THE IMPACT OF QUESTION ORDER ON MULTIPLE CHOICE EXAMS ON STUDENT PERFORMANCE IN AN UNCONVENTIONAL INTRODUCTORY ECONOMICS COURSE

Paul Kagundu and Glenwood Ross[1]

Abstract

We investigate the effect of question order on multiple-choice exams on students' performance in an unconventional introductory economics course. The course is an introduction to the *global economy* and comprises elements of principles of economics, introductory international trade and introductory international finance. The tests in two sections of the course were administered in four versions. On one of the versions, multiple-choice questions are ordered according to the order in which course material was offered, while questions on the other versions are randomly scrambled. Our empirical analysis reveals no statistically significant effect of question order on students' grades.

Key Words: question order, multiple-choice exams, students' performance, unconventional economics course

JEL Classification: A22

Introduction

In an effort to reduce cheating on multiple-choice tests, instructors often use several different versions of the same test. In many cases, a sequenced version of the test is accompanied by several randomly scrambled versions of the test. The questions on the sequenced test are presented in a logical fashion, based on the order in which the course material was offered. As a result, students may glean cues and prompts from a prior question or set of questions in a logical sequence to lead them to a correct response. If this is the case, an unintended consequence of this effort to minimize cheating could be the introduction of a bias in favor of those students who receive the sequenced version of the test versus those who receive the scrambled versions.

A number of studies have taken a look at the importance of question order on multiple-choice tests in introductory economics courses. The courses examined in these studies have typically been principles of economics, principles of macroeconomics, and principles of microeconomics courses in which material is generally presented in a logical building block manner. This study attempts to extend the work of previous studies by examining the impact of

[1] P. Kagundu, Senior Lecturer of Economics, Department of Economics, Penn State University; G. Ross, Clinical Associate Professor of Economics, Department of Economics, Georgia State University.

question order on multiple-choice tests in a *Global Economy* course, an unconventional introductory economics class.[2]

The class is unconventional in the sense that (1) it is a hybrid course comprised of roughly equal elements of economic principles, introductory international trade, and introductory international finance; and (2) not all the material is presented in a typical building block sequential manner of most introductory economics courses. The content in the first third of the course consists of standalone topics whose understanding is not dependent on the mastery of any other topic in the section. This is where a select group of basic economic concepts that are needed in the remainder of the course are introduced. These concepts include definitions of economics, marginal analysis, opportunity cost, supply and demand, gross domestic product (GDP), inflation, and the production possibilities frontier (PPF). Each of these topics is independent of the others and can be presented in any order.

The remainder of the course is much more structured in that the topics covered in international trade and international finance build upon each other. One can conceive of the first part of the course as a collection of random topics, while the last two-thirds of the course is more logically sequenced. This dichotomy allows us to relate the importance of question order to the nature of course content. *A priori*, it would seem that question order on multiple-choice tests should matter more for logically ordered and related course content than for course content that is randomly ordered and unconnected.

The next section presents a brief review of the literature on question order and student performance on multiple-choice tests, highlighting our contribution to this literature. This is followed by the empirical framework, the data and a discussion of simple inferences. Then we present and discuss the regression results for the impact of question order on student performance, followed by a conclusion.

Literature Review

Several studies examine the impact of the question order of multiple-choice exams on student performance in introductory economics courses. To date, however, no general consensus has emerged. Some studies conclude that question order may indeed matter. In one of the first studies to investigate this topic, Taub & Bell (1975) developed a regression model that included a dummy variable to indicate whether a multiple-choice test was randomly ordered or sequentially ordered. This variable proved to be significant, indicating that students who completed randomly ordered tests scored about 1.4 points lower than students who completed tests on which the questions followed the order of topics in the textbook and lectures.

Carlson & Ostrosky (1992) also concluded that question order might matter. They analyzed four exams in a large microeconomics principles class in which each exam was administered in two versions — one in sequential content-order and one in random order. When each exam was analyzed individually, Carlson and Ostrosky (1992) were unable to reject the null hypothesis that the means and the variances on the sequenced and random versions of the exam were equal. When the data were pooled, however, the null was rejected. This, combined with

---

[2] The *Global Economy Course* is part of the core curriculum at the university. It is designed for the economic novice and, as such, has no prerequisites. Most students who elect to take the *Global Economy* course are not economics majors. Since this class serves as a *"gateway course"* to the discipline, however, a number of students subsequently decide to become economics majors.

the fact that the mean score of the content-ordered exam exceeded the mean score of the scrambled exam in three of the four cases, led them to conclude that the mean level of performance may be higher on the content-ordered exam than the scrambled exam.

Doerner & Calhoun (2009) found mixed results. They conducted an experiment on three large introductory economics classes—one principles of microeconomics class and two principles of macroeconomics classes. The data were stratified between male and female students and three versions of a multiple-choice final exam were administered. One version was randomly ordered, a second was sequentially ordered, and the third version had questions that were ordered in a reverse sequential format. Their results indicated that females benefited from both sequentially ordered and reverse sequentially ordered exams. Question order did not matter for males.

Still other studies claim that question order doesn't matter. Bresnock, Graves & White (1989) examined the results of three multiple-choice tests given to a large section of undergraduate principles of economics class. They concluded that question order didn't matter, but that the pattern or distribution of correct answer responses on the multiple-choice tests did impact the degree of test difficulty. Gohmann & Spector (1989) randomly distributed sequenced and randomly ordered multiple-choice final exams to a large principles of macroeconomics class. In several specifications of linear regressions to determine whether exam performance could be attributed to the scrambling of exam questions, they found that question order had no significant effect on exam scores.

