International Journal of Education, Management, and Technology

e-ISSN : 3026-5444
p-ISSN : 3026-4200

Index: Harvard, Boston, Sydney University, Copernicus, Lens, Scilit, Semantic, Dimensions, etc

https://doi.org/10.58578/IJEMT.v4i2.9365

Test-Taker Behaviour under Uncertainty: Joint Behavioural Modelling of Guessing, Risk Aversion, and Score Validity in Negatively Marked Multiple-Choice Mathematics Tests

Hassan Muhammad[1], Mudassir Umar[2], Muhammad Abubakar Kole[3]
[1,2]Federal University of Education Kontagora, Niger State, Nigeria
[3]Gombe State University, Nigeria
profmaths781@fcekg.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Mar 28, 2026	Jul 26, 2026	Aug 7, 2026	Aug 12, 2026

Abstract

Although negative marking is intended to discourage random guessing in multiple-choice mathematics assessments, it may systematically influence test-takers' decisions under uncertainty and thereby affect score interpretation. This study examines how guessing behaviour and risk aversion jointly relate to item-level decision-making and overall performance in negatively marked multiple-choice mathematics tests. Quantitative observational data were obtained from 885 senior secondary school students. Logistic regression models were used to estimate item-level guessing decisions, while linear regression models incorporating interaction terms examined cumulative performance outcomes. Higher expected score gains were significantly associated with an increased probability of guessing, whereas greater risk aversion was associated with reduced guessing and increased omission behaviour. A statistically significant guessing-by-risk-aversion interaction (reported as $b = 0.29$, $p < .001$) was interpreted as indicating that the performance benefits associated with guessing diminished as risk aversion increased. These findings demonstrate that scores from negatively marked multiple-choice mathematics assessments may reflect behavioural decision strategies in addition to mathematical competence, thereby

IJEMT Journal is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

introducing construct-irrelevant variance into score interpretation. The study contributes to assessment research by clarifying the joint role of guessing and risk aversion in testing behaviour and underscores the need for assessment designers and score users to evaluate cautiously the validity of negatively marked scores, particularly in high-stakes mathematics testing contexts.

Keywords: Decision-Making Under Uncertainty; Guessing Behaviour; Mathematics Assessment; Negative Marking; Risk Aversion

INTRODUCTION

Multiple choice test is a very critical instrument in testing in the learning field and especially in subjects like mathematics where they are used in an attempt to gauge mental power and knowledge. Nevertheless, the character of these tests makes it unpredictable to the individuals engaging in tests that cause certain behaviors such as guessing and riskiness. Two important behaviors that determine the validity of test scores are guessing (choosing an answer without knowing the correct answer) and risk aversion (not providing the answers because of the fear of the punishment in case of wrong answer) (Budescu and Bo, 2014; Akyol et al., 2022). These tendencies are particularly notable in a high-stakes test when the students feel the pressure to perform well, which may cause them to be quite strategic in what they omit or even due to the tendency to guess the answers, making interpretation of the student scores a difficult task. The manner the people who will be administering the tests will deal with the ambiguity in such settings should be comprehended, so that more precise and fairer assessment instruments may be developed.

Although multiple choice tests are common and common, they have a number of biases such as guessing, risk aversion. This kind of undesirable behaviours is more likely to be multiplied because of the nature of negative marking which is common in such tests and the overall effect of which is the formation of misleading tests. This is especially an issue with mathematics where precision and high level of cognitive ability is of great essence. To make the test outcomes correct and unbiased, one ought to know the impact that guessing and risk aversion have on the test taking strategy. This paper tries to explore these behaviors and the effect of behaviors to test design, and test validity.

Despite the previous literature recognizing the existence of guessing behaviour and risk aversion during multiple-choice testing, such behaviour has been studied independently

or taken as a second cause of measurement error. Thus, the current models of assessment do not explain much about systematic differences in scores based on behavioural dispositions but equal mathematical knowledge of test-takers. Such a limitation restricts the correct interpretation of negatively marked multiple-choice mathematics scores and prompts the desire of having empirical models that mutually explain a combination of decision-making under uncertainty.

The research can find application in teaching as well as research design. The study will improve the validity and reliability of the multiple-choice tests by testing the effect of guessing and risk aversion. It is aimed at providing lessons in such a way that would allow creating scoring systems, as well as items formats, which would allow mitigating the effect of such behaviors to achieve a more precise representation of the mathematical abilities of the students. The results will also be relevant to the general argument of how to redesign high stakes assessment in order to enhance more learning and critical thinking and to reduce the number of the strategies, which do not indicate the real academic competence.

