THE EFFECTIVENESS OF PEER TUTORING ON STUDENT ACHIEVEMENT AT THE UNIVERSITY LEVEL

Vincent G. Munley
Lehigh University
vincent.munley@lehigh.edu
610 758 3428
(Corresponding Author)

Eoghan Garvey*
National University of Ireland, Galway

Michael J. McConnell
Economic Research Service
U.S. Department of Agriculture
MMCCONNELL@ers.usda.gov
202 694 5158

Several studies over the past decade have investigated the importance of peer effects on student learning. Bruce Sacerdote (2001) and David J. Zimmerman (2003) both take advantage of randomness in the assignment of freshman roommates to avoid the estimating problems that usually arise because of endogeneity in the self selection of one's peer group. Ralph Stinebrickner and Todd R. Stinebrickner (2008) take this strategy one step further by using information about whether a randomly assigned roommate brings a video game to college as an instrument in identifying the impact of a freshman's studying effort on grade point average.

Another 'peer effect' of potential significance in the learning process, one that seems to be increasing in its utilization across campuses, has yet to receive explicit attention in the empirical literature.[1] This is the use of peer tutors, especially in subjects like economics where developing problem solving skills is a key factor in mastering the subject material. Peer tutoring programs typically employ upper level undergraduates who have successfully completed a course to lead problem solving sessions for currently enrolled students. Because participating in peer tutoring sessions is typically a voluntary activity, the econometric issues arising from endogeneity of the self selection process are likewise something that must be addressed in assessing the effect of this program.

I. Empirical Model

We examine the effect of participating in the peer tutoring program on a student's final course grade in two ways. First we consider a simple specification that posits that either participation in the peer tutoring or the level of participation, measured by the total number of tutorial hours over the course of a semester, will affect the final grade exogenously. We next recognize that participation in the peer tutoring program is voluntary and that the level of tutoring participation and the course grade received are interrelated outcomes and must be modeled jointly.

The initial model specification is

(1) $$\text{Grade}_i = x'_i\beta + \delta T_i + e_i \,,$$

where Grade_i is the numerical value on a four point scale of the letter grade (A through F including plus and minus) received in a course. T_i is participation by person i in the peer tutoring program, measured alternatively as a binary variable to indicate that a student participated in the program during the semester or as a continuous variable measuring the number of hours tutored. X_i is a set of control variables containing information about a student's academic profile and demographic characteristics that may be relevant in explaining the grade that a student receives as well characteristics of the course offering itself.

In order to model the tutoring participation decision jointly with the grade determination outcome, we adopt a treatment model as described in William H. Greene (pp 889-891).

If we assume that the choice to participate in the peer tutoring program is affected by identifying variables as well as some, or all, of the control variables in (1), then we can specify:

(2) $$T^*_i = w'_i\gamma + u_i \text{ and } T_i = 1 \text{ if } T_i^* > 0, T_i = 0 \text{ otherwise}.$$

where X_i is a subset of W_i. If the error terms from the two equations are normal and are correlated, then

(3) $$E[\text{Grade}_i \mid T_i = 1, x_i, w_i] = x'_i\beta + \delta + \rho\sigma_e\lambda(-w_i\gamma)$$

and

(4) $$E[\text{Grade}_i \mid T_i = 0, x_i, w_i] = x'_i\beta + \rho\sigma_e \frac{-\phi(w_i\gamma)}{1-\Phi(w_i\gamma)}$$

λ in (3) is the inverse Mills ratio, while the normal distribution and cumulative normal are referenced in the ratio at the end of (4).

If the level of participation, measured in total tutorial hours over the course of a semester, is specified as the treatment, rather than simply participation in the program, then a tobit treatment model is appropriate (Jeffrey M. Wooldridge, pp. 571-572) rather than the more common probit variant. In this model step 1 is to estimate

$$\text{Hours}^*_i = w'_i \gamma + u_i \tag{5}$$

In place of the unobserved latent dependent variable in (5), actual reported hours are used on the left hand side for tobit estimation with an observable lower limit of zero hours. In step 2, the residuals from step 1 are added to the right hand side of (3) for tutorial participants and, as before, the estimated ratio at the end of (4) is included for non-participants (except that the term $w_{ij}\gamma$ is divided by the standard error as estimated by the tobit).

