Who "Teaches" College Economics Students?
Student Peers or Professors?

by

Sherman Folland
Department of Economics
Oakland University
Rochester, MI 48309
586-752-5738

February 2006

Abstract

A value added panel and a larger posttest sample are applied to test hypotheses regarding the common influences on student performance on the TUCE III. Common influences, those shared by all students in a given classroom, included: day vs. night time of course; class size; peer effects; and teacher effects.

Samples of 500 pre-test students and 400 post tests, collected over a three year period, were matched to the Registrar's student records, Admissions Office records, responses to a survey of the teachers and to department records. Operating by random chance, 76 pretest students were resampled on posttests, these form a value added panel. Complete data for posttest students included 232 cases.

These researches find that: The observed beneficial effects on both value added and posttest TUCE III scores apparently due to the stimulus of the student's peers is most likely spurious, a reflection of the common influence of the teacher. Instead the largest influences on student success in economics are probably the teacher and the student's own intellectual human capital.

JEL: A2

Keywords: Economic education; productivity.

Research to find keys to measure and ultimately to improve student outcomes has distinguished between student-family-neighborhood influences and those influences shared in common by everyone in the classroom. Class size is a tremendously controversial issue in K-12 education, and it receives research attention in proportion (Hanushek, 1997, 1998; Hoxby, 2000). The time of day at which the class is held generates concern among professors, but research in college economic education has generally neglected this issue. Peer effects, where found, appear to be beneficial as a rule, but they are also the most difficult of the influences to identify econometrically (Hanushek et al, 2003). The potential existence of peer effects also raises the blood pressure of politics at the high school policy level; they may be both good news and bad news. For example, if the gains found in some charter schools arise due to better quality peers, the gains may be offset by corresponding losses at the public schools. Meanwhile, little research addresses peer effects in college economics. Finally, teachers potentially provide the dominant common influence on the class: if teacher effects prove dominant in producing student value added, then this would urge us toward better efforts to discriminate among, promote and assign the best teachers to the most critical courses.

These four issues are pursued in the present paper. First, the research develops estimates of the contributions of each common influence to long-term student gains on TUCE III. The span from the pretest, done at the beginning of the micro principles course, to the posttest, taken at the end of the last core economics course, varied for students from four to 36 months, most completing the series of three economics core courses within one half to one and a half years. The study is designed to identify the characteristics of students who best succeed in our economic program, a service program provided to all business students. It identifies elements of instruction that cause economic concepts to be retained long term, that is, the focus is on what "sticks to the ribs."

Second, the larger posttest sample is put to the same tests. Although the paper shows how value added is superior both conceptually and as a satisfactory econometric specification, the posttest results are also useful. They provide a larger sample to develop more precise estimates in cases where value added and posttest tend to corroborate each other. The two often corroborate each other, suggesting that much of the signal in the value added score comes from the posttest

With both samples, peer effects appear to be a strong common influence in specifications that describe the teachers by their experience. However, by alternate means of controlling for teacher influences, one is able to show that "peer effects" disappear, suggesting at best a very weak role for student peer effects in these college economics data. In the final analysis, teachers dominate the common influences.

Section I examines the education literature with special emphasis on these common influences and their effects. Section II describes the data, its method of collection, and issues that potentially threaten sample bias and how these were addressed. Section III discusses the theory of estimation and methods to identify peer effects and teaching effects. Section IV generates the econometric results for value added, with related data tables. Section V does the same for the posttest sample. Section VI discusses the results, offers conclusions, and suggests avenues for further research.

I. The Literature

While most published estimates of class size effects address the high schools, there are a few that study college economics. Kennedy and Siegfried (1997) measured student achievement by performance on the TUCE III. Focusing on classroom effects, they assembled a large number of observations averaged to the class level. These authors found that class size has no effect on achievement, and they further claimed that no class characteristics controllable by teachers or department chairs have any effect. An earlier study (Raimondo and Esposito, 1990) had found adverse effects of class size on macro principles, though this design was not as extensive as that of Kennedy and Siegfried.

"No class size effect" appears to be the general consensus of the K-12 literature, too; though there the high stakes riding on the results have generated a more thorough scrutinizing of the issue as well as more econometrically refined methods. Eric Hanushek (1997, 1998) has surveyed this literature on a frequent basis, comparing, most recently, 277 estimates from 59 studies. He concludes that "there is no strong or consistent relationship between school inputs and student performance." Caroline Hoxby (2000), studying a large sample of elementary schools draws a similar conclusion: "class size does not have a statistically significant effect on student achievement". Jepsen and Rivkin (2000) find some benefit from small class size, but show this to be contradicted by the reductions in teacher quality that accompany the reduction in class size.

The consensus I have suggested to exist is not unanimity. Alan Krueger (2003) disputes Hanushek's meta-analysis, arguing in part that the grouping of studies in Hanushek is arbitrary and that an alternative consolidation leads to contrary size results. Krueger and Whitmore (2000) in analyzing data from the Tennessee STAR experiment find that students experiencing early, small classes later on scored higher on their SATs and subsequent college tests.

The give and take in the K-12 literature has generated a better understanding of the dangers of misspecifying the estimating equations and of failure to properly identify the class size effect. This is especially made clear in Hanushek, Kain and Rivkin (1998) and in Hoxby (2000). Where multiple schools are sampled, the student and family's choice of school district can affect class size, rendering it endogenous. This effect may prove difficult to remove, whether through instrumental variables or the use of fixed effects models. While multi-school studies are advantageous for generalizability when successfully done, they add difficulties for identifying the effects that are focused on.

The difference between day-time and evening courses is rarely a focus of college economics studies, however, it is potentially relevant to both administrators, who target cost efficiency, and to teachers, who worry about student learning. The presumption in the literature is that this and other available policy instruments have no effect on student performance. Yet, the duration of evening classes is much longer than day sessions, and many of these students work during the day. Is the difference in performance negligible for economic classes when controlling for individual student characteristics and for the other common influences?

