sciendo

Acta Educationis Generalis
Volume 14, 2024, Issue 3

DOI: 10.2478/atd-2024-0022

Developing Distractors for Mathematics Multiple Choice Items: A Literature Review

Ato Kwamina Arhin*

Received: April 9, 2024; received in revised form: May 16, 2024; accepted: May 17, 2024

Abstract:
Introduction: This article aimed at digging deep into distractors used for mathematics multiple-choice items. The quality of distractors may be more important than their number and the stem in a multiple-choice question. Little attention is given to this aspect of item writing especially, mathematics multiple-choice questions. This article provides an engaging but succinct literature review on the development of plausible distractors for mathematics multiple choice items.
Methods: A literature review was conducted on developing effective distractors for mathematics tests. The review explored potential strategies for generating distractors that effectively assess students' understanding and problem-solving skills in mathematics.
Results: Alternative sources of distractors other than students' misconceptions are provided to aid the development of plausible distractors for mathematics multiple-choice items.
Discussion: Practical guidelines for judging distractors fit for mathematics questions are provided to help teachers improve their item writing skills based on literature and experience as a mathematics teacher. Common pitfalls in distractor development were identified to enable mathematics teachers to have a clear path for their work.
Limitations: This article focused on the development of plausible distractors specifically for mathematics multiple choice questions.
Conclusions: Test constructors must ensure aligning the distractors with the objectives of each lesson in order to make the distractors relevant to the demands of the item.

* Ato Kwamina Arhin, Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Department of Interdisciplinary Studies, Kumasi, Ghana; akarhin@aamusted.edu.gh; ORCID: 0000-0002-7702-5869

Key words: distractors, errors, mathematics, misconceptions and stem.

Introduction

Multiple-choice questions (MCQs) consist of a question or an incomplete statement, known as the stem, and a set of two or more options that contain possible answers to the question. The use of multiple-choice in assessing students' knowledge has become more popular in recent times due to the following reasons: Teachers with a large enrollment find the ease of scoring MCQs appealing, and well-constructed MCQs can yield test scores as reliable as a constructed-response test, allowing for a quick assessment of a large portion of the course's content coverage (Bacon, 2003). The student's task is to choose the option that provides the best or correct answer to the question posed. The best or correct answer is the key, and the remaining options are distractors. Only one option should be unequivocally correct, and the others should be unmistakably incorrect (Quagraine & Arhin, 2017). This description is the most appropriate for the options of a multiple-choice question. Distractors must be unambiguously wrong. Distractors should not confuse the test-taker, and they should be clearly and definitively incorrect. They should not be open to any other interpretation or discussion. If the distractors are not clearly and definitively incorrect, it can lead to confusion and uncertainty for the test-taker, making it difficult for them to accurately select the correct answer. This can ultimately undermine the validity and reliability of the assessment. Therefore, it is important for test writers to ensure that distractors are unambiguously wrong in order to effectively assess the test-taker's knowledge and understanding. The goal of distractors is to challenge students' understanding while maintaining fairness and validity in assessment.

It is relatively difficult to write MCQs, especially, creating excellent distractors (Quagraine & Arhin, 2017); it takes about twenty to sixty minutes to craft a quality multiple-choice item free from errors (Rush et al., 2016, as cited in Arhin et al., 2021). It is clear that writing MCQs is not a simple task; it becomes very elaborate when searching for or designing distractors that are plausible. For instance, when a mathematics teacher writes 50 MCQs with 5 options, he must tackle the challenging task of creating 200 distractors and 50 answers. In fact, the availability of high-quality distractors determines an MCQ's appropriate value. A successful distractor should be able to discriminate between informed and uninformed students.

Mathematical literacy is a critical attribute for individuals who live more effectively as constructive, concerned, and reflective citizens. Each level of mathematics may contribute differently; however, the students attain complementary skills and knowledge as they transition from one class to the next. Mathematics, therefore, ensures a child's total cognitive development. This necessitates the proper assessment of mathematical knowledge at all learning levels. The importance of mathematics in our lives leaves no room for poor assessment practices. Teaching and assessing students' knowledge in mathematics is of utmost importance to society at large. A wide range of purposes can utilize test results, such as evaluating school districts' performance, comparing students to their national peers, diagnosing learning difficulties, selecting students for programmes, grouping students for instruction, planning instructional activities, and changing curriculum (Etsey, 1997). This necessitates that the test used to assess students be error-free, so that the test results can be supported. A valid test result is dependent on the quality of individual test items.

