

Scrambled Questions Penalty in Multiple Choice Tests: New Evidence from French Undergraduate Students

Morgan Raux
Marc Sangnier
Tanguy van Ypersele

Scrambled questions penalty in multiple choice tests: New evidence from French undergraduate students*

Morgan Raux^a Marc Sangnier^b Tanguy van Ypersele^c

December 2016

Abstract

This note evaluates the scrambled questions penalty using multiple choice tests taken by first-year undergraduate students who follow a microeconomics introductory course. We provide new evidence that students perform worse at scrambled questionnaires than at logically ordered ones. We improve on previous studies by explicitly modeling students individual skills thanks to a fixed effects regression. We further show that the scrambled questions penalty does not differ along gender but varies along the distribution of students' skills and mostly affects students with lower-intermediate skills.

KEYWORDS: Multiple choice tests, scrambled questions, student performance.

JEL CLASSIFICATION: A10, A20, A22.

*We thank staff and faculty members of Aix-Marseille University's Faculty of Economics and Management who provided us with the data used in this note. We are especially grateful to Marion DAVIS, Simon Granier, Véronique Marques and Alain Paraponaris.

^aAix-Marseille Univ. (Aix-Marseille School of Economics), CNRS, EHESS and Centrale Marseille.
Mail: morgan.raux@univ-amu.fr

^bAix-Marseille Univ. (Aix-Marseille School of Economics), CNRS, EHESS and Centrale Marseille.
Mail: marc.sangnier@univ-amu.fr

^cAix-Marseille Univ. (Aix-Marseille School of Economics), CNRS, EHESS and Centrale Marseille.
Mail: tanguy.vy@univ-amu.fr

1 Introduction

Scrambled questionnaires, i.e. questionnaires whose questions are arranged in some random order, are frequently used to prevent cheating by students. The question of whether scrambled questionnaires affect students performance as received lots of attention by scholars.¹ Most studies conclude that students perform worst when facing a scrambled rather than a logically ordered test. However, some studies fail to uncover any statistically significant difference in scores of students facing different types of tests. The former findings raise questions about the fairness of such questionnaires as students who are requested to complete them frequently seat next to fellows that face a non-random arrangement of the same set of questions. Conclusions of the second group of studies suggest that this is not so much of an issue. In this note, we provide new evidence that students perform worse at scrambled questionnaires.

We use multiple choice tests taken by first-year students who follow the microeconomics introductory course at Aix-Marseille University's Faculty of Economics and Management. We improve on previous studies by explicitly modeling students individual skills thanks to a fixed effects regression. The use of student fixed effects has two advantages. First, while some past studies do use students' observable characteristics to account for potential confounding factors, none of them explicitly account for students individual skills in a flexible way. Second, it makes more likely to detect the effect of scrambled questionnaires if this effect does actually exist but has a small magnitude relative to the overall variation of the data.

We further contribute to the literature by investigating whether the average scrambled questions penalty varies with gender or along the distribution of students' skills. Our estimates indicate that female and male students suffer similar penalties. In contrast, we find that low- and high-skill students are unaffected by the order in which questions are presented, while students with lower-intermediate skills appear to suffer the most from scrambled questionnaires.

The remainder of this note is organized as follows. Data are presented and analyzed in sections 2 and 3, respectively. Section 4 briefly concludes.

2 Data

At Aix-Marseille University's Faculty of Economics and Management, the microeconomics introductory course takes place during the very first term of students' curriculum. It is organized around lectures and tutorials. As for students' evaluation, students are requested to take two multiple choice questionnaires in addition to an intermediate exam and a final exam.

Questions used in multiple choice questionnaires are entered by the course's main instructors via an online application. When requested by the staff to prepare the next test, instructors—two of this note's authors—generally create questions by going through lectures slides, what mechanically creates tests whose questions follow a logical order. For instance, students are asked to find the definition of an indifference curve before

¹See Mitchell (1974), Taub and Bell (1975), Laffitte (1984), Pettit et al. (1986), Balch (1989), Bresnock et al. (1989), Chidomere (1989), Gohmann and Spector (1989), Carlson and Ostrosky (1992), Neely et al. (1994), Russell et al. (2003), Sue (2009), Vander Schee (2009), Khan et al. (2013), Kagundu and Ross (2015) and Ollennu and Etsey (2015) among others.

