Tanzania Journal of Science 48(4): 760-772, 2022
ISSN 0856-1761, e-ISSN 2507-7961

Quantifying the Effects of Guessing, Position Bias and Prior Knowledge in Multiple Choice Exams

Almasi S Maguya
Department of Computing Science Studies, Mzumbe University P. O. Box 85, Mzumbe, Tanzania *Email address: asmaguya@mzumbe.ac.tz*
Received 7 Apr 2022, Revised 28 Oct 2022, Accepted 13 Nov 2022, Published Dec 2022
DOI: https://dx.doi.org/10.4314/tjs.v48i4.4

Abstract

"Can an examinee pass a multiple-choice (MC) exam by chance?" Many studies have tried to address this question. However, these studies ignore different types of position bias associated with picking an item from a list. Despite the presence of considerable evidence on the existence of position bias in guessing answers in MC exams, these studies assume an examinee chooses answers with equal probability. This paper seeks to fill this gap by quantifying the chance of success in MC exams due to guessing when different types of position bias and prior knowledge are taken into consideration. The paper proposes a probabilistic model for position bias and uses it to conduct a series of computer simulations for quantifying the chance of passing an MC exam. Results show that the chance of passing an MC exam by pure random guessing is generally negligible even for a poorly set MC exam. Furthermore, results show that chances for an examinee with a fair amount of prior knowledge passing an MC exam of acceptable standard are between very high and perfect. Since a typical examinee is expected to possess some amount of prior knowledge, these results imply that despite their popularity, MC exams should be avoided particularly in high-stake exams where they can lead to false positives.

Keywords: Multiple-choice exams; multiple-choice questions; position bias; computer simulation.

Introduction

Multiple-choice (MC) exams are popular and play a key role as a tool of assessment at virtually all levels of education ranging from primary schools through universities (Moss 2002, Parkes and Zimmaro 2016). However, despite their practical utility as assessment tools, MC exams have one major drawback: the possibility of an examinee to answer a multiple-choice question (MCQ) correctly due to partial prior knowledge or by random guessing (by chance) (Lee 2019). Consequently, it is theoretically possible for an examinee to pass an MC exam by random guessing, leading to what are known as false positives.

This drawback undermines the value of MC exams as assessment tools especially in science, technology, engineering and mathematics (STEM) subjects in which creative thinking, rather than memorisation of facts is emphasized (Stanger-Hall 2012, Keinänen et al. 2018, Kleemola et al. 2022). Despite a great deal of efforts made to ameliorate this drawback (Burton and Miller 1999, Dubins et al. 2016)—such as putting forward different guidelines for setting multiple-choice questions (MCQs)—evidences from the literature show that setting MCQs of acceptable quality is so demanding and time consuming that these guidelines are usually not adhered to by majority of MC exam setters (Dehnad et al.

2014, Kilgour and Tayyaba 2016, Brown and Abdulnabi 2017). Therefore, it is not surprising that researchers have been extensively investigating this MCQ problem for a long time in an effort to quantify the extent to which an examinee can pass an MC exam by random guessing (Burton 2001, Gu and Schwartz 2018, McKenna 2019).

The fact that an increasing number of schools and universities around the world are now adopting e-learning to deliver their courses means that MC exams, due to their objectivity and simplicity of grading, will continue to be the means of choice for assessment in schools and universities for a long time to come. Moreover, the outbreak of COVID-19, for instance, has seen an unprecedented rate of adoption of e-learning by many universities and schools around the globe (Slade et al. (2022). Because assessment in e-learning courses relies heavily on MCQs, this increasing trend of e-learning adoption calls for even more research on the suitability and robustness of MC exams against random guessing and other weaknesses (Kumar et al. (2021).

The guessing problem in MC exams has been studied relatively extensively in the past, with most researchers focusing on the guessing behaviour of examinees when taking MC exams (Attali and Bar-Hillel 2003, Lee 2019, Akyol et al. 2022) as well as quantifying the probability of passing an MC exam by random guessing (Burton 2001, Gu and Schwartz 2018, McKenna 2019). The former aspect tries to model or describe the guessing behaviour of an examinee with respect to attempting an MCQ, while the latter tries to estimate the pass rate attributable to guessing.

The main limitation of previous studies aiming to quantifying the pass rate in MC exams is the fact that they do not take into consideration the examinees' behaviour, particularly position bias. Position bias, which is a tendency of an examinee to prefer choice items at specific locations such as at the top, middle, or bottom of the choice list (Gu and Schwartz (2018) bias is used in modelling the behaviour of examinees who rely solely on random guessing when taking MC exams.

Previous studies assume a naive (uniform) probability for each choice item in an MCQ. This assumption is inconsistent with the findings of many previous studies on examinee behaviour when attempting MCQs which indicate the presence of strong position bias towards certain positions in the choice item list (Clark 1956, Fagley 1987, Attali and Bar-Hillel 2003, Kuhn et al. 2020, Akyol et al. 2022). Another limitation is that the sample sizes used in many of these studies have been relatively small—something that makes generalising the results of these studies challenging.