Unfortunately, test constructors, both in low-stakes and high-stakes settings, regularly believe that excellent items are simple to craft. Research shows that effective item writing is a challenging process, and even the highest-stakes tests include poorly written items (Haladyna & Rodriguez, 2013). One drawback of MCQs is that they can encourage students to memorize answers by reusing stems, thus posing a threat to the validity of the test (Wood, 2009). This particular drawback necessitates that test writers rewrite the distractors to meet the lesson objectives and present them in a plausible way so that they can challenge students' understanding, particularly in mathematics.

Most MCQ stem developers lack the expertise and training required to create high-quality MCQ stems (Tarrant et al., 2006). It is correct to infer from (Tarrant et al., 2019) that the inability to construct effective stems also includes distractors. The development of plausible distractors is a critical aspect of creating effective multiple-choice test items. Plausible distractors are the incorrect answer options provided alongside the correct answer (key), aiming to challenge students' understanding and assess their misconceptions or incomplete knowledge. Content specialists and test developers describe distractor development as challenging and daunting, yet the test development community seems to prioritize it less (Gierl, Bulut, Guo, & Zhang, 2017). Research on item development appears to take an imbalanced approach, prioritizing the task of creating the stem and correct option over the task of developing distractors. This paper contributes to literature by filling an important gap. Research on distractor development frequently focuses on languages, such as distractor generation in non-mathematical subjects like English (Susanti et al., 2018; Zu et al., 2023), but this work focuses on developing distractors specifically for mathematics tests. This article aims to conduct a literature review on a neglected area of multiple-choice item construction: strategies for developing distractors that can distinguish between high and low achievers in mathematics tests. This article fills this gap in the literature, as most works on distractors are not specific to mathematics.

1 Literature review

1.1 Importance of distractors to a multiple-choice item

Most test developers and users believe that the stem and the correct option serve as the most important parts of MCQs (Thissen, Steinberg, & Fitzpatrick, 1989). Before teachers can analyze assessment results, they must create a valid and reliable assessment. While much effort is typically spent on crafting rigorous questions and identifying correct answers, developing incorrect answers (distractors) is equally important. Reflecting on distractors allows teachers to deeply consider the learning targets, potential student misunderstandings, and common mistakes. When writing an individual item, the number of potentially meaningful, relevant distractors is far more limited than the universe of items; the law of diminishing returns quickly takes hold. Once you've found three or four beneficial distractors, it becomes futile to find more (Wesman, 1971, p. 99-98).

Stiggins and DuFour (2009, p. 643) argue that teacher-made assessments are highly effective in determining student achievement because they allow teachers to combine their expertise to make informed instructional decisions based on the results. However, the results they refer to go beyond just the percentage of students who scored proficient. By thoroughly analyzing both correct and incorrect answers (distractors), teams can gain significant insights into students and their learning processes. Distractors help teachers provide additional instruction based on the most common errors made by individual students or groups (Popham, 2000, p. 244).

According to Gierl, Bulut, Guo, and Zhang (2017), distractors produce an important part of the context required to solve a multiple-choice item that can affect item quality and learning outcomes. The implausibility of distractors compromises the measurement of examinees' knowledge of a particular subject. Distractors reflect the learning objective. Impossible distractors reduce the effective number of choices. Within this context, a complex relationship exists between the correct and incorrect options due to the fact that students are required to make a distinction among response options in order to select the correct response (Hambleton & Jirka, 2006). The effects of partial knowledge on response performance nurture this multifaceted relationship, which in turn interacts with the plausibility of each distractor and can affect the psychometric properties of the correct and incorrect options.

Gierl et al. (2017) further indicated that this complex relationship (testing effect) between the correct and incorrect options can affect learning. When an assessment enhances memory retention, it triggers the "testing effect" (Roediger & Karpicke, 2006). Research has demonstrated that multiple-choice items can produce the testing effect by eliciting beneficial retrieval processes that, in turn, result in improved performance on subsequent examinations. However, the testing effect's benefits are dependent on the quality of the distractors. Competitive multiple-choice questions, where the wrong answer could be true and shares important information with the right answer, help students remember things when they have to study for future tests that test information about both the right answer and the wrong answer (Little, Bjork, Bjork, & Angello, 2012).

However, distractors unrelated to the correct option can introduce missing information into the assessment process, leading to a decrease in memory retention. The number of unrelated distractors used for an item can also adversely affect memory retention (Brown, Schilling, & Hockensmith, 1999; Butler et al., 2006). Most test developers have failed to establish the relationships distractors have with the knowledge of testees and, therefore, do not spend time and effort developing plausible distractors.