More recently, Sue (2009) focused her analysis on a small class setting. Heretofore, most of the other studies that investigated the role of question order on multiple-choice exams in economics courses focused on large classes. She analyzed data for three sections of principles of microeconomics and three sections of principles of macroeconomics. The average class size was less than 30 students. In regression analysis she was unable to reject the null hypothesis that "scrambling the content order of questions in a multiple choice test does not affect student performance on the test."

We contribute to this literature by examining the impact of question order on student's performance in an unconventional introductory economics course on the global economy.

Empirical Framework

To estimate the effect of question order on multiple-choice tests on students' grades we compare the average performance of a random group of students who took a version of the test with questions ordered in the order in which the course content was covered in class to that of a control group with scrambled versions of the test. We refer to the version of the test that is ordered consistent with the lecture coverage of the course content as the "sequenced" version. The other versions of the test are, together, referred to as the "scrambled" version. Students taking the "sequenced" version of the test are our *treatment* group, while those taking the "scrambled" version are our *control* group. As such, we estimate a linear regression equation of the following form.

$$G_{ij} = \alpha_0 + \alpha_1 V_{ij} + \alpha_2 X_{ij} + \varepsilon_{ij} \qquad (1)$$

The left-hand side variable, G_{ij}, represents student i's grade on the multiple-choice section of test j (j = 1, 2, 3, 4). The variable V_{ij} represents the version of the test (sequenced or scrambled) and X_{ij} represents a number of control variables including proxies for the student's academic ability and/or prior knowledge of the subject. The last term in equation (1) denotes the idiosyncratic random error term. The major variable of interest is V_{ij}, defined as a binary dummy variable equal to 1 if student *i* took the *sequenced* version on test *j*, and 0 otherwise. Therefore, a positive and statistically significant estimate of α_1 suggests a bias in favor of students taking the *sequenced* version. In other words, student performance benefits from ordering questions consistent with class coverage of the course material.

The Data

The data were collected in two sections of an introductory course on the *Global Economy* at a large public university in the southeastern United States. The data were collected in the Fall 2011 semester. The two sections (015 and 035) were taught by the same instructor. Table 1 presents descriptive statistics of key variables by section, test, and version. Students in each section took a total of four tests during the course of the school semester. The last of the four tests is the comprehensive final exam.

Table 1: Descriptive Statistics by course section, exam, and test version

Section	**Test**	**Test Version**	**Obs**	**Mean**	**Standard Deviation**
Econ2100-015	1	Sequenced	16	77.06	11.23
	1	Scrambled	46	69.09	12.64
	2	Sequenced	16	81.31	16.37
	2	Scrambled	46	76.65	16.62
	3	Sequenced	15	62.20	10.88
	3	Scrambled	44	69.50	15.09
	4 (Final)	Sequenced	14	67.21	9.96
	4 (Final)	Scrambled	46	70.54	13.07
Econ2100-035	1	Sequenced	17	70.94	14.36
	1	Scrambled	54	71.11	13.63
	2	Sequenced	17	79.88	13.66
	2	Scrambled	52	77.40	14.31
	3	Sequenced	19	68.58	12.30
	3	Scrambled	46	65.35	16.09
	4 (Final)	Sequenced	18	65.39	12.61
	4 (Final)	Scrambled	49	69.90	12.65

Tests consisted of a multiple-choice section and a problem-type question. Only students' scores on the multiple-choice part of the test are used in our estimations. Each test was composed of four versions, one "sequenced" and three "scrambled".

On each test, about one-quarter of students taking the test had the "sequenced" version. In Table 1, there does not seem to be a statistically significant difference in average scores between those students who took the "sequenced" version and those that took the "scrambled" versions.

In fact, average scores are higher for the "scrambled" versions in two of the four tests. Formal tests for the difference in mean scores for the two sub-samples are presented below.

Note that the distribution of test versions in the two classes was random. That is, the test version taken by student *i* depended entirely on where in the classroom the student sat. If students maintained their seating positions in the classroom throughout the semester, it is conceivable that the same students received the "sequenced" version of the test in all or most of the 4 tests. This does not bias our estimates if students initially randomly self-selected their seating positions. But, if students' self-selected seating positions are correlated with ability, assignment of test versions based on seating positions biases our results.

To exclude this possibility, Table 2 presents sample probabilities associated with student *i* receiving a "sequenced" version of the test on more than one test. These probabilities are compared with joint probabilities under a purely random assignment. For example, only 5 students (row 1, columns 1 & 2) had the "sequenced" version of the test on both the first two tests. This translates into a probability of 0.0189 (row 1, column 4) that student *i* had the sequenced version on both test 1 and test 2 compared to the joint probability of the same event of 0.0352 (row 1, column 5). Going down to the bottom of the table, we show the probabilities are even small that student *i* got the "sequenced" version on more than two exams. We are, therefore, confident that our regression estimates are not biased by the assignment of the test versions to students.

