Journal of Education for Sustainable Development Studies
ISSN 3048-2054 (online) & 3048-2259 (print)
December 2025, Vol. 2, No. 2, pp. 156–174

Research Article

From Stated Knowledge to Certainty of Thought: A Study on Four-Level Tests in Mathematics Teaching

Senad Orhani[1]

[1]Faculty of Education, University of Prishtina, Prishtina, Kosovo

Abstract
Traditional approaches to assessment in mathematics are often limited to measuring final results, without providing sufficient insight into students' cognitive processes and confidence in their answers. The Four-Tier Diagnostic Test (FTDT) represents an innovative assessment instrument designed to evaluate not only the correctness of knowledge but also students' confidence levels, reasoning accuracy, and ability to distinguish between accurate understanding and misconceptions. This study aims to analyze the application of FTDT in mathematics teaching, focusing on identifying conceptual difficulties and the structure of students' beliefs regarding their knowledge. The research was conducted with 70 lower secondary school students from "Heronjtë e Lumës" in Vërmica, Prizren, Kosovo, through a specially designed diagnostic instrument covering topics such as algebra, geometry, probability, and statistics. Data were analyzed using descriptive and inferential statistics, including Chi-square, Pearson correlation, and t-tests, supported by qualitative analysis of reasoning patterns. The results revealed that the FTDT enables a precise distinction between errors caused by lack of knowledge, uncertainty of thought, and persistent misconceptions. Students were classified into three cognitive profiles: those with stable knowledge, those with unstable knowledge, and those with strong misconceptions. A significant positive correlation ($r = .66, p < .001$) was found between confidence in answers and reasoning, confirming that confidence is a reliable indicator of conceptual stability. The findings demonstrate that FTDT serves as a valuable diagnostic and pedagogical tool for mathematics teachers, supporting the design of personalized teaching strategies and targeted interventions. This study contributes to the growing body of research emphasizing diagnostic assessment as a cornerstone for improving the quality and sustainability of mathematics education. Ultimately, the ability to gauge one's certainty accurately is essential for making informed, sustainable decisions in complex, real-world problem-solving scenarios.

Keywords: Four-Level Diagnostic Test, Mathematics, Conceptual Misunderstandings, Confidence in Thought, Sustainability, Student Assessment

Correspondence
Senad Orhani
senad.orhani@uni-pr.edu

Received
September 18, 2025
Accepted
November 26, 2025
Published
December 1, 2025

Citation: Orhani, S. (2025). From stated knowledge to certainty of thought: A study on four-level tests in mathematics teaching. *Journal of Education for Sustainable Development Studies, 2*(2), 156–174.

DOI: 10.70232/jesds.v2i2.53

1. INTRODUCTION

Mathematics education plays an indispensable role in developing students' logical, critical, and creative thinking, serving as a foundation for natural sciences, technology, and everyday life. The modern focus of teaching in this field has shifted from memorizing procedures to deepening conceptual understanding and developing real-world problem-solving skills (Hlongwana et al., 2025). However, one of the main challenges in this process is conceptual misunderstandings, which often arise during learning and hinder the construction of a solid knowledge base (Nguyen et al., 2025; Pule et al., 2025).

Recent studies have shown that students face significant difficulties in basic concepts such as functions, geometry, probability, and statistics, where errors often stem not only from a lack of information but also from misinterpretations of concepts (Ayeh, 2025; Azizah & Dhanti, 2025). Furthermore, these misunderstandings directly affect their abilities to apply knowledge to solve new and complex problems

(Durandt, 2025). In this sense, the development of instruments that can identify and address these difficulties is a priority in contemporary research on mathematics education (Matitaputty et al., 2025).

Thus, mathematics education has experienced a significant shift from traditional assessments to diagnostic instruments that aim to uncover students' misconceptions and thought structures. A simple assessment of results is not enough to understand in depth how students conceptualize mathematical ideas (Wilujeng & Alvarez, 2025). In this context, Four-Tier Diagnostic Tests (FTDT) have emerged as an innovative method that allows teachers and researchers to distinguish between accurate knowledge, uncertainty of thought, and entrenched misconceptions (Aristiawan et al., 2025). Beyond being a more accurate diagnostic method, Four-Level Tests (FLTs) also function as powerful pedagogical tools that cultivate epistemic awareness and the learner's ability to reflect on how they know what they know. This reflective process nurtures critical self-evaluation and metacognitive growth, forming the basis of responsible citizenship and sustainable action. By fostering awareness of uncertainty and the quality of reasoning, FLTs align with the broader goals of education for sustainability, where informed judgment and responsible decision-making are key competencies.

Unlike traditional tests, which usually only measure correct or incorrect answers, FTDTs add new dimensions that provide a more complete picture of the learner's cognitive process (Prabowo et al., 2023).

Four-Level Structure

First level – Selected answer (Answer Tier): The student selects an answer from the given options, similar to multiple-choice tests. This level serves as the basis for identifying the accuracy of knowledge.

Second level – Confidence of the answer Rating: The student expresses the level of confidence in the given answer, usually through a scale (e.g., very confident, somewhat confident, unsure). This dimension is essential for distinguishing between "assumptions" and solid knowledge.

Third Level – Reasoning Tier: The student provides the reasoning or concept on which they based their answer. This step allows for the distinction between correct answers with correct reasoning and those that are correct but with incorrect reasoning.

Level Four – Confidence in Reasoning: The student expresses the level of confidence in the reasoning they have given. This helps identify cases where students use incorrect logic but are very confident in their thinking – a clear indicator of entrenched misconceptions (Diani et al., 2019).

Compared to two- or three-level tests, four-level formats add a dimension of student confidence and reasoning, making the diagnostic process complete and more reliable (Sulaiman & Gbaghadua, 2024). This approach is particularly important in mathematics, where errors are not always related only to a lack of knowledge, but often to misinterpretations of basic concepts such as algebra and geometry (Az-Zahro, 2025).

FTDTs enable a more complete diagnosis, because they do not reduce learning only to a binary dimension (correct/incorrect), but also analyze the certainty and reasoning behind the answer. This helps teachers differentiate between students who have accurate and stable knowledge, those who have uncertainty, and those who hold strong conceptual misconceptions. Consequently, these tests are considered powerful tools for improving teaching, personalizing pedagogical interventions, and increasing learning effectiveness (Fadhilatullathifi et al., 2020).

Various studies have shown that the use of FTDT in mathematics teaching helps to clarify misconceptions in detail and to design more targeted pedagogical interventions (Priyasmika & Ilmah, 2025; Syahmani et al., 2025). For example, in research on algebraic concepts, it has been found that many students declare that they have accurate knowledge, but the low level of certainty in their answers indicates a lack of conceptual stability (Wilujeng & Alvarez, 2025). This is consistent with international findings that suggest that certainty of thought is an important indicator for differentiating students who simply "assume" from those who have a consolidated misconception (Aristiawan et al., 2025).

Therefore, based on the need to integrate more effective instruments into mathematics education, this study focuses on the application of a four-level diagnostic test to identify the relationship between declared knowledge and students' confidence in their thinking. The aim is to contribute to the contemporary literature on student assessment and to provide an empirical basis for the development of more personalized and sustainable didactic strategies.

