Grades and student evaluations of teachers

Anthony C. Krautmann *, William Sander

Department of Economics, DePaul University, 1 E. Jackson Blvd., Chicago, IL 60604, USA

Received 20 May 1996; accepted 14 October 1997

Abstract

Understanding the relationship between grading practices and student evaluations is especially important in higher education because of the increasing importance of this instrument in the promotion process. If evaluations can be increased by giving higher grades, then they are a flawed instrument for the evaluation of teaching. Further, this process may be contributing to the inflation of grades in higher education if faculty have an incentive to increase their evaluations. Also, grade inflation dilutes the signaling role of educational credentials in screening workers for the labor market. In this paper, we revisit the determinants of student evaluations in a model that allows for the possibility that (expected) grades are simultaneously determined. We estimate evaluations using both OLS and two-stage least squares (TSLS) and find that grades do affect an instructor’s evaluation. These results are consistent with the hypothesis that instructors can “buy” better evaluations through more lenient grading.

1. Introduction

Can instructors increase their student evaluations by giving out easier grades? While a number of studies suggest that there is a positive relationship between the grades students expect to receive and student evaluations (Aigner and Thum, 1986; Ditts, 1983; Kau and Rubin, 1976; Mehdizadeh, 1990; Nelson and Lynch, 1984; Nichols and Soper, 1972; Zangenehzadeh, 1988), a few studies find no such relationship (Seiver, 1983; Decanio, 1986). One possible reason for these conflicting results is that grades might be an endogenous determinant of student evaluations (Nelson and Lynch, 1984; Seiver, 1983; Zangenehzadeh, 1988). In this case, ordinary least squares (OLS) estimates of the effect of grades on evaluations would be biased.

Understanding the relationship between grading practices and student evaluations is especially important in higher education because of the increasing importance of this instrument in the promotion process. If evaluations can be increased by giving higher grades, then they are a flawed instrument for the evaluation of teaching. Further, this process may be contributing to the inflation of grades in higher education if faculty have an incentive to increase their evaluations (see Schultz, 1981). Also, grade inflation dilutes the signaling role of educational credentials in screening workers for the labor market (Blaug, 1993).

In this paper, we revisit the determinants of student evaluations in a model that allows for the possibility that (expected) grades are simultaneously determined. We estimate evaluations using both OLS and two-stage least squares (TSLS) and find that grades do affect an instructor’s evaluation. These results are consistent with the hypothesis that instructors can “buy” better evaluations through more lenient grading.

2. THE MODEL AND DATA

There are many factors associated with the assessment of an instructor on the student evaluations (EVAL). The
variable of key interest in this paper is grades — particularly the grade that a student expects to receive at the time he or she is filling out the evaluation. The endogeneity issue arises from the possibility that grades are correlated with an unobservable, such as productivity in teaching. One might argue that higher productivity in teaching results in students earning higher grades. Better teaching might also result in students giving higher evaluations. If productivity (an omitted explanatory variable) causes higher evaluations (the dependent variable) and higher grades (an included explanatory variable), then OLS estimates will be biased and inconsistent. While there exists little previous research controlling for the productivity of teachers, at least one study found that more skilled professors do not necessarily get higher student evaluations (Gramlich and Greenlee, 1993).

Characteristics of the instructor would also be expected to affect evaluations. For example, Aigner and Thum (1986) found that full professors tend to receive lower marks than others. Characteristics of the course may also have an effect on EVAL. For example, Kau and Rubin (1976) found that courses in the major field are significantly related to student ratings, while Aigner and Thum (1986) found that the difficulty of a course, the amount of homework in the course, and the percentage of seniors in a course all have an effect on evaluations. As previous research suggests, many other factors might also affect student evaluations. For this study, data on these other factors are unavailable.

In this study, we estimate student evaluations with the following model:

$$\text{EVAL} = \beta_0 + \beta_1 \text{EXPGRADE} + \beta_2 \text{VISITOR}$$

$$+ \beta_3 \text{EMERITUS} + \beta_4 \text{MALE} + \beta_5 \text{TENURED}$$

$$+ \beta_6 \text{SIZE} + \beta_7 \text{CORE} + \beta_8 \text{GRADUATE} + \epsilon$$

We use data from student evaluations of economics courses taught at DePaul University in Chicago. The grade variable, EXPGRADE, was obtained from the student’s response on the evaluation as to the grade he or she expected to receive in the course. The variables related to the instructor (MALE, TENURED, VISITOR, and EMERITUS) are all dummy variables equal to one if the faculty member is a male, a tenured, a visiting or an emeritus professor. The variables related to the course include: SIZE, equal to the number of students in the class; CORE, a dummy variable equal to 1 if it is a core (rather than elective) course; and GRADUATE, a dummy variable equal to 1 if it is a graduate (rather than undergraduate) course.¹

Our data set spans the 1994 through 1996 academic years, a time period chosen because the University began evaluating all courses (regardless of an instructor’s tenure status) in 1994. The raw data on student evaluations of teaching goes from 1 to 5 (lowest to highest). The unit of observation in this study is the individual course; hence, class-wide averages are used to measure both EVAL and EXPGRADE.² Summary statistics for the data set are presented below in Table 1.

