch nicht untersucht wurde, geprüft, ob bestimmte Charakteristika von Studenten (beispielsweise das Geschlecht, das Elternhaus oder eine vor Studienbeginn absolviert Ausbildung) Einfluss auf die Wahrscheinlichkeit haben, dass ein Student auf eine Notenverbesserung (d. h. eine Verbesserung der zum Zeitpunkt der Evaluation erwarteten oder bereits erhaltenen Note) mit besseren Lehrbewertungen reagiert. In diesem Fall könnte die Möglichkeit, bessere Lehrbewertungen zu „erkaufen“, auch von der Kurszusammensetzung abhängen. Des Weiteren wird geprüft, ob in Bezug auf die Wahrscheinlichkeit für eine solche Reaktion signifikante Unterschiede zwischen verschiedenen Dozenten bestehen. Dies würde bedeuten, dass auch bei gleicher Kurszusammensetzung einige Dozenten in stärkerem Maße als ihre Kollegen die Möglichkeit hätten, bessere Lehrbewertungen zu „erkaufen“.

Im folgenden zweiten Abschnitt werden die untersuchten Hypothesen hergeleitet. Danach wird im dritten Abschnitt das methodische Vorgehen erläutert. Dabei werden auch Schwierigkeiten aufgezeigt, die bei der Untersuchung der Frage bestehen, ob Dozenten bessere Lehrbewertungen „erkaufen“ können. Im vierten Abschnitt werden die Ergebnisse präsentiert, die dann im fünften Abschnitt diskutiert werden. Die Arbeit schließt mit einem Fazit.

2. Hypothesen

Die Frage, ob Dozenten bessere Lehrbewertungen „erkaufen“ können, wird von Krautmann/Sander (1999), McPherson (2006), McPherson/Jewell (2007) und McPherson/Jewell/Kim

(2009) bejaht.⁴ Krautmann/Sander (1999, S. 61) berichten, dass die durchschnittliche Lehrbewertung um 0,34 bis 0,56 Notenpunkte besser ausfällt, wenn die erwartete Durchschnittsnote der Kursteilnehmer um einen Notenpunkt besser ist. McPherson/Jewell/Kim (2009, S. 43) zeigen, dass um einen Notenpunkt bessere Erwartungsnoten die Lehrbewertungen um 0,27 Notenpunkten verbessern. Unter der Annahme, dass der ‚Käuflichkeitsansatz‘ – wie von den genannten Autoren angeführt – eine denkbare Erklärung für beobachtbare Zusammenhänge zwischen Noten und Lehrbewertungen ist, lässt sich die *Hypothese H1* aufstellen.

H1: Die Wahrscheinlichkeit, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertungen reagiert, ist umso höher, je stärker die Notenverbesserung ist.

Außerdem ist vorstellbar, dass Charakteristika von Studenten ebenfalls die Wahrscheinlichkeit beeinflussen, mit der sie bei der Bewertung einer bestimmten Vorlesung auf bessere Noten mit besseren Lehrbewertungen reagieren. Diese Vermutung ist in *Hypothese H2* formuliert. Sie wird gestützt von empirischen Studien, die derartige Faktoren als Einflussfaktoren auf studentische Evaluationsergebnisse identifizieren. So berichten Hancock/Shannon/Trentham (1993, S. 244) und Spooren (2010, S. 122) von einem signifikanten Einfluss des Geschlechts des evaluierenden Studenten auf die Lehrbewertungen. Dass auch das Alter von Studenten die Evaluationsergebnisse beeinflussen kann, wird von Nasser/Hagtvet (2006, S. 572) und von Spooren (2010, S. 128) gezeigt. Ein Einfluss dieser Faktoren auf die Lehrbewertungen bedeutet jedoch nicht zwangsläufig, dass sie auch die Wahrscheinlichkeit beeinflussen, mit der ein Student auf eine Notenverbesserung mit besseren Lehrbewertungen reagiert. Analog zu *Hypothese H2* basiert *Hypothese H3* auf den Ergebnissen von empirischen Studien, die spezifische Merkmale von Dozenten, wie ihr Alter und ihr Geschlecht (vgl. McPherson/Jewell/Kim 2009, S. 45) sowie ihre physische Attraktivität (vgl. Felton/Koper/Mitchell/Stinson 2008, S. 49; Rosar/Klein 2009, S. 632) als Einflussfaktoren auf studentische Evaluationen identifizieren.

H2: Bestimmte Charakteristika von Studenten beeinflussen die Wahrscheinlichkeit, dass sie auf eine Notenverbesserung mit besseren Lehrbewertungen reagieren.

⁴ Centra (2003, S. 516) kommt hingegen zu dem Ergebnis, dass bessere studentische Evaluationsergebnisse nicht „kaufbar“ sind. Brockx/Spooren/Mortelmans (2011), die ebenfalls einen Zusammenhang zwischen Noten und Evaluationsergebnissen finden, lassen die Interpretation dieser Beziehung offen und betonen, dass diesbezüglich weiterer Forschungsbedarf besteht (Brockx/Spooren/Mortelmans, S. 289 und S. 302).

H3: Die Wahrscheinlichkeit, dass Studenten auf Notenverbesserungen mit besseren Lehrbewertungen reagieren, unterscheidet sich zwischen verschiedenen Dozenten.

3. Methodik

3.1. Probleme bei Untersuchungen, ob bessere Lehrbewertungen „kaufbar“ sind

Der bereits in einer Vielzahl vorheriger Studien erbrachte Nachweis eines Zusammenhangs zwischen Noten und Evaluationsergebnissen (siehe u. a. Krautmann/Sandner 1999; Isely/Singh 2005; McPherson/Jewell/Kim 2009; Brockx/Spooren/Mortelmans 2011) belegt nicht zweifelsfrei, dass gute Lehrbewertungen „kaufbar“ sind. So könnte argumentiert werden, dass bessere Lehre sowohl zu besseren Noten als auch zu besseren Evaluationsergebnissen führt (vgl. Marsh 1984, S. 722) und dass der beobachtbare Zusammenhang genau dies belegt. Ob Dozenten tatsächlich bessere Lehrbewertungen „erkaufen“ können, lässt sich nicht ohne Weiteres untersuchen. So ist weder ein Experiment noch eine reale Situation vorstellbar, in der Dozenten (in glaubwürdiger Art und Weise) Studierenden bessere Noten für bessere Lehrbewertung anbieten. Direkte Befragungen von Studierenden sind zwar möglich, sie sind jedoch insofern problematisch, als die Antworten der Studenten stark verzerrt sein könnten. Zum Beispiel könnte man Studierenden eine Notenskala vorlegen und ihre Lehrbewertung bei unterschiedlichen (erwarteten oder bereits erhaltenen) Noten erfragen. Bei derartigen Befragungen besteht jedoch die Gefahr, dass die Befragten keine ehrlichen, sondern stattdessen „sozial erwünschte“ Antworten (die Lehrbewertungen werden nicht von den eigenen erwarteten oder bereits erhaltenen Noten abhängig gemacht) abgeben. Dies dürfte erklären, warum derartige Befragungen offenbar noch nicht durchgeführt wurden. Stattdessen werden häufig Evaluationsergebnisse zu einem bestimmten Zeitpunkt betrachtet. Bei diesem Vorgehen ist aber unklar, wie die befragten Studenten die jeweilige Vorlesung bewerten würden, wenn sie zum Zeitpunkt der Evaluation eine bessere Note erwarten würden oder bereits erhalten hätten. In Unkenntnis dessen könnte ein beobachtbarer Zusammenhang zwischen Noten und Evaluationsergebnissen auch darauf beruhen, dass Studierende, die unabhängig von spezifischen Eigenschaften von Lehrveranstaltungen und Dozenten gute Noten erwarten (z. B. aufgrund ihres bereits vorhandenen Wissens oder ihrer hohen Lernfähigkeit), bessere Lehrbewertungen abgeben (vgl. Marsh 1984, S. 722).