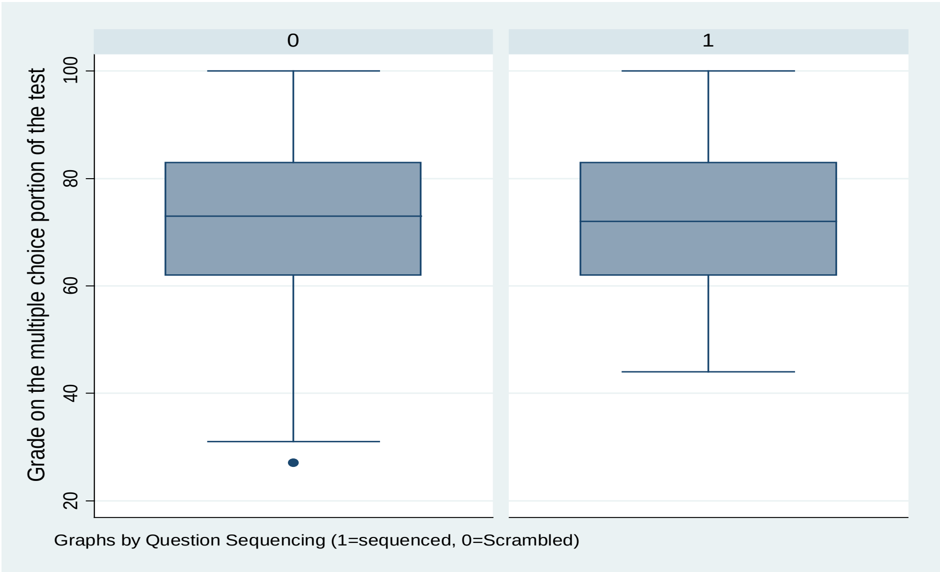
Table 2: Distribution of Test Versions (Sequenced versus Scrambled)

Tests	Number of Students with Sequenced Version	Total Number of Students	Probability	Joint Probability (purely random distribution of exams)
1 and 2 only	5	264	0.0189	0.0352
1 and 3 only	7	257	0.0272	0.0352
1 and 4 only	10	257	0.0389	0.0352
2 and 3 only	8	255	0.0314	0.0352
2 and 4 only	9	255	0.0353	0.0352
3 and 4 only	9	248	0.0363	0.0352
1, 2 and 3 only	1	388	0.0026	0.0117
2, 3, and 4 only	2	379	0.0053	0.0117
1, 3, and 4 only	2	381	0.0052	0.0117
1, 2, 3, and 4	1	515	0.0020	0.0039

Simple Inference

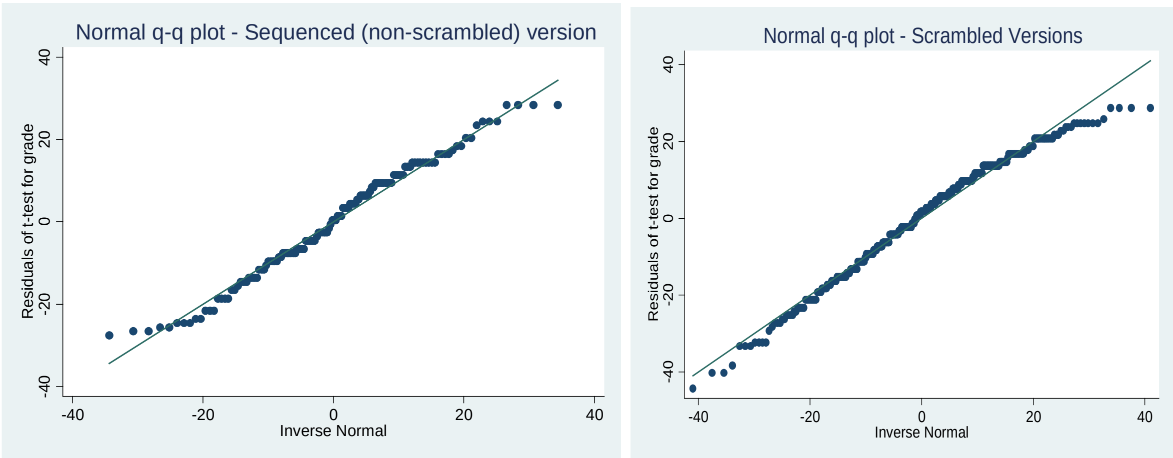
Here we use sample statistics to test for possible differences in population parameters between the "sequenced" and the "scrambled" versions of the tests. In particular, we conduct an F-test for differences in variances as well as a t-test for differences in mean scores. These tests assume that the two samples ("sequenced" and "scrambled") are independent and drawn from normally distributed populations with equal variances. We begin with a simple visual test in the form of a boxplot for grades in the two samples. The boxplot provides an informal test of the independent samples assumption.

Figure 1: Grade on the Multiple Choice Portions of the Tests



From Figure 1, the two samples do not seem to differ much in regard to students' grades. Further, the distributions for both groups seem symmetric enough to justify a t-test for difference in mean scores between the two populations. Nevertheless, we check for normality using the *normal quantile* plot of *residuals* of "grade" (G_{ij}). The appropriate residuals here are computed as the difference between the observed grade on the multiple-choice portion of the tests (G_{ij}) and the group-specific mean grade (group ≡ "sequenced", "scrambled"). The normality assumption is satisfied if the quantiles of the residuals are linearly related to the quantiles of the normal distribution. Figure 2 below presents the *normal quantile* plots for the "sequenced" sub-sample, and the "scrambled" sub-sample.

Figure 2: Normal Quantile Plots



From Figure 2, we conclude that the points on the *normal quantile* plots (representing residuals of "grade") are very close to the straight line. Therefore, it is very plausible that both samples ("sequenced" and "scrambled") are from normally distributed populations. We now proceed to conduct F-tests for differences in variances and independent samples t-tests.