Second, distractor analysis can assist test developers and instructors in understanding why students make mistakes, providing guidance for diagnostic inferences about test performance. For instance, distractor analysis can reveal students' misconceptions, which can then guide the type of instruction and remedial lessons required to overcome these errors in thinking, reasoning, and problem solving (Briggs, Alonzo, Schwab, & Wilson, 2006). Based on the results of distractor analysis, instructors can identify content areas that require improvement and provide students with remedial instruction. The selection of a distractor provides the teacher with hidden information about students learning that, hitherto, would have escaped the teacher's watchful eyes. This calls for a more comprehensive analysis of all distractors. Non-functioning distractors are options that are rarely selected by examinees (<5%) or otherwise do not perform as expected. Therefore, we should either remove these options from the items or replace them with more plausible ones. A review of 477 functioning distractors across four MCQ assessments, revealed that students eliminated over 38% of these distractors due to a mere 5% selection rate (Haladyna & Downing, 1993). Overall, the percentage of items with three functioning distractors ranged from only 1.1% to 8.4% of all items. Conducting distractor analysis helps to sanitize the test and also gives the test constructor a deeper insight into the essence of constructing more functioning distractors.

Third, distractors serve as gatekeepers for multiple-choice test items. A plausible distractor is capable of separating examinees who have learned the material from those who have failed to learn. The purpose of a test is to give students a more credible and objective picture of their performance with concrete feedback concerning their strengths and weaknesses. To be able to provide such lucid feedback to students regarding the use of MCQs, distractors must provide excellent services. Therefore, plausible distractors help to determine the depth and accuracy of students' knowledge and identify gaps in their learning. A function of formative assessment is to employ quality distractors when using MCQs.

2 Assessing the quality of mathematics multiple-choice items

Teachers frequently measure mathematics assessment based on the educational concepts of mathematical content, learning, and equity. These educational concepts may appear to contradict the usual psychometric, technical, and practical principles used to determine the quality of tests and other evaluations. Recently, there has been a growing trend to consider the concepts of content, learning, and equity as derived from the psychometric traditions of the past. We may never definitively address the quest for the defining features of a high-quality mathematics test item. The literature on this subject has explored, espoused, and documented a variety of perspectives. Effective mathematics questions facilitate idea development, identify misconceptions, and present practical and theoretical applications (Massachusetts Department of Education). On the other hand, Romberg (1992) identified three primary thematic areas as criteria for evaluating effective mathematics questions as follows.

2.1 Alignment with goals

Good mathematics questions should align with mathematics education's goals. Consider whether the questions reflect the desired learning outcomes and curriculum objectives. Avoid questions that diverge from the intended focus. Test items should align with the contemporary understanding of the essence of mathematics. This perspective emphasizes the importance of students being able to recognize mathematical linkages in real-world problems and self-regulate their thought processes in order to solve problems effectively. Test items should align with contemporary knowledge of student learning. Current instructional theory posits that, students should be active participants in the learning process, responsible for constructing their own understanding.

2.2 Authentic performance assessment

Romberg (1992) emphasizes the importance of assessing authentic performance in mathematics. Authentic assessment goes beyond rote memorization and routine procedures. Questions should require students to apply mathematical concepts, solve problems, and demonstrate deeper understanding. Test items should be consistent with effective classroom instruction and not susceptible to manipulation or deviation from the curriculum. According to Hubbard (2001), effective mathematics items are ones that prompt students to contemplate outcomes rather than just achieve them.

2.3 Quality criteria

Romberg (1992) suggest that mathematics item writers consider factors such as clarity, relevance, cognitive demand, and alignment with instructional goals when evaluating the quality of an item. Well-constructed questions challenge students' thinking and provide meaningful insights into their mathematical abilities. Quality items promote students' development of relational knowledge, a process-oriented approach, and higher-level learning skills. Students' responses to high-quality questions should demonstrate the intellectual processes they used to solve them. Quality products motivate students to engage in both critical thinking and practical application (Hubbard, 2001). Well-posed questions and their answers will enhance a conducive setting for deep contemplation (Niss, 1993). The quality of assessment questions directly impacts students' learning experiences and the effectiveness of mathematics education. To maximize student learning, mathematics teachers must use quality items to assess students' mathematical knowledge, following Romberg's suggestions.

