With So Many Proxies (for Student Aptitude), Which is a Scholar to Choose? Does it Matter?

Wayne A. Grove, Assistant Professor,
Economics Department, LeMoyne College, wgrove@lemoyne.edu

Tim Wasserman, Assistant Director,
Institutional Research and Data Management,
Center for Support of Teaching & Learning, Syracuse University, thwasser@syr.edu

Andrew Grodner, Research Associate,
Center for Policy Research, Syracuse University, agrodner@syr.edu.

Abstract

Although academic ability is the most important explanatory variable in studies of student learning, researchers control for it with a wide array of proxies and combinations thereof. In this paper we investigate how the proxy choice affects estimates of student learning from graded versus non-graded problem sets. Based on a natural experiment data set, we test over 150 specifications of a single model, each including a different combination of the 11 scholastic aptitude measures. Our results indicate that the proxy choice causes estimated treatment effects to bounce around by meaningful amounts. Next, based on three techniques, principal components analysis, factor analysis, and a recently-proposed "post-hoc" estimator, we find that collegiate GPA data uniformly function as the best proxies of students' individual propensities to learn economics—a result that runs counter to researchers' actual proxy choices.

Key words: Undergraduate economics, Proxy choice, Student learning, Methodology
JEL Code: A22, C81, I21,

We would like to acknowledge the valuable comments by Dan Black, Kevin Rask and Darren Lubotsky and by participants at our panel of the 2002 ASSA meetings in Atlanta and at seminars at Colgate University and LeMoyne College. We are grateful to the administrators at Syracuse University for their cooperation with our research and to LeMoyne College for a Research and Development grant.

Although academic aptitude control variables are the only consistently significant and meaningful explanatory variables of student learning of economics (Becker, 1997), scholars proxy for it with such a wide assortment of measures that in a survey selection of "best practice" studies most papers contained a unique set of ability measures. Despite this heterogeneity investigators seldom explain the logic of their control variable choices and rarely indicate how the use of alternative ability measures affects their estimates of learning.[1] In this paper we have two objectives: (1) to determine whether and how the choice of academic ability proxies matters and (2) to identify which measure(s) function as the best controls for scholastic ability.[2] For a popular press discussion of scholastic proxy choices, see, for example, the recent *Wall Street Journal* article about the varied criteria used by college admissions' committees at a number of elite institutions of higher education in the United States.[3]

To determine the effect of the proxy choice on estimates of student learning we use data from a natural experiment in which an instructional technique was in effect for one group of students but not for the other and test over 150 specifications of a single model. Each specification contained a different combination of the 11 scholastic aptitude measures—high school GPA and rank and variants of college GPA and SAT scores. Our results indicate that the proxy choice alone causes estimated learning gains to bounce around

---

[1] For a notable exception, Kennedy and Siegfried (1997, p. 388) argue for using the composite SAT score instead of math- and verbal-SAT scores separately, ACT scores, or GPA. Also, for example, Ferber et al (1983, 36) explain not using high school rank "because of the great variation in quality among different institutions."

[2] We do not attempt to replicate published articles as, for example, Dewald, Thursby, and Anderson (1986) did for all published articles in *The Journal of Money, Credit, and Banking* during a two-year period in the early 1980s (with discouraging results).

[3] In "Why Colleges Scoff at Your Kid's GPA," Anne Marie Chaker summarizes how GPA is accepted as reported (Columbia University and University of North Carolina), recalculated without, for example, pluses and minuses, freshman year grades (Johns Hopkins University), or those from certain courses, and ignored and rank used instead (Georgetown University and Haverford College; 2003, pp. D1-2).

by meaningfully different amounts in the context of college grades, ranging from a third of a letter grade increase (from a B to a B+) to two-thirds of a letter grade gain (from a B to an -A).[4] Even that degree of variation probably underestimates the true range of estimated learning gains since our analysis excludes ACT or TUCE scores, both of which are commonly used and have been criticized as poor proxies of student ability to learn economics.[5]

To provide some prescriptive guidance about which aptitude control variables to include in student learning studies, we use three multivariate techniques—principal components analysis, factor analysis, and a newly-proposed "post-hoc" estimator—to determine which of our 11 aptitude variables best proxies for students' academic ability.[6] All methods uniformly identify collegiate GPA as the superior proxy for individual students' propensities to learn economics.[7] While a substantial proxy choice literature exists regarding the use of index variables to summarize and replace larger sets of data (see Lubotsky and Wittenberg, 2003; Boivin and Ng, 2003), we know of no other study that, as ours does, evaluates the proxy choice effect and then identifies which individual variables proxy best for unobserved, latent measures.[8]

---

[4] We obtain a comparable range of estimated learning gains when conducting this analysis with a sample of non-freshman and a sample of only freshman.
[5] The *Test of Understanding of College Economics* or TUCE, an exam of 30 multiple choice questions, or 33 with the international questions, assess students' net gain in economic knowledge by comparing the performance on an exam at the beginning and again at the end of a course.
[6] For both principal components and factor analysis, see Johnson and Wichern (1992), and regarding the "post-hoc" estimator, see Lubotsky and Wittenberg (2003).
[7] Results from over 300 regressions suggest that collegiate GPA determines the level of the estimated treatment effects and that SAT scores and high school variables only marginally alter estimated learning gains. See Figure 1 which summarized 100 estimated treatment effects based on our sample of 117.
[8] For example, Stock and Watson (2002) showed that extracting three factors from around 150 disaggregated

SURVEY SELECTION OF BEST PRACTICE APTITUDE PROXY CHOICE

To obtain a profile of best practice use of academic ability control variables we selected the 37 published papers cited in "Research on Teaching College Economics" by Siegfried and Walstad (1998; citations come from pages 147-158) that empirically estimate college undergraduate student learning of economics.[9] As this survey selection indicates, scholars of economic education adhere to no standard protocol for measuring student academic ability—in fact, one set of authors argued for including no such information.[10] Researchers in our survey selection used college entrance exam scores most frequently (70 percent), TUCE in half the papers, college GPA in over 40 percent, and high school measures in twenty percent (with equal numbers using high school GPA and rank).[11] In sixty percent of the studies surveyed authors used a unique set of student academic ability control variables; this 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). Two studies used data from national databases (Kennedy and Siegfried, 1997; Lopus and Maxwell, 1995), Prince et al's (1981) from two schools, and the remainder from a single institution.

