Poster Presented at the April (2009) 37th Annual Western
Pennsylvania Undergraduate Psychology Conference. Moon Township, PA.

Pattern Guessing on Multiple-Choice Exams: Is Your Guess As Good As Mine?[1]

Maria Olausson
Robert Morris University

Studies have shown that various factors influence multiple-choice exam performance. The present study examined the influence of students' ability to identify patterns in multiple-choice exams on their final performance. If students have an expectation that answers will be distributed randomly, then a response pattern that is not perceived as random could yield a change of answer from right to wrong. To test this hypothesis three exam versions consisting of 40 multiple-choice questions were constructed: (1) Long pattern, (2) short pattern, and (3) control (random pattern). It is predicted that test scores will vary as a function of the degree to which the expectation for randomness is violated.

Introduction

Multiple-choice examination-format is one of the most frequently used testing formats. It is an easy way to test the knowledge of students in a large group. However, researchers (Balch, 1989; Bresnock, Graves & White, 1989; Brown, Schilling and Hockensmith, 1999; Gaskins, Forte, Wood & Riley, 1996; Higham and Gerrard, 2005; McClain, 1983) have been concerned with the validity of this examination format. It appears that there are various factors (e.g., item-order, number of response options, changing answer, perception on changing answer, test-taking strategies) that influence the quality of multiple-choice testing.

The most obvious relationship between students and exam performance is with regard to the quality of the student. McClain (1983) identified a relationship between the level of course performance and test-taking behavior of students that suggested "A" students possess superior reading skills and are better at verbalizing their thoughts than lower-performing students. McClain examined the multiple-choice test-taking strategies of "A," "C," and "F" students. Students were measured on performance score, answers read per question, number of anticipated answers, critiques of incorrect answers per question, number of questions initially skipped. Findings showed that "A" students use different strategies than "C" and "F" students. Better-performing students' behaviors appeared to be more thorough when contemplating alternative answers than lower-performing students. Clearly, then, students generate strategies to assist them on exams.

Among the most commonly cited strategies reported anecdotally among students is that the first guess is usually the best guess. Nonetheless, students do occasionally change their answers. Several researchers (e.g., Gaskins et al., 1996; Shatz & Best, 1987) have specifically examined the reasons students provide for justifying changing their answers. Shatz and Best hypothesized that when guessing was the reason for changing an answer, the outcome would not be as beneficial as when an answer was changed for other reasons (e.g., mismarking, misreading, clue in a later question). Sixty-five students were given a 62-items multiple-choice exam and asked to mark their changes and provide reasons for their changes after the exam. As hypothesized, the results showed that the most detrimental changes were made when students guessed.

Gaskins et al. (1996) also examined the reasons students change responses and when students benefited the most from changing answers. Gaskins et al.'s outcomes converged on the results of Shatz and Best (1987). That is, answer changes based on guessing resulted in the least beneficial outcome. Answer changes based on other strategies tended to result in favorable outcomes on performance (cf. Geiger, 1996; Skinner, 1983). The benefits of changing answers

[1] Address correspondence to: Stephen T. Paul, Ph.D., 6001 University Blvd., Moon Township, PA 15108-1189, or via email at: paul@rmu.edu.

on multiple choice exams was examined in depth by Haase, Riley, Dunn, and Gaskins (1992) who reviewed 22 studies and found a consistency of beneficial outcomes. Nonetheless, students tend to persist in their belief that their initial answers are more likely to be the right ones. This perspective only appears to be true when answer changes are the result of guessing. Therefore, identification of the cues that lead to guessing may be valuable in better understanding, and possibly reducing, students' reliance on this poor exam strategy.

One approach to identifying the cues likely to trigger answer changing may be derived from the work of Higham and Gerrard (2005). These researchers found that when students encounter confusing response alternatives they adopt the belief that changing their responses will not benefit them. In this instance, such a belief precludes test-takers from correcting their mistakes. However, as the number of response alternatives increases on a multiple-choice exam, the likelihood that students will guess increases as well. Consequently, the more alternatives students have to choose from, the more incorrect responses they get (Rosenthal & Rubin, 1989). Based on these findings, the challenge would be to construct relatively few, but clearly articulated response alternatives that do not diminish the quality of the exam (Delgado & Prieto, 1998). This is especially challenging given that exposure to incorrect response alternatives puts students at-risk for the negative suggestions effect (Remmers & Remmers, 1926; Roediger & Marsh, 2005). That is, if student are not given feedback from multiple-choice examinations they tend to mix false knowledge with real knowledge.

