A Note on Multiple Choice Exams, with Respect to Students’ Risk Preference and Confidence

DOUGLAS M. WALKER, Department of Economics, Georgia College & State University, Milledgeville, Georgia, USA
JOHN S. THOMPSON, Department of Economics, Louisiana State University, Baton Rouge, Louisiana, USA

ABSTRACT A standard multiple-choice exam format is compared with two modified regimes which provide instructors with information on students’ risk preferences (with respect to score and uncertainty) and students’ confidence in their answers. In one format, students answer each question twice. This format provides information on students’ risk preference. The other exam format allows students to assign a point value to the questions. The point values they choose are clear signals of their relative confidence in their answers. Both alternative exam formats have advantages to students (increased choice) and faculty (low-cost information), while they do not have a significant effect on exam scores.

Introduction

At many universities, introductory courses are taught in large classes, sometimes with hundreds (or even thousands) of students. Although the relative benefits of multiple choice (MC) and essay exams are debatable (see Walstad, 1998; Welsh & Saunders, 1998 for discussions), many instructors use the MC exam format simply because of its ease in grading. One drawback of the traditional MC exam is that it provides the instructor with little information about how well the students perceive that they understand the material tested. In fact, the traditional MC exam provides little (if any) information other than which questions were answered correctly. If the students could easily signal their confidence in understanding various concepts, the instructor could improve the course instruction in the areas of lower confidence.

Given that many introductory courses are examined using the MC format, it is interesting to consider alternative grading schemes for MC exams. Do students exhibit
a preference for or aversion to risk when they are allowed to demonstrate their preferences on exams? Is there a MC exam format that allows the instructor to determine students’ understanding and confidence better than the traditional format? In addition to suggesting some alternative exam formats, these questions are addressed using data from several different MC exam formats. The paper begins by first explaining the various testing regimes that have been employed, along with the risk preference theory that the different regimes address. In addition, it is explained how the various testing methods can be useful to instructors in assessing students’ understanding and confidence. The third section explains the empirical model and results, and the final section concludes the paper.

**Testing Regimes**

The typical scoring process on MC exams involves calculating a student’s score as the percentage of correct responses to the total number of questions. (We call this a ‘standard exam’.) Each question on the standard exam receives equal weight, and answers are either correct or incorrect; there is no partial credit. For simplicity, it is also assumed that only one of the possible answers is indeed correct. (Alternatively, instructors could write MC tests that have two or more correct answers.) Some innovative testing and scoring mechanisms were introduced on our MC exams. Each of the alternative testing and grading procedures is described below. In each testing scheme, students are concerned with maximising their percentage score.

**Hedging**

Under the ‘hedging’ format, students may try for partial credit on MC questions by choosing two answers. Each question is asked twice, as indicated in the example below.

25, 26. A price ceiling set below the equilibrium price causes:

a. a surplus;
b. a shortage;
c. quantity demanded to fall;
d. quantity supplied to rise.

Thus, if a student is sure the correct answer is ‘D’, he would answer ‘D’ twice. If he is correct, each answer earns him a point and he receives full credit. Alternatively, the student can spread out the risk associated with answering ‘D’ if he thinks another answer may possibly be correct. In this case, he could answer the question with two different answers. Suppose the student has narrowed down the correct answer to either ‘B’ or ‘D’. If he chooses both and the correct answer is ‘D’, he only receives one of the two points. If he is certain that one of the two answers is correct, he may elect to forgo the risk associated with trying for full credit in favour of the certainty of getting half credit. (We call this ‘hedging’.) His choice, of course, would depend on his preference for risk, among other things.

The less certain a student is about a particular question, the more attractive hedging becomes. It should be made clear, however, that hedging limits a student’s maximum score on a particular question to half credit. (He would receive half credit if one of the two answers was correct and no credit if neither was correct.) As an extreme example, a student who hedges on every question could receive no higher than a 50% exam score.
Several points about the hedging exam are worth noting. First, an exam formatted in this way requires little effort from the instructor beyond that in preparing and grading a standard MC exam, assuming a computer-graded (or automated) answer sheet is used [1]. Second, the student still has the option of taking a standard exam, simply by not hedging [2]. Thus, having the option to hedge cannot make students worse off.