Ignoring position bias in these studies can be attributed to the fact that only few models of position bias exist (e.g., Blunch 1984), and that these models are both relatively old and, in some cases unwieldy, making it difficult to apply them in practical settings. In this paper, we fill this gap by quantifying the chance of success in MC exams when the effect of position bias is taken into consideration. To achieve this goal, we first develop a simple probabilistic model for position bias and use it to conduct a series of computer simulations for quantifying the chance of success in MC exams due to guessing.

Materials and Methods

In this section, the test data used in this study are described and a probabilistic model for position bias is developed. The developed model is then used in subsequent computer simulations for quantifying the chance of success attributable to guessing in MC exams.

Test data

The test data used in this study consisted of a real exam. The exam is a 2015 national standard VII mathematics exam administered by the National Examination Council of Tanzania (NECTA) annually as part of qualifying examinations for entering secondary education (NECTA 2022a). The exam consisted of 50 multiple-choice questions and each question had 5 choice items (A–E). This exam was used because it was readily available to the author and it served the purpose of the study well.

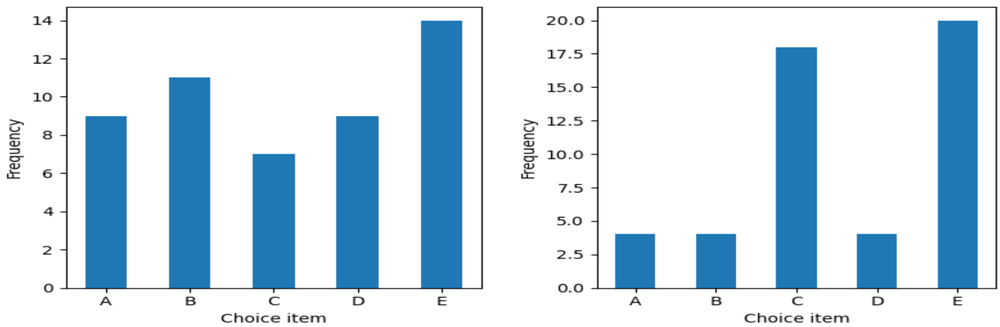
Additionally, so as to study the effect of position bias, a poorly set MC exam was needed. Here, "poorly set" means the distribution of answers in the answer key is not balanced—some answers have a much higher frequency than others. Since this kind of exam was not readily available, we resorted to modifying the original exam's answer key by arbitrarily permuting the answers so that choice items C and E occur with a relatively higher frequency. The answer key for the original exam was

CDDEABCEABEDEABDECEBDAABEAECBBDBEACDAEDDBCEABECEBE.

while the answer key for the modified exam was

ACECECCEECEBEECDECECDEACEAECCBDBEEADCECEBCCECECECE.

Figure 1 shows the distribution of answers in these answer keys. As seen in the figure, modification of the original exam was performed so as to favour two specific types of position bias namely centre bias (choice item C) and lower edge bias (choice item E).



(a) Original exam (b) Modified exam

Figure 1: Distribution of answers in the answer keys of the test exams.

Modelling position bias

The model for position bias proposed is based on the observation that there are generally two main types of position bias reported in the literature, namely edge bias and centre bias (Attali and Bar-Hillel (2003).

An examinee with edge bias exhibits a tendency to prefer choice items located on the edges of the choice item list, while an examinee with centre bias prefers items located in the middle of the choice item list. For example, in a four-item MCQ (choices A, B, C, and D), an examinee exhibiting edge bias would prefer either choice item A or choice item B, while an examinee exhibiting centre bias would prefer either choice item B or choice item C. By contrast, an examinee without position bias would not have any particular preference. Here, prefer means that the probability that the examinee will choose items in those locations is much higher compared to the probability of choosing items in the other locations. Therefore, in light of this fact, we can treat the item chosen by an examinee who tries to answer an MCQ by random guessing as a random variable. In doing this, the problem of modelling position bias reduces to the problem of finding a suitable probability distribution for modelling each of the two types of position bias described above.

For this purpose, we make use of the beta distribution (Forbes et al. 2011) for modelling each type of position bias. The probability density function of this distribution is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (1)$$