The heterogeneous assortment of scholastic aptitude controls resulted from a combination of authors excluding available data,[12] institutional constraints,[13] and international

---

series reduced forecast error and Forni and Reichlin (1998) found it helpful to form two factors from 450 disaggregated series.

[9] See the reference section for a list of the papers included in the selection, indicated with asterisks. An alternative survey of the literature would examined the empirical papers in the *Journal of Economic Education* (see McCloskey and Ziliak, 1996).

[10] Gohmann and Spector (1989) argue that ability is likely to be orthogonal to their order variable (238).

[11] For more details, see our working paper, Grove, Wasserman and Grodner (2003).

[12] Paul (1982), for example, used SAT scores to verify a similar distribution of student characteristics each semester over a four-year period but dis not utilize that college entrance exam data in his empirical estimates of student learning.

differences in educational practices.[14] In addition, scholars measured high school and college performance and college entrance exam scores in many different ways. Success in high school is measured by cumulative GPA (Borg et al, 1996), senior year GPA (Reid, 1983), and class rank (Fizel, 1986; Schmidt, 1983). Collegiate GPA was measured for both the semester being studied and cumulatively and in both cases might have included or excluded the economics course grade. Only rarely do scholars precisely indicate the calculation of college grades.[15] Finally, college entrance exam scores appear in six forms with the following frequencies: (1) ACT scores, 10, (2) both the individual math- and verbal-SAT scores (MSAT and VSAT), 8, (3) composite SAT (TSAT), scores 7, (4) VSAT scores only, 1, (5) ACT rank, 1, and (6) dummy variables for student's with above average ACT or SAT scores, 1.[16]

Whether or not this eclectic set of academic ability proxies is problematic depends upon how much the proxy choice influences estimated treatment effects. Before turning to that matter, though, let us consider the data and model we use.

---

[13] Institutional constraints comprise both restricted access to student records and the lack of data collection, namely variations in the type 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. Investigators with limited or no access to official student records at their own institutions use survey data to obtain aptitude measures (a universal outcome with multi-school studies) in which student's systematically overestimate their achievement (what Maxwell and Lopus, 1994, label the "Lake Wobegon" effect).

[14] Instead of college entrance exam scores, for example, Canadian scholars used performances in grade 11, 12 or 13 math and English courses (see Anderson et al, 1994 and Myatt and Waddell, 1990). Bauer and Zimmermann (1998) use high school grades for a study of learning of economics in a German university.

[15] Cardell et al (1996, 458) and Caudill (1991, 305) both report using "most recent GPA," Brasfield et al (1993, 102) "current college GPA," and Ferber et al (1983, 30-31) "university GPA." Bonello et al (1984, 206) use end of the semester cumulative GPA minus cumulative GPA "as projected by the University's admissions office." As best as we can tell, researchers tend to use cumulative, not individual semester, GPA.

[16] Jensen and Owen (2000) use another obvious choice, math SAT unaccompanied by verbal scores. Charkins et al (1985) used verbal SAT, Park and Kerr (1990) ACT percentile rank, and Borg et al (1989) dummy

A NATURAL EXPERIMENT AND THE DATA

During the fall of 1998, 239 undergraduate students enrolled in and completed four sections of introductory microeconomics taught by one of the authors at Syracuse University, a large, private residential university in the northeast.[17] A "natural experiment" occurred separating students into one group (of 143) whose course grades were based on problem set performance and another (of 96) whose course grades were not.[18] 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 encouragement to practice economics by solving them as important preparation for the exams. Students in both groups received the answer keys when the problem sets were handed in. Mastery of the course material was measured by performance on three required exams of equal value, each of which was assessed with a 0 to 100-point scale. To ensure as much uniformity as possible, the same grader evaluated each exam question for all 239 students. Thus, the only 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.[19]

---

variables for above average SAT and ACT scores.

[17] For more details, see Grove and Wasserman (2002).

[18] This occurred because a colleague fell seriously ill as classes began and one of the authors of this paper volunteered to teach his section for a few class periods. At the end of the second week of classes it became apparent that he 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 with a control group and a treatment group. For a discussion of sample selection bias, see Grove and Wasserman (2002).

[19] 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.

All the data used in this study come from university records, not from student surveys which have been shown to overstate actual performance.[20] Due to missing SAT scores, high school rank, or high school GPA data, a common set of academic aptitude variables exists for 117 students, of whom 71 were in the experimental group and 46 in the control group. In Table 1 we provide descriptive statistic (means and standard deviations for each variable used by group membership). Since mean exam scores and some ability measures were not equivalent in the experimental and control groups, it is appropriate for us to condition on ability in our empirical analysis—see Table 1. For a correlation matrix, see Table 2.