Difference in variances

We test the null hypothesis that the two samples are drawn from populations with equal variances against the alternative that population variances are not equal. That is, $\sigma_1^2 - \sigma_2^2 = 0$

Table 3: **F-tests for difference in Variances (Sequenced versus Scrambled) by Course Section and Test**

(a) Section 015

	Version	n	f	DF	F-critical value	Pr(F>f)
Test1	Scrambled	46	1.2668	45, 15	2.5650	0.6357
	Sequenced	16				
Test 2	Scrambled	46	1.0312	45, 15	2.5650	0.9989
	Sequenced	16				
Test 3	Scrambled	44	1.9253	43, 14	2.6618	0.1835
	Sequenced	15				
Final	Scrambled	46	1.7242	45, 13	2.7601	0.2861
	Sequenced	14				

(b) Section 035

	Version	n	f	DF	F-critical value	Pr(F>f)
Test 1	Scrambled	54	0.9	53, 16	2.4635	0.7386
	Sequenced	17				
Test 2	Scrambled	52	1.0986	51, 16	2.3995	0.8751
	Sequenced	17				
Test 3	Scrambled	46	1.7104	45, 18	2.3635	0.216
	Sequenced	19				
Final	Scrambled	49	1.0067	48, 17	2.4115	0.9623
	Sequenced	18				

against the alternative that $\sigma_1^2 - \sigma_2^2 \neq 0$.[3] In this case, σ_1^2 and σ_2^2 are the population variances for the "scrambled" versions of the tests and the "sequenced" version of the tests respectively. In all the tests reported in Table 3, we fail to reject the null of equal population variances. Next, we proceed to test for differences in population means based on the assumption of normality and equal population variance.

Difference in Means

The test results presented in Table 4 show no statistical evidence of a bias in favor of students who took the "sequenced" version. The null hypothesis for the t-test is that there is no statistical difference in mean scores across versions: $\mu_1 - \mu_2 = 0$, where μ_1 and μ_2 denote population means for the "scrambled" and the "sequenced" versions respectively. On the other hand, the alternative hypothesis is that mean scores on the "sequenced" version are higher than those on the scrambled versions ($\mu_1 - \mu_2 < 0$). The t-tests fail to reject the null hypothesis in all cases except one – test 1 in section 015. Overall, the t-tests suggest that students that took the scrambled versions of the tests performed at least as well as those that took the "sequenced" version.

Regression Analysis

Tests based on mean scores and sample variances may not fully exclude the likelihood that the order of questions on the test has an effect on students' performance. A linear regression analysis that controls for other possible determinants of students' performance is a more effective way to tease out the impact of any given factor on students' grades while holding other factors constant. We estimate several regression specifications based on equation (1). Our baseline specification takes the following form:

$$Grade_{ij} = \alpha_0 + \alpha_1 Attend_{ij} + \alpha_2 Seq_{ij} + \alpha_3 Priorecon_i + \alpha_4 Section015_i + \varepsilon_{ij} \quad (2)$$

Again, $Grade_{ij}$ denotes student *i*'s grade on the multiple choice section of test *j*. $Attend_{ij}$ denotes the proportion of class sessions attended by student *i* in which material covered on test *j* was covered in class lectures. In other words, student *i*'s attendance during the first weeks of the semester (before test 1) is matched with the student's grade on the multiple-choice section of test 1.[4] Seq_{ij} is a dummy variable equal to 1 if student *i* took the "sequenced" version of test *j*, and 0 elsewhere. This is our variable of interest. Equation (2) also includes a dummy variable, $Priorecon_i$, to control for the effect of prior knowledge of the subject on performance. This dummy variable is equal to 1 if student *i* had taken any college-level economics courses before enrolling for the *Global Economy* course.

---

3 The test statistic is given by the ratio of the two sample variances $\left(\frac{S_1^2}{S_2^2}\right)$. S_1^2 and S_2^2 are the sample variances. We reject the null hypothesis if $\frac{S_1^2}{S_2^2} \geq f_{0.05, n_1-1, n_2-1}$.

4 To further clarify on the attendance variable, we should emphasize that this variable is constructed separately for each test. For example, if a student missed the first 2 weeks of the semester but attended the rest of the class sessions during the semester, her attendance corresponding to test 1 would be about 0.5 (50 percent) while her attendance corresponding to the other three tests would be 1 (100 percent). Please also see the variable description in the appendix.

Table 4: T-tests for difference in Means (Sequenced versus Scrambled) by Course Section and Test[5]

(a) Section 015

	Version	n	Mean	t	DF	T-critical value	Pr(T<t)
Test1	Scrambled	46	69.09	**-2.2329**	**60**	**1.6706**	**0.0146**
	Sequenced	16	77.06				
Test 2	Scrambled	46	76.65	-0.9699	60	1.6706	0.168
	Sequenced	16	81.31				
Test 3	Scrambled	44	69.5	1.7226	57	1.672	0.9548
	Sequenced	15	62.2				
Final	Scrambled	46	70.54	0.8767	58	1.6715	0.8079
	Sequenced	14	67.21				

(b) Section 035

	Version	n	Mean	t	DF	T-critical value	Pr(T<t)
Test 1	Scrambled	54	71.11	0.0443	69	1.6673	0.5176
	Sequenced	17	70.94				
Test 2	Scrambled	52	77.4	-0.6265	67	1.6679	0.2665
	Sequenced	17	79.88				
Test 3	Scrambled	46	65.35	-0.7844	63	1.6694	0.2179
	Sequenced	19	68.58				
Final	Scrambled	49	69.9	1.2939	65	1.6686	0.8999
	Sequenced	18	65.39				

