MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2024, 20(10), em2519
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/15431

An analysis of calculus application questions in grade 12 mathematics exams in South Africa

Eric Machisi [1*]

[1] Fujairah Boys' School, Fujairah, UAE

Received 31 July 2024 ▪ Accepted 10 August 2024

Abstract

This study employed a quantitative survey methodology to analyze the 2021-2023 grade 12 calculus application exam questions in the context of South Africa. A total of 11 grade 12 teachers and 220 grade 12 students participated in the study. Analysis of the survey data and students' test results points to the view that examiners sometimes misjudge the level of difficulty of the exam questions and include questions that are very difficult, which fail to discriminate between strong and weak candidates. It was noticed that calculus applications questions that most students could not correctly answer included multiple mathematical skills that were compacted in a short statement. This paper recommends that mathematics examiners should continuously reflect on their own work and find ways to improve their skills in setting future exam papers. The inclusion of item difficulty and item discrimination indices in post-exam analysis of results could be part of the examiners' reflective practice. Feedback from the grade 12 mathematics teachers on the standard of each exam question is just as important to the examiners as the examiners' reports are to the teachers.

Keywords: calculus application, exam questions, item difficulty, expert opinion, discrimination index

INTRODUCTION

The Importance of Learning Calculus

Calculus is an essential branch of mathematics that plays a crucial role in various fields such as physics, chemistry, engineering, commerce, business, and computer science. It deals with rates of change in systems and the summation of infinitely many small quantities, making it a foundational subject for students pursuing majors in these disciplines. According to Georgetown University (2022), calculus is a prerequisite for entrance into renowned academic programs. Therefore, excelling in high school calculus can significantly enhance a student's chances of gaining acceptance into top-tier institutions (Burdman, 2019). Mastering calculus not only prepares students for college or university-level mathematics but also opens numerous opportunities for future academic and professional success. For these reasons, calculus has long been considered the gold standard of high school mathematics curricula (Dennon, 2022).

South African Students' Performance on Calculus Application Questions

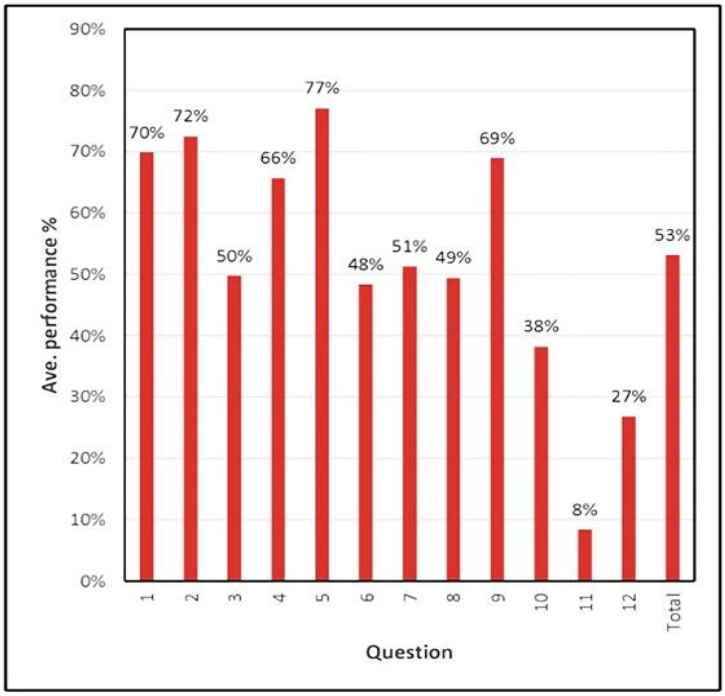
In the grade 12 mathematics paper 1 exam in South Africa, calculus is one of the two topics that enjoy the biggest share in terms of mark allocation. It is therefore crucial for high school students to perform well in this area to secure a higher score on the exam. Unfortunately, examination diagnostic reports released by the South African Department of Basic Education in the period 2021-2023 indicate that students perform worse in calculus application than in any other aspect of the grade 12 mathematics paper 1 exam. According to **Figure 1**, just 8% of the sampled grade 12 mathematics paper 1 candidates correctly solved the calculus application question (Q.11) in 2021. This was the lowest among the average performance per question percentages.

A similar trend was reported in 2022 and 2023 (Department of Basic Education, 2023a, 2024). This trend

✉ e.machisi@yahoo.com **(*Correspondence)**

Contribution to the literature

- This study analyzed grade 12 calculus application exam questions in South Africa to see if they meet the standards for good exam questions.
- Contrary to the popular belief that students' poor performance in mathematics exams is a consequence of poor teaching in the classroom, this study revealed that mathematics examiners can also contribute to the high failure rate by setting questions that are so difficult that even the stronger candidates cannot attempt.
- The findings of this study prompt mathematics examiners to continuously reflect on their work and find ways to improve their skills in setting future exam papers.



Q	Topics
1	Equations, Inequalities & Algebraic Manipulation
2	Number Patterns & Sequences
3	Number Patterns & Sequences
4	Number Patterns & Sequences
5	Functions & Graphs
6	Functions & Graphs
7	Functions & Graphs
8	Finance
9	Calculus
10	Calculus
11	Calculus
12	Probability & Counting

Figure 1. The average performance per question (Department of Basic Education, 2022a, p. 186)

limits students' chances of gaining access to university admission, as a poor performance in calculus can significantly lower their overall grade in mathematics. According to South Africa's national senior certificate examination diagnostic data, most candidates did not attempt the calculus application questions in 2021 and 2022 (Department of Basic Education, 2022a, 2023a). The diagnostic reports urge that teachers devote adequate time to the calculus application concept to ensure that students thoroughly understand the idea. Thus, the examiners believe that students' underperformance on calculus application questions is caused by teachers not doing justice to the topic. On the other hand, teachers are not given the opportunity to review exam questions and provide helpful suggestions for future changes. While examiners consider syllabus requirements while setting examination questions, there is a human element to drafting exam questions (Huntley et al., 2010). The examiners use a significant amount of personal judgement to determine "how difficult the question will be for the students" (Huntley et al., 2010, p. 145). It is likely that students' underperformance on calculus application questions could be due to an examination standard that is set too high for them. As a result, it is critical to analyze the exam questions themselves to see if their depth and complexity fit the students' cognitive abilities.

