Effective testing techniques I

Running Head: SHORT ANSWER COMPARED WITH MULTIPLE CHOICE

Effective testing techniques: Short Answer compared with Multiple Choice

Names here

Northern Illinois University

Comment [B1]: Not a perfect paper but an actual grop paper.
Effective testing techniques: Short Answer compared with Multiple Choice

At universities across the country, multiple choice testing is the popular method of choice in measuring student’s knowledge. Multiple choice testing has become the norm due to convenience, efficiency, and reliability between graders. Much research has been done to find the effectiveness of this type of testing. Alternatively, the use of short answer testing technique has diminished. Research has also been done to compare short answer and multiple-choice exams. For all types of testing, when feedback is given, positive affects are enhanced and negative affects are reduced. There has been research done studying the results of retrieval time spacing on retention, which also plays a large role on test performance.

Roediger and Marsh (2005) have studied the positive and negative effects of multiple choice testing. They use several other studies to show the positive consequence of multiple choice testing, the main theory proposed being the testing effect. The testing effect states that previous testing will increase the individual performance on later exams. One way to describe it is as a positive transference of information during learning. However, as with positive effects on multiple-choice tests, there are negative effects too.

Another theory proposed is the negative suggestion effect, associated with the number of lures given in each multiple-choice question in hindering later test performance. The negative suggestion effect is the tendency of an individual to believe in false statements, it occurs when a test taker believes a wrong choice on a multiple choice exam to be the correct choice to the propose question. If one chooses the wrong answer on the initial test, without feedback, they will most likely chose the same wrong answer on later tests.

Roediger and Marsh (2005) hypothesize that the testing effect and the negative suggestion will occur within their study. They predicted that participants would do better on a
cued recall test later if they had taken a preliminary multiple choice test than if they had not, exemplifying the testing effect. There was also the idea that a negative effect would arise from the greater the number of decoys.

Their study consisted of twenty-four undergraduate students who were instructed to read passages that were used to construct a one hundred forty-four question multiple-choice test. The final cued recall test consisted of two hundred and sixteen questions. Two independent variables included passage status (read and unread) and number of possible answers given (0, 2, 4, 6). The dependent variable was the number correct on the later cued recall test. This study was done within subjects. The first phase consisted of ninety seconds to read eighteen passages. In the second phase the participants were given twenty-two minutes to complete the multiple-choice exam. Next they were then given a five-minute filler task. Lastly they were given thirty-five minutes to take the cued recall exam.

Results supported the predictions made. Without previous multiple choice tests, participants responded correctly 28% of the time on the cued recall test, in contrast participants answered 46% of the questions correctly when they were seen before on the multiple choice test. When participants were only given two possible choices on the multiple choice exam they answer 51% of the answers correctly on the cued recall section, when given four possible answers they answered 45% correctly, and 43% correctly with six possible answers. Seventy-five percent of the false responses recorded by participants on the cued recall test were lures from the previously given multiple-choice exam.

Overall previous multiple-choice tests have a positive effect on student’s retention of information. Although, multiple-choice exams tend to also produce false knowledge that incorrect information is actually true. Upon reviewing the research of Roediger and Marsh
Effective testing techniques

(2005) interest has been sparked as to the positive and negative effects of multiple choice testing competing with that of short answer testing.

The study of Butler and Roediger (2008) demonstrates the importance of feedback on student’s correct knowledge of information. Feedback increases the learning of correct information especially if the question from the first exam is present in identical form on the later one. This research emphasizes the importance for educators to provide feedback in order to increase retrieval of information on later testing.

As stated earlier, multiple-choice tests are generally beneficial to student’s learning; they are even more beneficial when feedback is given which is supported by the research by Hanna (1973). However, it is much more likely that professors will not give feedback on multiple-choice exams because they are computer graded. On the other hand, short answer exams are highly likely to include feedback because they are instructor graded rather than computer graded. Participants taking short answer tests with feedback perform significantly better when retested at a later date.

Experiments done by Karpicke and Roediger (2007) have looked into how timing of retrieval affects short and long term retention. They propose that the more closely the first test is given after learning, the more likely it will be part of short term retention and the more postponed the first test, the more likely the long term retention will be improved. In our study we want to study how timing effects retrieval based on different testing methods. We would like to find the effectiveness of short answer exams in comparison to multiple-choice exams with feedback on the retention of information over an extended period of time.

