

A test taker's gamble: The effect of average grade to date on guessing behaviour in a multiple choice test with a negative marking rule



Authors:

Aylit T. Romm¹ 
 Volker Schoer¹ 
 Jesal C. Kika² 

Affiliations:

¹AMERU, School of Economic and Business Sciences, University of the Witwatersrand, Johannesburg, South Africa

²Research, Coordination, Monitoring and Evaluation, Department of Basic Education, Pretoria, South Africa

Corresponding author:

Aylit Romm,
 aylit.romm@wits.ac.za

Dates:

Received: 26 June 2018
 Accepted: 16 Jan. 2019
 Published: 11 Apr. 2019

How to cite this article:

Romm, A.T., Schoer, V. & Kika, J.C., 2019, 'A test taker's gamble: The effect of average grade to date on guessing behaviour in a multiple choice test with a negative marking rule', *South African Journal of Economic and Management Sciences* 22(1), a2542.
<https://doi.org/10.4102/sajems.v22i1.2542>

Copyright:

© 2019. The Authors.
 Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Background: Multiple choice questions (MCQs) are used as a preferred assessment tool, especially when testing large classes like most first-year Economics classes. However, while convenient and reliable, the validity of MCQs with a negative marking rule has been questioned repeatedly, especially with respect to the impact of differential risk preferences of students affecting their probability of taking a guess.

Aim: In this article we conduct an experiment aimed at replicating a situation where a student enters an examination or test once they already have an average from previous assessments, where both this and previous assessments will count towards the final grade. Our aim is to investigate the effect of a student's aggregate score to date on their degree of risk aversion in terms of the degree to which they guess in this particular assessment.

Setting: A total of 102 first-year Economics students at the University of the Witwatersrand volunteered to participate in this study. The test used in this study did not count as part of the students' overall course assessment. However, students were financially compensated based on their performance in the test.

Methods: Following an experimental design, students were allocated randomly into four groups to ensure that these differed only with respect to the starting points but not in any other observed or unobserved characteristic that could affect the guessing behaviour of the students. The first group consisted of students who were told that they were starting the multiple choice test with 53 points, the second group were told they were starting with 47 points, and the third and fourth groups were told that they were starting with 35 and 65 points respectively.

Results: We show that entering an assessment with a very low previous score encourages risk seeking behaviour. Entering with a borderline passing score encourages risk aversion in this assessment. For those who place little value on every marginal point, entering with a very high score encourages risk seeking behaviour, while entering with a very high score when a lot of value is placed on each marginal point encourages risk aversion in this assessment.

Conclusion: The validity of MCQs combined with a negative marking rule as an assessment tool is likely to be reduced and its usage might actually create a systematic bias against risk averse students.

Keywords: Average score; risk aversion; risk seeking; guessing; multiple choice questions.

Introduction

A multiple choice test with a negative marking rule is a typical situation in which an individual who does not have complete knowledge of the correct solution is faced with a gamble of either omitting a question and receiving no reward, or answering a question and being penalised for an incorrect response. Hence, the performance or the associated payoff from such a test is dependent upon whether the student chooses to answer a question when confronted with such a gamble.

The use and grading of multiple choice questions (MCQs) is a well-established and reliable method of assessing knowledge in standardised tests and examinations within the education space. These multiple choice tests are advantageous to both the instructor and the student. From an instructor's perspective, these tests offer increased accuracy and reliability in scoring (Walstad & Becker 1994) as well as objectivity of the grading process (Becker & Johnston 1999). In addition, Buckles and Siegfried (2006) reveal that these tests enable instructors to cover a wide range of

subject material, and facilitate the availability of comparative statistical analysis. From a student's perspective, the objectivity of the grading process is welcomed, since it ensures consistent scoring, and eliminates instructor bias (Kniveton 1996). The consistency and convenience of MCQ testing are the main reasons for conveners of large university classes like first-year Economics to adopt this assessment strategy.

However, multiple choice testing is not without its critiques. Incorrect answer options expose students to misinformation, which can influence subsequent thinking about content (Butler & Roediger III 2008). In addition, students expecting to write a multiple choice test also spend less time preparing for the test (as opposed to essay-based tests) as the answer is selected, and not generated (Roediger III & Marsh 2005). Thus, MCQ testing encourages surface learning rather than deep learning, with students relying on memorising answers to previous MCQ tests instead of working through problems and understanding concepts (Williams & Clark, in: Betts et al. 2009). Furthermore, providing the student with the correct answer option amid a number of incorrect distractors still provides the option to simply guess the answer, which affects the validity of MCQs as an assessment tool (Betts et al. 2009). Especially with partial knowledge, students can increase the probability of guessing the correct answer to a question by eliminating unlikely choices (Bush 2001).

Formula scoring rules, also known as negative marking, are frequently adopted as a means to discourage guessing by subtracting points for incorrect responses; unanswered questions are neither penalised nor rewarded (Holt 2006). The penalty for an incorrect response serves to augment the reliability of tests through a reduction in the measurement errors induced by guessing. A basic property of such formula scoring rules is that the expected value of a pure guess is the same as the expected value from omitting a response (Budescu & Bar-Hillel 1993). However, Davis (1967) asserts that a limitation of the formula scoring rule is the failure to take into account the partial knowledge of examinees, which enables them to eliminate one or more solution options. As students with partial knowledge eliminate one or more solution options, the expected value of guessing exceeds the expected value of omitting a response, which therefore results in guessing being the optimal strategy. However, taking this gamble is a function of the student's risk preference. Bliss (1980) reveals that risk averse students are more likely to omit items for which they have partial knowledge regardless of the positive expected payoff, and these examinees are thus at a disadvantage compared to other students.