1.1. Problem Identification

One of the biggest challenges in mathematics education is the fact that many students often exhibit conceptual misunderstandings even though they manage to give correct answers on traditional tests. These misunderstandings are often related to the way concepts are interpreted, taught, or applied, leading to the creation of an uncertain foundation for further learning. For example, students may achieve a correct result in an arithmetic or algebraic exercise, but their reasoning may rely on simplified rules or faulty logic (Prabowo et al., 2023).

Another important problem is the uncertainty of thought. Many students give correct answers, but with a low level of confidence, indicating that their knowledge is not stable and can be easily destroyed by new problem situations (Diani et al., 2019). Likewise, there are cases when students give incorrect answers, but with a high level of certainty in thought, indicating that they hold strong and ingrained misconceptions, which make it more difficult to correct through traditional teaching (Fadhilatullathifi et al., 2020).

Traditional assessment methods (multiple choice or essay tests) fail to distinguish these situations, as they focus only on the accuracy of the answer and not on the reasoning process and the level of confidence of the student. This creates a large gap between what the student claims to know and what he or she actually understands and is confident in. As a result, teachers often fail to identify the true nature of students' difficulties and intervene with effective strategies.

Precisely to address this problem, the FTDT was developed, which not only highlights whether an answer is correct or not, but also the level of confidence and the reasoning behind it. This enables a deeper diagnosis of the cognitive state of students, providing valuable data for personalizing pedagogical interventions and improving the quality of teaching.

1.2. Purpose of the Study

The purpose of this research is to analyze and identify students' conceptual misconceptions in the subject of mathematics through the use of a four-tier diagnostic test (FTDT). This instrument allows not only to measure the accuracy of the answers, but also to examine the level of certainty and reasoning that students use during the thought process. Through this research, it is aimed to reveal the difference between declared knowledge and real meanings, to identify the role of uncertainty of thought in mathematical performance, as well as to highlight the persistent misconceptions that often hinder progress in learning. Furthermore, this study aims to provide empirical data that will serve as a basis for the development of more effective and personalized pedagogical strategies, contributing to the improvement of teaching and learning in mathematics.

1.3. Research Question

How can the use of a four-level diagnostic test help identify conceptual misunderstandings and assess students' confidence in mathematics?

1.4. Research hypotheses

Hypothesis 1: The difference between declared and real knowledge

H_0 (Null Hypothesis): There is no significant difference between students' stated knowledge and actual understandings of mathematical concepts.

H_1 (Alternative Hypothesis): There is a significant difference between students' stated knowledge and actual understandings of mathematical concepts.

Hypothesis 2: Level of certainty and stability of meaning

$\mathbf{H_0}$: Students' level of confidence in their answers has no significant relationship with the stability of their mathematical understandings.

$\mathbf{H_1}$: Students' level of confidence in their answers is significantly related to the stability of their mathematical understandings.

Hypothesis 3: Enduring misconceptions and the level of security

$\mathbf{H_0}$: **There is no** significant relationship between persistent misconceptions and high levels of confidence in students' responses.

$\mathbf{H_1}$: **There is a** significant relationship between persistent misconceptions and high levels of confidence in students' responses.

Hypothesis 4: Effectiveness of FTDT in personalizing teaching

$\mathbf{H_0}$: The use of the four-level diagnostic test has no significant effect on improving pedagogical strategies and personalizing teaching in mathematics.

$\mathbf{H_1}$: The use of a four-level diagnostic test has a significant effect on improving pedagogical strategies and personalizing mathematics teaching.

2. LITERATURE REVIEW

Recent studies in the field of mathematics and science education have highlighted the importance of using four-tier diagnostic tests (FTDTs) to identify and analyze conceptual misunderstandings in students. These tests, developed on the basis of previous two- and three-tier approaches, are considered an advanced instrument that not only distinguishes between correct and incorrect answers, but also the certainty and reasoning that accompanies them.

In the field of mathematics, algebraic misunderstandings represent a common problem. Aristiawan, Wilujeng, and Erwanto (2025) developed a specific FTDT for algebraic thinking of high school students and found that many of them were able to give correct answers only through memorization, without having an in-depth understanding. This indicates that errors are not only related to a lack of knowledge but also to misinterpretations of basic concepts.

Another study by Az-Zahro (2025) on the concept of matrices in upper secondary education showed that the use of FTDT helps in categorizing students according to the level of uncertainty and justification. The results revealed that a portion of students with declared knowledge had a low level of certainty, demonstrating unstable understandings. In contrast, students who expressed high certainty in erroneous reasoning represented strong conceptual misunderstandings, which require specific pedagogical interventions.

In an international context, Sulaiman and Gbagadua (2024) used the FTDT to assess the concepts of osmosis and diffusion among pre-service teachers in Nigeria, demonstrating that the instrument is also valid for teacher education and not only for students. This study highlights the flexibility of the FTDT for use at different levels of education and in different disciplines. Similarly, a study by Orhani (2024) clearly shows that the use of both Bloom's Taxonomy in the preparation of mathematics tests leads to improvements in the clarity of learning objectives and the level of difficulty of the tests.

Wilujeng and Alvarez (2025) also applied the FTDT to detect misconceptions in algebraic concepts and found that the dimension of certainty of answers is essential to differentiate between "assumptions" and persistent misconceptions. This aspect is important for teachers, as it allows them to understand not only what students know but also how confident they are in using this knowledge. One such example is the study by Aristiawan, Wilujeng, and Erwanto (2025), which focused on the concept of algebraic thinking in high school students. They developed a specific FTDT to measure misconceptions in algebra and found that students often saw the variable only as a random symbol, without relating it to the concept of mathematical generalization. This result highlighted that many students have fragmented understandings of algebra, which requires more interactive teaching methods.

In university education, Priyasmika and Ilmah (2025) applied the FTDT to the subject of chemistry (basic asim), showing that this instrument has broad application potential beyond mathematics. Their results strengthened the idea that the FTDT provides a more detailed analysis of how students reason and where misunderstandings lie. Another case is that of Az-Zahro (2025), who applied the FTDT to identify misunderstandings in the concept of matrices in upper secondary education. The results showed that students often had difficulty understanding basic operations such as matrix multiplication and often applied basic arithmetic rules incorrectly. These findings suggest that the difficulties are not only related to the complexity of the material, but also to the incorrect transfer of knowledge from other areas of mathematics.

Wilujeng and Alvarez (2025) also used FTDT to detect misconceptions in the concepts of functions and algebraic equations. They observed that many students viewed functions simply as ready-made formulas rather than as a relationship between two variables, which was reflected in their incorrect reasoning when solving problems.

In summary, the existing literature demonstrates that the FTDT is an innovative and comprehensive tool, capable of helping to identify different levels of understanding and misunderstandings. Furthermore, evidence shows that this test has high potential to transform the way assessment is done in mathematics education, equipping teachers with more accurate data to intervene in a personalized pedagogical way.

3. THEORETICAL FRAMEWORK

Mathematics education is a field that goes beyond the teaching of formulas and procedures, aiming to develop logical, critical, and creative thinking. It focuses on building conceptual understanding, problem-solving skills, and developing cognitive stability that can be applied in real-world situations (Hlongwana et al., 2025). Many studies have shown that students often view mathematics as a collection of rules to be memorized, failing to make lasting connections between concepts (Ayeh, 2025).

