Multiple-Choice Exam Question Order Influences on Student Performance, Completion Time, and Perceptions

Terry F. Pettijohn II and Matthew F. Sacco

We conducted 2 studies to investigate undergraduate performance, perceptions, and time required in completing sequentially ordered, randomly ordered, or reverse ordered exams in introductory psychology classes. Study 1 compared the outcomes and perceptions of students ($N = 66$) on 3 non-comprehensive multiple-choice exams which were sequentially, randomly, and reverse ordered and Study 2 investigated the outcomes and perceptions of students on a multiple-choice final exam. We also measured perceived test difficulty, test anxiety, and understanding of material. There were no statistically significant differences between the scores on the different exams or the time required to complete the exam versions, but perceptions of difficulty were influenced by the version of the exam assigned. Professors should consider these findings when testing students.

To prevent cheating on exams, many professors will mix up the order of multiple-choice test questions from exam to exam without thought of the consequence the order may have on student exam performance and perceptions. Textbook companies even provide randomization options for preparing exams using electronic test banks to assist in this common practice. Some research suggests that the different exam versions can have a significant effect on student performance. According to Balch (1989), students score higher on multiple-choice exams when the questions are presented in the same order that the material was presented in lecture and text as opposed to when questions are randomly grouped by chapter or in completely random order. Providing an advantage to one group of students who take the sequential versus a random test question order exam is problematic and unfair. Balch suggests that sequentially ordered exams provide retrieval cues which may help with memory recall, consistent with

Terry F. Pettijohn II and Matthew F. Sacco, Mercyhurst College.

Correspondence concerning this article should be addressed to Terry F. Pettijohn II, Department of Psychology, Mercyhurst College, 501 E. 38th St., Erie, PA 16546-0001; email: terrypet@usa.com.

encoding specificity. The context of surrounding information used in encoding is utilized in information retrieval, and the sequential test question order provides a situation where context of encoding and retrieval are similar. In addition, Balch found that there was no significant difference in completion times between these versions of the exam. Other researchers have challenged this rationale and these findings.

Neely, Springston, and McCann (1994) conducted a three study follow-up to Balch (1989) in which student performance on sequential and random order multiple-choice question exams in an introductory psychology class were compared and the influence of test anxiety was also considered. The results of the three studies showed no significant difference between the sequential and random order multiple-choice question tests. However, the researchers did report a significant interaction such that high-anxiety students performed "somewhat better" on the sequential question order test and low-anxiety students performed "substantially better" on the random question order test. Similarly, Peters and Messier (1970) also found no differences in performance on sequential versus random question order multiple-choice tests in a class of graduate students studying research methods, and those students who reported

high levels of anxiety performed worse on the random question order test compared to the sequential question order test.

Perlini, Lind, and Zumbo (1998) further investigated the effects of item order and item difficulty on test performance in four studies. In experiment 2, the investigators found no advantage in student performance on sequenced chapter-order multiple-choice question tests over random or reverse question order tests. In experiment 3, researchers varied chapter question order and within chapter question order, but again found no performance differences between conditions. Item and chapter question arrangements were found to have little or no effect on test performance. Perlini, Lind, and Zumbo also arranged test questions with respect to item difficulty in their fourth study: easy-to-hard, hard-to-easy, or random. Again, there was no significant difference between difficulty arrangements. In a similar investigation, Laffittee (1984) created four versions of an introductory psychology multiple-choice test: easy to difficult by topic, easy to difficult across chapters, randomly within chapters, and randomly across chapters. Laffittee reported that presentation order had no effect on achievement test scores or student perception of test difficulty. Contrary to the notion that starting students off with easy questions to build confidence improves test scores, Skinner (1999) presented results suggesting that students may actually perform better if tests begin with difficult questions and students are given immediate feedback.

