Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 1, 173 - 190
e_ISSN: 2798-2351
DOI: 10.58524/jasme.v6i1.1056

Development of a rule-based adaptive four-tier diagnostic quiz system for identifying misconceptions in geometrical optics

Muhammad Luqman
Universitas Mulawarman, INDONESIA

Akhmad Irsyad*
Universitas Mulawarman, Indonesia

Muhammad Ibadurrahman Arrasyid Supriyanto
Universitas Mulawarman, INDONESIA

Riskan Qadar
Universitas Mulawarman, INDONESIA

Article Info

Article history:

Received: Jan 24, 2026
Revised: Feb 15, 2026
Accepted: March 10, 2026

Keywords:

Assessment; Adaptive; Four-tier test; Misconceptions; Physics education.

Abstract

Background: Misconceptions in geometric optics remain persistent, including among prospective physics teachers, and are often difficult to detect using conventional assessments that focus only on final answers. Although four-tier diagnostic tests provide deeper conceptual information, their implementation is generally static and may reduce assessment efficiency.
Aims: This study aimed to develop and evaluate a rule-based adaptive four-tier diagnostic quiz system capable of identifying misconceptions more efficiently while maintaining stable diagnostic classifications.
Method: The study employed a research and development approach using the ADDIE model. A web-based adaptive diagnostic was developed by integrating a four-tier scoring scheme with transparent rule-based adaptive decisions. The platform was tested on 36 prospective physics teacher students. Data were analyzed descriptively to examine adaptive test length, diagnostic pathways, and classification stability.
Results: The adaptive system reduced the average test length from 15 static items to 11.5 items, representing an efficiency gain of 23.3%. A total of 77.8% of participants reached diagnostic stability before the maximum item limit, and the classification consistency rate reached 83.3%. The system also revealed variations in misconception patterns across topics, with concave mirror concepts showing the highest proportion of strong misconceptions.
Conclusion: The rule-based adaptive four-tier system improved diagnostic efficiency while maintaining stable classification outcomes. The transparent adaptive mechanism makes the system suitable for formative diagnostic assessment in physics education, although further studies with larger samples are recommended.

To cite this article: Luqman, M., Irsyad, A., Supriyanto, M. I. A., & Qadar, R. (2026). Development of a rule-based adaptive four-tier diagnostic quiz system for identifying misconceptions in geometrical optics. *Journal of Advanced Sciences and Mathematics Education, 6*(*1*), 173-190.

INTRODUCTION

Misconceptions remain a persistent challenge in physics education, particularly in the domain of geometrical optics. Numerous studies report that students, including prospective physics teachers, may select correct answers while still holding inaccurate conceptual explanations when required to justify their reasoning (Kaltakci-Gurel et al., 2017; Soeharto et al., 2019). Research in physics education has consistently shown that students' conceptual difficulties often persist even after formal instruction and may hinder meaningful conceptual understanding (Ivanjek et al., 2016; Neidorf et al., 2020). These misconceptions frequently arise because students construct intuitive explanations that conflict with scientifically accepted models, making them difficult to identify using conventional assessment approaches (Rusilowati, 2017; Soeharto et al., 2019). Therefore, diagnostic assessment instruments that can capture students' reasoning processes and levels of confidence are

---

* **Corresponding author:**
Akhmad Irsyad, Universitas Mulawarman, INDONESIA
akhmadirsyad@unmul.ac.id

needed to identify conceptual misunderstandings more accurately (Hasan et al., 2017; Jang et al., 2019).

Conventional multiple-choice tests have limited ability to reveal misconceptions because correct responses may be produced through guessing or incomplete reasoning. As highlighted by Ivanjek et al. (2016), traditional assessment formats often fail to capture the underlying conceptual structures that shape students' understanding of physical phenomena. To address this limitation, multi-tier diagnostic assessments have been developed to evaluate not only students' answers but also their reasoning and confidence levels (Peşman & Eryilmaz, 2018). Among these formats, the four-tier diagnostic test has gained considerable attention because it integrates conceptual answers, reasoning, and confidence judgments within a single assessment framework (Kaltakci-Gurel et al., 2017). This structure allows researchers to distinguish between scientifically correct responses supported by appropriate reasoning and responses generated through guessing or partial conceptual understanding.

Empirical studies demonstrate that four-tier diagnostic instruments are effective in identifying various categories of conceptual understanding and misconceptions in science learning. Previous studies have applied four-tier diagnostic tests to investigate misconceptions in various physics topics, including Newton's laws, fluid mechanics, and harmonic motion (Kanıawatı et al., 2019; Putranta & Afifah, 2025; Tumanggor et al., 2020). Similar approaches have also been used to diagnose misconceptions in heat, temperature, and electromagnetism (Aini et al., 2025; Festiana et al., 2019). More recent studies have continued to develop and apply four-tier diagnostic instruments in physics learning contexts to identify students' misconceptions more precisely (Çelikkanlı & Kızılcık, 2022; Kiray & Simsek, 2021; Rusilowati et al., 2024). These findings confirm that misconceptions remain prevalent even among university students and prospective teachers, highlighting the need for diagnostic approaches capable of capturing students' conceptual structures more comprehensively.

Alongside the development of diagnostic assessment approaches, advances in educational technology have encouraged the development of adaptive assessment systems. Technology-enhanced assessment environments can improve measurement efficiency and support individualized feedback during the learning process (Spector et al., 2016). Moreover, recent developments in artificial intelligence and digital learning technologies indicate that adaptive assessment can support data-driven learning environments that respond to students' individual characteristics (Holmes et al., 2019; Zawacki-Richter et al., 2019). Computer-based assessment systems also provide opportunities to dynamically adjust assessment items based on learners' responses, allowing assessment processes to become more flexible and responsive (Sargazi Moghadam et al., 2024; Shute & Rahimi, 2017).

Despite these developments, many adaptive assessment systems are primarily designed to estimate students' ability levels rather than diagnose conceptual misconceptions in depth. In other words, adaptive testing is often focused on performance measurement rather than identifying the conceptual reasoning underlying students' answers. At the same time, most implementations of four-tier diagnostic tests are still conducted in static formats, where all students complete the same sequence of items regardless of their conceptual profiles (Rusilowati, 2017). Such static structures may reduce assessment efficiency and increase cognitive load during diagnostic testing.

Therefore, a gap remains between the diagnostic depth offered by four-tier tests and the efficiency provided by adaptive assessment systems. Four-tier diagnostic tests provide detailed information about students' conceptual understanding, yet their implementation remains largely static. In contrast, adaptive assessment systems emphasize efficiency and personalization but often lack the ability to diagnose misconceptions in detail. Bridging this gap requires the development of an assessment approach that integrates the conceptual diagnostic capabilities of four-tier tests with an adaptive mechanism that is transparent, efficient, and applicable in classroom contexts.

Based on this rationale, the present study aims to develop and evaluate a rule-based adaptive four-tier diagnostic quiz system designed to identify misconceptions in geometrical optics more efficiently while maintaining stable diagnostic classifications. The proposed system integrates the diagnostic strength of four-tier assessments with a rule-based adaptive mechanism that dynamically adjusts item difficulty according to students' diagnostic responses. This study contributes to physics education research in three ways. First, it proposes a rule-based adaptive diagnostic framework that combines conceptual diagnosis with adaptive assessment efficiency. Second, it introduces a transparent adaptive decision mechanism that can be implemented in classroom-based formative assessment without requiring complex probabilistic calibration models. Third, the study provides empirical evidence regarding the efficiency and diagnostic stability of an adaptive four-tier diagnostic system in identifying misconception patterns among prospective physics teachers.