College entrance examination scores, typically thought to measure raw intelligence, a stock of knowledge, and/or a general aptitude for learning, have the virtue of being uniform and methodical but, when used to control aptitude in college courses, have the disadvantage of providing a measure at a point in time in the past. We use three SAT scores: SAT math (MSAT), SAT verbal (VSAT), and SAT combined (TSAT). Since virtually none of the students in our sample reported ACT scores, we can not evaluate the effect of ACT scores on estimated student learning.[21] This is particularly unfortunate because Kennedy and Siegfried (1997), in a multi-school study, and Feber et al (1983, 36), in a single institution study, argue that ACT scores fail to reflect the basic ability to learn economics.[22]

College grade point average (GPA), an institution of higher education's measure of academic success, directly measures success in course work at the same college or university and draws on the attributes associated with the SAT, but performance in college coursework

[20] Scholars survey students to obtain information about student academic success either because data is not available from administrators or to ease the data collection process in multi-school studies. Maxwell and Lopus (1994) label students' systematic overestimation of self-reported aptitude information as the "Lake Wobegon effect," so-named for Garrion Keillor's fictional town where "all children are above average."

[21] We have ACT scores for 3 of our 239 students.

[22] Kennedy and Siegfried (1997, 388), in a multi-school study, challenge the interchangeable use of ACT and

also reflects the application, throughout each academic term, of good study skills, motivation, organization, industriousness, perseverance, and consistency.[23] We use two temporal measures of collegiate GPA: grades earned during the semester evaluated in this study and cumulative GPA.[24] Same semester GPA provides a coincident measure of student academic success during the semester being studied, including information about shocks (e.g., prolonged or serious illnesses, severe personal problems, or grave family crises) that may affect a student's potential scholastic achievement.[25] From a practical perspective, much of the economic education research addresses the principles of economics course even though during the fall term freshman have no prior college cumulative GPA.[26] Since 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).[27] We include the same semester GPA with the economics grade (SemGPA) and the cumulative

---

SAT scores, reporting that estimates student learning with ACT had an R^2 of 0.12 versus 0.86 with SAT.

[23] Jencks (1979) demonstrates 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) establish that a "clean-home measure" is predictive of own and children's earnings 25 years later and children's schooling. On this point, note that the only student with a perfect math-SAT score from our full sample, not the sub-sample of 117 students used for this study, failed to hand in the required four problem sets.

[24] Caudill and Gropper (1991) view 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" (305). In the fall term, though, freshman have no cumulative GPA, whereas sophomores, juniors and seniors have two, four or six previous semesters of grades, respectively. Hence, we create a non-freshman sample to test the effect of prior cumulative GPA.

[25] A longitudinal analysis of the Syracuse University class of 2001, comprising 2,552 students enrolled for up to 8 semesters, reveals that student's experience large negative deviations from their 4-year cumulative GPA, especially during the second semester freshman and first semester sophomore terms: ten 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 two percent as 1.0 or more above.

[26] Our cumulative GPA measure includes the semester of the study. We construct a prior cumulative GPA variable with a sample that excludes freshman (non-freshman sample).

[27] For the correlations, see Table 2. Jerry Evensky suggested excluding the economics grade. Other papers that use this GPA variation include Evensky et al (1997) and Chizmar and Ostrosky (1998).

GPA with the economics grade (CumGPA) in our analysis 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 since those "independent" variables contain some or all of the dependent variable.[28]

While grades distinguish student performance within a course, unlike the systematic nature of SAT scores, students take such a heterogeneous mix of courses taught by an assortment of professors than differences in grade distributions between departments, courses, and faculty members may obscure the information provided by GPA.[29] For example, one student's B+ might represent the lowest grade in a course whereas another's B+ might be the top course grade. To permit meaningful comparison of between class grades, we created z-score GPA measures which calculate a student's grade deviation from the distribution mean for each course—such "standardized GPA" data provide more useful information than raw GPA.[30] Rather than access to each sample member's transcript, though, construction of z-score GPA data requires access to the transcripts of every student enrolled in each course taken by a member of the sample group.[31]

Some researchers have expressed skepticism about using cumulative high school GPA (HSGPA) to measure cognitive ability either because of long time lags or potentially large variations in the standards between and the quality of schools, school districts and states (Ferber et al, 1983, 36), as well as admissions committees at many institutions of

[28] Note that our dependent variable is mean exam scores, not course grade.
[29] 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).
[30] 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.
[31] Database systems dramatically reduce the cost of such calculations but creation of zCumGPA-ECN remains a daunting task: for seniors, for example, an investigator must calculate a z-score for each of the dozens of courses based on GPA data for every student in each of those courses.

higher education (Chaker, 2003, D2). High school rank (HS%) may provide a better proxy—an option akin to Georgetown University's and Haverford College's use high school rank, instead of GPA, in their admissions' decisions (ibid).

In addition to the absence of ACT scores, we lack TUCE scores since our data do not come from a planned study. Although, typically, the post-TUCE score is used as the dependent variables and the pre-TUCE as an independent variable to obtain a measure of the net value added, how can a researcher ensure that pre-TUCE scores measure student knowledge of economics since grading it (unlike grading the post-TUCE) seems out of the question? In addition, scholars have criticized the use of TUCE as a narrow gauge of learning (30 multiple choice questions, or 33 with the international questions included).[32] Without the pre-TUCE score (as some have done e.g., Kennedy and Siegfried, 1997), the TUCE becomes a short standardized multiple-choice test which facilitates multi-school studies rather than a measure of the net gain in economics knowledge.

OUR BASIC MODEL

We hypothesize that basing course grades on regular problem set performance throughout the semester improves student's exam performance, controlling for their academic ability and demographic characteristics.[33] To check for inequalities between the two groups that might call into question their presumed equivalence, we compared the distribution of gender, race/ethnicity (white versus other), and class standing (freshman versus other) for the experimental and control group via chi-square analysis. The

[32] This is one of Becker's (1997) four explanations for the lack of empirical support for the efficacy of a variety of instructional methods and techniques. 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.

