

## The Use of Test and Non-Test Techniques in the Assessment of Arabic Term Comprehension in PAI Learning at SDIT Salsabila Al Muthi'in

Rizki Sanjaya<sup>a,1,\*</sup>, Dedi Saputra<sup>b,2</sup>, Muh. Aditya Putra. P<sup>c,3</sup>, Muhammad Nur Hafidzul Qur'an<sup>d,4</sup>, Azhara<sup>e,5</sup>, M. Husni Arsyad<sup>f,6</sup>.

<sup>a,b,c,e,f</sup> Tarbiyah Madani College of Sciences Yogyakarta

<sup>d</sup> UIN Raden Mas Said Surakarta

<sup>1</sup> rizkisanjay003@gmail.com ; <sup>2</sup> d4zuyes08@gmail.com ; <sup>3</sup> mhdadyaaa22@gmail.com ; <sup>4</sup> quranhafidzul1@gmail.com ;

<sup>5</sup> ouyuwaaa@gmail.com ; <sup>6</sup> husniarsyad@stitmadani.ac.id .

\*Correspondent Author

Received: 23-08-2026

Revised: 11-09-2026

Accepted: 18-10-2026

### KEYWORDS

test and non-test techniques;  
understanding of Arabic terms;  
learning assessment;

### ABSTRACT

The assessment of Islamic Religious Education (PAI) learning in elementary schools requires evidence that can distinguish the understanding of the meaning of Arabic terms from memorization and learning engagement. This study aims to describe the planning, implementation, and utilization of assessment results through test and non-test techniques at SDIT Salsabila Al Muthi'in. The research uses a qualitative approach with a case study design in PAI learning in grades III and VI. The main informants are PAI teachers who are selected purposively. Data were collected through learning observation, semi-structured interviews, and document review, then analyzed thematically. The validity of the data was checked through the comparison of evidence sources and interpretation confirmation to the informant. The results showed that assessment planning refers to the material, learning targets, and grade level, but the grid was not available to be examined during the study. The implementation of the assessment included thematic written questions, daily tests, informal oral questions, and observation of students' activities, attitudes, and responses. Non-test observations did not use structured instruments so that evidence of learning engagement could not be equated with evidence of understanding of meaning. The results of the assessment were used to identify students who had not achieved the target and provide additional guidance. The obstacles found included time constraints and differences in the level of understanding of students. This study showed the connection between the assessment and mentoring processes, but could not be assured of its effectiveness on the understanding of Arabic terms because indicators, specific errors, and developments after follow-up have not been documented. The development of assessments requires explicit indicators of meaning, structured instruments, and reassessment to examine the success of mentoring.

This is an open-access article under the CC-BY-SA license.



## Introduction

Islamic Religious Education (PAI) in elementary schools requires assessments that are able to describe students' understanding while providing a basis for learning improvement. Assessment does not stop at the provision of grades, but includes planning, evidence collection, interpretation,

and follow-up of learning outcomes. (Sawaluddin & Muhammad, 2020) places follow-up as part of the PAI evaluation series, while (Black & Wiliam, 2009) shows the importance of using assessment evidence to adjust learning. Thus, the use of assessment techniques needs to be assessed based on their suitability with the competencies assessed and the learning decisions produced. One of the relevant aspects of understanding in PAI is the meaning of the Arabic term used to explain religious concepts. Terms such as faith, piety, zakat, and ablution are examples of terms that can be found in PAI materials; the terms that are the object of research still need to be determined based on the material actually taught. The ability to pronounce or memorize terms does not in itself indicate understanding of the concept. In the vocabulary study, (Nation, 2022) emphasizes the scope of word knowledge and learning, while offering an assessment framework that goes beyond measuring vocabulary separately. The framework is relevant as a basis for formulating an assessment of the meaning of terms in the context of PAI, although its use needs to be adjusted to the learning objectives of religion and the age of students.

The understanding of terms in this study is directed at the ability to recognize meaning, explain meaning in one's own language, and connect terms with appropriate examples in PAI material. This direction considers the difference between partial knowledge and more precise knowledge, depth of knowledge, and receptive and productive abilities as stated (Henriksen, 1999). (Schmitt, 2014) also shows that the breadth and depth of vocabulary knowledge depend on how both are conceptualized and measured. Meanwhile, (Nafiati, 2021) distinguishes the process of remembering, understanding, and applying in the revision of Bloom's taxonomy. Based on this foundation, the ability to mention translation needs to be complemented by evidence of explanation of meaning; application examples are used to clarify conceptual understanding, not to claim overall Arabic language mastery. Comprehension assessment can use written or oral tests that are compiled based on specific indicators. (Putri et al., 2022) discuss the form of descriptive and objective tests as cognitive assessment instruments. In vocabulary assessment, (Schmitt et al., 2001) and (Webb et al., 2017) show the importance of developing and validating the instrument so that the results can be interpreted appropriately. The English vocabulary instruments in both studies cannot be directly applied to Arabic terms in PAI. However, the principle of conformity between the assessment objectives, the form of the assignment, and the interpretation of the answer provides a basis for teachers to choose questions that are in accordance with the material and the class level.

