MJP Journal of Education and Teaching Learning
Vol.4, No.2 (2026): 1290 – 1303
https://jurnalilmiah.co.id/index.php/MJPJETL



Mechanisms of Quizizz-Mediated Formative Assessment in Islamic Religious Education: Evidence, Feedback, and Pedagogical Follow-Up

[1)]Khoiril Yahya, [2)]Muhammad Abu Kholil

[1,2)] Khozinatul Ulum Islamic Institute, Blora, Indonesia

***Corresponding Author:** khoirilyahya@iaiku.ac.id

Article Info

Keywords:
Quizizz; formative assessment; Islamic Religious Education; evidence of learning; feedback; pedagogical follow-up.

ABSTRACT

This study aims to analyze the mechanisms of Quizizz-based formative assessment in Islamic Religious Education (PAI) for 12th-grade students at SMK PSM Randublatung, specifically how evidence of student understanding is obtained, interpreted, used for feedback, and followed up on pedagogically. The study employs a qualitative approach with a case study design. Data were collected from three Quizizz assessment sessions comprising 24 items and 720 student response records, four observation sessions, 17 interview sessions with 10 informants, 90 follow-up assignment responses, and instructional planning documents. The analysis focused on the relationship between digital response patterns, teacher interpretations, follow-up checks, feedback, and pedagogical actions. The results indicate that Quizizz provides preliminary evidence at the item level; however, correct or incorrect answers alone are insufficient to explain students' reasoning and learning needs. Teachers interpret response patterns by comparing question content, diagnostic questions, dialogues, and student explanations. These interpretations then form the basis for providing feedback and follow-up in the form of reinforcement, clarification, practice, and instructional adjustments. This study positions Quizizz as part of an evidence-based formative assessment cycle, rather than merely a scoring tool.

Article Information

Keywords:
Pendidikan Agama Islam; bukti pembelajaran; umpan balik; tindak lanjut pedagogis.

ABSTRACK

Abstrak ini bertujuan untuk menganalisis mekanisme asesmen formatif berbasis Quizizz pada mata pelajaran Pendidikan Agama Islam (PAI) bagi siswa kelas XII di SMK PSM Randublatung, khususnya bagaimana bukti pemahaman siswa diperoleh, diinterpretasikan, digunakan untuk umpan balik, serta ditindaklanjuti secara pedagogis. Penelitian ini menggunakan pendekatan kualitatif dengan desain studi kasus. Data dikumpulkan dari tiga sesi asesmen Quizizz yang mencakup 24 butir soal dan 720 catatan respons siswa, empat sesi observasi, 17 sesi wawancara dengan 10 informan, 90 respons penugasan tindak lanjut, serta dokumen perencanaan pembelajaran. Analisis berfokus pada hubungan antara pola respons digital, interpretasi guru, pemeriksaan tindak lanjut, umpan balik, dan tindakan pedagogis. Hasil penelitian menunjukkan bahwa Quizizz menyediakan bukti awal pada tingkat butir soal; namun, jawaban benar atau salah saja tidak cukup untuk menjelaskan penalaran dan kebutuhan belajar siswa. Guru menginterpretasikan pola respons dengan membandingkan isi pertanyaan, pertanyaan diagnostik, dialog, dan penjelasan siswa. Interpretasi tersebut kemudian menjadi dasar untuk memberikan umpan balik dan tindak lanjut berupa penguatan, klarifikasi, latihan, dan penyesuaian pembelajaran. Penelitian ini memposisikan Quizizz sebagai bagian dari siklus asesmen formatif berbasis bukti, bukan sekadar alat penilaian skor.

Article History

Received: 15/08/2026
Revised : 10/09/2026
Accepted: 21/10/2026

✉ ***Corresponding Author:*** (1) Khoiril Yahya, (2) Department of Islamic Education, (3) Khozinatul Ulum Islamic Institute, (4) Blora, Central Java, 58214, Indonesia, (5) Email: khoirilyahya@iaiku.ac.id

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
Copyright © 2026 - [1)] Khoiril Yahya, [2)] Muhammad Abu Kholil

1. Introduction

Digital transformation in education has shifted assessment practices from activities primarily designed to measure learning outcomes at the end of instruction toward a process that is increasingly integrated with learning activities, allowing evidence of learning to be collected more quickly, repeatedly, and interactively [1]. In this context, gamification has emerged as a strategy that integrates game elements into learning and assessment activities to enhance student engagement and motivation, particularly in digital learning environments [2]. The benefits of gamification cannot be viewed as an automatic consequence of the presence of game features, as its effectiveness is heavily influenced by the alignment of the design with student characteristics, learning objectives, and the context of its implementation [3]. A meta-analysis of 35 interventions even showed that gamification does have a positive effect on intrinsic motivation, but the overall effect size is relatively small and varies across the dimensions of learner autonomy, relatedness, and competence [4]. This finding suggests that the pedagogical value of gamification is determined more by how game elements are designed and integrated into the learning and assessment processes than simply by the use of the technology itself [5].

Quizizz can be understood not merely as a game-based quiz app, but as a digital assessment environment that integrates assessment activities with gamification elements and provides a space for students to respond to tasks interactively [6], [7]. In the context of formative assessment, this platform is capable of generating more detailed information than just a final score, as Quizizz reports can include data on attendance, accuracy rates, points and scores, the number of attempts, completion time, and the specific questions answered incorrectly by students. The availability of this information enables teachers to quickly gather evidence of learning while identifying response patterns, mastery levels, and areas of the material that require further attention; meanwhile, the real-time feedback system also allows students to recognize their weaknesses and review their mistakes after completing the assessment [8]. This potential makes Quizizz relevant not only as a gamification tool to boost motivation, satisfaction, or engagement, but also as an instrument that can be integrated into more structured assessment practices, although the results of its use may vary depending on the learning context and student characteristics[9]. Quizizz is beginning to take on a more substantive role as a gamified formative assessment environment one that not only facilitates quiz activities but also generates digital evidence that can potentially be used to support the assessment process and subsequent learning decisions [8].

To date, research on gamification in education has been dominated by studies examining the relationship between the use of game elements and students' motivation, engagement, and academic achievement. A meta-analysis of 22 experimental studies indicates that gamification generally has a moderate positive effect on academic performance, although the magnitude of the effect varies depending on the learning context, educational level, subject matter, and game elements used [10]. A similar pattern is observed regarding motivation, where gamification has been shown to increase intrinsic motivation as well as perceptions of autonomy and relatedness, but has only a minimal effect on competence; thus, its psychological benefits cannot be assumed to be uniform across all dimensions of learning [4]. Experimental evidence also shows inconsistent results; the use of leaderboards in online formative quizzes does increase the number of quizzes completed, but does not result in significant differences in learning outcomes or student engagement compared to non-gamified assessments [11]. Conversely, another study on leaderboard-based formative assessments found a positive association with student achievement and engagement, although these benefits may not sustain attention in the long term [12]. This inconsistency is further reinforced by a systematic review indicating that engagement in gamification is multidimensional and cannot be reduced solely to student participation or emotional responses [13]. Current evidence no longer sufficiently supports the simple question of whether gamification is "effective," but rather directs attention toward the conditions, processes, and pedagogical mechanisms that determine how this technology yields learning benefits.

Although research on gamification and Quizizz continues to evolve, most studies remain focused on outcomes following the use of the technology such as motivation, engagement, perceptions, learning experiences, and academic performance while the pedagogical processes that link assessment data to learning interventions have not received commensurate attention [3], [8]. This focus raises a significant issue because the ability of digital systems to record student responses, scores, accuracy rates, or error patterns does not, in and of itself, explain how that information is interpreted and utilized in the learning process. A systematic review of game-based assessment indicates that game-based environments do indeed enable assessment activities while capturing student performance data; however, the assessment design must remain aligned with educational objectives, game characteristics, and the competencies to be measured [14]. This issue becomes increasingly important because the validity of game-based assessment is determined not only by the technology's ability to generate data, but also by the alignment between the game design, the tasks assigned, and the evidence required to represent the constructs or competencies being assessed [15]. The availability of digital assessment data is not synonymous with its formative use; the same information may remain merely as a score, be presented as a correction, be used to identify learning difficulties, or be translated into feedback, re-teaching, remediation, or learning adjustments. In the context of Quizizz, this gap shifts the research focus from questions about the impact of gamification toward a more fundamental issue: how technology-generated evidence of learning is interpreted and transformed into pedagogical actions.