3 Assessing the quality of distractors for mathematics MCQs

Providing a definite list of criteria for judging the quality of a distractor is a non-exhaustive and uphill task, and I make no claim to have accomplished it in this paper. Instead, this paper outlines some typical qualities to consider when creating distractors for a mathematics multiple-choice test. According to Goodrich (1977), some qualities are readily apparent; the general reasonability of a distractor as a possible solution to the question posed is perhaps the most fundamental requirement. Measuring the reasonability of a distractor may be accomplished in several ways. To avoid immediate exclusion, the difference in value between a distractor and the key is potentially a beneficial baseline; a distractor should be context-specific, similar in magnitude to the true solution.

Zhang et al. (2020) state that an adequate distractor must meet the following criteria: (1) it must be an incorrect answer to the question; (2) it must be grammatically correct and consistent with the underlying article; (3) it must be semantically related to the correct answer; and (4) it must provide distraction, requiring some understanding of the underlying article to identify the correct answer. Even though Zhang et al. (2020) provided these criteria for judging the quality of a distractor do not relate to mathematics to a greater extent, these requirements tilt more towards language and social sciences. In a mathematics test, we can judge the quality of distractors using only two out of the four criteria given by Zhang et al. (2020).

The importance of quality distractors reduces the likelihood of students arriving at the correct response by eliminating other choices and, equally important, may allow identification of widespread student misunderstandings or inclinations that could lead to curricular or instructional changes and improvements. A good distractor can help to assess gaps in students' understanding. Based on experience and literature, I suggest that distractors fit for mathematics MCQs should be guided by the following criteria:

1. *Provide a rationale for each distractor*

For each distractor, write a rationale, such as a common misconception the response option represents or the process by which a student could have arrived at that incorrect answer. This will help the teacher to eliminate irrelevant or non-functional options before the test will be administered. Also, it helps the teacher in classifying each distractor so that it explains why a student has selected and incorrect option.

For example:
Where does the graph of y= $(\chi - 1)(\chi + 5)^2$ cross the x-axis?
A. -5 [the student takes the intercept from the wrong factor $(\chi + 5]$
B. -1 [the student uses the correct factor, but makes a mistake with the sign]
C. 0 [the student thinks that crossing any axis means $\chi = 0$]
D. 1 [the correct answer]
E. 5 [the student uses the incorrect factor, and makes a mistake with the sign]

This will serve as a good feedback to the teacher and students. The provision of rationales for each option will enable immediate feedback to be given to students, thereby, facilitating interactive formative assessment where individuals are urged to analyse and learn from their mistakes (Zendesk, 2022).
According to King, Gardner, Zucker, and Jorgensen (2004) Pearson's distractor rationale taxonomy for mathematics indicates four levels of understanding: level one includes the most fundamental errors; levels two and three correspond to responses that, while incorrect, indicate increasing sophistication in the student's response; and level four represents the correct response (key). Using the distractor rationale taxonomy, mathematics teachers can design distractors to contribute to an analysis of a student's pattern of misunderstanding in mathematics.

Table 1

A distractor rationale taxonomy for mathematics items

Level of understanding	*Student Error*
Level 1	Makes errors that reflect combinations of conceptual misunderstanding and unfamiliarity with how to relate operational procedures to problem contexts. Student attempts to implement strategies that are unrelated to the task at hand. These errors may indicate that the student has an inordinate dependence on information that is explicitly stated in item stimuli, and is lacking the sophistication to correctly answer the question being asked.
Level 2	Makes errors that reflect some sophistication and computational ability, but that demonstrate an inadequate conceptual framework and flawed reasoning in support of conclusions or inferences.
Level 3	Makes errors that reflect definite levels of sophistication in analysis and conceptual knowledge, but that are flawed by inconsistent reasoning or computational weakness.
Level 4	Correct response.

An example of a mathematics MCQ using the distractor taxonomy in Table 1.

Kofi spent GHc 222.46 to buy a climbing rope that sold for GHc3.50 per meter. What is the greatest number of meters of the rope he could buy at that rate?