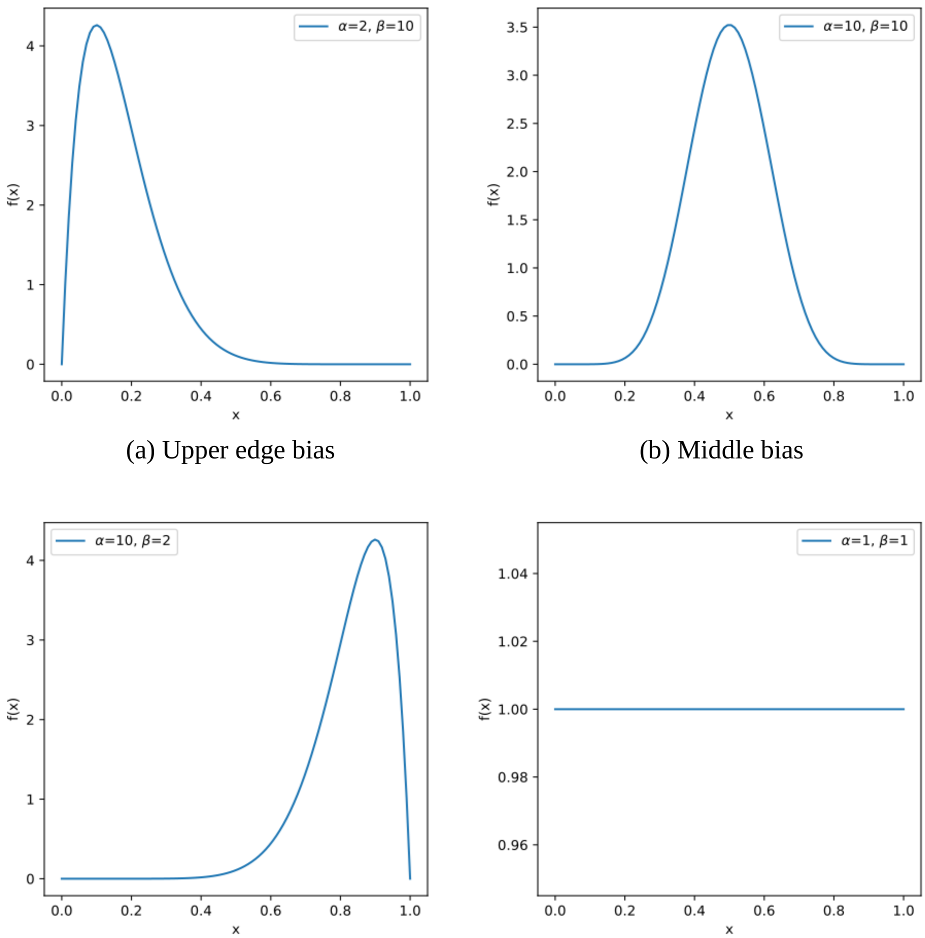
where $B(\alpha, \beta)$ is the beta function with arguments α and β given by

$$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du \quad (2)$$

The free parameters α and β in f(x) are called shape parameters—they control the shape of the beta distribution and they contribute to the versatility of this probability distribution.

It follows that, by choosing suitable values for α and β we can obtain probability distribution functions (PDFs) with desired shapes to model the random variables that represent position bias. Figure 2 shows sample beta distributions tuned for representing the two types of position bias described above. In this paper, we distinguish edge bias into two independent biases that represent the upper and the lower edges of the MCQ choice item list.

Since the PDFs shown in Figure 2 are continuous, whereas the random variables of interest are discrete, we need to derive probability mass functions (PMFs) from these continuous PDFs for modelling these variables. In the next section we show how these PMFs are derived from corresponding beta PDFs. After showing how to derive the PMFs for each type of position bias, we describe how to sample from the resulting PMFs.



(a) Upper edge bias (b) Middle bias

(c) Lower edge bias (d) No bias (naive) case

Figure 2: Beta distributions with parameters tuned to model common types of position bias shown when an examinee tries to answer an MCQ by random guessing (a–c) and a beta distribution tuned to model absence of position bias (d).

Deriving PMFs from beta distribution PDFs

To derive PMFs from the beta distributions, we begin by partitioning the sample space of each beta distribution into N intervals of equal width where N is the number of choice items per MC question. For example, for a five-item MCQ (choices A, B, C, D, and E), each interval is going to be 1/5 in length. These intervals correspond to the choice items, i.e., the first half-open interval [0–1/5) corresponds to the first choice item (item A), the second half-open interval [1/5–2/5) corresponds to the second choice item (item B), and so on. Figure 3 illustrates partitioning of the sample space of a beta distribution tuned for modelling centre-bias into five intervals that correspond to an MCQ with five choice items (A–E).

The next step after partitioning the sample space is to compute the probabilities for guessing each choice item. This step is done for each type of bias shown in Figure 2. For example, the probability for guessing a given choice item k = A, B, ⋯, E in Figure 3, for example is given by

$$p_k(x) = \int_i^{i+1/N} f(x)dx \qquad (3)$$

where i = 0, 1/N, ..., 1-1/N, and i+1/N are the start and end points of the interval for a choice item, respectively, and f(x) is the PDF corresponding to the beta distribution for a particular type of position bias. Table 1 and Figure 4 show the PMFs and probabilities, respectively, obtained after computing individual probabilities corresponding to each choice item for the example beta PDFs shown in Figure 2.