Within the framework of formative assessment, the mere presence of scores, digital reports, or automated feedback is not sufficient to make an assessment activity formative; the formative function emerges when evidence of learning is interpreted and used to determine steps that can improve the subsequent learning process [16]. Empirical evidence shows that formative assessment and feedback in general can support learning outcomes, but the strength of their impact depends on their form and method of implementation; in fact, Karaman's meta-analysis found that computer-initiated feedback produces a smaller effect compared to student-initiated, mixed, or adult-initiated feedback [17], [18]. Therefore, the focus on feedback needs to shift from merely "providing" information to the process of how that information is received, interpreted, and used by students to make subsequent learning decisions [19]. In digital assessment environments, a similar issue arises when

systems can provide real-time performance data, but that data is not always easy to interpret or can be directly translated into relevant pedagogical actions by teachers [16]. A systematic review of the use of technology in formative assessment highlights three key prerequisites: conceptual clarity regarding formative assessment, alignment between digital tools and pedagogical practices, and teachers' data literacy to interpret information and translate it into actions that support learning [20]. Thus, technology-based formative assessment is best understood as a continuous mechanism that links the collection of learning evidence, the interpretation of that evidence, the provision of feedback, pedagogical decision-making, and subsequent learning responses rather than merely as a process of generating scores or providing automated corrections.

The integration of Quizizz into Islamic Religious Education has evolved beyond its function as a quiz tool, with previous research highlighting its diverse applications ranging from enhancing motivation and student engagement to assessing learning outcomes and developing digital learning resources. The use of Quizizz gamification in Islamic Religious Education (IRE) has been reported to create a more interactive learning experience while supporting students' motivation to engage in the learning process [21]. Quantitative studies also indicate that the use of Quizizz gamification, combined with student engagement, has a significant relationship with IRE learning outcomes, demonstrating that academic performance is a prominent outcome in the existing literature [22]. Quizizz has been used as a tool to support PAI learning and assessment, with findings highlighting improvements in motivation, engagement, and learning outcomes, while also identifying supporting factors and barriers to its implementation [23]. The implementation of Quizizz in Islamic education has also shown a high level of acceptance and has been deemed suitable as an interactive evaluation tool that supports student engagement [24]. Research then shifted toward the development of Quizizz-based PAI media through a systematic approach, with results showing increased engagement, participation, and academic achievement after the media was implemented [25]. Collectively, these findings indicate that research on Quizizz in Islamic Education has developed significantly and consistently focuses on motivation, student activity, engagement, feasibility of implementation, and learning outcomes; therefore, the novelty of future research is no longer sufficient if it is based solely on the use of Quizizz in the context of Islamic Education, but must be directed toward pedagogical processes that have not yet been thoroughly explored.

Recent studies indicate that the use of Quizizz in Islamic Education has extended to the context of vocational high schools, so the Islamic Education vocational high school–Quizizz context can no longer be considered a research gap; existing studies have revealed a generally positive student reception, although some students still question the depth of spiritual content within the gamified format [26]. Experimental research in similar contexts has also examined the impact of Quizizz on motivation to learn PAI, so the motivational aspect is not an unexplored area either [27]. In fact, Quizizz has been explicitly positioned as a gamified formative assessment tool in madrasah settings, with findings that primarily highlight increased motivation, engagement, and learning outcomes following its implementation [28].

Other PAI literature has also discussed Quizizz as an adaptive digital assessment tool that offers higher participation and immediate feedback, but the discussion remains centered on the technology's functions and potential as an assessment tool [29]. Thus, the more substantive gap does not lie in the presence of Quizizz in PAI formative assessment, but rather in the lack of adequate explanation regarding how the learning evidence generated by the platform is interpreted by teachers, transformed into feedback, and translated into pedagogical follow-up actions; this issue aligns with findings that research on technology-assisted formative assessment still tends to focus more on testing the effectiveness of the tools than on documenting the process of utilizing assessment information in teaching practice [20].

The use of Quizizz in Islamic Education has shown potential in supporting motivation, engagement, participation, and learning outcomes, including in the context of vocational education [26], [27]. Even when used as formative assessment, the primary focus tends to remain on effectiveness, scores, and student responses, while the process of how assessment data is interpreted and translated into pedagogical actions has not been adequately explained [28], [29]. In fact, digital data does not automatically constitute formative evidence; its formative value emerges when the information is interpreted, used to provide feedback, and acted upon in the learning process [20]. Based on this gap, this study shifts the focus from the effectiveness of the technology to the mechanisms of its use by examining the collection of learning evidence, the interpretation of results, the provision of feedback, and pedagogical follow-up. Its main contribution is to formulate a conceptual model of the mechanisms of Quizizz-based formative assessment in Islamic Religious Education (IRE) learning, so that Quizizz is understood not merely as digitized testing, but as part of an evidence-based learning decision-making cycle.

This study aims to analyze the mechanisms of Quizizz-based formative assessment in 12th-grade Islamic Religious Education (PAI) at SMK PSM Randublatung by examining how evidence of student understanding is obtained, interpreted, used as the basis for providing feedback, and followed up through pedagogical decisions. Based on these objectives, this study seeks to answer three research questions, namely: (1) How is Quizizz used to obtain evidence of student understanding in formative assessment for 12th-grade Islamic Religious Education (PAI) at SMK PSM Randublatung?; (2) How are the assessment results generated by Quizizz interpreted and used by teachers in providing feedback to students?; and (3) How does the Quizizz-based formative assessment mechanism take shape through the relationship between the collection of learning evidence, the interpretation of results, the provision of feedback, and pedagogical follow-up in Islamic Religious Education (PAI) instruction? Theoretically, this study is expected to expand the body of research on digital formative assessment in Islamic Religious Education by formulating a conceptual model that explains the transformation of assessment data into evidence of learning, feedback, and pedagogical actions. Practically, this study provides a framework for Islamic Religious Education (IRE) teachers to utilize Quizizz not only as a scoring tool but as part of a formative assessment cycle that supports the diagnosis of understanding, learning improvement, and evidence-based decision-making.

2. Methods

2.1 Research Design

This study employs a qualitative case study design to examine the mechanisms of implementing Quizizz-based formative assessment in Islamic Religious Education (PAI) instruction. The case study approach was chosen because the research focuses on processes occurring within a real-world learning context, specifically how evidence from digital assessments is obtained, read, and interpreted by teachers; communicated through feedback; and followed up upon in the learning process. The research case is limited to the implementation of formative assessment in 12th-grade PAI classes at SMK PSM Randublatung. Within this case, three formative assessment episodes were analyzed sequentially: E01: Ikhtiyar and Tawakkal, E02: Patience and Corrective Actions, and E03: Qadar and Responsibility. These three episodes are not treated as three separate cases but as analytical episodes within the same case. The primary unit of analysis in this study is not Quizizz as a technological platform but rather the formative assessment episodes that link student responses, teacher interpretations, the provision of feedback, and pedagogical follow-up. Thus, the study is not aimed at measuring the effectiveness of Quizizz or proving that the use of Quizizz leads to improved learning outcomes. The focus of the study is to reconstruct the mechanisms that explain how digital assessment data can be transformed into pedagogical information and subsequently used to inform instructional decisions.

2.2 Research Location and Participants

This study was conducted in a Grade 12 Islamic Religious Education (PAI) class at SMK PSM Randublatung, Central Java. There were 32 research participants, consisting of 30 Grade 12 students, one PAI teacher, and one school principal. Of these participants, interviews were conducted with 10 informants: one school principal (P), one PAI teacher (T), and eight 12th-grade students (S01, S04, S07, S10, S14, S18, S23, and S28). The interviews were conducted in 17 individual sessions with a total duration of approximately 421 minutes. Four students were interviewed more than once to explore their interpretations, responses, and experiences regarding more than one assessment episode.