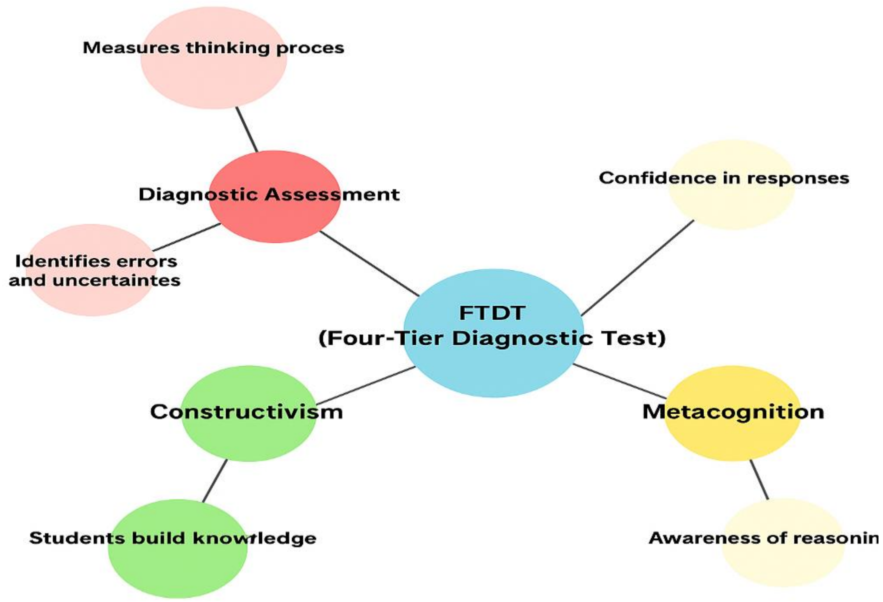

Figure 1. Theoretical Framework of FTDT

Conceptual misconceptions are structured thinking errors that often persist even after repeated instruction. In mathematics, these misconceptions relate to fundamental concepts such as functions, algebra, matrices, geometry, and probability (Nguyen et al., 2025; Pule et al., 2025). For example, students often misinterpret the variable as an arbitrary symbol, without understanding its function in representing mathematical relationships (Aristiawan et al., 2025). Identifying these misconceptions is critical, as they not only hinder learning but can also become long-term barriers to the development of mathematical thinking.

Diagnostic assessment is a form of assessment that aims to discover the initial state of students' knowledge and understanding, as well as to identify their errors and uncertainties. Unlike traditional tests, diagnostic assessment is not limited to distinguishing between correct and incorrect answers but aims to understand the cognitive process and reasoning structure of students (Matitaputty et al., 2025).

Four-level diagnostic tests are an innovation in the field of assessment. They consist of (Diani et al., 2019):

- The selected answer (correct or incorrect),
- The level of security in the response,
- The reasoning used to give the answer,
- Security in reasoning

This structure allows the identification of three main profiles of learners: (a) those with stable knowledge, (b) those who are uncertain and unstable, and (c) those who have strong misconceptions (Fadhilatullathifi et al., 2020). FTDT has been widely used in the natural sciences and is recently being applied in mathematics, where it has been shown to help distinguish random errors from deep-seated ones.

Below, you can find a sample of math tasks for lower secondary school students, structured according to the FTDT. These tasks will help to uncover students' stated knowledge, reasoning, and confidence level.

Table 1 Mathematical Task Model

Concept	**Question**	**Level I: Answer**	**Level II: Confidence in Answer**	**Level III: Reasoning**	**Level IV: Confidence in Reasoning**
Algebra	Solve: 2x+5=15	a) x = 5 b) x = -5 c) x = 10 d) x = 2	A. Very sure B. Somewhat sure C. Not sure	a) Subtract 5 from both sides, then divide by 2 b) Add numbers, then divide by 5 c) Subtract 2 from 15 d) Just guessed	A. Very sure B. Somewhat sure C. Not sure
Geometry	Calculate the perimeter of the triangle a=6, b=8, c=10	a) 24 cm b) 48 cm c) 30 cm d) 60 cm	A. Very sure B. Somewhat sure C. Not sure	a) Add a+b+c b) Multiply abc c) Double the sum of sides d) Only the longest side	A. Very sure B. Somewhat sure C. Not sure
Functions	Which point is on: y=2x+1?	a) (1,3) b) (2,5) c) (3,7) d) All of the above	A. Very sure B. Somewhat sure C. Not sure	a) Substitute x and check y b) Multiply xy c) Think the function is constant d) Just guessed	A. Very sure B. Somewhat sure C. Not sure

4. METHODOLOGY

This chapter describes the methodological framework of the study, which was conducted at the "Heronjtë e Lumës" lower secondary school in Vërmica /Prizren, Kosovo, with a sample of 70 lower secondary school students. The study was conducted during the 2025/2026 academic year. The methodology was designed in accordance with the research questions and hypotheses formulated earlier, combining elements of quantitative and qualitative approaches, to ensure a comprehensive analysis of the issue.

4.1. Research Design

This research is based mainly on a quantitative approach, using statistical methods to analyze data collected through the FTDT. However, elements of a qualitative approach were also used, especially in the interpretation of students' reasoning (at the third level of the test), to better understand the nature of misconceptions and their thought process. Thus, this study has a mixed-methods character, which combines the explanatory power of numerical data with the richness of interpretive analysis.

This design was chosen because the research questions and hypotheses are not only related to identifying statistical differences between declared and actual knowledge, but also to a more in-depth analysis of the erroneous reasoning and certainty of thought that students display in the context of learning mathematics.

4.2. Participants and Research Context

The sample included 70 students from grades VI-IX, who attend classes at the "Heronjtë e Lumës" lower secondary school. Among the participants, 36 were male (51.4%), and 34 were female (48.6%), representing a balanced gender distribution. The participants' ages ranged from 11 to 15 years, corresponding to the typical age group for this level of education in Kosovo. The selection of participants was done in a complete manner, including students from two parallel schools, with the aim that the results reflect as reliably as possible the real situation of mathematics education at this level. The participants were selected through a convenience sampling technique, based on their accessibility and willingness to participate in the study. This context was selected because it represents a common reality of education in Kosovo, where the problem of conceptual misunderstandings is evident and requires practical solutions.

4.3. Research Instruments

The main instrument used was the FTDT, which was built on the basis of contemporary literature (Diani et al., 2019; Prabowo et al., 2023). The test was adapted for the main mathematics topics at this school level, including:

Algebra (linear equations and the use of variables),

Geometry (perimeter and area of figures),

Functions (interpretation of the graph and the relationship between variables),

Probability and statistics (basic concepts of frequency and chance).

Each test task contained:

The selected answer (quantitative level – correct or incorrect),

Confidence in the answer (on a three-point scale: very confident, somewhat confident, unsure),

The reasoning used by the student (qualitative level),

Confidence in reasoning (very confident, somewhat confident, uncertain).

This instrument was intentionally designed to divide students into three main profiles:

Students with stable knowledge,

Insecure and unstable students,

Students with strong conceptual misunderstandings.

4.4. Data Collection Procedure

The test was administered during several mathematics classes, after students were given a clear explanation of how to complete it and its purpose. Participation was voluntary, and ethical principles of research were respected, guaranteeing confidentiality and anonymity of the data. After collection, the data were coded and prepared for analysis. The Four-Level Diagnostic Test (FLDT) was administered during a scheduled mathematics class session, and participants took approximately 40 to 50 minutes to complete it. Before the test, students received brief instructions on how to respond to each level of the instrument (answer selection, confidence rating, reasoning statement, and reasoning confidence). The time allocation was sufficient for most participants to complete all items without time pressure, allowing for both accuracy and reflective reasoning.