In contrast to day/night issues, the question of peer effects in the classroom has drawn substantial interest as well as controversy. It is clear that students are influenced by their peers in other settings. Studies of randomly assigned college roommates (Sacerdote, 2000; Williams and Zimmerman, 2003) reveal that roommates who differ in intellectual ability influence the "other's" performance, though in a nonlinear fashion. For example, a student with mid level SATs gets poorer grades if he shares a room with someone of low level SATs. If peer influences are clear at the individual level, however, what effects do they have in the classroom?

Much of the work on peer effects in the classroom focuses on K-12 education. This is natural as the policy implications are weightier, especially for school vouchers policy. Under an open voucher program, the better students may choose private schooling; if peer effects are important to performance, those left behind may suffer academically.

Angrist and Long (2002) studied peer effects in the Boston Metco program, a plan designed to desegregate the schools. Using instrumental variables to remove endogeneity, they found that the negative effects of an influx of lower achieving peers are small and tend to influence only a subset of students. They concluded: "Any peer effects from Metco are modest and short-lived". Hanushek and colleagues (2003) investigated panel data from the UTD Texas Schools Project. Their paper is especially useful in defining the econometric problems that must be overcome to identify classroom peer effects. Most notably the peer influences are simultaneously determined, implying that the peer characteristics measures are endogenous. Also, individual student, family, and school influences must be extracted so as to reveal the peer effects. These authors propose and apply solutions to such problems, and they find that higher achieving student peers have a positive effect on student achievement throughout the school system.

Peer effects in the economic college classroom have been less often studied. I found no college economics education study focused on peer effects on student achievement. However, Ashworth and Evans (2001) found both peer effects and teacher effects relevant to the student's decision to study economics. An indirectly applicable paper (Brewer, Eide, and Ehrenberg, 1996) inquired whether students gained an advantage in attending an elite private college, where presumed student peer quality and teacher quality is higher; they found positive effects after controlling for student characteristics.

Teachers contribute a common influence on student learning, and they pose a possible confounding effect when studying peer effects. Teacher characteristics have been shown to influence student performance as well as choice of study. Gender and ethnicity function to make some teachers more effective role models for some students (Jensen and Owens, 2001). Other studies attempt to measure teacher quality by testing the marginal effects of selected teacher characteristics such as educational background (Lynch, 1990; Bosshardt and Watts, 1994). Laband and Piette (1995) explore the role of the teacher indirectly, inferring a material difference between community college teaching and four year college teaching by examining the relative performance of students who transferred in from a community college. More directly, Kennedy and Siegfried (1997) search for teacher effects within an economic college program, though they found no effects. Recently, Hanushek and colleagues (2001) found, in a research context intended to meticulously extract confounding effects, that teachers make a substantial contribution to high school student intellectual progress.

On examining this literature it is clear that each of the targeted common influences on student performance have been studied, sometimes very thoroughly. It also becomes clear that the investigations of these focus effects rarely address the college economics classroom.

In the several studies discussed, related variables have been employed as controls. Prior studies present mixed effects for gender with the more common result being an advantage to males. Anderson, Benjamin and Fuss (1994), for example, find that males handle principles of economics more effectively than females; and Greene's (1997) study tends to discount previous claims that females showed better reading comprehension skills in economics. Williams, Waldauer and Duggal (1992), in contrast, find no significant differences between the sexes.

Ethnicity effects have been studied for college courses generally with the well-known results of positive advantages for whites and Asians. As a first guess, one might expect that these results apply to economics as well. However, this presumption is qualified by a study by Simkins and Allen (1999), which compared African American and white students and found that while the black students began at a lower level of economic knowledge, their value added was greater than for the white students.

Measures of student preparedness and/or ability, such as the SAT and ACT, usually show positive and significant associations with student achievement in economics (Kennedy and Siegfried, 1997; Simkins and Allen, 1999), though at least one study (Laband and Piette, 1995) has found no association. A related measure, the student's prior cumulative GPA, may represent a combination of ability and student behavior patterns that contribute to academic achievement. This variable sometimes proves significant (Laband and Piette, 1995; Simkins and Allen, 1999; Anderson, Benjamin, and Fuss, 1994) though in one value added study it did not (Kennedy and Siegfried, 1997).

Many states require four-year public universities to accept transfers from qualified community college students. At issue is whether these students are adequately prepared for a university level education. Anecdotal evidence suggests that many highly qualified individuals matriculate from the community colleges, but is the average of such students sufficiently prepared? A study by Laband and Piette (1995) finds a significant negative effect of transfer status on economics achievement.

Other variables: Bishop (1998) examines the effect of the curriculum-based external exit exam; Saunders and Powers (1995) propose content changes in course materials; Borg and Shapiro look at student personality type (1996). Others have tried laboratory approaches (King and LaRoe, 1991). A collaborative learning project was evaluated by Johnston, James and Lye (2000); they reported improved student morale yet little or no gain on tests of economic knowledge. Finally, a high school background in economics as well as high school academic success has been shown to affect college economics performance (Anderson, Benjamin and Fuss, 1994; Brasfield, Harrison, and McCoy, 1993).

II. The Data

To reexamine these and related hypotheses, the process of data collection was begun in Fall of 1999, and during the following two years 500 students were tested at the beginning of their principles of microeconomics course, a testing period which ended in Fall 2001. Beginning in Winter 2000 and continuing through Fall 2002, 409 posttests were given at or near the end of the three course economics core, which is required of all students in the School of Business Administration. A panel was formed by matching students in the pre and posttest pools. This panel includes 76 students, and forms the basis for value added measures and related hypothesis testing.