In order to isolate the impact of the peer tutoring program, the set of control variables, X_i, should comprise as many conditioning measures of a student's academic and demographic characteristics as possible. To accomplish this we include all information available from the university's student record system. Also included as explanatory variables are several characteristics of the course offering. Together these are:

High School Rank Index – Index scaled from 0 to 120 that is based on both a student's graduation rank in High School and the size of the High School class.

SAT-Math – Student's SAT mathematics score.

GPA – Student's beginning of semester grade point average.

Semester Credits – The number of credit-hours for which the student is enrolled.

Freshman – binary variable equal to one if the student is a freshman.

Sophomore – binary variable equal to one if the student is a sophomore.

2004-05 – binary variable equal to one for the 2004-05 academic year.

2005-06 – binary variable equal to one for the 2005-06 academic year.

Recitation – binary variable equal to one if the course includes a one hour per week recitation class with a graduate student teaching assistant in addition to lectures.

Lab – binary variable equal to one if the course includes a weekly laboratory.

Male – binary variable equal to one if the student is a male.

African American – binary variable equal to one if the student is African American.

Hispanic – binary variable equal to one if the student is Hispanic

Greek – binary variable equal to one if the student belongs to a fraternity or a sorority.

Finally, we include a set of binary variables for each semester course offering to capture any fixed effects that may be associated with a particular course or the instructor who taught it in a given semester.

A crucial element in estimating the treatment model is to specify an identifying variable in step 1 that affects the selection choice into the tutoring program but not the determination of the final course grade. Two policies adopted by Lehigh University, source of the data used to estimate the model, in the area of intercollegiate athletics are most helpful in this regard.

The first policy relates to the academic profile of recruited student-athletes. Lehigh University is a member of the NCAA Division 1 Patriot League. A founding principle of the Patriot League by its University Presidents was "admitting athletes who are academically representative of their class."[2] For this reason the relationship estimated by (3) and (4) between the set of academic and demographic characteristics of students and the grade received in a particular course should be the same for student-athletes as for (non-athlete) matriculating students.

Secondly, Lehigh University has opted to provide tutoring services to student-athletes through the Office of Student Affairs rather than a program operated directly through the athletics department. This is in contrast to the growing trend among NCAA Division 1 institutions to operate a separate program for student-athletes that reports to the athletic department rather than the academic stem of the university.[3] Lehigh uses its NCAA Academic Enhancement Fund revenues partly to help pay the expenses of this peer tutoring program available to all students enrolled in the university with no fee at point of service, with varsity coaches advising all student-athletes of the services available through the peer tutoring program. Given this institutional arrangement we expect student-athletes to participate more actively in the peer tutoring program.

II. Data

To analyze the pattern of student participation in these areas we collected data for all students enrolled in twenty individual courses over three academic years, 2003-04 through 2005-06.[4] Some, but not all, of these courses are offered in both the fall and spring semesters. This yielded 83 courses with a total enrollment of 18294 students. Because the empirical model includes a measure of each student's grade point average (GPA) at the beginning of a semester, we exclude first semester students thereby reducing the sample to 14761 observations. We also exclude observations where a student withdrew from a course before the semester end and did not receive a final grade, reducing the sample further to 13835 observations.[5]

III. Results

In interpreting the effect of explanatory variables it is important to remember that the dependent variable is the numerical value of the letter grades A through F on a four point scale, including pluses and minuses. Because 0.3 is the numerical value of a plus or a minus, the estimated impact of any explanatory variable must yield an expected value greater than 0.15 for its interpretation to correspond to the magnitude needed to actually result in a change in grade.

Table 1 presents the model OLS estimation results for the simple specification that tutoring participation, or the number of tutorial hours over the course of a semester, exogenously affects a student's course grade.[6] The first column of the table lists the results with the binary variable indicating that a student participated in the program during the semester. The second column lists the results when participation is measured as the total number of tutorial hours during the semester. Column three of the table lists the results when quadratic and cubic terms are included for total tutorial hours.

Most of the control variable estimated coefficients are statistically significant with a plausible direction of effect. These estimates, moreover, are quite consistent as the specification for tutorial participation changes across the three columns. Neither Table 1, nor Table 2 below, presents the estimated coefficients for the lengthy list of course specific dummy variables for space reasons. It is worthwhile to note, though, that many of these are statistically significant from the arbitrarily omitted base course section and each other, indicating that course and instructor effects both influence the decision to participate in the peer tutoring program and manifest themselves in the distribution of final grades.