METHOD

Research Design

This study employed a research and development (R&D) approach using the ADDIE model, which consists of five stages: analysis, design, development, implementation, and evaluation. The ADDIE model was selected because it provides a systematic framework for designing, implementing, and evaluating educational products and assessment systems (Branch, 2009; Richey & Klein, 2014).

The purpose of this study was not to test statistical hypotheses, but to design and evaluate the functionality, rule performance, and diagnostic characteristics of a rule-based adaptive four-tier assessment system. The analysis stage involved identifying common misconceptions in geometric optics. The design stage included the development of four-tier diagnostic items and adaptive decision rules. The development stage involved system programming and expert validation. The implementation stage consisted of a limited field trial. The evaluation stage focused on system performance, adaptive pathways, and diagnostic classification outcomes. The overall development process following the ADDIE framework is illustrated in Figure 1.

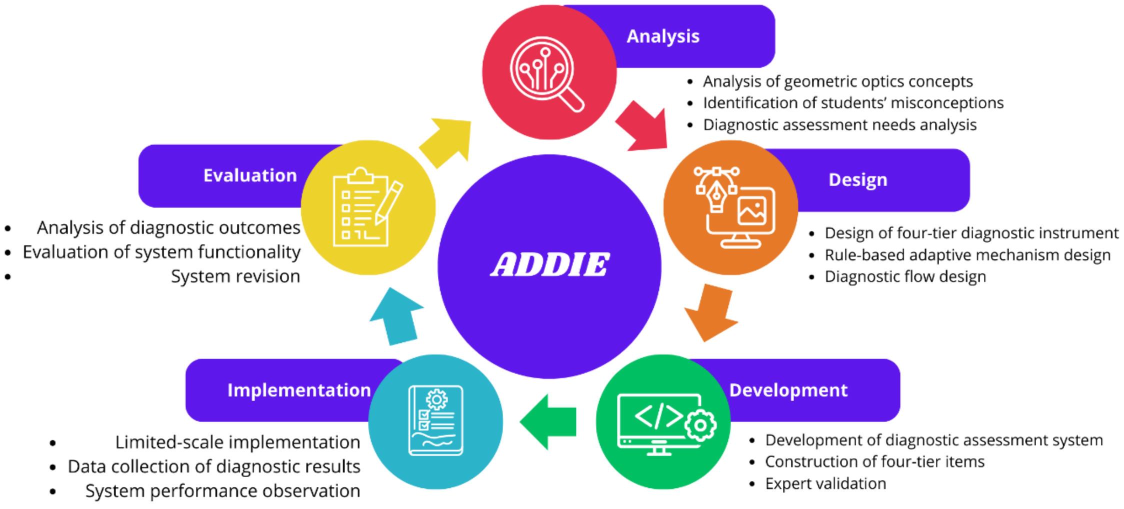


Figure 1. Stages of diagnostic assessment system development using the ADDIE model.

Participants

The participants in this study were prospective physics teachers enrolled in a physics education program at a state university. A total of 36 students voluntarily participated in the limited field trial phase. All participants had previously completed a course in geometric optics. Participation was conducted with informed consent, and all data were anonymized to ensure confidentiality.

The sample size of 36 participants was considered sufficient for functional system testing, rule verification, and diagnostic pathway evaluation, which were the primary objectives of this development study. In design-based and prototype-oriented research, limited samples are commonly used to evaluate system logic, termination criteria, and rule stability before large-scale deployment (Branch, 2009; Richey & Klein, 2014; Annisa et al., 2024).

Therefore, the sample size was not intended for statistical generalization, but for evaluating system functionality and diagnostic decision consistency. This constitutes a methodological limitation, and further studies with larger and more diverse samples are recommended.

Population and Sampling

The population of this study consisted of all prospective physics teachers in the physics education program. Convenience sampling was employed because the main objective of the study was the development and functional evaluation of the adaptive diagnostic system rather than statistical generalization.

In development-oriented research, limited and accessible samples are commonly used to test system functionality, rule performance, and termination mechanisms under controlled conditions (Branch, 2009; Richey & Klein, 2014). Accordingly, the number of participants was determined based on the needs of system testing, including verification of adaptive rules, diagnostic stability, and classification consistency.

Instruments

The diagnostic instrument used in this study was a four-tier diagnostic test covering five topics in geometric optics: plane mirrors, concave mirrors, convex mirrors, convex lenses, and concave lenses. Each item consisted of four levels:

1. Conceptual multiple-choice question
2. Confidence level in the conceptual answer
3. Multiple-choice reasoning question
4. Confidence level in the reasoning

This structure allows the instrument to identify not only answer correctness but also reasoning quality and confidence level, thereby enabling more detailed diagnosis of misconceptions (Caleon & Subramaniam, 2010; Gurel et al., 2015). The structure of the four-tier instrument is summarized in Table 1.

Table 1. Four-tier diagnostic test structure and scoring scheme

Tier	Description	Response Type	Diagnostic Function
Tier 1	Conceptual question related to geometrical optics concepts	Multiple-choice answer	Identifies students' initial conceptual responses
Tier 2	Confidence level for the Tier 1 answer	Confidence scale (e.g., confident / not confident)	Distinguishes between confident answers and guesses
Tier 3	Reasoning question explaining the Tier 1 answer	Multiple-choice reasoning	Reveals students' underlying conceptual reasoning
Tier 4	Confidence level for the Tier 3 answer	Confidence scale (e.g., confident / not confident)	Determines the stability of students' conceptual reasoning

Diagnostic Scoring Scheme

The diagnostic score for each item was computed using a rule-based four-tier classification scheme. Let:

- C_1 = correctness of conceptual answer (1 = correct, 0 = incorrect)

- R = correctness of reasoning (1 = correct, 0 = incorrect)
- K_1 = confidence in conceptual answer (1 = confident, 0 = not confident)
- K_2 = confidence in reasoning (1 = confident, 0 = not confident)

The diagnostic score was determined using a rule-based classification scheme rather than a direct mathematical formula. Each combination of conceptual answer correctness, reasoning correctness, and confidence levels was mapped into one of four diagnostic categories: sound understanding, partial understanding, weak misconception, or strong misconception. The complete classification rules are presented in Appendix A. The final classification is interpreted using the following rules:

Table 2. Classification Rules

Score	Category
3	Sound conceptual understanding
2	Partial understanding
1	Weak misconception
0	Strong misconception

This scoring approach integrates correctness and confidence dimensions to differentiate levels of conceptual stability, consistent with four-tier diagnostic principles (Caleon & Subramaniam, 2010; Gurel et al., 2015). A detailed rule table and pseudo-code for the adaptive mechanism are provided in Appendix A.

Content Validity *and Reliability*

The content validity of the instrument was evaluated through expert judgment involving three physics education experts. Each item was assessed in terms of conceptual relevance, clarity, distractor quality, and alignment with geometric optics concepts.

The Content Validity Index (CVI) was calculated at the item level (I-CVI) and scale level (S-CVI). All items obtained I-CVI values above 0.80, and the overall S-CVI exceeded 0.90, indicating strong expert agreement and high content validity (Lynn, 1986; Polit & Beck, 2006).