A. 7786.1 m [Level 1: incorrect operation with place-value error]
B. 778.61 m [Level 2: incorrect operation, correctly applied]
C. 635.6 m [Level 3: correct operation with place-value error]
D. 63.56 m* [Level 4: correct response]

2. *The orientation (format) of the options must be different from the well-known presentations*

Students some of the times are fixed to one way of presentation of mathematical operations. For example, brackets () can be used in place of multiplication (x) and also division (÷) can be replaced by /. In the presentation of options teacher are therefore to vary the usage of these operations. Students are used to this type of orientation $\frac{\sqrt{3}}{\sqrt{2}}$ so instead of this presentation you can use $\sqrt{3} \div \sqrt{2}$ in providing options. This I have used often in mathematics test and you find students select the wrong options because the familiar presentations are not available. Also, instead of using (x) in this example 3a x 5c we can write it as 3a(5c) or 3a.5c.

For example:
Rationalize $\frac{\sqrt{2}}{\sqrt{3}}$
A. $3 \div \sqrt{6}$
B. $\sqrt{6} \div 3$
C. $\sqrt{6} \div 2$
D. $\sqrt{3} \div \sqrt{2}$

3. *It must agree with the stem (homogenous in content)*

The distractors must be homogenous with the stem, i.e., from the same area of content. The less homogenous the distractors the more easily they give away the key. That is, it is more difficult for test takers to discriminate between items that are parallel than items that are dissimilar. Options that are dissimilar from the correct answer may be easier to eliminate. This suggests that "test performance could result from test-wiseness rather than knowledge of the test material, (Ascalon, et. al., 2007).

For example:
Which of the following is a plane figure?
a. Cube
b. Cylinder
c. Sphere
d. Triangle

The answer (triangle) is too obvious, the rest of the options are all 3-demensional figures.

An improved version regarding the options:
Which of the following is a plane figure?
a. *Circle
b. Sphere
c. Tetrahedrone
d. Triangular prism

The student has to decide between circle and square. Circle is a round figure and has curved boundary. It's two-dimensional but it's not made up of straight lines.

4. *The key must be relevant to the stem's content*

Detailed content and cognitive domain descriptions must be organized for each item. Items aligned with the content topics and cognitive domains must be designed to collect evidence about what students know and are able to achieve. This helps the classroom teacher to assess students' performance.

5. *It must be capable of digging deep into the students' knowledge base*

The essential role of the incorrect options is to serve as the answer to students who did not achieve the learning outcome but ignored by students who did achieve the learning outcome. A distractor must be capable of preventing students from engaging in blind guessing. A student should not just eliminate a distractor without solving the problem at hand. If the distractors cannot 'painstakingly search' into the student's knowledge base, then the correct answer could be concluded easily. As a result, the discrimination of the question will be lowered, and the test will also lose the ability of the assessment.

6. *It has to be culturally and gender-sensitive to avoid bias*

A distractor must not favour or appeal especially to a group of students. Also, it should not seem to be negatively affecting a section of students. Distractors that are appealing to boys but unappealing to girls must be avoided. The bias review guarantees that distractors are clearly worded, are of appropriate difficulty and interest level, are unbiased, and will result in a full range of responses. A good question a teacher can ask: "Will these options favour or unfairly punish examinees on the basis of personal characteristics such as gender, race, ethnicity, religion, socioeconomic status, disability, or geographic region?" Distractors must be bias free.

4 Errors committed by item writers

A consequence of poorly constructed distractors is that it leads to inaccurate evaluation of students' performance. Use between three to five alternatives per question. Research shows that three-choice items are about as effective as four or five-choice items, mainly because it is difficult to come up with plausible distractors. Limiting the number of options enhance item writing by preparation of fewer distractors and save examinee's time to complete exam (Haladyna, Downing, & Rodriguez, 2002; Tarrant, Ware, & Mohammed, 2009).

Here are some common errors committed by test constructors when writing plausible distractors for mathematics tests. These errors reduce the plausibility of the distractors.

1. Overly obvious distractors: Distractors that are obviously incorrect can make it simple for students to eliminate them, reducing the effectiveness of the multiple-choice item.
2. Distractors unrelated to the concept under assessment can confuse students and result in an inaccurate assessment of their understanding.

3. Inclusion of correct elements: Distractors that contain elements of the correct answer can make it easier for students to identify the correct response. The inclusion of elements of correct responses will easily give the correct option out. A test-wise student will quickly spot the correct option.
4. Failing to consider students' common errors and misconceptions can lead to distractors who do not effectively assess their understanding.
5. Distractors that don't sufficiently cover the range of concepts under assessment can reduce the multiple-choice item's validity.
6. Failure to update distractors: Using the same distractors repeatedly without updating them can lead to familiarity and reduced effectiveness in assessing student understanding.
7. Distractors that do not align with the specific learning objectives under assessment may not effectively measure students' mastery of the material.
8. Using contradictory distractors: Contradictory distractors are those that negate the key or other options. For instance, if the answer is "a square is a rectangle" and there is an option which states that "a square is not a rectangle" is contradictory because it is false. To avoid this, ensure that each option is distinct and independent from the other options.

