Research in Economic Education

In this section, the *Journal of Economic Education* publishes original theoretical and empirical studies of economic education dealing with the analysis and evaluation of teaching methods, learning, attitudes and interests, materials, or processes.

PETER KENNEDY, Section Editor

Choosing a Proxy for Academic Aptitude

Wayne A. Grove, Tim Wasserman, and Andrew Grodner

Abstract: Although academic ability is the most important explanatory variable in studies of student learning, researchers control for it with a wide array and combinations of proxies. The authors investigated how the proxy choice affects estimates of undergraduate student learning by testing over 150 specifications of a single model, each including a different combination of 11 scholastic aptitude measures—high school grade point average (GPA) and rank and variants of college GPA and Scholastic Aptitude Test (SAT) scores. Proxy choices alone cause the magnitude of the estimated learning gains to vary by large and meaningful amounts, with increases ranging from a C+ to less than a B− or to a B. The authors found that collegiate GPA data offer the best proxy for students' individual propensities to learn economics—a result that runs counter to researchers' actual proxy choices. The results suggest that scholars should control for academic aptitude with college grades and either SAT scores or high school GPA or rank.

Key words: methodology, proxy variables, student learning, undergraduate economics
JEL codes: A22, C81, I21

***Wayne A. Grove** is an associate professor of economics at Le Moyne College (e-mail: grovewa@lemoyne.edu). **Tim Wasserman** is assistant director, Institutional Research and Data Management, Center for Support of Teaching and Learning, Syracuse University. **Andrew Grodner** is an assistant professor of economics at East Carolina University.* *The authors thank three anonymous reviewers, Peter Kennedy, William Becker, Dan Black, Kevin Rask, Paul Blackley, and participants at the 2002 ASSA meetings in Atlanta and the seminars at Colgate University and Le Moyne College. They are grateful to the administrators at Syracuse University for their cooperation and to Le Moyne College for a Research and Development grant.*

A perennial problem in economic research is the lack of good consistent assessments of individual ability or of access to such data. However, scholars of higher education investigate enterprises that maintain a rich database of information for a variety of internal and external reporting purposes that contain a uniform set of data used to screen applicants and then continually to assess matriculants' academic performance. Despite this, and even though academic aptitude variables are the only consistently significant and meaningful explanatory variables of undergraduate student learning of economics (Becker 1997), scholars use a wide array of control measures for academic performance and seldom explain the logic of their proxy choices or how alternative ability measures affect their estimates of cognitive achievement.[1] According to a 1998 survey of selected studies of college student learning of economics that we conducted, in researchers proxy for academic aptitude most frequently by using college entrance exam scores (70 percent), then with TUCE (Test of Understanding College Economics) scores (50 percent), followed by college grade point average (GPA) (over 40 percent), and finally with high school measures (GPA and rank, 10 percent each). Beside surveying researchers' proxy choices, we sought to answer two questions. First, does the choice of scholastic ability proxies matter for estimates of student learning, and if so, how? Second, what prescriptive guidance can be offered regarding the best proxy choice for studies of undergraduate student learning?

To determine the effect of the proxy choice on estimates of student learning, we used data from a natural experiment in which an instructional technique was in effect for one group of students but not for the other.[2] Using a single data set, we tested over 150 specifications of the same model, each containing a different combination of 11 scholastic aptitude measures—high school GPA and rank and variants of college GPA and Scholastic Aptitude Test (SAT) scores.[3] Our results indicated that the proxy choice alone causes the magnitude of estimated learning gains to fluctuate by meaningfully different amounts in the context of college grades, ranging from a 60 percent chance of a third of a letter grade increase (from a C+ to a B-) to two-thirds of a letter grade gain (from a C+ to a B). These results probably underestimate the true range of estimated learning gains because our analysis excluded ACT and TUCE scores, both of which were used by over 60 percent of the studies in our survey selection but have been criticized as poor proxies of student ability to learn economics.[4]

To identify the best proxies for student academic aptitude, we first extracted an estimate of latent academic ability from our rich set of proxy measures using factor analysis. The best proxies, then, are those most highly correlated with estimated latent aptitude. Recent theoretical work by Lubotsky and Wittenberg (forthcoming) established the proxy-choice criteria for accurately estimating treatment effects: The researcher should include additional proxies that are most highly correlated with the latent variable, but not with each other, to avoid correlated measurement errors. Factor analysis indicates that collegiate GPA is the superior proxy for individual students' propensities to learn economics. In contrast, researchers have relied most heavily upon SAT and ACT scores. Our results suggested that scholars of student learning studies, with single-institution data sets, should control for academic aptitude with collegiate GPA and then either with SAT scores or high school GPA or rank. We know of no other student learning study in which the researchers evaluate the effect of the proxy-choice decision and attempt to identify which variables best control for the unobserved, latent attribute.

SURVEY SELECTION OF BEST PRACTICE APTITUDE PROXY CHOICE

To obtain a profile of the best practice use of academic ability control variables, we selected the 37 published articles cited in *Research on Teaching College Economics* by Siegfried and Walstad (1998, 147–58) in which researchers empirically estimate college undergraduate student learning of economics.[5] In 60 percent of the studies we surveyed, the authors used a unique set of student academic ability control variables, despite the fact that scholars planned many of those studies and that the typical institution of higher education maintains all of these student records (except for TUCE scores). Although authors of one article (Gohmann and Spector 1989) argued against including any aptitude proxies at all, researchers in our survey selection used college entrance exam scores most frequently (70 percent), TUCE in half the articles, college GPA in over 40 percent, and high school measures in 20 percent (high school GPA and rank equally).

The heterogeneous assortment of scholastic aptitude controls used by scholars appeared to result from a combination of institutional constraints,[6] international differences in educational practices,[7] and authors excluding available data.[8] Scholars measured success in high school in many ways, namely by cumulative GPA (Borg and Shapiro 1996), senior year GPA (Reid 1983), and class rank (Fizel and Johnson 1986; Schmidt 1983), but most often did not specify how they calculated college grades.[9] College entrance exam scores appear in our survey selection in six forms with the following frequencies: (*a*) ACT scores, 10, (*b*) both the individual math- and verbal-SAT scores (MSAT and VSAT), 8, (*c*) composite SAT (TSAT) scores 7, (*d*) VSAT scores only, 1, (*e*) ACT rank, 1, and (*f*) dummy variables for students with above average ACT or SAT scores, 1.[10]