Non-test techniques can provide supporting evidence through observation of the way students use or explain terms during learning activities. (Mukmin & Nuraini, 2024) and (Abdurahman & Erihadiana, 2024) discuss the need for a balance of test and non-test in a comprehensive PAI evaluation. However, observation of discipline, enthusiasm, or obedience is not enough to conclude an understanding of the meaning of the term. Observations need to be directed at behaviors that show understanding, such as the accuracy of spontaneous explanations in discussions. (Panadero & Jonsson, 2013) discusses the use of rubrics for formative purposes, so that clear observational criteria can be considered to aid in the recording and interpretation of evidence. Oral questions that are deliberately structured as tests are still categorized as oral tests; the classification of techniques is determined based on the procedure, not just the medium. In addition to the form of the instrument, the function of assessment is determined by the use of the results. (Sadler, 1989) describes the relationship between formative assessment and quality standards and performance improvement, while (Bennett, 2011) emphasizes the need for conceptual clarity and adequate evidence in the discussion of formative assessment. (Nurlitasari & Hamami, 2023) discusses the assessment of, for, and as learning in PAI at the upper secondary level, while Subehi and (Subehi & Sriyanto, 2021) examine its application in online learning of PAI in junior high school. The literature shows the relevance of the discussion of objectives and the use of assessment in PAI. In this study, formative functions are understood as the use of evidence to determine the next learning step; this function is not automatically attached to all non-test activities.

Previous studies on PAI evaluation have explained the use of various assessment techniques. (Widiyanto & Inayati, 2023), through research at SMK Batik 2 Surakarta, found the use of oral or written knowledge tests, practical assessments, and remedial for students who have not met the criteria for completeness. (Mukmin & Nuraini, 2024) examines the integration of tests and non-tests through literature studies, while (Abdurahman & Erihadiana, 2024) examines the balance of the two techniques in the evaluation of Islamic education. The three studies provide a basis for discussing the assessment of PAI comprehensively. However, the target of the discussion has not been specifically directed at the assessment process of understanding Arabic terms in elementary school students. Other research highlights the planning and quality of assessment instruments. (Azis et al., 2022) describes the planning and implementation of online-based assessments in PAI at the high school level, including setting goals and compiling grids. (Puspitasari et al., 2023) developed a Quizizz-based PAI assessment tool at the junior high school level by examining the feasibility and characteristics of question items. Both studies show that the use of instruments needs to be accompanied by quality and planning considerations. In elementary schools, (Ijarmana & Putra, 2021) found difficulties for teachers in applying authentic assessment techniques to attitude competencies, knowledge, and skills. However, these studies have not specifically explained the connection between the choice of Arabic terms, indicators of comprehension, test and non-test evidence, and follow-up of students' difficulties in learning PAI at the location of this study.

In the field of Arabic, (Mohamad et al., 2021) examined the mastery of receptive and productive academic vocabulary in public university students in Malaysia. The results of the productive test in the study were higher than the results of the receptive test; the researcher attributed the possible explanation to the characteristics of the vocabulary and the format of the test. This finding reminds that the results of the assessment need to be interpreted by paying attention to the form of the instrument. However, the study used an academic vocabulary test on students, so the findings cannot be transferred directly to elementary school students or religious terms in PAI. Differences in objectives, age, material context, and the use of assessment results are reasons for the need for a more contextual study. The formative assessment study also shows that the collection of evidence is not necessarily followed by the right teaching decision. (Heritage et al., 2009), in the context of mathematics learning, found that teachers are better able to interpret students' levels of understanding rather than determine the next step of instruction. (Schildkamp et al., 2020) identified teacher prerequisites for the use of formative assessment through systematic review. (Yen et al., 2020) demonstrated the benefits of formative assessment in K–12 education, with outcomes dependent on its implementation features. These studies serve as a basis for tracing the relationship between assessment outcomes and follow-up, without making them evidence that a particular technique is necessarily effective in the context of the PAI being studied.

Follow-up requires information that can help students improve their comprehension. (Hattie & Timperley, 2007) and (Shute, 2008) show the importance of the type, content, and manner of giving feedback. (Wisniewski et al., 2020) through meta-analysis also shows variations in the influence of feedback in education. Meanwhile, (Andrade, 2019) discusses student self-assessment as a practice that can support the learning process. This foundation directs research attention to how teachers explain misunderstandings, provide comparative examples, or provide follow-up guidance according to evidence of students' difficulties. These forms are aspects that will be explored, not activities that are assumed to have been carried out in the research school. Preliminary data from the mini-research at SDIT Salsabila Al Muthi'in showed that PAI teachers in grades III and VI used written tests in the form of thematic questions and tests, accompanied by observations of student activities, attitudes, and responses. Field records stated that non-test observations had not used structured instruments. Interviews also revealed time constraints, differences in students' understanding levels, and guidance to students who had not reached the target. This information shows that there is an assessment and follow-up process that can be traced further. However, the initial data does not detail the Arabic terms assessed, comprehension indicators, or evidence of special difficulties related to the term. Thus, this initial condition is the

basis for the focus of the research, not evidence that students have had difficulty understanding certain terms.

Based on the literature reviewed, the knowledge gap lies in the unexplained relationship between the planning, implementation, and utilization of the assessment of Arabic term comprehension in the context of elementary school PAI learning where the research is located. The PAI assessment study provides an explanation of the techniques and domains of assessment, while the vocabulary study provides a basis for distinguishing the level of knowledge of meaning. There is still a need for a contextual explanation of how teachers determine indicators, select and interpret evidence from tests and observations, and then use them to determine learning aids. This gap is limited to the sources studied and the context of the research; it is not intended as a claim that there has been no similar research in the entire literature. The novelty offered by this study is that it focuses on the interconnectedness of the three stages of assessment for one specific object, namely the understanding of Arabic terms in PAI. The research places the suitability of indicators, evidence of understanding, and follow-up as a series of analyses. The novelty does not rest solely on the use of tests and non-tests or location differences, but on an attempt to explain how both techniques provide information about the meaning of terms and how that information underlies teacher decisions. This contribution is contextual and directed at producing an overview of assessment practices that can be examined, rather than new models that are assumed to have proven effective.