Table 1: Probabilities of choosing a given choice item under different types of position bias

Type of bias	Choice item				
	A	**B**	**C**	**D**	**E**
upper	0.677877	0.291889	0.029499	0.000733	9.216e-07
middle	0.001579	0.184513	0.627816	0.184513	0.001579
lower	9.216e-07	0.000733	0.029499	0.291889	0.677877
naive	0.200000	0.200000	0.200000	0.200000	0.200000

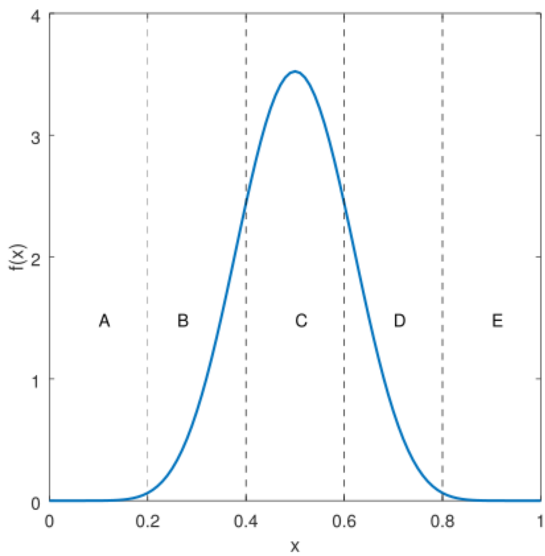
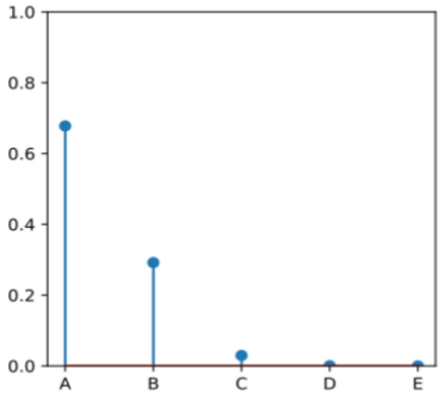
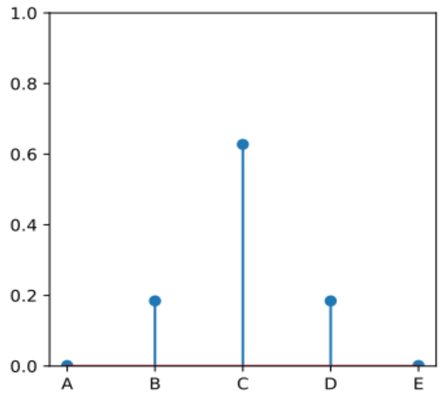


Figure 3: Partitioning the sample space of a beta distribution into intervals of equal length and mapping those intervals to MCQ choice items.

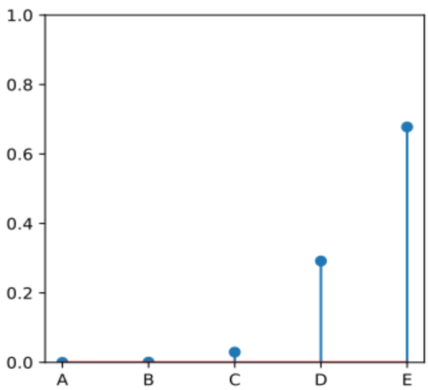
The obtained PMFs are then used in the next step which is simulating the process of randomly guessing the correct answer of an MCQ. It turns out that simulating this process amounts to sampling from the derived PMFs.



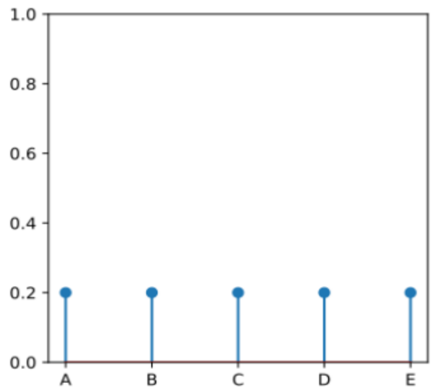
(a) Upper edge bias PMF



(b) Middle bias PMF



(c) Lower edge bias PMF



(d) No bias (naive) case PMF

Figure 4: Final PMFs obtained after computing individual choice item probabilities for each Beta PDF shown in Figure 2.

Sampling from derived PMFs

The process of sampling from the derived PMFs simulates the process of randomly picking a choice item in an MCQ with a given type of position bias in mind. Thus, we consider a discrete random variable X whose PMF has been derived as described in the previous section and whose sample space is the set of choice items in the MCQ.

Sampling from the derived PMFs can be achieved either through the use of existing software tools such as R and Python or through the use of existing sampling techniques such as importance and rejection sampling (Martinez and Martinez 2016). The latter method requires a PDF; hence we opted for sampling indirectly through the original beta density as described below.

Sampling from the derived PMFs is done in two steps. First, we generate a sample y from the corresponding beta distribution. Second, we determine the value of X by assigning to it the choice item corresponding to the interval on which y lies (see Figure 3). Generally, if the sample space of the beta distribution is partitioned into N intervals of equal width, we assign X the choice item corresponding to the half-open interval [i, i+1/N) in which the value of y lies. For example, in Figure 3 if y = 0.6 we assign the choice item C to X.