Table 1: Estimation of the scrambled questions penalty.

Dependent variable: Score at multiple choice questionnaire				
	(1)	(2)	(3)	(4)
Scrambled questionnaire	-0.312*** (0.113)	-0.086 (0.102)	-0.308* (0.167)	-0.324** (0.154)
Test fixed effects	Yes	Yes	Yes	Yes
Student fixed effects	Yes		Yes	Yes
Sample restriction			Females	Males
Observations	5,110	5,110	2,322	2,788
Students	2,677	2,677	1,228	1,449
Adjusted R ²	0.159	0.577	0.588	0.568

*** p<0.01, ** p<0.05, * p<0.1. White heteroskedastic standard errors in parentheses. OLS regressions. Each column presents estimates from a separate regression. All regressions include a constant term.

being asked whether two of such curves can cross each other. Once all questions have been entered, the application generates two additional alternative versions of the test by randomly re-ordering questions. The three versions of the test are then distributed to students who have 30 minutes to complete the exam. A third of them face the sequentially organized test, while the two other thirds face one of the randomly ordered version. Each test is made of 20 questions. A correct answer buys 1 point, an incorrect ones 0. Students are provided with their grade on 0–20 scale within the next days that follow the test.

We collected the 5,110 scores obtained by all the 2,677 students who participated to multiple choice tests from 2012–13 to 2015–16.²

3 Data analysis

In order to investigate whether or not students perform worse at randomly ordered tests, we estimate the following expression using ordinary least squares:

$$s_{ij} = \alpha + \beta \text{Scrambled questionnaire}_{ij} + \mathbf{I}_j + \mathbf{I}_i + \varepsilon_{ij}, \quad (1)$$

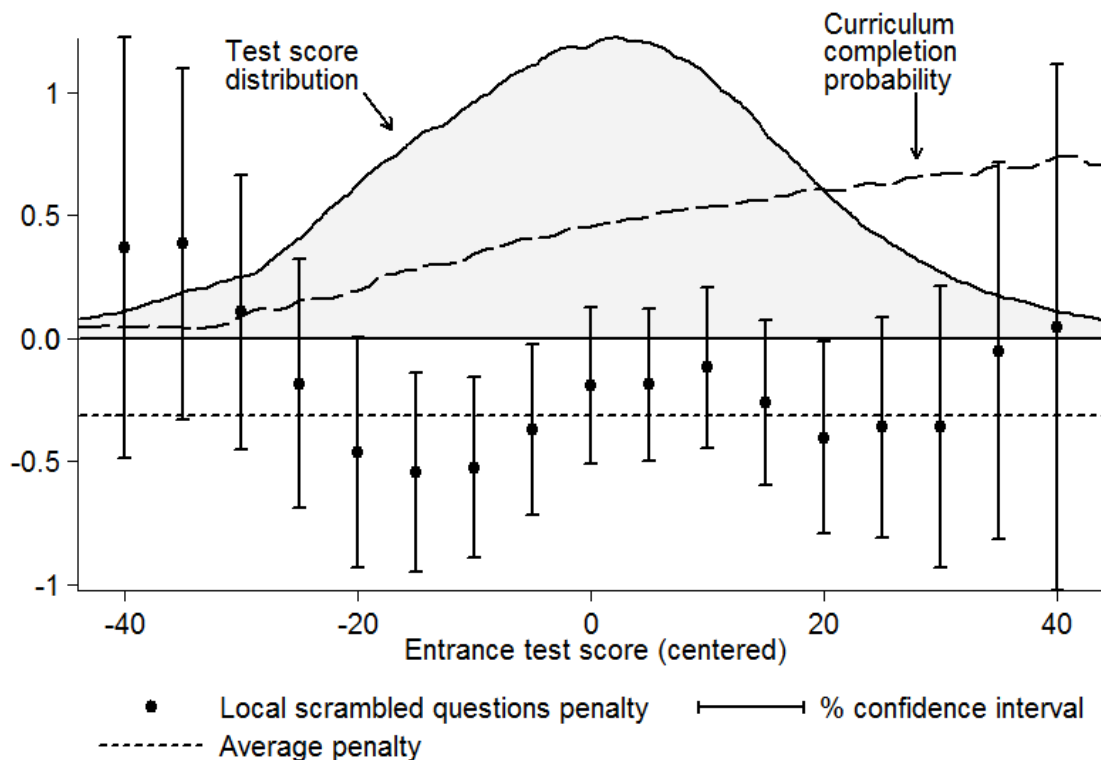
where s_{ij} denotes student i score at test j , $\text{Scrambled questionnaire}_{ij}$ is a dummy variable equal to 1 if student i faces a randomly ordered version of test j , \mathbf{I}_j and \mathbf{I}_i are vectors of test and student fixed effects, respectively, α is a constant term, and ε_{ij} is the error term.

