Research in Economic Education

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

PETER KENNEDY, Section Editor

Test Structure, Human Capital, and Student Performance on Economics Exams

Steven B. Caudill and Daniel M. Gropper

Budgetary considerations have led to the widespread use of large class sections to teach principles of economics. Multiple-choice tests are frequently the preferred method of examination for such classes, because they can be quickly and objectively graded. Unfortunately, multiple-choice tests may increase the potential for cheating. A solution to this problem is to scramble the order of the questions to create different versions of the same exam. These procedures will increase the expected cost of cheating, but the structure of the test may affect student performance; thus, students' scores may reflect something other than solely their knowledge of economics.

The effect of question order on student performance in economics classes has been investigated by a number of researchers. Taub and Bell (1975) found that students who took a test where the question order was scrambled scored slightly lower than those who took a chronologically ordered exam. This issue has been reexamined in recent articles by Bresnock, Graves, and White (1989) and by Gohmann and Spector (1989). Their findings suggest that question order does not have a significant effect on exam scores. The article by Bresnock, Graves, and White contains a recent review of the related literature.

Steven B. Caudill and Daniel M. Gropper are professors at Auburn University. *The authors are grateful to Dan Berry and Doug Bunn for assistance with the data set.*

Although it may be of central importance, none of the previously published literature has included a measure of student ability in their investigations. In this article, we extend this research by including variables that provide some measure of the student's human capital. We also consider student performance on all exams during the course rather than only a single exam.

Data and Model

The data set upon which this study was based consisted of all of the exam scores for 211 students in a principles of microeconomics class. In the course, four 1-hour exams and a final exam were given. Because students missed some of the exams, the data set consisted of 1,021 exam scores. Some observations were deleted because of missing values for some variables. Each exam score was mean adjusted or curved so that the class average was equal to seventy-two. The dependent variable was the exam score.

Two versions of each exam were used to test the students. One version contained questions presented in the same order as the lecture material, or chronologically. The other version had the questions rearranged so that they were in no particular order and could be considered random. The questions on the exams were independent, and the answers to the questions were not reordered. Students also did not have assigned seats, so the distribution of exams could be considered random.

To determine the effects on student performance of question order, we estimated several specifications of a regression model in which the exam score was the dependent variable. The primary independent variable, CHRONO, had the value 1 if the questions on the exam were presented in chronological order, and 0 otherwise. Some additional information on student characteristics was available and was included in our model. Consistent with previous research (Lumsden and Scott [1987]; Gohmann and Spector [1989]), we included a dichotomous variable for gender that was equal to 1 if the student was male and 0 if female. One factor, which had not been included in previous studies, was some measure of the individual student's human capital. This variable was not included in the model estimated by Gohmann and Spector (1989), and its omission may account for their insignificant findings. Although Gohmann and Spector rightly claim that the omission of this variable will not bias their results because ability is likely to be orthogonal to their order variable, they are ignoring the omitted-variable effects on the estimate of the variance. Ability is certainly an important variable in any model estimating performance, and its omission will lead to an upward bias in the estimate of the variance. The consequence is that t ratios will be too low, and the results of hypothesis tests too pessimistic. Other studies investigating test structure and student performance use a chi-square test of significance, which requires the assumption that the two groups receiving a treatment are relatively homogeneous; they differ primarily in that, in this case, they receive tests that have a different structure (Bresnock, Graves, and White 1989). Although it is true that the law of large numbers makes it reasonable to assume no dif-

ferences in student ability between the two groups, with our data this becomes a testable rather than a maintained hypothesis. In our investigation, student grade point average (GPA) was included as a measure of ability. Grade point average at the time of the course was not easily obtainable, but the most recent GPA of the student was obtained without much difficulty. This substitution was not troublesome because the GPA was being used as a proxy for ability, and the most current GPA was 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. We expected GPA to be positively related to exam performance.