In der vorliegenden Studie wurde daher ein Ansatz gewählt, bei dem dieselben Studenten Vorlesungen zu mehreren Semesterzeitpunkten evaluiert haben (siehe Abschnitt 3.2.). Mit Hilfe der hierbei gewonnenen Daten wurde anschließend eine mögliche „Tauschsituation“ von besseren Noten gegen bessere Lehrbewertungen simuliert (siehe Abschnitt 3.3). Um zu untersuchen, ob das Ausmaß einer Notenverbesserung sowie bestimmte Charakteristika von Studenten und Dozenten Einfluss auf die Wahrscheinlichkeit haben, dass eine Notenverbesserung zu besseren Lehrbewertungen führt, wurden binäre logistische Regressionen berechnet.

3.2. Datenerhebung

In 21 Wirtschaftsvorlesungen der Westfälischen Wilhelm-Universität Münster wurden zu mehreren Semesterzeitpunkten Studentenbefragungen durchgeführt. Die Befragungsteilnehmer wurden unter anderem nach ihrer Gesamtbewertung der Vorlesung sowie der Klausurnote gefragt, die sie zum Zeitpunkt der Evaluation erwarteten oder bereits erhalten hatten. In jeder Vorlesung fanden zu mindestens drei Zeitpunkten Befragungen statt: kurz (ca. eine Woche) vor der Klausur, kurz nach der Klausur und nach Notenbekanntgabe. Zusätzlich zu diesen Befragungen wurden einige Vorlesungen auch einen Monat vor der Klausur evaluiert. Zu den Befragungszeitpunkten nach der Klausur wurden Online-Fragebögen an alle Kursteilnehmer versendet. Während des Vorlesungszeitraums wurden zumeist Online-Fragebögen, in einigen Vorlesungen aber auch Papier-Fragebögen verwendet. Um die Antworten der einzelnen Teilnehmer in den verschiedenen Befragungen zusammenführen zu können, wurden die befragten Studierenden gebeten, bei jeder Befragung eine nur ihnen bekannte Kennung anzugeben. Diese wurde so konzipiert, dass sie, wie auch die Matrikelnummer, individuell verschieden ist. Dazu wurde auf die Geburtsdaten der Eltern der befragten Studierenden zurückgegriffen.

Die Datenerhebung erfolgte über einen Zeitraum von sechs Semestern (Wintersemester 2009/2010 bis Sommersemester 2012). Bei den Befragungsteilnehmern handelte es sich überwiegend um Bachelor-Studenten, die kein rein wirtschaftswissenschaftliches Fach studierten. Dazu zählen z. B. Studenten des Studienfachs „Politik und Wirtschaft“ oder „Economics and Law“. Um die Gesamtbewertungen der Vorlesungen zu erheben, wurde den befragten Studenten eine fünfstufige Skala vorgelegt, die sich am deutschen Schulnotensystem orientiert. Die erwarteten oder bereits erhaltenen Noten wurden auf der gleichen elfstufigen Skala erhoben, die auch die wirtschaftswissenschaftliche Fakultät der Universität Münster verwendet (1,0; 1,3; 1,7; 2,0; 2,3; 2,7; 3,0; 3,3; 3,7; 4,0; 5,0).

3.3. Simulation eines „Tausches“ von guten Noten gegen gute Lehrbewertungen

Um eine Situation nachzustellen, in der Studenten möglicherweise allein aufgrund ihrer zum Evaluationszeitpunkt erwarteten oder bereits erhaltenen Noten bessere Lehrbewertungen vergeben haben, wurden in mehreren Schritten bestimmte Fälle ausgewählt. Zunächst einmal wurden nur Studenten berücksichtigt, die in einem Semester zu mindestens zwei Zeitpunkten an den Befragungen in einer Vorlesung teilgenommen haben (und bei denen Werte für alle in Tabelle 3 dargestellten Variablen vorliegen). Für diese Teilnehmer wurden alle möglichen Kombinationen von jeweils zwei Evaluationszeitpunkten gebildet. Anschließend wurden aus diesen Beobachtungen alle Fälle extrahiert, in denen Studenten zum späteren Evaluationszeitpunkt eine bessere Note aufwiesen als zum früheren Zeitpunkt (d. h. diese Studenten erwarteten zum späteren Zeitpunkt eine bessere Note oder ihre Note war besser als zuvor erwartet). Für diese Fälle wurde dann geprüft, ob die Notenverbesserung mit einer besseren Lehrbewertung einhergeht. Außerdem wurde untersucht, ob bestimmte Faktoren (das Ausmaß der Notenverbesserung, aber auch demografische Faktoren) Einfluss auf die Wahrscheinlichkeit haben, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert. Hierzu wurde mit SPSS ein binär logistisches Regressionsmodell geschätzt. Als abhängige Variable wurde eine 0-1-kodierte Variable verwendet, in der allen Fällen, in denen ein Student auf die Notenverbesserung mit einer besseren Lehrbewertung reagierte, der Wert „1“ zugewiesen wurde. Die Schätzung erfolgte gestuft über fünf Modelle. Das Grundmodell enthält ausschließlich die Variable *Notenverbesserung*. Diese Variable gibt an, um wie viel besser die Note zum späteren Evaluationszeitpunkt im Vergleich zum früheren Zeitpunkt ausfällt. In den übrigen Modellen wurden schrittweise weitere erklärende Variablen hinzugenommen.