Carrying out simulations

The proposed model for position bias is meant to be used in modelling the behaviour of an MCQ examinee when randomly guessing answers. Consequently, the model can be used in simulating the probability of passing an MC exam due to random guessing when taking position bias into consideration. The probability of passing an MC exam estimated this way is generally more plausible than the naive one used in existing studies. This is due to the probabilistic behaviour shown by people when choosing items from a list (Kuhn et al. 2020).

In carrying out the simulations, it was assumed that an examinee possesses varying levels of partial prior knowledge about the subject matter. This knowledge was expressed in percentages. With this assumption, for example, we assume an examinee with a 10% prior knowledge can score 10% in an exam (10 out of 100 questions) either legitimately or through educated guessing (see e.g., Wu et al. 2019). To incorporate this prior knowledge, we adopted the strategy used by Dubins et al. (2016). Using the strategy, an examinee is awarded points according to their level of knowledge. For example, if an examinee has a 10% level of prior knowledge, we award the examinee 10% (5 questions) and the remaining 90% (45 questions) is obtained through simulation (guessing). We further assume that an examinee does not answer the questions in any particular order and the type of position bias the examinee possesses is consistent, that is, it applies to all questions in the exam.

This setting somehow looks similar to Bernoulli trials except that the probabilities of success and failure change in each trial. For example, consider an examinee with middle bias; in guessing one question whose correct answer is A, the probability of success will be approximately 0.0016, while in answering the next question whose answer is, say C, the probability of success this time is approximately 0.6278 (see Table 1), which differs from that of the previous trial.

Another factor that makes Bernoulli trials unsuitable for this problem is the sheer number of possibilities in answering the questions. For example, how many ways can an examinee answer ten questions correctly? What about fifteen? etc. Answers to these questions are easy to find, but the problem is that there are just too many cases like this to consider to get the desired outcome.

We overcome these challenges by employing computer simulations. Computer simulations have been used in many contexts in education, including simulations aiming to study the effect of guessing in MC exams (e.g., Dubins et al. 2016). With the widespread availability of time-tested simulation algorithms and software tools, computer simulations have proven to be an indispensable tool when an analytic solution to a problem is either intractable or expensive to achieve, as it is the case with the problem being addressed in the present study.

In this study, computer simulations were carried out in Python (version 3.6.9) running under Ubuntu GNU/Linux 18.04. Eight different cases corresponding to the four types of position bias (including the naive case) and the two test exams were considered. As the cases were independent to one another, simulations were ran in parallel using separate processes to speed up the simulation process and take advantage of multiple processor cores in the host computer.

The simulation algorithm starts by initialising key parameters that control the simulation process. These parameters, their role in the simulation process and the values used in this study are shown in Table 2. The maximum number of simulation rounds to run (10000000) was determined experimentally by striking a balance between computing time and convergence of computed probability values.

Table 2: Global parameters used to control the simulation algorithm. Parameters shown with asterisks change as the algorithm progresses while the rest remain constant.

Parameter	**Meaning**	**Initial value used**
prior_knowledge	Examinee's percentage of prior knowledge	5–50 at increments of 5
exam_length	Number of questions in the exam	50
*num_simulations**	Number of simulation rounds to run	10000000

For each type of bias, and for each level of prior knowledge, the algorithm then runs num_simulations simulations against the test exam. In each simulation round, the following are performed: (1) drawing exam_length samples from a corresponding PMF; (2) translating the samples to a key of item choices; (3) comparing the sampled answers to the respective exam answer key and incrementing by one all min_questions items for which at least their number of questions were answered correctly. For example, if we compare the two keys and find the number of questions guessed correctly is 13, we increment the counters for the 5 and 10 min_questions by one. The final value of the counters is used to compute relative frequencies in the next step; and (4) computing relative frequencies of scoring different number of questions in min_questions (the probability of achieving different scores). Upon converging the algorithm returns relative frequencies for scoring at least each number of questions in *min_questions*.

Drawing samples from corresponding PMFs and translating the samples to an answer key was done as described in the previous section. Comparing the sampled answer key to the respective exam answer key was done by converting the answer keys to numeric vectors whereby the elements of the vectors became ASCII codes of the letters in the answer keys and then subtracting the two vectors. After subtracting the two vectors the number of zero elements, which denotes correct answers, was computed. Counter variables corresponding to each element of min questions were then incremented by one if the number of correctly answered questions was greater or equal to the corresponding element.