The objectives of the research are to:

I. Approximate the predictors of the multiple-choice mathematics test guessing behaviour in the conditions of scoring uncertainty.

II. Examine how the risk aversion has influenced the decisions to take tests, particularly to determine whether to guess or skip questions that are negatively-marked.

III. Compare the relationship between guessing behaviour and risk aversion and establish its impact on the performance in all tests.

In this paper, the following research questions were discussed:

I. What are the variables that predict gambling in multiple choice mathematics tests in case one works with negative marking conditions?

II. How does the risk aversion impact on the decision of taking tests, in particular, the decision to guess and/or are-leave answers when there are doubts?

III. How does guessing behaviour interact with risk aversion in cumulative test performance in multiple choice mathematics tests?

.Literature Review

Conceptual Framework

In the case of negative marking of multiple-choice tests, the conceptualisation of test-taking as a decision-making process in the conditions of uncertainty can be used. According to both the behavioural decision theory and the expected utility theory, the test-takers are supposed to weigh the potential benefits of a correct answer against the potential costs of a wrong answer. Here, subject-matter knowledge is not the only determinant of observed responses since people have their inclination towards risk and loss.

This behavioural framing is consistent with the unified theory of validity in the work of Messick, especially the concept of construct-irrelevant variance, where the test scores are affected by other factors that have nothing to do with the target construct. Guessing and omission are examples of strategic behaviour in multiple-choice tests where the decisions are determined by the scoring rules and not necessarily by mathematical knowledge. As a result, such behaviours should be modeled to help comprehend the outcomes of performance, as well as the validity of the decisions made in the design of assessments in mathematics education.

The two associated behavioural processes through which the test takers respond to uncertainty in multiple choice tests are guessing and risk aversion. Guessing is a willingness to assume the probability risk in the pursuit of possible score advantages particularly when the partial information may be leveraged to have distractors eliminated. Risk aversion on the other hand is an act in which, individuals shun things which they feel will result in loss and that tends to result in omission of items, which are being omitted in a strategic manner despite them being subject to partial information. None of these behaviours is independent and all of them interact to define the means through which the test-takers transform the knowledge and uncertainty into conspicuous forms of responses.

Assessment wise, such strategic reactions add construct irrelevant variance in the test results in that the test performance results will be based on behavioural choice-making in addition to mathematical talent. It can thus be seen that the interaction between guessing and risk aversion can be seen to ascertain the validity of multiple choice mathematics tests especially in high stakes cases where scoring policies increase the risks of decision making and consequently decision-contingent response behaviour is reiterated.

The literature review is the one that studies the precious behaviors of guessing and risk aversion to taking tests especially in multiple choice tests in mathematics. These actions are the major aspects of the attitude of test takers to uncertainty and their decision making and the general performance. The review judges accesses research accessible on these factors, contrasts the many models that interpolate the behavior of persons taking the tests and generates gaps left by the existing literature and gives avenues of further research.

Guessing in Test-taking

Guessing can also be taken to be a compensatory factor that enables the test taker in the multiple choices to obtain the maximum marks with minimum knowledge of what to answer correctly to the test. To illustrate, one can select the correct option that is more likely, when the test-taker has an opportunity to drop one or multiple items without possessing excellent knowledge of the material (Folson and Swanzy, 2024). Nevertheless, it has significant biases in that it can be applied in situations where high stakes are involved in exams, i.e., in assessing higher-order cognitive abilities, i.e. problem-solving skills and reasoning, and not simple recognition or memorization (Rios and Deng, 2021). This will not be indicative of the true ability of the students in the case that they are in a position to guess what is correct without learning it as such. It also can be called construct-irrelevant variance, a situation when students are guided by the approaches that will work in the situation of passing a test to become involved in the actual test measurement of their knowledge and skills (Coniam et al., 2021).