Im Gegensatz zu den meisten vorherigen Studien werden bei dem hier gewählten Ansatz jeweils die Noten und Lehrbewertungen zweier Evaluationszeitpunkte verglichen. Die (Ausgangs-) Frage, ob Dozenten die Möglichkeit haben, bessere Lehrbewertungen zu „erkaufen“, lässt sich dadurch klarer beantworten als bei Betrachtung eines einzigen Zeitpunkts. Mit dem hier gewählten Ansatz kann nicht nur untersucht werden, ob bessere Noten mit besseren Lehrbewertungen einhergehen, sondern es kann auch überprüft werden, ob dieselben Studenten nach einer Notenverbesserung die jeweilige Lehrveranstaltung wohlwollender bewertet haben. Etwaige Befunde dieser Art lassen sich dabei nur sehr eingeschränkt als Hinweis auf eine mögliche Verbesserung der Lehrqualität interpretieren, da die Evaluationszeitpunkte nah beieinander liegen. Auch lassen sie sich nicht damit erklären,

dass individuelle Eigenschaften der Studierenden sowohl ihre Noten als auch ihre Lehrbewertungen beeinflussen. Auf diese Weise beobachtete Zusammenhänge stellen somit ein stärkeres Indiz für die eventuell bestehende Möglichkeit dar, bessere Lehrbewertungen zu „erkaufen“, als Zusammenhänge, die gefunden werden, wenn die Lehrbewertungen nur zu einem Zeitpunkt erhoben werden.

4. Empirische Ergebnisse

4.1. Lehrbewertungen derselben Studenten zu zwei Zeitpunkten

Bei paarweiser Betrachtung jeweils zweier Evaluationszeitpunkte liegen insgesamt 128 auswertbare Fälle vor, in denen ein befragter Student zum späteren Zeitpunkt eine bessere Note aufweist als zum früheren Zeitpunkt (d. h. dass er zum späteren Zeitpunkt eine bessere Note erwartet oder dass er vor der Evaluation zum späteren Zeitpunkt besser bewertet wurde als zuvor erwartet). Studenten, die eine Vorlesung zu mehr als zwei Zeitpunkten bewertet haben oder mehrere Vorlesungen zu mindestens zwei Zeitpunkten evaluierten, stellen (entsprechend der Anzahl an Vorlesungen und Kombinationsmöglichkeiten zweier Zeitpunkte) mehrere Fälle dar. Unter den 128 Fällen sind 45 Kennungen einmal, 23 Kennungen zweimal, sechs Kennungen dreimal und zwei Kennungen viermal vertreten; eine Kennung kommt fünf- und eine andere Kennung sechsmal vor. Unter der Annahme, dass die Kennung, wie auch die Matrikelnummer individuell verschieden ist, lässt sich die Aussage treffen, dass sich die Gesamtfallzahl auf 78 Studenten bezieht.

Je nach Evaluationszeitpunkt (einen Monat vor der Klausur, kurz vor der Klausur, kurz nach der Klausur oder nach Notenbekanntgabe) werden sowohl erwartete als auch bereits erhaltene Noten betrachtet. In 66,4 % der Fälle wird eine zum früheren Evaluationszeitpunkt erwartete Note mit einer zum späteren Evaluationszeitpunkt bereits bekannten Note verglichen. Für die Evaluationsergebnisse bedeutet dies, dass Lehrbewertungen vor Notenbekanntgabe mit Lehrbewertungen nach Notenbekanntgabe verglichen werden. In den übrigen Fällen (33,6 %) können die befragten Studenten zu beiden Evaluationszeitpunkten nur schätzen, welche Note sie in der Klausur erhalten werden. Für beide Notenvergleiche wurden mit SPSS Korrelationen berechnet, deren Ergebnisse in Tabelle 1 dargestellt sind. Es zeigt sich, dass miteinander verglichene Noten in beiden Fällen auf dem 0,1 %-Niveau signifikant korreliert sind. Die Koeffizienten fallen für beide Korrelationen ähnlich aus ($r = 0,912$ und $r = 0,855$) und lassen jeweils auf einen starken Zusammenhang schließen. Dieser rechtfertigt das

Vorgehen, bei nachfolgenden Analysen nicht zwischen erwarteten und tatsächlichen Noten zu unterscheiden.

		Zum späteren Evaluationszeitpunkt	
		erwartete Note	bekannte Note
Zum früheren Evaluationszeitpunkt	Korrelationskoeffizient	0,912***	0,855***
erwartete Note	N	43	85

Hinweis: *** kennzeichnet statistische Signifikanz auf dem 0,1 %-Niveau.

Tabelle 1: Spearman-Korrelation von erwarteten und bekannten Noten

Beim Vergleich des früheren mit dem späteren Evaluationszeitpunkt ist die Differenz der Noten signifikant positiv mit der Differenz der Lehrbewertungen korreliert ($p \leq 0,05$; $r = 0,205$; $N = 128$). Zum späteren Evaluationszeitpunkt (d. h. nach Notenverbesserung) wird signifikant besser evaluiert als zum früheren Zeitpunkt. Ein Wilcoxon-Test, dessen Ergebnisse in Tabelle 2 dargestellt sind, findet einen auf dem 1 %-Niveau signifikanten Unterschied. Zum späteren Zeitpunkt fallen die Noten im Durchschnitt um 0,79 Notenpunkte und die Lehrbewertungen um 0,20 Notenpunkte besser aus. Fälle, in denen zum späteren Zeitpunkt besser evaluiert wird ($N = 36$), sind mehr als doppelt so häufig vertreten, wie gegenteilige Fälle ($N = 16$). Es ist aber auch erkennbar, dass in über der Hälfte der Fälle (59,4 %) die jeweilige Vorlesung zu beiden Zeitpunkten gleich bewertet wird, obwohl in allen Fällen die Noten der evaluierenden Studenten zum späteren Zeitpunkt besser ausfallen.

Noten-verbesserung (Durchschnitt)	LB_Z0 minus LB_Z1 (Durchschnitt)	Wilcoxon-Test			
		LB_Z0 > LB_Z1	LB_Z0 = LB_Z1	LB_Z0 < LB_Z1	Signifikanz
0,786	0,195	36	76	16	0,002

Tabelle 2: Lehrbewertungen (LB) von Studenten, die dieselbe Vorlesung zu zwei Zeitpunkten evaluiert haben, wobei sie zum späteren Zeitpunkt (Z1) eine besseren Note aufweisen als zum früheren Zeitpunkt (Z0)

4.2. Betrachtete Einflussfaktoren auf die Wahrscheinlichkeit, dass Studenten auf eine Notenverbesserung mit besseren Lehrbewertungen reagieren

Tabelle 3 gibt einen Überblick darüber, für welche Faktoren untersucht wurde, ob sie die Wahrscheinlichkeit beeinflussen, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert.