Data for the study were collected from two sections of the course, such that we control for possible section effects on students' performance that may arise from a difference in the class size, the time class meets, etc. The binary dummy variable, $Section015_i$, is equal to 1 if student *i*

---

5 The t-statistic is given by: $t = \frac{\overline{x_1}-\overline{x_2}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$; $S_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$; Where, $\overline{x_1}\ and\ \overline{x_2}$ are the sample means for the "scrambled" and "sequenced" samples respectively; n_1 and n_2 are the sample sizes; and s_1^2 and s_2^2 are the sample variances. We reject the null hypothesis if: $t-statistic \leq -t_{0.05,n_1+n_2-2}$.

was enrolled in section 015 of the course, and is equal to 0 elsewhere. As is typical, the last term on the left-hand side of equation (2) denotes the usual idiosyncratic error term to capture the effect of purely random unobservable factors.

We also run additional specifications of equation (2) that include various proxies for the students' academic ability as well as student classification (freshman, sophomore, junior, and senior). Measures of students' academic ability prior to enrolling for the course include grade point average (GPA) prior to enrolling for the course and SAT scores.

Regression Results

The first set of results, presented in Table 5, consists of OLS estimates by test. Given the structure of the course as described earlier in the paper, test 1 covered an overview of key economic concepts – defining economics, marginal analysis, opportunity cost, demand and supply, GDP, and inflation. Since these topics may not appear to the novice to be connected, we postulated that the benefit of a "sequenced" version of the test would be smaller or nonexistent on test 1. On subsequent tests - where lecture topics were covered in a more logical sequence – we expect question sequencing to affect students' performance. We therefore run four specifications of regression equation (2) by test (test1, test2, test 3, and the final exam).

Table 5: Linear Regression Estimates by Exam

	Test 1 Grade	**Test 2 Grade**	**Test 3 Grade**	**Final Exam Grade**
Covariates				
Attendance	12.197	7.720	8.657	-0.841
	(8.768)	(6.602)	(7.012)	(3.465)
Sequenced	**3.679**	**1.011**	**-2.152**	**-2.698**
	(3.003)	**(2.797)**	**(2.900)**	**(1.990)**
Section015	3.132	-1.048	-0.882	-1.060
	(2.742)	(2.422)	(2.633)	(1.783)
Prior College Econ	7.090**	-0.104	-1.479	3.469*
	(2.877)	(2.647)	(2.864)	(1.890)
Prior tests average[6]		0.769***	0.614***	0.700***
		(0.095)	(0.106)	(0.077)
Constant	56.562***	10.812	16.451*	20.637***
	(8.592)	(6.902)	(9.075)	(4.743)
R-Squared	0.10	0.46	0.34	0.57
No. of Observations	90	90	86	84

Standard errors in parenthesis; ***: significant at 1 percent; **: significant at 5 percent; *: significant at 10 percent.

[6] "**Prior tests average**" represents the student's average grade on previous tests during the semester. For Test 2, prior tests average is simply the student's grade on Test 1. For Test 3, it is the student's average grade on Test 1 and Test 2. For the final exam, it's the student's average grade on Test 1, Test 2, and Test 3.

In column 2 of Table 5, we present estimates of equation (2) only for test 1. Based on our estimates, the order of questions on the test does not have a statistically significant influence on the students' performance on the multiple-choice section of the test. Among all the explanatory variables in column 1, only "prior college econ" (equal to 1 if a student had taken any college-level economics prior to enrolling for the Global Economy, and 0 elsewhere) had a significant impact on performance. On average, a student with prior college-level economics classes scored 7 percentage points higher on the multiple choice section of test 1 than those who had not taken a college-level economics course prior to enrollment in this class.

Estimates for the subsequent tests (tests 2, 3 and the final exam) show similar results. Common to all is the finding that the order of questions on tests (sequenced versus scrambled) does not influence students' performance.

In specifications for tests 2, 3, and the final exam, we include previous test average grades in the course as a proxy for ability. We see some consistency in performance. Students test 1 scores are a good predictor of their performance on test 2. We find similar results for test 3 and the final exam.

In Table 6, we present estimates of equation (2) with a pooled sample as opposed to test-specific subsamples. Also included are a number of proxies for ability that were not included in Table 5. First, our estimates from the pooled sample show no statistically significant impact of test question order (sequenced versus scrambled) on performance. Although class attendance seems to be highly correlated with student's performance, we cannot immediately infer causation based on our results reported in Table 6. In three of five regression specifications reported in Table 6, the coefficient on attendance is positive and statistically significant. It's important to note that when all the proxy variables for students' ability are added to the model (Model 6), the attendance variable is no longer statistically significant. This may suggest that attendance is correlated with ability and/or motivation. Specifically, attendance is likely positively correlated with ability in the sense that more able students tend to be more motivated to attend class relative to less able students. In this case, students who regularly attend class may perform better not just because of attendance but also because they are more able students. Therefore, the regression coefficients on the attendance variable reported in Tables 5 & 6 should be interpreted as a correlation rather than a causal effect. Our results in Table 6 notwithstanding, prior studies with varying degrees of statistical rigor have found a small to moderate positive effect of class attendance (Thatcher, et al. (2007); Romer (1993); Rodgers (2001))

Our proxies for students' academic ability are, for the most part, statistically correlated with students' performance. We include six such variables in the six specifications shown in Table 6. The data on these variables are self-reported by students on surveys administered by the authors. The first one is "*Below 2.5 GPA*". This is a dummy variable equal to 1 if a student's cumulative grade point average at the end of the previous school semester prior to taking the *Global Economy* class was less or equal to 2.5. The second variable, "*Above 3.5 GPA*" is defined similarly and is equal to 1 for students whose cumulative grade point average going into the *Global Economy* class was 3.5 or more. The reference group is the middle group – those with a cumulative GPA higher than 2.5 but less than 3.5. As expected, students coming into the class with lower GPA (2.5 or below) performed poorly relative to the reference group (2.5<GPA<3.5). On average, students with a GPA of 2.5 or below coming into the *Global Economy* class scored about 10 percentage points lower than the reference group. In contrast, students coming into the *Global Economy* class with a 3.5 or better GPA scored about 8 percentage points higher than the reference group (students with 2.5<GPA<3.5).

