

# Authentic Assessment in Islamic Education: A Study of Models, Instruments, and Implementation Factors

Nazarudin Feri Fariansyah, Muhammad Yusron Maulana El-Yunusi

Universitas Sunan Giri Surabaya, Indonesia

## ARTICLE INFO

### Article history:

Received 17 February 2023

Revised 4 Mei 2023

Accepted 13 June 2023

### Key words:

Authentic assessment,

Islamic religious education,

Affective assessment,

Spiritual assessment,

Assessment instruments,

Implementation,

Islamic religious education teachers.

## ABSTRACT

The evaluation of Islamic Religious Education (IRE) learning, which has largely focused on cognitive aspects, is considered incapable of measuring the affective and spiritual dimensions that are its main objectives. This academic paper, based on a literature review, aims to examine authentic evaluation models in IRE, focusing on their characteristics, instruments, and supporting and inhibiting factors for their implementation. The study was conducted using a qualitative approach with a thematic synthesis method of relevant literature. The results of the study show that authentic assessment offers principles such as relevance to real tasks and continuous assessment that can be applied to assess attitudes, worship practices, and morals. Various instruments such as attitude scales, observations, portfolios, and performance assessments can be developed by considering validity and reliability. The implementation of this model faces supporting factors such as school leadership and teacher competence, as well as obstacles such as teacher workload and the pressure of a standardized evaluation system. This paper concludes that the development of an authentic PAI evaluation model requires a systemic approach involving policy adjustments, strengthening teacher capacity, and building understanding with all stakeholders so that assessment truly reflects the success of comprehensive education.

## INTRODUCTION

Learning assessment is an integral component of the overall educational process, serving as both a measuring tool and a guide for improvement. In the contemporary educational paradigm, assessment is no longer viewed solely as a final, judgmental activity, but rather as part of a continuous and formative learning process. This paradigm shift has led to the emergence of an authentic assessment approach that seeks to measure students' abilities comprehensively in situations relevant to real life. Authentic assessment emphasizes the application of knowledge, skills, and attitudes in complex and meaningful tasks, so that it can describe actual learning achievements beyond mere mastery of facts or memorization.

In the domain of Islamic Religious Education (PAI), the learning objectives are not limited to cognitive aspects such as mastery of Fiqh, history, or memorization of verses. PAI systematically strives to prepare students to recognize, understand, appreciate, believe in, be pious, be noble, and

practice the teachings of Islam (Saputra, 2022). The more substantive objectives lie in the affective and spiritual realms, namely the formation of attitudes, character (morals), and the appreciation of Islamic values in everyday life. The formation of students' morality is a crucial mission of PAI, where academic success in PAI and the dynamics of involvement in the social environment, including social media, significantly influence the formation of character (Al Mursyidi & Darmawan, 2023). PAI aims for students to not only know about Islam but also to practice its teachings with full awareness and sincerity. Therefore, the success of PAI learning should be measured by the extent to which there are positive changes in the attitudes, worship habits, and good character behavior of students. This requires an evaluation model that is able to reach dimensions that are internal and difficult to observe directly.

The reality of PAI learning evaluation in many educational institutions is still often concentrated on measuring cognitive aspects that are easy to quantify. Multiple-choice or essay tests on

\* Corresponding author, email address: [yusronmaulana@unsuri.ac.id](mailto:yusronmaulana@unsuri.ac.id)

definitions, laws, and memorization are still the main instruments. Meanwhile, assessments of religious attitudes, consistency in worship, honesty, patience, social awareness, and other spiritual dimensions are often carried out in a highly subjective, impressionistic, and unsystematic manner. The gap between comprehensive learning objectives and partial evaluation instruments poses a serious problem, because what is measured will determine what is taught and prioritized by teachers and students. Narrow evaluation has the potential to reduce the meaning of religious education to mere theoretical knowledge divorced from practice. These risks reducing the relevance of religious education in the eyes of the community, which expects educational outcomes that can bring about real changes in attitudes and behavior, as reflected in the community's response to the objectives of studying Islamic religious education programmed (Majid et al., 2023).