Eight students were purposively selected to represent a variety of response patterns and learning trajectories throughout the series of formative assessments. This selection took into account differences in students' responses on the Quizizz assessments, how they explained or interpreted feedback, and the responses demonstrated in follow-up tasks across three learning episodes. These variations included students who demonstrated adequate responses on follow-up tasks, students who showed changes in their responses across tasks, students who tended to provide partial justifications, and students who still struggled to explain concepts after receiving feedback. By selecting these diverse cases, the study was able to explore the process of using assessment evidence and feedback without assuming that all students exhibit the same response patterns or changes in understanding. Four students were interviewed more than once to deepen the exploration of their interpretations, experiences, and responses across several assessment episodes.

2.3 Data Sources and Data Collection Methods

The research data were obtained from several complementary sources, namely Quizizz response data, classroom observations, interviews, student follow-up assignments, and lesson plans. The digital assessment data were derived from three sessions of Quizizz use. Each session consisted of eight multiple-choice questions, so the entire assessment comprised 24 items and yielded 720 student response records from the 30 students who participated in all three sessions. The test items were collaboratively developed by an Islamic Religious Education (PAI) teacher and the researcher, then reviewed before being used in the classroom. All three assessments were conducted in real time in the classroom (*live mode*) with students working individually. The time limits on the questions and the leaderboard were disabled. The duration of each quiz was recorded at approximately 12, 12, and 13 minutes. The three assessment episodes were used to identify patterns in student responses to the material studied. Since each episode contained different items and focused on different material, differences in the distribution of answers across episodes were not directly interpreted as an increase or decrease in student understanding. Quizizz assessment results were exported in Excel/CSV format and processed using a spreadsheet application. The processing involved matching students' answers to the answer key and calculating the distribution of responses at the item level. The processed data is used to identify patterns of correct and incorrect answers as well as the status of responses recorded in the dataset. These patterns are then analyzed alongside observational data, interviews, follow-up assignments, and lesson plans to understand how assessment results are interpreted and used in pedagogical decision-making.

The observations were conducted over four learning sessions, each lasting approximately 90 minutes, for a total duration of 360 minutes. The observations focused on the administration of assessments, how teachers observed and interpreted student responses, interactions following the assessments, the provision of feedback, and changes or adjustments to learning activities after assessment information was obtained. Interview data were collected through 17 individual sessions with 10 informants, with a total duration of approximately 421 minutes. The interviews were used to explore how teachers interpreted assessment results, the considerations underlying their pedagogical decisions, as well as students' experiences and responses to the assessments, feedback, and follow-up learning activities. Supplementary data consisted of 90 student responses to follow-up assignments given after three assessment episodes. These responses were analyzed to trace how evidence from previous assessments related to students' responses in subsequent learning activities, while also examining the extent to which teachers' interpretations of students' difficulties were supported by evidence from subsequent learning. Instructional planning documents were used as supporting sources to trace the relationship between assessment results, instructional decisions, and planned follow-up activities.

2.4 Data Analysis Techniques

Data analysis was conducted iteratively by treating the three assessment episodes as a series of processes within a single case. The analysis employed *a hybrid deductive–inductive coding* approach, which integrated the conceptual framework of

formative assessment with patterns identified from the research data. Deductively, the analysis drew upon the literature regarding the use of data from various sources in digital formative assessment [16], formative assessment practices and feedback [17], empirical evidence on formative assessment practices [18], students' processing of feedback [19], and the prerequisites for teachers' use of technology in formative assessment [20]. This literature serves as a conceptual foundation for examining the relationship between the acquisition of learning evidence, teacher interpretation, the provision of feedback, and pedagogical follow-up—not as a standardized coding tool applied mechanically.

Based on this framework, the a priori categories include assessment evidence, teacher interpretations or diagnoses, feedback, and pedagogical follow-up. Inductive coding was then used to identify and elaborate on patterns emerging from Quizizz data, observations, interviews, and follow-up assignments, including the distribution of student responses, teachers' attention to specific responses, diagnostic questions, differences between mastery of definitions and application of concepts, instructional adjustments, and difficulties that persisted after follow-up. Both approaches were integrated by using theoretical categories as a framework for analysis, while empirical patterns were used to clarify how these categories manifested in Islamic Education (PAI) teaching practice. The analysis did not rigidly separate data sources but brought them together within each episode to trace the temporal and pedagogical relationships between student responses, teacher interpretations, feedback, and subsequent instructional actions.

Next, a cross-episode analysis was conducted to identify recurring mechanisms as well as variations that emerged across different learning materials or contexts. This comparison was used to distinguish consistent patterns from processes specific to certain episodes. The analysis also included negative cases and deviations, such as students who continued to struggle after receiving feedback and completing follow-up tasks. These cases were used to test the interpretive limits of the identified patterns, ensuring that the formulated mechanisms did not assume that every error in Quizizz automatically resulted in an accurate diagnosis, effective feedback, or a change in understanding. Thus, the analysis aimed to explain how assessment evidence is interpreted and used in pedagogical decision-making, while also identifying conditions under which this process did not produce the expected learning responses.

2.5 Data Validity and Research Ethics

Data validity was ensured through triangulation of methods and sources by comparing data from Quizizz responses, classroom observations, interviews, student follow-up assignments, and lesson plans. The triangulation process was conducted at the episode and student-case levels by examining the consistency and discrepancies between responses recorded in the assessment, students' explanations of their responses, teachers' actions following the assessment, and students' responses during follow-up activities. When multiple sources provided consistent information, this consistency was used to strengthen the interpretation of the formative assessment process. Conversely, when discrepancies were found, the data were not immediately harmonized or determined based on a majority vote. The researcher reexamines the context of the test items, statements from interviews, observation notes, and follow-up evidence to consider possible alternative explanations. Through this procedure, discrepancies between sources are treated as part of the findings that need to be analyzed, rather than merely as data inconsistencies.

For example, the contrast between S04 and S10 highlights the importance of distinguishing between the accuracy of an answer and the underlying process. In S04, a correct answer must be interpreted alongside the note that the student felt uncertain; thus, the accuracy of the answer is not automatically interpreted as solid understanding. In S10, the incorrect answer must be considered alongside the note regarding an oversight in reading the question, so that the error is not immediately concluded to be a misconception. Both cases demonstrate that Quizizz responses provide evidence of performance outcomes but are not always sufficient to explain the reasoning behind students' answers. Therefore, interpretations regarding confidence, conceptual understanding, or the causes of errors are made by considering other available data sources. If evidence from multiple sources is insufficient to support a single explanation, the interpretation is kept limited and is not forced into a definitive conclusion. Analysis of cases that deviate from general patterns is also used to test the limits of explanations regarding the relationship between assessment evidence, teacher interpretation, feedback, and pedagogical follow-up.

The reconstruction of episodes, coding decisions, and considerations in interpreting both the similarities and differences in the data were documented as part of the study's analytical trail. Source triangulation included the perspectives of teachers, students, and informants responsible for the curriculum, while methodological triangulation was conducted by combining data from digital assessments, observations, interviews, follow-up assignments, and learning documents. The findings are presented as a contextual explanation of the mechanisms of formative assessment in the case studied, not as a statistical generalization regarding all Islamic Religious Education instruction or all uses of Quizizz. The research was conducted in accordance with the access procedures in effect at the school. The identities of the school, teachers, students, and informants are protected in the presentation of the findings, while digital assessment data are treated as restricted research data and used solely for the purposes of the research analysis.

3. Results and Discussion

3.1. Gathering Learning Evidence through Quizizz in Formative Assessment of Islamic Education

In this study, Quizizz was used in three formative assessment sessions for 12th-grade Islamic Education (PAI) classes, held on September 2, 9, and 16, 2026. All three sessions explored faith in qada and qadar through different topics: the relationship between human effort (ikhtiar) and trust in God (tawakal), the attitude of patience when facing failure, and responsibility for one's actions. The questions in each session were designed to assess students' ability to recognize concepts and apply them to evaluate situations relevant to their own experiences.