Variable	Beschreibung	Codierung	Fälle	%
Noten-verbesserung	Mittelwert = 0,79 Minimum = 0,30 Maximum = 2,30		128	
Geschlecht		0: Frauen 1: Männer	93 35	72,7 27,3
Alter	Mittelwert = 22,34 Minimum = 18 Maximum = 27		128	
Ausbildung	vor Studienbeginn absolvierte Ausbildung	0: keine Ausbildung 1: Ausbildung	91 37	71,7 28,9
Nebenjob	Studentischer Nebenjob	0: kein Nebenjob 1: Nebenjob	60 68	46,9 53,1
Eltern_HA	Hochschulabschluss (HA) der Eltern	0: kein Elternteil 1: ein Elternteil 2: beide Elternteile	49 41 38	38,3 32,0 29,7
Eltern_WiWi	Arbeitsplatz der Eltern im kaufmännischen oder wirtschaftswissenschaftlichen Bereich	0: kein Elternteil 1: mindestens ein Elternteil	71 57	55,5 44,5
Dozent	Dozent, dessen Vorlesung evaluiert wurde	1: Dozent 1 2: Dozent 2 3: Dozent 3 4: Dozent 4	13 75 26 14	10,2 58,6 20,3 10,9
Relativ_bessere_Note	vNote: Selbstbewertung Z0 und Z1: früherer bzw. späterer Evaluationszeitpunkt	0: $(\text{NoteZ0} - \text{vNoteZ0}) \leq (\text{NoteZ1} - \text{vNoteZ1})$ 1: andernfalls	35 93	27,3 72,7
Note_bekannt	Die Klausurnote ist zum späteren Evaluationszeitpunkt bekannt.	0: nein 1: ja	43 85	33,6 66,4

Tabelle 3: Übersicht über die betrachteten Variablen

Die Variable *Notenverbesserung* gibt an, um wie viel besser die Note zum späteren Evaluationszeitpunkt im Vergleich zum früheren Zeitpunkt ausfällt. Dabei ist zu berücksichtigen, dass ausschließlich Studenten betrachtet werden, deren Noten zum zweiten Evaluationszeitpunkt besser ausfallen als zum ersten Zeitpunkt. Neben den Noten wird auch

die Selbstbewertung der Studierenden (als „verdient“ angesehene Note) berücksichtigt. Dazu wurde die Differenz von Note und Selbstbewertung ($vNote$) berechnet, wobei jeweils die Selbstbewertung von der Note subtrahiert wurde. Dabei gilt: Je kleiner der sich daraus ergebende Wert ist, desto besser ist die Note im Vergleich zur Selbstbewertung. So bedeutet ein negativer Wert, dass der befragte Student erwartet, zu gut bewertet zu werden oder dass er seiner Meinung nach bereits zu gut bewertet wurde. Fällen, bei denen die Differenz zwischen Note und Selbstbewertung zum späteren Evaluationszeitpunkt kleiner ist als zum früheren Evaluationszeitpunkt ist in der 0-1-kodierten Variable *Relativ_bessere_Note* der Wert „1“ zugewiesen.⁵ Daneben gibt die Variable *Note_bekannt* an, ob die Note zum späteren Evaluationszeitpunkt bereits bekannt war.

4.3. Binäre logistische Regressionen

Um zu untersuchen, welche Faktoren Einfluss auf die Wahrscheinlichkeit haben, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert, wurden binäre logistische Regressionen berechnet. Die Ergebnisse sind in Tabelle 4 zusammengefasst. Als abhängige Variable dient jeweils die 0-1-kodierte Variable *bessere_Lehrbewertung_nach_Notenverbesserung*. In dieser Variablen sind alle Fälle, auf die dies zutrifft, mit dem Wert 1 kodiert. Erklärende Variablen sind die in Tabelle 3 dargestellten Variablen.

Das Ausmaß der *Notenverbesserung* hat in allen Modellen einen signifikant positiven Einfluss auf die Wahrscheinlichkeit, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert. Das Signifikanzniveau variiert zwischen 1 % und 0,1 %. Modell 1, in dem die Notenverbesserung die einzige erklärende Variable ist, erklärt 7,5 % der Varianz der abhängigen Variablen. Im Vergleich dazu hat Modell 5, in dem alle (in Tabelle 3 dargestellten) Faktoren berücksichtigt sind, einen deutlich höheren Erklärungsgehalt (30,6 %). Die Variable *Geschlecht* ist ebenfalls in allen Modellen signifikant. Die Wahrscheinlichkeit, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert, ist bei männlichen Studenten signifikant höher als bei weiblichen Studenten. In den Modellen 2 und 5 liegt das Signifikanzniveau bei 5 % und in den Modellen 3 und 4 bei 1 %. Die Variable *Alter* ist in keinem der fünf Modelle signifikant. Werden neben der Notenverbesserung auch das Geschlecht und das Alter mit in die Regression aufgenommen, so steigt der erklärte Varianzanteil von 7,5 % auf 11,8 %.

⁵ Die zwischen den Variablen *Notenverbesserung* und *Relativ_bessere_Note* beobachtbare signifikante Korrelation ($p \leq 0.01$ und $r = 0,385$) ist gemäß Brosius 2011 (S. 523) als schwach einzustufen.

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
Konstante	-1,81/1,63 *** [0,41]	-0,27/0,76 [2,10]	-4,75/0,01 (*) [2,78]	-5,55/0,00 (*) [2,88]	-7,46/0,00 * [3,26]
Notenverbesserung	1,06/2,87 ** [0,41]	1,30/3,67 ** [0,45]	1,44/4,23 ** [0,48]	1,82/6,14 *** [0,55]	2,00/7,29 *** [0,59]
Geschlecht (männlich)		0,93/2,54 * [0,47]	1,54/4,68 ** [0,60]	1,54/4,68 ** [0,60]	1,46/4,31 * [0,62]
Alter		-0,09/0,91 [0,10]	0,07/1,07 [0,13]	0,13/1,13 [0,13]	0,20/1,23 [0,15]
Eltern_HA					
Eltern_HA1			0,41/1,51 [0,55]	0,48/1,62 [0,56]	0,38/1,46 [0,61]
Eltern_HA2			1,07/2,91 (*) [0,58]	1,21/3,35 * [0,61]	1,26/3,52 * [0,63]
Eltern_WiWi			1,08/2,94* [0,50]	0,96/2,61 (*) [0,50]	1,06/2,90 * [0,54]
Ausbildung			-1,13/0,32(*) [0,65]	-1,22/0,30 (*) [0,66]	-1,61/0,20 * [0,73]
Nebenjob			-0,06/0,94 [0,46]	-0,16/0,86 [0,47]	-0,24/0,79 [0,50]
Relativ_bessere_Note				-0,90/0,41 [0,55]	-0,78/0,46 [0,60]
Note_bekannt				-0,19/0,83 [0,51]	-0,41/0,66 [0,55]
Dozent					
Dozent 1					-0,65/0,52 [1,04]
Dozent 3					0,34/1,40 [0,71]
Dozent 4					1,57/4,81 * [0,79]
R ² (Nagelkerke)	0,075	0,118	0,227	0,253	0,306
Beobachtungen	128	128	128	128	128

Hinweise: Die Standardfehler sind in Klammern angegeben. Vor dem Schrägstrich ist jeweils der (nicht standardisierte) Regressionskoeffizient angegeben; dahinter das Chancen-/Quotenverhältnis. (*)/**/*** kennzeichnet Signifikanz auf dem 10/5/1/0,1 %-Niveau.