The estimated coefficient of participating in the peer tutoring program in column one is negative and statistically significant, though not large enough in magnitude to change a student's course grade. Because this unexpected result does not control for voluntary participation in the program, however, a possible explanation is that participation is greater among students who find the course material difficult. The estimated coefficient of participation hours in the second column is positive but not statistically significant. In the cubic specification for column three, however, all three terms are statistically significant (at the six percent level or better).

TABLE 1 – OLS Estimates without Selection

COEFFICIENT	Final Grade- Participation OLS	Final Grade- Hours Linear OLS	Final Grade- Hours Nonlinear OLS
High School Rank Index	0.00799*	0.00801*	0.00800*
	(7.30)	(7.31)	(7.30)
SAT-Math	0.00140*	0.00142*	0.00141*
	(11.8)	(11.9)	(11.8)
Previous Term Cum. GPA	1.018*	1.019*	1.018*
	(69.5)	(69.6)	(69.5)
Term Credit Hours Earned	0.0853*	0.0854*	0.0853*
	(31.9)	(31.9)	(31.9)
Freshman	0.0809*	0.0796*	0.0815*
	(2.97)	(2.93)	(3.00)
Sophomore	0.0631*	0.0621*	0.0632*
	(3.56)	(3.51)	(3.57)
2004-05	0.0110	0.0133	0.00835
	(0.093)	(0.11)	(0.070)
2005-06	0.0497	0.0448	0.0519
	(0.40)	(0.36)	(0.42)
Recitation Courses	-0.0974	-0.100	-0.0967
	(0.82)	(0.84)	(0.81)
Lab Courses	-0.435*	-0.442*	-0.432*
	(3.09)	(3.13)	(3.06)
Male	0.0237*	0.0255*	0.0240
	(1.67)	(1.80)	(1.69)
African-American	-0.285*	-0.287*	-0.286*
	(6.20)	(6.23)	(6.22)
Hispanic	-0.0423	-0.0442	-0.0439
	(1.04)	(1.09)	(-1.08)
Greek Participant	-0.0736*	-0.0743*	-0.0737*
	(5.60)	(5.66)	(5.61)
Tutoring Participant	-0.0607*		
	(2.04)		
Hours Tutored		0.0000418	-0.0393*
		(0.015)	(2.86)
Hours Tutored Squared			0.00390*
			(2.45)
Hours Tutored Cubed			-0.0000799
			(1.87)
Constant	-3.128*	-3.145*	-3.135*
	(19.6)	(19.8)	(19.7)
Observations	13835	13835	13835
R-squared	0.48	0.48	0.48

Absolute value of t statistics in parentheses, * $p<0.05$

Table 2 presents estimates for the treatment model, with the binary measure of participation in the first and second columns and the continuous measure of number of hours tutored in the third and fourth columns. Note first that athletic participation performs quite well as a selection identifying variable, statistically significant at the one percent level in the first stage equation for both the binary and the continuous measures of tutoring participation. [7]

While the estimated coefficient of the binary measure of participation in the peer tutoring program presented in column one is now positive with the selection correction, neither it nor the selection test statistic is significantly different than zero. In columns three and four where the level of participation is measured continuously by the number of hours tutored over the course of a semester, using the tobit treatment procedure, the coefficients of both the number of hours tutored and the selection test statistic are both significant at the one percent level.[8,9] When we include fixed effects for individual students over the six semester period,[10] again using the tobit selection procedure, the estimated coefficients for both the number of hours tutored and the selection test statistic remain significant at the one percent level with the estimated magnitude of the former dropping slightly from 0.0166 to 0.0131.

In order to examine for the existence of gender specific effects, we re-estimated the tobit treatment model, interacting the number of hours tutored with a set of dummy variables that capture the four pairwise cases. The estimated coefficients and their respective t values are:

Female tutoring Female	0.0208*	(2.66)
Female tutoring Male	0.0178 *	(2.68)
Male tutoring Female	0.000099	(0.05)
Male tutoring Male	0.0107	(1.71)

These results suggest that the gender dimension of the peer tutoring process may indeed represent an important and intriguing phenomenon, one that surely warrants further investigation.