**Risk Preference**

Frank (1997, chapter 6) offers a standard discussion of choice and risk preference. One application of the theory was performed by Miller and Westmoreland (1998), who tested students’ behaviour regarding selective grading of homework assignments in college economics courses. ‘Selective grading’ means that the professors would assign frequent homework problem sets, but would only grade one or two of the problems. The students would not know beforehand which problems would be graded. Miller and Westmoreland (1998, p. 200) found no evidence that selective grading has an impact on the number or quality of assignments submitted by students. Adapting risk-preference theory to our application, risk averse students are more likely to hedge. Interestingly, some students appear to be better off with a lower score expectation. That is, some students may prefer to hedge because they want to increase the probability of getting at least partial credit, and are willing to sacrifice the possibility of receiving full credit to do so.

Suppose a student is taking a hedging exam where each question has only two answer choices (say, true/false). Alternatively, suppose the student has narrowed down four MC answers to two. In this simplest case, the student can hedge to get a certain 50% score on a particular question. From a pure grade-maximising standpoint (ignoring risk preferences) one wonders whether it ever pays to hedge. For example, suppose the student believes there is a 70% chance of the correct answer being ‘true’, and a 30% chance of the answer being ‘false’. The expected value of the grade from not hedging (picking ‘true’ only) is 0.7 [3]. Since the expected value of the grade from hedging is 0.5, the student who only cares about maximising his grade should not hedge. In fact, if he always picks the answer that he thinks has the higher probability of being correct, his expected grade can never be lower than hedging between the two answers [4].

Thus, on a true/false question, students can always guarantee themselves an expected grade of 50% by hedging. If a student chooses not to hedge, his expected grade will never be lower than 50% (the expected score from guessing), and it will generally be higher than 50% if the goal is a high score, and if he has some knowledge of the material being tested. Despite the fact that a student cannot increase his expected grade by hedging, it still may be desirable to hedge. A risk-averse student may hedge simply because the certainty of the score (lower risk) is worth the trade-off of accepting a lower expected grade on that particular question. That is, the risk-averse student will only be willing to choose a lower expected grade if there is greater certainty of that grade.

An exam with four or five answer choices is slightly more complicated. In this case, a student presumably has greater uncertainty, and we are therefore more likely to see hedging on these exams. If a student is randomly guessing answers, then the expected score from hedging will equal that of not hedging.

**Students’ Confidence**

Instructors could improve the quality of their teaching if they have some information about how well students understand the lecture material. Generally, though, an instructor
has little information other than each student’s score and which questions were most commonly missed. On a standard exam, the instructor has no way of determining the cause of frequently missed questions. It could be that the questions are very difficult given the material, or that the material itself was difficult to understand. Nor does the hedging exam help the instructor in this regard. If many students hedge on a particular question, again it may be the difficulty of the question or of the material, or it may be a demonstration of risk-aversion [5].

One way to assess whether students understand a particular topic is to ask both a ‘simple’ and ‘difficult’ question on the topic. However, for an exam with a limited number of questions, using this scheme would reduce the number of topics that could be tested.

A better way to gain evidence on students’ confidence is described below.

Value Choice

The ‘value exam’ contains MC questions in standard format, i.e. each question is asked only once. The grading, however, is unique. Students have the option of attaching different weights to each of the questions by choosing the point value of each question, from 1 to 3 points. (Each question is worth at least one point, and students must answer all questions.) They do this by either writing the point value in a blank provided next to the question number, or by circling the value (see the example below). If the student circles neither 1 nor 3, the question takes a default value of 2 points.

(1) (3) 25. A price ceiling set below the equilibrium price causes:

a. a surplus;
b. a shortage;
c. quantity demanded to fall;
d. quantity supplied to rise.

Each student’s score is the percentage of correct points (i.e. the number of points earned divided by points attempted) [6]. Like the hedging exam, the student is still free to take a standard exam, by simply choosing the same point value for all questions (be it 1, 2, or 3); in the example question above, this is done most easily by not circling a point value, so that each question carries a 2-point weight.