To gather more detailed evidence, each session combined questions about the meaning of concepts with case-based questions requiring their application. In the first session, students were asked to explain the meaning of tawakal, then evaluate the statement that a person does not need to study because exam results are predetermined. The second session

linked the concept of patience to a case involving a student who stopped improving their practice results after experiencing failure. In the third session, questions about "tanggung jawab" (responsibility) were followed by a case involving a group member who used the concept of "takdir" (fate) to justify their negligence. This structure allowed for a comparison between responses to definition-based questions and those to application-based questions. The differences between the two provided initial clues about which aspects of understanding needed further exploration, but were not sufficient to determine why students chose a particular answer.

After each quiz, the Quizizz response report in this study presented students' choices for each item, the number of students selecting each answer option, and whether the answer was marked correct or incorrect. This item-level information allowed researchers to identify questions where many students selected the same distractor, but it did not explain the reasoning behind those choices. This limitation also applied to correct answers. S04 explained that:

"In the report, it looked correct. My doubts only arose when the teacher asked why option A was incorrect. So the correct answer doesn't tell the whole story of what I was thinking." (S04)

This statement indicates that response records should be read as initial clues about students' understanding, not as a complete picture of their thought processes. The distribution of answers for the item that was the focus of each session is presented in Table 1.

Table 1 Distribution of answers on the focus items of the PAI formative assessment using QUIZIZZ

Episode and Concept	Item Code	Question Prompt	Correct Answers	Students Answered Correctly	Most frequently selected distractor
E01: Tawakal	Q1-06	Recognizing the meaning of tawakal accompanied by effort	A	23/30 (76.7%)	B: 3 students
E01: Effort and tawakal	Q1-04	Assessing the attitude of not studying on the grounds that the outcome has already been determined	B	12/30 (40.0%)	A: 14 students consider this attitude to be tawakal
E02: Patience	Q2-01	Recognizing the meaning of patience when facing learning difficulties	C	24/30 (80.0%)	B: 3 students
E02: Patience and Improvement	Q2-04	Evaluating the decision to stop correcting failed practice results	C	13/30 (43.3%)	A: 12 students view giving up as an expression of patience
E03: Responsibility	Q3-01	Recognizing actions that demonstrate responsibility in group tasks	A	25/30 (83.3%)	B and C: 2 students each
E03: Fate and Responsibility	Q3-04	Assessing the Use of Fate as a Justification for Negligence	D	11/30 (36.7%)	A: 13 students believe that fate absolves one of responsibility

Table 1 shows that in all three episodes, the number of correct answers on concept-recognition items (23–25 students) was higher than on case-based application items (11–13 students). This pattern indicates that concept and case items yield different response distributions, but it is not sufficient to conclude that there is a direct difficulty in concept transfer. These differences may also be related to the complexity of the language or the length of the questions, the position of the items in the quiz, and the limitations of multiple-choice questions in capturing students' contextual reasoning. Notes regarding Student S10, who skipped part of the question, reinforce the need to consider factors related to question comprehension, not just conceptual mastery. Therefore, these findings are treated as preliminary indications that need to be interpreted in conjunction with data from interviews, observations, and follow-up tasks, rather than as standalone evidence of difficulties in concept transfer.

In the first episode, Q1-04 asked students to evaluate Rafi's statement that he did not study because he believed the exam results had already been determined. Fourteen students chose option A, which stated that Rafi had already placed his trust in God, while twelve chose option B, which distinguished between accepting the results and the obligation to make an effort. This finding becomes more meaningful when read in conjunction with Q1-06, where 23 students recognized a definition of tawakal that includes responsible effort. This distribution draws attention to possible difficulties in applying the concept in specific situations, but does not prove that the fourteen students who chose A had identical reasons. To determine those reasons, the on-screen choices need to be linked to the students' explanations of the case content and the basis for their answers.

In the first episode, S04's correct answer which was still accompanied by uncertainty indicated that the on-screen choice did not fully reflect all of his considerations. A different situation was observed with S10: he selected the incorrect answer after skipping a portion of the sentence in the case, then identified the source of his error when asked to read and explain the question. Thus, the responses of S04 and S10 have different answer statuses but both reveal the limitations of quiz reports in

uncovering students' reasoning. Teacher T's remarks reinforce these limitations. According to T, the clustering of answers around certain options is important to note, but this pattern does not explain each student's individual reasoning. Therefore, the distribution of options serves as a starting point for assessing understanding, not as a basis for determining the cause of errors.

"The screen didn't show the part of the sentence I missed. I only realized the source of my mistake when the teacher asked for my reasoning and I read the case aloud." (S10)

"I don't start with the students' rankings. I look at which questions have answers clustered around the same choice. On Rafi's question, fourteen students chose that he had already put his trust in God, while twelve chose the answer that distinguishes trust in God from giving up effort. That pattern needs to be discussed, but it doesn't yet explain each student's reasoning." (T)

In the second episode, Q2-04 presents a situation where Arif stopped improving his practical project after failing and described his decision as an act of patience. Thirteen out of 30 students chose option C, which calls for evaluation and improvement of the effort, while twelve chose option A, which views stopping the effort as an appropriate acceptance of the situation. In Q2-01, 24 students chose a definition of patience that included self-control and the continuation of effort. This discrepancy indicates that responses to definition questions are not sufficient to explain how students evaluate an action. S23 recognized a similar limitation in his own answer: choosing an alternative considered correct still needs to be accompanied by a rationale so that his understanding can be further examined.

"Not yet. With Arif, I can see the improvement. In other cases linking effort to certain outcomes, I might still be confused. I need to explain my reasoning, not just pick a letter." (S23)

In the third episode, Q3-04 tested responses to Bima, who did not complete a group assignment and used fate to justify his negligence. Thirteen students chose A, which absolves Bima of responsibility, while eleven chose D, which calls for acknowledgment and correction. On Q3-01, 25 students selected the appropriate definition of responsibility. The difference in responses to these two items indicates an area of understanding that needs to be explored through students' reasoning, without assuming that the questions are of equal difficulty. In a follow-up interview, S18 also distinguished between knowing the option explained by the teacher and being able to accept and explain the reasoning behind it. Thus, simply selecting the correct answer is not sufficient to conclude that understanding is solid.

"I know the teacher explained option D in Bima's case. Knowing option D is not the same as accepting it or being able to explain all the reasons behind it in my own words." (S18)

Overall, the three item pairs demonstrate how Quizizz provides preliminary evidence regarding which parts of the material require further investigation. The item-level reports show the choices made by each student, the number of students who selected a particular distractor, and differences in response patterns between the introductory questions and the case-based questions. However, this evidence does not yet explain whether incorrect choices stem from a lack of conceptual understanding, misreading the question, or other considerations; even correct choices are not always accompanied by sound reasoning. Thus, the answer to the first research question is that Quizizz helps obtain detailed response traces, while students' explanations are necessary to determine the pedagogical significance of those traces.

The pedagogical value of Quizizz does not primarily lie in the automation of scoring, but in its ability to generate response patterns that can serve as the focus of teachers' attention and diagnostic reasoning. This interpretation aligns with Black and Wiliam's framework for formative assessment, which positions evidence of learning as information that must be interpreted before it can be used to determine next steps in learning [48]. In all three episodes, the value of the Quizizz report lies in the visibility of item-by-item responses, not in the score as a final conclusion about understanding. Kannan and Zapata-Rivera discuss the importance of linking information from multiple sources in digital assessment to support formative learning [16]. Therefore, item-by-item answers, question content, and student explanations need to be analyzed together. This relationship explains why "correct" or "incorrect" labels are useful for identifying areas that need review, but are insufficient for determining the rationale behind students' responses.

This explanation also highlights the importance of alignment between the concepts to be assessed and the tasks used to elicit evidence. Bijl and colleagues' study on game-based assessment emphasizes the relationship between task design, recorded evidence, and the competencies being assessed [15]. In the context of Islamic Education in this scenario, definition questions provide insight into the recognition of terms, while case-based questions ask students to evaluate actions using those concepts. Both remain responses to classroom tasks. Neither the answer to a single item nor the combined results of all six items can prove students' internalization of values or religious behavior outside of the classroom. This limitation must be kept in mind when quiz reports are used in subsequent discussions.