Because the system employed an adaptive structure in which participants did not receive identical item sequences, traditional internal consistency measures such as Cronbach's alpha were not emphasized. Instead, reliability was evaluated using classification consistency indicators, including:

1. Consistency of diagnostic categories across consecutive items within a topic
2. Stability of classifications across adaptive pathways
3. Repeatability of diagnostic outcomes across different topics

These indicators align with recommended reliability approaches for adaptive and diagnostic assessment systems, where decision stability is more relevant than internal consistency coefficients (Shute & Zapata-Rivera, 2012; Van der Linden & Glas, 2010).

Adaptive Diagnostic System

The diagnostic system was implemented as a web-based application integrating four main components:

1. User interface module
2. Four-tier diagnostic engine
3. Rule-based adaptive decision module
4. Database for storing responses and results

The system workflow is illustrated in Figure 2.

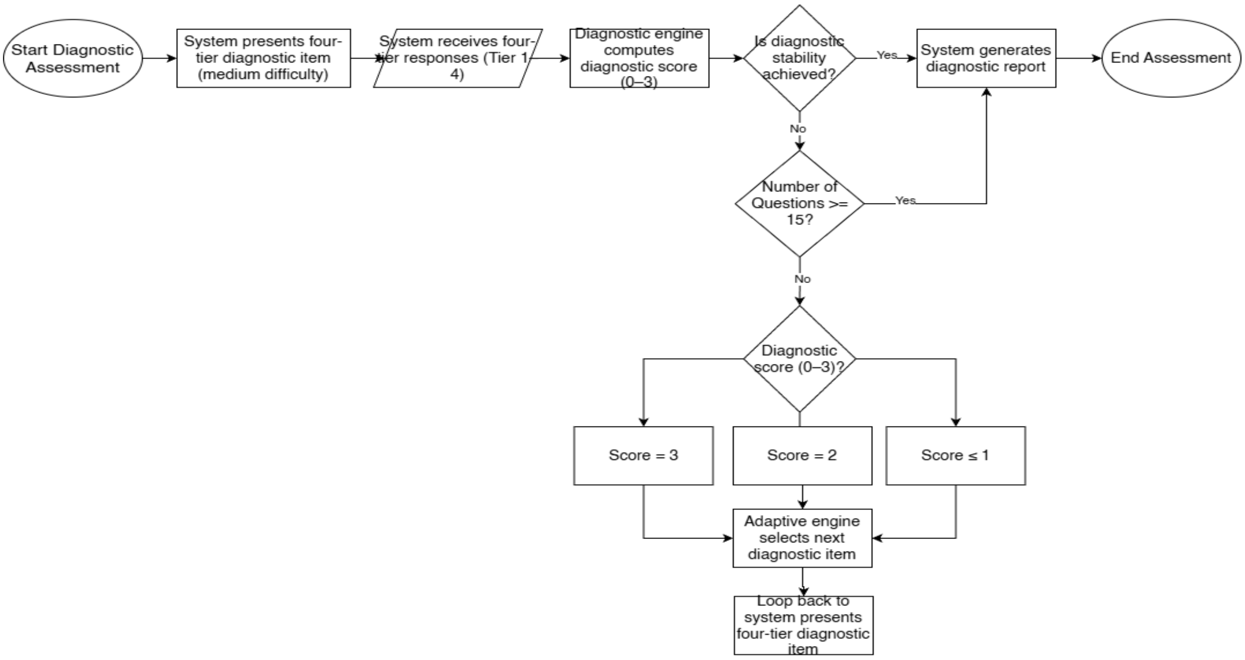


Figure 2. Rule-based adaptive decision-making flow of the diagnostic assessment system.

Termination Criteria

The system used a dual termination mechanism. The assessment was automatically terminated when one of the following conditions was met:

1. Diagnostic stability achieved
2. Maximum of 15 items reached

Diagnostic stability was operationally defined as the repetition of the same diagnostic category across the last two consecutive items within a topic. This rule ensured that final classifications were based on consistent response patterns rather than single-item outcomes.

Adaptive Decision Rules

If the termination criteria were not met, the system executed a rule-based adaptive mechanism to determine the next item. The adaptive decision logic was defined as follows:

Adaptive decision rules:

1. Start with a medium-difficulty item.
2. Compute diagnostic score S.
3. If $S = 3$: → move to a higher-difficulty item
4. If $S = 2$: → present a confirmation item at the same difficulty level.
5. If $S \leq 1$: → move a lower-difficulty or remedial item.
6. After each item, check termination criteria:
    - If diagnostic stability achieved → terminate
    - Else if item count ≥ 15 → terminate
    - Else → continue adaptive selection

The adaptive logic flow is presented in Figure 3.

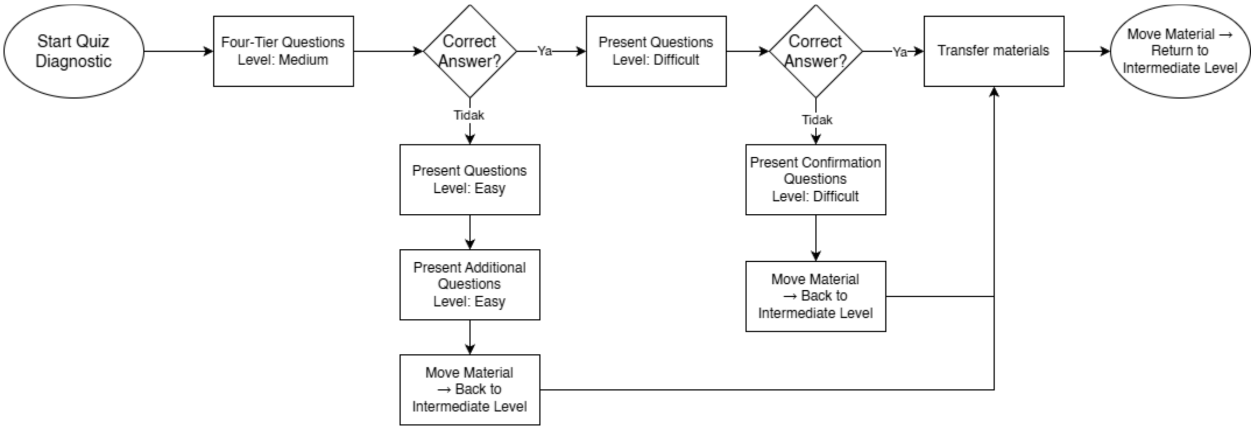


Figure 3. Adaptive Logic Flowchart for Item Difficulty Determination

System Architecture

The adaptive decision-making process operates within a multi-layered system architecture, as illustrated in Figure 4. The architecture consists of four layers:

1. User Layer
2. Presentation Layer
3. Logic Layer
4. Database Layer

The Logic Layer integrates three core components:

- Four-tier diagnostic engine
- Rule-based adaptive engine
- Learning analytics module

This architecture supports diagnostic validity by ensuring that each adaptive decision is directly derived from the four-tier diagnostic score and rule-based classification logic. The separation of diagnostic processing, adaptive decisions, and analytics ensures transparency of classification rules and traceability of diagnostic pathways.