In addition, such construction of the multiple choices questions is also a pertinent factor in stimulating or preventing the effect of guessing. Suppose that poorly-constructed distractors, (i.e. incorrect answers), can also be more predictable to guess, since the participants who take the test can easily detect the incorrect answers to the test, the guessing behavior will be predictable (Stenlund et al., 2017). This does not count to the test, because this promotes a type of test taking methodology that is not genuinely assessing knowledge of a learner but test taking wisdom (Ravand et al., 2025). On the same note, the aspect of designing objects and marking scripts to mitigate the effects of guessing is also a significant area that can be utilized to advance the validity of testing, especially within the tests that are likely to be used to measure higher-order thinking and not low-level recognition (Eduafo, 2014).

Risk Aversion in Test-taking

Risk aversion is a quality of the test takers which implies that such takers tend to escape the uncertainty and this factor in most instances results in failure to answer the question rather than to make guesses. This is largely experienced where they are doing tests which require negative marking system where they are punished on false answers. In fact, the price of making a false guess might be overstated against the returns that might have been attained with an accurate guess and risk-averse individuals will simply forego (Direr, 2020). The main characteristic of high stakes testing is risk aversion whereby people are more likely to prefer security despite the possibility of a score increment.

Test taking risk aversion may have a significant effect on the performance results. It would cause the test-taker to do nothing with the questions that would otherwise increase the probability of obtaining the correct choice over the risk of obtaining a penalty in case he/she makes the erroneous choice (Bhekiswayo & Moleko, 2023). The psychological outlook of risk perception and making up of the test drives this tendency. Risk-aversion can be also increased when stakes are also high especially in the exams when the cost of failure is higher. Indicatively, a research study has established that women respondents in a test are risk aversive than their counterparts in multiple choices tests which take the form of negative marking (Akyol et al., 2016). The difference between the risk aversion of the two genders stands a possibility of creating disparities in the performance in that the female students are likely not to respond to the few questions they have an inkling of, thus scoring less than their male counterparts who are capable of guessing the answers.

Moreover, risking aversion will tend to bring some distortion in knowledge of the actual knowledge of a test-taker. This dynamism involves the methodology of scoring. Systems of scoring that over-penalize errors as wrong answers like weak scoring systems can also produce omission particularly on short-item tests. It may cause the artificial representation of the actual level of knowledge and cognitive level of a student, specifically with those students who are not the riskiest or more conservative in their decision (Direr, 2020). Not only is the presence of risk aversion likely to influence the personal test scores but also the overall validity of the test itself as the element of strategicity that can contribute to falsifying the actual competence of a test-taker is also introduced (Vanderoost et al., 2018). This dynamic implies the methodology of scoring. Weak scoring systems like scoring systems that excessively penalize errors as incorrect answers can create omission particularly on short-

item tests. It may result in the artificial illustration of the actual knowledge level and cognitive level of a student, especially among the students who are not so risky or conservative in their selection (Direr, 2020). Not only the presence of risk aversion influences personal test scores but also the overall validity of the test itself as it is based on the aspect of strategicity that may result in the falsification of the actual ability of the students (Vanderoost et al., 2018).

The intersection in the Guessing and Risk Aversion.

The interplay between the guessing and the risk aversion is very eminent to the modeling of the test-taking behavior in a comprehensive sense. The two are likely to collaborate because they have their influence on the response that the test takers exhibit against uncertainty in multiple choice tests. In spite of the study traditions when it is necessary to consider the issues of guessing and risk-aversion as independent ones, the latest research has started to investigate the context of the two concepts overlap and influence on the decision-making process of the high-stakes tests. To provide an example, students with a risk-averse nature will leave a question without an answer, believing that the punishment in case of giving the wrong answer is too high, yet, despite the high likelihood of giving the right answer (Akyol et al., 2022). On the other hand, less risk-averse individuals are free to indulge in guessing over the strategies, and one is not irritated by the indecision in making the decisions and even enjoy the statistical likelihood of coming up with the right guess. This is especially acute in multiple-choice format whereby the test set up prescribes behaviors of such behaviors to occur.

In this case, men who test have been found to make guesses more than women and this indicates a difference in risk-taking between the two genders (Iriberri and Rey-Biel, 2020). The division may have serious effects in the overall performance in the examinations especially whereby marks are given out. Since females are more predisposed to a greater degree of risk aversion, there exists the risk of females who do not choose to guess and who neglect the question having lower rankings yet they have conditional knowledge (Akyol et al., 2016). In addition to this, such gendered propensities can be further augmented by devising of multiple choice tests. Among them is the fact that the less the distractors in questions or distractor quality, the better the informed guess can be given but the wrongly identified tests may spur random guessing and this will further affect the test outcomes(Stanger-Hall, 2012).

