
FACULTY FORUM

Introductory Psychology Student Performance: Weekly Quizzes Followed by a Cumulative Final Exam

R. Eric Landrum
Boise State University

Students in an introductory psychology course took a quiz a week over each textbook chapter, followed by a cumulative final exam. Students missing a quiz in class could make up a quiz at any time during the semester, and answers to quiz items were available to students prior to the cumulative final exam. The cumulative final exam consisted of half the items previously presented on quizzes; half of those items had the response options scrambled. The performance on similar items on the cumulative final was slightly higher than on the original quiz, and scrambling the response options had little effect. Students strongly supported the quiz a week approach.

The introductory psychology course is a popular course for nonpsychology majors completing general education (or core) requirements and a required course for nearly all undergraduate psychology majors nationwide (Perlman & McCann, 1999). In my previous approach to teaching the course, I divided the content into “units” covering two to four chapters each, followed by a major exam, with the course containing five or six units and exams. This format led to large reading assignments for students (often more than 100 pages). Given that researchers have found that introductory students retain little after the semester is over (Rickard, Rogers, Ellis, & Beidleman, 1988; VanderStoep, Fagerlin, & Feenstra, 2000), I began thinking about alternative approaches to teaching the course. I wanted to redesign the course to reduce student anxiety and enhance student performance, so I changed my teaching and testing strategy dramatically.

I implemented a teaching approach that covered one chapter per week with a quiz at the end of each week. At the end of the course, I gave students a cumu-

lative final exam. In an effort to reduce anxiety, I also implemented a lenient and generous make-up policy, such that I allowed students to make up quizzes at any time during the semester prior to the last day of class. Students knew from the beginning of the course that from each 20-item weekly quiz, 10 items would appear again on the cumulative final exam; students could also retrieve (from teaching assistants) a printout with their quiz answers and the correct answers. Of the items appearing on the cumulative final exam, half of the items (75) had scrambled response options (a, b, c) compared to the original quiz item. Although software programs are certainly capable of scrambling response options (e.g., Harnisch & Rotheroe, 1986), to my knowledge there is no information available about the change in item correctness (proportion of students answering an item correctly) when response options are scrambled in a test-retest situation.

Some of the inspiration for making the course changes came from Weimer (2002) and her work on learner-centered teaching; other researchers have tested different ideas. For instance, Grover, Becker, and Davis (1989) found that students preferred a frequent testing program (chapter by chapter) over a unit testing program (four chapters). Although student performance was similar for both approaches, only students in the frequent testing program indicated that they would choose that option again. Grabe (1994) found that whether students had one or more attempts at taking exams earlier in the course had no effect on cumulative final exam performance. Kahn (2000) found that students who missed required exams scored lower on comprehensive final tests and had more overall class absences.

I was interested in answering the following questions: (a) What is the relation between a lenient (i.e., generous) make-up quiz policy and student scores? (b) How was overall quiz performance related to cumulative final exam performance? (c) What is the effect of scrambling the response options for half of the cumulative final exam items?

Method

Participants

General psychology students ($N = 253$) enrolled in my Fall 2003 course were the participants in this study. Of the 3,795 possible quizzes (15 quizzes \times 253 students), students completed 3,385 quizzes during the scheduled quiz time (89.2%). Students completed 249 quizzes as make-up quizzes (7.4%); 161 missed quizzes were not made up (4.2%).

Materials

I created a 20-item multiple-choice quiz with three response options (a, b, and c) for each chapter every week of the course. Each week covered one textbook chapter from Lahey (2004). Students completed quizzes on Scantron bubble sheets; the following week students could pick up printouts of their quiz performance indicating their answers and the correct answers. At the end of the semester students completed a cumulative final exam. The items for the cumulative final exam appeared in the same sequence as the topics appeared during the course.

Procedure

Students completed weekly quizzes in class. I told students at the beginning of the semester that the quiz items would comprise the cumulative final exam items, but that some of the response options would be scrambled.

Results and Discussion

First, I examined make-up quiz performance versus quiz performance of students taking the weekly quiz at the scheduled time. Because make-up quizzes were hand-scored, these quizzes were easily identifiable. Second, I compared average quiz performance scores with performance on the cumulative final exam.

Lastly, I analyzed item correctness statistics of the original quiz items compared to the repeated cumulative exam items, attempting to detect the effect (if any) of response option scrambling.