Past studies provide conflicting findings with respect to multiple-choice question order effects on student performance, although the majority of research supports the conclusion that there is no difference between sequential, random, chapter contiguous, and reverse test question orders. The concern for fairness in exam performance has driven these investigations, but the differences in time required to complete exams, student perceptions of test difficulty, understanding of course ma-

terial, and anxiety levels are also important consequences of alternative order versions of exams. With a clearer understanding of each of these variables, professors may choose to develop exams to meet the needs of their individual classes and students.

The current research was designed to provide a conceptual replication of previous research with respect to the effects of multiple-choice test question order on student performance and completion time. Specifically, in line with past findings, we predicted that there would be no significant differences in student performance and completion times between sequential (S), random (RA), and reverse (RE) question order exams. The reverse test question order condition has not been included in all previous investigations of test question order, although this condition provides an opportunity to determine the effect of testing the most recent information learned first and working backwards in a reverse sequential order.

The current research was also conducted to consider how different multiple-choice question order exams influence student perceptions about test difficulty, anxiety, and understanding of material. These perceptions have not been fully investigated in past studies. Specifically, we predicted that students would perceive sequential question order exams the least difficult, followed by reverse question order exams, and random question order exams would be perceived as the most difficult. Difficulty perceptions were expected to influence perceptions of anxiety and material understanding. We predicted that students would report the greatest post-exam anxiety in reaction to the random question order exams, followed by the reverse question order exams, and students would report the least anxiety in reaction to the sequential question order exams. Finally, we predicted that students would report the greatest material understanding for sequential question order exams, followed by reverse question order exams, and students would report the

least material understanding for random question order exams. We tested these general hypotheses in two studies.

Study 1

Study 1 was designed to test our general hypotheses using a repeated measures design by investigating student performance, completion time, and perceptions across three non-comprehensive unit tests.

Method
Participants

Sixty-six undergraduates, 17 men and 49 women, from two concurrent sections of Introduction to Psychology classes ($n = 32$, 34) taught by the first author (TFP) participated in this study. The majority of students were Caucasian (97%) and freshmen (77%). The class rosters were used to determine the overall composition of sex and class rank to describe the sample.

Materials

TFP wrote the exams, composed of 50 multiple-choice questions of varying difficulty, focusing on information presented in class lectures and from the assigned textbook chapter readings. We arranged the S question order condition to correspond chronologically to class information presentation and the RE question order condition completely reversed the sequential order for each of the three exams. For the RA question order condition, we randomly selected question numbers and arranged the test questions in this random order for each of the three exams.

After completing each exam, students completed a three question post-exam questionnaire. Students were asked to respond to "How difficult was this exam for you?" on a 5-point Likert scale (5 = *very difficult*, 4 = *difficult*, 3 = *average*, 2 = *easy*, 1 = *very easy*), "How anxious were you about this exam?" on a 4-point Likert scale (4 = *extremely anxious*, 3 = *moderately anxious*, 2 = *somewhat anxious*, 1 = *not at all anxious*),

and "After having completed the exam, how well do you feel you understood the material?" on a 5-point Likert scale (5 = *very well*, 4 = *well*, 3 = *average*, 2 = *somewhat*, and 1 = *not at all*).

Design and Procedure

We completed the investigation over one academic term within the context of ordinary Introduction to Psychology course requirements. We randomly assigned an equal number of participants to the S, RA, and RE conditions for the first exam. For exams two and three, we counterbalanced the conditions to insure that all participants completed all three test question order conditions and that the order of conditions was balanced in a repeated measures design. We covertly recorded how long it took each student to complete each exam. When students were finished with each exam, we distributed a post-exam questionnaire to assess student perceptions of test difficulty, anxiety, time spent studying, and understanding of test material.

Results and Discussion

Once all data was collected, we calculated mean exam scores for the S, RE, and the RA question order group for each exam. We conducted a repeated measures ANOVA (test question order condition: S, RE, RA) for the dependent variable exam score. Participants performed similarly on the different versions of the exams, $F(2, 63) = .89$, *ns* (see *Table* for descriptive statistics). Individual comparisons revealed no significant differences between test question order conditions on test score. In separate analyses, we found that test condition order and sex did not interact with test performance.