Further interaction between guessing and risk aversion may also be complicated because of other individual differences including cognitive style, anxiety and perceived test difficulty besides gender. The risk-avertive also can evade the guess to save on the potential losses, which cause the omission of strategy compared to those who may exploit the opportunities of success of a guess, particularly in cases where the perceived cost of the error is smaller (Akyol et al., 2016). Therefore, the interaction between the two kind of behavior is complex and subtle and requires additional empirical studies to provide the knowledge of the effects of various test designs and scoring system and individual influences of the test takers on the test taking techniques.

Together, the existing literature body proposes that both guessing and risk aversion are both a part of the test-taking behaviour in the state of uncertainty even though the interplay of the two is not always explicitly modeled in empirical models. Most of the past studies independently separate one of the behaviours and assume the other one as either a background or control variable. A limitation is the fact that it is limiting to explain the reason behind the appearance of systematically different performance results in the case of test-takers that have an equal level of knowledge.

Research Gaps and Future directions.

Although numerous studies have been conducted on guessing and risk aversion, limited empirical research works have been conducted in the joint analysis of these behaviours in the same analysis framework in mathematics assessments with negative marking. In particular, there are few studies that have been conducted on the value of the interaction between guessing and risk aversion to construct-irrelevant variance in measured test scores. The paper aims at addressing this gap by empirically exploring the joint and interactive effects of guessing and risk aversion on performance on multiple choice mathematics tests.

Addressing this gap is essential because failure to model guessing and risk aversion interactively limits the interpretability of multiple-choice mathematics scores and obscures the behavioural mechanisms through which scoring policies distort observed performance.

METHODOLOGY

Research Design

The research design was a quantitative and explanatory one based on the behavioural decision theory and educational measurement. The design investigated their joint influence of guessing behaviour and risk aversion on decision-making and performance in multiple-choice mathematics test with negative marking condition. The conceptualisation of test-taking was that of making decisions in the state of uncertainty and responses were both a measure of mathematical competence and behavioural responses to scoring rules.

Population and Sample

These students were senior secondary students who had previous experience in multiple-choice mathematics tests. Stratified random sampling was used to select 885 students to represent proportions of students both in terms of gender and ability. The process was voluntary, and ethical concerns were taken into consideration.

The participants were familiar with multiple-choice mathematics testing formats, including prior exposure to negatively marked assessments within school-based examinations. Although the test was administered for research purposes, the scoring rules were explicitly communicated and designed to mirror high-stakes testing conditions in which penalties for incorrect responses influence decision-making.

Instrumentation and Measures.

A 40-item multiple-choice mathematics test based on researcher designed question of four options was used to collect data. It featured items of different level of difficulty, conceivable distracters, and negative mark scheme to cause decision risk. The empirical determination of item difficulty was through item p-values.

The patterns of responses, based on the difficulty of items, the correctness and the probability of success were inferred to guess the behaviour. The risk aversion was assessed with the help of a composite index of behavioural levels of omission frequency, sensitivity to magnitude of penalties, and comparison of expected values between the guessing and omission.

The operationalisation of risk aversion involved a revealed-preference methodology based on the behavioural decision theory where omission rates and sensitivity to penalty level is viewed as visible measures of risk aversion in uncertainty. This method is in line with the

previous psychometric and economic studies that have shown that strategic non-response in negatively marked tests is an indication of rational avoidance of anticipated loss but not incompetence. Although self-reported risk measures have not been used, behavioural indicators provide a context-specific validity, as they provide measurement of risk responses in the context of making actual test-taking decisions.

Validity and Reliability

The content validity was determined by review by experts (mathematics educators and assessment specialists). Cronbach alpha (0.70) was used to test internal consistency reliability of the test. The index of behavioural risk aversion also had good internal consistency.

Data Collection

The test was conducted in a single session in standardised conditions and the participants were aware of the rules of scoring. The data were analysed to estimate the influence of item difficulty, anticipated gains, penalties, and risk aversion on the guessing and omission decisions using the analytical procedure based on the descriptive statistics and probabilistic choice models involving regression analysis. Data analysis employed logistic regression models to estimate guessing probabilities and linear regression models with interaction terms to examine the joint effects of guessing behaviour and risk aversion on test performance. Model assumptions were evaluated using standard diagnostic checks, and statistical significance was assessed at the 0.05 level.

