Economics Bulletin

Volume 36, Issue 1

Tuition Fees and Instructional Quality

Bastian Gawellek
University of Leipzig

Bernd Süssmuth
University of Leipzig; CESifo

Daniel Singh
University of Munich

Abstract

This study examines the impact of introducing modest tuition fees on perceived instructional quality in a publicly financed system of higher education. Relying on a difference-in-differences strategy and controlling for additional teaching financed from collected fees, we find that the introduction of fees has a significant positive impact on faculty evaluations that is robust and amounts to about one third of an evaluation grade.

We thank Donna K. Ginther, an anonymous referee, Alexander Fink, Andreas Haufler, Susanne Hoffmann, Christian Holzner, and Marco Sunder for many helpful comments and suggestions.

Citation: Bastian Gawellek and Bernd Süssmuth and Daniel Singh, (2016) "Tuition Fees and Instructional Quality", *Economics Bulletin*, Volume 36, Issue 1, pages 84-91

Contact: Bastian Gawellek - gawellek@wifa.uni-leipzig.de, Bernd Süssmuth - suessmuth@wifa.uni-leipzig.de, Daniel Singh - danielsingh@outlook.com.

Submitted: July 17, 2015. **Published:** February 04, 2016.

1. Introduction

Since the last decade, an extensive accountability movement is on the rise in higher education across developed countries (Goldin and Katz 2001). It affects systems regardless of being more or less privately funded. As part of the movement instructional evaluations substantially gain in importance. This study examines the impact of the introduction of modest tuition fees on perceived instructional quality in a system that has been exclusively financed by public funds.

Nowadays instructional evaluations are—for good or ill (Sproule 2002)—widely used at the faculty level to make comparisons among courses and teachers. They determine the granting of teaching awards as well as faculty salary, promotion, and tenure decisions on both sides of the Atlantic Ocean (e.g. Haskell 1997; Langbein 2008; Mandel and Süssmuth 2011). Equally important, in a fee financed system instructional evaluations might provide a mechanism of control in an environment otherwise lacking direct control over faculty. According to Haskell (1997), they are often seen as a powerful tool assuring classroom changes that lead to the retention of student tuition dollars by assenting to student consumer demands and of parents who foot the tuition bill. Additionally, a growing literature in economics of education employs student evaluations of teaching as educational outcome (e.g. Hamermesh and Parker 2005; Süssmuth 2006; Bedard and Kuhn 2008; Mandel and Süssmuth 2011), which bears a potential bias given a systematic relationship between fees and ratings.

To the best of our knowledge, no study exists that analyzes whether and if so how tuition fees affect instructional evaluation ratings. To study the relationship, we make use of a recent policy experiment. In the public system of higher education in post-war Germany, college tuition fees existed and were to some degree subject to institutional discretion up to the late 1960s. They have been legally banned at the end of the 1960s, for example, at the University of Munich as of Fall 1968. Since then post-secondary education was basically free of charge for the following decades. After a Federal Constitutional Court ruling in 2005 several German states and universities, among them the University of Munich, re-introduced tuition fees amounting to approximately 1,000 euros per year as of Summer 2007. In contrast, Berlin and its universities did not re-introduce tuition fees. We make use of this dichotomous treatment of students at two renowned economics faculties in Munich and Berlin to quantify the effect introducing tuition fees has on perceived instructional quality.

2. Data and estimation strategy

2.1. Student evaluation data

Data for this study comprises economics classes offered at the University of Munich (Ludwig Maximilian University, henceforth: LMU) and at its peer institution the Humboldt University Berlin (henceforth: HU) from Winter 2004/05 to Winter 2008/09. During this period of 9 semesters, 1,701 economics classes in 386 different courses were offered by 488 instructors at the two institutions. Our data include information about class size, the semester that each course was offered, the level of the class (lower/upper division), the instructor, and the average evaluation score (Table 1). Evaluation data are made available, corresponding to the natural unit of observation, in the form of student evaluation scores aggregated to class means:

$$E_{j,i,c,t} = \frac{1}{R_{j,i,c,t}} \sum_{k=1}^{R_{j,i,c,t}} e_{j,i,c,t,k}, \tag{1}$$

where e denotes individual student evaluation scores, E is the average class evaluation score, R is the number of evaluation responses, j denotes the institution (LMU/HU), t denotes semesters ($t = 2004/2005, \ldots, 2008/2009$), c denotes $c = 1, \ldots, C$ different courses, and $i = 1, \ldots, I$ denotes instructors. See Süssmuth (2006) and Mandel and Süssmuth (2011) for further detail and discussion.