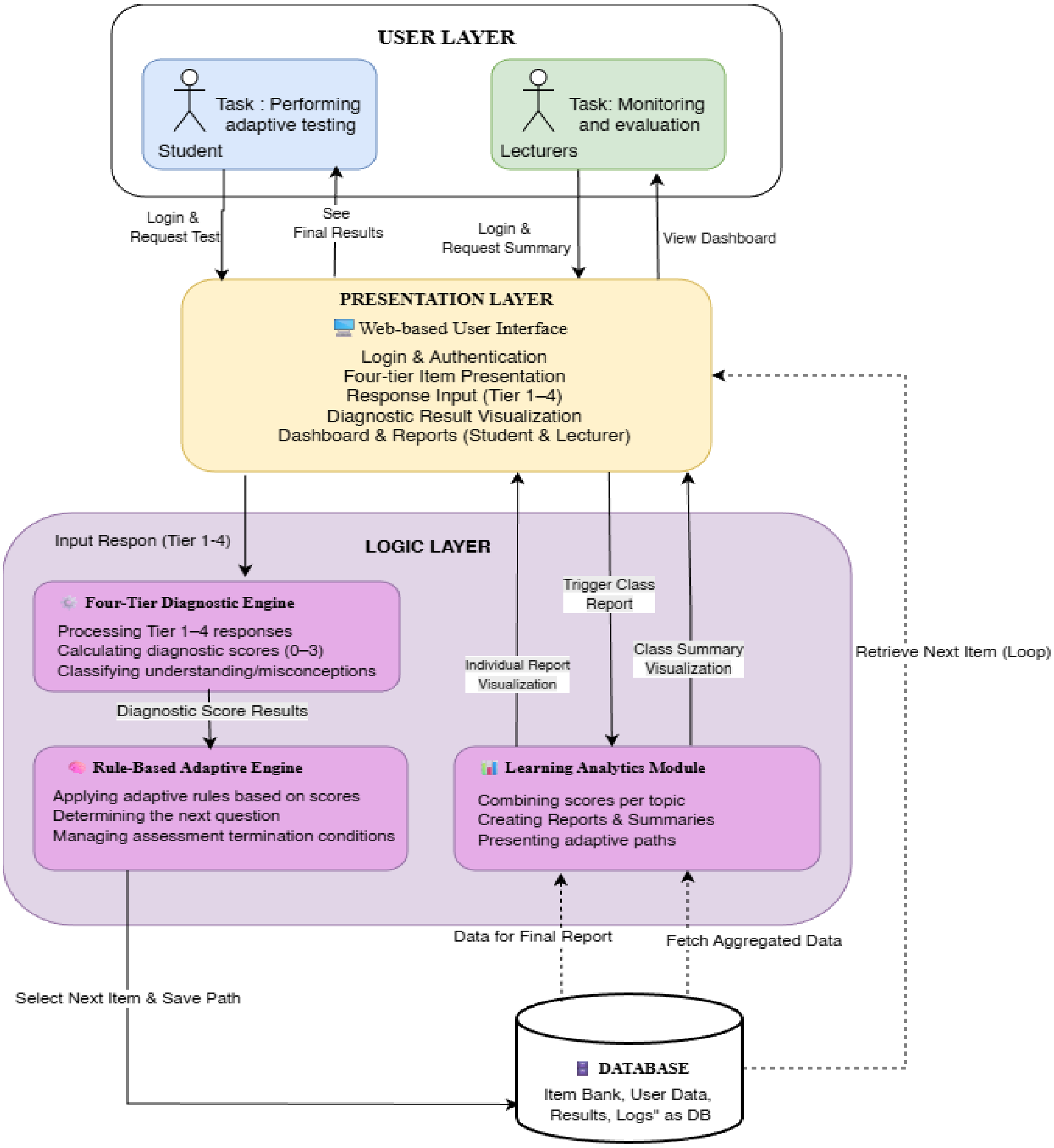


Figure 4. User-oriented system architecture of the web-based adaptive four-tier diagnostic assessment system.

Procedures

The research procedures were conducted in four stages:

1. Development and expert validation of four-tier diagnostic items
2. Formulation of adaptive rules based on diagnostic score thresholds
3. Implementation and functional testing of the web-based system
4. Individual completion of the adaptive diagnostic quiz by participants

During the implementation stage, the system automatically recorded:

- Student responses
- Diagnostic scores
- Adaptive pathways
- Final misconception classifications

Data Analysis

Data were analyzed using descriptive statistical methods. The analysis focused on:

1. Number of items completed by each participant
2. Adaptive diagnostic pathways
3. Distribution of conceptual understanding categories

Descriptive analysis was considered appropriate because the primary objective of the study was to evaluate system functionality, rule performance, and diagnostic characteristics rather than to test inferential hypotheses (Fraenkel et al., 2012).

RESULTS AND DISCUSSION

Results

System Performance and Adaptive Test Efficiency

The rule-based adaptive four-tier diagnostic system functioned properly during the implementation phase. All participants were able to access the web-based platform and complete the diagnostic assessment without technical difficulties. The system successfully recorded student responses, applied adaptive decision rules, and generated diagnostic outputs in real time.

To evaluate the efficiency and stability of the adaptive system, several performance indicators were calculated. A comparison between the static four-tier test length and the adaptive test outcomes is presented in Table 2.

Table 2. Efficiency and Diagnostic Stability Indicators of the Adaptive System

Indicator	Definition	Result
Static test length	Total items if using full four-tier instrument	15 items
Adaptive test length (min–max)	Range of items completed per participant	10–14 items
Mean adaptive test length	Average number of items completed	11.5 items
Average item reduction	Reduction compared to static test	3.5 items (23.3%)
Participants reaching stability before max items	Terminated by stability rule	28 of 36 (77.8%)
Participants reaching max item limit	Terminated by item limit	8 of 36 (22.2%)
Classification consistency rate	Stable category in last two items	30 of 36 (83.3%)
Mean confirmation items per participant	Average number of confirmation items	1.2 items
Maximum confirmation items	Highest number observed	3 items

The static diagnostic instrument consisted of 15 items across five geometric optics topics. In contrast, the adaptive system assigned between 10 and 14 items per participant, with an average of 11.5 items. This indicates an average reduction of 3.5 items, or approximately 23.3% compared to the static test length.

In terms of termination mechanisms, 28 out of 36 participants (77.8%) reached diagnostic stability before the maximum item limit was reached, while only 8 participants (22.2%) completed the full item limit. This suggests that the stability-based termination rule functioned effectively in shortening the diagnostic process.

The classification consistency rate, defined as the proportion of participants achieving the same diagnostic category in the final two consecutive items, reached 83.3%. These findings indicate that most classifications were derived from stable response patterns rather than isolated answers.

Additionally, the system triggered an average of 1.2 confirmation items per participant, with a maximum of three confirmation items observed in more complex diagnostic pathways. This confirms the role of confirmation items in stabilizing classification decisions, consistent with adaptive diagnostic principles (Shute & Zapata-Rivera, 2012).

Characteristics of Adaptive Diagnostic Pathways

Analysis of individual diagnostic pathways revealed that all participants initially received items with moderate difficulty levels. Based on the four-tier diagnostic scores, the system subsequently adjusted item difficulty in accordance with the predefined decision rules.

When students obtained high diagnostic scores, the system directed them to higher-difficulty items to test the depth of their conceptual understanding. Conversely, students who demonstrated weak understanding or misconception indicators were presented with lower-difficulty or remedial items.