Ethical Considerations

Data collection was done after receiving an ethical approval. The anonymity, confidentiality and right to withdraw of the participants was guaranteed and test results did not have any academic implications.

RESULTS

Preliminary Analysis and Descriptive Statistics

A total of 885 respondents of senior secondary schools were screened and their data retained to investigate the study. Initial empirical checks showed that there were no distributional assumptions violations that were important enough to discredit model estimation.

Table 1: Descriptive Statistics of Test Performance and Behavioural Indicators (N = 885)

Variable	Mean	SD	Min	Max
Total Test Score	54.72	12.84	18	86
Proportion of Items Guessed	0.29	0.14	0.05	0.68
Proportion of Items Omitted	0.17	0.11	0.00	0.54
Risk Aversion Index	0.61	0.19	0.12	0.94

The descriptive findings reveal that there is a lot of variation in the guessing behaviour as well as in the risk aversion, which imply that there are different decision-making strategies by the test-takers. It is important to note that rates of omission were also not trivial that assumed negative marking triggered risk-sensitive response behaviour.

RQ1: What are the variables that predict the guessing probability on multiple-choice mathematics tests?

To answer RQ1, a probabilistic choice model was fitted, which forecasted the probability of guessing at a given item difficulty, the expected improvement in score, the penalty severity and personal risk aversion.

Table 2: Determinants of Guessing Probability

Predictor	B	SE	Z	p
Item Difficulty	−0.47	0.06	−7.83	< .001
Expected Gain from Guessing	0.62	0.08	7.75	< .001
Penalty Magnitude	−0.39	0.07	−5.57	< .001
Risk Aversion Index	−0.51	0.05	−10.20	< .001

The **o**utcomes suggest that the probability of guessing is less predicted with a high difficulty of the item and that high expected gains will greatly influence guessing. Notably, risk aversion had a significant negative influence, which confirmed that more risk-averse people did not guess quite often though it was statistically profitable to do so.

RQ2: How does risk aversion influence decision-making concerning the issue of guessing and omission?

In order to address this question, the participants were classified into low, moderate, and high risk-aversion depending on tertiles of the risk aversion index.

Table 3: Response Strategy by Risk Aversion Level

Risk Aversion Group	Guessing (%)	Omission (%)	Attempted Responses (%)
Low Risk Aversion	42.6	8.9	91.1
Moderate Risk Aversion	28.3	16.7	83.3
High Risk Aversion	14.5	31.8	68.2

The findings reveal a very obvious monotonic trend: the greater the risk aversion, the lower the guessing behaviour becomes and the higher the omission rates become. Test-takers who were risk-averse omitted almost one-third of items, which means that they are afraid of possible punishment.

RQ3: How does guessing relate to risk aversion, and what is the overall effect of the two on test performance?

To answer RQ3, an estimated interaction model was developed, and test score was the dependent variable and guessing behaviour as well as risk aversion and the interaction between these two variables were used as predictors.

Table 4; Interaction Effects of Guessing and Risk Aversion on Test Performance

Predictor	β	SE	t	P
Guessing Behaviour	0.34	0.06	5.67	< .001
Risk Aversion	−0.41	0.05	−8.20	< .001
Guessing × Risk Aversion	−0.29	0.07	−4.14	< .001

The interaction term was statistically significant, indicating that the performance benefit of guessing diminishes as risk aversion increases. While moderate guessing improved scores for low-risk individuals, highly risk-averse students failed to capitalise on partial knowledge, leading to systematically lower scores despite comparable ability levels.

DISCUSSION

The paper discussed the joint effects of guessing behaviour and risk aversion on decision-making and performance in negatively marked multiple-choice mathematics tests. The simultaneous modelling of these behaviours, therefore, furnishes empirical evidence that the scores in the test observed are not just being reflections of mathematical competence alone but in a systematic way dependent upon behavioural responses to uncertainty scoring rules.

Answering the first question of the research, the findings indicate that the perceived likelihood of gaining scores, difficulty of the items, the amount of the penalty, and personal risk aversion are highly conditioned to guess the behaviour. The expected gain has a positive influence on estimating the probability as found in decision-theoretic predictions such that test-takers weigh potential rewards against expected losses in case of uncertainty. In contrast, higher item difficulty and penalties decrease the possibility of guessing by the participants,

and this is a sign of increased perceived risk. Notably, risk aversion imposed a significant and independent suppressive influence on the occurrence of guess, confirming that although guessing may be a statistically beneficial behavior, loss-aversive people are less likely to perform it. The results enhance previous studies by showing that only item specifications do not determine strategic responding in mathematics tests but the existence of consistent behavioural risk propensity.