[33] Student learning is typically modeled as a production function in which exam performance results from student human capital inputs, demographic characteristics, student effort, and treatment effects.

distributions by group are roughly equal for gender ($\chi^2_{(1)} = 0.0409$, $p > .84$) and race/ethnicity ($\chi^2_{(1)} = 0.0417$, $p > .84$), but not for class standing ($\chi^2_{(1)} = 2.9017$, $p < .10$). Consequently, to avoid estimation bias we include in our model a dichotomous, independent variable for academic class standing (1 if freshman; 0 if not). To test whether graded problem sets improved student learning, we use OLS to regress mean exam grades in Economics 101 on freshman status, membership in the control and treatment groups, and then, in succession, various scholastic aptitude proxy(ies), as shown in the following model.

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

In Table 3 we provide regression results for four specifications of this model. Since 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 SemGPA-ECN (first column) indicates that students in the treatment group improved their performance by 3.38 points, i.e., by a third of a letter grade. Together those four regressions indicate that freshman status is not significant.

DOES THE PROXY CHOICE AFFECT ESTIMATED TREATMENT EFFECTS?

We estimate our basic student learning model alternatively with each of our 11 academic aptitude proxies alone and then with all sensible combinations of them, for a total of over 150 different regressions. The results of the four specifications of the model in Table 3, each including a different individual proxy, reveal estimated learning gains ranging from over 3 points to 5.5 points. As with all of our regression results, the estimated treatment effect and the aptitude proxy coefficients of the four versions in Table 3 are positive and statistically significant at the 5- or 1-percent levels.

We summarize 100 of the estimated treatment effect coefficients in Figure 1, using a legend to indicate which variables are included as aptitude controls in each specification. The figure succinctly indicates that the aptitude proxy choice does cause the magnitude of the estimated student learning gains to bounce around enough for most people to acknowledge that it constitutes a substantive difference in students' grades. The lowest estimated learning gain of 2.5 points (R^2 > .5) implies that 60 percent of experimental group students would have received a third of a letter grade increase for the course (e.g., B to a B+), whereas the highest coefficient estimates a grade increase of 5.8 points (R^2 = .11), indicating that 90 percent of the students would have experienced a two-thirds of a letter grade increase (e.g., B to an A-). Although a 130 percent variation in estimated learning gains due merely to the proxy choice represents a meaningfully large difference in course grades in this context, the estimated treatment effects, as a function of the controls used, are not statistically significantly difference from one another, in the sense that 95 percent confidence intervals created around the estimated treatment effects overlap.[34]

The highest range of estimates of improvement in exam performance (and the lowest R^2s) due to the instructional methods utilized, of 5.2 to 5.6 points, come from models in which academic ability is controlled only with pre-college measures and without collegiate GPA data. Such a pre-college-controls-only specification, as occurred in over 20 percent of the papers in our survey, yields estimated learning gain magnitudes most similar to that from

[34] We attribute this lack of statistical significance to the large standard errors associated with a relatively small sample size. The confidence intervals are the parameter estimate plus and minus twice the standard error.

a model with no academic ability control variables at all.[35] Without offering evidence to substantiate the claim, Gohmann and Spector (1989) argue that student learning due to instructional methods can be estimated without measures of scholastic aptitude.[36] Their specification generates the largest estimate of student learning (5.8 points) with a set of results that explain the least amount of variation in students' exam performance (11 percent; see Figure 1).

As a check on the robustness of these results, we also estimate our basic model with a non-freshman population and with a freshman-only sample.[37] The additional over 200 regressions provide similar evidence that the proxy choice matters for estimates of student learning.[38]

WITH SO MANY PROXIES, WHICH IS (ARE) A SCHOLAR TO CHOOSE?

Having examined what difference the proxy choice makes, we now consider the prescriptive and policy-oriented question of which proxy(ies) best control for academic ability. 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 step).[39] That pattern indicates that collegiate GPA measures primarily determine estimated treatment

---

[35] Those pre-college parameter estimates were, on average, 7.5 percent lower than with using no aptitude control variables at all and had adjusted R^2's twice to three times.

[36] They justify excluding measures of human capital on the basis that ability is likely to be orthogonal to their order variable (238).

[37] Since for freshman concurrent semester GPA equaled cumulative GPA, we constructed a "prior cumulative GPA" measure by eliminating freshman from the sample (the non-freshman sample; n=82). The estimated treatment effects range from 1.1 to 4.6 points, a range of estimated learning gains in excess of 170 percent (thus, a 50 percent chance of a two-thirds of a letter grade increase to the certainty of an entire letter grade boost). Since, as we show elsewhere, first year students' disproportionately benefited from this particular instructional method (Grove and Wasserman, 2002), we also estimated the same specifications of the model with a freshman-only sample (n=35) and find a 90 percent range of estimated learning gains.

[38] For more details, see our working paper (Grove, Wasserman and Grodner, 2003).

effects whereas pre-college 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 increases the amount of the variation in students' exam scores that can be explained (increases the AdjR^2).[40] Given the range of estimated learning gains due to the proxy choice, note the stability of the coefficients with the inclusion or exclusion of particular proxy variables.

While our proxy choice exercise, as summarized in Figure 1, suggests that college GPA is the most important aptitude proxy, we directly test it using three multivariate techniques—principal components analysis, factor analysis, and a recently proposed "post-hoc" estimator. Each method is used to derive an index from the 11 aptitude measures and, then, the explanatory power of each proxy is determined by comparing its correlation with each method's index. Note that the goal of the extensive proxy choice literature, in contrast to our purpose here, is to extract a few factors from a large number of variables and use those index variables instead of the actual data in regressions.[41]

Principal component regression analysis finds uncorrelated linear combinations of the correlated proxy variables that explain the highest variance-covariance structure of the proxies.[42] The first principal component is a weighted average of the independent proxy

---

[39] Whereas any single GPA measure, combined with any or all pre-college variables, generates a 10-20 percent range of estimated learning gains, the range of treatment effects estimated with SAT and high school measures, but excluding collegiate GPA, exceeds 100 percent.