Because this reduction in cases is dramatic, it will help to discuss the reduction steps and establish the degree to which they may or may not introduce bias. First, the 409 posttests were reduced in number to 232 because of missing data for needed items. There is no known reason that these missing items would be related to study variables. The Registrar's Office provided very rich data, but while most items are carefully filled in, incomplete items happen in commonplace and must be accepted by the Registrar. Attempting to match the Registrar's data with the test data, also resulted in non-matches of persons, mainly due to transfers out of the program, drop outs from school, or from unusually short or long student delays in registering for the follow-up class. These latter events may cause him not to appear even within the wide time window of the data available from the Registrar. In summary, the large majority of lost cases are missing data problems without any known bias introduced. The possible exception, dropouts, which could result from poor student performance or from loss of income, could introduce bias but were the minor portion of the lost cases. The larger posttest sample provides a useful comparison. Researchers have found that the value added equations provide similar empirical results to the generally larger posttest results, a plausible result since pretests are so difficult for most students as to be largely random.

Both the value added panel and the posttest sample have been augmented in four ways. First, a survey of the participating faculty was conducted, which yielded data on teacher experience and research activity, as well as personal background and teaching philosophy. Second, class, classroom, and related course characteristics were collected from school and department sources. Third, the University Registrar's Office provided the needed data on most student characteristics, both demographic and academic data.

Finally, the University Admissions Office provided records of the student's academic characteristics upon admission to the university including ACT scores.

TABLE 1 ABOUT HERE

The data items are listed and defined in Table 1. Table 2 presents the descriptive statistics of both study samples. As noted, the data items matching process revealed missing data for the posttest sample, reducing the number of observations to a complete variable set to 232. ACT data, which are collected at this university on a voluntary basis, have more missing values; thus, the ACT regressions are based on a lesser number, 131 observations.

TABLE 2 ABOUT HERE

Faculty participation was voluntary throughout. Each teacher was asked to offer students a small academic incentive to take the TUCE test diligently. Specifically, each was recommended to offer the equivalent of between .05 and .10 point addition to the student's final course grade for completing the test, with the full credit obtained for scoring in the top half of his or her class. In almost all cases, an incentive was given, though in several cases the teacher's choice was to give the academic bonus solely for taking the test. Of the 16 classes given the posttest, only one was given without an incentive that benefited the final course grade; and, the use of incentives in the pretest was at a similar level. Over the 16 post test courses, an average of 83 percent of the students enrolled in the course were tested on the TUCE. The percentages ranged from 69 to 91 percent among the classes.

Student scores were recorded item-by-item giving the opportunity to examine performance on clusters of questions. Briefly, I have selected groups by subject matter: "No Free Lunch", questions that required understanding of opportunity costs; "Supply & Demand", questions that could be answered by students with a good grasp of supply and demand curves; "Externalities & Public Goods", questions on concepts that are relatively easy to teach; "Marginals and Elasticity", questions of a more technical and purely economic character; "International", questions focusing on comparative advantage or international monetary systems. These subsections were of different size, and to make the groups easier to compare, they were recalibrated in Table 3 to the zero/one range, which can be interpreted as the fraction of the question group that was answered correctly.

TABLE 3 ABOUT HERE

With 3.128 additional correct over the three course series, a modest improvement of just over 30 percent, these sampled students and their program seem comparable to other reported studies. Published studies of value added in college economics often report average gains of four to six points (questions) on the 33 point TUCE III test, but these typically test at the beginning and the end of the single course. In contrast, the present study, ranging from the beginning to the end of a three course sequence finds gains of only about 3 points. In principle, the comparison could go either way, that is, the span allows enough time for either depreciation of knowledge capital or further investment through relevant coursework and other communication. The TUCE, however, is focused primarily on the principles topics, and the case here, which indirectly suggests that a depreciation of knowledge, should not be unexpected. It is odd that more studies do not address long-term economic learning, because this is where lies our greater challenge, and it is closer a typical school's mission.

These pretest data, collected at the beginning of their first principles course, show that students without any formal preparation do their best on the supply and demand questions, though they also make their smallest gain in this concept area. On the more technical subjects of marginals and elasticity, they quite naturally do their worst (note that pure chance would yield an expectation of 0.25 correct). Their gains are also modest on such topics. In contrast, it appears that the message many economists consider to be the most fundamental, "no free lunch", is also the area of greatest gain.

Section III. Theory and Econometric Specification Issues

Several econometric roadblocks must be overcome to identify the contribution of each of the targeted common influences. Consider first that each may be part of a system that partially determines their value, thus implying possible endogeneities that must be removed. For example, the effects of families and schools are problematic issues where multiple schools are sampled; families and students may arrange themselves spatially in a Tiebout-like process (Rivkin, 2001) making school and student related in this matching. The present three-year sample of a single school and student body avoids this problem, though at some cost; the multi-school approach has an easier claim to generalisability, provided the inherent problems are addressed.

A related problem can arise if there are omitted variables describing either student or class. I include several student variables, such as *GPA, ACT* scores, transfer status, and gender, but unknown other relevant variables may exist that are omitted. Let the determinants of student achievement be, $S_i, SC_{ik}, F_i, Cl_{ij}, ClR_{il}, T_{im}$, the student, school, family, class peers, classroom, and teacher characteristics. Let the achievement function in period *t* be defined by Equation (1) where vectors of characteristics are shown in product with vectors of constant coefficients, here subscripts for individuals and the various influences have been suppressed for simplicity.