As shown by column 1 of Table 1, coefficient β , that denotes the scrambled test average penalty, is found to be equal to -0.312 (p-value = 0.006). This suggests that students score about 1.5% lower when requested to complete a randomly ordered test rather than a sequentially ordered one. The presence of student fixed effects further ensures that this effect acts within individuals and does not result from lucky draws from the joint distribution of tests and students' skills.

Would we remove student fixed effects from equation (1), we would obtain an estimate

²The total number of observations is slightly different from two times the number of students for two reasons: (i) during the academic year 2012–13, students were requested to take three (rather than two) multiple choice questionnaires; (ii) some students occasionally don't show up at a test.

Figure 1: Scrambled questions penalty at different levels of students' skills distribution.



The average penalty corresponds to the full sample estimate of β in equation (1). Local scrambled questions penalties correspond to estimates of β over each interval $[l - 10, l + 10]$, where l moves along the universe of entrance score by steps of 5 points. Entrance test score have been harmonized across cohorts thanks to a simple regression of students individual scores on a set of academic year fixed effects. The curriculum completion probability is the average observed probability that a student with a given entrance test score registers successfully complete her first-year at Aix-Marseille University's Faculty of Economics and Management.

of β that is equal to -0.086 (p-value = 0.400) as illustrated by column 2 of Table 1. This stresses the importance of student fixed effects for the empirical identification and illustrate how failing to flexibly account for students' skills may lead to the erroneous conclusion that the average penalty cannot be considered to be different from 0.

Columns 3 and 4 of Table 1 display estimates of the scrambled question penalty for female and male students, respectively. The two coefficients are of similar magnitude. A formal test of equality of the two estimates provides us with a p-value of 0.479, what confirms that female and male students experience identical penalties.

While the magnitude of the effect we uncover is comparable to estimates provided by other scholars, our data further allow us to investigate whether the scrambled questions penalty varies with student's skills. We approximate skills thanks to entrance tests students are asked to take by the very beginning of the academic year at Aix-Marseille University's Faculty of Economics and Management.³ We then uncover the scrambled questions penalty at different levels of the test score distribution by estimating equation (1) over each interval $[s - 10, s + 10]$, where s moves along the universe of entrance scores by steps of 5 points.

³The entrance test consists in 100 questions that cover general skills such as reading, writing, basic mathematics and logic. This test is not selective and is only used to construct homogeneous tutorial groups.

Figure 1 plots the series of local β s together with the test score distribution. This figure makes clear that the previously estimated average scrambled questions penalty—represented by Figure 1’s dotted horizontal line—hides much heterogeneity. Despite estimates are less precise in areas with low density, low- and high-skill students do not seem to suffer from facing randomly ordered test questions (presumably for obviously different reasons). In contrast, students with lower-intermediate skills appear to experience statistically significant negative penalties that go up to -0.546 . The concentration of the scrambled questions penalty on these students makes it even more problematic for two reasons. First, they represent a high share of the population. Second, they face a high risk of not completing their first year as shown by the curriculum completion probability represented by the dashed line of Figure 1 that does not exceeds 0.5 for students with lower-intermediate skills.