Similar to the U.S. practice (Hamermesh and Parker 2005; Bedard and Kuhn 2008), the rating forms both at LMU and HU include: 'Overall, my personal impression is that the course was excellent (1); very good (2); satisfactory (3); unsatisfactory (4); very unsatisfactory (5).' For interpretive ease, we reverse and transform the average class evaluation scores E to lie in the interval $[0, 100]$, using $\widetilde{E}_{j,i,c,t} = \frac{1}{4}\left|E_{j,i,c,t} - 5\right| \times 100$, such that $\widetilde{E}_{j,i,c,t} \in [0, 100]$, where $\widetilde{E}_{j,i,c,t} = 0$ denotes the poorest and $\widetilde{E}_{j,i,c,t} = 100$ the best performance. Besides E, the published evaluation summary contains information on the number of participants in the class, the title of the class, and the instructor. This information allows us to estimate instructor fixed effects models that control for time-invariant instructor heterogeneity and instructor and course fixed effects models that control for both instructor and course-specific heterogeneity.

2.2. Empirical identification

To estimate the effect introducing tuition fees has on average student evaluations, we consider the following three models

$$\tilde{E}_{j,i,c,t} = a_{j,i,c} + \lambda_t + \beta \times \mathrm{LMU}_j \times \lambda_t + \gamma \times \mathrm{LMU}_j \times F_{j,t} + X'_{i,j,c}\delta + u_{j,i,c,t} \tag{2}$$

$$\text{with } a_{j,i,c} = a \qquad \text{(P-OLS)} \tag{3}$$

$$a_{j,i,c} = \alpha_{j,i} \qquad \text{(I-FE)} \tag{4}$$

$$a_{j,i,c} = \alpha_{j,i} + \alpha_{j,c} \qquad \text{(I-C-FE)}, \tag{5}$$

where $\alpha_{j,i}$ and $\alpha_{j,c}$ is a vector of instructor and course fixed effects, respectively; $F_{j,t}$ represents a tuition fee treatment dummy taking on a value of 1 beginning with Summer 2007. λ_t are semester fixed effects. $F_{j,t}$ and λ_t each are also included interacted with our treatment group (LMU) identifier LMU in (2). The corresponding coefficients are γ and the ones contained in vector β, respectively. Matrix X contains the following course characteristics: level of class (lower/upper division), a dummy identifying whether the class is a lecture or lecture accompanying tutorial, a class size polynomial $\sum_{m=1}^{M} \eta_m \left(S_{j,i,c,t}\right)^m$, where S denotes respective class size, with $M = 3$ (Mandel and Süssmuth 2011), and the number of tutorials offered in combination with a particular lecture ('N tutorials'). The latter two variables control for the fact that since Summer 2007 fees collected at LMU essentially were used to finance additional teaching in the form of additional lecture accompanying tutorials. Finally, u is the usual error term. P-OLS, I-FE, I-C-FE denote pooled OLS, instructor fixed effects, and instructor and course fixed effects, respectively.

Specification (2) in combination with (3) to (5) allows for a maximum of flexibility with regard to differences between the two institutions over time. These differences are accounted for in addition to the isolated tuition fee treatment effect.