TABLE 2 – Treatment Model Estimates

	Participation: Binary		Participation: Number of Hours	
COEFFICIENT	Final Grade (Second Stage)	Participation (First Stage)	Final Grade (Second Stage)	Tutoring Hours (First Stage)
High School Rank Index	0.008*	-0.00428	0.00800*	-0.0784
	(7.35)	(1.21)	(7.30)	(1.34)
SAT-Math	0.001*	-0.00382*	0.00143*	-0.0653*
	(11.04)	(9.79)	(12.0)	(9.68)
Previous Term Cum. GPA	1.020*	-0.149*	1.018*	-2.321*
	(69.19)	(3.14)	(69.6)	(2.94)
Term Credit Hours Earned	0.085*	-0.0177*	0.0853*	-0.289*
	(31.94)	(2.03)	(31.9)	(2.01)
Freshman	0.078*	0.297*	0.0815*	4.426*
	(2.83)	(3.21)	(3.00)	(2.88)
Sophomore	0.061*	0.222*	0.0629*	3.585*
	(3.40)	(3.51)	(3.55)	(3.40)
2004-05	0.017	-0.258	0.0153	-4.232
	(0.14)	(0.71)	(0.13)	(0.69)
2005-06	0.037	1.113	0.0440	19.32*
	(0.30)	(2.77)	(0.36)	(2.84)
Recitation Courses	-0.105	0.477	-0.104	8.973
	(0.89)	(1.30)	(-0.87)	(1.45)
Lab Courses	-0.452*	1.301*	-0.442*	21.98*
	(3.19)	(3.50)	(3.13)	(3.55)
Male	0.029	-0.328*	0.0252	-5.404*
	(1.90)	(7.51)	(1.77)	(7.29)
African-American	-0.290*	0.088	-0.291*	1.714
	(6.28)	(0.74)	(6.33)	(0.89)
Hispanic	-0.047	0.182	-0.0462	2.97
	(1.16)	(1.72)	(1.14)	(1.71)
Greek Participant	-0.076*	0.173*	-0.0737*	2.691*
	(5.70)	(4.11)	(5.61)	(3.83)
Athletic Participation		0.195*		3.295*
		(3.71)		(3.79)
Tutoring Participant	0.102			
	(0.61)			
Hours Tutored			0.0166*	
			(3.12)	
Selection Test Statistic	-0.080		-1.785*	
	(0.99)		(3.68)	
Constant	-3.172*	0.457	-3.141*	6.988
	(19.25)	(0.88)	(19.7)	(0.80)
Observations	13835	13835	13835	13835
R-squared	0.48		0.48	

Absolute value of t statistics in parentheses, * p<0.05

IV. Conclusion

Figure 1 plots the expected effect of the number of hours of peer tutoring on grade point average for the alternative estimating procedures. It seems reasonable to pose that the 'take away' from these combined results is that peer tutoring does indeed produce a positive effect on student learning outcomes. For this effect to translate into an increase in a student's grade in a particular course, though, the results suggest that it is necessary to engage in the activity for ten to twenty hours over the course of a fourteen week semester, something akin to an hour per week. While this result may seem intuitively most plausible to many engaged in education at the university level, it also may provide most helpful empirical evidence to advocates of peer tutoring programs when tightening budgets prompt administrators to consider such programs for the chopping block.