A NATURAL EXPERIMENT

During the fall of 1998, 239 undergraduate students enrolled in and completed four sections of introductory microeconomics taught by Grove at Syracuse University, a large, private residential university in the northeast (Carnegie Classification: Doctoral Research Universities II—Extensive). A “natural experiment” separated students into one group (of 143) whose course grades were based on problem-set performance and another group (of 96) whose course grades were not.[11] Students with graded problem sets received a course grade that included the average of the best four of five possible problem-set grades. All lectures, handouts, exams, and review sessions were as identical as possible. All students received the problem sets at the same time and encouraged to practice economics by solving them as important preparation for the exams. When the problem sets were handed in, students in both groups received the answer keys. We measured mastery of the course material by performance on three required exams of equal value, each of which we assessed with a 0-to-100-point scale. To ensure as much uniformity in grading as possible, the same grader evaluated each exam question for all 239 students. Thus, the most discernible difference between the four sections of introductory microeconomics was that students in three sections had a direct grade-based incentive to practice economics problems throughout the semester (the treatment group), whereas those in the control group received neither reward nor penalty for completing problem sets.[12]

All the data used in this study came from faculty or university records, not from student surveys, which have been shown to overstate actual performance.[13] Because of missing SAT scores and high school data, a common set of academic aptitude variables existed for 117 students, of whom 71 were in the experimental group and 46 in the control group. We provide descriptive statistics (means and standard deviations) for each variable used for each group in Table 1. Mean exam scores and college GPA measures were not equivalent in the experimental and control groups (Table 1). We had three types of academic aptitude variables: college GPA, SAT scores, and high school measures. According to the correlation matrix we present in Table 2, the collegiate GPA measures were highly correlated with each other (.72 to .96) as were high school GPA and rank (.7), but each of

TABLE 1. Descriptive Statistics: Means, Standard Deviations, and *p* Values for *t* Tests and Chi Square ($N = 117$)

Dependent and independent variable	Graded group (N = 71) *Mean*	Graded group (N = 71) *SD*	Nongraded group (N = 46) *Mean*	Nongraded group (N = 46) *SD*	*p* values[a]
Mean exam score	83.0	8.12	77.0	8.94	.000[b]
SemGPA	3.12	0.57	2.78	0.64	.003[b]
SemGPA–ECN	3.14	0.64	2.79	0.78	.009[b]
*z*SemGPA–ECN	0.20	0.63	−0.10	0.84	.023[b]
CumGPA	3.21	0.49	3.05	0.47	.087
CumGPA–ECN	3.18	0.55	2.99	0.53	.056
*z*CumGPA–ECN	0.23	0.55	0.04	0.64	.091
MSAT	590.1	81.4	579.8	70.2	.480
VSAT	567.2	81.5	559.3	66.1	.586
SATsum	1157.3	143.5	1139.1	114.1	.470
HS rank	75.3	14.90	72.0	14.63	.238
HSGPA	3.40	0.43	3.35	0.37	.570
White	0.79	0.41	0.80	0.40	.838[c]
Men	0.63	0.49	0.65	0.48	.840[c]
Freshman	0.35	0.48	0.22	0.42	.102[c]

Note: Concurrent SemGPA = semester grade point average; ECN = economics; Cum = cumulative; MSAT = math score on Scholastic Aptitude Test; VSAT = verbal score on SAT; SATsum = composite score on SAT; HS = high school; HSGPA = high school grade point average.

[a]values for continuous variables from *t* tests and chi squares for dichotomous variables.
[b]Means differ at the .05 Type I error level.
[c]Chi square analysis.

TABLE 2. Correlation Matrix (N = 117)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Mean Exam Score														
*z*Men	.18													
White	.15	.06												
Freshman	.16	.14	−.04											
SemGPA	.76	.10	.19	.20										
SemGPA–ECN	.65	.11	.27	.18	.96									
*z*SemGPA–ECN	.66	.09	.16	.16	.92	.91								
CumGPA	.60	.01	.24	.13	.76	.75	.72							
CumGPA–ECN	.60	.05	.24	.19	.86	.87	.84	.94						
*z*CumGPA–ECN	.65	.08	.19	.15	.86	.86	.92	.88	.94					
SATsum	.45	.22	.29	.20	.40	.36	.33	.48	.43	.42				
MSAT	.48	.33	.30	.17	.40	.38	.34	.44	.41	.40	.87			
VSAT	.30	.05	.19	.18	.29	.24	.24	.39	.33	.32	.86	.51		
HS rank	.35	−.01	.13	.09	.35	.34	.34	.38	.35	.38	.41	.44	.28	
HSGPA	.44	.08	.14	.15	.36	.34	.37	.36	.34	.39	.39	.43	.24	.70

Note: Concurrent SemGPA = semester grade point average; ECN = economics; Cum = cumulative; SATsum = composite score Scholastic Aptitude test; MSAT = math score on SAT; VSAT = verbal score on SAT, HS = high school; HSGPA = high school GPA.

those types was not strongly correlated with variables in the other groups (less than .45). MSAT and VSAT verbal scores had a much lower correlation (0.51) with each other and even less with other aptitude variables.

ACADEMIC APTITUDE PROXIES

College entrance examination scores, typically thought to measure raw intelligence, a stock of knowledge, or a general aptitude for learning, have the virtue of being uniform and methodical but, when used to control for aptitude in college courses, have the disadvantage of providing a measure at a point in time in the past.[14] We used three SAT scores: MSAT, VSAT, and SATsum. Unfortunately, virtually none of the students in our sample reported ACT scores.

GPA, an indicator of academic success, directly measures success in course work and draws on the attributes associated with the SAT, but performance in college coursework also reflects the application, throughout each academic term, of good study skills, motivation, organization, industriousness, perseverance, and consistency.[15] We used two temporal measures of collegiate GPA: grades earned during the semester under study and cumulative GPA including the semester under study.[16] Same semester GPA, a measure of student academic success during the semester being studied, included information about positive or negative shocks that may have affected a student's potential scholastic achievement.[17] From a practical perspective, much of the economic education research has addressed the principles of economics course even though during the fall term freshman have no prior college cumulative GPA. If the dependent variable in learning studies is the course grade or is highly correlated with it, the appropriate GPA measure would be the concurrent semester's GPA minus the economics grade (SemGPA–ECN) or the cumulative GPA minus the economics grade (CumGPA–ECN).[18] We included in our analysis the same semester GPA with the economics grade (SemGPA) and the cumulative GPA with the economics grade (CumGPA) *exclusively for purposes of replication* with usage in the literature—unequivocally, SemGPA and CumGPA *should not* be considered for inclusion in a study of student learning because those "independent" variables contain some or all of the dependent variables.[19]

Academic success as measured by grades is obscured by students' heterogeneous mix of courses and the variability of grade distributions by professors, courses, and departments. Thus, a B+ might represent a low grade in one course but a top score in another.[20] To improve the comparison of between-class grades, we created z score GPA measures that calculated a student's grade deviation from the distribution mean for each course.[21] Construction of such "standardized GPA" data required access to the transcripts of every student enrolled in each course taken by a member of the sample group.[22]

Some researchers and college admission committees have expressed skepticism about using cumulative high school GPA to measure cognitive ability because of long time lags and large variations in the standards and quality of schools, school districts, and states (Ferber, Birnbaum, and Green 1983, 36; Chaker 2003, D2). Georgetown University and Haverford College, for example, use high school rank, rather than GPA, in their admissions' decisions (Chaker 2003).