As an alternative to using grade point averages as a measure of human capital, we also estimated the model with a series of dichotomous variables to separately identify all students in the class. This approach allowed the intercept of the estimated regression equation to differ for each student.

Because grade point averages depend in part on major, a dummy variable for being an engineering major (ENGINEER) was included in the model. Our experiences have led us to conclude that, regardless of GPA, engineering stu-

TABLE 1
Test Score Regression Results

Variable	Model				
	1	2	3	4[a]	5[a]
CONSTANT	41.69	40.14	30.84	70.42	52.75
	(28.63)	(22.24)	(5.73)	(39.53)	(5.57)
CHRONO	0.13	0.13	−0.73	0.44	0.38
	(0.18)	(0.18)	(0.26)	(0.65)	(0.56)
GPA	11.86	11.68	15.23		
	(22.34)	(20.26)	(7.46)		
CLASS		0.55	3.90		10.25
		(0.98)	(1.75)		(2.41)
MALE		0.97	5.71		−4.81
		(1.25)	(1.91)		(0.73)
ENGINEER		2.19	−1.43		−4.58
		(1.63)	(0.19)		(0.76)
CHRONO*GPA			0.35		
			(0.33)		
CLASS*GPA			−1.27		
			(1.54)		
MALE*GPA			−1.86		
			(1.64)		
ENGINEER*GPA			1.48		
			(0.56)		
F	249.63	101.78	57.24	5.85	6.03
R^2	.3323	.3373	.3409	.5958	.6089
$\bar{R}^2$	.3310	.3339	.3349	.4940	.5080
df	1,006	1,006	1,006	1,020	1,015

Note: Figures in parentheses are absolute values of t ratios.
[a] Models 4 and 5 include 205 student dummy variables. Information on those variables is available on request from the authors.

dents do very well in microeconomics, perhaps because of their quantitative backgrounds. The class (freshman, sophomore, junior, or senior) of the student at the time of the course (CLASS) was also included. This variable might reflect background or maturity, which we believed to be positively related to exam performance. To account for additional interaction effects, we interacted GPA with CHRONO, MALE, CLASS, and ENGINEER.

EMPIRICAL RESULTS

The regression results are presented in Table 1. Several specifications of the model were estimated. Our primary measure of ability, GPA, was highly significant, which is evidence that the results of other studies may be questionable because of omitted-variable bias. However, the results indicated that even when accounting for differences in ability, scores were not significantly higher for students who were given chronologically ordered exams. The result was present when GPA or individual student dummy variables were used. The explanatory power of the regression models was greatly improved by the inclusion of the student ability variables. The adjusted R^2 values increased to between .332 and .611, compared with values between .006 and .047 in comparable models, estimated by Gohmann and Spector, that did not include any measure of ability.

CONCLUSION

Previous studies have generally found that question order does not significantly affect student exam scores. We have extended the work done recently by others in this area by estimating a more fully specified model of student performance. We also found that the order of questions has no statistically significant effect on exam scores, even after including variables that reflect differential student-specific human capital characteristics. Our results reaffirm findings in previous studies and suggest that instructors need not worry that some students have an unfair advantage because of the version of the exam they receive. In any case, newer computer software makes scrambling exams an increasingly easy precaution to implement.

REFERENCES

Bresnock, A. E., P. E. Graves, and N. White. 1989. Multiple-choice testing: Question and response position. *Journal of Economic Education* 20 (Summer): 239–45.

Gohmann, S. F., and L. C. Spector. 1989. Test scrambling and student performance. *Journal of Economic Education* 20 (Summer): 235–38.

Lumsden, K. G., and A. Scott. 1987. The economics student reexamined: Male-female differences in comprehension. *Journal of Economic Education* 18 (Fall): 365–75.

Taub, A. T., and E. B. Bell. 1975. A bias in scores on multiple-form exams. *Journal of Economic Education* 7 (Fall): 58–59.