[40] Consider the case, for example, of SemGPA-ECN: the estimated learning gain with the proxy by itself is 3.38 points (adjusted R^2 of .44), adding only high school rank lowers the coefficient by 2 percent (to 3.31; adjusted R^2 of .46), and adding instead TSAT and both high school variables raises the estimated learning gain by 12 percent (to 3.78; adjusted R^2 of .53).

[41] See, for example, the papers discussed in Lubotsky and Wittenberg (2003) and Boivin and Ng (2003).

[42] The set up of a principal components analysis is to have the input proxy variables on the right hand side explaining the latent factors. See Webster (2001) for an application of principal component analysis to the *U.S.*

variables in which the weights are chosen to make the composite variable account for as much of the variability in the proxies (ignoring the relationship with the dependent variable) as is possible to capture via a linear combination, and each succeeding component accounts for as much of the remaining variability as possible.[43] Factor analysis, an extension of principal components analysis, adds a random component (an error term) to the deterministic part of the proxy variables so the proxies are not exact linear combinations of latent factors (in fact, the deterministic part of factor analysis describes a principal component model).[44] The decomposition of the factors into the deterministic part and the random part allow us to interpret retrieved latent factors as independent, "source," variables, rather than as a weighted average of the proxy variables (as with principal components).[45] Principal components analysis uses all variability in a proxy, while in factor analysis uses only the variability a proxy shares in common with the other proxies.[46] As in most cases, these two methods yield very similar correlations of the generated psuedo-ability variables with our scholastic ability proxies: the generated factors are most highly correlated with college GPA measures (.81-.88), followed by math and the combined SAT (.71-.77), and the least with high school variables and verbal SAT (.55-.61)—see Table 4.

News & World Report rankings of colleges and universities.

[43] The first principal component captures the most variance among the proxies, 55 percent, and the first three components collectively capture 85 percent. Of the 11 aptitude proxies examined, college GPA variables have the highest coefficients on the first principal component (greater than 0.4) suggesting that college grades, however measured, represents the most important proxy type.

[44] Note that the set up of a factor analytic model is different from the principal components model in that now the input proxy variables are on the left hand side, and in some sense become dependent variables, to be explained by the latent factors. Factor analysis is a richer model with more relaxed assumptions (see Johnson and Wichern, 1992, Chs. 8 and 9).

[45] The first factor alone explains over 80 percent of the total variance in the aptitude control variables.

[46] Principal component analysis is often preferred as a method for data reduction, while factor analysis to detect structure.

Lubotsky and Wittenberg's (2003) newly-proposed posthoc estimator entails running a regression with all 11 proxies included and using their coefficients to compute a weighted average. The weights used are the ratios of the covariances between each independent variable and the dependent variable, to the covariance between the first independent variable and the dependent variable.[47] This method minimizes the error variance of the aggregate set of proxies and achieves the greatest reduction in attenuation bias (i.e., biased coefficients that result from poorly measured independent variables).[48] This procedure generates a composite index from the individual proxies which if used instead of them in the regression will provide the same posthoc estimator coefficient.[49] The correlation between the proxies and a composite index constructed from the separate proxies provides a much stronger degree of differentiation between the proxy choices and a slightly different ranking of them than obtained with the two other methods. While college GPA variables remain the most strongly correlated with the posthoc index variable (.83-.89), high school GPA joins SAT math and SAT sum in the second cluster with the highest correlation (.61-.65), and the least correlation exists with high school rank and SAT verbal (.43-.48)—see Table 2. The posthoc estimator gap between these three clusters of proxies is twice as large with principal components and factor analysis.[50] Finally, note that only with the posthoc selection technique do z-score

---

[47] This index has the same interpretation as the factors in factor analysis and has the units of the first variable used in the analysis.
[48] Although the posthoc estimator retains the ordinary least squares bias, the index variable approximates the true value of students' ability since the optimal weighting provided by least squares means that no linear combination of the proxies will produce a less biased estimate.
[49] The *post hoc* index, x(p), is constructed as follows: x(p)=(1/b (p))*sum(j=1,k)[x(j)b(j)] where b(p) is the coefficient on the post estimator and b(j) is the *j*-th regression coefficient (for more details, see Lubotsky and Wittenberg, 2003, p. 12).
[50] The difference between the average correlation of the variables in the first cluster and the index variable, and the average correlation between the variables in the second cluster and the index variable are roughly .1 with principal component and factor analysis compared with .25 for the posthoc estimator; and the average correlations between the variables in the first and third clusters and the index variables are approximately .2 and .4, respectively.

GPAs have the highest correlation with the proxy index.[51] Although z-scores have the virtue of improving between-course grade comparisons, many scholars and researchers will not have sufficient access to student record databases to construct such standardized measures of college grades.