More importantly, risk preferences are likely to be non-randomly distributed which could introduce systemic bias. Ben-Shakar and Sinai (1991), Marin and Rosa-Garcia (2011), Burns, Halliday and Keswell (2012) and Hartford and Spearman (2014) show that females are more risk averse than males in the context of economic assessments under a negative marking rule. Hartford and Spearman (2014) show further that it is better performing female students that are most biased against under a negative marking rule, since

better students are more likely to have partial knowledge regarding any particular question.

The purpose of assessments is to measure the students' knowledge through their responses to test questions. If the score obtained through MCQs not only reflects the student's content knowledge but is also a function of other factors that affect the student's probability of taking a guess, then the validity of MCQs as an assessment tool is reduced.

In this article we conduct an experiment to show how performance in previous assessments influences a student's degree of risk aversion in a subsequent assessment. Our experiment aims to replicate a situation where a final grade for a course is calculated by finding the average of more than one assessment. In particular, this experiment looks at how a student might behave in a particular assessment once they already have an average from previous assessments, where both this and previous assessments will count towards the final grade. This is important for the literature on risk taking in general. More often than not, gambles do not occur in isolation but rather in the context of multiple gambles, where a final outcome will be a cumulative result of all such gambles. For example, a financial investment will probably be seen and assessed in the context of any other investments that an agent has undertaken.

The experimental nature of this study offers an advantage over a study based on observational data that was not derived from a controlled experiment. Differences in a student's guessing behaviour as captured by observational data from an actual examination might be reflective of factors other than the student's aggregate mark when entering the examination, while the controlled environment of the experiment aims to isolate the effect of the student's aggregate mark.

Thus, the contribution of this article is twofold: firstly, the experiment has been set up in such a way as to allow us to isolate guessing behaviour by taking partial knowledge out of the equation. This is an advancement in the field of educational science in the sense that it has proven difficult to distinguish guessing behaviour from the effect of partial knowledge. Secondly, guessing behaviour in any particular assessment has, up to now, been looked at in isolation. This study looks at the effect of framing (Kahneman & Tversky 1979), with respect to the effect of the score with which the student enters the assessment. In particular, we investigate the effect of the entry score on the degree of guessing in this particular assessment. This kind of framing effect, that is, aggregate score to date, has not been investigated in the economic or educational literature up to now and adds to the debate on the validity of MCQs as an assessment tool.

Our findings show that entering an assessment with a very low previous score encourages risk seeking behaviour. Entering with a borderline passing score encourages risk aversion in this assessment. For those who place little value on every marginal point, entering with a very high score encourages risk seeking behaviour in the particular assessment,

while entering with a very high score when a lot of value is placed on each marginal point leads to risk averse behaviour.

This article will proceed as follows: the next section provides a description of the experiment conducted in this study. An analysis of the data and a discussion of the corresponding results is presented next and the last section concludes.

Experiment

First-year Economics students from the University of the Witwatersrand were invited to participate voluntarily in the experiment. Participating students were randomly allocated to four different treatment groups as they entered the venue for the experiment. However, an attempt was made to stratify allocation by gender by giving male and female participants different colour tickets with their respective group allocations to ensure that there was a sufficient number of male and female students within each group, thus permitting the analysis of possible treatment heterogeneities along gender lines.

The decision-making experiment was conducted in a classroom setting. A multiple choice test (see Appendix 1) consisting of 10 questions was presented to the students, and each question comprised of 3 possible solution options; the student needed to either choose one correct answer or omit the question. The test consisted of a number of questions that could not be answered, that is, none of the alternative solutions provided was correct, and a response to such a question was reflective of a pure guess (Bereby-Meyer, Meyer & Flascher 2002; Slakter 1969). In order to create the perception of legitimacy of the multiple choice test, the first question was solvable and contained a correct solution; however, this solvable question was not considered in the final analysis of this study. A negative marking rule was adopted where each student received 2 points for a response that was predetermined by the research team as the 'correct' answer, and 1 point was subtracted for each response that was predetermined by the research team as an 'incorrect' response. In addition, no points were gained or lost for each omitted response. Prior to the test, all students were informed about the negative marking rule and the potential payoffs (see Appendix 1).

The expected points from guessing are computed using the equation $EP = \frac{1}{C}G + \frac{(C-1)}{C}L$, in which G represents the number of points gained for a correct response, L represents the number of points lost for an incorrect response and C denotes the number of possible solution options. Since each question comprised of three possible solution options and the student gained 2 points for a correct response, and lost 1 point for an incorrect response, the expected points from a random guess was 0 ($EP = 0$). A standard risk averse expected utility maximiser would not have guessed when the expected number of points from guessing was 0.

Participants were allocated to one of four experimental groups and were treated equally, aside from the points that they received before commencing with a multiple choice test. The first group consisted of students who were told

that they were starting the multiple choice test with 53 points, the second group were told they were starting with 47 points, and the third and fourth groups were told that they were starting with 35 and 65 points respectively. The possible results of the multiple choice assessment ranged from 25 points to 85 points, encompassing the points that could be lost or gained in addition to the points that the students entered the experiment with. Since this was not a real test counting towards the final grade for the course, in order to incentivise students to try and maximise their total scores (so that the situation would replicate a real test situation) a financial payout to each student was provided on a rand (R1) per point basis. In addition, participants were told that a bonus of R50 would be given to each student who reached a score of 50 points or above after completion of the test. The possible results of the multiple choice assessment, ranging from 25 points to 85 points, corresponded to a range with regard to the monetary payoff, from a low of R25 to a high of R135 (taking into account the R50 bonus). The bonus of R50 for students reaching or surpassing a score of 50 points created a reference for each group. Two groups started off being close to the reference point (starting points 47 and 53) while the other two groups started off further away from the reference point (starting points 35 and 65). This allows us to investigate if the framing (relative distance to the reference point) affects the guessing behaviour of the students.