Purpose of the Study

This paper examines the calculus application questions set in the South African grade 12 mathematics national examinations in 2021, 2022, and 2023 to see if they meet the standards for good exam questions. The parameters for the analysis are item difficulty index, expert opinion deviation, and discrimination index. The findings of this study may help mathematics examiners reflect on their work and find ways to improve how they set exams in future.

LITERATURE REVIEW

Challenges in Setting Good Exam Questions

Setting good exam questions is a difficult task. It is an art that requires specific skills and creative thinking on

the part of the examiner. Huntley et al. (2010) proposed two techniques to generate good exam questions: proactive and retroactive. In the proactive approach, examiners follow a set of guidelines to ensure that the exam questions are valid, clear, and appropriate for the cognitive level of the students. However, following guidelines does not guarantee that the exam questions will be good for the students (Huntley et al., 2010). There have been numerous instances where examiners wrote exam questions in good faith, believing them to be relatively straightforward, only to be astonished by how difficult the questions turned out to be for the students. The performance of South African grade 12 students on calculus application exam questions in recent years serves as an example. Post-exam diagnostic reports show that grade 12 students in South Africa have consistently performed poorly on calculus application exam questions (Department of Basic Education, 2022a, 2023a, 2024). As a result, there is increasing pressure on mathematics teachers in South Africa to enhance their pedagogy.

Although encouraging teachers to enhance their calculus application teaching is a good thing, we should also acknowledge that calculus application is a higher-order thinking skill (Djidu et al., 2021). Calculus application in secondary school includes word problems in which students must derive the optimum function from the given information and then find values that optimize the function. Several studies have found that most students struggle with word problems in mathematics (Chinn, 2020; Geisler & Rolka, 2021). As a result, students' difficulties with calculus application questions may be exacerbated by the exam question style. This is because the process of creating exam questions at the appropriate cognitive level is not automatic. Examiners use intuition to determine the level of difficulty when constructing exam questions (Gierl et al., 2017). It is possible to under or overestimate the exam question's difficulty level. Exam questions that are either too easy or too difficult might demotivate students and fail to distinguish between high and low achievers (Lee et al., 2024).

The Retroactive and Proactive as Complementary Approaches to Setting Good Exam Questions

Clearly, a proactive approach to exam question writing is not enough. The retroactive approach is needed to complement the proactive approach. The retroactive approach is based on post-exam item analysis, which examines students' performance using metrics such as mean, standard deviation, item difficulty, and item discrimination (Kumar et al., 2021). Post-examination item analysis will assist examiners in improving question construction in future examinations, which will benefit students (Hartati & Yogi, 2019). In South Africa, post-exam item analysis has traditionally concentrated on calculating average performance per question as indicated in diagnostic reports (Department of Basic Education, 2022a, 2023a, 2024). The South African post-exam diagnostic reports do not include parameters such as item difficulty, item discrimination, or expert opinion deviation. Thus, it is considered that examiners always do an excellent job and that students' difficulties with exam questions are due to insufficient education in the classrooms. (Department of Basic Education, 2022a, 2023a, 2024). This is a limited view of the nature of the problem. The researcher in this study believes that the complexity of the calculus application exam questions should also be investigated to ensure that the examiners improve their exam question writing skills. No study has examined South Africa's grade 12 mathematics examination papers to assess variables such as item difficulty, item discrimination, and expert opinion variation to enhance exam question construction in the future. This study aims to fill this research gap.

THEORETICAL FOUNDATION

In this study, the researcher adapts the item analysis model created by Huntley et al. (2010) as a theoretical lens for analyzing the grade 12 calculus application exam questions. The adapted model uses item difficulty index, expert opinion deviation, and item discrimination index as variables for analyzing the suitability of the exam questions.

According to Boateng et al. (2018), the item difficulty index measures the proportion of correct answers to a particular question. It ranges from 0 to 1. A high item difficulty index implies that the question was simple, and a higher number of candidates successfully answered it. An item difficulty index around 0 indicates that just a few candidates properly answered the question. Acceptable item difficulty indices range from 0.2 to 0.8 (Mitra et al., 2009). Good exam questions should have a difficulty index of 0.3 to 0.6 (Patil et al., 2016). Items with difficulty indices that fall outside of this range should be revised because they are considered too difficult or too easy.

Expert opinion deviation is the difference or gap between the expected student performance based on experts' judgements of the exam question's level of difficulty and the actual student performance assessed using the item difficulty index (Huntley et al. 2010). The expert opinion deviation is represented by the vertical distance from the line of best fit for each exam question. A bigger variance shows a mismatch between experts' expectations and candidates' experience on the exam question as shown in **Figure 2**. Point (x_i, y_i) lies far away from the line of best fit which means that the experts misjudged the difficulty level of the question as per student performance.