Make-Up Quiz Performance Versus Regularly Scheduled Quiz Performance

I conducted a 15 (weekly quiz) \times 2 (quiz type: make-up vs. in class) ANOVA to examine quiz performance. There was a significant main effect of quiz, $F(14, 3355) = 8.40$, $p < .001$. Quiz 5 performance was unusually low, and Quizzes 1, 2, 4, and 7 exhibited high overall scores. There was a significant main effect of quiz type, $F(1, 3355) = 6.01$, $p < .02$. Overall, students taking make-up quizzes scored significantly higher ($M = 15.3$, $SD = 3.4$) than students taking the quiz at the regularly scheduled time ($M = 14.7$, $SD = 3.0$).

There was also a significant interaction between the weekly quiz and quiz type, $F(14, 3355) = 2.58$, $p < .005$ (see Table 1). The pattern of change over time did not appear to be stable. In 9 of 15 quizzes, students in the make-up condition performed better; in the remaining 6 of 15 quizzes, students taking the quiz in class performed better. This unpredictable pattern of quiz performance may be due to the varying difficulty of chapter topics covered each week.

Average Quiz Performance Compared to Cumulative Final Exam Performance

I calculated an average quiz score for each student and compared it with each student's cumulative final exam score. There was a significant correlation between average quiz score and cumulative final exam score, $r(222) = 0.65$, $p < .001$. I then divided students, according to their quiz score averages, into thirds (top, middle, bottom) and then compared their quiz score ranks with cumulative final exam performance to test the notion that "those who have the most to gain, gain the most." After converting average quiz score and cumulative final exam scores into percentages, I performed a 3 (top, middle, bottom) \times 2 (percent quiz score, percent final exam score) repeated measures ANOVA to detect differential changes across test performance dependent on the ranks. As expected, the rank-ordered groups did differ on final exam performance (top third, $M = 95.4$, $SD = 5.5$; middle third, $M = 86.8$, $SD = 7.7$; bottom third, $M = 78.6$, $SD = 12.7$), $F(2, 214) = 220.58$, $p < .001$. For all groups, there was an overall increase in percentage correct from

Table 1. Means and Standard Deviations for Quiz Scores by Week and Quiz Type

Quiz Week	In class		Make UP		In Class <i>N</i>	Make Up <i>N</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
1	16.07	2.61	17.88	2.89	231	9
2	17.15	2.59	16.22	3.49	229	9
3	14.48	2.65	16.33	3.70	226	12
4	15.38	2.66	17.66	2.40	221	9
5	12.97	2.93	11.88	4.26	222	9
6	13.36	3.44	14.75	4.03	221	12
7	16.98	2.49	14.77	3.53	222	9
8	14.39	2.75	16.18	3.62	202	22
9	13.89	3.07	14.83	2.41	210	12
10	15.42	2.61	15.04	3.28	196	23
11	13.13	2.93	14.07	3.76	191	28
12	15.87	1.95	15.21	2.80	186	33
13	14.03	2.88	15.25	3.75	189	27
14	14.18	3.01	14.04	2.92	193	25
15	14.56	3.00	15.40	3.10	197	10

quizzes ($M = 74.0, SD = 9.8$) to the cumulative final exam ($M = 86.9, SD = 11.4$). However, there was also a significant interaction between quiz rank and change from quiz percentage correct to final exam percentage correct, $F(2, 214) = 6.89, p < .002$. The lowest third of quiz scorers increased from an average of 62.8% correct ($SD = 4.8$) to an average of 78.6% correct on the final exam ($SD = 12.7$). Middle-ranked quiz scorers increased from 74.2% quiz performance ($SD = 2.3$) to 86.8% final exam performance ($SD = 7.7$), and top-ranked quiz scorers increased from 85.0% correct on quiz performance ($SD = 4.0$) to 95.4% correct on final exam performance ($SD = 5.5$). Tukey's post-hoc tests indicated that all of the quiz-final exam paired means were significantly different from one another, $p < .05$.

Effects of Response Option Scrambling on Changes in Item Correctness

The 150 items on the cumulative final exam had an initial average item correctness of .81 ($SD = 0.13$). That is, students answered these items correctly 81.0% of the time when they appeared on weekly quizzes. When these items appeared again on the cumulative final exam (with scrambled response options), item correctness increased to .87 ($SD = 0.09$). This change was significant, $t(149) = -9.49, p < .001$. Thus, students answered the reoccurring multiple-choice items more accurately on the cumulative final exam (average 87.4% correct) than on the original 15 quizzes (average 81.0% correct).