The experimental nature of this study also enables a causal interpretation between the initial number of points received by each student and the tendency to guess. Firstly, this relationship can be attributed to the exclusion of content knowledge as a covariate as participants could not reduce the number of possible solution options through their partial knowledge or obtain the correct answer to the question, since none of the alternative solutions provided was correct. Secondly, the random allocation of participating students into the four groups tried to ensure that no other observable and unobservable characteristics could explain differences in the guessing behaviour of students. Clearly, our interpretation of the results as a causal link depends on the assumption that the randomisation was successful in balancing our four groups in all other characteristics that might affect the guessing behaviour. While we try to test the balance of the four groups on a set of selected observed characteristics that are likely to be correlated with the students' guessing behaviour, we cannot say with absolute certainty that the participants in the four groups only differ with respect to the allocation into the four groups. As such, it is possible that our results could still suffer from omitted variable bias.

We then looked separately at participants willing to accept a wage of R500 a month, and those not willing to accept this wage. Since the payoff in this study was in monetary terms, it was expected that students in greater financial need would attach greater value to every point. We proxied a student's financial need status by the answer they gave to the question regarding whether or not they were willing to work for a wage of R500 a month. Those willing to work for R500 a month

(considered an extremely low wage) were assumed to be in greater financial need than those who were not. As such, the aim was to analyse how guessing behaviour changed, depending on the value students put on each mark. Obviously students in more financial need would value each rand, and therefore each point in the test, more than a student who is financially better off. In a real test situation, students will differ in how they value each marginal point. While some students will be satisfied knowing they have a passing score (and not value each marginal point above that score), others will value each point they get over and above a passing score. Similarly, some students who fail will not care by how many points they fail, whereas others, even though they know they might fail, will still value each point and attempt to maximise their score (perhaps because they would then be eligible for a supplementary – second chance – exam).

In accordance with the university's ethics policy, participation in this study was completely voluntary and participation or non-participation did not affect the students' academic performance in any credit-bearing course. Furthermore, the identity of participants remained completely anonymous as the findings are reported in aggregate format. In addition, participating students were made aware of the full range of possible financial payoffs before they made their decision on participation.

Ethical consideration

This article was written with ethical clearance (protocol number: CECON/1031).

Analysis

Descriptive statistics

The study sample consisted of 102 participating students. Group 1 consisted of 28 students, group 2 of 25 students, group 3 of 24 students and group 4 of 25 participating students.

As part of the experimental design, students were allocated randomly to the four groups. This is to ensure that the four groups differ only with respect to the starting points but not in any other observed or unobserved characteristic that could affect the guessing behaviour of the students. As a test (reported in Table 1), we investigate if the four groups are balanced for a set of observed characteristics that are likely to affect their guessing behaviour; specifically we look at gender, their willingness to work for R500 per month, as well as their matriculation scores in English and Mathematics.

As can be seen in Table 1, despite the small number of participants in each group, the four groups are relatively balanced in their observed characteristics with respect to gender, willingness to work for R500 per month, as well as their academic performances in the final secondary schooling Mathematics exams (matriculation). This was confirmed by the statistically insignificant *F* stats of the analysis of variance testing for these characteristics. However, the variance testing shows that group 3 does differ with respect to the final secondary schooling English exam marks compared to the other groups at a 10% level of significance. We address the potential differences in the groups in the regression analysis by including additional sets of covariates.

Figure 1 shows the outcome variable as the frequency of guesses taken by each group. Students that were allocated to groups 3 (starting point 35) and 4 (starting point 65) were significantly more likely to take a guess compared to students that were allocated to groups 1 (starting point 53) and 2 (starting point 47). Figure 1 also suggests that the difference is across the distribution and not driven by the guessing behaviour of just a few students but due to a significant shift of students that take a higher number of guesses.

As is shown in Table 2, both the mean and median number of guesses show that group 1, that is, those entering with 53 points, guess the least, while group 3, that is, those entering with 35 points, guess the most. This is also supported by the fact that only group 1 has a number of

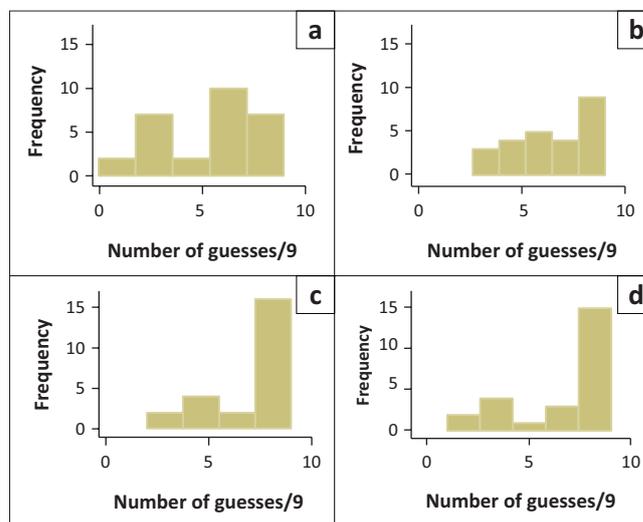


FIGURE 1: Frequency of guesses taken by allocated group. (a) Group 1: Starting point 53; (b) Group 2: Starting point 47; (c) Group 3: Starting point 35; (d) Group 4: Starting point 65.

TABLE 1: Characteristics of composition of each group.