Item discrimination is a statistic that indicates how well an exam question distinguishes between high and low-performing students (Huntley et al., 2010). In an

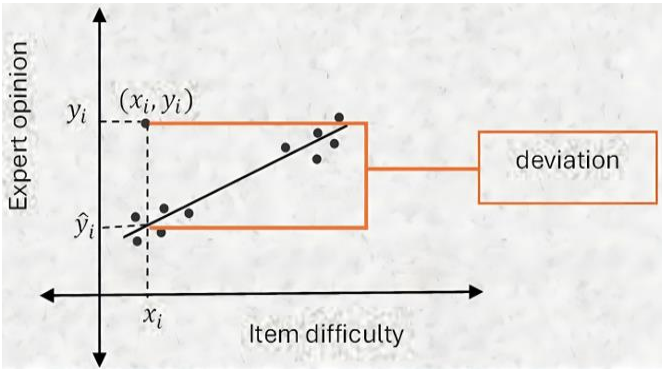


Figure 2. Illustration of expert opinion deviation (adapted from Huntley et al., 2010, p. 158)

ideal circumstance, stronger candidates should outperform weaker candidates on exam questions. Item discrimination is computed by taking the difference between the number of students who got the question correct in the upper group and the number of students who got the item correct in the lower group, divided by the total number of students who took the test (Huntley et al., 2010). The discrimination index value ranges from -1 to 1. A negative discrimination index indicates that more low-performing students successfully answered the question, which is an unusual situation. Questions with a negative discrimination index are invalid and should be removed (Gamage et al., 2019). If an equal number of high and poor performing students answered the question correctly, or if none of the students in both groups answered correctly, the discrimination index is 0. This demonstrates that the question is ineffective in distinguishing between strong and weak candidates. Such questions need to be revised. Good questions should have a discrimination index close to 1, but values greater than 0.2 are acceptable (Kheyami et al., 2018).

The decision to discard, revise, or include a question in an assessment should be based on an interpretation of the difficulty index and the item discrimination index in tandem. Padua and Santos (1997) provided a set of criteria to assist researchers in analyzing item difficulty and discrimination indices collaboratively. **Table 1** displays the proposed criteria. If a test question is difficult and does not differentiate between stronger and weaker students, it should be removed. However, if it is difficult but accurately distinguishes between stronger and weaker students, it should be included in the assessment. A moderately difficult item that is moderately discriminating or not discriminating between stronger and weaker students should be revised.

The next section describes the methodology used to collect and analyze research data.

METHODOLOGY

This study adopted the quantitative survey approach.

Table 1. Table for interpreting item difficulty and item discrimination indices (Padua & Santos, 1997)

Difficulty index	Interpretation	Discrimination index	Interpretation
0.00 → 0.20	Very difficult	−1.00 → −0.60	Questionable
0.21 → 0.40	Difficult	−0.59 → −0.20	Not discriminating
0.41 → 0.60	Moderately difficult	−0.19 → 0.20	Moderately discriminating
0.61 → 0.80	Easy	0.21 → 0.60	Discriminating
0.80 → above	Very easy	0.61 → 1.00	Very discriminating

Difficulty level	Discriminating level	Action
Difficult	Not discriminating	Discard
	Moderately discriminating	Revise
	Discriminating	Include
Moderately difficult	Not discriminating	Revise
	Moderately discriminating	Revise
	Discriminating	Include
Easy	Not discriminating	Discard
	Moderately discriminating	Revise
	Discriminating	Revise

Participants and Context

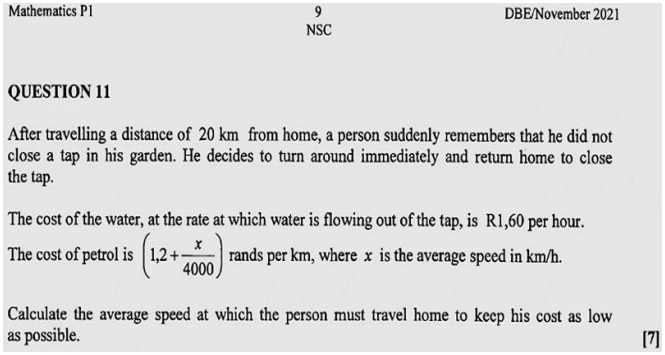
This study included 11 grade 12 mathematics teachers and a sample of 220 grade 12 students from schools in the Limpopo Province of South Africa. The teachers had more than 5 years of grade 12 mathematics teaching experience and were also involved in marking grade 12 mathematics paper 1 national exams. The teacher participants were recruited using snowballing or chain referral sampling. Snowballing is a sampling approach in which the researcher selects one or two people who meet the required criteria and asks them to recruit others or introduce the researcher to other possible participants. This approach is recommended in cases where participants with the required characteristics may be hard to reach (Dosek, 2012). The student participants consisted of ten high performers and ten low-achieving students in grade 12 mathematics from each teacher's class. The sample size of students per school was determined using the upper 27% and lower 27% criteria for determining item difficulty and item discrimination (Rush et al., 2016).

Instruments

Data were gathered using a closed response survey instrument adapted from Huntley et al. (2010) and a test instrument created by the researcher using 2021, 2022, and 2023 calculus application exam questions. The survey instrument requested teacher participants to rate each of the three calculus application exam questions, as follows:

Option 1. Students should find the question easy

Option 2. The question is of average difficulty



Mathematics P1 9 DBE/November 2021
NSC

QUESTION 11

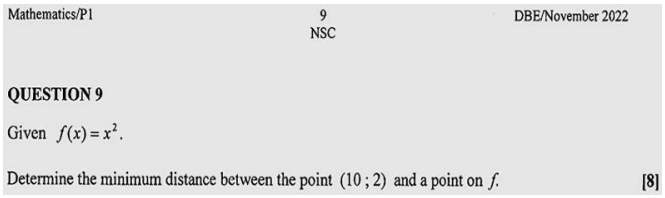
After travelling a distance of 20 km from home, a person suddenly remembers that he did not close a tap in his garden. He decides to turn around immediately and return home to close the tap.

The cost of the water, at the rate at which water is flowing out of the tap, is R1,60 per hour.

The cost of petrol is $\left(1{,}2+\frac{x}{4000}\right)$ rands per km, where x is the average speed in km/h.

Calculate the average speed at which the person must travel home to keep his cost as low as possible. **[7]**

Figure 3. 2021 grade 12 calculus application exam question (Department of Basic Education, 2021, p. 9)



Mathematics/P1 9 DBE/November 2022
NSC

QUESTION 9

Given $f(x)=x^2$.