(1) $A_t = a_{ot} + Sa_{1t} + SCa_{2t} + Fa_{3t} + Cla_{4t} + ClRa_{5t} + Ta_{6t} + u_t$

Yet, many of these same characteristics do not change between pre and posttests, for example, gender and parents. An ever present problem in specifying education production functions is that many student, student family or school characteristics are difficult to observe or if observed, difficult to quantify and extract. Thus they become omitted variables that may introduce biases. So, by defining the dependent variable as the value added measure, $A_t - A_{t-1}$, these influences in principle are netted out. This generates the value added equation as follows here:

$$V_t = a_o^{'} + [Cl_t - Cl_{t-1}]a_4 + [ClR_t - ClR_{t-1}]a_5 + [T_t - T_{t-1}]a_6 + \upsilon_t$$

A practical modification is that some student characteristics do change, such as *Months* between pre and post test, and these are applied in the analysis. However, to better compare this study with previous work, the variables *Male, Ethnic White, GPA* and *Transfer* status are included in separate regressions.

Potential problems of endogeneity are of several types. First, class characteristics, *Cl*, might not be randomly distributed across classes, for example, students may cluster in classes with friends who share similar intellectual interests and abilities. Perhaps more important, the student peer effects create an endogeneity in themselves, the individual simultaneously affects and is affected by his peers. The present design and database cannot address this potential problem. Hanushek and colleagues (2003) suggested a means to address this by regressing value added against lagged peer characteristics *inter alia,* but that approach, while suitable for high school studies is not practicable for college studies where peer groups change from class to class.

Second, class size is also measured in the *Cl* vector. The potential endogeneity of class size has already been described, however, that it is more plausibly a problem of multi-school studies. There many high school options are available to students and parents in a spatial sorting process; a process that might result in the more popular schools exhibiting, at least temporarily, larger class sizes.

Third, classroom characteristics, *ClR*, specifically the day vs. night issue, will likely have little effect on the student peer group. This public university is a commuter school where nearly all students have jobs, most take both night and day courses, and teachers report little difference in average grades. Included student characteristics should be adequate to control for unanticipated differences.

Finally, teachers will draw some students who follow them, but the classes studied are all required courses and for our working students time and day of the week are at least as important as the teachers. Teachers have seen the micro TUCE III briefly, but no one has or has expressed any incentive to teach to the test. Teacher effects here are measured several ways, alternatively by teacher responses to a survey as well as by fixed effects.

Section IV. Value Added and Influences Common to the Class

The common influence measures include day vs. night courses, (*Preday and Postday);* measures of the number of registered students in the class, *PreSize and PostSize*; *Peer Score* (measured as the average TUCE score of fellow students), and *Teacher Experience* (measured initially by years of experience and then alternatives are treated). Students/family characteristics netted out by assumption, except that alternate equation versions are presented with several student specific items included.

TABLE 4 ABOUT HERE

Table 4 presents the results from OLS regressions with *Value Added* as the dependent variable. Several facets are noted in particular. First, the addition of student personal characteristics in Equation (2), improves R Square on the margin though most of these variable are insignificant as theorized. An alternative hypothesis may be that some personal/family characteristics contribute to learning in ways not otherwise accounted for in differencing the TUCE scores. Though the change is small, I will continue to show the constrasts in what follows.

Second, peer effects, which fail to appear in Equations (1) and (2), become positive and strong where day/night and class size variables are combined. Other regressions (not shown) remove the size variables with the same result. We get another look at this conflicting relationship when applying the larger posttest sample, but here, peer effects appear to be possibly important though somewhat unreliable.

The day vs. night results appear counter to popular intuition: night students register greater not lesser gains. This could indicate that the nighttime long class hours are more a negative for teachers than for the students. Alternatively, unmeasured student/family characteristics are important to student gains in the night courses. This would require as well some explanation of why students with these characteristics would cluster in the night courses.

Finally, I considered that the irregular time frame of the study in which students effectively choose the span from pre to posttest might influence the results: Do students who take their last core course (and thus the posttest) later do better (worse?) on their posttest. To test this possibility a measure was generated for each student of the number of months between pretest and posttest; this measure was then included in both value added and posttest regressions. This time span had neither any significant effect nor any noticeable influence on the other coefficients. As an alternative way to look for this potential effect, we see in Table 5 the cross-tabulation of the numbers of students in each category of time span crossed with their average test scores. Few differences are noted and the pattern is statistically indistinct regarding the time span.

TABLE 5 ABOUT HERE

Section V. Posttest and Effects Common to the Class

Given the probabilistic method of matching students, pretest to posttest, the value added sample was relatively small compared to the posttest sample; the posttest sample in contrast is based on 232 students. We note, in Table 6, that this larger sample renders day/night and class size insignificant, and it reports significant peer effects. This suggests that the previous, mixed results for peer scores might be related to the value added sample. We also note that the insertion of an alternative peer influence measure, the mean GPA of one's fellow students competes poorly with the current peer scores. This suggests that the peer effects, if they exist, may be primarily related to some current experience of the class, rather than human capital characteristics of the class.

TABLE 6 ABOUT HERE

The student characteristics, *GPA* and *Male*, are both significant contributors to TUCE III scores. The results both correspond to findings in prior literature, and though GPA makes good sense as a measure of intellectual ability and academic determination, the role of gender is not well understood and is sometimes disputed.

Does the nature of TUCE III itself distort the results? The TUCE is heavily weighted toward principles level questions; critics also ask whether portions of the test are too technique dependent or too jargon laden to truly be measure understanding of concepts. One way to examine the issue is to break down the TUCE into parts, some technical or jargon-involved and some more conceptual strongly intuitive.

The five theme groups of questions selected from the TUCE III serve this purpose: "No Free Lunch", "Externalities & Public Goods", "Supply & Demand", "Marginals & Elasticities", and "International Trade".

Student mean scores in these groups, as reported in Table 7, suggest that basic, intuitive economics sticks with you better than technique; "No Free Lunch", "Externalities & Public Goods" and "Supply & Demand" record the highest mean percentage correct. If males have an advantage, it appears to be on the technical side, questions such "Marginals & Elasticities" and "International Trade". The advantage across the board goes to "good students", those with a higher GPA.