The system rarely assigned high-difficulty items consecutively to participants exhibiting strong misconceptions. Instead, confirmation items were employed when ambiguous diagnostic scores occurred. These confirmation items allowed the system to stabilize the classification before moving to a new difficulty level.

An example of an individual adaptive diagnostic pathway is illustrated in Figure 5, showing how item difficulty *and classification decisions evolved across the assessment sequence.*

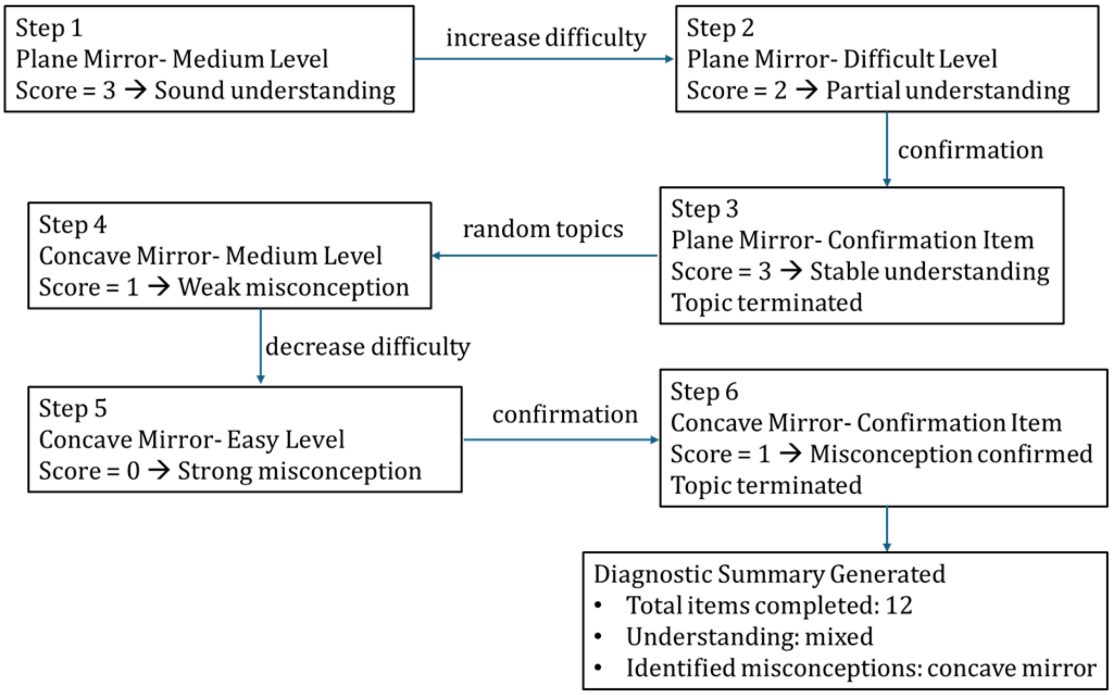


Figure 5. Example of an individual adaptive diagnostic pathway

Misconception Profile Distribution

The analysis of diagnostic results showed that students' levels of conceptual understanding varied across geometric optics topics. Differences in diagnostic scores and category distributions indicated that conceptual difficulties were not evenly distributed, but rather concentrated in certain subtopics. The distribution of conceptual understanding categories across topics is presented in Figure 6.

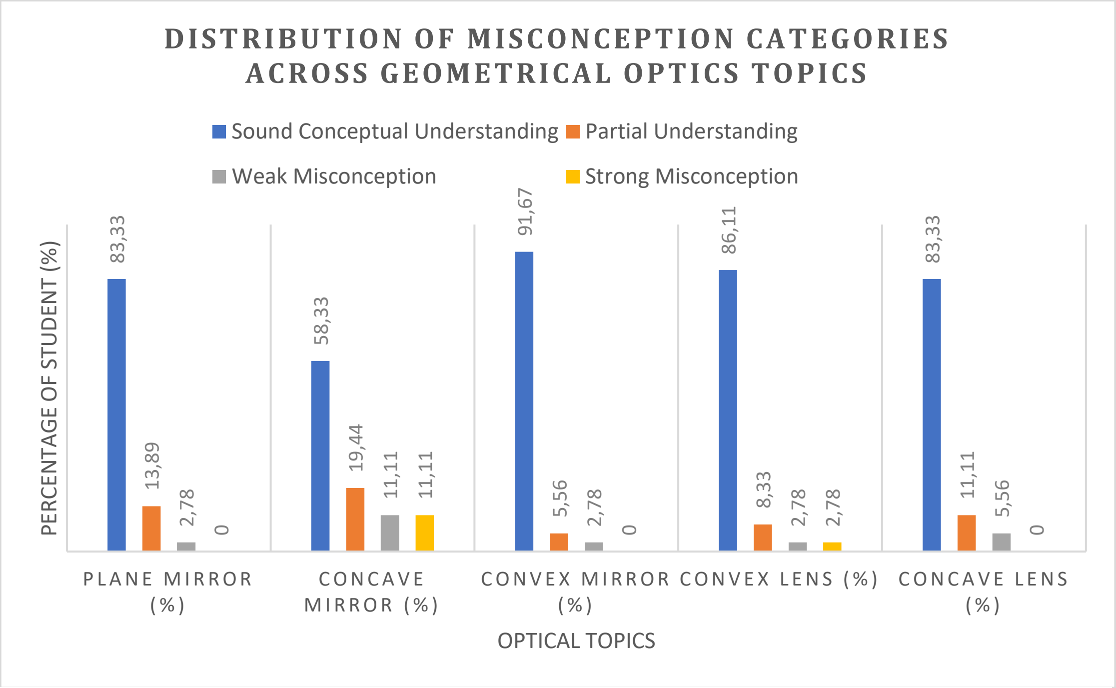


Figure 6. Distribution of misconception categories across geometrical optics topics

The overall distribution of conceptual understanding categories across topics is presented in Figure 6. Overall, the Sound Conceptual Understanding category dominated most topics, with an average proportion above 80%. The topic of convex mirrors recorded the highest level of conceptual understanding at 91.67%, followed by convex lenses at 86.11%. These results indicate that these topics were relatively well mastered by the majority of participants. For a more detailed numerical representation of the category distribution, the data are summarized in Table 3.

Table 3. Distribution of Conceptual Understanding Categories Across Geometric Optics Topics

Topic	Sound Understanding (%)	Partial Understanding (%)	Weak Misconception (%)	Strong Misconception (%)
Plane mirror	83.33	13.89	2.78	0
Concave mirror	58.33	19.44	11.11	11.11
Convex mirror	91.67	5.56	2.78	0
Convex lens	86.11	8.33	2.78	2.78
Concave lens	83.33	11.11	5.56	0

Table 3 presents the numerical distribution of conceptual understanding categories across geometric optics topics. The data show that most topics are dominated by the sound conceptual understanding category, with proportions above 80%. However, the concave mirror topic exhibits the lowest conceptual understanding and the highest proportions of partial understanding and strong misconception categories.

However, a significant anomaly was observed in the concave mirror topic. This topic recorded the lowest conceptual understanding percentage at 58.33%, compared to other topics. It also showed the highest proportion of partial understanding (19.44%) and strong misconception (11.11%) categories.

The presence of high percentages in these non-scientific categories suggests that the complexity of image formation in concave mirrors represents a major conceptual barrier for students. This finding confirms previous studies reporting persistent misconceptions in geometric optics, particularly in image formation by mirrors and lenses. A more specific visualization of misconception patterns for the mirror topics is shown in Figure 7.