Determine the minimum distance between the point $(10\,;2)$ and a point on f. **[8]**

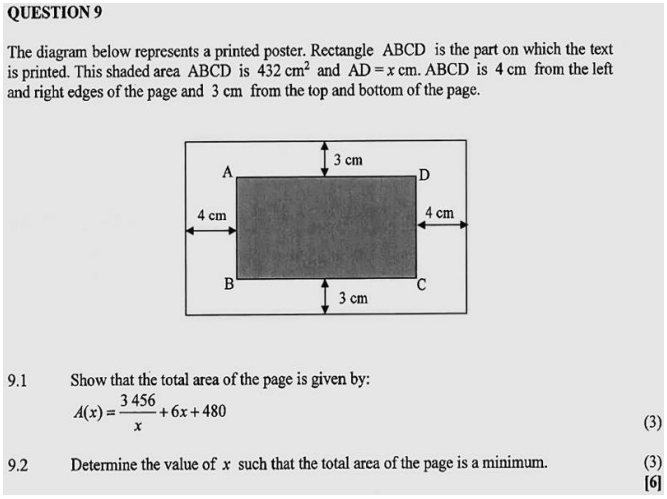
Figure 4. 2022 grade 12 calculus application exam question (Department of Basic Education, 2022b, p. 9)

Option 3. Students should find the question very difficult

The survey instrument's validity was evaluated by a panel of seven university professors in mathematics education. The experts were asked to score the survey's content on a scale of 1 (not relevant), 2 (somewhat relevant), 3 (quite relevant), and 4 (highly relevant). Each survey item's content validity index (CVI) was computed by dividing the number of experts who rated it 3 or 4 by the total number of experts. The total scale CVI was calculated by averaging the item CVIs. An average CVI of 1 was obtained, indicating that all seven experts agreed that the survey instrument accurately measured what it was designed to measure. The Cronbach alpha reliability coefficient was used to examine the survey instrument's internal consistency. Cronbach's alpha reliability coefficient of 0.74 was achieved, which is within acceptable limits (Taber, 2018).

The test instrument used to assess students' performance was just a compilation of the 2021, 2022, and 2023 calculus application questions from South Africa's grade 12 exam papers. **Figure 3**, **Figure 4**, and **Figure 5** display the questions that were used:

A panel of ten grade 12 mathematics teachers reviewed the test instrument's validity and unanimously concluded that the test items were clear and relevant to the construct under inquiry. An average CVI of 1 was obtained, indicating complete agreement among the raters that the test items measured exactly what they were meant to measure. The internal consistency of the test instrument was tested using a test-retest reliability technique with a sample of 20 grade 12 students from non-targeted schools. Two weeks were permitted between the first and second testing. A Pearson



QUESTION 9

The diagram below represents a printed poster. Rectangle ABCD is the part on which the text is printed. This shaded area ABCD is 432 cm^2 and AD = x cm. ABCD is 4 cm from the left and right edges of the page and 3 cm from the top and bottom of the page.

9.1 Show that the total area of the page is given by:

$$A(x)=\frac{3\,456}{x}+6x+480$$ (3)

9.2 Determine the value of x such that the total area of the page is a minimum. (3)
[6]

Figure 5. 2023 grade 12 calculus application exam question (Department of Basic Education, 2023b, p. 8)

correlation coefficient of 0.85 was obtained, indicating that the test instrument had excellent internal consistency (Noble et al., 2021).

Data Collection Procedures

Data were collected from April to July 2024. The purpose of the study was communicated to the teachers during the recruitment process. They were informed that their participation was voluntary and that they had the opportunity to withdraw from the study at any time without explanation. The teachers were informed that the participants' names would remain anonymous throughout the study. The participants signed informed consent forms at the start of the data collection process.

The survey was administered to the eleven grade 12 teachers online using google forms and the response rate was 100%. The selected teachers administered the test in their classrooms during regular lessons. The test was taken after the calculus applications topic. The test was part of the teachers' end of topic informal assessment of students' performance. As a result, the administration of the test did not interfere with the normal functioning of the school. All twelfth-grade mathematics students in the selected teachers' classes took the test. However, only the top and bottom ten students were considered for analysis. The duration of the test was one hour. The teachers marked the test scripts. The teachers were asked to mark students' scripts according to the guidelines they use for the twelfth-grade final exams. The test results were recorded on a uniform score sheet prepared by the researcher, which required the teachers to write down the students' scores for each question. The ten schools were coded A through K. The top ten students from school A were assigned the codes AU1, AU2, …, and AU10. The bottom ten students from school A were assigned the codes AL1, AL2, …, and AL10. A similar coding system was used for the top and bottom ten students from schools B through K. The teachers were asked to submit their test scores to the researcher by July

20, 2024. All scores were submitted by the due date. The researcher wrote emails to the teachers thanking them for their participation and assistance in the study.

Data Analysis Procedures

The data were analyzed using the item difficulty index, expert opinion deviation, and item discrimination index. The item difficulty index was calculated by taking the average score per item divided by the total mark allocation for the item (Nitko, 2004):

$$IDI = \frac{Average\ score\ per\ item}{Total\ marks\ allocated\ to\ the\ item}, \quad (1)$$

where *IDI* is item difficulty index. The results were evaluated using the criteria established by Padua and Santos (1997).

Expert opinion as used in this study refers to the rating that had the highest frequency in each survey question. Expert opinion deviation was determined by plotting the item difficulty indices against transformed and adapted expert opinion ratings from the survey data. The expert opinion survey used a scale of [1, 3] where 1 represented 'easy', 2 represented 'average difficulty', and 3 represented 'difficult'. The survey scale was standardized so that the expert opinion ratings could be represented by values between 0 and 1 using the transformation formula suggested by Huntley et al. (2010):

$$y = \frac{x-a}{b-a}, \quad (2)$$

where x is the dominant expert rating; a is the minimum value of the original scale; and b is the maximum value of the original scale. The transformed ratings were then reversed by subtracting them from 1 so that they match the difficulty index scale where values close to or equal to 0 represent difficult questions and values close to or equal to 1 represent easy questions. A scatter plot was then drawn using the item difficulty indices and the standardized and adapted expert opinion ratings. A line of best fit was added to the data. A point that is distant from the line of best fit indicates a concerning discrepancy between what experts believe about the calculus application question and actual student performance. This becomes a subject for additional discussion.

