Economics of Education Review 31 (2012) 141–154

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Estimating the impact of relative expected grade on student evaluations of teachers

Andrew M. Ewing*

Eckerd College, Collegium of Behavioral Sciences, 4200 54th Avenue South, St. Petersburg, FL 33711, United States

ARTICLE INFO

Article history:
Received 20 July 2010
Received in revised form 16 July 2011
Accepted 14 October 2011

JEL classification:
I20
I21
I23

Keywords:
Student evaluations of teachers
Grades

ABSTRACT

Grade inflation over the past few decades has been a concern for many universities. Course evaluation scores are known to be positively correlated with students' expected grades, and this paper tests whether or not there is an incentive for the instructor to "buy" higher evaluation scores by inflating grades. To test this hypothesis, I use unique data from the University of Washington's Office of Educational Assessment that includes a measure of each student's relative expected grade in the course. I find that there is an incentive for instructors to grade leniently after accounting for the potential endogeneity of the relative expected grade variable due to unobserved teacher productivity and unobserved heterogeneity of instructors and departments. Instructor fixed effects account for a significant part of the measured effect of relative expected grade on evaluations, and by not including them, the estimated impact of relative expected grade on evaluations is biased upwards.

1. Introduction

At many universities, there is a strong positive correlation between students' expected grades in a college course and their evaluations of the instructor for that course (Johnson, 2003). One suggested reason for this correlation is that if a student expects to fail the course, she most likely will not give an "excellent" evaluation of the teacher, regardless of whether or not that teacher really is excellent; if a student expects to get an "A" in the class, she may be nicer on the evaluation form. This relationship presents a problem since one of the major factors in deciding on promotions and merit pay raises in higher education is the quality of student teaching evaluations (Hostetler, Sawyer, & Prichard, 2004). This implies that instructors may have an incentive to inflate grades to get higher evaluations from their students.

This paper analyzes the link between expected grades and students' evaluations of teachers (SETs) to determine the effects of an exogenous variation of grading leniency on SETs. I use both a two-stage least squares/instrumental variables (TSLS/IV) approach and a fixed effects (FE) approach as two methods of controlling for unobserved instructor characteristics. The end result of this study is a range of estimates for the impact of grading leniency on SETs that control for both department and instructor characteristics. These measurements are beneficial to the college or university as they consider the issue of grade inflation and decide whether or not to adjust instructors' final SET scores.

There are two intertwined, policy relevant issues at play. One is the measurement of the quality of the student (grades) and the other is the measurement of the quality of the instructors (SETs). Grades act as a market signal and grade inflation dilutes this signal by compressing grades at the upper end of the distribution. In addition to the potential influence of the SET system on grades, there have been other hypotheses put forth for the trend in grade inflation over time, including a decades-long arms race for colleges to appear selective and the hypothesis that student quality has simply improved over this time span (Rojstaczer & Healy, 2010). The primary focus of this paper, however,

* Tel.: +1 727 864 8250; fax: +1 727 864 7967.
E-mail address: ewingam@eckerd.edu

0272-7757/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2011.10.002

is on measuring the quality of instructors. Determining whether or not grading leniency can potentially bias SET scores will enable university administrators to account for this bias in tenure, promotion, and other personnel decisions.

There are several variables that are used as grades in previous studies. Some authors use actual, final grades given in the course in their models, whereas some use expected grades as given by students on the SET form. While each measure has its own merits, the preferred measure in the literature is expected grade. Students do not know their final grade at the time of the evaluations, but we assume they have some expectation of the grade they will receive. Therefore, expected, and not actual, grades are determinants of SETs. Additionally, questions have been raised about how accurately final grades measure expected grades, as students tend to be irrationally overconfident in their grade expectations (Nowell & Alston, 2007). This difference between expected and actual grades may be exploited by the instructor.

The expected grade variable can be further refined by using a relative measure of expected grades instead of an absolute level. That is, the preferred measure in terms of representing a determinant of SETs, is the difference between a student's expected grade and what that student is accustomed to receiving (Isely & Singh, 2005). For example, an "A" student may be upset if she receives a "B" grade, while a "C" student would be pleased with that same absolute grade. Unfortunately, researchers typically calculate this measure of "relative expected grade" after the course is over by using the actual cumulative GPAs of the students at the section level. This raises selection concerns since evaluations are typically anonymous and some students simply do not fill out the forms. If there is a systematic difference between those students that do fill out evaluations and those that do not, this version of relative expected grade that is calculated after the course is finished can be biased due to selection.

To the best of my knowledge, my dataset is unique in that it includes a question on the evaluation form that directly asks students how their expected grade relates to what they *individually* are accustomed to receiving. By asking this question directly to the student, my dataset bypasses the above selection bias problem. This gives a direct indication of how an individual student's relative expected grade affects her evaluation of her instructor.

I find that after using TSLS/IV to account for the potential endogeneity of the relative expected grade variable due to unobserved teacher productivity, the estimated impact of relative expected grade on evaluations is large and significant. This suggests that there remains an incentive for instructors to grade leniently even after instrumenting for the effect of teacher productivity. When estimating the model with instructor fixed effects, I find that the incentive to grade leniently remains significant but of a smaller magnitude than the TSLS/IV estimates. If these fixed effects are not included, the estimated impact of relative expected grade on SETs is biased upwards.

2. SETs and the "principal–agent–client" problem

The problems associated with SETs in higher education fall under the umbrella of the broader "principal–agent–client" problem (Klitgaard, 1991). In general, introducing hierarchies into an agency problem increases information and monitoring costs (Garicano, 2008). In this example, the principal (department/college) introduces an evaluation instrument (SET) that is to be used by the clients (students) in order to gather information about the agent's (instructor's) performance. A problem arises given that the instructor has two methods of achieving better SET scores. The department would prefer that the instructors improve SET scores by actually improving their teaching, but instructors may find it less costly to increase SET scores by grading leniently. The issue of a principal finding an appropriate evaluation instrument for the agent that does not also introduce perverse incentives lies at the heart of Personnel Economics.[1]

2.1. Students – the clients

Presumably, students take courses for the learning that is produced and/or to satisfy degree requirements. Part of the learning produced is due to student effort in the class and the other part is due to the instructor's contribution. Whatever the productive combination may be, overall learning in the course is ubiquitously measured by the students' grades in the course. Therefore, students will value the absolute value of the grade they get in the class since it is a signal to others of what they have learned in that class and certain grade levels will be necessary for their degree requirements. It could be the case that students who achieve high grades (potentially through lenient grading) believe that they learned more, and reward their instructors by giving higher SET scores. Students also value the grade they receive in a class relative to what they as an individual student are accustomed to receiving (Isely & Singh, 2005).

As far as the answers given on the SET form are concerned, students are able to evaluate the instructor and the course given all of the information above except for their final grades. Since the SET form is typically handed out with a portion of students' grades yet to be determined, students answer the questions armed with only their *expected* grade, and thus, expected grade is a determinant of SETs (McPherson, 2006).

2.2. Instructors – the agents

Different instructor characteristics will be important determinants of their behavior as far as their attitudes to improving their evaluation scores. In particular, the rank of the instructor is important since department deliberations over promotion, tenure, and merit pay raises depend on ranks. For example, teaching assistants (TAs) and lecturers may have more incentive to inflate grades since evaluations

[1] See Chapter 8 of Lazear (1995) for an overview of issues in evaluation.