The urgency of this research lies in the need to maintain the accuracy of the interpretation of students' abilities and the relevance of the guidance provided. Assessments that only record memorization or activeness have the potential to not provide enough information about the understanding of meaning. On the other hand, evidence that is appropriate to indicators can help teachers distinguish the ability to recognize terms, explain meanings, and relate them to the context of the material. Academically, this study is expected to bring together the discussion of PAI evaluation with the assessment of vocabulary knowledge in the context of elementary school. Practically, the results are expected to provide input for the preparation of indicators, observation recording, and follow-up in accordance with the needs of students. Based on this description, this study aims to: (1) describe how teachers plan the assessment of understanding of Arabic terms through test and non-test techniques in PAI learning at SDIT Salsabila Al Muthi'in; (2) analyze how teachers implement both techniques to assess students' understanding of Arabic terms; and (3) explain how teachers use the assessment results to follow up on students' difficulties in understanding Arabic terms in PAI learning. These three objectives place the assessment process and the use of evidence as the focus of the research.

### **Test and Non-Test Techniques**

Learning assessment is the process of collecting and interpreting information to determine students' achievements and determine learning follow-up. This information can be obtained through test and non-test techniques. The selection of both needs to be adjusted to the learning objectives, the characteristics of the competencies assessed, and the form of evidence needed. In Islamic Religious Education (PAI) learning, the use of both techniques can help teachers obtain a more comprehensive picture of students' abilities, including knowledge, attitudes, and skills (Mukmin & Nuraini, 2024). Test techniques are a way of obtaining information about students' abilities through questions or tasks that must be answered or done based on certain assessment criteria. Tests provide opportunities for students to demonstrate the abilities that are the target of assessment. Therefore, the preparation of tests needs to start from setting goals, formulating indicators, selecting question forms, and preparing scoring guidelines. (Sawaluddin & Muhammad, 2020) explained that the evaluation of learning outcomes includes planning, data collection and processing, interpretation, and follow-up. Thus, tests are part of the evaluation process whose results need to be used to support learning decisions.

Based on how it is implemented, tests can be in the form of written tests, oral tests, and practice tests. Written tests ask students to give written answers, while oral tests ask for answers through verbal communication. Practice tests ask students to demonstrate abilities through the execution of certain tasks or actions. These forms of responses do not automatically determine the domain of ability being assessed; the main determinants are the indicators and demands of the task. For example, oral tests can be used to assess the ability to explain the meaning of a term, while practical tests can be used to assess the skills of performing worship. In written tests, the form of questions can be divided into objective tests and description tests. Objective tests include multiple-choice, true-wrong, matchmaking, and short answers. This form allows assessment based on predetermined answers. Meanwhile, the description test provides students with the opportunity to formulate answers, explain the relationships between concepts, and present reasons. Scoring a description test requires clear guidelines so that assessment does not depend only on the assessor's impression (Putri et al., 2022).

In the assessment of the comprehension of Arabic terms in PAI, the test needs to reveal more than the ability to memorize sounds or translate terms. Teachers can formulate questions that ask students to recognize meanings, explain in their own language, or relate terms to examples in the material. As an illustration, questions about *zakat* can ask students to choose the right meaning, explain their meaning, or distinguish it from alms according to the material that has been studied. (Read, 2000) emphasizes the need to clarify the aspects of vocabulary knowledge that are the target of assessment. (Schmitt, 2014) also shows that the number of known words and the depth of knowledge of words are aspects that need to be considered in vocabulary measurement.

Non-test techniques are a way of collecting information about students through procedures other than testing with a set of questions or test assignments. These forms include observation, interviews, questionnaires, attitude scales, and development notes. (Abdurahman & Erihadiana, 2024) explained that non-test techniques can provide information about attitudes, behaviors, and characteristics of students that complement test results. Its use still requires compatibility between the information collected and the purpose of assessment. Observation is carried out by observing students' behavior during learning activities. In order for the results to be accounted for, teachers need to determine the aspects observed, indicators, observation times, and how they are recorded. Observations without indicators have the potential to produce assessments based on general impressions. In the assessment of the understanding of Arabic terms, observation can be directed to the accuracy of students using terms when discussing, the ability to explain meaning spontaneously, or the suitability of the examples presented. On the other hand, discipline and enthusiasm only show aspects of learning behavior and are not enough to prove understanding of meaning.

Interviews or conversations can be used to explore students' explanations and difficulties in more depth. Questionnaires can help obtain information about students' learning experiences, perceptions of difficulties, or self-assessment. However, a student's statement that he or she understands a term is information about self-perception that needs to be examined along with evidence of actual ability. (Andrade, 2019) discusses self-assessment as an activity that can support learning if students use criteria to assess and improve their work. The boundary between tests and non-tests needs to be set based on the implementation procedure. Verbal questions that are deliberately designed to test the mastery of meaning and are scored are oral tests. Meanwhile, recording students' spontaneous explanations during discussions can be part of non-test observation. Similarly, practical activities are not always categorized as non-tests: practices that are carried out as structured testing tasks can be practice tests. Clarity of classification is needed so that the techniques used in the research are explained consistently.

### **Assessment of Arabic Term Comprehension in PAI Learning**

The assessment of the understanding of Arabic terms in Islamic Religious Education (PAI) learning is the process of collecting and interpreting information about students' ability to

understand the meaning of terms used in religious materials. This assessment is directed to determine students' ability to recognize meanings, explain meanings in their own language, and connect terms with the context of the material being studied. The scope is adjusted to the learning objectives of PAI, so it is not intended to measure all students' Arabic language skills. Arabic terms in PAI learning are related to religious concepts that need to be understood appropriately. Terms such as *faith*, *piety*, *zakat*, and *ablution* can be examples if they are found in the material taught. Some of these terms have been used in Indonesian, but the introduction to sounds or writing is not always accompanied by an understanding of the meaning. Students may be able to mention a term or memorize its definition, but still have difficulty when asked to explain the meaning using their own language. Therefore, assessment needs to obtain evidence that shows an understanding of the concept behind the term.