The discrimination index was calculated by taking equally sized groups of high and low performing students and applying the following formula:

$$Discrimination\ index\ (DI) = \frac{H-L}{N}, \quad (3)$$

where *H* is the number of high performing students that answered the question correctly; *L* is the number of the low performing students that answered the question correctly and *N* is the total number of students in both groups. The output was interpreted using the criteria set by Padua and Santos (1997).

Table 2. Item difficulty indices for the 2021-2023 calculus application exam questions

Exam question	November 2021	November 2022	November 2023
Average score	1.5	0.75	3.5
Total mark allocation	7	8	6
Difficulty index	**0.21**	**0.09**	**0.58**

RESULTS AND DISCUSSION

Item Difficulty Results

Table 2 displays the item difficulty indices for the calculus application test questions based on the sample data. Using the standards established by Padua and Santos (1997), the 2021 question is classed as 'difficult'. The 2022 question is categorized as 'very difficult', while the 2023 question is rated as 'moderately difficult'.

Although the item difficulty index alone does not indicate if a question is good or not, several researchers concur that items with difficulty indices below 0.2 should be discarded (Date et al., 2019; Ntumi et al., 2023; Sharma, 2021). Based on this criterion, the 2022 grade 12 calculus application exam question was not fit to be included in the national examination in South Africa. This result does not mean that the concept of calculus application should not have been examined in that particular year. Very difficult exam questions are not fair to the students and fail to serve the purpose of assessment and evaluation of students' performance. Such questions should be avoided by the examiners in future examinations. It may be important to examine the 2022 grade 12 calculus application question here to determine what makes the question so difficult for students. The question reads, as follows:

Given $f(x) = x^2$. Determine the minimum distance between the point (10; 2) and a point on f (Department of Basic Education, 2022b, p. 9).

No diagram was provided to help candidates understand the question. Candidates were expected to come up with their own sketch diagram to visualize the problem, imagine their own point $P(x; y)$ on f, generate their own optimum function using distance formula, determine the value of x that minimizes the distance, and finally compute the minimum distance. The generated optimum function $\left(d = \sqrt{x^4 - 3x^2 - 20x + 104}\right)$ contains a fourth-degree polynomial, the derivative of which is a cubic function. It then requires knowledge of the factor theorem to obtain the x value that minimizes the function. Thus, the question contained multiple concepts compacted by the examiner into a short statement. While the question may pique the interests of the examiner, it does not do good for the students. The question itself is divorced from real life scenarios which adds to its complexity. All these factors amplify the cognitive load of the question and

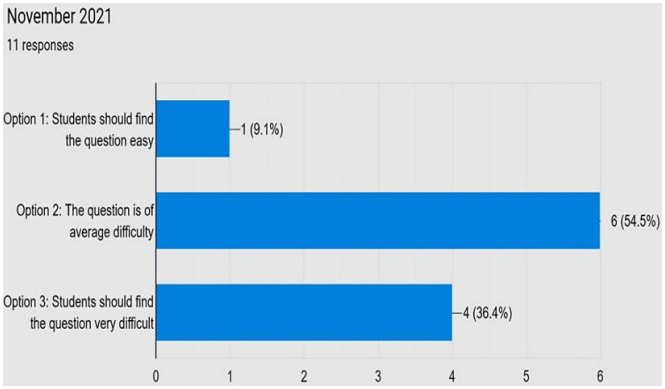


Figure 6. Experts' opinions on the 2021 grade 12 calculus application exam question (Source: Author's own elaboration)

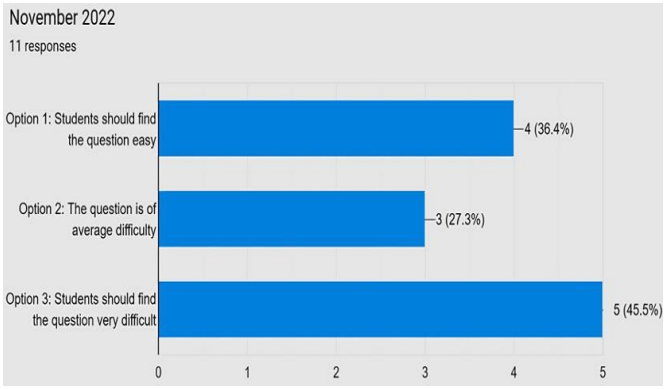


Figure 7. Experts' opinions on the 2022 grade 12 calculus application exam question (Source: Author's own elaboration)

increase its level of difficulty. A possible way of moderating the complexity of the 2021 calculus application exam question and reducing its cognitive load is to include a diagram and then split the question into separate parts to give students multiple chances of getting part marks on the question. This explains why the 2023 grade 12 calculus application exam question has a 'good' item difficulty index. In 2023 grade 12 mathematics paper 1 exam, the calculus application question included a diagram, and the question was split into two parts. Students who failed to derive the optimum function in the first part of the question could still get some marks from the second part of the question, which is a good style of assessing students' performance. Now, let us see what the selected grade 12 mathematics teachers thought about the 2021-2023 calculus application exam questions.

Expert Opinion Results

Figure 6 shows that most of the survey respondents believed the question was of 'average difficulty' while students' performance on the question shows that the question was 'difficult'. Thus, the experts' views differed from students' actual performance. The 2021 grade 12 examination diagnostic report indicated that most candidates did not attempt this question which means it was difficult for them (Department of Basic Education, 2022a). Thus, it is possible for experts including examiners to intuitively misjudge the level of difficulty of an exam question. The takeaway point here is that examiners should not rely on intuition alone in determining the level of difficulty when setting exam questions. Analysis of past records of students' performance per each question may help examiners set future exam questions at the appropriate cognitive level.