TABLE 7 ABOUT HERE

Another issue worth pursuing is whether student intellectual ability is adequately measured by the GPA or college boards such as ACT would add significant information. ACT data at this university is voluntary, thus reported mean ACTs will be somewhat inflated from the true level. However, regressions within the data set reporting ACTs should indicate their importance on the margin.

TABLE 8 ABOUT HERE

As shown in Table 8, the addition of *ACT* scores adds explanatory power to the equations. The best performing measure of the *ACT*s is the combined measure, *ACT-C*. The verbal/reading score, *ACT-R*, makes essentially the same contribution as the math, *ACT-M*. They cause no deterioration in the performance of either the student's *GPA* or the other effects. Most notably, the observed peer effects remain strong.

TABLE 9 ABOUT HERE

Are These "Peer Effects" Merely Reflections of Good Teaching?

Critics of reported peer effects point to the need to extract confounding teacher effects to identify the role of peers properly. So far, we have treated teaching experience as the measure of teacher quality. This needs to be challenged; the issue is approached in two ways. First, alternative measures of teacher quality are inserted into the regression. Then, fixed effects are entered for the teachers.

Table 9 reports regressions including alternative teacher measures (variables not presently focused on the teacher peer effects issue are suppressed). Professor characteristics are alternately measured as (1) *Teaching Experience* as before (years since attaining the PhD); (2) *ARTs* number of refereed articles published, as indexed by EconLit; (3) *Rank* of the teacher's graduate school (*Rank* is the reciprocal of the published ranking and thus increases with school quality; and (4) *Full Time*, a dummy variable distinguishing between full time and part time faculty.

The performance of the *Full Time* faculty dummy is noteworthy, because it tends to reduce the significance of *Peer Score*. The university mission, a mix of research and teaching, puts a premium on the teaching quality of full time faculty, they are screened upon hiring for indications of high teaching ability. Thus the present result suggests that teacher quality effects, when they are omitted, may cause an upward bias in the coefficient of *Peer Score*.

TABLE 10 ABOUT HERE

In Table 10, the peer score coefficient does well in Equation (1) but is substantially diminished with the addition of fixed effects representing the individual professors in Equation (2). The *t* value for *Peer Score* becomes insignificant and its coefficient takes the "wrong" sign. We can conclude tentatively that teacher effects are important and may be the underlying cause for the apparent peer effects. This much is consistent with the research by Hanushek and colleagues (2003) who found for high school students that peer effects exist but are modest, while teacher effects are substantial and even so are probably understated in currently published research.

To pursue this issue further, I developed at this stage of the analysis, a substantially larger data set consisting of only the focus variables. This was possible because the smaller set of variables selected—*GPA*, *Peer Score*, Teachers indicators—are complete for many more student records. The newly omitted variables were typically insignificant in the final equations and at best are only weakly correlated with the variables to be included. While there is a cost to dropping variables, the benefit is that this provided 161 additional complete cases.

This larger data set served two purposes: 1) this gives can provide more precise estimates of the coefficient for *Peer Score*; and 2) this provides sufficient numbers of cases for individual teachers so that separated regressions could be tested "within group", that is, within the group of students what had taken a given teacher. These latter regressions treat only those teachers who taught multiple classes.

Table 11, which reports the regressions on the enlarged data set, compares the performance of *Peer Score* under three different versions of the equations. Entered alone, this variable is easily significant but controlling for student GPA, changes this result very little. Entering fixed affects for teachers, in Equation (3), much improves the equation's explanatory power, but reduces the performance of Peer Effects substantially.

TABLE 11 ABOUT HERE

Under certain assumptions, grouping students by teacher provides a better test. When the dummy variable is applied to represent teachers, this perhaps too rigidly forces the teacher effects on the student scores to be a fixed, constant amount; if teacher/student interactions are more individualized and complex, we would improve the estimation by allowing the for a more general effect. Further, if the teacher contributions are similar across their own classes, grouping students by teacher eliminates teacher effects entirely. Table 12 presents the regressions by individual teacher, T-i, for the five teachers who taught multiple sections. These results suggest that the apparent peer effects were spurious, that is, they were probably biased upward due to an insufficient measure of teacher characteristics.

TABLE 12 ABOUT HERE

VI. Discussion and Conclusions

What can be concluded? The present study finds that observed student peer effects are reduced and even eliminated in these samples. This result is generally consistent with that of Hanushek (2003), who finds that while peer effects exist they are small compared to teacher effects in K-12.

Here teachers are the most important of the common effects on college student performance in economics, and this suggests the need to reemphasize efforts to put the best teachers in the right classrooms. But this is a commonplace to deans and department chairs. It may be more productive to address the question of how one assesses which teachers are the 'best".

According to reports, it remains common throughout the country to evaluate teaching by Student Evaluations of Teachers (SETs), without resort to other supporting materials. The SETs alone policy is weak, because these student self reports are poorly correlated with student objective performance; the psychology literature, where much of the research on student testing is done, reports that global SET scores attain only a 0.40 correlation with objective final exams or standardized tests. Used alone they could account for only 16 percent of the variation in student end of term skill performance tests

(Abrami et al, 1990). The SATs have a similarly weak ability to predict, in this case to predict college freshman grades; the irony is that SATs are most commonly used by universities only in conjunction with supporting evidence. This practice of portfolio assessment makes sense as well for assessing teacher quality.

A bigger caution derived from the present research, however, is that very little of the variation in these students' objective performance could be explained by teachers and all of the other common influences that were examined. These inquiries reinforce the view that very few policy instruments can be found that would allow the dean or chair of an economics department to use to improve student outcomes (Kennedy and Siegfried, 1997; Hanushek et al, 2003). Eric Hanushek even asked: "What if no best practice exists?" (2004).