3.2 Interpretation of Quizizz Results and Feedback Provided by the PAI Teacher

The results in the previous subsection show that Quizizz responses provide initial clues about answer patterns that can direct teachers' attention to specific parts of the material, but they are not sufficient to explain the reasons behind students' choices. The next stage in the formative assessment process is the teacher's interpretation of these response patterns. In this study, incorrect answers were not automatically treated as evidence of a lack of conceptual understanding. Teachers examined prominent response patterns through questions, case discussions, and students' oral explanations. This process combined Quizizz data with other learning information to build a more contextual interpretation, without assuming that the distribution of answers in and of itself proved the cause of errors or changes in students' understanding.

The teacher explained that the analysis of Quizizz results does not start with student rankings, but with the patterns of responses to specific questions:

"I don't start with the students' rankings. I look at which questions had answers clustered around the same choice. On the question about Rafi, fourteen students chose that he had already put his trust in God, while twelve chose the answer that distinguishes putting one's trust in God from giving up on one's efforts. That pattern needs to be discussed, but it doesn't yet explain each student's reasoning." (T)

This statement indicates that teachers use the distribution of responses as a starting point for further exploration, rather than

as a final diagnosis. Teachers observe the concentration of answers on specific response options to determine which parts of the material need further discussion. However, the meaning behind these patterns must still be examined through students' explanations. This is important because students who select the same response option do not necessarily share the same reasoning or level of understanding.

To illustrate the relationship between response patterns and the teachers' interpretation process, the results of the analysis of the three episodes are summarized in Table 2.

Table 2 Teachers' interpretations of students' response patterns in three formative assessment episodes

Episode	Preliminary Evidence from Quizizz	Response patterns observed	Focus of the teacher's interpretation	Follow-up Assessment
E01: Effort and Trust in God	Q1-04: 12 out of 30 students answered correctly; 14 students chose option A	Most responses focused on the option that views not studying because the outcome is predetermined as tawakal	Do students understand the relationship between tawakal and the obligation to make an effort, or do they merely recognize the term tawakal declaratively?	Questions about the reasons for choosing an answer and a review of Rafi's case
E02: Patience and Improvement	Q2-04: 13 out of 30 students answered correctly; 12 students chose A	Responses are divided between persevering after failure and viewing giving up as a form of patience	Do students understand patience as self-control that is accompanied by continued effort and improvement?	Discussion of Arif's case and a request for an explanation of the reasoning behind the answer
E03: Destiny and Responsibility	Q3-04: 11 out of 30 students answered correctly; 13 students chose A	Some students use fate as a basis for absolving the character of responsibility	Can students distinguish between a belief in destiny and the avoidance of responsibility for one's actions?	Discussion of the Bima case and explanation of the reasons for choosing an alternative

Table 2 shows that teachers' interpretations shifted from response patterns to diagnostic questions. In E01, for example, the difference between Q1-06 and Q1-04 indicates that the ability to recognize the meaning of tawakal is not necessarily accompanied by the ability to apply that concept to concrete situations. A total of 23 students were able to correctly answer the item identifying the meaning of tawakal, while only 12 students answered correctly when asked to evaluate Rafi's case. However, this difference should not be used to immediately conclude that the other 18 students do not understand the concept. Teachers need to understand the reasons behind their choices.

The need to examine these reasons is evident in S04's experience. Although the system marked his answer as correct, he still had doubts when asked to explain the basis for his choice.

"In my report, it looked correct. My doubts only arose when the teacher asked why option A was incorrect. So, the correct answer didn't convey everything I was thinking." (S04)

Statement S04 shows that the teacher's interpretation is not directed solely at students who gave incorrect answers. Even correct answers can be subject to further inquiry if there is doubt regarding the reasoning behind them. Thus, the correct or incorrect status serves as initial information, while the student's explanation provides additional insight into the quality of understanding underlying that response.

Case S10 illustrates a different aspect. S10 chose an incorrect answer after skipping a specific part of the sentence in the case. When the teacher asked the student to reread the question and explain their reasoning, S10 was able to identify the source of their error.

"The screen didn't show the part of the sentence I missed. I only realized the source of the error when the teacher asked for an explanation and I read the example aloud." (S10)

This case is important because it shows that a wrong answer is not always synonymous with a conceptual error. Quizizz's feedback displays the student's final selection, while the post-assessment dialogue allows the teacher to distinguish between errors in understanding and errors in reading or processing the information in the question.

In E02, the interpretation process is evident in the case of Arif, who stopped revising his practice results after experiencing failure and viewed that decision as an act of patience. A total of 13 students chose answers that emphasized evaluation and improvement of their efforts, while 12 students chose alternatives that viewed stopping their efforts as an appropriate form of acceptance. Compared to Q2-01, where 24 students were able to recognize the meaning of patience, this pattern indicates that there is an aspect of understanding that needs further examination. This examination does not stop at providing the correct answer. S23 shows that the issue relates to the ability to explain the reasoning behind the choice:

"Not yet. With Arif, I can see the improvement. In other cases that link effort to definite results, I might still be confused. I need to explain the reasoning, not just choose a letter." (S23)

This statement indicates that teachers need to know whether students are able to connect the concept of patience with concrete actions in specific situations. In other words, the focus of interpretation shifts from "which answer was chosen?" to "how does the student explain that choice?"

In E03, the same process emerges in the discussion of the

relationship between fate and responsibility. Q3-04 shows that 13 students chose alternative A, which absolves Bima of responsibility, while 11 students chose D. Meanwhile, 25 students were able to identify actions that demonstrate responsibility in Q3-01. This pattern prompts teachers to assess whether students understand the concept of responsibility when it is applied to situations involving a belief in fate.

S18 shows that knowing the answer the teacher has explained does not always mean being able to explain the conceptual reasoning independently:

"I know the teacher explained D in Bima's case. Knowing D is not the same as accepting it or being able to explain all the reasons behind it in one's own words." (S18)

This statement reinforces the finding that teachers' interpretations need to consider more than just changes in answer choices. Teachers need to assess whether students can provide coherent reasoning and connect concepts to the given situations. Thus, the diagnostic process in this study is best understood as a step-by-step examination of the meaning behind responses, rather than a simple classification between students who are correct and those who are incorrect.

Overall, the three episodes show a relatively consistent pattern of interpretation. Teachers used the Quizizz response distribution to identify parts of the material that required attention, then used questions and discussions to elicit explanations not available in the quiz reports. At this stage, digital data served primarily to reveal patterns in student responses, while the pedagogical meaning of those patterns was shaped through teacher-student interactions. This finding also explains why the same answers can lead to different interpretations once teachers obtain additional information from students.

This model clarifies the teacher's role in Quizizz-based formative assessment. Teachers are not merely recipients of digital reports but act as interpreters of assessment evidence, linking response data to the learning context and students' explanations. This role is crucial because assessment evidence does not automatically carry pedagogical meaning. Within Black and Wiliam's framework for formative assessment, evidence of learning must be interpreted before it can be used to determine the next steps in the learning process [32]. Thus, the findings of this study indicate that Quizizz's formative function is not inherent in the platform's features alone but emerges through teachers' interpretive work on the evidence generated by the platform.

These findings also reinforce research on digital assessment that positions teachers' interpretations as the link between data and pedagogical actions. Kannan and Zapata-Rivera [16]. demonstrate that digital assessment environments can generate data from various learning activities, but this data must be integrated and interpreted to support formative learning. In this study, this need was empirically evident when students' response patterns on Quizizz items were used not as a final diagnosis but as a signal to conduct further assessment through questions, case discussions, and student explanations. Thus, this study demonstrates that digital information acquires diagnostic meaning when placed within the context of pedagogical interaction.

These findings can be further explored in the systematic review by Børte et al. [20], which identifies teachers' data literacy as one of the key prerequisites for using technology in formative assessment. Data literacy in this context is not merely the ability to read a dashboard or understand scores, but the ability to examine, interpret, and translate assessment information into instructional knowledge and learning actions. The findings of this study provide empirical support for this proposition. Teachers do not treat the distribution of Quizizz answers as final information but rather explore the meaning behind response patterns by considering the content of the questions, the concentration of answer choices, students' explanations, and the context of the learning interaction. In other words, this study demonstrates that data literacy is not an abstract teacher competency but rather an interpretive practice that takes place within the assessment process.