Term understanding can be explained through the study of vocabulary knowledge. (Nation, 2022) discusses word knowledge through aspects of form, meaning, and use. In PAI learning, this framework can be used to pay attention to the relationship between terms that students recognize, the meaning they understand, and the accuracy of their use when explaining the material. However, its application needs to be adjusted to the characteristics of PAI, because the mastery of terms serves to support the understanding of religious concepts. (Read, 2000) also explains the importance of determining the aspects of vocabulary knowledge that are the target of assessment. Thus, assessments of pronunciation, writing, and understanding of meaning need to be differentiated so that the results are not misinterpreted. Knowledge of a term develops gradually. (Henriksen, 1999) suggests the development from partial knowledge to more precise knowledge, depth of knowledge, and receptive ability to be productive. The framework shows that students can recognize terms when reading or hearing them, but are not necessarily able to explain or use them appropriately. (Schmitt, 2014) also shows that the breadth of vocabulary knowledge and depth of understanding are aspects that need to be considered in measurement. In the context of PAI, the number of terms that students can mention is not enough to show the quality of their understanding of the concepts learned.

Comprehension also needs to be distinguished from the ability to remember. (Nafiati, 2021), in the discussion of Bloom's taxonomy revision, explains the difference in the cognitive process of remembering, understanding, and applying. Remembering is related to the ability to retrieve information that has been learned, while understanding requires students to build the meaning of the information. Based on these differences, the assessment of understanding of Arabic terms needs to provide opportunities for students to explain the definition, distinguish related terms, or identify appropriate examples. For example, students are not only asked to mention the meaning of *zakat*, but also explain the difference between alms according to the material and level of learning. Assessment planning begins with determining the terms that are actually contained in the material and formulating indicators of understanding. These indicators can include the ability to recognize meanings, explain meanings in their own language, distinguish related concepts, and connect terms with examples in PAI learning. This formulation is an operationalization for research that needs to be adjusted to learning objectives, not a standard instrument that is directly taken from one theory. Teachers need to determine the elements of meaning that must appear in the answers, so that the accuracy of understanding is not measured based on the similarity of students' sentences with the definitions in the book alone.

Evidence of comprehension can be collected through test and non-test techniques. Written tests in the form of multiple-choice or matchmaking can be used to obtain information about the recognition of meaning, while short answers and descriptions allow students to explain the meaning or relationships between concepts. (Putri et al., 2022) discuss objective tests and descriptions as a form of cognitive assessment instruments. The selection of the form of the question needs to follow the ability to be assessed. If the learning objective requires students to explain the meaning, the assessment needs to provide an assignment that allows the student to demonstrate these skills. Oral tests can be used to obtain students' explanations and trace unclear

answers through follow-up questions. Meanwhile, non-test observation can record the accuracy of the use of terms that arise spontaneously in discussions or learning activities. The two need to be differentiated based on their implementation procedures. Questions that are deliberately given as tests are oral tests, while recording students' explanations during learning activities is observation. Observations of activeness, discipline, or enthusiasm can provide information about learning behavior, but are not enough to prove an understanding of the meaning of the term.

Assessment requires clear criteria so that answers and observations can be interpreted consistently. These criteria need to explain the elements of the right meaning, missing parts, and forms of conceptual errors. Rubrics can be considered to help teachers distinguish between correct, partially correct, and incorrect answers. (Panadero & Jonsson, 2013) discuss the use of rubrics for the purpose of formative assessment. In assessing term understanding in PAI, rubrics can help clarify the reasons for scoring as well as identify the needs of students' guidance. The results of the assessment also need to be interpreted by paying attention to the characteristics of the assignments given. (Mohamad et al., 2021), through research on Arabic academic vocabulary in students, showed that the results of receptive and productive tests can be different and need to be read taking into account the characteristics of vocabulary and test format. The findings cannot be applied directly to elementary school students, but they remind that one form of testing does not necessarily provide a complete picture of a learner's abilities. Teachers can compare written answers, oral explanations, and observation notes that refer to the same indicators to obtain a more adequate basis for interpretation.

The use of assessment results is directed at determining follow-up according to the difficulties found. Students who do not recognize meaning need different help from students who have recognized meaning but are not yet able to distinguish examples. Follow-up can be in the form of re-explanation, giving comparative examples, grouping situation exercises, or guided conversations. (Black & Wiliam, 1998) explained that the formative function of assessment lies in the use of evidence to determine the next step of learning. Therefore, the results of the assessment need to be used as a basis for providing assistance, followed by the collection of evidence again to determine the development of students' understanding. Feedback is an important part of the follow-up. (Shute, 2008) emphasizes the importance of specific feedback and supports learning improvement. In PAI learning, teachers can explain the parts of the understanding that have been corrected, point out misunderstandings, and provide instructions to correct them. Giving grades or statements that the answer is wrong does not provide enough information for students to understand where the error lies.