We are left with the question: Is the glass three quarters empty or one quarter full? Avenues for optimism may lie with specific, focused interventions, such as classroom experiments, games, writing, projects, on-line study guides, and experiments. Research, both on what works and what does not work in college economic education serves as a stimulus to innovation in this field.

Acknowledgements: Carl Liedholm provided Assistance through suggestions and comments. .

References

Abrami, PC et al. 1990. Validity of Student Ratings of Instruction, Journal of Educational Psychology 82:219-231.

Anderson, G., D. Benjamin, and M.A. Fuss. 1994. The Determinants of Success in University Introductory Economic Courses. *Journal of Economic Education* 25(Spring): 99-119.

Angrist, J.D. and Lang, K. 2002. How Important Are Classroom Peer Effects? National Bureau of Economic Research Working Paper 9263.

Ashworth, J. and Evans, J.L. 2001. Modeling Student Subject Choice. Journal of Economic Education 32:311-20.

Bishop, J.H. 1998. The Effect of Curriculum-Based External Exit Exam Systems on Student Achievement. *Journal of Economic Education* 29(Spring): 171-82.

Bonesronning, H. 2004. Can Effective Teacher Behavior Be Identified? Economics of Education Review 23(June):237-47.

Borg, M.O. and S.L. Shapiro. 1996. Personality Type and Student Performance in Principles of Economics. *Journal of Economic Education* 27(Winter): 3-25.

Bosshardt, W. and Watts, M. 1994. Instructor Effects in Economics in Elementary and Junior High Schools, Journal of Economic Education 25:195-211.

Brasfield, D.W., D.E. Harrison, and J.P. McCoy. 1993. The Impact of High School Economics on the College Principles of Economics Course. *Journal of Economic Education* 24(Spring): 99-111.

Brewer, D.J., Eide, E. and Ehrenberg, R.G. Does It Pay To Attend An Elite Private College? National Bureau of Economic Research Working Paper 5613.

Greene, B. 1997. Verbal Abilities, Gender and the Introductory Economics Course: A New Look at an Old Assumption. *Journal of Economic Education* 28(Winter): 13-30.

Hanushek, E.A. JF Kain, JM Markham, SG Rivkin. 2003. Does Peer Ability Affect Student Achievement? Journal of Applied Econometrics 18:527-544.

.

Hanushek, E.A. SG Rivkin, JF Kain. 1998. Teachers, Schools, and Academic Achievement. Econometrica, 23(2): 417-58.

Hanushek, E.A. 2004. What If There Are No Best Practices? *Scottish Journal of Political Economy* 51:156-72

Hanushek, E.A., 1999. Evidence on Class Size. *Earning and Learning: How Schools Matter* (Washington, D.C.: Brookings Institution,), pp. 131-168.

Hoxby, C. 2000. The Effects of Class Size in Student Achievement. Quarterly Journal of Economics 115:1239-1285.

Jensen, EJ and Owen, AL. 2001. Pedagogy, Gender and Interest in Economics, Journal of Economic Education 32:323-43.

Jepsen, C. and Rivkin, S. 2002. What Is The Tradeoff Between Smaller Classes and Teacher Quality? National Bureau of Economic Research, Working Paper 6205.

Johnston, C.G., R.H. James, and J.N. Lye. 2000. An Evaluation of Collaborative Problem Solving for Learning Economics. *Journal of Economic Education* 31(Winter): 13-29.

Kennedy, P.E., and J.J. Siegfried. 1997. Class Size and Achievement in Introductory Economics: Evidence from the TUCE III Data. *Economics of Education Review*. 16(4): 385-394.

King, P.G. and R.M. LaRoe. 1991. The Laboratory-Based Economics Curriculum. *Journal of Economic Education* 22(Summer): 285-92.

Krueger, A. 2003. Economic Considerations and Class Size. Economic Journal 113:F34-63.

Krueger, A. and Whitmore, D. 2000. The Affect of Attending a Small Class in the Early Grades on College Test Taking., National Bureau of Economic Research Working Paper 7656.

Laband, D.N., and M.J. Piette. 1995. Does Who Teaches Principles of Economics Matter? *American Economic Association Papers and Proceedings* 85(May): 335-338.

Lynch, GJ. 1990. The Effect of Teacher Course Work on Student Learning, Journal of Economic Education 21:287-86.

Raimondo, H.J., L. Esposito, and I. Gershenberg. 1990. Introductory Class Size and Student Performance in Intermediate Theory Courses. *Journal of Economic Education* 21(Fall): 369-381.

Rivkin, J. 2001. Tiebout Sorting Aggregation of the Estimation of Peer Group Effects, Economics of Education Review 20(June): 201-09.

Sacerdote, B. 2001. Peer Effects with Random Assignment, Quarterly Journal of Economics 116:681-704.

Saunders, P., and J.R. Powers. 1995. Reallocating Content Coverage in Principles of Microeconomics to Increase Student Learning. *American Economic Association Papers and Proceedings* 85(May): 339-342.

Scott, L.C., and P.M. Mitias. 1996. Trends in Rankings of Economics Departments in the U.S.: An Update. *Economic Inquiry* 34(April): 378-400.

Simkins, S. and S. Allen. 1999. Cognitive Learning in Principles of Economics at Historically Black and Traditionally White Colleges and Universities: Differences in Value Added. Paper presented at the National Economics Association meetings in January.

Williams, GC and Zimmerman, D.J. Peer Effects in Higher Education. National Bureau of Economic Research Working Paper 9501.

Williams, M.L., C. Waldauer, and V.G. Duggal. 1992. Gender Differences in Economic Knowledge: An Extension of the Analysis. *Journal of Economic Education* 23(Summer): 219-31.