Tabelle 4: Binäre logistische Regressionen mit der 0-1-kodierten Variable bessere_Lehrbewertung_nach_Notenverbesserung als abhängiger Variable

In Modell 3 werden zusätzlich zu den erklärenden Variablen in Modell 2 (Notenverbesserung, Geschlecht und Alter) auch die Antworten auf die Fragen nach einem Hochschulabschluss der Eltern (*Eltern_HA*), einer Berufstätigkeit der Eltern im kaufmännischen oder wirtschaftswissenschaftlichen Bereich (*Eltern_WiWi*), einer vor Studienbeginn abgeschlossenen Ausbildung und der Ausübung eines studentischen Nebenjobs berücksichtigt. Die Hinzunahme dieser Variablen erhöht den erklärten Varianzanteil von 11,8 % auf 22,7 %. Die Wahrscheinlichkeit, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert, ist signifikant höher, wenn beide Elternteile einen Hochschulabschluss haben (*Eltern_HA2*). Die Referenzgruppe bilden dabei Studenten, die die Frage nach einem Hochschulabschluss der Eltern verneinten. In Modell 3 ist die Variable *Eltern_HA2* nur schwach signifikant; in den Modellen 4 und 5, die weitere Einflussfaktoren beinhalten, ist sie auf dem 5 %-Niveau signifikant. Studenten, bei denen ein Elternteil einen Hochschulabschluss hat (*Eltern_HA1*), unterscheiden sich hingegen nicht signifikant von der Referenzgruppe.

Eine Berufstätigkeit mindestens eines Elternteils im kaufmännischen oder wirtschaftswissenschaftlichen Bereich (*Eltern_WiWi*) erhöht die Wahrscheinlichkeit, dass nach einer Notenverbesserung bessere Lehrbewertungen vergeben werden. In Modell 3 und in Modell 5 (das alle Faktoren berücksichtigt) ist die Variable *Eltern_WiWi* auf dem 5 %-Niveau signifikant; in Modell 4 ist sie nur schwach signifikant (auf dem 10 % Niveau). Die Variable *Ausbildung* ist in Modell 5 (das alle Faktoren einbezieht) auf dem 5 %-Niveau signifikant; in den Modellen 3 und 4 ist sie schwach signifikant. Die negativen Vorzeichen der Regressionskoeffizienten lassen darauf schließen, dass eine vor Studienbeginn absolvierte Ausbildung die Wahrscheinlichkeit senkt, mit der eine Notenverbesserung zu einer besseren Lehrbewertung führt. Ein studentischer Nebenjob hat hingegen keinen signifikanten Einfluss.

In Modell 4 werden zusätzlich zu den in Modell 3 berücksichtigten Variablen die Variablen *Relativ_bessere_Note* (gemeint ist, dass die Differenz zwischen Note und Selbstbewertung zum späteren Evaluationszeitpunkt kleiner ist als zu früheren Evaluationszeitpunkt) und *Note_bekannt* (d. h. zum späteren Evaluationszeitpunkt sind die Noten bekannt) berücksichtigt. Beide Variablen sind jedoch insignifikant. Modell 5 entspricht Modell 4, ergänzt um die 0-1-kodierten Variablen für die Dozenten. Die Hinzunahme dieser Variablen erhöht den erklärten Varianzanteil von 25,3 % (Modell 4) auf 30,6 %. Die Wahrscheinlichkeit, dass ein Student auf eine Notenverbesserung mit einer besseren Lehrbewertung

reagiert, ist bei Dozent 4 signifikant höher als bei der Referenzkategorie (Dozent 2). Die Dozenten 1 und 3 unterscheiden sich hier hingegen nicht signifikant von der Referenz.

5. Diskussion

Die Ergebnisse dieser Studie zeigen, dass ein Student mit einer umso höheren Wahrscheinlichkeit auf eine Notenverbesserung mit einer besseren Lehrbewertung reagiert, je stärker die Notenverbesserung ist. Dieses Ergebnis entspricht der in *Hypothese H1* formulierten Erwartung. Bestimmte Charakteristika von Studenten (Geschlecht, Elternhaus sowie eine vor Studienbeginn absolvierte Ausbildung) haben aber ebenfalls Einfluss auf die Wahrscheinlichkeit, dass ein Student nach einer Notenverbesserung eine positivere Lehrbewertung abgibt. Außerdem zeigen sich diesbezüglich Unterschiede zwischen Dozenten. Somit können auch die *Hypothesen H2* und *H3* bestätigt werden.

Die Ergebnisse deuten darauf hin, dass sich Zusammenhänge zwischen Noten und Lehrbewertungen nicht allein auf den ‚Effektive-Lehre-Ansatz‘ (vgl. Marsh 1984, S. 722) zurückführen lassen. So werden ausschließlich Studierende betrachtet, deren Noten zum späteren Evaluationszeitpunkt besser sind als zum früheren Zeitpunkt. Dabei ist zu beobachten, dass die Lehrbewertungen zum späteren Zeitpunkt signifikant besser ausfallen. Eine Verbesserung der Lehre, die eine Erklärung dafür sein könnte, ist nur zwischen dem ersten und dem zweiten Befragungszeitpunkt möglich (d. h. wenn Lehrbewertungen, die ca. einen Monat vor der Klausur erhoben wurden, mit Lehrbewertungen verglichen werden, die kurz vor der Klausur erhoben wurden). Fälle, in denen diese beiden Zeitpunkte verglichen werden, haben aber nur einen Anteil von 7,0 % an der Gesamtfallzahl. Motiviert durch die Erkenntnis, durch die Vorlesung besser auf die Klausur vorbereitet worden zu sein als ex ante erwartet, haben einige Studenten aber möglicherweise nach der Klausur oder nach der Notenbekanntgabe bessere Lehrbewertungen abgegeben. In diesen Fällen wird jedoch nicht die didaktische Qualität der Vorlesung bewertet, sondern nur das Ausmaß, in dem die Vorlesung die Studierenden auf die Klausur vorbereitet hat. Nach diesem Bewertungskriterium können aber auch Vorlesungen, in denen Studenten sehr wenig lernen, sehr gute Lehrbewertungen erzielen. Fälle, auf die diese Begründung zutrifft, lassen sich somit eigentlich nicht mehr mit dem ‚Effektive-Lehre-Ansatz‘ erklären. Der ‚Studenten-Charakteristika-Ansatz‘ argumentiert hingegen, dass von vornherein bestehende Unterschiede in den Charakteristika der Studenten sowohl ihre Noten als auch ihre Lehrbewertungen beeinflussen (vgl. Marsh 1984, S. 722). Mit diesem Ansatz lassen sich die Ergebnisse kaum erklären, da jeweils Lehrbewertungen

derselben Studenten verglichen werden. Die Ergebnisse dieser Studie liefern somit einen klaren Hinweis für das Zutreffen des ‚Käuflichkeitsansatzes‘. Dies schließt jedoch keineswegs aus, dass die anderen beiden Ansätze in anderen Kontexten – beispielsweise bei Vergleichen der Lehrbewertungen verschiedener Studenten – ebenfalls als Erklärungen für beobachtbare Zusammenhänge zwischen Noten und Evaluationsergebnissen in Betracht kommen.