For example, which of the following is a true statement about the function $f(x) = x \sin x$. The function is f(x)=xsinx.

A. an even periodic function
B. is even but not periodic
C. *odd but not periodic
D. an odd periodic function

5 Strategies for developing distractors for mathematics MCQs

When it comes to generating distractors for questions, there are various methods to consider, depending on the type and purpose of the question. For instance, you can create distractors for mathematical items by altering the sign, unit, order, or magnitude. In recent times distractors can be generated using automated algorithms but majority of classroom teachers cannot afford this sophisticated method of generating distractors. Existing works conduct some attempts on generating short distractors (Stasaski & Hearst, 2017; Guo et al., 2016). Therefore, there must be a way to help mathematics to generate plausible distractors.

1. Distractors could be based on common misconceptions or errors that students are likely to make when solving a problem. For example, in a question about fractions, a distractor could be a common error, such as

adding the numerators and denominators instead of finding a common denominator.

Question: If a cake recipe calls for ⅓ cup of sugar and you want to double the recipe, how much sugar will you need?

A. $\frac{1}{6}$ cup
B. $\frac{2}{3}$ cup
C. $\frac{3}{3}$ cup

2. Related concepts: Distractors could be based on related mathematical concepts that are similar to the correct answer but not exactly the same. This strategy focuses on similarities between the answer and distractors. The key is manipulated to provide new set of options which become the distractors. For example, the key can be negated, multiplied or divided by a factor. Here the item writer will exercise his/her discretion as to how to manipulate the key. Unlike the first strategy this is not based on common errors or misconceptions. Because of this limitation students with incomplete knowledge on the topic may be able to guess the correct option.
According Dave, Owen, Pursel, and Giles (2021) some common ways to manipulate the key are listed below:
   - Change One Sign: Randomly pick one coefficient or constant in an equation / inequality and multiply by -1.
   - Change Two Signs: Randomly pick two coefficients or constants of an equation/inequality and multiply by -1.
   - Most Frequent Number: Use the most frequent number in the equation/inequality as a distractor.
   - Nearest Multiple: Randomly pick a coefficient or a constant in an equation/inequality and change it to the nearest multiple of 2, 3 or 5.
   - Random Drop: Randomly drop one of the coefficients or constants in an equation/inequality.
   - Invert Range: Invert the solution range of the inequality, e.g., change [0, 1] to (-∞, 0) ∪ (1, ∞).
   - Trivial Solution: For inequality problems, one of the distractors can be chosen from {φ}, (-∞, ∞), or No Solution. For equations, choices are from 0, -1 or 1-.
   - Flip brackets: Change an open bracket in answer to closed and vice and versa. In the question, order of operations can be changed by changing the position of brackets.

3. Incorrect operations: Distractors could be based on applying the wrong mathematical operation to the given problem. For example, in a question about order operations, a distractor could be the result of adding instead of subtracting or multiplying instead of dividing.

   For example:
   Question: Solve: 10 ÷ 2 + 3 - 5 x 1
   A. 5
   B. 3
   C. 0
   D. -3

Overall, developing plausible distractors for mathematics multiple choice items requires a deep understanding of students' misconceptions, common errors, and related mathematical concepts in order to effectively assess their understanding and problem-solving skills. The test constructor's ability to align the items to the requirements of the curriculum is key. Crafting effective distractors requires careful consideration, especially in mathematics. By following these principles, educators can create high-quality MCQs that accurately assess students understanding and problem-solving abilities.

6 Negative practices in writing distractors

These are errors that test constructors often commit during item writing, and they affect the quality of distractors. Test writers use two such errors in the hope that it will help them develop plausible distractors.
Extraneous information: The stem may include irrelevant or extraneous information (window dressing), which may confuse students and lead them to choose the wrong answer. In the context of MC item stem construction, the term "window dressing" refers to extraneous information, such as words, phrases, or entire sentences that are not relevant to the question's stated problem (Board & Whitney, 1972). For example, in a question about solving word problems, a distractor could include unnecessary data that does not affect the solution. This approach to writing distractors is disingenuous in the sense that the test constructor has provided irrelevant information with the sole aim of confusing the students. All item writers should avoid using window dressing (excessive verbiage) (Haladyna et al., 2002). MCQs containing window dressing (or excess verbiage) have increased difficulty, are less reliable, and are less valid (Koepf, 2018). Parkes and Zimmaro (2016, p. 34-23) contend that window dressing can negatively affect validity by confusing the student, as well as causing the student to "spend time reading and deciding on the relevance of the information in which they don't actually need to engage in order to answer the question".