Variable	Group 1		Group 2		Group 3		Group 4		Observations
	<i>N</i>	Standard deviation							
Male	54%	0.51	56%	0.51	54%	0.51	52%	0.51	102
Willing to work for R500 per month	54%	0.51	48%	0.51	54%	0.51	44%	0.51	102
Mathematics score (Matric)	79.19	10.90	75.33	11.06	77.09	12.62	78.10	12.22	91
English score (Matric)	76.00	6.51	74.27	7.57	72.39*	6.43	74.86	8.23	92
Number of guesses (Mean)	5.4	2.8	6.3	2.1	7.4***	2.2	6.8**	2.5	102

*, significance at 10% level; **, significance at 5% level; ***, significance at 1% level.

TABLE 2: Guessing behaviour by group allocation.

Group	Average number of guesses	Median number of guesses	% of students responding to 0 questions	% of students responding to 9 questions
Group 1: 53 points	5.4	6	7.1	17.9
Group 2: 47 points	6.3	7	0	16
Group 3: 35 points	7.4	9	0	54.2
Group 4: 65 points	6.8	8	0	32

TABLE 3: Effects by willingness to accept a monthly wage of R500.

Group	Willing to accept a monthly wage of R500		Not willing to accept a monthly wage of R500	
	Average number of guesses	Median number of guesses	Average number of guesses	Median number of guesses
Group 1: 53 points	5.5	6	5.3	5
Group 2: 47 points	5.5	5	7	7
Group 3: 35 points	7.5	9	7.3	9
Group 4: 65 points	5.5	5	7.9	8

TABLE 4: Gender effects.

Group	Number of guesses: Males		Number of guesses: Females	
	Average	Median	Average	Median
Group 1: 53 points	5.1	6	5.8	6
Group 2: 47 points	6.3	7	6.3	6
Group 3: 35 points	7.9	9	6.8	8
Group 4: 65 points	7.4	8	6.3	7.5

students who decided not to guess at all, whereas group 3 has by far the largest percentage taking a guess at all of the 9 unsolvable questions.

Table 3 shows the frequency of guesses for those willing to work for R500 a month, and those who are not. We see in the first two columns of Table 3 that when students place a higher value on every point, that is, in this case they are willing to work for R500 a month, students in group 2 (47 points) and group 4 (65 points) now guess less, that is, are more risk averse. In fact, their level of risk aversion now seems very similar to the group entering with 53 points. Thus for the group entering with 65 points, even though it is impossible for them to fall below the 50 point level (at which point there would be a huge penalty), they still do not want to lose marginal points. For those entering with 47 points, even though they could jump above the 50 point level (which would result in a huge benefit), they are less willing to take the chance on losing marginal points. Looking at Table 3, when students place less value on each point, the level of risk seeking of group 4 (65 points) now seems to be very similar to that of group 3 (35 points). Those in group 4 seem to care less about each marginal point, knowing that while they may lose marginal points, they can never fall below the 50 point level.

Table 4 shows how these effects differ across gender. It seems that males and females act similarly near the 50 point reference mark where the stakes can be thought to be high, that is, in groups 1 and 2 the guessing behaviour of males and females is very similar. As one moves further away from the 50 point mark, that is, groups 3 and 4, females guess less than males, that is, they are more risk averse.

Regression analysis

With a successful randomisation, the analysis is simply a comparison of the differences in means. The analysis is therefore conducted using an ordinary least squares (OLS) approach, with the average number of guesses for the nine unsolvable items in the multiple choice assessment as the dependent variable (Y) regressed against the allocation into the four different groups (G) which enter the experiment with differential points. We use group 1 which received 53 points as the reference group.

However, should the randomisation not have led to a balancing of other covariates, we need to test if the inclusion of additional covariates (D) mitigates the effect of being in any one of the four assignment groups. We initially test this by adding only gender and the willingness to work for R500 per month. Additionally, we obtained English and Mathematics matric scores but only for a reduced sample (for 90 students in total). As a robustness check, we reduce our sample to students for whom we have a full set of characteristics and test if the inclusion of the full set of covariates affects the group coefficients. Thus, we estimate:

$$Y = \beta_0 + \beta_1 G + \beta_2 D + \varepsilon \quad [\text{Eqn 1}]$$

In Equation 1, ε represents the disturbance term.

We conduct a second set of regressions which allows for the analysis of possible heterogeneities along the lines of gender and income status, in terms of how initial points affects guessing behaviour. This is done by using interaction terms such that:

$$Y = \beta_0 + \beta_1 G + \beta_2 G * X + \varepsilon \quad [\text{Eqn 2}]$$

In Equation 2, $G * X$ is the interactive term between the group and a dummy variable representing either gender or participants' willingness to accept a monthly wage of R500.

Table 5 shows the regression results of the OLS regression represented by Equation 1 for the full sample of participating students. Column 1–3 reports the regression output for the full sample of 102 students. The first column in Table 5 includes only the group allocation and shows that the guessing behaviour of students allocated to group 1 (53 points) and group 2 (47 points) do not statistically differ from one another, while students allocated to group 3 (35 points) and group 4 (65 points) took on average more guesses than students that were allocated to group 1. The point estimates for the groups are robust even when we control for gender and willingness to work for R500 per month. While the point estimate for males is positive, indicating that they are more likely to take a guess compared to females, the difference is not statistically significant. On the other hand, students that indicated that they are willing to work for R500 per month are significantly less likely to take a guess.

The additional robustness test reported in columns 4 and 5 confirms the consistency of these findings. The average number of guesses taken by students in group 3 and group 4 remain above the number of guesses taken by students in group 1.

TABLE 5: Regression results.