Ob Dozenten die Möglichkeit haben, bessere Lehrbewertungen zu „erkaufen“, hängt den Ergebnissen zufolge nicht nur von ihrer Notengebung ab, sondern auch von der Kurszusammensetzung. So lassen die Ergebnisse der binären logistischen Regressionen darauf schließen, dass das „Erkaufen“ guter Lehrbewertungen umso eher möglich ist, je höher der Anteil an Studenten ist, die männlichen Geschlechts sind, deren Eltern beide einen Hochschulabschluss haben und von denen mindestens ein Elternteil im kaufmännischen oder wirtschaftswissenschaftlichen Bereich arbeitet. In Vorlesungen, in denen mehr Studenten vor Beginn ihres Studiums eine Ausbildung absolviert haben, erscheint das „Erkaufen“ guter Lehrbewertungen hingegen schwieriger. Aber auch Charakteristika von Dozenten scheinen Einfluss auf die Wahrscheinlichkeit zu haben, dass Studenten auf eine Notenverbesserung mit besseren Lehrbewertungen reagieren. Die Ergebnisse liefern wichtige Erkenntnisse für den Hochschulalltag. Sie lassen darauf schließen, dass die Möglichkeit, bessere Lehrbewertungen zu „erkaufen“, für verschiedene Kurszusammensetzungen und Dozenten in unterschiedlichem Ausmaß besteht.

Obwohl ausschließlich Studenten betrachtet werden, deren Noten zum früheren Evaluationszeitpunkt schlechter ausfallen als zum späteren Zeitpunkt, geben nur 28,1 % der Studenten zum späteren Zeitpunkt eine bessere Lehrbewertung ab. Mehr als die Hälfte der Studenten evaluiert zu beiden Zeitpunkten gleich. Dass der Anteil der Studenten, der zum zweiten Zeitpunkt bessere Lehrbewertungen abgibt, nicht höher ausfällt, könnte daran liegen, dass für die Datenerhebung dieselben Skalen verwendet wurden, die auch die Universität Münster benutzt. Dies bedeutet, dass die Noten auf einer 11-stufigen und die Lehrbewertungen auf einer fünfstufigen Skala erhoben wurden. Von daher ist nicht auszuschließen, dass der Anteil der Studenten, die auf die Notenverbesserung mit besseren Lehrbewertungen reagiert haben, höher ausgefallen wäre, wenn die Lehrbewertungen ebenfalls auf einer 11-stufigen Skala erhoben worden wären. Möglicherweise hat aber auch die Auswahl der befragten Studenten zu einer Selbstselektion geführt, die den Anteil derjenigen Studenten reduziert haben könnte, die auf die Notenverbesserung mit besseren Lehrbewertungen reagieren haben. So wurden nur Studenten ausgewählt, die in einem Semester in mindestens einer Vorlesung zu zwei

Zeitpunkten an den Befragungen teilgenommen haben. Dies könnten aber genau diejenigen Studenten sein, die motivierter sind und stärker darum bemüht sind, objektive Lehrbewertungen abzugeben als ihre Kommilitonen, die sich gar nicht oder nur zu einem Zeitpunkt an den Befragungen beteiligt haben. Eine etwaige Selbstselektion könnte aber genauso gut andersherum stattgefunden haben. So könnten auch gerade diejenigen mehrfach an den Befragungen teilgenommen haben, die auf Notenverbesserungen mit besseren Lehrbewertungen reagieren. Dass sich die jeweils verglichenen Lehrbewertungen aufgrund der verschiedenen Erhebungszeitpunkte unterscheiden, erscheint wenig wahrscheinlich. Zwar findet Lütkenhöner (2012) zum Teil signifikante Unterschiede zwischen den zu zwei Zeitpunkten erhobenen Lehrbewertungen, die Ergebnisse zeigen aber auch, dass die Lehrbewertungen tendenziell umso besser ausfallen, je früher sie im Semester erhoben wurden (vgl. Lütkenhöner 2012, S. 24). Außerdem beobachtet Lütkenhöner (2012) keine signifikanten Unterschiede zwischen den – auch in der vorliegenden Studie betrachteten – Gesamtbewertungen einer Vorlesung, die von denselben Studenten zu verschiedenen Zeitpunkten evaluiert wurde (vgl. Lütkenhöner 2012, S. 21). Die Studien von Kohlan (1973, S. 591) und Pruitt/Dicks/Tilley (2011, S. 12) ergeben ebenfalls keine signifikanten systematischen Unterschiede zwischen den zu verschiedenen Zeitpunkten abgegebenen Lehrbewertungen.⁶ Somit lassen die Ergebnisse darauf schließen, dass die betrachteten Studenten zum zweiten Evaluationszeitpunkt allein aufgrund der Notenverbesserung bessere Lehrbewertung abgegeben haben. Dass dieselben Studenten zu einem bestimmten Evaluationszeitpunkt bessere Lehrbewertungen abgegeben hätten, wenn ihre Noten zu diesem Zeitpunkt besser gewesen wären, lässt sich nur vermuten. Für diese Vermutung spricht jedoch, dass nicht nur aufeinander folgende Evaluationszeitpunkte verglichen wurden, sondern paarweise jeweils alle möglichen Kombinationen von Zeitpunkten. Ob diese Vermutung zutrifft, ließe sich nur mit direkten Befragungen prüfen, was jedoch in Bezug auf die Qualität der so gewonnenen Daten problematisch ist. Bei dem hier verwendeten Ansatz ist für die befragten Studenten zwar auch ersichtlich, dass Zusammenhänge von Noten und Lehrbewertungen untersucht werden (da beides abgefragt wird), durch die Wahl mehrere Evaluationszeitpunkte hat der Ansatz jedoch eher den Charakter einer Feldstudie. Von daher ist zu erwarten, dass etwaige Verzerrungen durch ‚sozial erwünschtes‘ Antwortverhalten geringer ausfallen.

⁶ In der Studie von Pruitt/Dicks/Tilley (2011) zeigen sich allerdings signifikante Unterschiede, wenn zwischen „upper-devision-“ und „lower-devision-“ Studenten unterschieden wird. Letztere neigen dazu, am Ende des Semesters schlechtere Lehrbewertungen zu vergeben, während bei Ersteren das Gegenteil der Fall ist (vgl. Pruitt/Dicks/Tilley 2011, S. 12).