Ambiguous language: Using ambiguous or vague wording can create distractions, potentially leading to multiple interpretations of the problem. For instance, ambiguous language in a probability question could serve as a distractor, making it unclear what event the question is asking for. Ambiguous stems affect the ability of students to answer a question correctly. Ambiguous questions can confuse even knowledgeable students and cause them to answer incorrectly. Therefore, we must completely avoid using an ambiguous stem, as it invalidates the test.

7 Implications for practice

Developing plausible distractors for mathematics multiple choice tests is a crucial aspect of creating high-quality assessments to effectively assess students' mathematical understanding and problem-solving skills. The purpose of assessment is to diagnose, instruct, and grade students. If not done properly, the consequences of using assessment are grave. In order to achieve maximum impact, it is imperative to get the right mix of things regarding student assessment. Diagnostic decisions are made about individual students, as well as about group needs. Gathering information in a specific area enables the teacher to identify areas of progress and focus attention. The assessment results dictate the type of instruction provided, and conversely, the type of instruction influences the types of assessments used and their outcomes. Grades serve as a form of motivation for students and also provide feedback to parents, stemming from the quality of the assessment. All these decisions are based on credible measurement information, which requires an assessment tool devoid of errors that can easily be prevented by creating plausible distractors. Finally, there is general mathematics anxiety among students at all levels of education. It can stem from a variety of factors, including past negative experiences, a lack of confidence in one's abilities, fear of failure, and societal pressure. However, the list often overlooks poor assessment practices, as a contributing factor to mathematics anxiety. Poor assessment practices may include using inappropriate item formats and not adhering to suggestions for writing test items, resulting in poor-quality items. According to Arends, Winnaar, and Mosimege (2017), teachers' clarity, communication skills, content knowledge, and assessment procedures significantly impact students' achievement in mathematics.

Conclusions

In conclusion, the development of plausible distractors for mathematics multiple-choice questions is both art and science. The science of creating distractors includes the recent development of automated distractors. Creating distractors is also a multifaceted task that requires careful consideration of various factors, including difficulty, discriminative ability, and the impact on student performance. The test constructor must align the distractors with the objectives of each lesson to make them relevant to the item's demands.

The literature underscores the importance of creating functioning distractors and the potential impact of distractor quality on assessment validity. Multiple-choice items written using the mathematics distractor rationale taxonomy are capable of revealing a student's breakdown in understanding through the incorrect responses selected by the student. An assessment system that incorporates this methodology, can indicate a student's instructional needs in a subject area or the behavioural outcome and thereby contribute to the development of a special intervention plan.

It also important that test constructors, whether classroom or high-stakes, to invest resources in developing plausible distractors that effectively assess students' mathematical understanding and problem-solving skills. Further research should focus on the use of automated-generated distractors in mathematics to simplify the assessment process for teachers. Finally, there is a need for more research on distractor performance in multiple-choice tests from different perspectives, including observational and item-analytic viewpoints. Furthermore, we should conduct research at various educational levels, from pre-school to university.

References

Ascalon, M. E., Meyers, L. S., Davis, B. W. & Smits, N. (2007). Distractor similarity and item-stem structure: Effects on item difficulty *Applied Measurement in Education, 20*(2), 153-170.

Arends, F., Winnaar, L., & Mosimege, M. (2017). Teacher classroom practices and mathematics performance in south African schools: A reflection on TIMSS 2011. *South African Journal of Education, 37*(3). https://doi.org/10.15700/saje.v37n3a1362

Arhin, A. K., & Yamoah, J. (2023). Identifying item flaws to improve the quality of core mathematics multiple-choice items. *Journal of African Education, 4*(3), 247-274.

Arhin, A. K., Essuman, J., & Arhin, E. (2021). Analysis of item writing flaws in a communications skills test in a Ghanaian University. *African Journal of Teacher Education*, *10*(2), 121-143. https://doi.org/10.21083/ajote.v10i2.6762

Board, C., & Whitney, D. R. (1972). The effect of selected poor item-writing practices on test difficulty, reliability, and validity. *Journal of Educational Measurement, 9*(3), 225-233.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11,* 33-63. https:doi.org/10.1207/s15326977ea1101_2

Brown, A. S., Schilling, H. E., & Hockensmith, M. L. (1999). The negative suggestion effect: Pondering incorrect alternatives may be hazardous to your knowledge.