OUR BASIC MODEL OF STUDENT LEARNING

We hypothesized that basing course grades on regular problem-set performance throughout the semester improved student's cognitive achievement, controlling for their academic ability and demographic characteristics.[23] We compared the distribution of gender, race (white vs. other), and class standing (freshman vs. other) for the experimental and control group via chi-square analysis to check for inequalities between the two groups that might call into question their presumed equivalence. The distributions by group were roughly equal for gender, ($\chi^2 = 0.0409, p > .84$) and race/ethnicity ($\chi^2_1 = 0.0417, p > .84$) but not for class standing ($\chi^2 = 2.9017, p < .10$).[24] Consequently, to avoid estimation bias we chose to include in our model a dichotomous, independent variable for academic class standing (1 if freshman; 0 if not) alongside the treatment dummy and the ability measures. To test whether graded problem sets improved student learning, we use ordinary least squares (OLS) to regress mean exam grades in Economics 101 on freshman status, membership in the control (nongraded) group and treatment (graded) group, and then, in succession, various scholastic aptitude proxy(ies), as shown in the following model:

$$MeanExamScore = \alpha_0 + \alpha_1 Freshman + \alpha_2 GradedGroup + \sum_{j=1}^{k} \alpha_j AcademicAptitudeProxy(ies)_j + \in . \quad (1)$$

We provide sample regression results for five specifications of this model in Table 3. Because the dependent variable is exam performance on a 100-point scale, the coefficient on the graded problem-set variable (when controlling for academic ability with *z*SemGPA–ECN [second column] indicated that students in the treatment group improved their performance by 3.74 points [i.e., by more than a third of a letter grade]). The results of the other four specifications of the model, each using a different aptitude proxy, revealed estimated learning gains ranging from 3.38 points to 5.5 points (Table 3). Freshman status was not significant.

DOES THE PROXY CHOICE AFFECT ESTIMATED TREATMENT EFFECTS?

We estimated our basic student learning model alternatively with each of the 11 academic aptitude proxies alone and then with all sensible combinations of them, for a total of over 150 different regressions.[25] In all regressions, the estimated treatment effect and the aptitude proxy coefficients were positive, meaningfully large, and statistically significant at the 1 or 5 percent levels.

We summarized 100 of the estimated treatment effect coefficients (Figure 1), using a legend to indicate which variables were included as aptitude controls in

TABLE 3. Estimated Learning Gains for the Student Learning Model with the Four Most Common Individual Aptitude Proxies

Independent variable	SemGPA–ECN	*z*SemGPA–ECN	CumGPA–ECN	SATsum	HSGPA
GradedProblem SetGroup	3.38*	3.74**	4.22**	5.47**	5.53**
	(2.60)	(2.96)	(3.16)	(3.77)	(3.80)
Academic aptitude proxy (see column heading)	7.57**	7.50**	9.16**	0.03**	9.07**
	(8.40)	(8.85)	(7.57)	(5.29)	(5.18)
Freshman	0.40	0.52	0.36	0.54	1.01
	(0.29)	(0.70)	(0.25)	(0.34)	(0.64)
Adjusted R^2	0.45	0.47	0.4	0.28	0.27

Note: t statistics are in parentheses. Dependent variable: mean exam score (in points); (N = 117). SemGPA–ECN = concurrent semester grade point average excluding economics; Cum = cumulative; SAT = Scholastic Aptitude Test; HSGPA = high school GPA.
*Mean is different from zero at the .05 Type I error level. **Mean is different from zero at the .01 Type I error level.

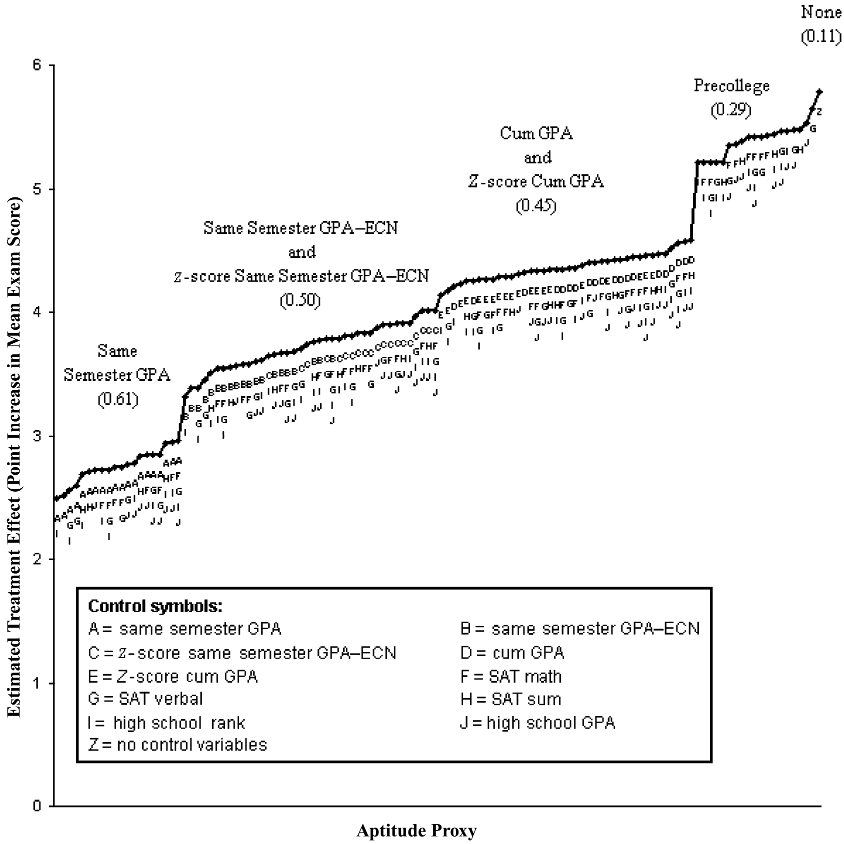

FIGURE 1. Estimated treatment effects according to the aptitude proxy(ies) included, dependent variable: mean exam scores (in points); N = 117. GPA–ECN = grade point average excluding economics; Cum = Cumulative; SAT = Scholastic Aptitude Test.