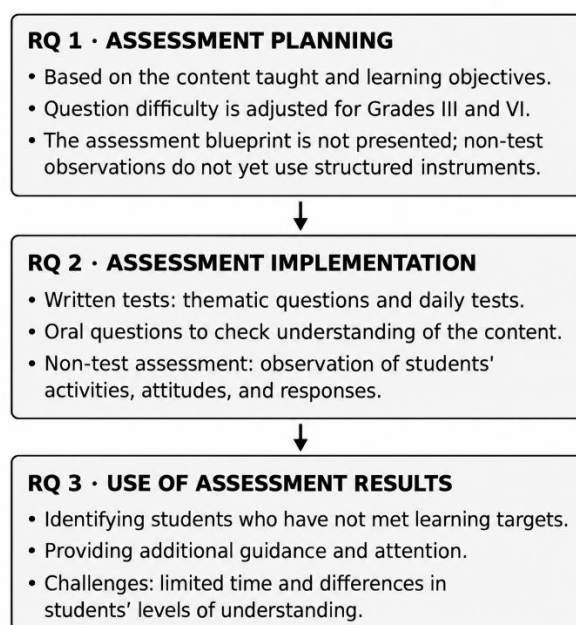
## Method

This study uses a qualitative approach with a case study design to understand the assessment process in the context of learning that is limited to specific locations and classes (Creswell & Poth, 2018). The research was carried out at SDIT Salsabila Al Muthi'in, Banguntapan, Bantul, Yogyakarta, with a focus on planning, implementing, and utilizing the results of the assessment of Arabic term comprehension in PAI learning in grades III and VI. The main informants were PAI teachers who were purposively selected based on their involvement in the assessment process. Data was collected through learning observation, semi-structured interviews, and analysis of available documents, such as lesson plans, evaluation questions, and student work results. The researcher acted as the main instrument by using observation, interview, and document review guidelines. Data analysis was carried out thematically through data introduction, coding, theme preparation, theme review, theme definition, and report preparation as described (Braun & Clarke, 2006). The validity of the data was checked through comparison of interview evidence, observation, and documentation as well as confirmation of interpretation to informants. Since preliminary data have not yet detailed the assessment of Arabic terms, further data collection is required so that the analyzed evidence is consistent with the focus of the study.

## Result

The results of the research are mapped based on three focuses, namely assessment planning, the implementation of test and non-test techniques, and the use of assessment results for learning follow-up. The following concept map presents the relationship of these three focuses in the practice of PAI assessment at SDIT Salsabila Al Muthi'in, including the basis for preparing evaluations, techniques used, and providing guidance to students.

Figure 1.1: Research Results



The concept map shows that the practice of PAI assessment at SDIT Salsabila Al Muthi'in includes planning based on learning materials and targets, the implementation of written tests and non-test observations, and the use of evaluation results to provide guidance to students. However, non-test observations have not used structured instruments, while evidence of assessment of Arabic term comprehension has not been specifically identified.

## Discussion

### Planning the Assessment of Arabic Term Comprehension through Test and Non-Test Techniques in PAI Learning

The results of the study show that the planning of PAI learning assessment at SDIT Salsabila Al Muthi'in relies on the suitability between the material taught, the learning target, and the characteristics of the students. Based on the interview, the teacher determines the instrument according to the material that has been delivered and adjusts the difficulty level of the questions to the grade III and VI levels. This finding is strengthened by observations that show the delivery of learning objectives at the beginning of the activity and the systematic presentation of the material. The documentation data also records the ownership of lesson plans or teaching modules, written evaluation questions, and daily test scores. This linkage indicates that the assessment is directed to check the achievement of learning, although the completeness of the instrument design cannot be fully examined by the researcher. This pattern can be understood through the concept of *constructive alignment* from (Biggs, 1996), which emphasizes the alignment of objectives, learning activities, and assessment tasks. In this study, the adjustment of the questions to the material showed that there was an alignment at the content level. However, the suitability of the material

did not automatically prove the suitability of the level of ability measured. Questions that discuss the same material can still measure different abilities, such as remembering, explaining, or applying concepts. Thus, the findings support the existence of an assessment orientation that is in harmony with learning, but it is not enough to state that all instruments have measured comprehension in depth.

In relation to the understanding of Arabic terms in PAI learning, this alignment needs to be translated into specific indicators of ability. (Nation, 2013a) explains that knowledge of a word has several dimensions and levels; recognizing the form of a word is not the same as understanding the meaning and being able to use it. This framework is relevant for interpreting the assessment of Arabic terms because learners can recognize or memorize religious terms without fully understanding the concepts they contain. Therefore, assessment planning needs to distinguish the ability to recognize terms, explain their meanings, and relate them to the context of the PAI material. However, the research document does not include a list of terms or examples of question items that specifically measure these abilities. This relationship is a theoretical implication of the findings, not evidence that the teacher has formulated all of these indicators in his instrument. In relation to the understanding of Arabic terms in PAI learning, this alignment needs to be translated into specific indicators of ability. (Nation, 2013a) explains that knowledge of a word has several dimensions and levels; recognizing the form of a word is not the same as understanding the meaning and being able to use it. This framework is relevant for interpreting the assessment of Arabic terms because learners can recognize or memorize religious terms without fully understanding the concepts they contain. Therefore, assessment planning needs to distinguish the ability to recognize terms, explain their meanings, and relate them to the context of the PAI material. However, the research document does not include a list of terms or examples of question items that specifically measure these abilities. This relationship is a theoretical implication of the findings, not evidence that the teacher has formulated all of these indicators in his instrument.

The adjustment of the difficulty level of the questions between grades III and VI shows that the teacher considers the differences in students' abilities. This consideration is related to the results of interviews which stated that students in lower grades still need intensive guidance. Interpretively, teachers realize that instruments cannot be applied uniformly without paying attention to learning readiness. However, the data has not shown the form of differences in questions, answer demands, or assessment criteria at the two levels. Therefore, the adjustments found are more appropriately understood as pedagogical considerations of the teacher, while the accuracy of the difficulty level cannot be ascertained through the analysis of the instruments. In the context of Arabic terms, planning can be developed through questions with concrete context and explanation demands that are according to the abilities of each class. The findings of the interviews also show that teachers plan an evaluation after all the material in a subject has been delivered. This statement indicates that the teacher's attention is mainly focused on checking the achievement of the target after the delivery of the material. However, observations show the presence of oral questions during learning to check students' understanding. Both data show the difference between assessment after discussion of the material and understanding examination that takes place in the learning process. According to Black and Wiliam (1998), assessment has a formative function when the information obtained is used to adjust learning. Thus, the oral questions found have the potential to support formative functions, but their existence has not proven that formative assessment has been systematically planned from the beginning.