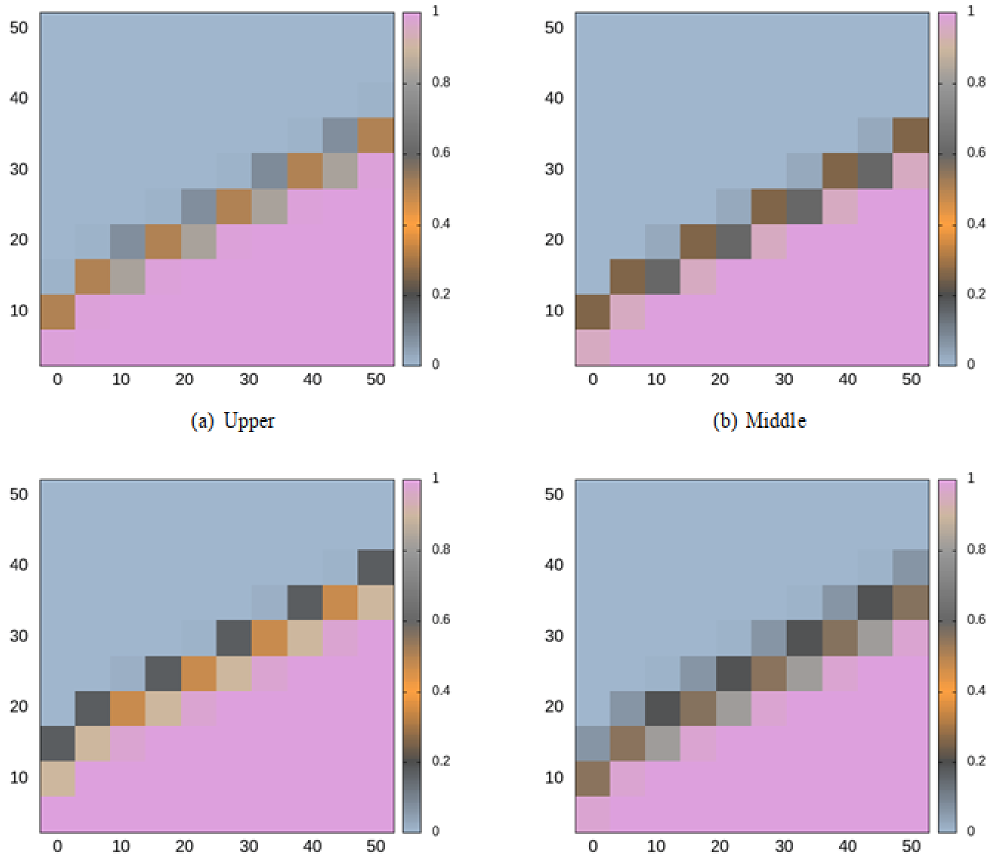
Relative frequencies of scoring at least the number of questions in each of *min_questions* were computed by dividing the corresponding counter variable by the value of *num_simulations* giving a vector of relative frequencies. Algorithm 1 summarizes the steps taken in a single simulation process.

Algorithm 1: Simulating the process of guessing answers in an MC exam. Variables shown in bold denote vector quantities.

```
min questions ← n              // Initialise parameters
exam length ← L
number of simulations ← S
for Each level of prior knowledge do
    counters ← 0
    for i ← 1, number_of simulations do
        Draw exam length samples from the respective PMF
        Translate the samples to choice items
        Get correct answers from samples and answer key
        correct_answers ← number of correct answers
        if correct answers ≥ min questions then
            counter ← counter + 1
        end if
    end for
    for i ← 1, numberofcounters do
        freq[i] ← counters/number of simulations
    end for
end for
return freq          // Vector of relative frequencies
```

Results and Discussion

Simulated joint PMFs for number of questions and level of prior knowledge for the two test exams and the four types of position bias (including the naive case) are shown in Figures 5 and 6. The general observation from the results is that the probability of passing an MC exam is higher than it might be anticipated. For example, to score 20 questions out of 50 an examinee just needs to have prior knowledge ranging between 25 and 30 percent even for a properly set exam (see Figure 5). This is an acceptable level of performance; for example, according to the National Examination Council of Tanzania (NECTA) (NECTA 2022b), a score of between 30 and 44 percent at ordinary secondary level (form I to form IV) is considered satisfactory (grade D) and obtaining four D grades in any non-divinity subject qualifies the candidate for diploma studies at a university level.



(a) Upper (b) Middle (c) Lower (d) Naive

Figure 5: Simulated joint PMFs showing the probability of scoring a given number of questions out of 50 (vertical axis) when possessing a given level of prior knowledge (horizontal axis) for the original test exam.

The present study aimed to quantify the chances of success in MC exams using computer simulations. Unlike many previous similar studies which ignored the presence of position bias normally inherent in examinees when attempting MCQs, in this study we took into account position bias with the aim of studying its impact in success rate when guessing in both a well set and a poorly set MCQ exams.