each specification, which clearly illustrated how the proxy choice caused the size of estimated treatment effects to fluctuate. The lowest estimated learning gain of 2.5 points ($R^2 > .5$) implied that 60 percent of the experimental-group students would have received a third of a letter grade increase for the course (e.g., C+ to B−), whereas the highest estimated coefficient of 5.8 points ($R^2 = .11$) indicated that 90 percent of the students would have experienced a two-thirds of a letter grade increase (e.g., C+ to B). The scholastic-ability proxy choice caused the magnitude of the estimated student learning gains to vary by more than 130 percent—enough for most people to acknowledge it as a meaningfully large difference in the context of students' grades. Nonetheless, these estimated treatment effects, as a function of the controls used and the small sample size, were not statistically significantly different, in the sense that 95 percent confidence intervals created around the estimated treatment effects overlapped.[26]

The results, summarized in Figure 1, suggest two more conclusions about the proxy-choice decision. First, collegiate GPAs are the most important aptitude variables. The stair-step pattern of Figure 1 reflects the clustering of estimated learning gains into five groups, resembling a flight of slightly upward-sloping stairs with each step determined by the choice of GPA variable used (or its omission in the case of the top range of estimates).[27] Second, and a corollary to the primacy of college GPA, precollege proxies complement collegiate GPA variables and substitute for each other. When combined with GPA measures, college entrance exam scores and high school aptitude measures reduce the possibility that observed relationships are chance occurrences (increases the *t* statistic) and increase the amount of the variation in students' exam scores that can be explained (increases the adjusted R^2).[28] The highest estimated learning gains (of 5.8 points) which have the lowest R^2 (of 11 percent) resulted from a model without any aptitude variables, as advocated by the authors of one of the articles in our survey selection (Gohmann and Spector 1989, 238).

WHICH SHOULD A SCHOLAR CHOOSE?

Because we found that the proxy choice meaningfully influenced the estimated treatment effect, which proxies should a scholar use to control for student academic aptitude? Our strategy was to use factor analysis to extract students' underlying academic aptitude from the proxies and then to determine the correlations between the estimated latent ability index and each individual proxy.[29] Lubotsky and Wittenberg (forthcoming) provided the proxy-choice criteria for deciding which academic ability control variables scholars should include in a regression of student learning: Scholars should include proxies that are most highly correlated with the estimated latent variable but not with each other, to avoid correlated measurement errors. We offer a brief, intuitive characterization of factor analysis that is typically employed to reduce the number of variables or to detect a structural relationship between variables (Johnson and Wichern 1992, chap. 8–9).[30]

Thinking of the correlations of the multiple control variables as defining a space, we could fit a plane through the data that represented the best summary of the relationship between the variables. The index variable that approximated the regression plane in such a plot would capture the essence of the items, essentially reducing many variables to one factor. Each proxy probably measured both a unique aspect of academic aptitude not quantified by the other measures and an underlying academic aptitude factor that the proxies shared in common. Factor analysis uses only the variability in a variable it has in common with the other proxies, allowing researchers to interpret retrieved latent factors as independent, source variables, rather than as a weighted average of the proxy variables.[31]

The first factor (i.e., the plane fitted through the space occupied by the proxies' correlations) explained the most variance among the control variables, in our case over 79 percent of the common variance.[32] The first factor, which we assumed was latent academic aptitude, was more highly correlated with all proxies than any successive factor.[33] The correlations between the individual academic aptitude proxies and the generated latent aptitude index are displayed in

TABLE 4. Correlations Between Academic Aptitude Proxies and the Unrotated First Factor (N = 117)

Proxy	First factor
zCumGPA–ECN	.96
zSemGPA–ECN	.91
SemGPA–ECN	.89
CumGPA–ECN	.88
HSGPA	.56
HS Rank	.56
MSAT	.55
VSAT	.42

Note: Factor analysis fits a plane through the space occupied by the academic aptitude proxies, retrieving the variability those measures share in common. The first factor, presumably latent academic aptitude, explains the most variance among the control variables (in our case over 79 percent of the common variance). This table shows the correlations between the individual academic aptitude proxies and the first analytic factor (for 8 proxy variables rather than all 11 since SemPGA and zSemGPA contain information in the dependent variable and SAT sum scores constitute a linear combination of MSAT and VSAT scores). SemGPA–ECN = concurrent semester grade point average excluding economics; Cum = cumulative; MSAT = Math Scholastic Aptitude Test; VSAT = verbal Scholastic Aptitude Test.

Table 4. The latent variables were extracted from 8 proxy variables rather than all 11—we excluded SemGPA and zSemGPA, which contain information in the dependent variable (and were only included to replicate the literature), and TSAT scores, a linear combination of MSAT and VSAT scores. Fisher's z tests (of whether pairs of proxies' correlations with the latent index were no different) established two conclusions suggested by Figure 1. First, collegiate GPA had a significantly stronger correlation with the latent index than did any of the other proxies (at the 0.01 Type I error level)[34]; hence, college grades determine the basic level of the treatment effects (i.e., the stair-step pattern in Figure 1). Second, we could not reject the null hypothesis that any of the four precollege control variables were differently correlated with the latent variable.[35] Thus, the scholastic aptitude proxy variables constituted two clusters of proxies: collegiate grades which were highly correlated (.88 to .96) and precollege measures, namely high school GPA and rank and MSAT and VSAT scores, which were less correlated (.42 to .56).

Our results yield one definitive conclusion: A scholar interested in controlling for academic ability in a study of student learning with a single-institution data set should include collegiate GPA. Because scholars should use only one of the highly correlated college GPA variables (Table 2), which collegiate GPA measure should be used? Although ideally researchers would use zCumGPA–ECN or zSemGPA–ECN, constructing z score GPAs will prove too daunting a task for many scholars.[36] Because SemGPA–ECN and CumGPA–ECN are not statistically different from each other (or from zSemGPA–ECN), SemGPA–ECN seems like the best choice for scholars without open access to institutional student databases.[37]

Although the four precollege control variables' correlations with the latent variable are not statistically different, it is probably prudent to include just one of them because using more alters estimated coefficients very little (Figure 1). The SATsum score was excluded from our analysis in Table 4, but could also be used. Thus, a reasonable summary of our empirical results is for researchers of undergraduate student learning, with a single institution data set, to proxy for academic aptitude with collegiate GPA and one of the precollege measures: SAT score or high school GPA or rank.

CONCLUSIONS AND IMPLICATIONS

In a study of student learning with a natural experiment data set, we found that merely altering the academic aptitude proxies caused the magnitude of treatment effects to fluctuate substantially: with estimated learning gains ranging from 60 percent of students experiencing an increase from, for example, a C+ to a B− to an increase of a C+ to a B. Such outcomes constituted a meaningful distinction in the context of student learning, although they were not statistically significantly different from each other.[38] To avoid correlated measurement errors (Lubotsky and Wittenberg forthcoming), our proxy choice criteria suggested including proxies that were most highly correlated with the latent ability variable but not with each other. We used factor-analysis estimation to extract an estimate of students' latent academic aptitude from our rich set of scholastic ability data. Collegiate GPA is the best proxy for students' academic ability (Table 4).