Planning for the use of test and non-test techniques shows the teacher's efforts to obtain information from more than one source. Tests in the form of thematic written questions and tests are used to see students' answers to the material, while non-test techniques are in the form of observations of activities, attitudes, and responses during learning. The two techniques have the potential to complement each other, but the objects assessed need to be differentiated. Attentiveness, activeness, or willingness to answer are information about learning engagement; these indicators do not in themselves prove understanding of Arabic terms. Active learners do not

necessarily provide precise explanations, while less active learners do not necessarily understand the material. Therefore, planning needs to establish evidence of understanding separately from evidence of attitude and participation. An important limitation is found in the readiness of assessment tools. Documentation data states that grids are not displayed formally because it takes time to find them. These findings cannot be interpreted as the teacher does not have a grid at all; what can be stated is that the document was not yet available for examination at the time of data collection. As a result, the relationship between the objectives, indicators, question form, and material scope has not been verified in detail. In contrast to the conditions of the grid, the field records clearly state that non-test observations were conducted in general without structured observation instruments. This indicates that the readiness of the non-test devices is still limited and the recording of the results still depends on the sensitivity of the teacher.

In the framework of assessment alignment (Biggs, 1996), these limitations indicate the need to clarify the relationship between the targeted ability and the evidence collected. Observations to assess the understanding of Arabic terms can be directed at the accuracy of students in explaining the meaning or associating the term with the PAI material, accompanied by clear criteria. Attitudes and participation are recorded through separate indicators. The development is a recommendation based on the gaps found, not a description of the tools that have been used by the teacher. Thus, assessment planning at SDIT Salsabila Al Muthi'in has shown a relevant basis through attention to the material, learning targets, grade level, and the use of test and non-test techniques. These strengths still need to be supported by indicators of explicit Arabic term comprehension, evaluable grids, and structured observation instruments. This conclusion is limited by the character of the data that still describes the assessment of PAI in general. Therefore, the study has not been able to confirm that the planning found has specifically and thoroughly measured the understanding of Arabic terms.

### Carry out test and non-test techniques to assess students' understanding of Arabic terms in PAI learning

The results of the study showed that PAI teachers at SDIT Salsabila Al Muthi'in used test and non-test techniques in the implementation of assessments in grades III and VI. The test technique was carried out through thematic written questions and daily tests, while the non-test technique was carried out through observation of students' activities, attitudes, and responses during learning. In addition, the teacher asked oral questions to check the understanding of the newly explained material. These findings indicated that the teacher obtained information about the students through written answers and direct interaction in the classroom. However, the available data still describe the PAI assessment in general, so its relationship with the understanding of Arabic terms needs to be interpreted in a limited way. The implementation of the written test showed the teacher's efforts to check the mastery of the material based on the students' answers. The interview stated that the questions were adjusted to the material that had been taught, while the documentation recorded the existence of evaluation questions and daily test scores for grades III and VI. This conformity is in line with the concept of *constructive alignment* from (Biggs, 1996), namely the relationship between learning objectives, learning activities, and assessment tasks. In the context of this study, the use of questions related to the material provides a basis for the assessment of learning achievement. However, the suitability of the content of the questions is not enough to ensure the depth of understanding measured because the data does not present examples of question items or students' answers in detail.

When associated with the understanding of Arabic terms, the written test needs to provide opportunities for students to demonstrate an understanding of meaning, not just the recognition or memorization of terms. (Nation, 2013b) explains that knowing a word covers various aspects and levels of knowledge, including the relationship between form and meaning as well as receptive and productive knowledge. Based on this framework, the ability to recognize religious terms cannot be equated with the ability to explain the concepts contained in it. The implication is that test answers

need to be examined based on the accuracy of the meaning and their relationship to the context of the PAI material. However, the research has not provided evidence that the thematic questions used have distinguished the ability to memorize terms and understand their meaning. Oral questions given during learning are an additional means for teachers to find out students' understanding directly. Observations showed that teachers asked questions to several learners in turn and provided opportunities to ask questions when the material was not understood. This interaction allowed teachers to recognize learning difficulties without waiting for daily repetition. The findings relate to thinking (Black & Wiliam, 1998), which places assessment information as the basis for adjusting learning activities. Thus, questions and answers have the potential to perform formative functions when learners' answers are used to determine additional explanations. However, the oral questions found to be more appropriate are called informal comprehension checks because there is no evidence of the use of oral test devices with structured questions and scoring guidelines.

In the non-test technique, the teacher observes the activities, attitudes, and ways students respond to learning. This observation provides information about student involvement that is not entirely visible through written test scores. However, field records show that observations are still carried out in a glimpse and general manner without structured observation instruments. Interpretively, these conditions show that non-test assessments still rely on the teacher's impression and sensitivity. This limits the traceability of assessment because there are no indicator records that allow researchers to know what behaviors are observed and how the results are determined. These limitations are even more important when observations are used to assess understanding of Arabic terms. Responsiveness, attention to explanation, and orderly attitude are evidence of learning engagement, but do not automatically indicate the accuracy of understanding. The framework (Nation, 2013b) helps to clarify that knowledge of terms needs to be proven through responses that show the relationship between the term and its meaning. Therefore, observation of engagement needs to be distinguished from observation of evidence of understanding, for example the accuracy of learners relating terms to material explanations. This distinction is a need for development based on findings, not practices that have been documented in research.