Brown, Schilling and Hockensmith (1999) investigated the negative suggestion effect by examining the consequences of students' exposure to misinformation during recognition (e.g. multiple-choice, true-false) and recall (cued-recall) testing. Exposure to misleading information hindered subsequent testing. Recall testing showed to be worse of than recognition in each condition (i.e., non-interpolated, interpolated). They concluded that memory impairment results from a reduction in confidence. This outcome suggests that students may make detrimental changes on multiple-choice exams (e.g., cumulative final exams) due to having retrieved misleading information from previous exams.

To minimize cheating, teachers sometimes use alternate test-forms to discourage students from looking at neighboring exams. Balch (1989) and Bresnock, Graves and White (1989) have shown that the outcome of an exam is dependent on the arrangement of response items on that exam. Specifically, Balch examined 404 general psychology students' performance on multiple-choice exams using three different exam formats. Each exam format (sequential, chapter contiguity, random) contained 75 questions and was assigned to students at random for the final exam. The sequential exam presented items in the order in which material was presented in the textbook and lectures. In the contiguity exams, items from same chapter were grouped together, however not sequentially. The random exam format presented the questions in an entirely random arrangement. The results showed that performance was best on the sequential exams relative to both the contiguity and random exams.

Bresnock et al. (1989) examined the effects on performance of altering the order and placement of questions and responses. Bresnock et al. created three exams (two midterms with 35 items each, one final exam with 70 items). The first exam set consisted of two jumbled versions, which compared students performance when exams were jumbled and when they were set in order of classroom lectures (i.e., answers distributed evenly among A, B, C, and D). Second exam set looked at examination performance and failing to hide the correct answer by generating two midterm versions with one version having an abnormal response distribution (i.e., higher percentage of A and D correct responses then B and C). The other version had just a slightly higher percentage of A and B correct responses than C and D. Lastly, the final exam compared a test with a high percentage of A responses to a test with high percentage of D responses. Bresnock et al. (1989) found no significant results from the first or second exam. However a more thorough investigation of the second experiment has suggested that students may have a tendency to select A responses over D responses, because A responses appear first and saves time compared to D responses. The final exam showed significance for asymmetrical response order, resulting in student presented with the exam containing a higher number of A

responses did better than student presented with the exam containing higher number of D responses. This indicated that altering response patterns (i.e., asymmetrical response pattern) changed the level of difficulty. Bresnock et al. (1989) suggest that altering the test versions to eliminate cheating should be done carefully.

The research by Bresnock et al. (1989) suggests that strategies students might employ during exam taking are to assess the relative proportion and distribution of the response alternatives. While Bresnock and colleagues examined the relative representation of response alternatives, no study has yet examined the relative distribution of response alternatives. Anecdotal reports from students suggest that they expect to see a random distribution of responses throughout their exams. To the extent that this expectation is confirmed or violated could have direct consequences with regard to exam performance. Therefore, students' ability to discriminate random from presumably non-random patterns has important implications for exam construction.

Regarding the human capacity for discriminating random from non-random patterns, Wasserman, Young, and Cook (2004) compared humans versus animals (i.e., pigeons, baboons). Wasserman et al. (2004) argued that a preference for, and therefore the ability to recognize randomness in the environment is an adaptive quality among humans and other animals. Essentially, the argument is that stability in the environment requires different responses compared with those to unstable or changing environments. Survival often depends on the ability to discriminate between these two environmental conditions. It seems clear that humans are capable of recognizing randomness in their environments. However, responses to perceived randomness (or non-randomness) are likely to vary with environmental expectation. In the case of multiple-choice exams, it is likely that students expect that there should be no recognizable underlying pattern to the order of correct responses. That is, the response pattern should be "random".

The present study examined the relationship between various degrees of patterned responses and students' test performance on multiple-choice examinations. To the extent that the response patterns represented on a multiple-choice exam violate students' expectancies, it is predicted that exam performance will suffer.

Method

Participants

Participants were 192 traditionally aged male and female college students from Robert Morris University (RMU) located in Moon Township, Pennsylvania. All participants were recruited from introductory level psychology classes (N = 116) and upper-level psychology classes (N = 78). Furthermore, students were volunteers and some earned extra credit for participating.

Design

The study used a simple one-way design in which the independent variable, Exam-Type, had three levels (i.e., long-pattern, short-pattern, random-pattern) and was manipulated between subjects. The dependent variable was the percent correct score earned on the exam.

Materials

A 32-item general psychology exam was developed from the test-bank of a popular psychology text (Wood, Wood, & Boyd, 2006). There were four response options for each item, and the arrangement of correct answers was such that each option occurred eight times (i.e., eight "A" answers, eight "B" answers, etc.). There were three versions of the exam. The random-pattern version resulted in no obvious underlying correct response pattern (order was determined randomly). That is, there were never more than two repetitions of any single response alternative. The short-pattern version ensured that correct responses alternated as ABCD continuously throughout the exam. For the long-pattern, correct responses resulted in the first eight items all having answer "A" while the second eight items used "B" as the correct answer, and so on. All items were presented in the same order for all participants. The only changes made were to the order of correct response alternatives.