Table 1. Variable Definitions and Short Names

ACT Composite	ACT composite score.
ACT Math	ACT score on mathematics
ACT Reading	ACT score on reading.
Articles	Teacher number of published articles indexed in EconLit.
Peert GPA	Average GPA for fellow students in post class.
Peer Score	Average TUCE score for fellow students in post class.
EthnicWhite	Equals 1 if student is white; zero else.
Full Time	Equals 1 if teacher is a full time faculty, adjunct=0.
GPA	Cumulative grade point average.
Male	Equals 1 if student is male; zero else
PostDay	Post-test course taken during the daytime; zero else.
PostScore	Total number correct in post-tested TUCE III out of 33.
PreDay	Pre-test course taken during the daytime; zero else
PreScore	Total number correct in pre-tested TUCE III out of 33.
PrinGrade	Student's grade point in microeconomic principles.
Rank	Ranking of school where teacher's PhD was attained
Transfer	The student transferred into the university.
ValueAdd	TUCE III score improvement from pre-test to post-test.
YearsPhD	Teacher years since attaining the PhD

Notes: The GPA value was taken as the cumulative measure just prior to the student beginning the first course in the economics sequence. Transfer is set to zero for students matriculating first time into this university. Cohorts values are calculated by removing the student's own value before calculation of the class mean. ValueAdd included a number of negative scores, suggesting that the test taking, memory, or other student conditions contain random influences.

Table 2. Descriptive Statistics for the Two Samples: Panel and Post Test Samples

	Panel (N=76)			Post Test Sample (N=232)		
Variable	Mean	Min	Max	Mean	Min	Max
ACT C	21.42	16.00	28.00	21.91	14.00	30.00
ACT M	22.25	15.00	30.00	22.59	15.00	33.00
ACT R	21.14	13.00	29.00	21.41	12.00	34.00
Articles	8.84	0.00	22.00	8.34	0.00	22.00
Peer Score	12.20	7.98	15.40	12.83	9.54	14.89
Peer GPA	3.04	2.66	3.31	3.09	2.88	3.29
EthnicWhite	0.824	0.00	1.00	0.847	0.00	1.00
Full Time	0.776	0.00	1.00	0.814	0.00	1.00
GPA	2.99	2.19	3.96	3.08	0.00	4.00
Male	0.407	0.00	1.00	0.452	0.00	1.00
PostDay	0.315	0.00	1.00	0.621	0.00	1.00
PostScore	12.42	6.00	29.00	12.948	4.00	30.00
PreDay	0.789	0.00	1.00	---	---	---
PreScore	11.32	4.00	25.00	---	---	---
PrinGrade	2.82	0.00	4.00	3.06	0.00	4.00
Rank	0.032	0.0138	0.111	0.034	0.013	0.058
Transfer	0.526	0.00	1.00	0.240	0.00	1.00
ValueAdd	1.096	-8.00	14.00	---	---	---
YearsPhD	15.92	1.00	30.00	15.40	1.00	30.00

Table 3. Performance on the TUCE in Whole and in Parts

Test Portion	Mean Pretest	Value Added	Mean Post Test
Whole Test (33 possible)	9.820	3.128	12.948
No Free Lunch (4)	0.299	0.165	0.464
Supply & Demand (6)	0.349	0.060	0.409
Externalities & Public Goods (5)	0.302	0.118	0.420
Marginal & Elasticity (5)	0.262	0.078	0.341
International (3)	0.286	0.099	0.385
N	500		232

Note: Whole test figures are number of correct out of 33 possible. Other figures are fractions correct out of total group number possible given in parentheses. NFL, "no free lunch", questions employing opportunity cost concepts (4); S&D, supply and demand questions (6); Ext & PG, "externalities and public goods" questions (5); MR & Elast, "marginals and elasticity questions (5); Internatl, international economics questions (3).

Table 4. Common and Student Effects Regressed on Student Value Added Score on TUCE III Over the Core Economic Series

Dependent Variable= ValueAdd; n=76

Independent variable	(1) Coeff (t)	(2) Coeff (t)	(3) Coeff. (t)	(4) Coeff (t)
Constant	20.99 (2.30)	12.56 (1.29)	-6.69 (2.48)	-4.47 (1.02)
PreDay	-1.63 (1.44)	-1.98 (1.76)	--	--
PostDay	-3.92 (2.35)	-3.51 (2.21)	--	--
PreDaySize	--	--	-0.564 (0.53)	-1.146 (1.08)
Pre Size	-0.026 (0.36)	0.013 (0.15)	--	--
Post Size	-0.392 (3.07)	-0.272 (2.20)	--	--
PostDaySize	--	--	0.641 (0.58)	-0.588 (0.55)
Male	--	-0.615 (0.69)	--	-0.854 (0.92)
Ethnic White	--	2.59 (2.15)	--	2.97 (2.32)
GPA	--	-1.02 (0.85)	--	-0.607 (0.49)
Principles Grade	--	0.263 (0.37)	--	-0.314 (0.45)
Peer Score	-0.083 (0.22)	0.137 (0.39)	0.694 (2.85)	0.597 (2.62)
Teacher Experience	0.416 (0.71)	0.019 (0.35)	-0.044 (0.86)	-0038 (0.82)
R Square (p value for F)	0.234 (0.004)	0.313 (0.007)	0.112 (0.072)	0.243 (0.035)

Table 5. Crosstabulation of Months Lag Between Pretest and Post Test and the Student's TUCE III Score.

Months/TUCE Score	< 11 Correct	11 to 16	> 16
Under 12 months	66	48	39
Over 12 months	21	19	9
% Under 12 months	75%	72%	81%

Table 6. Post Test Regressions of Student Characteristics and Common Influences on the Post Test Score.

Dependent Variable = *PostScore*, n=232.