Our suggestion for selecting the best set of proxies is to start with the variables that have the highest correlation with the generated index variables and be wary of the increased variance on the other variables in the model. The decision about which and how many control variables to include entails a trade-off between including all "powerful" proxies but not those with little explanatory power which lower the efficiency of the results and increase the attenuation bias. More specifically, a user would like to have proxies whose measurement error components are uncorrelated, meaning, additional covariates shall explain the latent variable but not other proxies or covariates in the model (Lubotsky and Wittenberg, 2003). Thus, if variables are grouped into highly-correlated clusters, researchers may want to use only one variable from the cluster because within the cluster it is likely that the measurement errors of proxies are correlated. Regarding the correlations between our type categories of aptitude proxies, note that (1) all collegiate GPA measures are highly correlated with each other (.86-.96) but not with SAT scores (.24-.43) or high school data (.34-.39); (2) high school GPA and rank are also highly correlated with each other (.70) but not with college grades (.34-.39) or SAT scores (.38-.44); and (3) SAT math and verbal have a correlation of .51 and both are highly correlated with the composite SAT score (.86 and .87). Based on the foregoing analysis, a scholar simply interested in controlling for academic ability in a study of some student learning should include one of the collegiate GPA variables

[51] For example, using zSemGPA-ECN instead of SemGPA-ECN increases the treatment effect by more than a quarter, improves the significance level from 5 to 1 percent, and nudges the adjusted R^2 up from .44 to .47 (see

and either SAT math or SAT sum and/or high school GPA. Thus, scholars may choose not to include high school rank and SAT verbal along with college grades even though they do contribute marginally to a better understanding of student learning.

CONCLUSIONS AND IMPLICATIONS

This study shows that the proxy choice alone causes students' course grade to increase from, for example, a B to a B+ or to an A-, a meaningful difference in the context of student learning, although not a statistically significant one (see Figure 1). While principal components analysis and factor analysis identify collegiate GPA as a better proxy for academic ability than SAT scores and high school variables, the superiority of collegiate GPA is especially pronounced with the newly-proposed "post-hoc" estimator (see Table 4).[52] Our results suggest that investigators in student learning studies should control for academic aptitude by including some version of college GPA data and SAT math or SAT sum scores and/or high school GPA. Beyond those, adding high school rank and SAT verbal scores provides little additional explanatory power.[53] These results differ from usage by scholars who, as indicated in our survey selection, most commonly used college entrance exam score proxies, in a quarter of the studies exclusively used pre-college measures, and included high school rank as often as high school GPA.

The major shortcoming of this study is our inability, for lack of data, to evaluate the role of ACT and TUCE scores as aptitude proxies.[54] Since both of those variables have been criticized as poor scholastic ability control variables, we expect that our results underestimate

---

Figure 1 and our working paper, Grove, Wasserman and Grodner, 2003).

[52] Regarding the "post-hoc" estimator, see Lubotsky and Wittenberg (2003).

[53] In contrast to this result, Georgetown University and Haverford College, for example, have eschewed high school GPA in favor of high school rank (Chaker, 2003, p. D2).

[54] Thirty percent of our survey selection relied upon ACT scores and over half used TUCE data.

the actual proxy choice-induced variability in estimated treatment effects in the literature.[55] Regarding the generalizability of our findings for the economics education literature, how does the proxy choice affect estimated treatment effects for other pedagogical techniques?[56] Of particular interest is whether college GPA can function as a control measure in a multi-school study given the variability of cross-institution academic standards.[57] If, as we find, 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 multi-school inquiries, this distinction will need to be carefully understood.[58] Our survey selection contained two multiple-institution studies—one of which controlled for ability with SAT scores (Kennedy and Siegfried, 1997) and the other with college grades (Lopus and Maxwell, 1995)—and 35 single-institution studies, two-thirds of which excluded college GPA.[59] Perhaps these results about the choice of aptitude proxies and its effect upon estimates of student learning will encourage a more systematic and transparent selection of scholastic control variables and reporting of the effect of scholars' proxy choices for their empirical findings.

---

[55] Regarding reservations expressed about TUCE data and ACT scores, see Kennedy and Siegfried (1997) and Saunders and Walstad (1990). Together this means we could not completely replicate over 60 percent of the aptitude proxies used in our survey selection.

[56] Replication, though, will require identifying data sets that overcome sample selection problems by containing a control group in order to permit estimation of learning gains (regarding this problem, see Becker 1997, 1366).

[57] In this context, we wonder whether z-scores, as we have used for a single university, can be used to overcome the problem of different grading norms between institutions of higher education.

[58] Kennedy and Siegfried's (1997) argue for using only the SAT score since GPA was insignificant when both were used and when individually used the R^2 with SAT was 0.86 versus 0.59 with GPA (388).

[59] The "multi-school" studies used national datasets. One of the other papers, Prince et al (1981), used data from two schools: James Madison University and Virginia Commonwealth University.

Figure 1
Estimated Treatment Effects According to the Aptitude Proxy(ies) Included
Dependent Variable: Mean Exam Scores (in points); N=117