We should expect the class average score on an exam with this format to be higher than either a standard or hedging exam [7]. This type of exam is relatively time consuming to grade, since computerised grading and scoring like this is not possible (at least at our schools). Because these exams must be graded and scored by hand, simple mathematical errors are possible.

This exam does not yield any information about a student’s attitude toward risk. This may be to its advantage, however. Since this exam does not address risk preference, the information on student confidence is much clearer. When a student chooses a point value, he is directly signalling his confidence about the answer he has chosen. Overall class confidence can be found by simply calculating the average score ‘wagered’ for a particular question, and comparing that to the average for other questions. As with previous exams, more information on this can be obtained by testing a particular concept using two questions of varied difficulty.
The Model and Results

Data were collected using the standard, hedging, and value exam formats at both Louisiana State University (LSU) and Georgia College & State University (GCSU) [8]. At LSU, the second author tested his macro-economics principles, micro-economics principles, and a survey of economics courses during the fall 1997 and spring 1999 semesters. The first author tested his survey courses during the spring 1999 semester at GCSU. Each term there were at least two sections (i.e., classes) of students per course, and there was at least one control group that was given the standard exam. Altogether, there are about 1400 exam scores included in the model. Four different regressions were run, one for each instructor’s course, according to the following format:

\[
SCORE = \beta_1 + \beta_2 HEDGE + \beta_3 VALUE + \beta_4 SECTION1 + \beta_5 SECTION2 \\
+ \beta_6 TEST1 + \beta_7 TEST2 + \mu
\]

Score is the percentage score earned on the exam. All of the explanatory variables are dummy variables, with value of 1 if, respectively, it is a hedging exam, a value choice exam, section 1 of the course, section 2 of the course, exam 1 of the term, and exam 2 of the term. In all courses, there were either 2 or 3 sections, and 3 exams per term. The results showed that none of the special exam formats had a significant effect on exam scores. The insignificance of the hedge variable indicates that, analogous to the findings of Miller and Westmoreland (1998), students on average are risk-neutral when it comes to answering exam questions (in terms of the risk-preference theory discussed in this paper). The value variable was also insignificant. Neither exam format appears to significantly affect students’ grades.

Conclusion

Often students are shy about providing feedback on how well they are learning material. Although their exam scores give general information, specific information on what is understood and what is confusing is not easy to obtain. One way of eliciting this information is to use written exams. In today’s universities, however, where large classes are common, grading written exams may be too time-consuming. There are alternatives to the standard MC exam that can provide useful information regarding students’ confidence and the question difficulty. In addition, students’ risk preference or aversion can be demonstrated on some exam types.

One would expect student reaction to the hedging exam format to be entirely positive, since their options are expanded, and they can still choose a standard exam format. However, we did experience occasional negative feedback, primarily when students did not understand the instructions. This appears to occur, in large part, because the standard format is so common to students. The reaction to the value choice exam was always positive. Students even suggested that this makes exams ‘fun’, because there is some strategy involved. It is important in either case that the exam format does not distract students from answering the questions.

None of the various exam regimes discussed here significantly affects average scores. Despite this, students clearly get a higher utility (or perhaps, lower disutility) from taking these exams, compared to the standard MC exam. This is true because they are given more control over their grade determination than under the standard exam format. Aside from this advantage, the instructor is able to elicit information on students’ confidence that is not generally available from using the standard MC exam format.
### Table 1. Empirical results [1]