As our data includes courses taught parallelly in a particular semester by the same instructor, we rely on least squares dummy variables (LSDV) regressions to estimate (2).[1] According to Bertrand *et al.* (2004) clustering standard errors should be done at the highest level of cross-sectional aggregation in order to account for autocorrelation within clusters, which is not captured by the considered temporal and cross-sectional fixed effects. In our case, the most aggregate level is the one of the university. However, Cameron and Miller (2015, p. 333) note in this context that the number of considered clusters should not be too low. Hence, we estimate our LSDV models both with standard errors clustered at the level of courses (Table 2) and at the university level (Table 3).[2]

To check the robustness of our estimates, we follow a similar strategy as e.g. Slusky (2015) by considering a counterfactual or placebo fee treatment at LMU from Summer 2005 to Winter 2006/07 leaving the remaining set-up unchanged (Table 4). Due to data limitations, we cannot go back beyond Summer 2005. Nevertheless, our strategy meets two crucial requirements of a robustness analysis. First, the considered placebo treatment period completely falls into the period of zero tuition fees at LMU. Secondly, the placebo treatment spans four semesters corresponding in length to our actually analyzed treatment sample.

Table 1: Summary statistics

	LMU Munich	HU Berlin
N classes	913	788
N instructors	262	226
N courses	165	221
Avg. evaluation score	79.72	67.40
Avg. class size	47.33	51.39
Avg. N tutorials	3.40	1.03
Upper division (share)	0.74	0.71
Lecture (share)	0.37	0.63

[1] Following Cameron and Miller (2015, p. 331), we multiply our estimated standard errors with a factor $\sqrt{(N-(K-1))\,/\,(N-G-(K-1))}$, where G denotes number of considered clusters, N the number of observations, and K the number of covariates, respectively.

[2] The total number of courses (and, thus, course clusters) at both institutions is smaller, i.e. more aggregate, than the one of instructors; see Table 1. Additionally, arguments for clustering at the level of instructors apply also for university clusters and so are "nested".

Table 2: DiD models (3)–(5), std. errors clustered at course level

	(3) P-OLS	(4) I-FE	(5) I-C-FE
LMU	10.16***	8.04***	−5.71
Size	−0.09***	−0.15***	−0.12***
$(\text{Size})^2/1{,}000$	0.38**	0.56***	0.45***
$(\text{Size})^3/1{,}000{,}000$	−0.38*	−0.57***	−0.47**
Upper division	−1.36	−1.66	−1.31
Lecture	−1.73*	−2.28	−3.11
N tutorials	−0.44	−0.31	0.19
Summer 2005	−0.04	−1.31	−4.70
Winter 2005/06	−0.06	−1.03	−0.96
Summer 2006	−0.53	−1.05	−4.27
Winter 2006/07	−0.58	0.11	0.75
Summer 2007	0.28	−1.84	−5.15*
Winter 2007/08	−1.18	−2.52	−1.72
Summer 2008	0.52	−1.22	−3.33
Winter 2008/09	2.17	2.46	3.61
LMU × Summer 2005	−0.46	0.74	3.47
LMU × Winter 2005/06	2.17	2.14	2.14
LMU × Summer 2006	0.79	1.71	3.89
LMU × Winter 2006/07	3.46	2.39	0.48
LMU × Winter 2007/08	3.45	3.02	−0.40
LMU × Summer 2008	1.21	2.03	1.00
LMU × Winter 2008/09	1.68	0.40	−2.59
LMU × Semester with tuition fees	1.61	4.91*	6.49*
Observations	1701	1701	1701
Adjusted R^2	0.27	0.64	0.74

* $p < .10$, ** $p < .05$, *** $p < .01$. Dependent variable: Student evaluation score.

Sample: Winter 2004/05 – Winter 2008/09.