We converted the time required to complete the 50 question test into minutes and conducted a repeated measures ANOVA (test question order condition: S, RE, RA) for the dependent variable of time. Overall, participants took about the same amount of time to

complete the different versions of the exams, $F(2, 54) = .29$, *ns* (see *Table* for descriptive statistics). Individual comparisons revealed no significant differences between test question order conditions on time required to complete the test.

We averaged the perceived difficulty, exam anxiety, and material understanding ratings and conducted separate repeated measures ANOVA (test question order condition: S, RE, RA) for each dependent variable (see *Table* for descriptive statistics). Participants reported different perceptions of test difficulty regarding the test question order conditions, $F(2, 43) = 7.90$, $p < .001$, $\eta_p^2 = .15$. Individual comparisons revealed that participants perceived the RA test to be more difficult than the S test ($p < .001$) and the RE test ($p < .01$). The RA and the RE exam conditions were perceived to be similar in difficulty.

We found a marginally significant difference in test anxiety reported between the test question order conditions, $F(2, 43) = 2.43$, $p = .09$, $\eta_p^2 = .05$. Participants reported slightly more anxiety after completing the RA condition compared to the RE condition ($p = .09$) and the S condition ($p = .20$). The anxiety reported for the RA and RE conditions were similar. Overall, participants reported equivalent understanding of test material in the different test question order conditions, $F(2, 43) = 1.62$, $p = .20$.

Study 2

Study 1 provided initial support for our hypothesis about no differences in test performance and completion time between test question order conditions, but support for the student perception hypotheses was mixed. Study 2 was designed to test our general hypotheses investigating student performance, completion time, and perceptions across three versions of a comprehensive final exam. We added an additional question to the post-exam questionnaire to assess student study method for the final exam. We predicted an interaction between question order condition and method of study such that performance would be greatest when study method and question order condition matched. For example, those who studied sequentially would perform best when given a sequential question order test.

Method

Participants

The same 66 undergraduates who participated in Study 1 were also in Study 2.

Materials

TFP wrote the final exam, composed of 100 multiple-choice questions of varying difficulty, focusing on information presented in class lectures and from the assigned textbook chapter readings from the entire term. We created three versions of the final exam to correspond to the S, RE, and RA question order conditions we used with each of the individual exams. After completing the final exam, students completed a similar post-exam questionnaire as used in Study 1, with the addition of a fourth question: "Which of the following best explains your method of studying for this exam?" and we asked students to select one of three options which corresponded to the question order conditions (*start with beginning information and continue in chronological order, start with last information and work backwards, randomly move from topic to topic*).

Design and Procedure

Using the same Introduction to Psychology classes from Study 1, we randomly assigned an equal number of participants to the S, RA, and RE conditions for the final exam. Again, we covertly recorded how long it took each student to complete the final and when students were finished with the exam, we distributed a questionnaire to assess their perceptions of test difficulty, anxiety, time spent studying, understanding of test material, and method of study.

Results and Discussion

For the final exam, we calculated mean exam scores for the S, RE, and RA groups and conducted a 1-way ANOVA for the dependent variable exam score. Participants performed similarly on the different versions of the final exams, $F(2, 63) = 1.46$, $p = .24$ (see *Table* for descriptive statistics). Individual comparisons revealed no significant differences between final exam versions on test score.

We recorded the time required to complete the 100 question final exam in minutes and conducted a 1-way ANOVA comparing the S, RE, and RA versions of the final exam for the dependent variable of time. Overall, participants took about the same amount of time to complete the different versions of the final exam, $F(2, 63) = .15$, *ns*. Individual comparisons revealed no significant differences between test question order conditions on time required to complete the final.