Variables	1		2		3		4		5	
	Average number of guesses	Robust standard errors	Average number of guesses	Robust standard errors	Average number of guesses	Robust standard errors	Average number of guesses	Robust standard errors	Average number of guesses	Robust standard errors
Group 2	0.887	0.675	0.879	0.676	0.834	0.670	0.720	0.725	0.558	0.722
Group 3	2.024***	0.682	2.022***	0.684	2.027***	0.677	1.850**	0.715	1.711**	0.719
Group 4	1.447**	0.675	1.453**	0.676	1.371**	0.671	1.386*	0.725	1.333*	0.721
Male			0.350	0.488	0.266	0.486	-	-	0.503	0.575
Accept R500 monthly wage	-	-	-	-	-0.839*	0.486	-	-	-0.852	0.518
English Matric mark (%)	-	-	-	-	-	-	-	-	-0.0276	0.0438
Mathematics Matric mark (%)	-	-	-	-	-	-	-	-	-0.0175	0.0258
Constant	5.393***	0.463	5.206***	0.533	5.700***	0.600	5.423***	0.484	9.079***	3.103
Observations	102	-	102	-	102	-	90	-	90	-
R-squared	0.090	-	0.095	-	0.122	-	0.081	-	0.145	-
F statistic	3.23	-	2.54	-	2.67	-	2.53	-	1.99	-
Prob > F	0.0257	-	0.0447	-	0.0265	-	0.0621	-	0.0879	-

*, significance at 10% level; **, significance at 5% level; ***, significance at 1% level.

TABLE 6: Regression results with interaction effects.

Variables	Average number of guesses (Willing to accept R500)		Average number of guesses (Not willing to accept R500)		Average number of guesses (Males)		Average number of guesses (Females)	
	#	Robust standard errors	#	Robust standard errors	#	Robust standard errors	#	Robust standard errors
Group 2: 47 points	0.750	1.772	0.800	0.980	1.897	1.382	1.733*	0.947
Group 3: 35 points	3.639***	1.503	2.000*	1.104	1.088	1.299	2.933***	1.071
Group 4: 65 points	1.250	1.664	2.000*	1.047	2.031**	0.862	2.933***	1.014
Gender	2.607*	1.405	-1.500	1.551	-	-	-	-
Group 2 * Gender	-1.857	2.046	2.367	1.999	-	-	-	-
Group 3 * Gender	-3.746	1.991	0.357	2.008	-	-	-	-
Group 4 * Gender	-2.964	2.111	1.300	1.719	-	-	-	-
Group 2 * R500	-	-	-	-	-1.917	1.505	-1.800	1.212
Group 3 * R500	-	-	-	-	-0.107	1.642	-0.111	0.931
Group 4 * R500	-	-	-	-	-2.657*	1.316	-2.500**	1.115
Constant	4.250***	1.344	6.000***	0.880	5.769***	0.719	5.067***	0.847
Observation (N)	51	-	51	-	47	-	55	-

*, significance at 10% level; **, significance at 5% level; ***, significance at 1% level.

While the randomisation seems to have been successful with respect to balancing the included observed characteristics, we can see that for the reduced sample the inclusion of the matric marks in English and Mathematics marginally reduces the point estimate for group 3. However, the overall pattern remains consistent.

The *R*-squared is low across the different specifications, which is not surprising given that a number of factors are likely to affect guessing behaviour. However, the aim of this study is not to predict guessing behaviour per se nor do we present a model that explains the variation in guessing behaviour. We are mainly interested in the relationship between the starting point at which the student enters the test (i.e. the allocation to one of the four groups) and the student's guessing behaviour. Similarly, while the *F* statistic is also low, the *p* value confirms that the variation between the groups is statistically significant at the 5% level of significance for the full sample and at the 10% level for the reduced sample. Nevertheless, our results and interpretation do depend on the assumption that the randomisation was successful in balancing observed and unobserved characteristics across the four groups.

Table 6 shows various interaction effects. Columns 1 and 2 represent the gender interaction effects for participants

who are willing and not willing to accept a monthly wage of R500. In both cases the gender interaction with the group is insignificant, that is, the effect of being in a specific group does not differ significantly across gender. However, it does seem that when females are financially constrained (i.e. they are willing to work for R500 a month), they guess more than males (see the gender coefficient in column 1 where this group of females has an average of 2.6 more guesses than males). Columns 3 and 4 show the interaction of financial need with the various groups for males and females. For both males and females, being in the position of being willing to work for R500 a month (i.e. being in financial need and valuing each additional rand) makes those in group 4 (i.e. those entering with 65 points) less likely to guess, or more risk averse. In particular, males in group 4 willing to work for R500 a month have an average of 2.7 guesses fewer than males in group 4 not willing to work for R500 a month, while females in group 4 willing to work for R500 a month have an average of 2.5 guesses fewer than females in group 4 not willing to work for R500 a month.

Discussion

In this article we conduct an experiment aimed at replicating a situation where students enter an assessment with an existing

average mark based on previous assessments. Our aim is to analyse the effect of the average grade to date on students' guessing behaviour in a multiple choice test with a negative marking rule. This is important for the literature on risk taking in general, in that a gamble should not be seen in isolation, but rather in the context of multiple gambles, where a final outcome will be a cumulative result of all such gambles.

The experimental nature of this study offers an advantage over a study based on observational data that was not derived from a controlled experiment. Differences in a student's guessing behaviour as captured by observational data from an actual examination might be reflective of factors other than the student's aggregate mark when entering the examination, while in the experimental setting we can isolate the effect of the aggregate mark.