The results of the study show that the rule-based four-tier adaptive diagnostic quiz system is capable of integrating in-depth conceptual diagnosis with assessment process efficiency. As shown in Figure 7, students work on questions with four levels of response, namely conceptual answers, level of confidence in the answers, reasoning, and confidence in the reasoning. This structure allows the system to not only identify whether an answer is correct or incorrect, but also the stability of the learner's conceptual understanding.

Waktu Tersisa: 00:29:52
Pertanyaan ke: 1
Topik: Cermin Datar

Muhammad Luqman
2309116068
Mahasiswa

Universitas
Universitas Mulawarman

Jurusan
Pendidikan Matematika

Kelas
-

Dosen
Dr. Riskan Qadar M.Si

[Tingkat 1: Pertanyaan Utama]

Seseorang berdiri 2 m di depan cermin datar. Berapa jarak antara orang tersebut dengan bayangannya?

○ 2 m
○ 4 m
○ 1 m
○ 0 m

[Tingkat 2: Keyakinan pada Jawaban]

○ Sangat yakin
○ Yakin
○ Tidak yakin
○ Sama sekali tidak yakin

[Tingkat 3: Penalaran]

○ Bayangan cermin datar berjarak setengah dari benda.
○ Bayangan berada di belakang cermin sejauh jarak benda → total jarak = 4 m.
○ Bayangan ada di permukaan cermin.
○ Bayangan selalu 0 m dari benda.

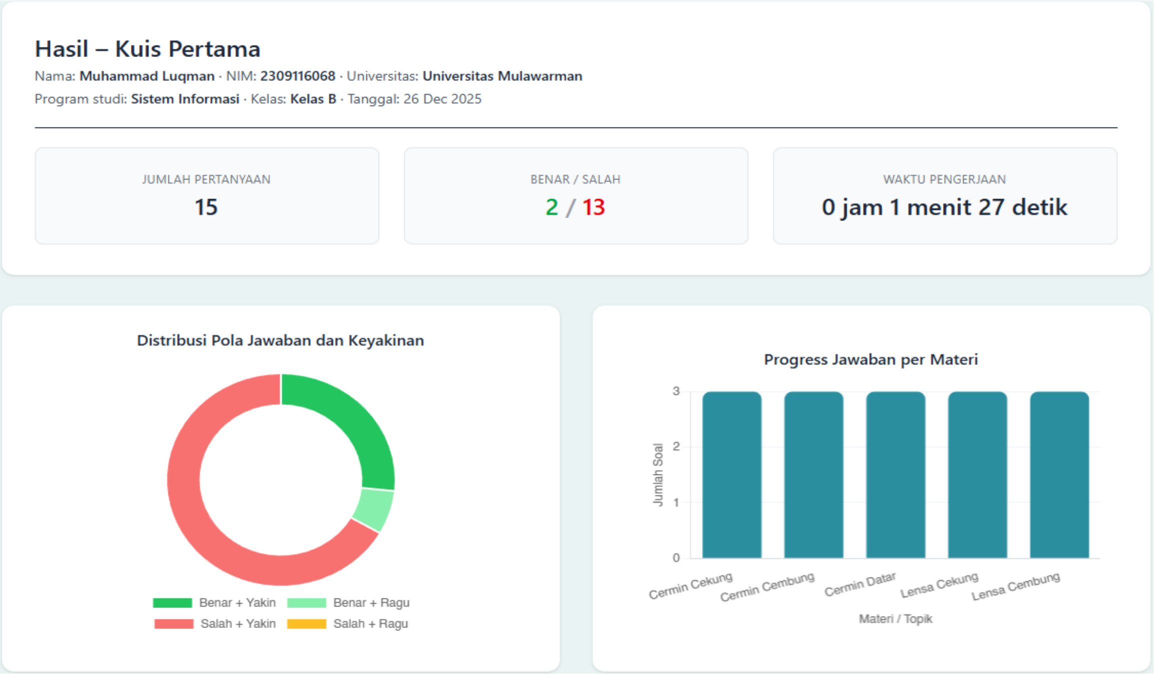
[Tingkat 4: Keyakinan pada Penalaran]

○ Sangat yakin
○ Yakin
○ Tidak yakin
○ Sama sekali tidak yakin

Figure 7. Example of a Four-Tier Diagnostic Item Interface in the Adaptive Quiz System

Classifying misconceptions into weak and strong categories provided more granular diagnostic insights than assessments based solely on final scores. This differentiation is pedagogically important because it enables educators to distinguish between learners who experience conceptual uncertainty and those who possess deeply rooted misconceptions, thus requiring different remediation strategies (Caleon & Subramaniam, 2010).

A summary of the assessment results and system decisions is shown in Figure 8. The system presents information on the number of items completed, the distribution of correct and incorrect responses, response time, and visualizations of students' answer patterns and confidence levels. Based on these data, the system automatically generates a diagnostic classification, such as weak misconception or sound understanding. This structured presentation allows educators to quickly interpret students' conceptual conditions in a concise and integrated format.



Hasil – Kuis Pertama

Nama: **Muhammad Luqman** · NIM: **2309116068** · Universitas: **Universitas Mulawarman**

Program studi: **Sistem Informasi** · Kelas: **Kelas B** · Tanggal: 26 Dec 2025

JUMLAH PERTANYAAN	BENAR / SALAH	WAKTU PENGERJAAN
15	2 / 13	0 jam 1 menit 27 detik

Figure 8. Summary of Students Diagnostic Results Generated by the Adaptive Four-Tier Quiz System

The diagnostic decision-making process is summarized in Figure 9. The final classification is derived from the four-tier diagnostic score, which integrates conceptual accuracy, reasoning, and confidence levels into categories of conceptual understanding and misconception strength. Figure 9 illustrates the comparison of misconception strengths across all geometric optics *topics.*

KEPUTUSAN SISTEM

Miskonsepsi Lemah

"Terdeteksi banyak keraguan. Kamu sering salah konsep tapi menyadarinya (tidak yakin)."

RATA-RATA SKOR

1.33/3.0

Proses Hitung

$$\frac{\text{Total Poin}}{\text{Jumlah Soal}} = \text{Skor Akhir}$$

$$\frac{20}{15} = 1.33$$

Tabel Indikator Penilaian

Rata-rata skor (0-3)	Kategori konsepsi	Makna
2.5 - 3.0	Paham benar	Jawaban dan alasan konsisten, keyakinan tinggi
1.5 - 2.4	Parsial / transisi	Kadang benar, alasan kurang stabil
0.5 - 1.4	**Miskonsepsi lemah**	Salah konsep tapi masih ragu
0.0 - 0.4	Miskonsepsi kuat	Salah konsep dengan keyakinan tinggi

Figure 9. Example of Scoring Process and Conceptual Category Determination in the Adaptive Four-Tier Diagnostic System

These visualizations consistently indicate that conceptual difficulties are concentrated in specific subtopics, particularly those involving image formation in concave mirrors, while other topics are generally dominated by sound conceptual understanding.