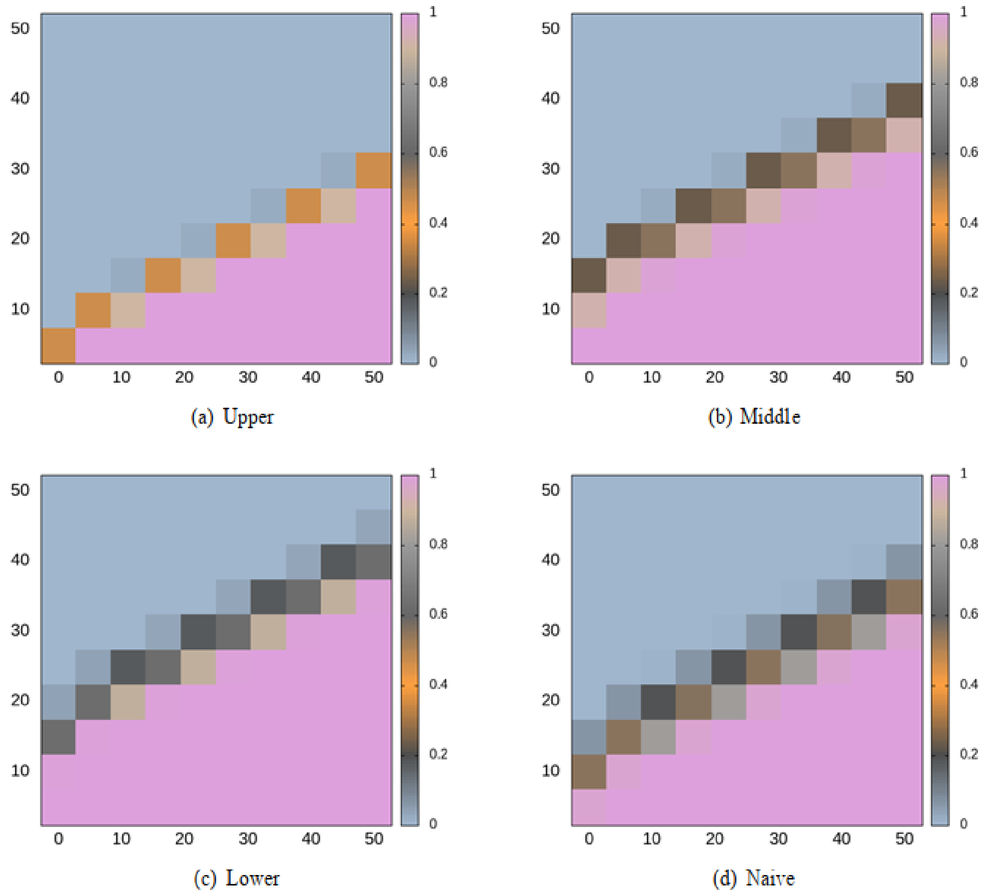


Figure 6: Simulated joint PMFs showing the probability of scoring a given number of questions out of 50 (vertical axis) when possessing a given level of prior knowledge (horizontal axis) for the modified test exam.

Typically, examinees would possess some partial knowledge that would give them an edge in the guessing process. Similarly, they can make use of educated guessing to eliminate some choice items as reported by Bailey et al. (2022). Thus, if we consider an average examinee to be the one with a 50% level of partial knowledge, these results show that an average examinee has an almost perfect chance of scoring 30 out of 50 questions in a well set MC exam (see Figure 5). This is an above-average performance corresponding to a C grade (classified as good performance by NECTA). This observation, together with the findings of Lions et al. (2022) that shows MC exam answer key distribution is often biased, clearly expresses a concern on the reliability of MC exams.

One way to alleviate this concern is to ensure that MC exams are set by following accepted standards and best practices to ensure their quality. Examples of such standards are given in Brame (2019) and Lions et al. (2022). Despite availability of such standards poorly constructed MCQs continue to make their way into the examination room. For example, a recent study by Kumar et al. (2021) analysed an MCQ bank meant for first year Bachelor of Medicine and Bachelor of Surgery Physiology students to determine its reliability found that 10% of questions in the bank were too easy with respect to their distractor effectiveness.

As expected, overall, the joint PMFs in Figure 5 show little variations compared to the ones in Figure 6. This observation exemplifies the effect of position bias. In the former case (the well set exam), position bias does not show a significant effect in the guessing process whereas in the latter case the effect is more pronounced. For example, looking at Figure 6 (b) and (c) we observe that it takes a relatively low level of prior knowledge to score the same score than is required in similar cases in Figure 5; for example, here an examinee can score more than 10 questions by pure guessing (0% prior knowledge) while it takes a higher level of prior knowledge to score the same number of questions in the original exam (Figure 5 (b) and (c)). This observation can be attributed to the fact that the modified test exam favours middle and lower edge bias, putting an examinee with other kinds of bias at a disadvantage (see Figure 6 (a)–(b)).

Several previous studies have reported relatively lower probabilities of passing MC exams through guessing than the ones reported in the present study (e.g., Dubins et al. 2016, Gu and Schwartz 2018). This difference is attributed to the fact that those studies assumed a naive (average) probability of picking an item choice. As reported in the literature (e.g., Attali and Bar-Hillel 2003), Kuhn et al. 2020), people usually tend to exhibit bias when choosing items from a list. The limitation of this assumption can be seen in Figures 5(d) and 6(d) which are identical despite the large difference in the distribution of answers in the two test exams.

Conclusion

This paper has addressed the question "what is the chance of passing a multiple-choice exam by random guessing?" through the use of computer simulations. Unlike many similar studies which ignore position bias, this paper takes into account position bias. Results show that the chance of passing a multiple-choice exam by random guessing is considerable even with a well set multiple-choice exam. A combination of pure chance and the partial/prior knowledge examinees possess makes this possible. Thus, though convenient in grading, multiple-choice exams can lead to false positives with regard to student understanding hence should be used with great caution especially in STEM subjects. The model for position bias developed in this paper can find uses in future similar studies, particularly those employing computer simulations.