The second question of the research was concerned with the impact of risk aversion on the response strategies, especially guessing and omission. The findings showed that the pattern was plain monotonic by the level of risk-aversion: the guessing rates decreased and omission rates increased significantly. Test-takers who are highly risk-averse elided a significant percentage of the items which shows that they used a systematic avoidance strategy due to fear of a penalty, rather than lack of partial knowledge. This trait of behaviour is congruent with previous results that negative marking does decrease response efforts by loss-sensitive individuals and contributes to the claim that such omission is a rational behaviour in the uncertain context not loss disengagement or effortlessness. In turn, omission behaviour turns out to be an imperative process by which scoring policies influence performance outcomes.

The most consequential finding of the study is the analysis of interactions that is based on the third research question. Guessing behaviour was found to work positively with overall test performance, but this advantage became markedly less with increased risk aversion. This is to say that the performance benefit of the partial knowledge into responses via guessing depended on the level of risk tolerance of an individual. Students that were highly risk-averse did not use partial information profitably, and they scored lower on average even though they were exposed to the test content in a similar manner. This is an interaction effect that points out the fact that; performance disparities between the test takers with similar knowledge level might be a result of behavioural decision strategies and not the actual differences in mathematical ability.

Regarding the aspect of validity, these results are significant. In multiple-choice tests of mathematics with negative marking, scores represent a behavioural aspect which is associated with decision-making in the face of uncertainty, and thus bring construct-irrelevant variance. Students who are risk-prone can score better by scoring by guess, whilst the equally knowledgeable risk-averse students can be punished by omission. Consequently,

test scores can confound mathematical ability and behavioural predispositions to risk which suggests that perceived performance reflects the intended construct. This is consistent with unified validity theory which underscores the importance of trying to explain systematic causes of construct irrelevant variance in explanation of the results of assessment.

In addition to the interpretation of the scores, the findings also create significant assessment design concerns. Although it is claimed that negative marking can be applied to prevent random guessing, the current outcomes suggest that negative marking regimes can place undue weight on risk-averse test-takers and exalt behavioural differences that are independent of mathematical ability. According to this trade-off, the scoring policies which are designed to increase fairness may unwillingly alter the performance distributions by introducing behavioural biases. Assessment designers must then be keen to consider whether the negative of discouraging blind guessing outweighs the negative of strategic omission and the repressed performance of loss-sensitive people.

There are a number of shortcomings that ought to be recognized when explaining these results. The risk aversion was determined behaviourally and not by independent psychological measures, and guessing pseudometrically in terms of response patterns instead of direct indications of knowledge states. Even though the methods are consistent with revealed-preference models and provide a high level of ecological validity, it may be improved in the future by adding complementary measures (self-reported risk preferences, response-time measures, or manipulation of the scoring rule in experiments, etc.). Further extending the analysis to other item-format or adaptive testing conditions would also help shed more light on the interaction between decision-making under uncertainty and assessment design.

On the whole, the given research makes a contribution to the research on mathematics assessment by showing that the two phenomena of guessing and risk aversion are mutually dependent decision-making processes that affect the performance in the context of negative marking systematically. The results by explicitly modelling their interaction help to explain why test scores can differ among equally capable students and the reasons why behavioural considerations must be introduced in both assessment design and validity interpretation.

CONCLUSION

The interaction between decision-making and behavioural guessing in the context of tasks based on under uncertainty multiple choice mathematics tests and risk aversion in performance has been discussed in this paper. Using the empirical evidence of the negatively marked assessment, the findings show that the scores of viewed tests are not only the measures of mathematical competence, but the systematic behavioural measures of scoring rules. In particular, it was noted that the more omission behaviour was associated with risk aversion and the less guessing showed to be associated with risk aversion, as well, performance benefits of guessing were found to decline with extent of risk aversion.

The correlation between the guessing and risk aversion identifies that the performance scores of the multiple choice mathematics tests are determined concerning the conditions of either decision strategy with the help of one or five conditions. The similar test takers with similar knowledge of math can therefore have different scores in the event of uncertainty and potential loss. What this observation points to is the fact that there is construct-irrelevant variance in tests that involves policies of rewarding strategic behaviour that is irrelevant to the construct of measurement.