As noted above, an important issue in estimating the effect of expected grades on student evaluations is the endogeneity problem. An approach to test for endogeneity is the procedure suggested by Hausman (1978). If one rejects endogeneity, OLS is the appropriate estimation technique. If, on the other hand, one finds evidence that some of the independent variables are jointly determined, then OLS estimates will be biased and inconsistent, and an alternative method of estimation is called for (such as TSLS).

The key modeling issue in undertaking a two-stage estimation procedure is the identification problem. To identify the model, one needs a set of variables that are strongly related to the suspected endogenous variable (i.e., expected grades) while not related to the error term in the base regression (i.e., the estimate of evaluations). To prove that one’s identifiers are correlated with the endogenous variable is straightforward. A necessary (but not sufficient) condition for establishing the independence of the identifiers from the error term is that the identifying variables are not correlated with the dependent variable in the base regression. That is, if the identi-

---

¹ Core courses are defined as the two undergraduate Principles courses and the undergraduate Money and Banking course. Graduate courses include the courses in the Graduate School of Business (Money and Banking, Managerial Economics, and Business Conditions Analysis), as well as courses in the Master of Arts program in economics.

² Because the unit of observation is the course, student-specific determinants (e.g., grade-point average, gender, years in college, etc.) cannot be included in the analysis.
The power of the Hausman Test is likewise limited by the quality of the identifying variables. To run a Hausman Test, one follows a very similar procedure as that corresponding to TSLS. Under both procedures, one must first properly identify and estimate a first-stage regression of the suspected endogenous variable (i.e., \( \text{EXPGRADE} \)):

\[
\text{EXPGRADE} = \delta_0 + \delta_1 X + \delta_2 Z + U
\]

where the vector \( Z \) constitutes the set of identifying variables, and \( X \) is the set of variables common to both (2) and the base regression. To conduct the Hausman Test, one then runs the following regression:

\[
\text{EVAL} = \beta_0 + \beta_1 \text{EXPGRADE} + \beta_2 X + \beta_3 \hat{U} + \epsilon
\]

where \( \hat{U} \) are the residuals from (2). One would then reject the null hypothesis that \( \text{EXPGRADE} \) is exogenous if the coefficient on \( \hat{U} \) is statistically significant.

To run TSLS, one runs the same regression of \( \text{EXPGRADE} \) given above in (2), then uses the fitted values of this regression in the following model:

\[
\text{EVAL} = \beta_0 + \beta_1 \text{EXPGRADE} \text{ fitted values} + \beta_2 X + \epsilon
\]

where \( \text{EXPGRADE} \text{ fitted values} \) is the fitted values obtained from (2). This Instrumental Variables (IV) approach to solving the endogeneity problem is valid as long as \( \text{EXPGRADE} \) is not correlated with \( \epsilon \).

The fact that both procedures involve the same first-stage regression exemplifies the critical importance of selecting good identifiers. If the identifying variables are weak, the statistical properties of both the Hausman Test and TSLS are uncertain (Bound et al., 1995). Because it is impossible to empirically establish the independence of the identifiers and the error term, test results on the endogeneity of \( \text{EXPGRADE} \) cannot be fully relied upon; and, as such, it is reasonable to check the robustness of one’s analysis by estimating the model under both the null and alternative hypotheses. Qualitatively similar results under both regimes would suggest that endogeneity is not a problem, leaving us with a greater degree of confidence in the inferences regarding \( \text{EXPGRADE} \).

For this reason, we present both OLS estimates (which are unbiased and efficient if \( \text{EXPGRADE} \) is exogenous) and TSLS estimates (which are consistent and asymptotically efficient if \( \text{EXPGRADE} \) is endogenous).

The quality of one’s identifying variables is an important consideration that has been ignored in related research. For example, Nelson and Lynch (1984) use an instructor’s rank and average present grade to identify the effect of grades on student evaluations. It is not very likely that such identifiers are valid because instructor’s rank and average present grade surely affect evaluations. The identifying strategies in Seiver (1983) and Zangenehzadeh (1988) are also problematic.