The incentive in this experiment is provided by way of a financial payout where money is paid for each marginal point, with a lump sum bonus being paid if the student's final score is 50 points or above. This aims to replicate a situation where the vast majority of students would aim to pass a course (where 50 points here replicates a passing mark), with some students placing greater value on each marginal mark above or below that level than others. Since the incentive provided here is monetary, we look at two groups of students, where one group is in greater financial need than the other. We assume students in greater financial need will place a higher value on each point, corresponding to each rand. We show that entering an assessment with a very low previous score (in this experiment 35 points) encourages risk seeking behaviour. Entering with a borderline passing score (53 points) encourages risk aversion in this assessment. Additionally, for those who place little value on every marginal point (here students in less financial need), entering with a very high score (65 points here) also encourages risk seeking behaviour in the particular assessment. In fact, the effect of valuing every marginal point is most prevalent for students entering with the highest amount of points (65), where students in this group who value every extra point are more risk averse, that is, guess less than students in this group who don't place as much value on each extra point.

The results can broadly be interpreted in the context of prospect theory as put forward by Kahneman and Tversky (1979), in which individuals seeing themselves in the loss domain are risk seeking, while those seeing themselves in the gain domain are risk averse. Clearly, in this context this is prevalent at certain points only in these domains, that is, most risk seeking at extreme loss (35 points) (whereas in prospect theory one would expect individuals to be most risk seeking at a small loss), and most risk averse at small gain (53 points). Whether or not students are more risk averse or more risk seeking at high gain (65 points) depends on the extent to which they value each point.

These results are important in light of literature advocating that risk averse students are biased against in multiple choice

settings, since in real test situations most students would have partial knowledge, so that the expected value of guessing would actually be positive. So far the literature has identified females and, in particular, better performing females as being biased against in this regard. Here we identify a student's existing average, apart from whether or not they are a good student (even though these are likely to be related), as a factor affecting risk aversion. Indeed students entering with a borderline pass are most risk averse from this perspective, and thus most biased against. Students with a borderline fail are second most risk averse. Students entering with very low scores seem to be the most advantaged by such a multiple choice test with negative marking (especially if they have some content knowledge) in that they have little to lose from guessing. Thus, previous results showing that worse students guess more could have a lot to do with the fact that they are entering with low scores. Whether or not students entering with very high scores are risk averse depends on the degree to which they value each marginal point. Students in this group who value each point are extremely risk averse, while those not valuing each point as much are extremely risk seeking. Thus, top students might be biased against for two reasons: (1) as identified by previous studies they have better content knowledge, making the expected value of a guess positive; and (2) they are likely to be entering with a very high score and, if they place a high value on each point, as shown in this study, they will be more risk averse for this reason too.

Conclusion

MCQ testing is a popular assessment strategy for large university classes like first-year Economics. However, in the light of our findings, the validity of multiple choice questions combined with a negative marking rule as an assessment tool is likely to be reduced and its usage might actually create a systematic bias against risk averse students. Budescu and Bar-Hillel (1993) suggest that the 'number-of-rights' scoring method, that is, students getting 1 mark for a correct answer and 0 for an incorrect answer (also referred to as 'lenient' marking), may be better to assess candidates. However, the implications of such lenient marking on reliability and validity needs to be further researched as both negative marking and number-of-rights marking introduce validity concerns: negative marking can lead to systematically biasing risk averse students, while number-of-rights marking encourages guessing. Course conveners need to consider this trade-off. Nevertheless, both marking regimes affect the validity of MCQ assessment marks as the final test scores not only reflect knowledge but also guessing behaviour. In response to this trade-off, Lesage, Valcke and Sabbe (2013) propose a number of alternative scoring rules which may also be considered when using multiple choice as an assessment tool.

Acknowledgements

Funding was provided by African Microeconomic Research Unit (AMERU) for this study.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contribution

J.K. was involved in data collection and the preliminary analysis. A.R. and V.S. were involved in the design, analysis and contextualisation.

References

- Becker, W.E. & Johnston, C., 1999, 'The relationship between multiple choice and essay response questions in assessing economic understanding', *Economic Record* 75(4), 348–357.
- Ben Shakar, G. & Sinai, Y., 1991, 'Gender differences in multiple choice tests: The role of differential guessing tendencies', *Journal of Educational Measurement* 28(1), 23–25.
- Bereby-Meyer, Y., Meyer, J. & Flascher, O.M., 2002, 'Prospect theory analysis of guessing in multiple choice tests', *Journal of Behavioral Decision Making* 15(4), 313–327.
- Betts, L.R., Elder, T.J., Hartley, J. & Trueman, M., 2009, 'Does correction for guessing reduce students' performance on multiple-choice examinations? Yes? No? Sometimes?', *Assessment & Evaluation in Higher Education* 34(1), 1–15. <https://doi.org/10.1080/02602930701773091>
- Bliss, L.B., 1980, 'A test of lord's assumption regarding examinee guessing behaviour on multiple-choice tests using elementary school students', *Journal of Educational Measurement* 17(2), 147–153.
- Buckles, S. & Siegfried, J.J., 2006, 'Using multiple-choice questions to evaluate in-depth learning of economics', *The Journal of Economic Education* 37(1), 48–57.
- Budescu, D. & Bar-Hillel, M., 1993, 'To guess or not to guess: A decision-theoretic view of formula scoring', *Journal of Educational Measurement* 30(4), 277–291.
- Burns, J., Halliday, S. & Keswell, M., 2012, *Gender and risk taking in the classroom*, Working Paper no 87, Southern Africa Labour and Development Research Unit, Cape Town.
- Bush, M., 2001, 'A multiple choice test that rewards partial knowledge', *Journal of Further and Higher Education* 25(2), 157–163.
- Davis, F.B., 1967, 'A note on the correction for chance success', *The Journal of Experimental Education* 35(3), 42–47.
- Eckel, C.C. & Grossman, P.J., 2008, 'Forecasting risk attitudes: An experimental study of actual and forecast risk attitudes of women and men', *Journal of Economic Behaviour and Organization* 68, 1–17. <https://doi.org/10.1016/j.jebo.2008.04.006>
- Hartford, J. & Spearman, N., 2014, *Who's afraid of the big bad Wolf? Risk aversion and gender discrimination in assessment*, Working Paper 418, Economic Research Southern Africa, Cape Town.
- Holt, A., 2006, 'An analysis of negative marking in multiple-choice assessment', *Proceedings of 19th Annual Conference of the National Advisory Committee on Computing Qualifications*, pp. 115–118, 7–10 July 2006, Wellington New Zealand.
- Kahneman, D. & Tversky, A., 1979, 'Prospect theory: An analysis of decisions under risk', *Econometrica* 47(2), 263–291.
- Kniveton, B.H., 1996, 'A correlational analysis of multiple-choice and essay assessment measures', *Research in Education - Manchester* 56(1), 73–84.
- Lesage, E., Valcke, M. & Sabbe, E., 2013, 'Scoring methods for multiple choice assessment in higher education – is it still a matter of number right scoring or negative marking?', *Studies in Educational Evaluation* 39(3), 188–193.
- Marin, C. & Rosa-Garcia, A., 2011, *Gender bias in risk aversion: Evidence from multiple choice exams*, Working paper 39987, Munich Personal RePEc Archive, Munich.
- Roediger III, H.L. & 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. <https://doi.org/10.1037/0278-7393.31.5.1155>
- Slakter, M.J., 1969, 'Generality of risk taking on objective examinations', *Educational and Psychological Measurement* 29(1), 115–128.
- Walstad, W. & Becker, W., 1994, 'Achievement differences on multiple-choice and essay tests in economics', *American Economic Review* 84(2), 193–196.