Table 6: Pooled OLS Estimates. The Dependent Variable is "Grade on the Multiple Choice Section of the test".

Covariates	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Attendance	9.000***	8.556**	5.806	8.622**	8.282**	4.431
	(2.327)	(3.628)	(3.704)	(3.648)	(3.560)	(4.032)
Section015	0.708	1.157	1.502	1.126	1.168	2.236
	(1.218)	(1.462)	(1.464)	(1.472)	(1.434)	(1.512)
Test1	0.044	0.060	0.018	0.053	0.125	1.352
	(1.779)	(2.046)	(2.033)	(2.049)	(2.007)	(1.993)
Test2	6.952***	6.920***	7.124***	6.913***	6.990***	8.552***
	(1.779)	(2.051)	(2.038)	(2.054)	(2.012)	(2.004)
Test3	-3.367*	-2.768	-2.678	-2.771	-2.793	-3.839*
	(1.765)	(2.058)	(2.046)	(2.061)	(2.019)	(2.026)
Sequenced	**-0.103**	**-0.859**	**-1.056**	**-0.843**	**-0.748**	**1.078**
	(1.400)	**(1.615)**	**(1.608)**	**(1.619)**	**(1.585)**	**(1.563)**
Freshman		3.631	4.814*	3.579	5.567**	4.111
		(2.735)	(2.729)	(2.751)	(2.731)	(2.735)
Sophomore		5.028*	5.575**	4.934*	7.060***	5.313**
		(2.672)	(2.650)	(2.717)	(2.676)	(2.625)
Junior		2.187	2.649	2.240	3.815	4.237
		(2.849)	(2.823)	(2.865)	(2.827)	(2.822)
GPA<2.5			-9.636***			-10.806**
			(3.305)			(4.475)
GPA>3.5						8.524***
						(1.797)
High School Econ				0.421		3.027
				(2.088)		(2.366)
Prior College Econ					5.833***	7.191***
					(1.535)	(1.658)
SAT<1200						2.561
						(2.422)
SAT>1600						7.914***
						(1.804)
Constant	62.794***	60.812***	63.039***	60.457***	57.353***	52.935***
	(2.092)	(4.256)	(4.272)	(4.611)	(4.273)	(4.990)
R-squared	0.10	0.10	0.10	0.10	0.14	0.41
No. of Observations	515	354	350	354	354	246

Notes: Standard errors in parenthesis; ***: significant at 1 percent; **: significant at 5 percent; *: significant at 10 percent.

The next pair of control variables captures prior exposure to economics. The first of these two variables, "High School Econ", is a binary dummy variable equal to 1 if the student took any economics classes in high school and 0 elsewhere. The second of the two variables, "Prior College Econ" is as defined earlier in the paper. Model 4, 5, and 6 in Table 6 presents estimates of these variables' effect on students' performance. Exposure to economics in high school had no statistically significant effect on students' performance. However, students with prior college-level exposure to economics performed better, on average, by 5 – 7 percentage points.

Finally, we added students' self-reported SAT scores as proxies for ability. The variable labeled "*Below 1200 SAT*" is a dummy variable equal to 1 if a student reported a total SAT score of 1200 or below and 0 elsewhere. Similarly, "*Above 1600 SAT*" is a dummy variable equal to 1 if a student reported a total SAT score of 1600 or higher. The reference group in this case includes students who reported SAT total scores between 1200 and 1600. Estimates of model 6 reported in Table 6 suggest no statistically significant difference in performance between those reporting a total SAT score of 1200 or less and the reference group (1200<SAT score<1600). However, those reporting a total SAT score of 1600 or higher outperformed the rest by an average of about 8 percentage points, all else constant. This result is not unexpected.

Further, we test if the regression coefficient estimates based on the "sequenced" portion of the sample are the same as those based on the "scrambled" portion of the sample. This is done using an F-test commonly referred to as the "Chow test". In each of the regression specifications reported in Table 6, each explanatory variable is interacted with the "Sequenced" variable. The interaction variables are then included in the respective regression specifications as additional explanatory variables. A joint F-test is then performed on the interaction variables. The null hypothesis for the test is that the coefficients from the two subsamples are the same, which suggests that coefficients on the interaction variables are, jointly, not statistically different from zero. The alternative hypothesis suggests different regression coefficients from the two subsamples. The regression results and the corresponding joint test results are reported in Table 7 below.