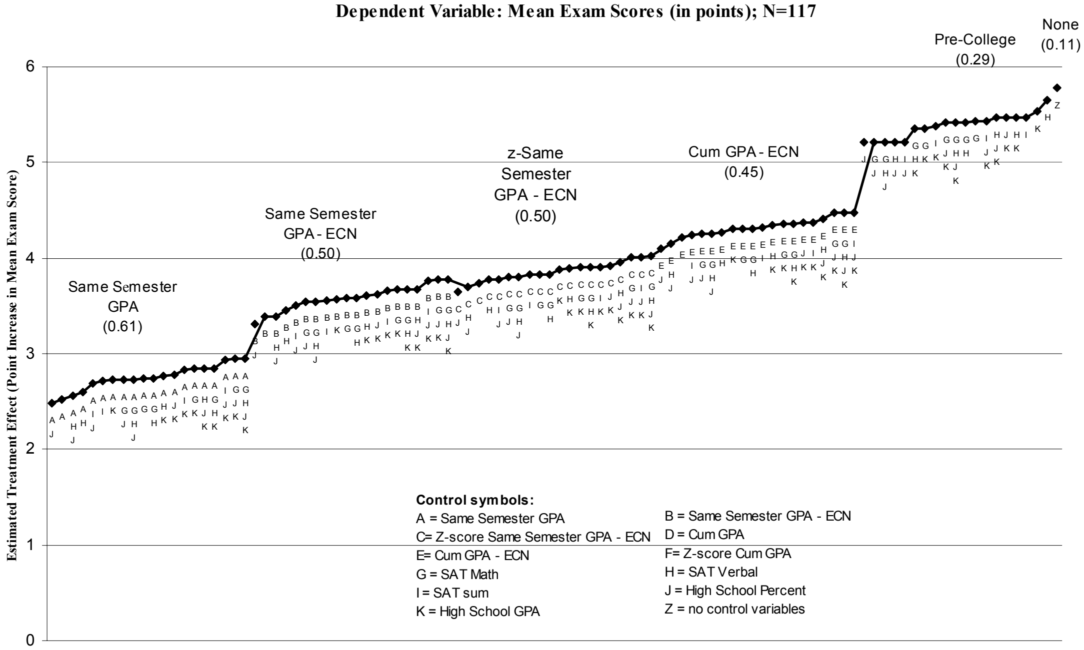

Table 1

Descriptive Statistics:
Means, Standard Deviations, and P-Values for T-tests and Chi-square
(N = 117)

	Means and in Parenthesis Standard Deviation		
Dependent and Independent Variables	**Graded Group (N=71)**	**Non-Graded Group (N=46)**	**P-values** [1]
Mean Exam Score	83.0 (8.12)	77.0 (8.94)	0.000*
SemGPA	3.12 (0.57)	2.78 (0.64)	0.003*
SemGPA-ECN	3.14 (0.64)	2.79 (0.78)	0.009*
zSemGPA-ECN	0.20 (0.63)	-0.10 (0.84)	0.023*
CumGPA	3.21 (0.49)	3.05 (0.47)	0.087
CumGPA-ECN	3.18 (0.55)	2.99 (0.53)	0.056
zCumGPA-ECN	0.23 (0.55)	0.04 (0.64)	0.091
MSAT	590.1 (81.4)	579.8 (70.2)	0.480
VSAT	567.2 (81.5)	559.3 (66.1)	0.586
TSAT	1157.3 (143.5)	1139.1 (114.1)	0.470
HS%	75.3 (14.90)	72.0 (14.63)	0.238
HSGPA	3.40 (0.43)	3.35 (0.37)	0.570
White	0.79 (0.41)	0.80 (0.40)	0.838[2]
Male	0.63 (0.49)	0.65 (0.48)	0.840[2]
Freshman	0.35 (0.48)	0.22 (0.42)	0.102[2]

[1] P-values for continuous variables from t-tests and for dichotomous variables from chi-square.
[2] Chi-square analysis.
* Means differ at the 5-percent level of significance.

Table 2

Correlation Matrix (N-117)

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

Table 3

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

Dependent Variable: Mean Exam Score (in points); N =117

Independent Variables	SemGPA-ECN	CumGPA-ECN	SATsum	HSGPA
GradedProblemSetGroup	3.38* (2.60)	4.22** (3.16)	5.47** (3.77)	5.53** (3.80)
Academic Aptitude Proxy (see column heading)	7.57** (8.40)	9.16** (7.57)	0.03** (5.29)	9.07** (5.18)
Freshman	0.40 (0.29)	0.36 (0.25)	0.54 (0.34)	1.01 (0.64)
Adjusted R^2	0.45	0.4	0.28	0.27

Note: t-statistics are in parentheses. The standard error is for the listed aptitude proxy.
* Mean is different from zero at the 5-percent level of significance.
** Mean is different from zero at the 1-percent level of significance.

Table 4

Correlations Between Academic Aptitude Variables and Index Variables Created from Three Multivariate Estimation Methods
(N =117)

	Multivariate Estimation Methods		
	Principal Components	**Factor Analysis**	**Posthoc Estimator**
zCumGPA-ECN	.87	.88	.89
zSemGPA-ECN	.81	.82	.89
SemGPA-ECN	.81	.81	.87
CumGPA-ECN	.84	.84	.83
MSAT	.71	.72	.65
TSAT	.75	.77	.62
HSGPA	.61	.55	.61
HS%	.61	.56	.48
VSAT	.59	.61	.43

Note: See text for details.

REFERENCES*

Note: Asterisks at the end of each citation indicate the 37 articles included in our survey selection.

Anderson, G., D. Benjamin, and M. Fuss (1994). The determinants of success in university economics courses. *Journal of Economic Education*, 25(2), Spring, 99-121.*

Bauer, Thomas and Klaus F. Zimmermann. 1998. "Learning Efficiency of Economics Students," Discussion Paper No. 23, IZA (Institute for the Study of Labor), Bonn, Germany.

Becker, W. E. (1997): "Teaching Economics to Undergraduates," *Journal of Economic Literature*, 35(3), 1347-1373.

Boivin, Jean and Serena Ng (2003) "Are More Data Always Better for Factor Analysis?" NBER Working Paper 9829, July 2003.

Bonello, F.J., T.R. Schwartz, and W.I. Davisson (1984) "Freshman-sophomore learning differentials: A comment," *Journal of Economic Education*, 15(3), (summer), 1205-1210.*

Borg, M.O., P.M. Mason, and S.L. Shapiro (1989), "The case of effort variables in student performance," *Journal of Economic Education*, 20(3), (summer), 308-313.*

Borg, M.O. and S.L. Shapiro (1996), "*Personality type and student performance in principles of economics," Journal* of Economic Education, 27(1), (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(2), (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(2), (May), 454-459.*

Cardell, N. Scott, Rodney Fort, Wayne Joerding, Fred Inaba, David Lamoreaux, Robert Rosenman, Ernst Stromsdorfer, and Robin Bartlett (1996). Laboratory-based experimental and demonstration initiatives in teaching undergraduate economics. *American Economic Review*. 84(2), (May), 206-210.*

Caudill, S.B. and D.M. Gropper (1991), "Test structure, human capital, and student performance on economics exams," *Journal of Economic Education*, 22(4), (Fall), 303-306.*

Chaker, Anne Marie (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(2), (Spring), 111-120.*

Chizmar, John F. and Anthony L. Ostrosky (1998), "The One-Minute Paper: Some Empirical Findings," *Journal of Economic Education*, 29(1), (Winter), 1-8.