References

Akyol P, Key J and Krishna K 2022 Hit or miss? test taking behavior in multiple choiceexams. *Ann. Econ. Stat*. 147: 3-50.

Attali Y and Bar-Hillel M 2003 Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *J. Educ. Meas*. 40(2): 109-128.

Bailey CD, Briggs JW and Irving JH 2022 Test-wiseness and multiple-choice accounting questions: Implications for instructors. *Issues Account. Educ*. 37(2): 1-14.

Blunch NJ 1984 Position bias in multiple-choice questions. *J. Mark Res*. 21(2): 216-220.

Brame CJ 2019 Writing exams: Good practice for writing multiple choice and constructed response test questions. In: Cynthia JB (Ed), Science Teaching Essentials (pp. 163-174), Academic Press.

Brown GT and Abdulnabi HH 2017 Evaluating the quality of higher education instructor-constructed multiple-choice tests: impact on student grades. *Front. Educ*. 2: 24.

Burton RF 2001 Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers. *Assess. Eval. High. Educ*. 26(1): 41-50.

Burton RF and Miller DJ 1999 Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainties attributable to guessing. *Assess. Eval. High. Educ*. 24(4): 399-411.

Clark EC 1956 General response patterns to five-choice items. *J. Educ. Psychol*. 47(2): 110-117.

Dehnad A, Nasser H and Hosseini AF 2014 A comparison between three-and four-option multiple choice questions. *Procedia Soc. Behav. Sci*. 98: 398-403.

Dubins DN, Poon GM and Raman-Wilms L 2016 When passing fails: Designing multiple choice assessments to control for false positives. *Curr. Pharm. Teach. Learn*. 8(5): 598-608.

Fagley NS 1987 Positional response bias in multiple-choice tests of learning: Its relation to testwiseness and guessing strategy. *J. Educ. Psychol*. 79(1): 95-97.

Forbes C, Evans M, Hastings N and Peacock B 2011 Statistical distributions, 4th Ed, John Wiley & Sons.

Gu Q and Schwartz M 2018 An automatic quantification of the randomness of answering correctly in taking traditional multiple-choice tests. *Theory Pract. Lang. Stud*. 8(9): 1152-1159.

Keinänen M, Ursin J and Nissinen K 2018 How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. *Stud. Educ. Eval.* 58: 30-36.

Kilgour JM and Tayyaba S 2016 An investigation into the optimal number of distractors in single-best answer exams. *Adv. Health Sci. Educ*. 21(3): 571-585.

Kleemola K, Hyytinen H and Toom A 2022 Exploring internal structure of a performance-based critical thinking assessment for new students in higher education. *Assess. Eval. High. Educ*. 47(4): 556-569.

Kuhn G, Pailhès A and Lan Y 2020 Forcing you to experience wonder: Unconsciously biasing people's choice through strategic physical positioning. *Conscious. Cogn*. 80: 102902.

Kumar D, Jaipurkar R, Shekhar A, Sikri G and Srinivas V 2021 Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Med J. Armed Forces India* 77(1): S85-S89.

Lions S, Monsalve C, Dartnell P, Blanco MP Ortega G and Lemarié J 2022 Does the response options placement provide clues to the correct answers in multiple-choice tests? A systematic review. *Appl. Meas. Educ*. 35(2): 133-152.

Lee CJ 2019 The test taker's fallacy: How students guess answers on multiple-choice tests. *J. Behav. Decis. Mak*. 32(2): 140-151.

Martinez W and Martinez A 2016 Computational statistics handbook with MATLAB, 3rd ed, Chapman & Hall/CRC.

McKenna P 2019 Multiple choice questions: answering correctly and knowing the answer. *Interact. Technol. Smart Educ*. 16(1): 59-73.

Moss E 2002 Multiple choice questions: Their value as an assessment tool. *Curr. Opin. Anaesthesiol*. 14(6): 661-666.

Parkes J and Zimmaro D 2016 Learning and assessing with multiple-choice questions in college classrooms, 1st ed, Routledge.

Slade C, Lawrie G, Taptamat N, Browne E, Sheppard K and Matthews KE 2022 Insights into how academics reframed their assessment during a pandemic: disciplinary variation and assessment as afterthought. *Assess. Eval. High. Educ*. 47(4): 588-605.

Stanger-Hall KF 2012 Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. CBE Life Sci. Educ. 11(3): 294-306.

NECTA (National Examinations Council of Tanzania) 2022a Primary school leaving examination. Available at: https://necta.go.tz/psle (Accessed: 06 April 2022).

NECTA 2022b Grading system. Available at: https://necta.go.tz/files/bango%20viwango%20Division.pdf (Accessed: 06 April 2022).

Wu Q, De Laet T and Janssen R 2019 Modeling partial knowledge on multiple-choice items using elimination testing. *J. Educ. Meas*. 56(2): 391-414.