Table 7: Regression Results with Interaction Variables and Joint F-tests

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Attendance	8.876**	8.735*	6.374	6.413	5.708	4.995
	(2.596)	(4.123)	(4.173)	(4.177)	(4.063)	(4.767)
Section 015	0.543	1.555	2.100	2.088	2.174	3.062+
	(1.410)	(1.702)	(1.714)	(1.716)	(1.669)	(1.770)
Test 1	-2.088	-2.507	-2.553	-2.549	-2.223	0.462
	(2.063)	(2.417)	(2.401)	(2.407)	(2.344)	(2.370)
Test 2	4.827*	4.610+	5.160*	5.162*	5.692*	8.169**
	(2.069)	(2.400)	(2.387)	(2.393)	(2.333)	(2.368)
Test 3	-4.118*	-4.357+	-3.831	-3.825	-3.052	-1.595
	(2.060)	(2.454)	(2.441)	(2.450)	(2.390)	(2.440)
Sequenced	-7.474	6.190	8.820	9.687	13.362	15.616
	(5.434)	(10.192)	(10.667)	(10.76)	(10.090)	(10.729)
Freshman		7.033*	8.158**	8.14**	10.716**	8.730**
		(3.094)	(3.088)	(3.111)	(3.079)	(3.137)
Sophomore		6.646*	7.223*	7.191*	10.291**	8.567**
		(3.026)	(3.003)	(3.062)	(3.028)	(3.010)
Junior		5.421	6.096+	6.105+	8.828**	8.517*
		(3.303)	(3.280)	(3.307)	(3.272)	(3.390)
GPA<2.5			-10.11**	-9.99**	-11.45**	-14.793*
			(3.581)	(3.341)	(3.262)	(5.745)
GPA>3.5						8.717**
						(2.091)
High School Econ				0.113		2.029
				(2.592)		(2.968)
Prior College Econ					7.221**	8.319**
					(1.803)	(1.931)
SAT<1200						3.377
						(2.791)
SAT>1600						8.436**
						(2.097)
Sequenced × Attendance	2.684	-0.197	-2.061	-2.572	-2.329	-0.189
	(5.872)	(8.686)	(9.321)	(8.770)	(8.470)	(8.938)
Sequenced × section 015	0.329	-1.988	-2.627	-2.436	-2.575	-4.612
	(2.790)	(3.293)	(3.288)	(3.382)	(3.205)	(3.435)
Sequenced × Test 1	8.247*	9.249*	9.444*	9.592*	9.050*	4.747
	(4.053)	(4.503)	(4.509)	(4.533)	(4.384)	(4.497)
Sequenced × Test 2	8.218*	7.886+	6.663	6.816	5.367	1.580
	(4.039)	(4.632)	(4.634)	(4.701)	(4.552)	(4.723)
Sequenced × Test 3	3.068	5.179	3.834	3.962	1.936	-5.727
	(3.978)	(4.469)	(4.486)	(4.539)	(4.444)	(4.597)
Sequenced × Freshman		-15.631*	-15.714*	-15.53*	-17.30**	-18.58**
		(6.562)	(6.562)	(6.553)	(6.420)	(6.413)

Sequenced × Sophomore		-8.641	-8.943	-8.654	-11.636+	-13.109*
		(6.520)	(6.462)	(6.579)	(6.349)	(6.410)
Sequenced × Junior		-13.747*	-14.692*	-14.71*	-17.52**	-17.10**
		(6.697)	(6.637)	(6.658)	(6.513)	(6.574)
Sequenced × (GPA<2.5)			0.693			9.094
			(9.553)			(9.662)
Sequenced × (GPA>3.5)						-0.567
						(4.075)
Sequenced × high School econ				-0.919		4.181
				(4.466)		(5.195)
Sequenced × Prior college econ					-3.558	-2.760
					(3.466)	(3.912)
Sequenced × (SAT<1200)						-0.039
						(5.735)
Sequenced × (SAT>1600)						-1.617
						(4.100)
Constant	64.248**	59.762**	61.352**	61.2**	56.601**	48.647**
	(2.302)	(4.807)	(4.794)	(5.315)	(4.819)	(6.114)
Observations	***515***	***354***	***350***	***350***	***350***	***246***
R-squared	***0.120***	***0.131***	***0.153***	***0.153***	***0.196***	***0.461***
F(k', n-k)	***1.43***	***1.52***	***1.38***	***1.38***	***1.62***	***1.46***
Prob>F	***0.211***	***0.148***	***0.197***	***0.197***	***0.11***	***0.127***

Standard errors in parentheses; ** p<0.01, * p<0.05, + p<0.1
k' denotes the number of restrictions under the null; *k* denotes the number of explanatory variables in the model (including the constant); *n* is the number of observations.

In all the regression specifications reported in Table 7, we fail to reject the null of equal coefficients at conventional confidence levels. In simple terms, the results presented in Table 7 do not suggest any differences in the behavior of the two groups in the sample (those that took the sequenced version of the tests versus those that took the scrambled versions of the tests).

Conclusions

We examined the impact of question order on multiple-choice tests on student performance in an unconventional introductory economics course. Our empirical estimates indicate that question order does not influence student performance. Therefore, instructors of introductory economic courses need not be concerned about introducing bias in multiple-choice exams by using scrambled and unscrambled tests. This finding reinforces the conclusions of earlier studies by Bresnock, Graves and White (1989), Gohmann and Spector (1989), and Sue (2009).

We also found no evidence that the structure of the course content influences the impact of question order on student performance. No systematic bias was found either when the course content consisted of unrelated standalone topics or when the course content was presented in a building block sequential manner.