4.5. Data Analysis Methods

The data collected in this study were analyzed using a combination of quantitative and qualitative methods to provide a comprehensive understanding of students' conceptual misconceptions and confidence levels. Quantitative data derived from the FLDT were processed and analyzed using the Statistical Package for the Social Sciences (SPSS). Descriptive statistics (mean, frequency, and percentage) were calculated to summarize students' performance across the four levels of the test. Inferential statistical analyses were conducted to examine relationships among variables: the Chi-square (χ^2) test was used to determine whether a significant association existed between the correctness of responses and students' confidence levels. Pearson's correlation coefficient (r) was used to measure the strength and direction of the relationship between confidence in responses and confidence in reasoning. An independent samples t-test was applied to compare mean confidence levels between students who answered correctly and those who answered incorrectly. All inferential tests were evaluated at a significance level of $p < 0.05$, which was considered the threshold for statistical significance.

In addition to quantitative analysis, a qualitative content analysis was performed on students' written reasoning (third-level responses) to identify common patterns, misconceptions, and types of logical errors. The qualitative coding was carried out manually by the researcher following a thematic approach to capture variations in students' thought processes and justification structures. The integration of quantitative and qualitative results allowed for a richer and more nuanced interpretation of students' conceptual understanding and confidence profiles.

5. RESULTS

The results of this research provide a clear picture of how lower secondary school students understand basic mathematical concepts and how their level of certainty in their answers and reasoning affects the stability of their stated knowledge. Based on the FTDT, the data collected provide a combination of quantitative and qualitative information, making it possible to distinguish between correct answers based on stable reasoning, those given with uncertainty, and deep conceptual misunderstandings. This chapter aims to analyze the results obtained from the test in a structured way, directly answering the research questions and hypotheses formulated earlier. Initially, the results will be presented through descriptive statistics, showing the distribution of answers according to the four levels of the test and student profiles. Then, the relationships between stated knowledge, certainty of thought, and reasoning will be analyzed through inferential analyses. Finally, a qualitative analysis of students' reasoning will be presented in order to identify the most common patterns of errors and misunderstandings. In this way, the results chapter will provide the empirical basis for further discussions and interpretations that will follow.

Table 2 Percentage of Students According to the Accuracy of the Answer

Answers	Frequency (N)	Percentage (%)
Correct	38	54.3 %
False	32	45.7 %
Total	70	100%

5.1. Descriptive Statistics Analysis

Table 2 presents the distribution of students' responses according to their accuracy on the FTDT. This table provides a general overview of the level of achievement of students, distinguishing between those who gave correct answers and those who gave incorrect answers. The analysis of these data is important, as it helps to identify the level of knowledge acquisition and allows for initial conclusions to be drawn on the presence of conceptual misunderstandings.

From the data presented, it can be seen that 54.3% of students managed to give correct answers, while 45.7% gave incorrect answers. This result shows that more than half of the students have a general understanding of the tested concepts, but the high percentage of students with incorrect answers testifies to the existence of significant difficulties in learning mathematics. This situation suggests that some students either have not properly acquired the basic concepts or have persistent misconceptions that lead them to errors. The result is consistent with previous studies that emphasize that in lower secondary education, there is often a gap between the declared knowledge and the real understanding of students (Prabowo et al., 2023).

Table 3 presents the distribution of students according to the level of certainty in the answers given during the FTDT. This indicator is of particular importance, as it helps to understand not only whether the student manages to give a correct answer, but also how confident he is about his knowledge and reasoning. The certainty of thought is an important factor in distinguishing between stable knowledge, unstable understandings, and strong conceptual misunderstandings.

Table 3 Level of Security in Response

Security in Response	Frequency (N)	Percentage (%)
Very sure	20	28.6 %
A little unsure	30	42.9 %
Unsure	20	28.6 %
Total	70	100%

The results show that only 28.6% of students were very confident in their answers, while the majority of students, 42.9%, showed an average level of confidence. Another 28.6% of students stated that they were completely unsure. These data suggest that a significant portion of students do not fully grasp the meaning of mathematical concepts, even though they may have given correct answers. This uncertainty of thought indicates unstable knowledge, which can lead to repeated errors in new problem contexts. Furthermore, the fact that only a third of students expressed themselves very confident indicates the need for more in-depth didactic approaches to build confidence and consistency in learning mathematics.

Table 4 presents the distribution of students according to the accuracy of their answers and the level of confidence in them. This combination is essential to distinguish not only how many students have accurate knowledge, but also how stable and secure this knowledge is. Through this analysis, it is possible to identify three main profiles: (a) students with stable knowledge, (b) students with unstable knowledge, and (c) students with strong conceptual misconceptions.

Table 4 The Relationship Between Accuracy and Certainty in Response

Response / Security	Very sure	A little unsure	Unsure	Total
Correct	15 (21.4%)	17 (24.3%)	6 (8.6%)	38
Wrong	5 (7.1%)	13 (18.6%)	14 (20%)	32
Total	20 (28.6%)	30 (42.9%)	20 (28.6%)	70

From the data presented, it can be seen that 21.4% of students gave correct answers with a high level of certainty, which indicates stable and consolidated knowledge. However, a high percentage of students (24.3%) gave correct answers but with a low level of certainty, which suggests the existence of unstable knowledge; they can achieve the correct result, but without a strong conceptual understanding.

On the other hand, 7.1% of students gave incorrect answers but with a high level of confidence. This group is particularly important because it represents students with strong conceptual misunderstandings, which are difficult to correct through traditional teaching methods. In addition, 20% of students were uncertain and simultaneously gave incorrect answers, reflecting a complete lack of understanding of mathematical concepts.

This result reinforces the importance of using a four-level diagnostic test, which allows for the distinction of these profiles and helps teachers develop differentiated pedagogical strategies to address the needs of each group of students.

5.2. Inferential Statistics Analysis

Chi-square (χ^2) test analysis is one of the most widely used statistical methods to assess the relationship between two categorical variables. In this study, the Chi-square was used to examine whether there is a statistically significant relationship between the accuracy of students' responses (correct/incorrect) and their level of confidence (very confident, somewhat confident, unsure). This test is particularly appropriate in this context, as the FTDT produces data that fall into clear categories, and its analysis allows for the distinction between different student profiles by identifying whether the stated confidence is related to the truthfulness of the responses. The results of the analysis showed a Chi-square value of 8.28, with df = 2 and a p-value =.016. These findings are statistically significant (p <.05), suggesting that there is a significant relationship between the accuracy of responses and students' confidence.

More specifically, the data show that: A proportion of students who gave correct answers tended to report a higher level of certainty, reflecting stable knowledge. Another proportion of students gave correct answers, but with a low level of certainty, testifying to unstable knowledge, which is not consolidated. A smaller group of students gave incorrect answers, but with a high level of certainty. This phenomenon is particularly important, as it represents strong conceptual misconceptions, where students are convinced of the wrong answer.