In dieser Studie werden ausschließlich Studenten betrachtet, die das Fach Wirtschaftswissenschaften zumindest als eines von mehreren Fächern belegt haben (v. a. Zwei-Fach-Bachelor-Studenten). In Anbetracht dieser Tatsache sollte nicht unberücksichtigt bleiben, dass die Ergebnisse einiger empirischer Studien darauf hindeuten, dass sich Wirtschaftsstudenten egoistischer und profitorientierter verhalten als Studenten anderer Fachdisziplinen (vgl. u. a. Selten/Ockenfels 1998, S. 530; Carter/Irons 1991, S. 176; Frank/Schulze 2000, S. 101). Dies könnte auch erklären, warum Studenten, deren Eltern im kaufmännischen oder wirtschaftswissenschaftlichen Bereich arbeiten, mit einer signifikant höheren Wahrscheinlichkeit auf eine Notenverbesserung mit besseren Lehrbewertungen reagieren. Zudem ist die Wahrscheinlichkeit einer solchen Reaktion bei Wirtschaftsstudenten möglicherweise höher als bei Studenten anderer Studiengänge.⁷ Für den Hochschulalltag wären derartige Unterschiede jedoch von untergeordneter Relevanz, da Vergleiche von Lehrbewertungen in der Regel innerhalb einer Fachdisziplin durchgeführt werden. Deutlich problematischer ist daher, dass die Ergebnisse dieser Studie darauf hindeuten, dass einige Dozenten aufgrund der Kurszusammensetzung und/oder eigener Charakteristika eher die Möglichkeit haben bessere Lehrbewertungen zu „erkaufen“ als andere Dozenten.

6. Fazit

Die Ergebnisse dieser Studie lassen darauf schließen, dass sich Zusammenhänge zwischen Noten und Evaluationsergebnissen nicht allein mit dem ‚*Effektive-Lehre-*‘ und/oder dem ‚*Studenten-Charakteristika-Ansatz*‘ erklären lassen. Vielmehr liefern sie einen klaren Hinweis darauf, dass Dozenten durchaus die Möglichkeit haben, sich bessere Lehrbewertungen zu „erkaufen“, indem sie bessere Noten vergeben oder indem sie Studenten bessere Noten erwarten lassen. Es zeigt sich aber auch, dass nicht nur das Ausmaß einer Notenverbesserung sondern auch Charakteristika von Studenten und Dozenten Einfluss auf die Wahrscheinlichkeit haben, dass ein Student nach einer Notenverbesserung eine bestimmte Lehrveranstaltung besser bewertet. Die Wahrscheinlichkeit hierfür ist signifikant höher, wenn der Student männlich ist, beide Elternteile einen Hochschulabschluss haben und mindestens ein

⁷ Unterschiede zwischen verschiedenen Studiengängen wurden in dieser Studie aus mehreren Gründen nicht untersucht. Studenten, die ausschließlich Wirtschaftswissenschaften (vor allem BWL) studieren, unterscheiden sich auch darin von den übrigen Studenten, dass sie sich fast alle bereits im Masterstudium befinden. Bei einem entsprechenden Vergleich wäre somit unklar, ob etwaige beobachtbare Unterschiede auf den Studiengang oder auf die Tatsache zurückzuführen sind, dass Bachelor- mit Masterstudenten verglichen werden. Zudem wurde der Studiengang oftmals nicht mit angegeben, so dass seine Berücksichtigung die Fallzahl deutlich reduzieren würde.

Elternteil im kaufmännischen oder wirtschaftswissenschaftlichen Bereich arbeitet. Bei Studenten, die vor Beginn ihres Studiums eine Ausbildung absolviert haben, erscheint das „Erkaufen“ guter Lehrbewertungen hingegen schwieriger. Die Ergebnisse liefern damit wichtige Erkenntnisse für den Hochschulalltag. Sie lassen darauf schließen, dass die Möglichkeit, bessere Lehrbewertungen zu „erkaufen“, für verschiedene Kurszusammensetzungen und Dozenten in unterschiedlichem Ausmaß besteht.

Literatur

- Brockx, Bert/Spooren, Pieter/Mortelmans, Dimitri (2011): „Taking the grading leniency story to the edge: The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education“, *Educational Assessment, Evaluation and Accountability* 23, S. 289-306.
- Brosius, Felix (2011): „SPSS 19“, Frechen: mitp.
- Carter, John R./Irons, Michael D. (1991): „Are economists different, and if so, why?“, *Journal of Economic Perspectives* 5, S. 171-177.
- Centra, John A. (2003): „Will teachers receive higher student evaluations by giving higher grades and less course work?“, *Research in Higher Education* 44, S. 495-518.
- Ewing, Andrew M. (2012): „Estimating the impact of relative expected grade on student evaluations of teachers“, *Economics of Education Review* 31, S. 141-154.
- Felton, James/Koper, Peter T./Mitchell, John/Stinson, Michael (2008): „Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com“, *Assessment and Evaluation in Higher Education* 33, S. 45-61.
- Frank, Björn/Schulze, Günther G. (2000): „Does economics make citizens corrupt?“, *Journal of Economic Behavior and Organization* 43, S. 101-113.
- Griffin, Bryan W. (2004): „Grading leniency, grade discrepancy, and student ratings of instruction“, *Contemporary Educational Psychology* 29, S. 410-425.
- Hancock, Gregory R./Shannon, David M./Trentham, Landa L. (1993): „Student and teacher gender in ratings of university faculty: results from five colleges of study“, *Journal of Personnel Evaluation in Education* 6, S. 235-248.
- Isely, Paul/Singh, Harinder (2005): „Do higher grades lead to favorable student evaluations?“, *Journal of Economic Education* 36, S. 29-42.
- Kohlan, Richard G. (1973): „A comparison of faculty evaluations early and late in the course“, *Journal of Higher Education* 44, S. 587-595.
- Krautmann, Anthony C./Sander, William (1999): „Grades and student evaluations of teachers“, *Economics of Education Review* 18, S. 59-63.
- Lütkenhöner, Laura (2012): „Effekte von Erhebungsart und -zeitpunkt auf studentische Evaluationen“, *Diskussionspapier des Instituts für Organisationsökonomik* 8/2012, Westfälische Wilhelms-Universität Münster, im Internet unter: https://www.wiwi.uni-muenster.de/io/forschen/downloads/DP-IO_08_2012.pdf (letzter Aufruf 30.07.2013).

- Marsh, Herbert W. (1984): „Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility“, *Journal of Educational Psychology* 76, S. 707-754.
- McPherson, Michael A. (2006): „Determinants of how students evaluate teachers“, *Journal of Economic Education* 37, S. 3-20.
- McPherson, Michael A./Jewell R. Todd (2007): „Leveling the playing field: Should student evaluation scores be adjusted?“, *Social Science Quarterly* 88, S. 868-881.
- McPherson, Michael A./Jewell, R. Todd/Kim, Myungsup (2009): „What determines student evaluation scores? A random effects analysis of undergraduate economics classes“, *Eastern Economic Journal* 35, S. 37-51.
- Nasser, Fadia/Hagtvet, Knut A. (2006): „Multilevel analysis of the effects of student and instructor/course characteristics on student ratings“, *Research in Higher Education* 47, S. 559-590.
- Pruitt, J. Ross/Dicks, Michael R./Tilley, Daniel S. (2011): „Divergence and magnitude of student evaluations of teaching“, im Internet unter: <http://ssrn.com/abstract=1887809>, Stand Februar 2012 (letzter Aufruf 30.07.2013).
- Rosar, Ulrich/Klein, Markus (2009): „Mein(schöner)Prof.de: Die physische Attraktivität des akademischen Lehrpersonals und ihr Einfluss auf die Ergebnisse studentischer Lehrevaluationen“, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 61, S. 621-645.
- Selten, Reinhard/Axel Ockenfels (1998): „An experimental solidarity game“, *Journal of Economic Behavior and Organization* 34, S. 517-539.
- Spooren, Pieter (2010): „On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching“, *Studies in Educational Evaluation* 36, S. 121-131.