Our results suggest that investigators in single-institution studies of student learning should control for academic aptitude with college GPA and then with either SAT scores or high school grades or rank. That set of proxies reduces the variance of estimated treatment effects displayed in Figure 1 from a range in excess of 3 points to less than 1. In contrast to these conclusions, scholars in most of the 35 single-institution studies of student learning from our "best practice" survey selection used standardized entrance exam scores rather than collegiate grades, a quarter of them relied exclusively on precollege measures. If, as we found, collegiate GPA proves to be an essential control variable for an individual college or university study but, if as Kennedy and Siegfried (1997) report, it is not for multischool inquiries, researchers will need to understand this distinction.[39] Of the two multiple-institution studies in our survey, one controlled for ability with SAT scores (Kennedy and Siegfried) and the other with college grades (Lopus and Maxwell 1995).[40]

Because our data set is unique (with an experimental and a control group) and rich (11 measures of academic aptitude from institutional records) but also limited—a small sample ($N = 117$) from a single institution, our results should be taken as suggestive. Because we could not, for lack of data, evaluate the role of ACT and TUCE scores as aptitude proxies, both of which have been criticized as poor scholastic ability control variables [ACT by Kennedy and Siegfried (1997, 388); the TUCE by Becker 1997, 1363–67 (1990)], our results probably underestimated the actual proxy choice-induced variability in estimated treatment effects in the literature.[41] Regarding the generalizability of our findings for the economics education literature, it would be useful to learn how researchers' proxy choices influence estimated treatment effects for other factors that faculty or administrators influence.[42] Of particular interest is whether college GPA can function as a control measure in a multischool study given the variability of cross-institution academic standards.[43] We hope this study will encourage scholars to select scholastic control variables more systematically and transparently and to report the effects of proxy choices for empirical findings.

NOTES

1. For an exception, see Kennedy and Siegfried (1997, 388) who argue for using the composite SAT score (SATsum) instead of math and verbal SAT (MSAT and VSAT) scores separately, ACT scores, or grade point averaged. Note that 35 of the 37 papers in our survey selection were conducted with data from a single school so that researchers may have been able to obtain the aptitude variables they wished, unlike the case with scholars using national data sets.
2. For an example of studies based on a designed experiment (rather than an unplanned experiment that naturally occurred as is the case with this study), see Becker and Salemi (1977) and Salemi and Tauchen (1982, 1987).
3. We did not attempt to replicate the results of published articles as, for example, Dewald, Thursby, and Anderson (1986) did during a two-year period in the early 1980s (with discouraging results).
4. The TUCE, an exam of 30 multiple-choice questions, or 33 with the international questions, assesses students' net gain in economic knowledge by comparing the performance on an exam at the beginning and again at the end of a course. Although Kennedy and Siegfried (1997, 388) criticized the ACT, Ballard and Johnson (2004) found ACT math scores to be a statistically significant determinant of students' grades and exam or both scores in an introductory microeconomics class.
5. The articles included in our survey selection are listed in the appendix. Regarding survey selections, see McCloskey and Ziliak (1996).
6. Institutional constraints comprise both restricted access to student records and the lack of data collection, namely variations in the types of high school data collected and whether schools accept SAT or ACT scores. For example, Chizmar and Ostrosky (1998) excluded ACT data because it was not collected from transfer students.
7. Instead of college entrance exam scores, for example, Canadian scholars used performances in grade 11, 12, or 13 math and English courses (Anderson, Benjamin, and Fuss 1994; Myatt and Waddell 1990). In a similar way, Bauer and Zimmermann (1998) used high school grades for a study of learning of economics in a German university.
8. Paul (1982), for example, used SAT scores to verify a similar distribution of student characteristics each semester over a four-year period but omitted that data from his empirical estimates of student learning.
9. Collegiate GPA might measure performance either in the semester being studied or cumulatively and either include or exclude the economics course grade. Cardell et al. (1996, 458) and Caudill and Gropper (1991, 305) report used the "most recent GPA," Brasfield, Harrison, and McCoy (1993, 102) used "current college GPA," and Ferber, Birnbaum, and Green (1983, 30–31) used the "university GPA." Bonello, Schwartz, and Davisson (1984, 206) used end of the semester cumulative GPA minus cumulative GPA "as projected by the University's admissions office." Researchers tend to use cumulative GPA.
10. Another obvious choice is MSAT unaccompanied by verbal scores (Jensen and Owen 2000). Charkins, O'Toole, and Wetzel, (1985) used VSAT, Park and Kerr (1990) used ACT percentile rank, and Borg, Mason, and Shapiro (1989) used dummy variables for above average SAT and ACT scores.
11. This occurred because of a professor's illness at the beginning of the semester. Grove agreed to teach his section for a few class periods. At the end of the second week of classes, it became apparent that the ill professor could not return to the classroom. Viewing the syllabus as a professor-student "contract" about the mutual expectations and responsibilities of a course, the newly assigned professor thought it inappropriate to impose new requirements so late in the semester. Hence, a natural experiment occurred, creating a control and a treatment group. For more details and a discussion of sample selection bias, see Grove and Wasserman (2006). In contrast to our natural experiment, Becker and Salemi (1977) and Salemi and Tauchen (1982, 1987) used a data set from an experiment with planned control and treatment groups.