Figure 7 shows that most of the survey respondents believed that the 2022 grade 12 calculus application exam in South Africa was very difficult. This matches perfectly with the item difficulty index obtained in **Table 2**.

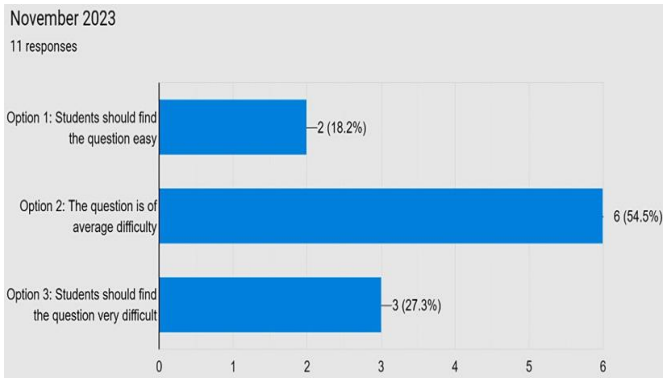


Figure 8. Experts' opinions on the 2023 grade 12 calculus application exam (Source: Author's own elaboration)

In the 2022 grade 12 examination diagnostic report, the examiners acknowledge that most students could not attempt the calculus application question and suggestions are made as to how teachers can help students improve their performance on the concept (Department of Basic Education, 2023a).There is no evidence of the examiners' acknowledgement of the fact that the 2022 grade calculus application exam question was very difficult. Thus, the examiners themselves do not see any problem with the standard of the exam question whereas teachers who had more than five years of teaching different cohorts of grade 12 students and marking grade 12 mathematics exams believed the standard was beyond reach for most students. This finding suggests that examiners should give teachers a platform to comment on the standard of exam papers as a reflective practice that will help examiners improve on setting future exam questions. Examiners are human beings and as such, it is possible for them to err in setting exam papers.

Figure 8 shows that most respondents believed that the 2023 grade 12 calculus application exam question was of average difficulty, and this matches perfectly with the item difficulty index obtained in **Table 2** presented earlier. Based on experts' opinions and item difficulty index, it appears that the 2023 grade 12 mathematics paper 1 examiner did a perfect job.

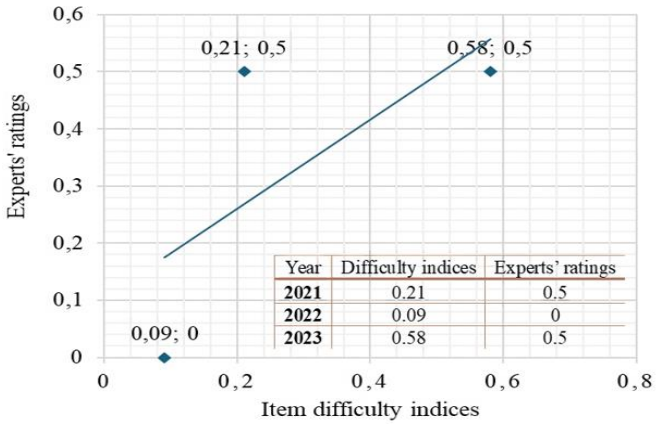


Figure 9. Scatterplot of item difficulty indices versus adapted expert ratings (Source: Author's own elaboration)

Table 3. Item discrimination indices of the 2021-2023 grade 12 calculus application exam questions

Exam paper	November 2021	November 2022	November 2023
Discrimination index	0.014	0.00	0.41

However, we will have to consider the item discrimination index to make a final determination on the standard of the 2023 calculus application exam question.

Figure 9 shows the relationship between the transformed and adapted experts' opinion ratings on each question and the actual performance of the participating students. The data summarizes the views expressed up to this point.

Point (0.09, 0) represents the 2022 grade 12 calculus application exam question which was confirmed to be very difficult by the item difficulty index and experts' ratings. The point lies very close to the origin (0, 0). Item discrimination will determine whether it was worthy including such a question in the 2022 grade 12 mathematics national senior certificate examination in South Africa. Point (0.21, 0.5) represents the 2021 grade 12 calculus application exam question whose difficulty level was underestimated by the expert raters. This point lies furthest from the trend line and can therefore be categorized as an outlier. Point (0.58, 0.5) which represents the 2023 grade 12 calculus application exam question lies very close to the trend line. This represents the ideal situation in terms of the difficulty index and the experts' rating. However, item discrimination will determine whether the 2023 exam question was ok, as it is or needed some revision.

Item Discrimination

Table 3 shows the item discrimination indices of the 2021-2023 grade 12 calculus application exam questions. Using the Padua and Santos (1997) criteria, only the November 2023 exam question (DI = 0.41) falls in the acceptable range of discrimination indices. Thus, the question was of a good standard and meets the widely accepted criteria for inclusion in the final examination of students' performance. The 2021 grade 12 calculus application exam question which was found to be at the 'difficult' level is moderately discriminating (DI = 0.014) and therefore the style of questioning used by the examiner falls short of the requirements for a good exam question. Examiners should revise such questions in future exam papers. The 2022 grade 12 calculus application exam question which was found to be very difficult is poorly discriminating (DI = 0.00) between high and low-performing students. Therefore, the style of questioning used by the examiners should also be revised in future exam papers.

Table 4. Final analysis of the 2021-2023 grade 12 calculus application exam questions

Exam question	Level of difficulty	Discrimination	Recommended action
2021	Difficult	Moderately discriminating	Revise
2022	Very difficult	Poorly discriminating	Revise or discard
2023	Moderately difficult	Discriminating	Include

Table 4 summarizes the findings of the study to guide examiners in future assessments.