Bisher erschienen:

Diskussionspapiere des Instituts für Organisationsökonomik

- DP-IO 7/2013** Können sich Hochschuldozenten bessere studentische Lehrevaluationen „erkaufen“?
Laura Lütkenhöner
Juli 2013
- DP-IO 6/2013** Scholars' Physical Appearance, Research Performance and Feelings of Happiness
Alexander Dilger/Laura Lütkenhöner/Harry Müller
Juni 2013
- DP-IO 5/2013** Vor- und Nachteile der W-Besoldung
Alexander Dilger
Mai 2013
- DP-IO 4/2013** Hochschulräte in NRW
Mehr Hochschulfreiheit oder Staatseinfluss?
Alexander Dilger
April 2013
- DP-IO 3/2013** Soll man das Handelsblatt-Ranking BWL boykottieren?
Alexander Dilger
März 2013
- DP-IO 2/2013** Composition Effects of the German Federal Government on the Average Top Income Tax Burden
Katrin Scharfenkamp
Februar 2013
- DP-IO 1/2013** Der Einfluss des Forschungsschwerpunkts auf den Zitationserfolg
Eine empirische Untersuchung anhand der Gesamtpublikationen deutschsprachiger Hochschullehrer für BWL
Harry Müller/Alexander Dilger
Januar 2013
- DP-IO 12/2012** Wettbewerbsvorteile aufgrund des Vornamens?
Feldexperimente auf dem Beziehungs-, Nachhilfe- und Wohnungsmarkt
Laura Lütkenhöner
Dezember 2012
- DP-IO 11/2012** The Impact of the Euro 2012 on Popularity and Market Value of Football Players
Stephanie Kiefer
November 2012
- DP-IO 10/2012** 2. Jahresbericht des Instituts für Organisationsökonomik
Alexander Dilger/Stephanie Kiefer
Oktober 2012
- DP-IO 9/2012** How (Not) to Pay Non-executive Directors
Alexander Dilger
September 2012
- DP-IO 8/2012** Effekte von Erhebungsart und -zeitpunkt auf studentische Evaluationsergebnisse
Laura Lütkenhöner
August 2012

- DP-IO 7/2012** Prolegomena zu einer Analyse ethischer und anderer Normen am Beispiel des Hochschulmanagements
Alexander Dilger
Juli 2012
- DP-IO 6/2012** The Impact of Physical Attractiveness on the Popularity of Female Tennis Players in Online Media
Stephanie Kiefer/Katrin Scharfenkamp
Juni 2012
- DP-IO 5/2012** Förderung von Wissenschaft zu nationalen und europäischen Fragen
Alexander Dilger
Mai 2012
- DP-IO 4/2012** Untersuchung von Indikatoren zur Qualitätsmessung von Reitschulen in Deutschland
Stephanie Kiefer
April 2012
- DP-IO 3/2012** Rigor, wissenschaftliche und praktische Relevanz
Alexander Dilger
März 2012
- DP-IO 2/2012** Socio-Demographic Characteristics and Human Capital of the German Federal Government's Members
Katrin Scharfenkamp/Alexander Dilger
Februar 2012
- DP-IO 1/2012** Die Zitationshäufigkeit als Qualitätsindikator im Rahmen der Forschungsleistungsmessung
Harry Müller
Januar 2012
- DP-IO 12/2011** Ein Forschungsleistungsranking auf der Grundlage von Google Scholar
Alexander Dilger/Harry Müller
Dezember 2011
- DP-IO 11/2011** Besonderheiten der Bewerbung um Promotionsstellen und -gelegenheiten
Alexander Dilger
November 2011
- DP-IO 10/2011** 1. Jahresbericht des Instituts für Organisationsökonomik
Alexander Dilger/Stephanie Kiefer/Katrin Scharfenkamp
Oktober 2011
- DP-IO 9/2011** Corporate Governance and Employee Power in the Boardroom
An Applied Game Theoretical Analysis
Benjamin Balsmeier/Andreas Bermig/Alexander Dilger/Hannah Geyer
September 2011
- DP-IO 8/2011** Ein Ranking von Hochschulen und (Bundes-)Ländern
am Beispiel der Betriebswirtschaftslehre
Harry Müller/Alexander Dilger
August 2011
- DP-IO 7/2011** Befragung der Kommission Hochschulmanagement zu VHB-JOURQUAL
Alexander Dilger
Juli 2011

- DP-IO 6/2011** Director Interlocks and Executive Turnover in German Public Corporations
A Hazard Analysis for the Period from 1996 to 2008
Benjamin Balsmeier/Achim Buchwald/Alexander Dilger/Jörg Lings
Juni 2011
- DP-IO 5/2011** Personalökonomik
Stärken, Schwächen und ihr Platz in der Personalwirtschaftslehre
Alexander Dilger
Mai 2011
- DP-IO 4/2011** Familienbewusste Personalpolitik und Unternehmenserfolg
Eine empirische Untersuchung
Christian Lehmann
April 2011
- DP-IO 3/2011** Welche Unternehmen berufen Vorstandsvorsitzende und andere Vorstände als externe
Kontrolleure?
Eine empirische Analyse der Präsenz von externen Vorständen in den Aufsichtsräten
deutscher Großunternehmen
Achim Buchwald
März 2011
- DP-IO 2/2011** Hat Julia aufgrund ihres Vornamens Wettbewerbsvorteile gegenüber Ayse und
Chantal?
Ein Experiment auf dem Beziehungs-, Nachhilfe- und Wohnungsmarkt
Laura Lütkenhöner
Februar 2011
- DP-IO 1/2011** Die dunkle Seite der Gerechtigkeit
Alexander Dilger
Januar 2011
- DP-IO 3/2010** On the Overconfidence-Effect in Teams
Hanke Wickhorst
Dezember 2010
- DP-IO 2/2010** Leistung, Identifikation oder die Unsicherheit über den Spielausgang – was zählt
wirklich?
Relevante Einflussfaktoren auf die Zuschauerzahlen in der Basketball-Bundesliga
Hannah Geyer
November 2010
- DP-IO 1/2010** A Citation Based Ranking of German-speaking Researchers in Business
Administration with Data of Google Scholar
Alexander Dilger/Harry Müller
Oktober 2010



Herausgeber:
Prof. Dr. Alexander Dilger
Westfälische Wilhelms-Universität Münster
Institut für Organisationsökonomik
Scharnhorststr. 100
D-48151 Münster

Tel: +49-251/83-24303

Fax: +49-251/83-28429

www.wiwi.uni-muenster.de/io