<table>
<thead>
<tr>
<th>Variable</th>
<th>LSU Macro</th>
<th>LSU Micro</th>
<th>LSU Survey</th>
<th>GC&amp;SU Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>69.96***</td>
<td>54.97***</td>
<td>70.61***</td>
<td>60.70***</td>
</tr>
<tr>
<td></td>
<td>(36.65)</td>
<td>(35.09)</td>
<td>(33.97)</td>
<td>(36.15)</td>
</tr>
<tr>
<td>Hedge</td>
<td>-2.53</td>
<td>-1.46</td>
<td>-2.18</td>
<td>-4.92</td>
</tr>
<tr>
<td></td>
<td>(-1.36)</td>
<td>(-0.84)</td>
<td>(-1.23)</td>
<td>(-1.39)</td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.31)</td>
<td></td>
</tr>
<tr>
<td>Section 1</td>
<td>-0.49</td>
<td>2.69*</td>
<td>-1.29</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>(-0.03)</td>
<td>(1.85)</td>
<td>(-0.73)</td>
<td>(1.47)</td>
</tr>
<tr>
<td>Section 2</td>
<td></td>
<td></td>
<td>-1.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-0.40)</td>
<td></td>
</tr>
<tr>
<td>Test 1</td>
<td>-9.03***</td>
<td>6.33***</td>
<td>-1.13</td>
<td>10.92***</td>
</tr>
<tr>
<td></td>
<td>(-4.18)</td>
<td>(3.07)</td>
<td>(-0.56)</td>
<td>(3.77)</td>
</tr>
<tr>
<td>Test 2</td>
<td>12.06***</td>
<td>10.14***</td>
<td>-0.19</td>
<td>-0.87</td>
</tr>
<tr>
<td></td>
<td>(5.49)</td>
<td>(5.38)</td>
<td>(-0.09)</td>
<td>(-0.29)</td>
</tr>
<tr>
<td>N</td>
<td>322</td>
<td>462</td>
<td>323</td>
<td>293</td>
</tr>
<tr>
<td>R²</td>
<td>0.24</td>
<td>0.07</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>F-stat.</td>
<td>25.48</td>
<td>8.46</td>
<td>0.50</td>
<td>8.73</td>
</tr>
<tr>
<td>(prob.)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.73)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

*Notes: t*-statistics are indicated in parentheses. *** indicates significance at the 0.01 level; ** at the 0.05 level; * at the 0.10 level.

[1] There are other things that affect scores, including student intelligence, amount of time spent studying, etc. It is assumed that the distribution of students with various qualities is randomly distributed throughout the course sections.

### Notes on Contributors

DOUGLAS M. WALKER is an Assistant Professor of Economics at Georgia College & State University. **Correspondence:** Department of Economics, CBX 14, Georgia College & State University, Milledgeville, GA 31061, USA. Tel: (478) 445-2596. Fax: (478) 445-1535. Email: walker@mail.gcsu.edu

JOHN S. THOMPSON is a visiting Assistant Professor of Economics at Louisiana State University.

**Correspondence:** Economics Department, 2107 CEBA, Louisiana State University, Baton Rouge, LA 70803. Tel: (225) 578-3780. Fax: (225) 578-3807. E-mail: jthomp6@lsu.edu

### NOTES

[1] If a computer-graded answer sheet (where answers are ‘bubbled in’ with a pencil) is employed, it is important to number each question twice (e.g., ‘25, 26’), rather than simply telling students to answer each question twice. This way, the numbers on the exam questions correspond to the numbers on their answer sheets.

[2] One student provided an interesting comment on teaching evaluations: “I don’t like the half credit exams. But I don’t ever go for half credit anyway so I guess it doesn’t matter”.
[3] The student has a 70% chance of receiving full credit, and a 30% chance of receiving no credit. Thus, his expected grade is $0.7(1) + 0.3(0) = 0.7$.

[4] His expected grade from hedging versus not hedging would be identical (0.5) only in the case where he attached equal probabilities (50%) to each answer being correct. For any other case where there is a greater probability of one being correct, his expected grade from not hedging would be higher than from hedging.

[5] Our empirical results suggest that students do not prefer nor avoid risk when given the opportunity.

[6] On a 30-question exam with point values of 1 to 3, a student’s exam may be worth from 30 to 90 points. It is worth stressing to students that the percentage of points correct—not the number of points chosen—is important. For example, a score of 28 of 30 is equal to 84 of 90.

[7] This assumes students’ confidence in knowing whether or not they understand a question is not consistently misplaced.

[8] The first author conducted experiments at Auburn University in 1995, but did not collect data.

REFERENCES