The study can be useful to the assessment research by showing that the decision mechanisms of behavioural visions can be a significant factor in determining the results of the test through joint modelling of the guessing and risk aversion. The findings indicate that one should be careful when using multiple-choice mathematics scores in high stakes settings, especially when negative marking is used and small deviations in scores have high stakes.

Recommendations

The use of negative marking in multiple-choice mathematics tests should be checked closely by the assessment designers because when scoring on a penalty basis, this method can discourage the performance of risk averse test-takers, but this does not necessarily represent their mathematical ability. The scoring schemes which will minimize the sensitivity to excessive losses, and decrease random guessing are likely to enhance the validity of the score.

Moreover, investigation organisations and teachers are advised to be cautious when testing results on multiple choice tests in high stake situations, as they realise that apparent performance can partially be related to the strategic decision making in uncertainty but not

solely in ability. The use of behaviour-sensitive analytic methods within the assessment evaluation can potentially increase the prejudice and interpretability of test result.

REFERENCES

Akyol, Ş. P., Key, J., & Krishna, K. (2016). *Hit or miss? Test taking behavior in multiple choice exams* (NBER Working Paper No. 22401). National Bureau of Economic Research. https://doi.org/10.3386/w22401

Akyol, P., Key, J., & Krishna, K. (2022). Hit or miss? Test taking behavior in multiple choice exams. *Annals of Economics and Statistics*, (147), 3–50. https://doi.org/10.2307/48684785

Awang Besar, M. N., Abd Aziz, K. H., & Abdul Halim, H. (2025). The validity of multiple-true-false and one-best-answer in the final professional undergraduate medical examination. *Education in Medicine Journal, 17*(2), 5–21. https://doi.org/10.21315/eimj2025.17.2.2

Bhekiswayo, N. M. (2023). *Balancing mathematics assessment tasks to accommodate problem-solving in Grade 7 classes* [Master's dissertation, University of the Free State]. KovsieScholar. https://hdl.handle.net/11660/12635

Budescu, D. V., & Bo, Y. (2015). Analyzing test-taking behavior: Decision theory meets psychometric theory. *Psychometrika, 80*(4), 1105–1122. https://doi.org/10.1007/s11336-014-9425-x

Coniam, D., Lee, T., & Lampropoulou, L. (2021). Identifying guessing in English language tests via Rasch fit statistics: An exploratory study. *English Language Teaching, 14*(5), 23–29. https://doi.org/10.5539/elt.v14n5p23

Direer, A. (2020). *Efficient scoring of multiple-choice tests* [Preprint]. HAL Open Science. https://hal.science/hal-02262181v2

Iriberri, N., & Rey-Biel, P. (2021). Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large-scale natural field experiment. *European Economic Review, 131*, Article 103603. https://doi.org/10.1016/j.euroecorev.2020.103603

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.

Rios, J. A., & Deng, J. (2021). Does the choice of response time threshold procedure substantially affect inferences concerning the identification and exclusion of rapid guessing responses? A meta-analysis. *Large-scale Assessments in Education, 9*, Article 18. https://doi.org/10.1186/s40536-021-00110-8

Romm, A. T., Schöer, 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), Article a2542. https://doi.org/10.4102/sajems.v22i1.2542

Stanger-Hall, K. F. (2012). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE—Life Sciences Education, 11*(3), 294–306. https://doi.org/10.1187/cbe.11-11-0100

Stenlund, T., Lyrén, P.-E., & Eklöf, H. (2018). The successful test taker: Exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education, 33*(2), 403–417. https://doi.org/10.1007/s10212-017-0332-2

Vanderoost, J., Janssen, R., Eggermont, J., Callens, R., & De Laet, T. (2018). Elimination testing with adapted scoring reduces guessing and anxiety in multiple-choice assessments, but does not increase grade average in comparison with negative marking. *PLOS ONE, 13*(10), Article e0203931. https://doi.org/10.1371/journal.pone.0203931

Wei, X., Zhang, S., & Zhang, J. (2024). Identifying student profiles in a digital mental rotation task: Insights from the 2017 NAEP math assessment. *Frontiers in Education, 9*, Article 1423602. https://doi.org/10.3389/feduc.2024.1423602