The results of the observation noted that teachers did not show significantly different treatment in providing assessments. The available questionnaire data also contained approval of the statement that the evaluation was in accordance with the material and the assessment was given fairly. Both data support a positive perception of the implementation of the assessment. However, the questionnaire did not include the number of respondents or the distribution of answers so that it could not be used to conclude the perception of all students. Similarly, the treatment that seemed fair during observation was not enough to ensure the consistency of scoring without checking the assessment guidelines. The implementation of the two techniques also faced time constraints and differences in the level of understanding of students. The teacher explained that some PAI materials have a wide scope, while students in lower grades need more intensive assistance. These findings show that the implementation of assessments takes place in conditions of diverse learning needs. Implications, questions and observations need to be focused on important indicators so that the information obtained remains meaningful in the available time.

Thus, the implementation of assessments has combined written tests, comprehension checks through oral questions, and non-test observations. These combinations provide a variety of sources of information, but they are not enough to ensure that the understanding of Arabic terms has been assessed specifically and systematically. Reinforcement is needed through questions that require explanations of meaning, clear scoring guidelines, and observation recordings that distinguish conceptual understanding from students' attitudes and participation.

### **Utilizing the Assessment Results to Follow Up on Students' Difficulties**

The results of the study showed that PAI teachers at SDIT Salsabila Al Muthi'in used the assessment results to identify students who had achieved the learning target and who still needed help. Based on the interviews, students who had not reached the target received further guidance and additional attention on the next learning, especially in the lower grades. These findings show that the assessment results are used as the basis for learning assistance. However, the data have not shown the form of follow-up activities in detail, the schedule of implementation, or the change in students' abilities after receiving guidance. In this study, the use of evaluation results to determine students who need assistance shows the existence of this function. Thus, the tests given after the completion of the material can also produce information for further learning improvements. However, this formative function can only be traced from the teacher's information regarding the provision of guidance, while its effectiveness cannot be ascertained because there is no data on reassessment.

Identification of difficulties is carried out through information from written tests and observations during learning. Thematic questions and daily tests provide an overview of students' answers, while observations provide information about their activities and responses in class. The use of these two sources has the potential to help teachers understand their learning needs more completely. However, field records show that non-test observations have not used structured instruments. Therefore, the information from observations still needs to be strengthened with notes that explain specific difficulties, not just general impressions about students' activeness or attitudes. In the context of understanding Arabic terms in PAI learning, the identification of difficulties needs to be directed to aspects of abilities that have not been mastered. (Nation, 2013b) explained that knowledge of words has various dimensions and levels, so the ability to recognize terms is not always accompanied by the ability to understand their meaning. The framework provides a basis for distinguishing difficulties in recognizing terms, explaining meanings, and relating them to the context of the material. This distinction is important because each difficulty requires different assistance. However, the research data have not detailed learners' mistakes in specific Arabic terms. Therefore, the grouping of difficulties is a theoretical implication for the development of follow-ups, rather than a documented diagnosis outcome.

The provision of additional guidance and attention showed the teacher's response to the different needs of the students. The interviews revealed that students in lower grades still needed intensive mentoring, while observations noted the opportunity to ask questions and provide additional explanations when the material was not understood. Both findings indicated that teachers provided assistance through learning interactions. However, additional explanations during learning could not be directly equated with structured remedial programs. The data did not show the existence of a specific remedial design, success criteria, or documentation of the results. Therefore, the follow-up found was more appropriately understood as integrated assistance in subsequent learning. The accuracy of the mentoring also depends on its suitability with the goals that have not yet been achieved. (Biggs, 1996), through the concept of *constructive alignment*, emphasizes the relationship between learning objectives, learning activities, and assessment. If applied to follow-up, assistance needs to be directed at abilities that students have not been able to demonstrate. As an implication, the difficulty of explaining the meaning of the term requires activities that help the formation of understanding, not just the repetition of the same test. In this study, teachers have stated that there is guidance, but the alignment between the students' mistakes and the form of assistance cannot be examined because the details of the two are not available.

Findings regarding students' responses also need to be interpreted carefully. Teachers mentioned that diligent students tend to take evaluations fluently, while students who do not like to learn or often joke tend to have difficulties. This information shows teachers' perceptions of the relationship between learning engagement and ability to participate in evaluations. However, the study has not provided enough evidence to establish a lack of seriousness as the main cause of difficulties. Inappropriate answers may also be related to understanding concepts, ability to read

questions, or other assistance needs. Thus, follow-up needs to rely on examining errors and learning needs, so that assessment of behavior does not replace the diagnosis of understanding. Findings regarding students' responses also need to be interpreted carefully. Teachers mentioned that students who are diligent tend to take evaluations fluently, while students who do not like to learn or often joke tend to have difficulties. This information shows teachers' perceptions of the relationship between learning engagement and the ability to participate in evaluations. However, the study has not provided enough evidence to establish a lack of seriousness as the main cause of difficulties. Inappropriate answers may also be related to understanding concepts, ability to read questions, or other need for help. Thus, follow-up needs to rely on fault checking and learning needs, so that assessment of behavior does not replace a diagnosis of understanding.

Thus, the use of assessment results at SDIT Salsabila Al Muthi'in has shown an improvement orientation through the identification of students who have not reached the target and the provision of additional guidance. This strength still needs to be supported by recording specific difficulties, assistance that suits needs, and reassessment. In relation to the Arabic term, follow-up needs to be directed to evidence of understanding the meaning and its relationship with PAI material. This study shows that there are mentoring efforts, but it cannot conclude its success in improving students' understanding because development data after follow-up is not yet available.