4 Conclusion

This note provides additional evidence that students perform worse at scrambled rather than at logically ordered tests. We improve on the existing literature by using student fixed effects and by showing that individuals with lower-intermediate skills suffer more than others from the scrambled question penalty.

An immediate policy recommendation follows from our findings. Questionnaires scrambled at the question level should not be used in the same time as logically ordered ones. Following this recommendation while preventing cheating can be done by using questionnaires scrambled at the answer level, i.e. scrambled only within each question. Alternatively, if the latter proves insufficient to prevent cheating, using question level scrambled versions of the set of questions would warrant fairness across students. This would however lead all students to perform worse than they would have in front of a logically ordered questionnaire. This obstacle can only be alleviated by tuning the overall test difficulty or by exogeneously adjusting the mean performance.

References

- Balch, William R. “Item order affects performance on multiple-choice exams.” *Teaching of Psychology* 16, 2: (1989) 75–77.
- Bresnock, Anne E., Philip E. Graves, and Nancy White. “Multiple-choice testing: question and response position.” *The Journal of Economic Education* 20, 3: (1989) 239–245.
- Carlson, J. Lon, and Anthony L. Ostrosky. “Item sequence and student performance on multiple-choice exams: Further evidence.” *The Journal of Economic Education* 23, 3: (1992) 232–235.
- Chidomere, Rowland C. “Test Item Arrangement and Student Performance in Principles of Marketing Examination: A Replication Study.” *Journal of Marketing Education* 11, 3: (1989) 36–40.
- Gohmann, Stephan F., and Lee C. Spector. “Test scrambling and student performance.” *The Journal of Economic Education* 20, 3: (1989) 235–238.

- Kagundu, Paul, and Glenwood Ross. "The Impact of Question Order on Multiple Choice Exams on Student Performance in an Unconventional Introductory Economics Course." *Journal for Economic Educators* 15, 1: (2015) 19–36.
- Khan, Junaid Sarfraz, Saima Tabasum, Osama Mukhtar, and Maryam Iqbal. "The Effect on Student Performance of Scrambling Questions and Their Stems in Medical Colleges Admission Tests." *Journal of the College of Physicians and Surgeons Pakistan* 23, 12: (2013) 904–906.
- Laffitte, Rondeau G. "Effects of item order on achievement test scores and students' perception of test difficulty." *Teaching of Psychology* .
- Mitchell, William E. "Bias in writing objective-type examination questions." *Journal of Economic Education* 58–60.
- Neely, Darlene L., Frederick J. Springston, and Stewart J.H. McCann. "Does Item Order Affect Performance on Multiple-Choice Exams?" *Teaching of Psychology* 21: (1994) 44–45.
- Ollennu, Sam Nii Nmai, and Y.K.A. Etsey. "The Impact of Item Position in Multiple-Choice Test on Student Performance at the Basic Education Certificate Examination (BECE) Level." *Universal Journal of Educational Research* 3, 10: (2015) 718–723.
- Pettit, Kathy L., Kenneth G. Baker, and Laura D. Davis. "Unconscious biasing of student examination scores: A case of sequential versus random information retrieval." *Journal of Marketing Education* 8, 3: (1986) 20–24.
- Russell, Michael, Michael J. Fischer, Carol M. Fischer, and Kathleen Premo. "Exam question sequencing effects on marketing and management sciences student performance." *Journal for Advancement of Marketing Education* 3, 1: (2003) 1–10.
- Sue, Della L. "The effect of scrambling test questions on student performance in a small class setting." *Journal for Economic Educators* 9, 1: (2009) 32–41.
- Taub, Allan J., and Edward B. Bell. "A bias in scores on multiple-form exams." *Journal of Economic Education* 7, 1: (1975) 58–59.
- Vander Schee, Brian A. "Test item order, academic achievement and student performance on principles of marketing examinations." *Journal for Advancement of Marketing Education* 14: (2009) 23–29.