Discussion

One of the main findings of this study is the reduction in the number of items required to reach stable diagnostic classifications when using the rule-based adaptive diagnostic system. While the static four-tier diagnostic instrument consisted of 15 items, the adaptive system required only 10–14 items, with an average of 11.5 items. This corresponds to an average reduction of approximately 23.3% in test length. Such efficiency indicates that adaptive mechanisms can shorten the diagnostic process without sacrificing the accuracy of classification outcomes. This result supports the fundamental principle of adaptive assessment, which aims to improve measurement efficiency by adjusting item difficulty and selection based on student responses (Sargazi Moghadam et al., 2024; Shute & Rahimi, 2017). Unlike conventional adaptive systems that rely on probabilistic models or item response theory, the system developed in this study operates using transparent rule-based decisions. This transparency makes the adaptive mechanism easier to interpret and implement in classroom contexts, particularly for formative diagnostic assessment (Spector et al., 2016).

The high proportion of participants who reached diagnostic stability before the maximum item limit further demonstrates the effectiveness of the termination rule implemented in the system. Specifically, 77.8% of participants achieved stable diagnostic classifications before reaching the maximum number of items. This finding suggests that stable classifications can be achieved without requiring all students to complete the full set of diagnostic questions. As a result, the adaptive system can reduce unnecessary testing time and minimize cognitive load during the assessment process. Reducing excessive cognitive demands during assessment is important because complex or lengthy tests may negatively affect students' concentration and response quality. Similar findings have been reported in studies on adaptive assessment systems, which indicate that adaptive item selection can improve testing efficiency while maintaining reliable diagnostic outcomes (Sargazi Moghadam et al., 2024; Shute & Rahimi, 2017). Therefore, the reduction in test length observed in this study reflects an important advantage of adaptive diagnostic systems in improving assessment efficiency.

In adaptive diagnostic systems, reliability is more appropriately interpreted as classification stability rather than internal consistency across identical item sets. Because participants follow different adaptive pathways, traditional reliability indices such as Cronbach's alpha become less meaningful. Instead, reliability can be evaluated through the consistency of diagnostic classifications across successive items. The classification consistency rate of 83.3% observed in this study indicates that most students reached stable diagnostic categories across consecutive diagnostic decisions. This suggests that the system's classifications were not based on isolated responses but on consistent conceptual patterns demonstrated by the students. The inclusion of confirmation items also contributed to this stability by preventing premature classification changes and ensuring that the final diagnostic outcomes reflected stable conceptual tendencies. Similar perspectives have been discussed in the literature on adaptive assessment and learning analytics, which emphasize classification stability and decision consistency as key indicators of reliability in adaptive diagnostic systems (Jang et al., 2019; Sargazi Moghadam et al., 2024).

The distribution of misconception categories across topics also revealed important diagnostic insights. Most topics were dominated by the sound conceptual understanding category, with proportions exceeding 80%, indicating that many students had relatively stable conceptual knowledge in those areas. However, the concave mirror topic showed substantially lower conceptual understanding and higher proportions of partial understanding and strong misconception categories. This pattern indicates that certain topics in geometrical optics remain conceptually challenging for students. Image formation by concave mirrors requires the integration of several abstract concepts, including focal relationships, ray tracing, and sign conventions, which often leads to persistent misconceptions even after formal instruction. Similar findings have been reported in previous research showing that geometrical optics concepts frequently produce persistent misconceptions among students and pre-service teachers (Kaltakci-Gurel et al., 2017; Soeharto et al., 2019). The diagnostic system developed in this study successfully detected these variations in conceptual understanding across different topics.

The adaptive four-tier system was also able to classify misconceptions into more detailed categories, including weak misconception, partial understanding, and strong misconception. This level of diagnostic differentiation is pedagogically important because it allows educators to distinguish between students who are uncertain about a concept and those who hold firmly rooted misconceptions. Such distinctions are crucial for designing targeted instructional interventions. Students who demonstrate partial understanding may benefit from clarification and conceptual reinforcement, whereas students with strong misconceptions may require more intensive conceptual change strategies. Similar conclusions have been reported in studies on diagnostic assessment, which emphasize the importance of identifying different levels of conceptual understanding in order to design effective remediation strategies (Soeharto et al., 2019; Rusilowati, 2017). Therefore, the ability of the adaptive diagnostic system to provide detailed misconception classifications represents an important contribution to diagnostic assessment practices in physics education.

Another important aspect of the proposed system is the transparency of its rule-based adaptive mechanism. Unlike many adaptive assessment systems that rely on complex probabilistic models, the system developed in this study applies explicit decision rules based on diagnostic scores obtained from the four-tier responses. This transparency enhances diagnostic validity because educators can clearly trace how each classification decision is generated by the system. In addition, the adaptive pathways are directly linked to students' conceptual understanding levels rather than abstract estimates of ability. Such interpretability is particularly important in formative and diagnostic assessment contexts, where the results must support instructional decision-making (Spector et al., 2016; Holmes et al., 2019).

The architecture of the system also contributes to the validity of the diagnostic process. The separation between the diagnostic engine, adaptive decision rules, and analytical modules ensures that classification outcomes are consistently derived from the four-tier diagnostic responses. This layered structure reduces the risk of inconsistent decisions and allows the diagnostic mechanism to remain transparent and interpretable. As a result, the rule-based adaptive framework not only improves assessment efficiency but also maintains stable and meaningful diagnostic classifications. Overall, the findings of this study demonstrate that integrating four-tier diagnostic assessment with a rule-based adaptive mechanism can provide a practical and efficient approach for identifying students' misconceptions in geometrical optics.

Implications

The results of this study suggest that rule-based adaptive four-tier diagnostic systems can provide efficient and meaningful diagnostic information about students' conceptual understanding. By reducing test length while maintaining stable classifications, the system enables a more practical implementation of diagnostic assessment in classroom settings. In addition, the ability to classify misconceptions into different strength levels provides valuable information for designing targeted instructional interventions. Such diagnostic insights are particularly important in physics education, where misconceptions are often persistent and difficult to address through traditional instruction. Overall, these findings indicate that transparent, rule-based adaptive diagnostic systems represent a practical and interpretable approach to formative assessment in physics education

Limitations

Despite the promising results, several limitations should be acknowledged. First, the sample size was relatively small and obtained through convenience sampling, which limits the generalizability of the findings to the broader population of prospective physics teachers, although it was sufficient for functional system testing and rule verification. Second, the reliability evidence in this study was based on classification consistency indicators rather than traditional internal consistency measures. While this approach is appropriate for adaptive diagnostic systems, future studies should involve larger samples and repeated measurements to examine test–retest stability more rigorously. Third, although the rule-based adaptive mechanism provides transparency and practical implementation advantages, it does not offer the same level of measurement precision as probabilistic adaptive models. Therefore, future research may explore hybrid approaches that integrate rule-based transparency with statistical calibration techniques to enhance diagnostic accuracy

Suggestions

Future studies are encouraged to expand the implementation of the rule-based adaptive four-tier diagnostic system across larger and more diverse samples of students to examine its robustness in different educational contexts. In addition, further research may investigate the application of the adaptive diagnostic framework to other physics topics beyond geometrical optics in order to evaluate its generalizability in diagnosing conceptual misconceptions. Researchers may also explore the integration of rule-based adaptive mechanisms with statistical or probabilistic calibration models to enhance measurement precision while maintaining transparency in diagnostic decision-making. Finally, future work could investigate how the diagnostic information generated by adaptive four-tier systems can be directly linked to targeted instructional interventions designed to support conceptual change and improve students' understanding of challenging physics concepts

CONCLUSION

This study developed and evaluated a rule-based adaptive four-tier diagnostic quiz system for identifying misconceptions in geometric optics. The system successfully generated individualized diagnostic pathways and dynamically adjusted item difficulty based on students' conceptual responses. The adaptive mechanism reduced the average test length from 15 static items to 11.5 items, representing an efficiency gain of approximately 23.3%. Most participants (77.8%) reached diagnostic stability before the maximum item limit, and the classification consistency rate reached 83.3%, indicating stable diagnostic decisions. It was also able to differentiate misconception patterns across topics, with concave mirror concepts showing the highest concentration of strong misconceptions. These findings demonstrate that a transparent rule-based adaptive approach can improve diagnostic efficiency while maintaining stable classification outcomes. The system has practical potential for formative assessment in physics education, although further studies with larger samples are recommended to examine generalizability and long-term reliability.