Table 3: DiD models (3)–(5), std. errors clustered at university level

	(3) P-OLS	(4) I-FE	(5) I-C-FE
LMU	10.16^{***}	8.04^{*}	-5.71^{**}
Size	-0.09^{*}	-0.15	-0.12^{**}
$(\text{Size})^2/1{,}000$	0.38^{*}	0.56^{**}	0.45^{*}
$(\text{Size})^3/1{,}000{,}000$	-0.38	-0.57^{***}	-0.47
Upper division	-1.36	-1.66^{*}	-1.31
Lecture	-1.73	-2.28	-3.11
N tutorials	-0.44	-0.31	0.19
Summer 2005	-0.04	-1.31	-4.70
Winter 2005/06	-0.06	-1.03	-0.96
Summer 2006	-0.53	-1.05	-4.27
Winter 2006/07	-0.58	0.11	0.75
Summer 2007	0.28	-1.84	-5.15
Winter 2007/08	-1.18	-2.52	-1.72
Summer 2008	0.52	-1.22	-3.33
Winter 2008/09	2.17	2.46	3.61
LMU × Summer 2005	-0.46	0.74	3.47
LMU × Winter 2005/06	2.17	2.14	2.14^{*}
LMU × Summer 2006	0.79	1.71	3.89
LMU × Winter 2006/07	3.46	2.39	0.48
LMU × Winter 2007/08	3.45^{**}	3.02^{**}	-0.40
LMU × Summer 2008	1.21^{**}	2.03	1.00
LMU × Winter 2008/09	1.68^{*}	0.40	-2.59^{**}
LMU × Semester with tuition fees	1.61	4.91^{**}	6.49^{*}
Observations	1701	1701	1701
Adjusted R^2	0.27	0.64	0.74

* $p < .10$, ** $p < .05$, *** $p < .01$. Dependent variable: Student evaluation score.

Sample: Winter 2004/05 – Winter 2008/09.

Table 4: DiD models (3)–(5): Four-semester placebo fees at LMU, starting in Summer 2005

	Std. erros clustered at course level			Std. erros clustered at university level		
	(3) P-OLS	(4) I-FE	(5) I-C-FE	(3) P-OLS	(4) I-FE	(5) I-C-FE
LMU	10.20***	8.89**	−0.03	10.20***	8.89*	−0.03
Size	−0.08**	−0.14***	−0.10	−0.08**	−0.14**	−0.10
$(\text{Size})^2/1{,}000$	0.31*	0.47***	0.35	0.31	0.47	0.35
$(\text{Size})^3/1{,}000{,}000$	−0.30	−0.47**	−0.38	−0.30	−0.47	−0.38
Upper division	−0.96	−1.39	−6.56	−0.96	−1.39	−6.56
Lecture	−1.76	−1.06	−2.41	−1.76	−1.06	−2.41
N tutorials	−0.44	−0.15	0.34	−0.44	−0.15	0.34
Summer 2005	−0.01	−0.93	−3.89	−0.01	−0.93	−3.89
Winter 2005/06	−0.03	−1.06	0.17	−0.03	−1.06	0.17
Summer 2006	−0.46	−0.49	−3.08	−0.46	−0.49	−3.08
Winter 2006/07	−0.55	0.17	0.84	−0.55	0.17	0.84
LMU × Winter 2005/06	2.62	1.88	−2.72	2.62**	1.88*	−2.72
LMU × Summer 2006	1.28	1.05	−1.74	1.28**	1.05	−1.74**
LMU × Winter 2006/07	3.96	2.50	−1.98	3.96***	2.50**	−1.98
LMU × Semester with placebo fees	−0.53	−0.20	3.41	−0.53	−0.20	3.41
Observations	946	946	946	946	946	946
Adjusted R^2	0.22	0.65	0.75	0.22	0.65	0.75

* $p < .10$, ** $p < .05$, *** $p < .01$. Dependent variable: Student evaluation score. Sample: Winter 2004/05 – Winter 2006/07.