We conducted separate 1-way ANOVAs comparing the test question order conditions on the dependent variables of perceived difficulty, exam anxiety, and material understanding ratings (see *Table* for descriptive statistics). Overall, participants reported similar perceptions of test difficulty when comparing the S, RE, and RA test question order conditions, $F(2, 63) = 1.87$, $p = .16$. However, the RA version of the final was perceived as slightly more difficult than the S version of the final ($p = .07$). We found no significant difference between the test question order conditions for anxiety ratings, $F(2, 63) = .20$, *ns*, and no significant difference between conditions for material understanding ratings, $F(2, 63) = .09$, *ns*.

In response to the additional question about how students studied, 54.54% indicated they studied in sequential order, 40.90% studied in reverse order, and only 4.50% of the students reported studying in random order. We conducted a 3 (test question order condition: S, RE, or RA) x 3 (method of study: sequential, reverse, or random) ANOVA for the dependent variable exam score. There were no main effects for test question order condition ($p = .50$) or method of study ($p = .44$), and there was no interaction effect ($p = .48$). Within the S final exam, participants who studied sequentially scored 1.76% higher than those who studied in reverse order, and within the RE final exam, participants who studied in reverse order performed 2.58% better than those who studied in sequential order. However, neither of these differences were statistically significant. Pairing students' method of study with a particular type of multiple-choice question order exam does not seem to significantly influence test scores.

Overall Discussion

As predicted, there were no significant differences between test performance across the sequential, reverse, and random question order exams in Study 1 or Study 2. Similarly, participants completed the S, RE, and RA ordered multiple-choice tests with no difference in completion times in Study 1 and Study 2. Support for our hypotheses concerning student perceptions of the different exam versions was mixed. Test difficulty perceptions generally supported our predictions, but anxiety perceptions varied from Study 1 to Study 2, and material understanding perceptions were not influenced by test version in either of the studies.

In Study 1, the RA question order exam condition was perceived to be more difficult than the S and the RE conditions, although there were no significant differences in difficulty perception between the RA and RE conditions. In Study 2, participants perceived the RA test question order final to be more difficult than the S final, but there were no significant differences between the other test question order conditions. In Study 1 and Study 2, the predicted order of perceived difficulty was found: S question order was rated as the least difficult, followed by the RE order, and the RA order was rated as the most difficult.

We found a marginally significant difference in post-exam anxiety reported between the test question order conditions in Study 1. Participants reported slightly more anxiety in the RA condition compared to the RE and the S conditions, but there was no significant difference between the RE and S conditions. In Study 1, the predicted order of reported anxiety was found: participants reported the lowest anxiety in the S question order condition, followed by the RE condition, and participants reported the highest anxiety in the RA condition. Study 2 found no differences in reported anxiety between test question order conditions, and the post-test anxiety ratings were actually highest for the S question order condition and lowest for the RA condition. Overall post-exam anxiety was higher in Study 2 than in was in Study 1, which makes sense because a final exam is

Table 1		
Studies 1 & 2: Mean Test Scores, Completion Times, and Ratings by Test Question Order Condition		
Variable	Study 1 Unit Exams	Study 2 Final Exam
Test Score (% correct)		
S	69.66 (12.57)	71.14 (8.79)
RE	68.48 (12.14)	66.30 (11.69)
RA	70.16 (12.48)	71.43 (13.08)
Completion Time (minutes)		
S	24.98 (5.17)	58.60 (12.14)
RE	25.38 (4.43)	56.85 (7.37)
RA	25.42 (4.95)	57.47 (12.38)
Test Difficulty		
S	3.53 (.54)	3.76 (.63)
RE	3.79 (.66)	3.87 (.55)
RA	3.96 (.60)	4.10 (.54)
Test Anxiety		
S	2.79 (.81)	3.52 (.68)
RE	2.85 (.76)	3.43 (.73)
RA	3.02 (.77)	3.38 (.81)
Material Understanding		
S	3.15 (.75)	3.33 (.58)
RE	2.93 (.80)	3.26 (.62)
RA	2.93 (.82)	3.33 (.73)