Did response option scrambling have any effect on changes in item correctness? I compared performance on the cumulative final exam to performance on the original quiz and calculated a change score for each of the 150 items. A positive change score indicated that the cumulative final exam item was easier to answer than the original quiz item. Items that kept the identical response option order from original quiz to cumulative final exam changed in item correctness an average of 0.067 ($SD = 0.08$), indicating that the cumulative final exam items were easier, answered on average 6.7% more accurately than previously. Items that had their response options scrambled from the original quiz to the cumulative final exam changed in item correctness an average of 0.060 ($SD = 0.81$), meaning that the scrambled items were also easier on the cumulative final exam, answered on average 6.0% more accurately than previously. Although the growth in increased performance for scrambled items was not as high for those with intact response options, this difference was not significant, $t(149) = 0.47, ns$. Scrambling the response options did not make cumulative final exam questions significantly harder; performance improved for both types of items (original and scrambled).

Conclusions

I was surprised that scores on the make-up quizzes were significantly higher than those on the in-class quizzes. Given the rush to take make-up quizzes at the

end of the semester, I expected that students did not prepare well and crammed these quizzes in at the last minute. Perhaps the additional time allowed students to prepare better; I also must consider that because previous quiz answers were available, students shared quiz results. In the future, I will use a different make-up strategy.

No matter what a student's quiz percentage ranking (top third, middle third, bottom third), student scores increased when comparing the change from overall quiz percentage to overall final exam percentage. However, the significant interaction indicated that those students with the most to gain (the lowest third) did gain the most. When comparing these percentages, students in the top third experienced a 10.4% growth in scores; the middle third, 12.6% growth; and the bottom third, 15.8% growth. The growth in the top third, however, may be limited in part to a ceiling effect (this group scored 95.4% correct on their cumulative final exam).

When comparing the same item correctness from original quizzes to cumulative final, students performed significantly better on the cumulative final exam (87.4% correct) compared to the original quiz items (81.0% correct). Given the changes that I made to my introductory course (e.g., one chapter and one quiz every week), I expected that students might not like this teaching approach. This particular course met Monday, Wednesday, and Friday at 8:40 a.m., and I gave a quiz every Friday at the end of class. About midway through the semester on a nonquiz day, I asked students attending class that day ($n = 165$) if I should continue using the quiz every week approach. Of those in attendance, 92.7% replied yes, 5.5% replied no, and 1.8% replied with a "don't know" or were nonresponsive.

I encourage other faculty members to experiment with their classes and try new approaches. I thought I was taking a substantial risk in redesigning my course, and I thought that the change might hinder student performance and student opinion might be negative.

The results of this study indicate that I still have as much to learn from my students as they do from me.

References

- Grabe, M. (1994). Motivational deficiencies when multiple examinations are allowed. *Contemporary Educational Psychology, 19*, 45–52.
- Grover, C. A., Becker, A. H., & Davis, S. F. (1989). Chapters and units: Frequent versus infrequent testing revisited. *Teaching of Psychology, 16*, 192–194.
- Harnisch, D. L., & Rotheroe, D. (1986). *ITEMBANK: An item bank package for the IBM PC*. Urbana, IL: Institute for Child Behavior and Development. (ERIC Document Reproduction Service No. ED298131)
- Kahn, I. (2000). Students' performance on makeup examinations and other evaluations. *Psychological Reports, 87*, 747–749.
- Lahey, B. B. (2004). *Psychology: An introduction* (8th ed.). Boston: McGraw-Hill.
- Perlman, B., & McCann, L. I. (1999). The structure of the psychology undergraduate curriculum. *Teaching of Psychology, 26*, 171–176.
- Rickard, H. C., Rogers, R., Ellis, N. R., & Beidleman, W. B. (1988). Some retention, but not enough. *Teaching of Psychology, 15*, 151–152.
- VanderStoep, S. W., Fagerlin, A., & Feenstra, J. S. (2000). What do students remember from introductory psychology? *Teaching of Psychology, 27*, 89–92.
- Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*. San Francisco: Wiley.

Notes

1. I appreciate earlier comments from the Editor and the three reviewers.
2. Send correspondence to R. Eric Landrum, Department of Psychology, Boise State University, 1910 University Drive, Boise, ID 83725-1715; e-mail: elandru@boisestate.edu.