Comparing Findings with Reports from Other Countries

The results of this investigation are consistent with reports from other nations. The 2016 Kenya certificate of secondary education examinations were found to be so difficult that the difference between strong and weak candidates was not discernible (Issah, 2022a). Both groups of candidates failed the exams equally, and it was determined that the national exams were flawed (Issah, 2022a). Since then, stakeholders have clamored for exam standardization so that they may reliably assess both academic abilities and variability of academic performance across students. The suggested standardization approach includes providing tryout exam questions to students in multiple schools across the country, allowing examiners to adjust the actual exam questions (Issah, 2022b). However, some academics suggest that tryout questions may lead to exam question leakage; instead, they advocate post-exam item analysis for future improvements (Rafi et al., 2023).

The presence of questions with unsatisfactory item difficulty and discriminatory indices in school exams is not unique to South Africa. An analysis of 237 mathematics examination scripts in Indonesia revealed that 20% of the exam items had unacceptable item difficulty indices and 45% had unacceptable item discrimination indices (Tjabolo & Otaya, 2019). Similar reports have surfaced in Botswana (Adedoyin & Mokobi, 2013), Nigeria (Amadi et al., 2019), Pakistan

(Syeda et al., 2021), and the Philippines (Balila & Cajilig, 2016). The apparent consensus among these reports is that very easy and very difficult exam questions do not reflect students' real academic performance. It is the moderately difficult exam questions that effectively distinguish between high- and low-performing students. These findings are consistent with those observed in this investigation.

Following these multiple reports, there is an increasing push for changes in the development of exam questions. For instance, Adedoyin and Mokobi (2013) and Tjabolo and Otaya (2019) suggested that examiners improve the quality of exam items by considering the characteristics proposed in the item response theory, such as item difficulty and item discrimination indices, while developing exam items. This is in line with the standardization processes proposed by Issah (2022b). Other scholars have advocated for the use of artificial intelligence to pick exam items from question banks (Gardner et al., 2021). This is a fascinating concept that warrants more examination. The takeaway message here is that there are numerous ways to improve the quality of exam questions and ensure that students' academic achievement is fairly evaluated. Exams will lose significance if they do not reveal how students' performance varies (Issah, 2022a).

RECOMMENDATIONS AND CONCLUSION

This study was prompted by a succession of poor performance by grade 12 mathematics students on calculus applications exam questions in South Africa. Contrary to the popular belief that students' failure to perform well on calculus application exam questions is due to poor teaching by teachers, the findings of this study revealed that examiners may err in their judgement of the question's level of difficulty and discriminating power. This results in the inclusion of very difficult questions that do not discriminate well between stronger and weaker students.

The researcher suggests that grade 12 mathematics post examination diagnostic reports should contain an analysis of the exam questions' difficulty level and item discrimination indices. In addition, grade 12 mathematics teachers should be given the opportunity to comment on the exam's standard as part of the post-exam review. Teachers' feedback should be given to examiners for consideration in future exam papers. Complex calculus application questions, such as those found on the 2022 grade 12 mathematics paper 1 exam, in which multiple mathematical skills were examined in one short question, should be avoided in future exams. Such questions have a high cognitive load, which will frustrate even the strongest candidates. The examiners may not find anything wrong with this type of questioning unless someone prompts them to reflect on their standards for setting exam questions. Examiners are human beings who can make mistakes and perform below expectations. Therefore, their job should be analyzed and appraised and constructive criticism of the examiners' work should be permitted.

Funding: No funding source is reported for this study.

Ethical statement: The author stated that the study did not require ethical approval. The study did not involve any experimental work or manipulation of the human participants. Written informed consents were obtained from the participants. The author further stated that the study did not interfere with the normal day to day operations of schools, teachers, and students. The collected data were treated as confidential information used exclusively for research purposes. It is impossible to identify the participants from the data used in this study.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the author.

REFERENCES

Adedoyin, O. O., & Mokobi, T. (2013). Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items. *International Journal of Asian Social Science, 3*(4), 992-1011.

Amadi, B. D., Ogunka, R. I., & Tandi, L. I. (2019). Estimating the psychometric properties of 50 multiple choice mathematics item for senior secondary II students in Ikwerre Local Government Area of Rivers State using classical test theory. *The African Journal of Behavioural and Scale Development Research, 1*(2).

Balila, J. S., & Cajilig, N. G. (2016). Psychometric analysis of mathematics achievement test using item response theory. *Jurnal Universitas Advent Indonesia, 1*(2), 34-44.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Burdman, P. (2019). Why calculus? Why indeed? *JustEquations*. https://justequations.org/blog/why-calculus-why-indeed

Chinn, S. (2020). *The trouble with maths: A practical guide to helping learners with numeracy difficulties* (4th ed.). Routledge. https://doi.org/10.4324/9781003017714

Date, A. P., Borkar, A. S., Badwaik, R. T., Siddiqui, R. A., Shende, T. R., & Dashputra, A. V. (2019). Item analysis as tool to validate multiple choice question bank in pharmacology. *International Journal of Basic & Clinical Pharmacology, 8*(9), 1999-2003. https://doi.org/10.18203/2319-2003.ijbcp20194106

Dennon, A. (2022). Want to get into Harvard? Ace calculus. *BestColleges.com*. https://www.bestcolleges.com/news/analysis/2022/02/28/college-admissions-calculus/

Department of Basic Education. (2021). National senior certificate grade 12 mathematics P1 November 2021. *Department of Basic Education*. https://www.education.gov.za/Curriculum/NationalSeniorCertificate(NSC)Examinations/2021NSCExamPapers.aspx

Department of Basic Education. (2022a). 2021 national senior certificate diagnostic report part 1: Content subjects. *Department of Basic Education*. https://www.education.gov.za/?fileticket=4JVOD_6ncds%3D&tabid=92&portalid=0&mid=4359