At this point, the research findings also clarify the distinction between digital assessment data and pedagogically meaningful assessment evidence. Quizizz data indicates whether a response is correct or incorrect and reveals patterns in students' choices; however, this data alone cannot explain whether a student holds a misconception, misread the prompt, is unable to apply a concept, or simply selected an answer without a sound rationale. Therefore, the teacher's interpretation serves as a transformative process that converts student response data into pedagogically meaningful information. By examining response patterns, question content, and student explanations, teachers determine which aspects of learning require attention and identify learning needs that form the basis for subsequent actions. This process aligns with the literature, which indicates that the success of technology-assisted formative assessment depends more on how teachers use the available information than on the presence of the technology itself [20].

These findings also help address the research gaps identified in previous studies on Quizizz in Islamic Education. Previous research has primarily examined Quizizz in relation to motivation, student engagement, participation, acceptance, and learning outcomes [21]–[29]. While these studies are important for demonstrating the technology's potential and acceptance, they have not yet provided a detailed explanation of how teachers interpret the feedback generated by Quizizz and translate it into pedagogical decisions. This study fills that gap by demonstrating that the formative mechanism does not end with the availability of digital evidence. The mechanism only comes into play when teachers interpret the evidence, examine its meaning through interactions with students, and identify actionable learning needs.

Thus, the analytical contribution of this subsection is not merely to show that teachers "use Quizizz data," but to reveal the interpretive mechanism that transforms digital visibility into pedagogical information. If the first stage of this research mechanism yields evidence of student responses, the second stage yields information about students' learning needs through the teachers' interpretive work. This information then serves as a prerequisite for the next stage—feedback—when the results of the interpretation are no longer merely used by teachers to understand students' conditions but are translated into information communicated to students to help them understand their mistakes, the reasons behind them, and the path to improvement.

3.3 The Quizizz-Based Formative Assessment Mechanism: From Evidence of Learning to Pedagogical Follow-Up

The findings in the previous two subsections show that the

formative assessment mechanism in this study does not end with the collection of Quizizz response data or with teachers' interpretation of the assessment results. These two stages are part of a broader process: how assessment information is used to identify learning needs, provided back to students in the form of feedback, and translated into follow-up learning activities. The mechanism identified in this study can be understood as a gradual process of transformation from digital evidence to pedagogical meaning and instructional actions. An illustration of this mechanism is presented in Figure 1 below:

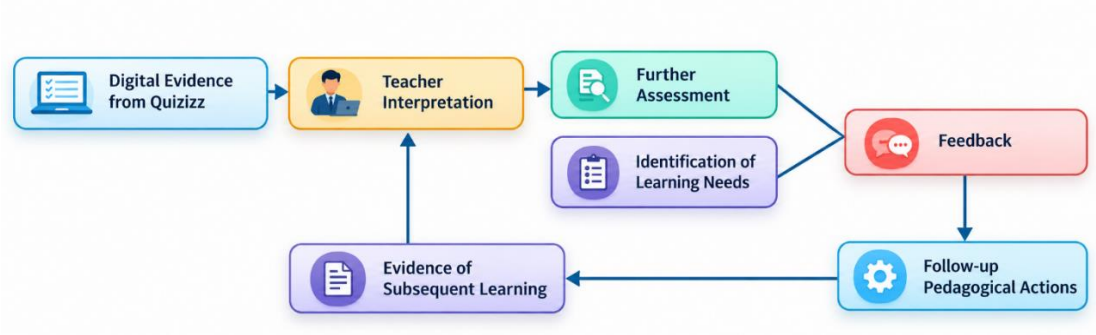


Figure 1: Conceptual Model of a Quizizz-Based Formative Assessment Mechanism in Islamic Religious Education

The research findings indicate that the Quizizz-based formative assessment mechanism does not stop at obtaining scores or identifying correct and incorrect answers, but proceeds through a step-by-step process that links digital evidence to the teacher's interpretation and instructional actions. Across three assessment sessions, Quizizz provides a record of responses at the item level, including students' choices and the distribution of responses for each alternative. In E01, for example, 23 out of 30 students answered correctly on an item testing their understanding of the meaning of tawakal accompanied by effort, but only 12 students answered correctly on a case requiring them to distinguish between tawakal and abandoning effort. A similar pattern was observed in E02, when 24 students recognized the meaning of "sabar" but only 13 selected the correct answer in a scenario involving a failed attempt, as well as in E03, when 25 students identified actions demonstrating responsibility but only 11 answered correctly in a scenario where "takdir" was used to justify negligence. These differences are treated as preliminary evidence regarding which parts of the material need further exploration, not as a longitudinal measure of student ability, since the content and difficulty levels of each item were not tested for equivalence. Quizizz in this study primarily served to increase visibility of student responses. This position aligns with Børte et al., who emphasize that the use of technology in formative assessment requires alignment between digital tools and pedagogical practices, as well as teachers' ability to examine and interpret the information generated by technology [20].

The next step is the teacher's interpretation of these response patterns. Quizizz data is not treated as an automatic diagnosis because a "correct" or "incorrect" status alone cannot explain why a student chose a particular answer. In E01, fourteen students selected the option stating that Rafi had placed his trust in God even though he did not study because the exam results were considered predetermined, while twelve students chose the answer that distinguished placing one's trust in God from abandoning one's efforts. The teacher used the concentration of these choices as a starting point for further examination, not as evidence that all students who chose option A were experiencing the same difficulty. This was reinforced by the experience of S04, who answered correctly but still had doubts, while S10 explained that his mistake arose from overlooking a specific part of the sentence in the case study.

The teacher also emphasized that the patterns of responses collected for certain options did not yet explain each student's reasoning. In E02, a similar pattern emerged when 13 students chose an answer that called for evaluation and improvement after failure, while 12 students viewed giving up as an expression of patience. S23 even realized that simply choosing the option considered correct was not enough, as their understanding needed to be demonstrated through the ability to explain their reasoning. In E03, eleven students chose an answer consistent with the principle of responsibility, while thirteen students selected an alternative that invoked fate to absolve themselves of responsibility; S18 distinguished between knowing the answer explained by the teacher and being able to accept and explain the reasoning in their own words. These findings demonstrate that teachers serve as mediators between assessment data and pedagogical meaning. This is consistent with Estaji et al., who showed that teachers' assessment literacy in digital environments encompasses the knowledge, cognitive processes, and contextual factors necessary for making assessment decisions not merely the ability to operate digital devices [30].

The next step involves further assessment to translate response patterns into insights about learning needs. Teachers use questions, case discussions, re-reading of questions, and requests for clarification as ways to examine the meaning behind the responses that appear on Quizizz. This stage is important because a single response can technically have different causes, while even a correct answer does not necessarily indicate a solid understanding. S04, for example, indicates that a correct answer may be accompanied by uncertainty, whereas S10 shows that an incorrect answer may be related to a missed part of the sentence, rather than merely a lack of conceptual understanding. In Arif's case, S23 indicates a need to explain the reasoning, not just select a letter, whereas in Bima's case, S18 highlights the difference between knowing the correct answer after a discussion and being able to explain the reasoning independently. Interpretation in this study can be understood as a process of diagnostic clarification that is, the process of clarifying the pedagogical meaning of initial signals obtained through digital assessment. These findings reinforce Hase and

Kuhl's view that data from digital learning platforms can provide information about students' performance, strengths, and weaknesses, but their use for instructional design still requires a decision-making process by the teacher [31]. Their review also indicates that research on the use of digital platform data for instructional interventions is still relatively nascent and often conducted in an exploratory manner; thus, the findings of this study contribute by illustrating this interpretive process at the level of Islamic Education classroom practice.