Procedure

The thirty-two-question test was distributed to introductory level general psychology classes and upper-level psychology classes. All students were required to record their responses on a scantron sheet (included in the exam packet). The three

exam-types were distributed randomly within each of the classes in which the study was conducted.

Results

A one-way analysis of variance was conducted on mean exam scores for all three conditions. The results indicated a significant main effect of exam type, $F(2,189) = 3.29$, $p < 0.05$.

The analysis revealed that random pattern exams resulted in better performance than specific pattern exams, as shown in Table 1.

Table 1. Main effect of exam type.

Condition	N	Mean	Standard Deviation
Long (AAAA)	**64**	**41.9**	**11.0**
Short (ABCD)	**64**	**45.9**	**12.9**
Random (CBDA)	**64**	**47.1**	**12.6**

Discussion

The results from the present study confirm the prediction that incorporated patterns in multiple-choice exams have an effect on the performance score. The more random patterned the exam appears the better participants did on the exam. This indicates that students have an expectation that there will be no underlying pattern on multiple-choice exams, therefore when presented with a pattern it goes against their expectation for random pattern and they either change their answers or select an answer that is consistent with their pattern expectation rather than the correctness of the response. The greater the deviation from perceptions of "random" the answer-patterns represent the more likely students are to modify their response strategies. When these response strategies are employed, performance tends to decline.

The relatively poor overall performance exhibited is troublesome. It is not clear whether this was due to the surprise nature of the test, the motivation of the students taking the test (no benefit or penalty associated with performance), or some other factor (e.g., particularly difficult questions, etc.). However, despite the relatively low performance, significant differences in performance were still observed.

An obvious implication of the present results is that teachers should take care not to employ automatic test-generators that produce response patterns that may be perceived as "non-random" by students. A corollary to this would be that students should be discouraged from employing strategies that might mislead them into selecting responses based on irrelevant information (such as response patterns).

Given the relatively poor performance found on the exams in the present study suggests an interesting direction for future research. Specifically, well-prepared (better performing) students might be less likely to employ potentially detrimental test strategies such as answer counting, guessing, answer-pattern matching, etc. In which case, the presence of an underlying answer pattern might benefit these students when they perceive that one or more of their answers violates the pattern. Students who perform poorly to begin with, might not be able to recognize an underlying pattern due to the degree to which their wrong answers has degraded the pattern.

References

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

Bresnock, A. E., Graves, P. E., & White, N. (1989). Multiple-choice testing: Question and response position. *Journal of Economic Education, summer*, 239-245.

Brown, A. S., Schilling, H. E. H., & Hockensmith, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge. *Journal of Educational Psychology, 91*, 756-764.

Delgado, A. R. & Prieto, G. (1998). Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment, 14*(3), 197-201.

Gaskins, S., Forte, L., Wood, F., & Riley, P. (1996). Student perceptions of changing answers on multiple choice examinations. *Journal of Nursing Education, 35*, 88-90.

Geiger, M. A. (1996). On the benefit of changing multiple-choice answers: Student

perception and performance. *Education, 117*, 108-116.

Haase, P. A., Riley, C. P., Dunn, L., & Gaskins, S. (1992). A review of literature on changing answers on multiple choice examinations. *Review of Research in Nursing Education, 5*, 107-121.

Higham, P. A. & Gerrard, C. (2005). Not all errors created equal: Metacognition and changing answers on multiple-choice tests. *Canadian Journal of Experimental Psychology, 59*(1), 28-34.

McClain, L. (1983). Behavior during examinations: A comparison of "A," "C," and "F" students. *Teaching of Psychology, 10*(2), 69-71.

Remmers, H. H. & Remmers, E. H. (1926). The negative suggestion effect of true-false examination questions. *Journal of Educational Psychology, 17*, 52-56.

Roediger, H. L., III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1155-1159.

Rosenthal, R. & Rubin, D. B. (1991). Further issues in effect size estimation for one-sample multiple-choice-type data. *Psychological Bulletin, 109*(2), 351-352.

Shatz, M. A. & Best, J. B. (1987). Students' reasons for changing answers on objective tests. *Teaching of Psychology, 14*(4), 241-242.

Skinner, N. F. (1983). Switching answers on multiple-choice questions: Shrewdness or shibboleth? *Teaching of Psychology, 10*(4), 220-222.

Wasserman, E., Young, M., & Cook, R. (2004). Variability discrimination in humans and animals: Implications for adaptive action. *American Psychologist, 59*(9), 879-890.

Wood, S. E., Wood, E. G., & Boyd, D. (2008). *Mastering the World of Psychology*, 3rd ed. Boston, MA: Allyn and Bacon.