Appendix starts on the next page →

Appendix 1:

Test Questionnaire:

Student number: _____

Gender: _____

As a student, would you be willing to accept a job that pays a monthly wage of R500?

Yes _____ No _____

Multiple Choice Questions:

1. A group of modern economists who believe that markets clear very rapidly and that expanding the money supply will always increase prices rather than employment are the:
A. Monetarists
B. Post-Keynesians
C. Keynesians
2. What is the next term in the following sequence?
1, 1, 2, 3, 19, 34, 83,
A. 162
B. 115
C. 247
3. The Independent Labour and Employment Equity Action Plan was drafted by Mbazima Sithole in which year?
A. 1996
B. 1994
C. 1999
4. Who is the author of the book titled: 'Random Walks and Business Cycles for Dummies'?
A. Norman Gladwell
B. Milton Savage
C. John Friedman
5. The Law of Diminishing Demand states that:
A. As more of a particular good is demanded by the economy, less of that good is demanded by an individual.
B. If good A is preferred to good B, then a higher demand for good B implies a lower demand for good A.
C. As more of a good is supplied in an economy, the less of that good is demanded by the economy.
6. The Depression of 1978 occurred as a result of:
A. Severe drought affecting subsistence agriculture and herding.
B. A banking panic which came about as a result of depositors simultaneously losing confidence in the solvency of the banks and demanding that their deposits be paid to them in cash.
C. A decline in the population growth rate.
7. A Pareto Supremum refers to the allocation of resources in which:
A. All resources are directed to a single individual and no one can be made better off.
B. It is possible to make all individuals better off.
C. A socially desirable distribution is acquired through all individuals having a higher income.
8. The principle of Malthusian Dominance states that:
A. Gains in income per person through technological advances dominates subsequent population growth.
B. An increase in the market price caused by an increase in demand dominates the higher price caused by a deficiency in supply.
C. Increased demand for subsistence consumption eliminates the non-productive elements of the economy.
9. The Population Poverty Index estimates:
A. The percentage of the population living in poor regions.
B. The number of people earning below \$1 a day.
C. The average worldwide population living below the poverty line.

10. A Walrasian Balanced Growth Path refers to:

- A. The act in which excess market supply counteracts excess market demand.
- B. A situation in which output per worker, capital per worker and consumption per worker are growing at a constant rate.
- C. An efficient allocation of goods and services in an economy, driven by seemingly separate decisions of individuals.

Preamble to various groups

Group 1

Good Day

Welcome to this decision-making experiment. My name is (author name). Before proceeding to the test questionnaire, please take note of the experimental instructions below. At the beginning of the test, you already have 53 points to start off with and you are now being placed in a test situation in which you may gain or lose points in addition to the 53 points. These additional points may be gained or lost through a multiple choice test with the following rules:

- You are required to answer a multiple choice test consisting of 10 questions in total.
- You will receive 2 points for each correct response; lose 1 point for each incorrect response; and no points will be gained or lost for each question that you choose to omit.
- Your final amount of points will be calculated as 53 plus the number of points you obtain in the test.
- The payoff you receive will be on a rand (R1) per point basis, i.e. you will receive R1 for each of your final amount of points. In addition, you will receive a bonus of R50 if your final score is above 50 points on completion of the test.
- For example: If you receive 8 points for the test, your final amount of points will be 61 (53+8). In this instance, you will receive R61 + R50 bonus since your final score is above 50 points. Therefore, your final payout will be R111.
- If however you receive -6 points (lose 6 points), for example, your final amount of points will be 47 (53-6). In this instance, you will receive a payout of R47 (you will NOT receive a bonus of R50 because your final score is below 50 points).
- Note that your total payoff can vary between R43 and R123.
- You have 20 min to complete the test.