Once learning needs have been identified, this information is translated into feedback that is more pedagogically oriented. In this study, feedback is not understood merely as automated information about right or wrong provided by the platform, but as a teacher's response based on an interpretation of the evidence and the students' needs. This distinction is important because, in Rafi's case, the student not only needs to know the correct answer but also needs to understand why giving up cannot be equated with resignation; in Arif's case, the student needs to understand why patience is not the same as ceasing to make improvements; whereas in Bima's case, the student needs to understand why belief in fate does not eliminate responsibility for one's actions. Recent literature supports this distinction. Brummer et al., through a meta-analysis of 116 interventions, demonstrated that the effect of digital feedback on learning performance is influenced by various factors, including the focus of the feedback, the subject area, the type of assessment, and the level of student control [32]. These findings indicate that the mere presence of digital feedback is insufficient to explain its pedagogical function; the nature and manner in which feedback is used also determine its contribution to learning.

Feedback is then linked to pedagogical interventions, ensuring that assessment information does not merely serve as a report on learning outcomes. When response patterns indicate difficulty in applying a concept, teachers use case discussions, clarifications, follow-up questions, and follow-up activities to provide students with opportunities to re-examine the concept through relevant contexts. Findings by Hase and Kuhl show that data from digital learning platforms can be used by teachers to gain insight into student performance, strengths, and weaknesses, which then serves as the basis for adjusting instruction although this process does not occur automatically in every context [31]. The findings of this study clarify these mechanisms in Islamic Education (PAI) instruction: Quizizz data serve as the basis for identifying sections that need to be revisited, while teachers determine the form of follow-up based on pedagogical insights gained through further assessment. This position is also consistent with Børte et al., who identify alignment between technology and pedagogical practices as one of the key prerequisites for technology-assisted formative assessment [20]. Thus, technology serves to provide information, while decisions regarding how that information is used remain within the teacher's pedagogical process.

This mechanism is iterative because follow-up learning provides new evidence for re-examining previous interpretations. In this study, 90 student responses to follow-up tasks were used as supplementary data to examine how these response patterns relate to subsequent learning activities. This data was not used to claim that Quizizz causally improves learning outcomes. From an evidential standpoint, follow-up responses serve as illustrative-confirmatory evidence in a qualitative case study: the data can support, elaborate on, or reveal inconsistencies with the teacher's interpretation, but it does not test the effectiveness of feedback or prove that specific pedagogical actions cause changes in student responses. The proposed mechanism which includes digital responses, teacher interpretations, follow-up checks, consideration of learning needs, feedback, pedagogical interventions, and evidence from subsequent learning is a process-based explanation derived from the interrelationships among data sources, not a tested cause-and-effect sequence. Cases S04, S10, S23, and S18 also demonstrate that response patterns need to be interpreted within the context of each individual student and do not always reflect uniform learning processes or development.

These findings reinforce the view that the use of digital data for learning requires teachers to have the capacity to interpret information and link it to appropriate instructional actions. Hase and Kuhl demonstrate that teachers can use platform data both in real time and after instruction to determine next steps, including revisiting a topic; however, not all studies show that reviewing data automatically leads to action [31]. Thus, the formative nature of this study arises precisely because of the iterative relationship between evidence, interpretation, action, and new evidence.

Overall, the findings of this study reconstruct Quizizz not merely as a gamified quiz app, but as an environment for acquiring digital evidence that serves a formative function through the teacher's pedagogical mediation. The study's contribution lies in its explanation of the process linking assessment data to learning actions: response data do not automatically become formative evidence; evidence must be interpreted; interpretations must be clarified through interaction; the results of this clarification form the basis for feedback; feedback is translated into pedagogical actions; and these actions generate new evidence of learning. The core of this mechanism does not lie in the technology itself, but rather in the relationship between technology, teacher interpretation, and pedagogical actions. This conclusion expands the existing literature on Quizizz which has largely focused on motivation, engagement, perceptions, and learning outcomes by explaining the process through which data generated by the platform acquires meaning and is utilized in formative assessment of Islamic Education. Theoretically, these findings support four propositions: digital data does not automatically constitute formative evidence; teachers' interpretations mediate the relationship between data and learning needs; pedagogical feedback is the result of interpretation, not merely a technical output of the platform; and formative assessment is iterative because learning actions generate new evidence. This framework aligns with the systematic review by Børte et al., which identified a clear definition of formative assessment, the alignment of digital tools with pedagogical practices, and teachers' data literacy as prerequisites for the productive use of technology in formative assessment [20].

4. Conclusion

This study shows that the Quizizz-based formative assessment mechanism in Grade 12 Islamic Education (PAI) classes at SMK PSM Randublatung does not stop at obtaining scores or determining right or wrong answers, but proceeds through a series of steps involving the collection of digital evidence, teacher interpretation, further review, feedback, and pedagogical follow-up. Data from 24 assessment items across three episodes show that Quizizz is capable of providing a

response trail at the item level that helps teachers identify answer patterns and areas of the material that need further exploration; however, this evidence is indicative and does not automatically explain the source of students' difficulties. New pedagogical meaning emerges when teachers compare response patterns, consider the content of the questions and the learning context, and then examine students' reasoning through questions, case discussions, and dialogue. The results of this interpretation are then used to provide feedback in the form of conceptual clarification, reinforcement of understanding, reflective questions, and guidance for improvement, which are subsequently translated into follow-up actions such as material reinforcement, case discussions, targeted exercises or assignments, and adjustments to learning strategies. Thus, the main contribution of this study lies in affirming that Quizizz is best understood as part of an evidence-based formative assessment cycle, rather than merely a digital quiz tool: its formative value emerges when response data are critically interpreted by teachers, transformed into feedback, and used to generate pedagogical actions and evidence of subsequent learning.

References

[1] A. N. M. Fuadi, A. I. A. Jaya, D. A. Fitriana, and I. bin Salem, "Strategic Management of Digital Transformation in Aqidah Akhlak Education at Islamic Senior High Schools," *Turots J. Pendidik. Islam*, vol. 7, no. 2, pp. 855–869, 2025, doi: 10.51468/jpi.v7i2.1142.

[2] A. Khaldi, R. Bouzidi, and F. Nader, "Gamification of e-learning in higher education: a systematic literature review," **Smart Learn. Environ. 2023 101*, vol. 10, no. 1, pp. 10–, Jan. 2023, doi: 10.1186/S40561-023-00227-Z.

[3] W. Oliveira *et al.*, "Tailored Gamification in Education: A Literature Review and Future Agenda," *Educ. Inf. Technol. 2022 281*, vol. 28, no. 1, pp. 373–406, June 2022, doi: 10.1007/S10639-022-11122-4.

[4] L. Li, · Khe, F. Hew, · Jiahui Du, K. Foon, and H. K. Hk, "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but has minimal impact on competency: a meta-analysis and systematic review," *Educ. Technol. Res. Dev. 2024 722*, vol. 72, no. 2, pp. 765–796, Jan. 2024, doi: 10.1007/S11423-023-10337-7.

[5] S. A. Triantafyllou, C. Georgiadis, and T. Sapounidis, "Gamification in education and training: A literature review," *Int. Rev. Educ. 2025 713*, vol. 71, no. 3, pp. 483–517, Apr. 2025, doi: 10.1007/S11159-024-10111-8.

[6] T. M. Lim and M. M. Yunus, "Teachers' Perceptions of the Use of Quizizz in the Teaching and Learning of English: A Systematic Review," *Sustain. 2021, Vol. 13, p. 6436*, Vol. 13, No. 11, p. 6436, June 2021, doi: 10.3390/SU13116436.

[7] P. Permana, I. Permatawati, and E. Khoerudin, "Gamification of Foreign Language Learning Using Quizizz: A Systematic Review Based on Students' Perceptions," *Eralingua Journal of Foreign Language and Literature Education, Foreign Languages and Literature*, vol. 7, no. 2, pp. 233–249, Aug. 2023, doi: 10.26858/ERALINGUA.V7I2.23969.

[8] Z. Zhang and J. Crawford, "EFL Learners' Motivation in a Gamified Formative Assessment: The Case of Quizizz," *Educ. Inf. Technol. 2023 295*, vol. 29, no. 5, pp. 6217–6239, July 2023, doi: 10.1007/S10639-023-12034-7.

[9] V. Hernanz, C. Latorre-Cosculluela, C. Suárez, and E. Lanchares-Sancho, "Revitalizing learning in three university contexts: Unleashing the power of the Quizizz app to increase self-efficacy, intrinsic motivation, satisfaction, and performance," *Educ. Inf. Technol. 2024 2917*, vol. 29, no. 17, pp. 23917–23935, June 2024, doi: 10.1007/S10639-024-12779-9.