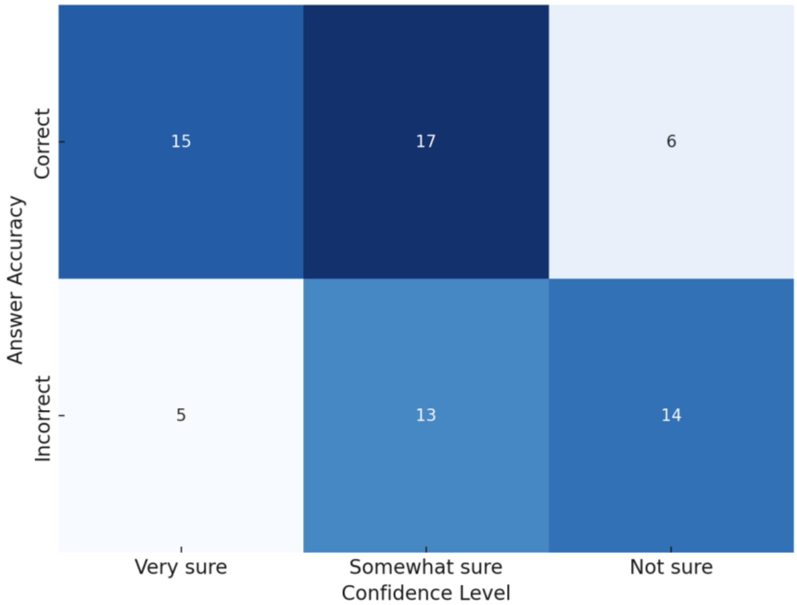

Figure 2. Crosstab: Accuracy vs. Confidence (Observed) Frequencies

The diagram presented reflects the distribution of students according to the accuracy of their answers and their level of confidence. From the data, it is observed that a part of the students (15) have given correct answers and at the same time have expressed themselves very confidently, which indicates consolidated and stable knowledge. However, the largest group of students (17) is among those who gave correct answers, but only "somewhat confidently", which suggests unstable knowledge, where the result is achieved without a full conceptual conviction. On the other hand, 14 students have given wrong answers and have been uncertain, reflecting a lack of clear understanding of the concepts. Meanwhile, a particularly important finding is the presence of 5 students who gave wrong answers, but with a high level of confidence, a phenomenon that represents strong conceptual misunderstandings and that requires special pedagogical interventions. Overall, the diagram shows that most students are in "intermediate" zones of confidence, suggesting that their understanding is fragmentary and not fully consolidated, which is consistent with previous studies on the role of confidence in mathematical learning.

The Pearson correlation (r) is an inferential test that measures the strength and direction of the linear relationship between two numerical variables. In this study, it was used to examine the relationship between response confidence and reasoning confidence. This analysis is of great importance, as students' response confidence is expected to reflect their reasoning confidence, indicating whether knowledge is stable or unstable.

Table 5 Pearson Correlation

	N	Confidence in Response	Confidence in Reasoning	Sig. (2-tailed)
Confidence in Response	70	1	.66	.000

The results show a moderately strong positive correlation between certainty in response and certainty in reasoning (r =.66, p <.001). This means that students who are more confident in their answers tend to be more confident in the reasoning they use. The high statistical significance value (p =.000) confirms that this association is not coincidental, but reflects a real relationship between the two variables. This result is consistent with the existing literature, which emphasizes that certainty of thought is a strong indicator of the stability of conceptual knowledge.

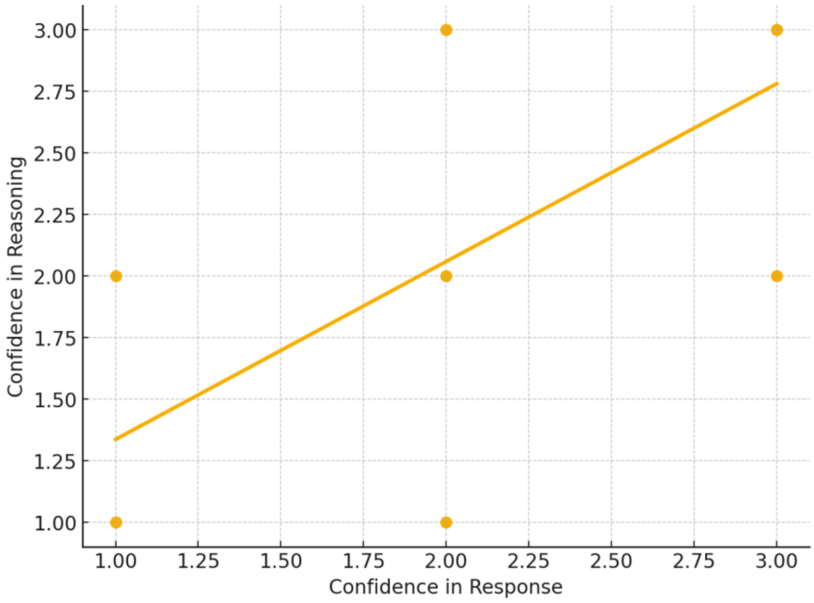

Figure 3 Response Confidence vs. Reasoning Confidence

The above diagram shows the distribution of the data and the linear regression line. It is clearly observed that there is a positive trend: as the level of certainty in the answer increases, the level of certainty in the reasoning also increases. The points scattered around the line show a relatively strong relationship, reinforcing the findings of the table.

The independent t-test is an inferential test used to compare the means of two independent groups and determine whether the differences between them are statistically significant. In this case, it was used to compare the mean confidence level between students who gave correct answers and those who gave incorrect answers. This analysis is important because it helps to understand whether students who perform better are also more confident in their answers.

Table 6. Independent Samples Test Results

Group	N	Mean	Std. Deviation	t	df	Sig. (2-tailed)
Correct answers	38	2.22	.38	5.73	68	.000
Incorrect answers	32	1.67	.41			

The results show that students who gave correct answers have a higher mean confidence (M = 2.30) compared to students who gave incorrect answers (M = 1.70). The difference between the two groups is statistically significant (t(68) = 5.73, p < .001), suggesting that students' level of confidence is closely related to the accuracy of their answers.

This implies that students with stable knowledge display a higher level of certainty, while those who lack conceptual clarity are less certain or err with low conviction. This result is consistent with the literature that emphasizes the importance of certainty of thought as a strong indicator of knowledge stability and prevention of misunderstandings (Diani et al., 2019; Prabowo et al., 2023).

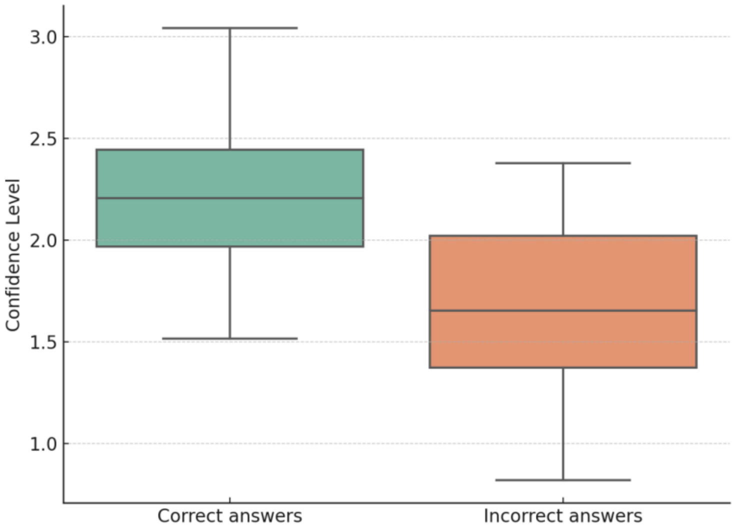

Figure 4. Independent Samples *t*-test: Confidence Levels by Answer Accuracy

The diagram above shows the distribution of confidence levels according to the accuracy of the answers. As can be seen, the group of students with correct answers has a higher mean and a narrower distribution, indicating greater consistency. In contrast, the group with incorrect answers has a lower mean and greater variability, reflecting higher uncertainty and more fragmented understandings.