Note. $N = 66$ participants. S = sequentially ordered test questions, RE = reverse ordered test questions, and RA = randomly ordered test questions. Means presented for the unit exams category are the average of three non-comprehensive unit tests. The unit exams were comprised of 50 multiple-choice questions each and the comprehensive final exam was comprised of 100 multiple-choice questions. Participants rated test difficulty and material understanding using a 5-point Likert scale and participants rated test anxiety using a 4-point Likert scale. Higher ratings indicate greater perceptions of test difficulty, test anxiety, and material understanding.

more important, and worth more points, than the unit exams. The high levels of anxiety reported after completing the final exam provided little room for differences between test question order conditions. The additional number of test questions on the final exam may also have increased anxiety as there were more opportunities to miss questions and there may have been a greater tendency to consider how performance would impact final course grade.

Participants did report slightly higher material understanding after completing the S test question order condition in Study 1, but overall participants reported equivalent understanding of test material in the different test question order conditions in both studies. Perceptions of how well students believed they understood course material do not appear to be influenced by the sequence of the multiple-choice test questions.

We recognize there are inherent limitations with our studies. We only investigated two classes of Introductory Psychology from a single college and we recognize that a number of factors influence student performance and perceptions. Individual test question item difficulty was not assessed in these studies, so it is possible that the placement of easy and difficult test questions may have influenced student performance, completion time, and perceptions. We only used 3 versions of the tests (S, RA, and RE), and additional versions of each could have been introduced to note differences. We only assessed test difficulty, test anxiety, and material understanding once immediately after taking each test. It would be interesting to see how these perceptions may change with time.

The null results of the current studies support previous research (Neely, Springston, & McCann, 1994; Perlini, Lind, & Zumbo, 1998) suggesting multiple-choice question order does not significantly influence test performance or test completion times. However, test question order variations do lead to differences in certain student percep-

tions. These perceptions may be an important byproduct of different test versions, and may be useful in different situations. For example, students who experience intense test anxiety may be given sequential order exams to help reduce test anxiety. Indeed, in pre-test assessments, past research (Neely, Springston, & McCann, 1994; Peters & Messier, 1970) found that high-anxiety students tended to perform better on S ordered exams than on RA ordered exams. Our post-exam anxiety assessments showed that participants reported greater anxiety in response to the RA test question order condition compared to the S condition in Study 1. Overly anxious students, students with learning disabilities, or students with other special circumstances may benefit from sequential test question order.

As a practical matter, these results show that professors can continue to utilize several versions of the same multiple-choice questions on exams and the sequencing of these questions will not lead to differences in student performance or completion times. However, if professors want students to believe course tests are more or less difficult, they may choose the appropriate test question order option to achieve their goals. Perceptions may also influence how much emphasis students place on studying for a course and professor evaluation. Future studies may further investigate these potential applications.

References

Balch, W. R. (1989). Item order affects performance on multiple-choice exams. *Teaching of Psychology, 16*(2), 75-77.

Laffittee, R. G. (1984). Effects of item order on achievement test scores and students' perceptions of test difficulty. *Teaching of Psychology, 11*(4), 212-214.

Neely, D. L., Springston, F. J., & McCann, S. J. H. (1994). Does item order affect performance on multiple-choice exams? *Teaching of Psychology, 21*(1), 44-45.

Perlini, A. H., Lind, D. L., & Mumbo, B. D. (1998). Context effects on examinations: The

effects of time, item order and item difficulty. *Canadian Psychology, 39*(4), 299-307.

Peters, D. L., & Messier, V. (1970). The effects of question sequence upon objective test performance. *Alberta Journal of Educational Research, 16*(4), 253-265.

Skinner, N. F. (1999). When the going get tough, the tough get going: Effects of item difficulty on multiple-choice test performance. *North American Journal of Psychology, 1*(1), 79-82.

Notes
1. Portions of this research were presented at the 11th Annual American Psychological Society Teaching Institute, Chicago, Illinois, May 2004.