[10] J. Zeng, D. Sun, C. K. Looi, and A. C. W. Fan, "Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from 2008 to 2023," *Br. J. Educ. Technol.*, vol. 55, no. 6, pp. 2478–2502, Nov. 2024, doi: 10.1111/BJET.13471;JOURNAL:JOURNAL:14678535;WEBSITE:WEBSITE:BERA-

[11] H. Cigdem, M. E. Korkusuz, and C. Karaçaltı, "Gamified learning: Assessing the influence of leaderboards on online formative quizzes in a computer programming course," *Comput. Appl. Eng. Educ.*, vol. 32, no. 2, p. e22697, Mar. 2024, doi: 10.1002/CAE.22697;REQUESTEDJOURNAL:JOURNAL:10990542;PAGE:STRING:ARTICLE/CHAPTER.

[12] H. Cigdem, M. Ozturk, Y. Karabacak, N. Atik, S. Gürkan, and M. H. Aldemir, "Unlocking student engagement and achievement: The impact of leaderboard gamification in online formative assessment for engineering education," *Educ. Inf. Technol. 2024 2918*, vol. 29, no. 18, pp. 24835–24860, June 2024, doi: 10.1007/S10639-024-12845-2.

[13] J. J. R. Ruiz, A. D. V. Sanchez, and O. R. B. Figueredo, "Impact of gamification on school engagement: a systematic review," *Front. Educ.*, vol. 9, p. 1466926, Dec. 2024, doi: 10.3389/FEDUC.2024.1466926/FULL.

[14] F. Su and D. Zou, "A systematic review of game-based assessment in education over the past decade," *Knowl. Manag. E-Learning: An Int. J.*, vol. 16, no. 3, pp. 451–476, Oct. 2024, doi: 10.34105/J.KMEL.2024.16.021.

[15] A. Bijl, B. P. Veldkamp, S. Wools, and S. de Klerk, "Serious games in high-stakes assessment contexts: a systematic literature review into the game design principles for valid game-based performance assessment," *Educ. Technol. Res. Dev.*, vol. 72, no. 4, pp. 2041–2064, Aug. 2024, doi: 10.1007/S11423-024-10362-0/TABLES/9.

[16] P. Kannan and D. Zapata-Rivera, "Facilitating the Use of Data From Multiple Sources for Formative Learning in the Context of Digital Assessments: Informing the Design and Development of Learning Analytics Dashboards," *Front. Educ.*, vol. 7, p. 913594, June 2022, doi: 10.3389/FEDUC.2022.913594/TEXT.

[17] R. Morris, T. Perry, and L. Wardle, "Formative assessment and feedback for learning in higher education: A systematic review," *Rev. Educ.*, vol. 9, no. 3, p. e3292, Oct. 2021, doi: 10.1002/REV3.3292;WEBSITE:WEBSITE:BERA-JOURNALS;ISSUE:ISSUE:DOI.

[18] P. Karaman, "The Effect of Formative Assessment Practices on Student Learning: A Meta-Analysis Study," *Int. J. Assess. Tools Educ.*, vol. 8, no. 4, pp. 801–817, Dec. 2021, doi: 10.21449/IJATE.870300.

[19] A. M. Lui and H. L. Andrade, "The Next Black Box of Formative Assessment: A Model of the Internal

Mechanisms of Feedback Processing," *Front. Educ.*, vol. 7, p. 751548, Feb. 2022, doi: 10.3389/FEDUC.2022.751548/TEXT.

[20] K. Børte, S. Lillejord, J. Chan, B. Wasson, and S. Greiff, "Prerequisites for Teachers' Use of Technology in Formative Assessment Practices: A Systematic Review," *Educ. Res. Rev.*, vol. 41, p. 100568, Nov. 2023, doi: 10.1016/J.EDUREV.2023.100568.

[21] S. Rahmania, I. Soraya, and A. S. Hamdani, "THE USE OF THE QUIZIZZ GAMIFICATION PLATFORM ON STUDENTS' LEARNING MOTIVATION IN ISLAMIC RELIGIOUS EDUCATION," *Tadbir J. Manaj. Pendidik.* vol. 11, no. 2, pp. 114–133, Aug. 2023, doi: 10.30603/TJMPI.V11I2.3714.

[22] I. Listiyani and M. N. Yamin, "The Effect of Quizizz Gamification and Learning Engagement on Learning Outcomes in Islamic Religious Education (PAI) at Muhammadiyah Mlangi Elementary School for the 2022/2023 Academic Year," *G-Couns J. Guidance and Counseling*, vol. 7, no. 03, pp. 773–786, Aug. 2023, doi: 10.31316/GCOUNS.V7I03.5029.

[23] F. N. Anggraini and T. A. Mustofa, "Quizizz-Based Learning Innovation to Improve Achievement in Islamic Religious Education," *Iseedu J. Islam. Educ. Thoughts Pract.*, vol. 8, no. 1, pp. 47–56, May 2024, doi: 10.23917/ISEEDU.V8I1.8843.

[24] S. Sundari, M. S. Nugraha, and U. Dedih, "Implementing Quizizz in Islamic Education: Enhancing Learning Engagement in Modern Classrooms," *J. Inov. Pendidik.* vol. 4, no. 1, pp. 19–36, Dec. 2024, doi: 10.15575/JIPAI.V4I1.31179.

[25] Sakdiah, Hayati, S. Mahmud, and M. Furqan, "Gamification in Islamic Religious Education: Developing Interactive Learning Media via Quizizz Using the ADDIE Model," **Edukasi: Journal of Education and Teaching*, vol. 12, no. 01, pp. 187–207, May 2025, doi: 10.19109/D48TV716.

[26] I. A. Yuliana, D. Siamika, T. Prayogi, and A. Miftahorrahman, "Reception Analysis of Quizizz-Based Gamification in Islamic Religious Education at Vocational High Schools," **EDURELIGIA J. Educators. Islamic Education*, vol. 9, no. 1, pp. 32–47, Jun. 2025, doi: 10.33650/EDURELIGIA.V9I1.10858.

[27] M. Shokhiful Fikri and A. Marzuki, "The Influence of Quizizz on Students' Learning Motivation in Islamic Religious Education (PAI) at Vocational High Schools," *AR-RASYID Journal of Education. Islamic Education*, vol. 5, no. 1, pp. 137–144, June 2025, doi: 10.30596/ARRASYID.V5I1.24520.

[28] N. Hidayati, S. Aripin, and S. Hina, "Gamified Formative Assessment with Quizizz to Enhance Learning Outcomes at Madrasah," *J. Islam. Educ. Res.*, vol. 6, no. 3, pp. 303–314, July 2025, doi: 10.35719/JIER.V6I3.492.

[29] I. Magfirah, F. Afiyati, and A. Bashith, "The Transformation of Digital-Based Islamic Education Assessment: Optimizing Quizizz as an Adaptive Assessment Tool," *Al-Qalam J. of Islamic Studies Islam and Educators*, vol. 17, no. 1, pp. 435–444, June 2025, doi: 10.47435/AL-QALAM.V17I1.3949.

[30] M. Estaji, G. T. L. Brown, and Z. Banitalebi, "The Key Competencies and Components of Teacher Assessment Literacy in Digital Environments: A Scoping Review," *Teach. Teach. Educ.*, vol. 141, p. 104497, Apr. 2024, doi: 10.1016/J.TATE.2024.104497.

[31] A. Hase and P. Kuhl, "Teachers' use of data from digital learning platforms for instructional design: a systematic review," *Educ. Technol. Res. Dev. 2024 724*, vol. 72, no. 4, pp. 1925–1945, Mar. 2024, doi: 10.1007/S11423-024-10356-Y.

[32] L. Brummer, H. de Boer, J. M. Mouw, and J. W. Strijbos, "A meta-analysis of the effects of context, content, and task factors of digitally delivered instructional feedback on learning performance," *Learn. Environ. Res. 2024 273*, vol. 27, no. 3, pp. 453–476, June 2024, doi: 10.1007/S10984-024-09501-4.