AUTHOR CONTRIBUTIONS STATEMENT

ML served as the principal investigator who designed the study, developed the adaptive diagnostic quiz system, collected and analyzed data, and drafted the initial manuscript. AI contributed as the principal advisor who provided conceptual and methodological guidance, as well as conducting critical review and revision of the manuscript. RS acted as the second supervisor who provided input on methodology, data analysis, and manuscript refinement. RQ contributed in the field of physics education, particularly in strengthening the pedagogical foundation, interpreting research results, and improving the substance of the discussion. All authors have read and approved the final version of the manuscript.

REFERENCES

Aini, H., Sabani, S., Solikin, S., Manik, K., Harahap, R. A., Putri, R. I. P., & Sibagariang, S. A. B. (2025). Analysis of Studentsâ€™ Cognitive Conceptual Understanding on Temperature and Heat Material Using a Four-Tier Multiple Choice Diagnostic Test. Lensa: Jurnal Kependidikan Fisika, 13(1), 73-83. https://doi.org/10.33394/j-lkf.v13i1.15822

Annisa, M., Budimansyah, D., Hidayat, M., Winarti, A., & Abrori, F. M. (2024). What can we learn from one-to-one trials in Instructional design? A case from module development. Research and Development in Education (RaDEn), 4(2), 816-826. https://doi.org/10.22219/raden.v4i2.35763

Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. https://doi.org/10.1007/978-0-387-09506-6

Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. International Journal of Science Education, 32(7), 939–961. https://doi.org/10.1080/09500690902890130

Çelikkanlı, N. Ö., & Kızılcık, H. (2022). A review of studies about four-tier diagnostic tests in physics education. Journal of Turkish Science Education, 19(4), 1220–1236. https://doi.org/10.36681/tused.2022.175

Festiana, I., Herlina, K., Kurniasari, L. S., & Haryanti, S. S. (2019). Damping Harmonic Oscillator (DHO) for learning media in the topic damping harmonic motion. In Journal of Physics: Conference Series (Vol. 1157, No. 3, p. 032062). IOP Publishing. https://doi.org/10.1088/1742-6596/1157/3/032062

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to design and evaluate research in education (8th ed.). McGraw-Hill.

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. Eurasia Journal of Mathematics, Science and Technology Education, 11(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Hasan, S., Bagayoko, D., & Kelley, E. L. (2017). Misconceptions and the certainty of response index (CRI). Physics Education, 34(5), 294–299. https://doi.org/10.1088/0031-9120/34/5/304

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign.

Ivanjek, L., Susac, A., Planinic, M., Andrasevic, A., & Milin-Sipus, Z. (2016). Student reasoning about graphs in different contexts. Physical Review Physics Education Research, 12(1), 010106. https://doi.org/10.1103/PhysRevPhysEducRes.12.010106

Jang, H., Kim, Y., & Reeve, J. (2019). Longitudinal test of self-determination theory's motivation mediation model in a naturally occurring classroom context. Journal of Educational Psychology, 111(5), 838–852.

Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. Research in Science & Technological Education, 35(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094

Kanıawatı, I., Fratıwı, N. J., Danawan, A., Suyana, I., Samsudın, A., & Suhendı, E. (2019). Analyzing students' misconceptions about newton's lws through four-tier newtonian test (FTNT). Journal of Turkish Science Education, 16(1), 110-122.

Kiray, S. A., & Simsek, S. (2021). Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test. International Journal of Science and Mathematics Education, 19(5), 935-955. https://doi.org/10.1007/s10763-020-10087-5

Lynn, M. R. (1986). Determination and quantification of content validity. Nursing Research, 35(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y., & Mai, T. (2020). Student misconceptions and errors in physics and mathematics: Exploring data from TIMSS and TIMSS Advanced. Springer Open. https://doi.org/10.1007/978-3-030-30188-0

Peşman, H., & Eryılmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. The Journal of educational research, 103(3), 208-222. https://doi.org/10.1080/00220670903383002

Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in nursing & health, 29(5), 489-497. https://doi.org/10.1002/nur.20147

Putranta, H., & Afifah, F. (2025). Development of the Four-Tier Diagnostic Test to Identify Student Misconceptions in the Static Fluids Chapter. Journal on Efficiency and Responsibility in Education and Science, 18(4), 268-281. https://doi.org/10.7160/eriesj.2025.180403

Richey, R. C., & Klein, J. D. (2014). Design and development research: Methods, strategies, and issues. Routledge. https://doi.org/10.4324/9780203826034

Rusilowati, A. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. Unnes Science Education Journal.

Rusilowati, A., Darsono, T., Sekarningtias, F. O., & Fariyani, Q. (2024). Development of a four-tier e-diagnostic test on the topic of momentum to measure and reduce student misconception. Jurnal Pendidikan Fisika Indonesia, 20(2), 173-187. https://doi.org/10.15294/jpfi.v20i2.1297

Sargazi Moghadam, T., Darejeh, A., Delaramifar, M., & Mashayekh, S. (2024). Toward an artificial intelligence-based decision framework for developing adaptive e-learning systems to impact learners' emotions. Interactive Learning Environments, 32(7), 3665-3685. https://doi.org/10.1080/10494820.2023.2188398

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. Journal of Computer Assisted Learning, 33(1), 1–19. https://doi.org/10.1111/jcal.12172

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. Adaptive technologies for training and education, 7(27), 1-35. https://doi.org/10.1017/CBO9781139049580.004

Soeharto, S., Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. Jurnal Pendidikan IPA Indonesia, 8(2), 247-266. https://doi.org/10.15294/jpii.v8i2.18649

Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L. J., Mukama, E., Warusavitarana, A., Lokuge Dona, K., Eichhorn, K., Fluck, A., Huang, R., Bridges, S., Lu, J., Ren, Y., Gui, X., Deneen, C. C., San Diego, J., & Gibson, D. C. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society, 19*(3), 58–71. http://www.jstor.org/stable/jeductechsoci.19.3.58.

Tumanggor, A. M. R., Supahar, S., Ringo, E. S., & Harliadi, M. D. (2020). Detecting students' misconception in simple harmonic motion concepts using four-tier diagnostic test instruments. Jurnal Ilmiah Pendidikan Fisika Al-Biruni, 9(1), 21-31. https://doi.org/10.24042/jipfalbiruni.v9i1.4571

Van der Linden, W. J., & Glas, C. A. (Eds.). (2010). Elements of adaptive testing (Vol. 10, pp. 978-0). New York: Springer. https://doi.org/10.1007/978-0-387-85461-8

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International journal of educational technology in higher education, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0