Independent variable	(1) Coeff (t)	(2) Coeff (t)
Constant	-10.46 (0.83)	-8.92 (1.87)
Post Day	-1.034 (1.03)	-0.392 (0.45)
Post Size	-0.062 (0.73)	-0.051 (0.45)
Male	2.176 (2.60)	2.318 (3.28)
Ethnic White	0.468 (0.38)	0.211 (0.21)
Transfer	0.287 (0.97)	0.125 (0.15)
GPA	3.67 (2.81)	4.64 (5.45)
Principles Grade	0.801 (1.05)	--
Peer Score	0.776 (2.47)	0.647 (2.84)
Peer GPA Average	0.391 (2.81)	--
Teacher Experience	0.012 (0.29)	0.135 (0.39)
R Square (p value for F)	0.255 (0.000)	0.250 (0.000)

.

Table 7. Student Performance on Subgroups of TUCE Questions

Variable	NFL	Ext & PubG	S&D	MR&Elast.	Internatl
Constant	-0.474 (1.59)	0.225 (1.03)	-0.232 (0.99)	-0.945 (2.74)	-0.945 (2.67)
Post Day	0.013 (0.26)	-0.043 (1.11)	-0,024 (0.58)	0.027 (0.63)	0.042 (0.66)
Post Size	0.004 (1.03)	-0.006 (1.95)	-0.005 (1.56)	0.001 (0.34)	0.003 (0.70)
Male	0.073 (1.68)	0.048 (1.48)	0.049 (1.42)	0.083 (2.28)	0.109 (2.09)
Ethnic White	0.003 (0.05)	-0.088 (2.01)	0.522 (1.10)	0.001 (0.03)	0.070 (0.98)
Transfer	-0.008 (0.17)	-0.051 (1.34)	0.029 (0.73)	-0.044 (1.03)	0.074 (1.23)
GPA	0.148 (2.81)	0.146 (3.73)	0.115 (2.77)	0.096 (2.18)	0.182 (2.89)
Peer Score	0.022 (1.55)	-0.003 (0.32)	0.034 (3.08)	0.005 (0.47)	0.037 (2.22)
Teacher Exper.	0.001 (0.12)	0.004 (2.97)	-0.002 (1.03)	0.002 (1.00)	0.001 (0.31)
R Square (p value for F)	0.089 (0.084)	0.195 (0.000)	0.137 (0.005)	0.087 (0.090)	0.129 (0.008)
Mean % Correct	46.7%	41.5%	40.1%	35.6%	38.7%

Notes: The dependent variables are the student's percentage correct answers to the following question groups (number of questions in parentheses): NFL, "no free lunch", questions employing opportunity cost concepts (4); S&D, supply and demand questions (6); Ext & PG, "externalities and public goods" questions (5); MR & Elast, "marginals and elasticity questions (5); Internatl, international economics questions (3).

Table 8. ACT Scores as an Alternative Measure of Human Capital

Variable	(1)	(2)	(3)	(4)
Peer Score	0.647 (2.43)	0.640 (2.17)	0.623 (2.06)	0.753 (2.53)
GPA	4.490 (5.81)	2.455 (2.08)	3.220 (2.89)	3.125 (2.67)
ACT Composite	---	0.448 (3.20)	---	---
ACT Reading	---	---	0.292 (2.77)	---
ACT Math	---	---	---	0.276 (2.72)
R Square (p value)	0.250 (0.000)	0.350 (0.000)	0.332 (0.000)	0.313 (0.000)

Table 9. Exploring the Impact of Teacher Characteristics on the Cohort Effect
.

Variable	(1)	(2)	(3)	(4)	(5)
Peer Score	0.766 (3.98)	0.751 (3.55)	0.769 (3.63)	0.435 (1.56)	0.755 (3.65)
Articles	---	-0.011 (0.19)	---	--	--
YearsPhD	---	---	-0.014 (0.39)	--	---
Full Time	---	---	--	2.792 (1.53)	---
Rank	---	---	--	---	4.466 (0.18)
R Squared (p value)	0.221 (0.000)	0.249 (0.000)	0.250 (0.000)	0.261 (0.000)	0.249 (0.000)

Note: The basic regressor set was included in each of these equations (see Equation (1) in Table 6. The teacher characteristics are: 1) Articles measures research experience as the number of articles teacher published which are indexed in EconLit; 2) YearsPhD measures the number of years since attaining the PhD; 3) Rank is the reciprocal of the most recent rank of the teacher's graduate school; and 4) Full Time equals one if the teacher was a full time faculty member, zero for adjunct faculty. N=232.

Table 10. Regression with Fixed Effects for Teachers is Contrasted with the Previous Results

Variable	(1)	(2)
Peer Scores	0.729 (3.97)	-0.455 (1.07)
GPA	3.132 (4.76)	3.221 (4.80)
Fixed Effects Entered?	NO	YES
R Square (p value for F)	0.152 (0.000)	0.196 (0.000)

Table 11. The Enlarged Sample Restores the Significance of the Peer Score

Variable	(1)	(2)	(3)
Peer Effects	0.014 (2.96)	0.015 (2.92)	0.009 (1.90)
GPA	---	-0.001 (0.48)	0.001 (0.16)
Teacher Fixed Effects Included?	NO	NO	YES
R Square (p value)	0.021 (0.000)	0.022 (0.011)	0.102 (0.000)
N of Cases	393	393	393

Table 12. Peer Effect Estimates When Students Are Grouped by Teacher, T-i.

Variable	T-A	T-B	T-C	T-D	T-E
Constant	-6.563	4.583	3.618	128.14	-5.428
Peer Score	0.137 (0.33)	-0.144 (0.26)	-4.590 (2.97)	-8.839 (5.39)	0.369 (1.10)
GPA	5.471 (2.99)	3.444 (3.17)	-0.001 (0.01)	2.806 (3.09)	4.035 (3.83)
R Squared (p value)	0.1642 (0.0135)	0.0820 (0.0079)	0.126 (0.0161)	0.414 (0.000)	0.176 (0.001)
N of Cases	51	116	64	56	75