Proposed experiment
In this experiment we are extending the research to look at the testing effect for short answer questions. All participants will read a series of passages prior to being randomly selected to take a multiple choice or a short answer exam. Half of the participants from each exam type group will be randomly selected to either take the exam immediately following the reading or one week later. Feedback for these exams will be given directly after completion of the test for all conditions. Four months from the date the passages were originally read, all participants will retake the original exam they were given, whether it is short answer or multiple-choice.

It is hypothesized that short answer exams with feedback will be more beneficial than multiple-choice exams with feedback because when information is presented in a way that is more difficult to recall, people tend to retain the information longer. We also hypothesize that the later a test is taken after initial learning; the more likely it will be retrieved in the long term. We believe this to be true because when a test is taken later rather than immediately, it is more likely to be stored in one’s long-term memory opposed to one’s short-term memory. We predict that there will be greater retention of learned information when taking the short answer test with feedback a week after reading the passages on the final test as opposed to the three other conditions.

**Methods**

**Participants**

Forty NIU psychology 102 students will be participating for course credit.

**Materials**

Twenty-four passages taken from a GRE practice test will be used. A fifty question multiple-choice test and a twenty-question short answer exam will be created; all tests will be in pencil and paper form. For each multiple-choice question, four possible answers will be
constructed. Twenty questions from the multiple-choice exam will be randomly selected and used for the short answer test. Feedback will consist of a written handout with the exact questions followed by the corresponding answers. Based on conditions participants will be previously placed into, they will be put in four separate identical rooms. In order to reduce the influence of confounds, it is planned to run each of the conditions multiple times and to rotate the conditions between the rooms for all of the different trials. Four months later, all participants will receive the same test they originally took. See appendix for example of materials.

**Procedure**

Before the study begins, participants will be given an informed consent sheet to sign. Phase one of the experiment will consist of reading twenty-four passages from the GRE practice test. They will be given forty-five minutes to complete this task. In phase two, red and yellow conditions will be dismissed and asked to return in a week. Blue and green conditions will be administered their multiple choice or short answer exam to take in one hour. Immediately following test completion, test participants will be given the feedback review sheet for twenty minutes, then asked to return it. Phase three will occur a week later, when red and yellow conditions return to the original classrooms they had read the passages in to take either a multiple choice or short answer test they will be randomly assigned to. They too will be given one hour. Again, these exams will be followed by immediate feedback in the form of a handout, which will be administered for strictly twenty minutes and then returned to the administrator. They will then be dismissed. Phase four involves a return of all four conditions four months after initial passage reading. They will then take the exact same test they received in their prior testing condition and again will be given one hour. In phase five the participants will be debriefed and given course credit.
Effective testing techniques 7

Design

This experiment has a 2 Type of exam (short answer or multiple choice) x 2 Retrieval time (immediate or one week) x 2 Test time (initial or final) between subjects design. The dependent variable measured will be the number correct on the test given four months after initial passage reading.

Predicted Results

Responses will be scored for accuracy. These data will then be submitted to a 2 Type of exam (short answer or multiple choice) x 2 Retrieval time (immediate or one week) x 2 Test time (initial or final) between subjects ANOVA. As shown in Figure 1, it is predicted that the results will replicate Roediger and Marsh’s (2005) findings. In the immediate multiple-choice condition the final test scores are expected to be lower than the initial test scores. With a week delay, the multiple-choice scores are expected to have higher final scores than initial scores. As shown in Figure 2, it is expected that the immediate short answer will show similar results to the immediate multiple-choice condition but increased overall score. The short answer after one week is predicted to have higher final scores than initial scores. Both multiple choice and short answer exam scores are expected to increase in the one-week later conditions. The short answer is estimated to have higher scores than multiple-choice in the one-week later condition.
Effective testing techniques

References


Example questions:

Multiple-choice test

1. It can be inferred from the passage that the pterosaur
   (A) resembled the bat
   (B) was a Mesozoic mammal
   (C) was a flying reptile
   (D) lived in the sea

2. The greatest number of forms of mammalian life is found in the
   (A) Triassic period
   (B) Eocene period
   (C) Oligocene period
   (D) Pliocene period

Short-answer test

1. What can be inferred from the passage about the pterosaur?

2. The greatest number of forms of mammalian life were found in what period?
Table 1. *Average percent correct for all test conditions.*

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Multiple Choice</th>
<th>Short Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>32%</td>
<td>40%</td>
</tr>
<tr>
<td>Final</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>One Week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>52%</td>
<td>60%</td>
</tr>
<tr>
<td>Final</td>
<td>66%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Figure 1. Average Correct on initial and final multiple choice exams for feedback conditions

Figure 2. Average Correct on initial and final short answer exams for feedback conditions