## Conclusion

This study shows that the assessment of PAI learning at SDIT Salsabila Al Muthi'in includes planning based on materials, learning targets, and grade levels; implementation through thematic written tests, daily tests, informal oral questions, and non-test observations; and the use of results to provide guidance to students who have not reached the target. However, grids are not yet available for review when non-test research and observations have not used structured instruments. Follow-up also faces time constraints and differences in students' understanding levels. Because the data have not detailed the Arabic terms assessed, comprehension indicators, and ability development after mentoring, the effectiveness of the two techniques in assessing and improving understanding of Arabic terms cannot be ascertained. Therefore, the development of assessments needs to be directed at clear meaning indicators, structured test and observation instruments, and reassessments to determine the success of follow-up.

## References

- Abdurahman, A., & Erihadiana, M. (2024). Comprehensive assessment: Balancing test and non-test methods in Islamic education evaluation. *Jurnal Inovasi Pendidikan Agama Islam (JIPAI)*, 4(2), 95–108. <https://doi.org/10.15575/jipai.v4i2.40752>
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, 87. <https://doi.org/10.3389/educ.2019.00087>
- Azis, A., Abou-Samra, R., & Aprilianto, A. (2022). Online assessment of Islamic religious education learning. *Tafkir: Interdisciplinary Journal of Islamic Education*, 3(1), 60–76. <https://doi.org/10.31538/tijie.v3i1.114>
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594x.2010.513678>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364. <https://doi.org/10.1007/BF00138871>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303–317. <https://doi.org/10.1017/s0272263199002089>
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24–31. <https://doi.org/10.1111/j.1745-3992.2009.00151.x>
- Ijarmana, F., & Putra, E. D. (2021). Analisis kesulitan guru dalam menerapkan penilaian autentik di sekolah dasar negeri. *NATURALISTIC: Jurnal Kajian Penelitian Pendidikan Dan Pembelajaran*, 6(1), 1050–1059. <https://doi.org/10.35568/naturalistic.v6i1.1366>
- Mohamad, N., Nor Amin Samsun Baharun, M., Bin Wan Abd Rahman, A., & Mohd Ramli, Z. (2021). Receptive and productive Arabic academic vocabulary mastery: A study of students in Malaysian public universities. *International Journal of Asian Social Science*, 11(3), 154–167. <https://doi.org/10.18488/journal.1.2021.113.154.167>
- Mukmin, M., & Nuraini, N. (2024). Integrasi penilaian tes dan non-tes dalam Pendidikan Agama Islam: Menuju evaluasi holistik untuk pembelajaran berkelanjutan. *Journal of Instructional and Development Researches*, 4(5), 370–379. <https://doi.org/10.53621/jider.v4i5.384>
- Nafiati, D. A. (2021). Revisi taksonomi Bloom: Kognitif, afektif, dan psikomotorik. *Humanika*, 21(2), 151–172. <https://doi.org/10.21831/hum.v21i2.29252>
- Nation, I. S. P. (2013a). Knowing a word. In *Learning vocabulary in another language* (2nd ed., pp. 44–91). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656.004>
- Nation, I. S. P. (2013b). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Nurlitasari, A., & Hamami, T. (2023). Assessment as, for, of learning pembelajaran pendidikan agama Islam tingkat menengah atas. *Humanika*, 23(2), 225–234. <https://doi.org/10.21831/hum.v23i2.61406>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Puspitasari, I., Suryani, A. I., & Arifin, S. (2023). Pengembangan alat penilaian mata pelajaran Pendidikan Agama Islam dengan menggunakan Quizizz di sekolah menengah pertama. *Jurnal Basicedu*, 7(1), 816–827. <https://doi.org/10.31004/basicedu.v7i1.4678>
- Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen penilaian hasil pembelajaran kognitif pada tes uraian dan tes objektif. *Jurnal Papeda: Jurnal Publikasi Pendidikan Dasar*, 4(2), 139–148. <https://doi.org/10.36232/jurnalpendidikandasar.v4i2.2649>
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/bf00117714>
- Sawaluddin, S., & Muhammad, S. (2020). Langkah-langkah dan teknik evaluasi hasil belajar

- Pendidikan Agama Islam. *Jurnal PTK Dan Pendidikan*, 6(1).  
<https://doi.org/10.18592/ptk.v6i1.3793>
- Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, 103, 101602.  
<https://doi.org/10.1016/j.ijer.2020.101602>
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88.  
<https://doi.org/10.1177/026553220101800103>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.  
<https://doi.org/10.3102/0034654307313795>
- Subehi, R., & Sriyanto, S. (2021). Implementasi assessment of, for, dan as learning dalam pembelajaran daring PAI di SMPN 8 Purwokerto. *Alhamra Jurnal Studi Islam*, 2(2), 111–122.  
<https://doi.org/10.30595/ajsi.v2i2.10632>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, 168(1), 33–69. <https://doi.org/10.1075/itl.168.1.02web>
- Widiyanto, A., & Inayati, N. L. (2023). Penerapan evaluasi pembelajaran tes dan non-tes Pendidikan Agama Islam di sekolah menengah kejuruan. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 4(2), 307–316. <https://doi.org/10.31538/munaddhomah.v4i2.439>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087.  
<https://doi.org/10.3389/fpsyg.2019.03087>
- Yen, C. C., Lee, C. Y., Chen, G. D., Wang, J. H., & ... (2020). A Digital Reality Learning Environment with Instant Assessment on Learning with Body and Visual Interaction. ... *on Advanced Learning* .... <https://ieeexplore.ieee.org/abstract/document/9155883/>