Cohn, E., S. Cohn, and J. Bradley (1995), "Notetaking, working memory, and learning in principles of economics," *Journal of Economic Education*, 26(4), (Fall), 291-307.*

Dewald, William G., Jerry G. Thursby and Richard G. Anderson (1986), "Replication in Empirical Economics: *The Journal of Money, Credit and Banking* Project," *American Economic Review*, 76(4), (Sept.), William G. 587-603.

Dunifon, Rachael, Greg J. Duncan and Jeanne Brooks-Gunn. 2001. "As Ye Sweep, So Shall Ye Reap," *American Economic Review*, 91(2), 150-154.

Durden, G. C., and L. V. Ellis (1995): "The Effects of Attendance on Student Learning in Principles of Economics," *American Economic Review*, 85 (2), 343-346.*

Evensky, Jerry, Duke Kao, Richard Fenner, Qing Yang, and Remi Falele. 1997. "Addressing Prerequisite Math Needs—A Case Study in Introductory Economics." *International Journal of Mathematical Education in Science and Technology*. 28(5): 629-639.

Ferber, M., B. Birnbaum, and C. Green (1983): "Gender Differences in Economic Knowledge: A Reevaluation of the Evidence," *Journal of Economic Education*, 14(2) (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(2) (Spring), 87-98.*

Forni, M. and L. Reichlin (1998), "Let's Get Real: a Factor-Analytic Approach to Disaggregated Business Cycle Dynamics," *Review of Economics Studies* 65, 453-473.

Gleason, J.P. and William B. Walstad (1988), "An empirical test of inventory model of student study time," *Journal of Economic Education*, 19(4) (Fall), 315-321.*

Gohmann, S.F. and L.C. Spector. 1989. Test scrambling and student performance. *Journal of Economic Education*, 20(2) (Summer), 235-238.*

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(2) (Spring), 139-152.*

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(2) (Spring), 139-152.*

Grove, Wayne and Tim Wasserman (2002), "Incentives and Student Learning: Results from a Natural Experiment," manuscript.

Hodgin, R.F (1984), "Information theory and attitude formation in economic education," *Journal of Economic Education*, 15(3) (Summer), 191-196.*

Jencks, Christopher. 1979. *Who Gets Ahead?* New York: Basic Book.

Jensen, Elizabeth J., and Ann L. Owen. 2000. "Why Are Women Such Reluctant Economists? Evidence from Liberal Arts Colleges," *American Economic Review*, 90(2): 466-470.

Johnson Richard A. and Dean W. Wichern. 1992. *Applied MultivariateStatistical Analysis*, Englewood Cliffs, New Jersey: Prentice Hall.

Kennedy, P. and J. Siegfried (1997), "Class size and achievement in introductory economics," *Economics of Education Review*, 16(4), 385-394.*

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(1) (Winter), 26-36.*

Lopus, J. S., and N. L. Maxwell (1995): "Should we Teach Microeconomic Principles Before Macroeconomic Principles?" *Economic Inquiry*, 38 (2), (April), 336-350.*

Lubotsky, Darren and Martin Wittenberg (2003) "Interpretation of Regressions with Multiple Proxies." Manuscript (http://www.business.uiuc.edu/lubotsky/papers.html).

Manahan, J. (1983), "An educational production function for principles of economics," *Journal of Economic Education*, 14(2) (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(3) (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(2), 167-178.*

McCloskey, Deirdre N., and Stephen T. Ziliak. 1996. "The Standard Error of Regressions," *American Economic Review*. 34:1, 97-114.

Miller, J.C. 1982. Technical efficiency in the production of economic knowledge. *Journal of Economic Education*, 13(3) (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(3) (Summer), 355-363.*

Park, K.H. and P.M. Kerr (1990), "Determinants of academic performance: A multinominal logit approach," *Journal of Economic Education*, 21(2) (Spring),101-112.*

Paul, H. (1982), "The impact of outside employment on student achievement in macro-economic principles," *Journal of Economic Education*, 13(3) (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(2) (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(4) (Fall), 369-381.*

Reid, R. (1983), "A note on environment as a factor affecting student performance in principles of economics," *Journal of Economic Education*, 14(4) (Fall), 18-22.*

Sabot, Richard and John Wakeman-Linn (1991), "Grade Inflation and Course Selection," *Journal of Economic Perspectives*, 5:1, 159-170.

Schmidt, R. M. (1983): "Who Maximizes What? A Study in Student Time Allocation," *American Economic Review*, 73 (2), (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*, by William B. Walstad and Philip Saunders (McGrawHill), 141-166.

Stock, J.H. and M.W. Watson (2002), "Diffusion Indexes," *Journal of American Statistical Association* 97:460, 1167-1179.

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(1) (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 (2), 336-340.*

Watts, M. and G.J. Lynch (1989), "The principles courses revisited," *American Economic Review*, 79(2), (May), 236-241.*

Webster, Thomas J. (2001), "A principal component analysis of the *U.S. News & World Report* tier rankings of colleges and universities," *Economics of Education Review*, 20(3), 235-244.

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(1), (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(3), (Summer), 219-231.*