12. In addition, some sections met on Mondays and Wednesdays and others on Tuesdays and Thursdays. Class times varied between 9:30 a.m. and 3 p.m.
13. Scholars survey students (*a*) to obtain information either because data is not available from administrators or not collected by them or (*b*) to ease the data collection process in multischool studies. Whereas Maxwell and Lopus (1995) labeled students' systematic overestimation of self-reported data as the "Lake Wobegon effect," so-named for Garrion Keillor's fictional town where "all children are above average," Ballard and Johnson (2004) reported that the small amount of overstatement involved did not bias regression results.
14. Rothstein (2004) argued that the SAT is a proxy for individual and high school demographic characteristics.
15. Jencks (1979) demonstrated that, net of background, formal schooling, and cognitive skills, personal traits such as industriousness, perseverance, and leadership have noteworthy associations with earnings and occupational status. With similar controls and housework time, Dunifon, Duncan, and Brooks-Gunn (2001) established that a "clean-home measure" is predictive of one's own and one's children's earnings 25 years later and of children's schooling. On this point, note that the only student with a perfect MSAT score from our full sample, not the subsample of 117 students used for this study, failed to hand in the required four problem sets.
16. Caudill and Gropper (1991, 305) viewed the prior semester's cumulative GPA as "probably a better measure than the GPA at the time of the course because the former measures the student's performance over a longer time period." In the fall term, freshman have no cumulative GPA, whereas sophomores, juniors, and seniors have two, four, or six previous semesters of grades, respectively, so that for freshman the semester and cumulative grades are the same. We obtained similar results with either a nonfreshman sample ($N = 82$) that included a "prior cumulative GPA" measure or with a freshman-only sample ($N = 35$).
17. The types of negative shocks that we hear about from students include prolonged or serious illnesses, severe personal problems, or grave family crises. A longitudinal analysis of the Syracuse University class of 2001, comprising 2,552 students enrolled for up to eight semesters, revealed the following deviations from students' four-year cumulative GPA: 10 percent of students experienced their greatest GPA deviation as -1.0 or more, 13 percent as -1.0 to -0.667, 27 percent as -0.667 to -0.334, 13 percent as -0.333 to 0, 11 percent as 0 to 0.333 above their mean GPA, 18 percent as $0.334-0.667$ above, 7 percent as $0.667-1.0$ above, and 2 percent as 1.0 or more above. For more about GPA over the academic life-cycle, see Grove and Wasserman (2004).
18. For the correlations, see Table 2. Evensky suggested excluding the economics grade from the GPA, a practice also found in Evensky et al. (1997) and Chizmar and Ostrosky (1998).
19. Note that our dependent variable was mean exam scores, not course grade.
20. Some of the grade inflation literature has shown higher average grades or grade compression in particular subject areas, rather than across the board (Sabot and Wakeman-Linn 1991).
21. The *z*-scores are calculated as the difference between the raw course grade and the sample mean course grade divided by the standard deviation of the course grades. We thank Kevin Rask for this suggestion.
22. Although database systems dramatically reduce the cost to investigators of calculating *z*-score GPAs, doing so remains an onerous task, especially, for example, when calculating *z*CumGPA–ECN for seniors, which requires knowledge of every grade given in each of the dozens of courses taken during the time a student was matriculated at the school.
23. Student learning is typically modeled as a production function with exam performance resulting from student human capital inputs, demographic characteristics, student effort, and treatment effects (Becker 1997).
24. The race and gender coefficients were small and had large *p* values.
25. By "sensible" we mean that we did not include, for example, both SATsum with MSAT or both current SemGPA and CumGPA that included current semester grades.
26. The confidence intervals were the parameter estimate plus and minus twice the standard error. We attributed this lack of statistical significance to the large standard errors associated with a relatively small sample size.
27. Whereas the choice of precollege variables generated a 10 to 20 percent range of estimated learning gains, the collegiate GPA choice caused estimated treatment effects to vary by over 100 percent.
28. Consider, for example, SemGPA–ECN: the estimated learning gain with the proxy by itself was 3.38 points (adjusted R^2 of .44), adding only high school rank lowered the coefficient by 2 percent (to 3.31; adjusted R^2 of .46), and adding instead SATsum and both high school variables raised the estimated learning gain by 12 percent (to 3.78; adjusted R^2 of .53).

29. We obtained almost identical results to those described later when we conducted the same investigation with principal components analysis. Using Lubotsky and Wittenberg's (forthcoming) newly proposed *post hoc* estimator strengthened the conclusion about the role of collegiate GPA and offered greater distinctions regarding the choice of precollege academic aptitude proxies. For those results, see Grove, Wasserman, and Grodner (2004).
30. For example, Forni and Reichlin (1998) found it helpful to form two factors from 450 variables, and Stock and Watson (2002) found it helpful to form three factors from around 150 disaggregated series. Also, see Boivin and Ng (2003).
31. Factor analysis adds a random component (an error term) to the deterministic part of the proxy variables so that factors can be decomposed according to the variability that proxies have in common with each other and that which is unique to each. See Johnson and Wichern, 1992, ch. 8–9.
32. Only the first factor had a corresponding eigenvalue larger than 1—the first factor was 4.8 and the second factor was less than one.
33. We did not rotate the factors (create variables as linear combinations of factors), a practice employed to identify the greatest correlation between each factor and groups of the proxies used to classify and interpret the factors (Fabrigar et al. 1999).
34. For example, the Fisher's z statistic comparing the correlation 0.96 (i.e., the factor analysis index and zCumGPA–ECN) and 0.55 (the factor analysis index and MSAT) was 10.02 with a p value of .00. Thus, we rejected the null and concluded that the two correlation coefficients were not equal.
35. For example, the Fisher's z statistic comparing the correlation 0.55 (i.e., the factor analysis index and MSAT scores) and 0.42 (the factor analysis index and VSAT) was 1.29 with a p value of 0.20. Thus, we accepted the null and concluded that the two correlation coefficients were not statistically different.
36. Using zCumGPA–ECN required obtaining the grades for every student in every class taken by a member of the study when matriculating at that institution, whereas zSemGPA–ECN would merely require the grades for every student in every class taken that semester by a member of the study.
37. Note that use of SemGPA–ECN generated roughly one grade point higher treatment effects than with CumGPA–ECN.
38. Ninety-five percent confidence intervals created around the estimated treatment effects overlapped because of the large standard errors associated with a relatively small sample size.
39. Kennedy and Siegfried (1997) argued for using only the SAT score because GPA was insignificant when both were used, and when individually used, the R^2 with SAT was .86 versus .59 with GPA (388).
40. The "multischool" studies used national datasets. Prince et al. (1981) drew data from two schools: James Madison University and Virginia Commonwealth University.
41. Thirty percent of our survey selection relied upon ACT scores and over half used TUCE data. We lacked TUCE scores because our data did not come from a planned study. Regarding the ACT, Ballard and Johnson (2004) found ACT math scores to be a statistically significant determinant of students' grades and exam scores in an introductory microeconomics class, although they did not offer a comparison with MSAT scores. Kennedy and Siegfried (1997, 388), in a multischool study, challenged the interchangeable use of ACT and SAT scores, reporting that estimated student learning with ACT had an R^2 of .12 versus .86 with SAT. Regarding TUCE scores, although originally intended to measure the net value added, typically the post-TUCE score was used as the dependent variable and the pre-TUCE as an independent variable or it was excluded. A vexing problem remained regarding the validity of the pre-TUCE scores as a measure of student knowledge of economics because grading it (unlike grading the post-TUCE) seemed out of the question. Without the pre-TUCE score (Kennedy and Siegfried 1997), the TUCE becomes a short standardized multiple-choice test that facilitates multischool studies rather than measuring the net gain in economics knowledge. Becker (1997) criticized its use as too narrow a gauge of learning (30 multiple-choice questions or 33 with the international questions included). In contrast with a short multiple-choice exam, the three tests used in our study contained a combination of problem-solving, short-answer, and multiple-choice questions, with the latter worth about 15 percent of each exam.
42. Replication will require identifying data sets that overcome sample selection problems by containing a control group to permit estimation of learning gains (Becker 1997, 1366).
43. In this context, we wondered whether z scores, as we used for a single university, can be used to overcome the problem of different grading norms between institutions of higher education.