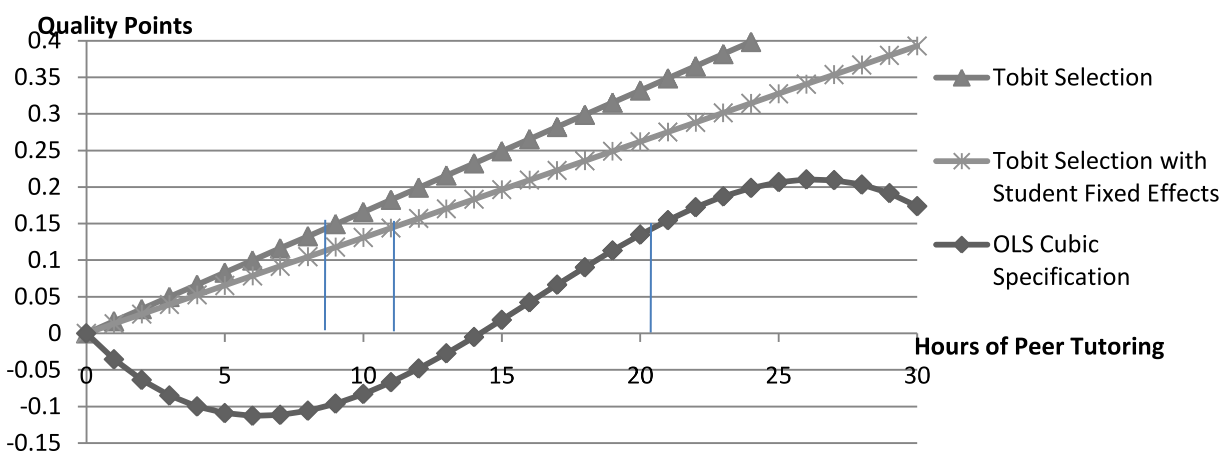

FIGURE1: EFFECT ON COURSE GRADE

References

Betts, Julian R. and Darlene Morell. 1999. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *Journal of Human Resources*, 34(2): 268-293.

Greene, William H. 2008. *Econometric Analysis*, Upper Saddle River, NJ, Prentice Hall.

Munley, Vincent G, Eoghan Garvey and Michael J. McConnell. 2009. "Participation in Peer Tutoring Services by NCAA Student-Athletes." *Unpublished manuscript*.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignments." *Quarterly Journal of Economics*, 116(2): 681-704.

Stinebrickner, Ralph and Todd R. Stinebrickner. 2008. "The Causal Effect of Studying on Academic Performance." *B.E. Journal of Economic Analysis and Policy*, 8(1) (Frontiers): Art. 14.

Thamel, P. 2006. "Athletes Get New College Pitch: Check Out Our Tutoring Center." *The New York Times* (November 4).

Wolverton, Brad. 2008. "Spending Plenty So That Athletes Can Make the Grade." *The Chronicle of Higher Education* (September 5).

Woolridge, Jeffrey M. 2001. *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA, MIT Press.

Zimmerman, David J. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*, 85(1): 9-23.

Endnotes

* Eoghan Garvey was diagnosed with mesothelioma during the later stages of this research project. His untimely passing occurred in April 2009. The final manuscript would have benefitted from his careful input. In addition to being a talented economist, Eoghan was a wonderful human being. He will continue to be missed by many. The usual disclaimer applies.

---

[1] Professional associations whose goals include promoting and enhancing this activity include: The National College Learning Center Association, Learning Support Centers in Higher Education, and The National Tutoring Association.

[2] See Section 2, Article V of the Patriot League Operational By Laws. Full League members include American, Army, Bucknell, Colgate, Holy Cross, Lafayette, Lehigh and Navy.

[3] See, for example, Thamel (2006) and Wolverton (2008).

[4] Student demand for peer tutoring services at Lehigh is concentrated in the areas of accounting, economics, finance, mathematics, chemistry, physics and mechanics.

[5] Summary statistics, not reported due to space limitations, are available from the authors upon request. Students in economics courses comprise over 20 percent of the observations.

[6] The dependent variable is measured on a four point scale corresponding to the letter grades A (4.0) through F (0.0) with a 'plus' adding 0.3 for grades B through D and a 'minus' subtracting 0.3 for grades A through D. Because the range for the dependent variable is bounded by zero and four, tobit estimation of the model with lower limit of 0.0 and upper limit of 4.0 is preferable to OLS. All OLS results presented for (1), both alone and as step 2 of the treatment model, were also estimated using the Tobit procedure. In all instances the degree of statistical significance for each of the estimated coefficients is the same. We present the OLS rather than Tobit results because doing so makes it easier to interpret the magnitude of the estimated effect.

[7] For an in-depth analysis of the relationship between participating in intercollegiate athletics and participating in peer tutoring see Munley, Garvey and McConnell (2009).

[8] We also estimated the treatment model with hours of tutorial participation entered as both a quadratic and a cubic relationship. In both cases none of the estimated coefficients for tutorial participation were statistically significant.

[9] A test of the assumption of identical coefficients of the probit and truncated normal components of the tobit procedure (Greene, 877-878) rejects their aggregation. When we re-estimated the treatment model using the Cragg hurdle variant, neither the coefficient of the linear specification of hours tutored nor the selection test statistic was statistically significant at the five percent level. With a nonlinear specification all three cubic terms of hours tutored exhibit statistical significance but the selection test statistic does not. This brings us back to the estimates presented in column three of Table 1.

[10] The fixed effects specification requires dropping all explanatory variables, for example gender and race, that do not vary over time. The results are not shown here due to space limitations, but they are available from the authors upon request.