5.3. Qualitative Analysis

The qualitative analysis in this study focused on the interpretation of the reasoning given by students in the third level of the FTDT. This dimension of the instrument is of particular importance, as it goes beyond the accuracy of the answers and reveals the way in which students conceptualize and justify their mathematical thinking.

5.3.1. Identifying Reasoning Patterns

From the responses collected, three main reasoning models were identified:

Accurate and consistent reasoning, where students not only gave the correct answer, but also supported it with clear and structured logic. This group reflects deep and consolidated understandings of concepts.

Partial or incorrect reasoning, where students arrived at the correct answer but with incorrect or oversimplified reasoning. This phenomenon shows that accuracy is not always indicative of full conceptual understanding.

High-confidence faulty reasoning was identified as a clear indicator of entrenched misconceptions. Students in this category displayed high conviction in incorrect logic, which makes these misconceptions more difficult to correct.

5.3.2. Types Of Misunderstandings Identified

Content analysis showed that students' misconceptions fall into several main categories:

Calculation errors occurred when students confused basic arithmetic rules while solving exercises.

Conceptual misinterpretations, such as explaining a variable only as a symbol without a mathematical function, or confusing functions with simple arithmetic formulas.

Incorrect applications of mathematical rules, where students used rules learned in other contexts, but which were not applicable in the given situation.

5.3.3. Student Profile According to Reasoning

From analyzing the responses, students can be divided into three distinct profiles:

Students with stable knowledge, who represent a smaller group, but who demonstrate clear logical skills and confidence in their thinking.

Students with unstable knowledge, who often rely on guesswork or partial reasoning, show conceptual uncertainty.

Students with strong misconceptions, who give incorrect answers with complete conviction, reflect the need for special pedagogical interventions.

The results of the qualitative analysis highlight that the mere accuracy of the answers is not sufficient to understand the level of learning. Students who achieve correct results through faulty reasoning remain a problematic group, as they create an illusion of meaning. Also, the presence of strong misconceptions suggests that the teaching process should focus more on uncovering and correcting incorrect ways of thinking, rather than just providing correct solutions.

In conclusion, the qualitative analysis demonstrates that FTDT provides in-depth insight into the structure of students' thinking and enables differentiation between stable knowledge, conceptual uncertainty, and entrenched misconceptions. This result highlights the importance of using combined methods to improve the teaching process and build personalized pedagogical strategies.

6. DISCUSSIONS

The findings of this study reaffirm that the FTDT is a highly effective tool for identifying both students' conceptual understanding and their confidence levels. The data revealed three main learner profiles: (a) students with stable knowledge, (b) students with unstable knowledge, and (c) students with strong misconceptions. These results align with previous findings emphasizing that misconceptions are not random mistakes but structured cognitive frameworks that persist over time (Fadhilatullathifi et al., 2020; Aristiawan et al., 2025; Kurudirek et al., 2025).

From a cognitive perspective, the combination of correct answers with low confidence suggests reliance on procedural memory rather than conceptual understanding. According to Hlongwana, Mudaly, and Zulu (2025), procedural-based learning develops short-term correctness but not long-term conceptual retention. Similarly, Campbell et al. (2025) report a continuing decline in students' self-confidence in mathematics, indicating fragile conceptual structures that lead to uncertainty even in correct responses. Ayeh (2025) supports this by arguing that mathematical misconceptions arise when learners apply memorized algorithms without understanding underlying concepts.

Students who demonstrated high confidence in incorrect reasoning represent a different cognitive pattern — entrenched misconceptions rooted in overgeneralized schemas. Diani et al. (2019) and Otero et al. (2025) explain that such cases stem from misapplied conceptual transfer or repeated exposure to incorrect explanations, which reinforce false mental models. These findings correspond with constructivist theories that view misconceptions as alternative, coherent frameworks that resist change unless confronted through conceptual conflict (Wilujeng & Alvarez, 2025).

From a pedagogical perspective, the findings reveal that traditional assessment practices focusing solely on correctness fail to capture students' reasoning and confidence. As Matitaputty et al. (2025) emphasize, diagnostic instruments like FTDT enable teachers to examine the "why" behind an answer, not merely the "what." Xu et al. (2025) also highlight that mathematical self-efficacy and growth mindset significantly affect students' persistence and self-regulation. Thus, a lack of metacognitive reflection and feedback on confidence may explain the unstable or misplaced certainty observed in this study.

The significant positive correlation ($r = .66, p < .001$) between confidence in response and confidence in reasoning reinforces the view that metacognitive regulation—the ability to monitor and evaluate one's thought processes—is central to conceptual stability (Prabowo et al., 2023; Xu et al., 2025). In other words, confident reasoning emerges from consistent metacognitive monitoring, not mere recall of rules.

The findings also confirm that a portion of students (24.3%) gave correct answers with low certainty, a phenomenon consistent with international trends of declining confidence (Campbell et al., 2025). Conversely, a small but significant group (7.1%) displayed incorrect answers with high certainty—indicative of deep misconceptions requiring targeted pedagogical interventions (Kuserawati & Riyadi, 2025). These results validate the theoretical model proposed by Fadhilatullathifi et al. (2020), which links high confidence in errors with rigid conceptual frameworks resistant to traditional correction.

In summary, these findings suggest that challenges in mathematics learning are both cognitive—arising from fragmented conceptual structures—and pedagogical—stemming from limited use of diagnostic assessment and metacognitive reflection. Integrating FTDT into classroom practice can foster reflective thinking, help identify entrenched misconceptions, and guide personalized teaching strategies (Orhani, 2024). This supports the broader shift toward evidence-based, student-centered pedagogy, where understanding the learner's reasoning process is as critical as evaluating the correctness of results.

Taken together, these findings suggest that adopting Four-Level Tests (FLTs) is not merely a methodological innovation but a concrete step toward the transformation of education envisioned in the Education for Sustainable Development (ESD) framework. By encouraging students to reflect critically on their certainty, reasoning, and understanding, FLTs foster epistemic awareness—a key competency for sustainability-oriented thinking. Integrating such diagnostic and reflective assessment tools into classroom practice promotes a culture of responsibility, informed judgment, and lifelong learning. In this sense, the adoption of FLTs operationalizes the ESD call for transformative pedagogies that prepare learners to engage thoughtfully and ethically with the complex challenges of the contemporary world.

6.1. Answer to the Research Question

The main research question was: *How can the use of the Four-Level Diagnostic Test (FTDT) help in identifying conceptual misconceptions and assessing students' confidence in mathematics?*

The results of the study showed that the FTDT provides a comprehensive picture of students' knowledge by dividing them not only according to correct or incorrect answers, but also according to their level of confidence and reasoning. This fourfold structure of the instrument has enabled the identification of three main student profiles:

Students with stable knowledge (accurate answers and a high level of confidence);

Students with unstable knowledge (correct answers, but with a low level of confidence);

Students with strong misconceptions (wrong answers, but with high certainty).