REFERENCES

Anderson, G., D. Benjamin, and M. Fuss. 1994. The determinants of success in university economics courses. *Journal of Economic Education* 25 (Spring): 99–121.

Ballard, C. L., and M. F. Johnson. 2004. Basic math skills and performance in an introductory economics class. *Journal of Economic Education* 35 (Winter): 3–23.
Bauer, T., and K. F. Zimmermann. 1998. Learning efficiency of economics students. Discussion Paper 23, IZA (Institute for the Study of Labor), Bonn, Germany.
Becker, W. E. 1997. Teaching economics to undergraduates. *Journal of Economic Literature* 35 (Sept.): 1347–73.
Becker, W. E., and M. K. Salemi. 1977. The learning and cost effectiveness of AVT supplemented instruction: Specification of learning models. *Journal of Economic Education* 8 (Spring): 77–92.
Boivin, J., and S. Ng. 2003. Are more data always better for factor analysis? National Bureau of Economics Research (NBER) Working Paper 9829.
Bonello, F. J., T. R. Schwartz, and W. I. Davisson. 1984. Freshman–sophomore learning differentials: A comment. *Journal of Economic Education* 15 (Summer): 1205–10.
Borg, M. O., P. M. Mason, and S. L. Shapiro. 1989. The case of effort variables in student performance. *Journal of Economic Education* 20 (Summer): 308–13.
Borg, M. O., and S. L. Shapiro. 1996. Personality type and student performance in principles of economics. *Journal of Economic Education* 27 (Winter): 3–27.
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): 99–112.
Cardell, N. S., R. Fort, W. Joerding, F. Inaba, D. Lamoreaux, R. Rosenman, E. Stromsdorfer, and R. Bartlett. 1996. Laboratory-based experimental and demonstration initiatives in teaching undergraduate economics. *American Economic Review* 84 (May): 206–10.
Caudill, S. B., and D. M. Gropper. 1991. Test structure, human capital, and student performance on economics exams. *Journal of Economic Education* 22 (Fall): 303–06.
Chaker, A. M. 2003. Why colleges scoff at your kid's GPA. *Wall Street Journal,* July, 24, D1–2.
Charkins, R. J., D. M. O'Toole, and J. N. Wetzel. 1985. Linking teacher and student learning styles with student achievement and attitudes. *Journal of Economic Education* 16 (Spring): 111–20.
Chizmar, J. F., and A. L. Ostrosky. 1998. The one-minute paper: Some empirical findings. *Journal of Economic Education* 29 (Winter): 1–8.
Dewald, W. G., J. G. Thursby, and R. G. Anderson. 1986. Replication in empirical economics: *The Journal of Money, Credit and Banking* project. *American Economic Review* 76 (Sept.): 587–603.
Dunifon, R., G. J. Duncan, and J. Brooks-Gunn. 2001. As ye sweep, so shall ye reap. *American Economic Review* 91 (May): 150–54.
Evensky, J., D. Kao, R. Fenner, Q. Yang, and R. Falele. 1997. Addressing prerequisite math needs—a case study in introductory economics. *International Journal of Mathematical Education in Science and Technology* 28 (Sept.–Oct.): 629–39.
Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4 (Sept.): 272–99.
Ferber, M., B. Birnbaum, and C. Green. 1983. Gender differences in economic knowledge: A reevaluation of the evidence. *Journal of Economic Education* 14 (Spring): 24–37.
Fizel, J. L., and J. D. Johnson. 1986. The effect of macro/micro course sequencing on learning and attitudes in principles of economics. *Journal of Economic Education* 17 (Spring): 87–98.
Gohmann, S. F., and L. C. Spector. 1989. Test scrambling and student performance. *Journal of Economic Education* 20 (Summer): 235–38.
Grove, W., and T. Wasserman. 2004. The life-cycle pattern of collegiate GPA: Longitudinal cohort analysis and grade inflation. *Journal of Economic Education* 35 (Spring): 162–74.
———. 2006 Incentives and student learning: Results from a natural experiment. *American Economic Review* 96 (May forthcoming).
Grove, W., T. Wasserman, and A. Grodner. 2004. With so many proxies (for academic aptitude), which is a scholar to choose? Working Paper, Economics Department, East Carolina University.
Jencks, C. 1979. *Who gets ahead?* New York, NY: Basic Book.
Jensen, E. J., and A. L. Owen. 2000. Why are women such reluctant economists? Evidence from liberal arts colleges. *American Economic Review* 90 (May): 466–70.
Johnson, R. A., and D. W. Wichern. 1992. *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.
Kennedy, P., and J. Siegfried. 1997. Class size and achievement in introductory economics. *Economics of Education Review* 16 (Oct.): 385–94.
Lopus, J. S., and N. L. Maxwell. 1995. Should we teach microeconomic principles before macroeconomic principles? *Economic Inquiry* 38 (April): 336–50.
Lubotsky, D., and M. Wittenberg. Forthcoming. Interpretation of regressions with multiple proxies. *Review of Economics and Statistics.*

Maxwell, N. L., and J. S. Lopus. 1995. A cost effectiveness analysis of large and small classes in the university. *Educational Evaluation and Policy Analysis* 17 (Summer): 167–78.
McCloskey, D. N., and S. T. Ziliak. 1996. The standard error of regressions. *American Economic Review* 34 (March): 97–114.
Myatt, A., and C. Waddell. 1990. An approach to testing the effectiveness of teaching and learning of economics in high school. *Journal of Economic Education* 21 (Summer): 355–63.
Park, K. H., and P. M. Kerr. 1990. Determinants of academic performance: A multinominal logit approach. *Journal of Economic Education* 21: 101–12.
Paul, H. 1982. The impact of outside employment on student achievement in macro-economic principles. *Journal of Economic Education* 13 (Summer): 51–56.
Prince, R., P. H. Kipps, H. M. Wilheim, and J. N. Wetzel. 1981. Scholastic effort: An empirical test of student choice models. *Journal of Economic Education* 12 (Spring): 15–25.
Reid, R. 1983. A note on environment as a factor affecting student performance in principles of economics. *Journal of Economic Education* 14 (Fall): 18–22.
Rothstein, J. M. 2004. College performance predications and the SAT. *Journal of Econometrics* 121 (July–August): 297–317.
Sabot, R., and J. Wakeman–Linn. 1991. Grade inflation and course selection. *Journal of Economic Perspectives* 5 (Winter): 159–70.
Salemi, M. K., and G. E. Tauchen. 1982. Estimation of nonlinear learning models. *Journal of American Statistical Association* 77 (Dec.): 725–31.
———. 1987. Simultaneous nonlinear learning models. In *Econometric modeling in economic education research,* eds. W. E. Becker and W. B. Walstad, Boston: Kluwer-Nijhoff. 207–23.
Schmidt, R. M. 1983. Who maximizes what? A study in student time allocation. *American Economic Review* 73 (May): 23–28.
Siegfried, J. J., and W. B. Walstad. 1998. Research on teaching college economics. In *The principles of economics course: A handbook for instructors*, eds. W. B. Walstad and P. Saunders, New York: McGrawHill. 141–66.
Stock, J. H., and M. W. Watson. 2002. Diffusion indexes. *Journal of American Statistical Association* 97 (Dec.): 1167–79.