Not surprisingly, our results show that academic ability does matter. For example, a student with a GPA of 2.5 or lower coming into the class will typically score about 11 percentage points below the course averages for those with GPAs between 2.5 and 3.5, and roughly 19 percentage points - several letter grades - below those of students with GPAs greater than 3.5, all else constant.[7]

Attendance is also positively correlated with performance. However, teasing out the causal effect of attendance requires dealing with potential endogeneity, since attendance may be driven by students' ability and motivation. Although we cannot conclusively confirm it, prior studies that focus on the role of attendance suggest a positive causal effect on academic performance. Encouraging students to attend class by providing incentives whenever possible is a likely worthwhile effort.

References

Bresnock, Anne E, Philip E Graves, and Nancy White. "Multiple-Choice Testing: Question and Response Position." *Journal of Economic Education*, 1989: 239-245.

Carlson, Lon J, and Anthony L. Ostrosky. "Item Sequence and Student Performance on Multiple-Choice Exams: Further Evidence." *Journal of Economic Education*, 1992: 232-235.

Doerner, William M., and Joseph P Calhoun. "The Impact of the Order of Test Questions in Introductory Economics." *Social Science Research Network.* April 2, 2009. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1321906 (accessed June 4, 2012).

Gohmann, Stephen F, and Lee C Spector. "Test Scrambling and Student Performance." *Journal of Economic Education*, 1989: 235-238.

Rodgers, J. R. *A panel-data study of the effect of student attendance on university performance.* Vol. 45. 3 vols. Australian Journal of Education, 2001.

Romer, David. *Do Students Go to Class? Should They?* Vol. 7. 3 vols. The Journal of Economic Perspectives, 1993.

Sue, Della L. "The Effect of Scrambling Test Questions on Student Performance in a Small Class Setting." *Journal for Economic Educators*, 2009: 32-41.

Taub, Allan J, and Edwaed B Bell. "A Bias in Scores on Multiple-Form Exams." *Journal of Economic Education*, 1975: 58-59.

Thatcher, Andrew, Peter Fridjhon, and Kate Cockcroft. *The Relationship between lecture attendance and academic performance in an undergraduate psychology class.* Vol. 37. 3 vols. South Africa Journal of Psychology, 2007.

[7] The university's grade scale is one with "pluses" and "minuses". We consider the "pluses" and "minuses" to be different letter grades. For example, A+, A, and A- are distinct letter grades.

Data Appendix

Table A1: Variable Description and Sources

Variable Label	Variable Description	Source
Grade (Dependent Variable)	Grade (out of 100) on the multiple choice section of the tests.	Instructor's grade book
Attendance	The proportion of classes attended before each test. For instance, if a student missed 2 of 8 class sessions before test1, attendance corresponding to the student's first test grade would be 0.75 (6/8). Attendance corresponding to the test 2 is based on class sessions after test1 and before test2. Attendance for test3 and test4 are similarly computed.	Instructor's attendance records.
Section015	This is a binary variable equal to 1 if a student was enrolled in course section 015, 0 elsewhere	Students' enrolment records.
Test 1	This is a binary variable equal to 1 for observations corresponding to test1, 0 elsewhere.	Instructor's grade book
Test 2	This is a binary variable equal to 1 for observations corresponding to test2 grades, 0 elsewhere.	Instructor's grade book
Test 3	This is a binary variable equal to 1 for observations corresponding to test3 grades, 0 elsewhere.	Instructor's grade book
Sequenced	This is a binary variable equal to 1 if student took the sequenced version of the test, 0 elsewhere.	Instructor's grade book
Freshman	This is a binary variable equal to 1 if the student is a freshman, 0 elsewhere.	Student self-reported information
Sophomore	This is a binary variable equal to 1 if the student is a Sophomore, 0 elsewhere.	Student self-reported information
Junior	This is a binary variable equal to 1 if the student is a Junior, 0 elsewhere.	Student self-reported information
Below 2.5 GPA	This is a binary variable equal to 1 if the student's GPA is 2.5 (out of 4) or below, 0 elsewhere.	Student self-reported information

Above 3.5 GPA	This is a binary variable equal to 1 if the student's GPA is 3.5 (out of 4) or above, 0 elsewhere.	Student self-reported information
High School Econ	This is a binary variable equal to 1 if the student took any economics courses in high school, 0 elsewhere.	Student self-reported information
Prior College Econ	This is a binary variable equal to 1 if the student took college-level economics course prior to the current course, 0 elsewhere.	Student self-reported information
Below 1200 SAT	This is a binary variable equal to 1 if the student reported SAT total score of 1200 or below.	Student self-reported information
Above 1600 SAT	This is a binary variable equal to 1 if the student reported SAT total score of 1600 or higher.	Student self-reported information

Table A2: Descriptive Statistics

Variable Name	Obs	Mean	Std. Dev.	Min	Max
Grade	515	71.369	14.516	27	100
Attendance	560	0.761	0.324	0	1
Section 015	560	0.464	0.499	0	1
Section 035	560	0.536	0.499	0	1
Sequenced Test Version	515	0.256	0.437	0	1
Freshman	360	0.311	0.464	0	1
Sophomore	360	0.367	0.483	0	1
Junior	360	0.233	0.424	0	1
Senior	360	0.089	0.285	0	1
GPA 2.5 or below	356	0.056	0.231	0	1
GPA between 2.5 & 3.5	356	0.539	0.402	0	1
GPA 3.5 or higher	356	0.236	0.425	0	1
Had Economics in HS	360	0.833	0.373	0	1
Had Economics in College	360	0.322	0.468	0	1
SAT score 1200 or below	256	0.172	0.378	0	1
SAT score between 1200 & 1600	256	0.438	0.497	0	1
SAT score 1600 or higher	256	0.391	0.489	0	1