3. Findings and interpretation

3.1. Findings

Our estimates of difference-in-differences (DiD) models (2) in combination with (3) to (5) are reported in Tables 2, 3. Note, the respective first row labeled 'LMU' gives our estimates for the level difference between LMU and HU for the reference semester Winter 2004/05. The last row before the penultimate line in both tables is the most decisive one as regards the fee treatment effect. It shows the respective γ coefficient estimates. As soon as I-FE and I-C-FE are considered, it is estimated statistically significant. It draws the highest significance for standard errors clustered at the university level and the I-FE specification (Table 3). For the P-OLS specification that does not control for time-invariant unobserved heterogeneity γ is estimated throughout as not significantly different from zero. Keeping in mind that the five distinct grades are transformed to the closed interval $[0, 100]$, the move from one full grade to another amounts to 20. Thus, e.g. in the two-way fixed effects (I-C-FE) specification, an estimated coefficient $\gamma = 6.49$ implies an improvement of about one third of a grade. In terms of size the γ estimates generally outweigh the other considered covariates' coefficients.

As regards our robustness analysis, the result is clear-cut. In none of the six considered placebo treatment cases, policy parameter γ comes out statistically significant (Table 4).

3.2. Interpretation

Improvement of evaluation ratings following the introduction of fees can ad hoc be explained in two ways. First, students might impute quality from the mere fact of paying. However, a recent and representative survey among high school graduates on behalf of the German Federal Ministry of Education by Heine *et al.* (2008) finds that only about one percent of respondents sees tuition fees as a signal of quality. Secondly, motivation and ambition of students might have increased and/or led to self-selection. However, in latest studies on the German policy experiment, the effect of introducing tuition fees both on overall enrollment and on inter-state mobility is found to be statistically insignificant and/or clearly negligible (Dwenger *et al.* 2012; Bruckmeier and Wigger 2014). Hence, quality imputation and selectivity seem not to be at the heart of the observed significant upward shift in instructional ratings. An alternative explanation could be an actual increase in teaching quality induced by a higher motivation on the part of instructors through the awareness of students paying for tuition.

4. Conclusion

Relying on a DiD strategy and controlling for additional teaching financed from collected fees, we find that the introduction of fees in a publicly financed system has a significant positive impact on faculty evaluations. The effect is robust and implies an improvement of about one third of a grade.

References

Bedard, K. and P. Kuhn (2008) "Where class size really matters: Class size and student ratings of instructor effectiveness" *Economics of Education Review* **27**, 253–265.

Bertrand, M., E. Duflo and S. Mullainathan (2004) "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* **119**, 249–275.

Bruckmeier, K. and B. Wigger (2014) "The effects of tuition fees on transition from high school to university in Germany" *Economics of Education Review* **41**, 14–23.

Cameron, A.C. and D.L. Miller (2015) "A practitioner's guide to cluster-robust inference" *Journal of Human Resources* **50**, 317–372.

Dwenger, N., J. Storck and K. Wrohlich (2012) "Do tuition fees affect the mobility of university applicants? Evidence from a natrual experiment" *Economics of Education Review* **31**, 155–167.

Goldin, C. and L. Katz (2001) "The legacy of U.S. educational leadership: Notes on distribution and economic growth in the 20th Century" *American Economic Review* **91**, 18–23.

Hamermesh, D.S. and A. Parker (2005) "Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity" *Economics of Education Review* **24**, 369–376.

Haskell, R.E. (1997) "Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st Century" *Education Policy Analysis Archives* **5**, 1–32.

Heine, C., H. Quast and H. Spangenberg (2008) "Studiengebühren aus der Sicht von Studienberechtigten" HIS Forum Hochschule number 15/2008.

Langbein, L. (2008) "Management by results: Student evaluation of faculty teaching and the mis-measurement of performance" *Economics of Education Review* **27**, 417–428.

Mandel, P. and B. Süssmuth (2011) "Size matters. The relevance and Hicksian surplus of preferred college class size" *Economics of Education Review* **30**, 1073–1084.

Slusky, D.J.G. (2015) "Significant placebo results in difference-in-differences analysis: The case of the ACA's parental mandate" *Eastern Economic Journal*, online first.

Sproule, R. (2002) "The underdetermination of instructor performance by data from the student evaluation of teaching" *Economics of Education Review* **21**, 287–294.

Süssmuth, B. (2006) "Beauty in the classroom: Are German students less blinded?" *Applied Economics* **38**, 231–238.