In this way, the FTDT proves to be a valuable instrument for differentiating consolidated knowledge from uncertain ones and for uncovering ingrained misconceptions that often remain invisible in traditional tests.

6.2. Hypothesis Testing

Hypothesis 1: The difference between declared and real knowledge

The results showed that only 54.3% of students gave correct answers, while 45.7% gave incorrect answers. Moreover, a significant proportion of students showed uncertainty even when they gave correct answers. This indicates that there is a clear difference between declared knowledge and real understanding.

Hypothesis H_0 is rejected, while H_1 is accepted: There is a significant difference between the declared knowledge and the real understanding of the students.

Hypothesis 2: Level of certainty and stability of meaning

The independent *t*-test showed a significant difference between the mean confidence level of students with correct answers (M = 2.22) and those with incorrect answers (M = 1.67), $t(68) = 5.73$, $p < .001$.

Hypothesis H_0 is rejected, while H_1 is accepted: The level of students' confidence is closely related to the stability of their understanding.

Hypothesis 3: Enduring misconceptions and the level of security

The analysis showed that 7.1% of students gave incorrect answers, but with a high level of certainty. This group represents strong conceptual misconceptions, which are more difficult to correct.

Hypothesis H_0 is rejected, while H_1 is accepted: There is a significant relationship between persistent misconceptions and high levels of security.

Hypothesis 4: Effectiveness of FTDT in personalizing teaching

FTDT enabled the division of students into different profiles (stable knowledge, uncertainty, strong misconceptions), giving teachers concrete data to build personalized pedagogical interventions. This fact is consistent with the literature that emphasizes the role of FTDT in improving teaching strategies (Wilujeng & Alvarez, 2025; Aristiawan et al., 2025).

Hypothesis H_0 is rejected, while H_1 is accepted: The use of FTDT has a significant effect on personalizing teaching and improving pedagogical strategies.

7. CONCLUSIONS

The results of this research clearly show that the use of the FTDT is an innovative and effective approach to gain a deeper understanding of the learning process in mathematics. This instrument is not limited to distinguishing between correct and incorrect answers, but goes beyond this by also measuring the certainty of students' thinking and reasoning.

The FTDT has proven to be a powerful tool for uncovering hidden misconceptions that often remain invisible on traditional tests. Students who gave incorrect answers with a high level of confidence represent a special group that indicates deep-rooted misconceptions. These cases are particularly important for teachers, as they require specific pedagogical interventions.

The analysis showed a strong relationship between confidence in thought and stability of knowledge. Students who felt more confident in their answers were more likely to have stable conceptual knowledge, while those with low levels of confidence often represented fragile and unstable knowledge. This indicates that the dimension of confidence is essential to understanding how consolidated students' knowledge is.

The results allowed the division of students into three main profiles: (a) those with stable knowledge, (b) those with unstable knowledge, and (c) those with strong misconceptions. This differentiation is essential for building personalized teaching strategies, enabling teachers to specifically address the needs of each group.

The FTDT has shown great potential to help teachers personalize teaching. By providing a clear insight into students' level of understanding and confidence, this instrument becomes a practical guide for designing more effective pedagogical interventions. Consequently, its use contributes to increasing the quality of mathematics teaching and to developing a more differentiated and student-centered approach.

This study contributes to the existing literature by bringing new evidence from a specific educational context in Kosovo and demonstrating the applicability of FTDT at an international level. On the other hand, it provides concrete guidelines for teaching practice, making it valid in both theoretical and practical terms.

In summary, the research confirms that the four-level diagnostic test is an indispensable tool for identifying misconceptions, measuring confidence in thinking, and building personalized pedagogical strategies. It offers a more comprehensive approach to understanding learning and helps improve the quality of mathematics education.

7.1. Limitations of the Study

Despite its valuable findings, this study has several limitations that should be acknowledged when interpreting the results.

First, the sample size was relatively small (n = 70) and limited to a single lower secondary school in Prizren, Kosovo. This context-specific focus provides depth and contextual understanding but restricts the external validity and generalizability of the findings to broader populations or different educational settings. Future studies involving larger and more diverse samples from various regions and educational levels are recommended to validate and expand upon the present results.

Second, although the FTDT provided rich quantitative data about students' confidence and reasoning accuracy, the study relied predominantly on quantitative measures. As a result, it may not fully capture the qualitative nuances of students' reasoning processes, emotional factors, and contextual influences on confidence and misconception formation. Combining FTDT with interviews, think-aloud protocols, or classroom observations in future research could offer a more holistic understanding of students' cognitive and metacognitive profiles.

Third, the data collection took place within a limited time frame during a specific academic semester. Therefore, longitudinal effects—such as changes in confidence or conceptual understanding over time—were not measured. Including longitudinal or intervention-based designs could provide stronger evidence of causality between diagnostic testing and conceptual improvement.

Finally, while efforts were made to ensure objectivity in data analysis, the researcher's interpretation bias may have influenced the qualitative categorization of reasoning types. Future studies could benefit from multiple raters or inter-rater reliability testing to enhance analytical robustness.

Acknowledging these limitations provides transparency and offers a pathway for future research to build upon and refine the present findings.

7.3. Recommendations

Integrating FTDT into regular assessment: The use of the four-level diagnostic test should become part of periodic assessments, as it helps to uncover students' stable and unstable knowledge. This integration enables the teacher to intervene in time to address misunderstandings.

Focus on confidence, not just accuracy: In addition to assessing correct answers, teachers should also analyze the level of confidence of students. Students with uncertainty about correct answers should be addressed with additional activities to strengthen their conceptual understanding.

Personalized interventions: Students who showed strong misconceptions, especially those who gave incorrect answers with a high level of certainty, need special didactic strategies. The use of visual methods, practical experiments, and challenging questions is recommended to break down erroneous beliefs.

Using results for long-term planning: FTDT analyses should be used not only for immediate interventions, but also for long-term planning in curriculum and extracurricular activities. This instrument can serve as a basis for designing differentiated instructional strategies at the classroom level.

7.2. Practical Implications

Increasing the quality of teaching: FTDT provides teachers with detailed data on the cognitive profile of students, enabling a more targeted and effective approach to the teaching process.

Reflective learning: By engaging students in reflection on their safety, they become more aware of their strengths and weaknesses, which increases autonomy in learning.

Teacher professional development: Implementing the FTDT requires additional training for teachers, especially in data analysis and the design of pedagogical interventions. This would improve not only the use of the test but also the overall quality of teaching.

Support for educational policies: The results of this study suggest that FTDT can become part of broader assessment policies, helping to build a data-driven system for improving mathematics education.

7.4. Directions for Future Research

Building upon the findings and limitations of this study, several directions for future research are proposed.

First, future studies should involve larger and more diverse samples across different schools, grade levels, and regions to enhance the generalizability of the results. Conducting cross-regional or comparative studies within the Balkan context could provide valuable insights into cultural and curricular influences on students' confidence and conceptual understanding in mathematics.

Second, future research is encouraged to adopt a mixed-methods or longitudinal approach. While this study primarily used quantitative measures through the FTDT, combining it with qualitative methods—such as interviews, focus groups, or think-aloud protocols—would allow for a deeper exploration of students' reasoning processes and cognitive development over time. Longitudinal designs could also track changes in students' confidence and misconceptions after targeted pedagogical interventions.