Please note that your participation in this experiment is completely voluntary, involves no risk and will not affect your academic results in any way. Your answers to these questions are completely confidential and your identity will remain anonymous in the analysis of this study. If you have any questions regarding the instructions above, please feel free to ask. Should you wish to withdraw from this experiment, you may do so at any stage. Thank you for your consideration to participate in this experiment. Should you wish to enquire about my study or access my final results, please feel free to contact me at (author email address). You may now proceed to the test questionnaire.

Kind regards
<author name>

Group 2

Good Day

Welcome to this decision-making experiment. My name is (author name). Before proceeding to the test questionnaire, please take note of the experimental instructions below. At the beginning of the test, you already have 47 points to start off with and you are now being placed in a test situation in which you may gain or lose points in addition to the 47 points. These additional points may be gained or lost through a multiple choice test with the following rules:

- You are required to answer a multiple choice test consisting of 10 questions in total.
- You will receive 2 points for each correct response; lose 1 point for each incorrect response; and no points will be gained or lost for each question that you choose to omit.
- Your final amount of points will be calculated as 47 plus the number of points you obtain in the test.
- The payoff you receive will be on a rand (R1) per point basis, i.e. you will receive R1 for each of your final amount of points. In addition, you will receive a bonus of R50 if your final score is above 50 points on completion of the test.
- For example: If you receive 8 points for the test, your final amount of points will be 55 (47+8). In this instance, you will receive R55 + R50 bonus since your final score is above 50 points. Therefore, your final payout will be R105.
- If however you receive -6 points (lose 6 points), for example, your final amount of points will be 41 (47-6). In this instance, you will receive a payout of R41 (you will NOT receive a bonus of R50 because your final score is below 50 points).
- Note that your total payoff can vary between R37 and R117.
- You have 20 min to complete the test.

Please note that your participation in this experiment is completely voluntary, involves no risk and will not affect your academic results in any way. Your answers to these questions are completely confidential and your identity will remain anonymous in the analysis of this study. If you have any questions regarding the instructions above, please feel free to ask. Should you wish to withdraw from this experiment, you may do so at any stage. Thank you for your consideration to participate in this experiment. Should you wish to enquire about my study or access my final results, please feel free to contact me at (author email address). You may now proceed to the test questionnaire.

Kind regards
<author name>

Group 3 Good Day

Welcome to this decision-making experiment. My name is (author name). Before proceeding to the test questionnaire, please take note of the experimental instructions below. At the beginning of the test, you already have 35 points to start off with and you are now being placed in a test situation in which you may gain or lose points in addition to the 35 points. These additional points may be gained or lost through a multiple choice test with the following rules:

- You are required to answer a multiple choice test consisting of 10 questions in total.
- You will receive 2 points for each correct response; lose 1 point for each incorrect response; and no points will be gained or lost for each question that you choose to omit.
- Your final amount of points will be calculated as 35 plus the number of points you obtain in the test.
- The payoff you receive will be on a rand (R1) per point basis, i.e. you will receive R1 for each of your final amount of points. In addition, you will receive a bonus of R50 if your final score is above 50 points on completion of the test.
- For example: If you receive 18 points for the test, your final amount of points will be 53 (35+18). In this instance, you will receive R53 + R50 bonus since your final score is above 50 points. Therefore, your final payout will be R103.
- If however you receive -5 points (lose 5 points), for example, your final amount of points will be 30 (35-5). In this instance, you will receive a payout of R30 (you will NOT receive a bonus of R50 because your final score is below 50 points).
- Note that your total payoff can vary between R25 and R105.
- You have 20 min to complete the test.

Please note that your participation in this experiment is completely voluntary, involves no risk and will not affect your academic results in any way. Your answers to these questions are completely confidential and your identity will remain anonymous in the analysis of this study. If you have any questions regarding the instructions above, please feel free to ask. Should you wish to withdraw from this experiment, you may do so at any stage. Thank you for your consideration to participate in this experiment. Should you wish to enquire about my study or access my final results, please feel free to contact me at (author email address). You may now proceed to the test questionnaire.

Kind regards
<author name>

Group 4 Good Day

Welcome to this decision-making experiment. My name is (author name). Before proceeding to the test questionnaire, please take note of the experimental instructions below. At the beginning of the test, you already have 65 points to start off with and you are now being placed in a test situation in which you may gain or lose points in addition to the 65 points. These additional points may be gained or lost through a multiple choice test with the following rules:

- You are required to answer a multiple choice test consisting of 10 questions in total.
- You will receive 2 points for each correct response; lose 1 point for each incorrect response; and no points will be gained or lost for each question that you choose to omit.
- Your final amount of points will be calculated as 65 plus the number of points you obtain in the test.
- The payoff you receive will be on a rand (R1) per point basis, i.e. you will receive R1 for each of your final amount of points. In addition, you will receive R50 for participating in this test.
- For example: If you receive 10 points for the test, your final amount of points will be 75 (65+10). Therefore, your final payout will be R125 (R75 + R50).
- If however you receive -10 points (lose 10 points), for example, your final amount of points will be 55 (65-10). In this instance, you will receive a payout of R105 (R55 + R50).
- Note that your total payoff can vary between R105 and R145.
- You have 20 min to complete the test.

Please note that your participation in this experiment is completely voluntary, involves no risk and will not affect your academic results in any way. Your answers to these questions are completely confidential and your identity will remain anonymous in the analysis of this study. If you have any questions regarding the instructions above, please feel free to ask. Should you wish to withdraw from this experiment, you may do so at any stage. Thank you for your consideration to participate in this experiment. Should you wish to enquire about my study or access my final results, please feel free to contact me at (author email address). You may now proceed to the test questionnaire.

Kind regards

<author name>