Journal of Educational Psychology, 91, 756-764. https://doi.org/10.1037/0022-0663.91.4.756

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology, 20*, 941-956. https://doi.org/10.1002/acp.1239

Dave, N., Owen, R. B., Pursel. B., & Giles, C. L., (2021). Math multiple choice question solving and distractor generation with attentional GRU networks. In *Proceedings of the 14th International Conference on Educational Data Mining (EDM21)* (pp. 422-430). International Educational Data Mining Society. Retrieved from https://educationaldatamining.org/edm2021/

Etsey, Y. K. (1997). *Teachers' and school administrators' perspectives and use of standardized achievement tests: A review of published research.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Goodrich, H. C. (1977). Distractor efficiency in foreign language testing. *TESOL Quarterly, 11*(1), 69-78.

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research, 87*(6), 1082-1116.

Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement 53*(4), 999-1010.

Haladyna, T. M., Downing, S. M., & Rodriquez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Hambleton, R. K., & Jirka, S. J. (2006). *Anchor-based methods for judgmentally estimating item statistics*. In S. M. Downing, & T. Haladyna (Eds.), *Handbook of test development* (pp. 399-420.). Mahwah, NJ: Erlbaum.

Hubbard, R. (2001). The why and how of getting rid of conventional examinations. *Quaestiones Mathematicae Supply, 1*, 57-64.

King, K. V., Gardner, D. A., Zucker, S., & Jorgensen, M. A. (2004). *The distractor rationale taxonomy: Enhancing multiple-choice items in reading and mathematics*. San Antonio, TX: Pearson.

Koepf, T. M., (2018). *The Effect of Item Stem and Response Option Length on the Item Analysis Outcomes of a Career and Technical Education Multiple Choice Assessment* (Doctoral thesis). Retrieved from https://scholarworks.wmich.edu/dissertations/3366

Massachusetts Department of Education. (1987). *The Massachusetts Educational Assessment Program*. Quincy: Massachusetts Department of Education.

Niss, M. (1993). *Investigations into assessment in mathematics education*. *An ICMI Study.* Kluwer Academic Publishers.

Parkes, J., & Zimmaro, D. (2016). *Learning and assessing with multiple-choice questions in college classrooms*. New York, NY: Routledge.

Popham, W. J. (2000). *Educational Measurement* (3rd ed.). Boston, MA: Allyn and Bacon.

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education, 4, 1301013*.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249-255. https://doi.org/10.1111/ j.1467-9280.2006.01693.x

Romberg, T. A. (1992). *Mathematics Assessment and Evaluation, Imperatives for Mathematics Educators*. State University of New York.

Stiggins, R., & DuFour, R. (2009). Maximizing the power of formative assessments. *Phi Delta Kappan, 90*(9), 640-644. https://doi.org/10.1177/003172170909 000907

Susanti, Y., Tokunaga, T., Nishikawa, H., & Obari, H. (2018). Automatic distractor generation for multiple-choice English vocabulary questions. *Research and Practice in Technology Enhanced Learning, 13*(1), 15.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high-stakes nursing assessments. *Nurse Education Today, 6*(6), 354-363. https://doi.org/10.1016/j.nedt.2006.07.006

Tarrant, M., Ware. J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education, 9,* 40.

Thissen, D., Steinberg, L., & Fitzpatrick, A. R. (1989). Multiple-choice models: The distractors are also part of the item. *Journal of Educational Measurement, 26*(2), 161-176.

Vyas, R., & Supe A. (2008). Multiple choice questions: A literature review on the optimal number of options. *National Medical Journal of India, 21*(3),130-133. PMID: 19004145.

Wesman, A. G. (1971). Writing the test item. In R. L. Thorndike (Ed.), *Educational Measurement (2nd ed.)* (pp. 81-129). Washington, D. C.: American Council on Education.

Wood, T. J. (2009). The effect of reused questions on repeat examinees. *Advances in Health Sciences Education 14,* 465-473. https://doi.or/10.1007/s10459-008-9129-z

Zhang, C., Yicheng, S., Chen, H., & Wang, J. (2020). Generating adequate distractors for multiple-choice. In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020)-1* (pp. 310-315). https://doi.org/10.5220/0010148303100315