APPENDIX

The 37 published articles cited in "Research on Teaching College Economics" by Siegfried and Walstad (1998, 147–58) are listed here.

Anderson, G., D. Benjamin, and M. Fuss. 1994. The determinants of success in university economics courses. *Journal of Economic Education* 25 (Spring): 99–121.
Bonello, F. J., T. R. Schwartz, and W. I. Davisson. 1984. Freshman–sophomore learning differentials: A comment. *Journal of Economic Education* 15 (Summer): 1205–10.
Borg, M. O., P. M. Mason, and S. L. Shapiro. 1989. The case of effort variables in student performance. *Journal of Economic Education* 20 (Summer): 308–13.
Borg, M. O., and S. L. Shapiro. 1996. Personality type and student performance in principles of economics. *Journal of Economic Education* 27 (Winter): 3–27.
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): 99–112.
Butler, J. S., T. A., Finegan, and J. J. Siegfried. 1994. Does more calculus improve student learning in intermediate micro and macro economic theory? *American Economic Review* 86 (May): 454–59.
Cardell, N. S., R. Fort, W. Joerding, F. Inaba, D. Lamoreaux, R. Rosenman, E. Stromsdorfer, and R. Bartlett. 1996. Laboratory-based experimental and demonstration initiatives in teaching undergraduate economics. *American Economic Review* 84 (May): 206–10.
Caudill, S. B., and D. M. Gropper. 1991. Test structure, human capital, and student performance on economics exams. *Journal of Economic Education* 22 (Fall): 303–06.
Charkins, R. J., D. M. O'Toole, and J. N. Wetzel. 1985. Linking teacher and student learning styles with student achievement and attitudes. *Journal of Economic Education* 16 (Spring): 111–20.
Cohn, E., S. Cohn, and J. Bradley. 1995. Notetaking, working memory, and learning in principles of economics. *Journal of Economic Education* 26 (Fall): 291–307.
Durden, G. C., and L. V. Ellis. 1995. The effects of attendance on student learning in principles of economics. *American Economic Review* 85 (May): 343–46.
Ferber, M., B. Birnbaum, and C. Green. 1983. Gender differences in economic knowledge: A reevaluation of the evidence. *Journal of Economic Education* 14 (Spring): 24–37.
Fizel, J. L., and J. D. Johnson. 1986. The effect of macro/micro course sequencing on learning and attitudes in principles of economics. *Journal of Economic Education* 17 (Spring): 87–98.

Gleason, J. P., and W. B. Walstad. 1988. An empirical test of inventory model of student study time. *Journal of Economic Education* 19 (Fall): 315–21.
Gohmann, S. F., and L. C. Spector. 1989. Test scrambling and student performance. *Journal of Economic Education* 20 (Summer): 235–38.
Grimes, P., and J. F. Niss. 1989. Concentrated study time and improved learning efficiency: An experiment using *Economics U$A*. *Journal of Economic Education* 20 (Spring): 133–38.
Grimes, P., T. L. Krehbeil, J. E. Nielsen, and J. F. Niss. 1989. The effectiveness of *Economics U$A* on learning and attitudes. *Journal of Economic Education* 20 (Spring): 139–52.
Hodgin, R. F. 1984. Information theory and attitude formation in economic education. *Journal of Economic Education* 15 (Summer): 191–96.
Kennedy, P., and J. Siegfried. 1997. Class size and achievement in introductory economics. *Economics of Education Review* 16 (Oct.): 385–94.
Lage, M. J., and M. Treglia. 1996. The impact of integrating scholarship on women into introductory economics: Evidence from one institution. *Journal of Economic Education* 27 (Winter): 26–36.
Lopus, J. S., and N. L. Maxwell. 1995. Should we teach microeconomic principles before macroeconomic principles? *Economic Inquiry* 38 (April): 336–50.
Manahan, J. 1983. An educational production function for principles of economics. *Journal of Economic Education* 14 (Spring): 11–16.
Marlin, J. W., and J. F. Niss. 1982. The advanced learning system, a computer-managed, self-paced system of instruction: An application in principles of economics. *Journal of Economic Education* 13 (Summer): 26–39.
Maxwell, N. L., and J. S. Lopus. 1995. A cost effectiveness analysis of large and small classes in the university. *Educational Evaluation and Policy Analysis* 17 (Summer): 167–78.
Miller, J. C. 1982. Technical efficiency in the production of economic knowledge. *Journal of Economic Education* 13 (Summer): 3–13.
Myatt, A., and C. Waddell. 1990. An approach to testing the effectiveness of teaching and learning of economics in high school. *Journal of Economic Education* 21 (Summer): 355–63.
Park, K. H., and P. M. Kerr. 1990. Determinants of academic performance: A multinominal logit approach. *Journal of Economic Education* 21: 101–12.
Paul, H. 1982. The impact of outside employment on student achievement in macro-economic principles. *Journal of Economic Education* 13 (Summer): 51–56.
Prince, R., P. H. Kipps, H. M. Wilheim, and J. N. Wetzel. 1981. Scholastic effort: An empirical test of student choice models. *Journal of Economic Education* 12 (Spring): 15–25.
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): 369–81.
Reid, R. 1983. A note on environment as a factor affecting student performance in principles of economics. *Journal of Economic Education* 14 (Fall): 18–22.
Schmidt, R. M. 1983. Who maximizes what? A study in student time allocation. *American Economic Review* 73 (May): 23–28.
Van Scyoc, L. J., and J. Gleason. 1993. Traditional or intensive course length? A comparison of outcomes in economics learning. *Journal of Economic Education* 24 (Winter): 15–22.
Watts, M., and W. Bosshardt. 1991. How instructors make a difference: Panel data estimates from principles of economic courses. *Review of Economics and Statistics* 73 (May): 336–40.
Watts, M., and G. J. Lynch. 1989. The principles courses revisited. *American Economic Review* 79 (May): 236–41.
Wetzel, J. N., W. J. Potter, and D. M. O'Toole. 1982. The influence of learning and teaching styles on student attitudes and achievement in the introductory economics course: A case study. *Journal of Economic Education* 13 (Winter): 33–39.
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.