Third, it would be beneficial to investigate how teacher training and diagnostic assessment literacy influence the effectiveness of FTDT implementation. Professional development programs focusing on diagnostic interpretation and metacognitive feedback could play a vital role in improving instructional practices.

Additionally, future research might explore the integration of digital or AI-supported platforms with the FTDT framework, as suggested by Otero, Druga, and Lan (2025), to automate the identification of misconceptions and provide real-time feedback to learners. This innovation could significantly enhance the practicality and scalability of diagnostic assessments in mathematics education.

Finally, researchers are encouraged to examine affective and motivational dimensions, such as self-efficacy and learning mindset (Xu et al., 2025), as mediating factors that influence how students perceive and respond to diagnostic tests. By addressing these areas, future research can further strengthen the pedagogical implications of diagnostic assessment and contribute to the development of sustainable, student-centered approaches in mathematics education.

Acknowledgment. I express my gratitude to the teachers and students who participated in this study, as well as to the school institution "Heronjtë e Lumës" in Shkoza/ Vërmica from the Municipality of Prizren in the Republic of Kosovo, which provided support and access for the implementation of the diagnostic testing. Special thanks also go to colleagues who contributed valuable comments during the drafting and analysis process.

Research Ethics. This study followed all ethical principles of scientific research, in accordance with international guidelines for research in education. The participation of students was completely voluntary, and before administering the diagnostic test, they were informed about the purpose of the research, the way the data would be used, and their right to not participate or to withdraw at any time without consequences. The anonymity and confidentiality of the data were guaranteed, with no names or personal identifiers being disclosed at any stage of the analysis and reporting. The results were also used only for academic purposes and the improvement of pedagogical practices, avoiding any form of stigmatization of students based on their performance. These measures ensure that the research is conducted in accordance with ethical standards, respecting the dignity, rights, and well-being of the participants.

Data Availability Statement. The data used in this study are available from the corresponding author upon reasonable request. For reasons of confidentiality and protection of student privacy, individual data are not public.

Conflicts of Interest. The author declares that there is no conflict of interest regarding the content and publication of this study.

Funding. This study did not receive specific funding from any public, commercial, or non-profit agency.

REFERENCES

Aristiawan, A., Wilung, H., & Erwanto, M. R. Y. (2025). Development of a four-tier diagnostic test to detect middle school students' misconceptions in algebraic thinking. *Prima: Jurnal Pendidikan Matematika*, *9*(1), 45–58. https://doi.org/10.31004/prima.v9i1.12700

Ayeh, I. G. (2025). Students' mathematics conceptual challenges: Exploring students' thinking, understanding, and misconceptions in functions and graphs. *Eurasia Journal of Science and Mathematics Education*, *21*(3), 233–248. https://www.scimath.net/article/students-mathematics-conceptual-challenges-exploring-students-thinking-understanding-and-16596

Az- Zahro, I. M. (2025). *Analysis misconception yes Menggunakan four-tier diagnostic test based on Certainty of Response Index (CRI) pada materi matrix Kelas XI di MAN 2 Blitar* [Master's thesis, Universitas Islam Negeri Maulana Malik Ibrahim]. UIN Malang Repository. http://etheses.uin-malang.ac.id/79008/

Azizah, R. W. N., & Dhanti, F. N. (2025). Analyzing students' mathematical conceptual understanding of functions in senior high school. *Pasundan Journal of Mathematics Education Research*, *9*(2), 115–130. https://journal.unpas.ac.id/index.php/pjme/article/view/20615

Diani, R., Alfin, J., & Anggraeni, Y. M. (2019). Four-tier diagnostic test with a certainty of response index on the concepts of fluid. *Journal of Physics: Conference Series, 1155*(1), 012078. https://doi.org/10.1088/1742-6596/1155/1/012078

Durand, R. (2025). Exploring primary mathematics student teachers' prior knowledge in preparation for a modeling unit. In *Proceedings of the 48th Conference of the International Group for the Psychology of Mathematics Education* (pp. 292–299). PME. https://eventos.cmm.uchile.cl/pme48/wp-content/uploads/sites/195/2025/07/preliminary_v2_pme48_proceedings_generalcontributions.pdf

Fadhilatullathifi, Z. N, Ardiyanto, B., & Sari, W. P. (2020). Four-tier diagnostic test method to identify conceptual understanding in calculus. *Journal of Physics: Conference Series, 1613*(1), 012075. https://doi.org/10.1088/1742-6596/1613/1/012075

Hlongwana, P., Mudaly, V., & Zulu, M. W. (2025). Enhancing geometry problem-solving through visualization for multilingual learners. *Eurasia Journal of Mathematics, Science and Technology Education*, 21(4), 411–425. https://doi.org/10.29333/ejmste/16564

Matitaputty, C., Nusantara, T., & Hidayanto, E. (2025). Examining pre-service mathematics teacher knowledge in error analysis of students' work in solving quadratic functions. *AIP Conference Proceedings, 3293*(1), 040006. https://doi.org/10.1063/5.0321234

Nguyen, T. D., Tai, N. L. T., & Thuy, N. T. M. (2025). Students' misconceptions in conditional probability: A case study in Grade 12. *Hue University Journal of Education Research, 18*(2), 76–89. https://csdlkhoahoc.hueuni.edu.vn/data/2025/8/Nguyen_Thi_Duyen.pdf

Orhani, S. (2024). Preparation of tests from the subject of mathematics according to Bloom's taxonomy. *International Journal of Research Publication and Reviews, 5*(2), 2335-2345. https://ijrpr.com/uploads/V5ISSUE2/IJRPR22812.pdf

Prabowo, A., Herman, T., & Fatimah, S. (2023). The characteristics of a four-tier diagnostic test: Classical test theory perspective. *International Journal on Emerging Mathematics Education*, *7*(2), 157–168. https://doi.org/10.12928/ijeme.v7i2.46837

Priyasmika, R., & Ilmah, M. (2025). Identification of misconceptions of students in the subject of acid-base using the four-tier diagnostic test (FTDT). *Ed-Humanistics: Journal Ilmu Pendidikan, 10*(2), 87–96. https://doi.org/10.33752/ed-humanistics.v10i1.8986

Pule, G. K., Khensane, M., & Given, M. A. (2025). Analyzing errors and misconceptions of 11th-grade learners in solving tangent-chord theorem problems: A case in Tshwane North District Secondary School. *Journal of Inquiry Based Activities*, *15*(1), 25–38. https://www.ated.info.tr/index.php/ated/article/view/260

Sulaiman, M. M., & Gabbadua, A. D. (2024). Assessment of pre-service teachers' conception of osmosis and diffusion using a four-tier diagnostic test in Ilorin, Nigeria. *Ilorin Journal of Science Education, 12*(3), 77–89. https://kwasuspace.kwasu.edu.ng/handle/123456789/3057

Syahmani, S., Rusmansyah, R., & Leny, L. (2025). Identifying students' undergraduate concept understanding and misconceptions on amino acids, proteins, and enzymes using a four-tier diagnostic test. *Journal of Mathematics, Science, and Education*, *6*(2), 134–150. https://doi.org/10.20527/jmsc.v6i2.13878

Wilung, H., & Alvarez, J. I. (2025). Students' misconceptions in algebraic concepts: A four-tier diagnostic test approach. *Journal Elemen*, *11*(2), 156–173. https://doi.org/10.33541/jelemen.v11i2.27604