Department of Basic Education. (2022b). National senior certificate grade 12 mathematics P1 November 2022. *Department of Basic Education*. https://www.education.gov.za/Curriculum/NationalSeniorCertificate(NSC)Examinations/2022NovemberExams.aspx

Department of Basic Education. (2023a). National senior certificate 2022 diagnostic report part 1: Content subjects. *Department of Basic Education*. https://www.education.gov.za/?fileticket=4JVOD_6ncds%3D&tabid=92&portalid=0&mid=4359

Department of Basic Education. (2023b). National senior certificate grade 12 mathematics P1 November 2023. *Department of Basic Education*. https://www.education.gov.za/Curriculum/NationalSeniorCertificate(NSC)Examinations/2023NSCNovemberpastpapers.aspx

Department of Basic Education. (2024). 2023 diagnostic report book 1. *Department of Basic Education*. https://www.ecexams.co.za/2023_NSC_Exam_Results/Diagnostic%20Report%202023%20Book%201.....pdf

Djidu, H., Jailani, J., & Retnawati, H. (2021). Higher-order thinking skills among male and female students: An experimental study of the problem-based calculus learning model in secondary schools. *Beta: Jurnal Tadris Matematika, 14*(2), 107-125. https://doi.org/10.20414/betajtm.v14i2.432

Dosek, T. (2012). Snowball sampling and Facebook: How social media can help access hard-to-reach populations. *PS: Political Science & Politics, 54*(4), 651-655. https://doi.org/10.1017/S104909652100041X

Gamage, S. H., Ayres, J. R., Behrend, M. B., & Smith, E. J. (2019). Optimising Moodle quizzes for online assessments. *International Journal of STEM Education, 6*, Article 27. https://doi.org/10.1186/s40594-019-0181-4

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning, 37*(5), 1207-1216. https://doi.org/10.1111/jcal.12577

Geisler, S., & Rolka, K. (2021). "That wasn't the math I wanted to do!"–Students' beliefs during the transition from school to university mathematics. *International Journal of Science and Mathematics Education, 19*(3), 599-618. https://doi.org/10.1007/s10763-020-10072-y

Georgetown University. (2022). The case for de-emphasizing calculus in college admissions. Access and affordability. *Georgetown University*. https://feed.georgetown.edu/access-affordability/the-case-for-de-emphasizing-calculus-in-college-admissions/

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. https://doi.org/10.3102/0034654317726529

Hartati, N., & Yogi, H. P. (2019). Item analysis for a better quality test. *English Language in Focus, 2*(1), 59-70. https://doi.org/10.24853/elif.2.1.59-70

Huntley, B., Engelbrecht, J., & Harding, A. (2010). A model for measuring the quality of a mathematics question. *Far East Journal of Mathematical Education, 5*(2), 141-171.

Issah, Z. (2022a). Education in Kenya: The problem with KCPE and KCSE. *Medium*. https://zinahissa.medium.com/education-in-kenya-the-problem-with-kcpe-and-kcse-9ff91b921730

Issah, Z. (2022b). Standardizing Kenya's national exams. *Medium*. https://zinahissa.medium.com/standardizing-kenyas-national-exams-d1d907d429b1

Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos University Medical Journal, 18*(1), 68-74. https://doi.org/10.18295/squmj.2018.18.01.011

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India, 77*(1), 85-89. https://doi.org/10.1016/j.mjafi.2020.11.007

Lee, H. Y., Yune, S. J., Lee, S. Y., Im, S., & Kam, B. S. (2024). The impact of repeated item development training on the prediction of medical faculty members' item difficulty index. *BMC Medical Education, 24*(1), Article 599. https://doi.org/10.1186/s12909-024-05577-x

Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *IeJSME, 3*(1), 2-7. https://doi.org/10.56026/imu.3.1.2

Nitko, A. (2004). *Educational assessment of students* (4th ed.). Pearson/Merill Prentice Hall.

Noble, S., Scheinost, D., & Constable, R. T. (2021). A guide to the measurement and interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences, 40*, 27-32. https://doi.org/10.1016/j.cobeha.2020.12.012

Ntumi, S., Agbenyo, S., & Bulala, T. (2023). Estimating the psychometric properties ("item difficulty, discrimination and reliability indices") of test items using Kuder-Richardson approach (KR-20). *Shanlax International Journal of Education, 11*(3), 18-28. https://doi.org/10.34293/education.v11i3.6081

Padua, R., & Santos, R. G. (1997). *Educational evaluation and measurement: Theory, practice, and application*. KATHA Publishing.

Patil, R., Palve, S. B., Vell, K., & Boratne, A. V. (2016). Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. *International Journal of Community Medicine and Public Health, 3*(6), 1612-1616. https://doi.org/10.18203/2394-6040.ijcmph20161638

Rafi, I., Retnawati, H., Apino, E., Hadiana, D., Lydiati, I., & Rosyada, M. N. (2023). What might be frequently overlooked is actually still beneficial: Learning from post national-standardized school examination. *Pedagogical Research, 8*(1), Article em0145. https://doi.org/10.29333/pr/12657

Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education, 16*, 250. https://doi.org/10.1186/s12909-016-0773-3

Sharma, L. R. (2021). Analysis of difficulty index, discrimination index and distractor efficiency of multiple choice questions of speech sounds of English. *International Research Journal of MMC, 2*(1), 15-28. https://doi.org/10.3126/irjmmc.v2i1.3

Syeda, Z. F., Nazir, H. M., & Shah, A. A. (2021). Semi standadrdization of a test of secondary school mathematics. *Pakistan Journal of Humanities & Social Sciences Research, 4*(2), 1-50. https://doi.org/10.37605/pjhssr.v4i2.61

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Tjabolo, S. A., & Otaya, L. G. (2019). Quality of school exam tests based on item response theory. *Universal Journal of Educational Research, 7*(10), 2156-2164. https://doi.org/10.13189/ujer.2019.071013