In this study, we use \( \text{GRADUATE} \) and \( \text{CORE} \) as identifiers. Both variables are highly correlated with \( \text{EXPGRADE} \) (i.e., they explain a relatively large percentage in the variation in expected grades). Estimates of \( \text{EXPGRADE} \) are presented using \( \text{GRADUATE}, \text{CORE}, \text{SIZE} \), and the four variables relating to the instructor; the \( R^2 \) for this regression is 0.29 (see Table 2). When \( \text{CORE} \) AND \( \text{GRADUATE} \) are excluded from the regression, the \( R^2 \) decreases to 0.03. Further, our identifying variables are not highly correlated with \( \text{EVAL} \) (see the OLS estimates in Table 3). Although it is impossible to empirically prove that \( \text{CORE} \) and \( \text{GRADUATE} \) constitutes a valid set of identifiers, they appear to be reasonable candidates.

### 3. THE EMPIRICAL RESULTS

Given the insignificance of the coefficient on the residual in (3) above, we conclude that \( \text{EXPGRADE} \) is not endogenous, implying that OLS estimates of (1) are appropriate. But given the earlier discussion pertaining to the inherent difficulty of establishing the power of the Hausman Test, we also estimate (1) using TSLS (see Table 3).

Both the OLS and TSLS estimates imply that student evaluations are positively related to expected grades; that is, the better grade a student expects to receive in a course, the higher he or she rates the instructor. Our results imply that a one-grade increase in classroom GPA results in an increase in the evaluation of between 0.34 and 0.56 — a moderate effect given that the standard deviation in \( \text{EVAL} \) is about 0.6. We also found that \( \text{TENURED}, \text{EMERITUS}, \text{and VISITOR} \) have significant negative effects on \( \text{EVAL} \). Finally, our results indicate that neither gender differences nor class size have a significant effect on evaluations.

### 4. Conclusions

Our findings indicate that grades affect student evaluations, suggesting that faculty have the ability to “buy” higher evaluations by lowering their grading standards. The positive effect of expected grades on evaluations likely implies that a similar relationship exists regarding actual grades. If so, then the long-term trend to inflated grades in higher education may be due in part to the increasing importance of student evaluations.
Table 2
OLS estimates of EXPGRADE (t-statistics in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>With identifiers</th>
<th>Without identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core course</td>
<td>-0.122*** (2.5)</td>
<td>—</td>
</tr>
<tr>
<td>Graduate course</td>
<td>0.249*** (5.3)</td>
<td>—</td>
</tr>
<tr>
<td>Emeritus professor</td>
<td>-0.036 (0.6)</td>
<td>0.021 (0.28)</td>
</tr>
<tr>
<td>Tenured professor</td>
<td>-0.069 (1.5)</td>
<td>-0.028 (0.5)</td>
</tr>
<tr>
<td>Visiting professor</td>
<td>-0.0001 (0.01)</td>
<td>0.129*** (2.04)</td>
</tr>
<tr>
<td>Male professor</td>
<td>-0.002 (0.05)</td>
<td>0.024 (0.43)</td>
</tr>
<tr>
<td>Class size</td>
<td>0.0007 (0.4)</td>
<td>-0.0006 (0.29)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.130*** (45)</td>
<td>3.13*** (39)</td>
</tr>
<tr>
<td>R²</td>
<td>0.29</td>
<td>0.03</td>
</tr>
<tr>
<td>Number of observations</td>
<td>258</td>
<td>258</td>
</tr>
</tbody>
</table>

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Table 3
Estimates of student evaluations of teachers (t-statistics in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>TSLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core course</td>
<td>-0.033 (-0.3)</td>
<td>—</td>
</tr>
<tr>
<td>Graduate course</td>
<td>-0.105 (-1.1)</td>
<td>—</td>
</tr>
<tr>
<td>Expected grade</td>
<td>0.564*** (4.6)</td>
<td>0.344* (1.7)</td>
</tr>
<tr>
<td>Emeritus professor</td>
<td>-0.669*** (5.4)</td>
<td>-0.673*** (5.5)</td>
</tr>
<tr>
<td>Tenured professor</td>
<td>-0.289*** (3.1)</td>
<td>-0.304*** (3.4)</td>
</tr>
<tr>
<td>Visiting professor</td>
<td>-0.433*** (4.0)</td>
<td>-0.434*** (4.0)</td>
</tr>
<tr>
<td>Male professor</td>
<td>0.024 (0.3)</td>
<td>0.026 (0.3)</td>
</tr>
<tr>
<td>Class size</td>
<td>0.004 (1.1)</td>
<td>0.003 (0.9)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.367*** (5.8)</td>
<td>3.04*** (4.6)</td>
</tr>
<tr>
<td>R²</td>
<td>0.193</td>
<td>—</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.4***</td>
<td>9.3***</td>
</tr>
<tr>
<td>Number of observations</td>
<td>258</td>
<td>258</td>
</tr>
</tbody>
</table>

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level

Acknowledgements
The authors would like to thank the College of Commerce, DePaul University, for research support, and two anonymous reviewers for their comments.

References
Aigner, D.J., Thum, F.D., 1986. On student evaluations of teaching ability. Journal of Economic Education 17 (Fall), 243–266.