The development of an authentic PAI learning evaluation model is an urgent necessity. The model needs to be specifically designed to capture the complexity and depth of Islamic religious education objectives, with a focus on valid and reliable measurement of affective and spiritual dimensions. This study seeks to explore, critique, and map various models and evaluation instruments that have been developed, both in the context of general education and religious education, which can be adapted to assess the attitudes, worship practices, and moral changes of students. Social education shapes a broader global awareness, in which understanding and appreciation of values, including religious values, become an important foundation (Hariani & Mardikaningsih, 2022). Thus, it is hoped that a conceptual and practical framework can be obtained that can enrich and refine evaluation practices in PAI learning.

A fundamental problem in PAI learning evaluation is the dominance of cognitive orientation, which neglects affective and psychomotor objectives. Evaluation must be based on actual facts (data and evidence), not manipulated or fabricated results (Muntu, 2022). However, many PAI teachers experience operational difficulties in designing and implementing assessments of attitudes and spirituality. These difficulties are partly due to the lack of clear guidelines and standardized instruments that can be used as references. As a result, attitude assessment is often reduced to a cursory observation of visible behaviors such as diligence in congregational worship or neatness in dress, without systematic data collection methods

such as structured observation, reflection journals, or portfolios that can record the internal development of students (Kusaeri, 2014). In fact, affective dimensions such as sincerity, devotion, or motivation to worship are complex constructs that require sophisticated measuring tools.

The second problem lies in the validity and reliability of the affective-spiritual assessment instruments used. Many instruments are developed ad hoc without undergoing a rigorous validation process, whether content, construct, or predictive validation. For example, attitude assessment questionnaires compiled by teachers may contain leading questions or questions that do not measure the intended construct. In addition, the tendency of respondents to give socially desirable answers (social desirability bias) in matters related to religion also poses a challenge. Students may answer based on what they consider to be normatively good and correct, rather than based on their actual practices and beliefs. This causes the data obtained to not reflect the real conditions, so that the policies or follow-up learning taken are not on target (Mueller, 2016).

Another challenge is that teachers' already heavy workloads often become a reason to avoid authentic assessment, which is seen as more complicated and time-consuming. Authentic assessment, such as continuous observation, portfolio assessment, or project-based assessment, does require more intensive planning, implementation, and analysis than correcting multiple-choice answer sheets. On the other hand, administrative demands and accountability to schools or parents often require clear and easily interpretable quantitative scores. The combination of time constraints, skill limitations, and administrative pressure traps teachers in their comfort zone, causing them to continue using conventional forms of assessment, even though they realize that these assessments are not comprehensive (McMillan, 2010). This situation is also related to legal guarantees of children's rights to quality education, including fair and comprehensive assessment, which are often not optimally fulfilled (Hariani et al., 2021). As a result, innovation in PAI evaluation is hampered.

This study is important because comprehensive and authentic evaluation is key to improving the quality and relevance of Islamic Religious Education. Learning evaluation should be carried out continuously and comprehensively in order to be effective and efficient so that educators can obtain a complete picture of the progress and learning outcomes achieved by students (Anwar, 2021).

Without an evaluation system capable of measuring affective and spiritual dimensions, it is impossible for educational institutions to know whether the main objective of PAI, namely the formation of people who are faithful, pious, and noble, has been truly achieved. Character building based on noble values, such as in the framework of building Pancasila character with religious harmony, requires assessment methods that can measure the internalization of values, not just knowledge about them (Hariyadi et al., 2023). The development of valid and reliable instruments to measure these aspects will provide accurate and accountable data on the real impact of PAI learning on students' lives. This data can then be used to improve the curriculum, learning methods, and educational policy as a whole.

This study can contribute to the development of Islamic education, particularly in the field of educational evaluation. By mapping and analyzing various existing models and instruments, this study can produce a synthesis or recommendations on the most appropriate authentic evaluation model for the characteristics and objectives of Islamic Education. The findings of this study are expected to serve as a guide for teachers, curriculum developers, and Islamic education researchers in designing, implementing, and researching more meaningful and behavior-oriented learning evaluations. Ultimately, this effort is aimed at realizing religious education that not only educates the mind but also shapes the hearts and behavior of students in accordance with Islamic values.

This study aims to conduct a systematic literature review to examine and map authentic PAI learning evaluation models, with a focus on developing instruments to measure affective and spiritual dimensions. The study seeks to identify relevant principles of authentic evaluation, analyses various types of assessment instruments that have been used, and explore factors that influence their implementation. Theoretically, this study is expected to enrich the scientific knowledge of Islamic education evaluation by providing a critical and integrative review. Practically, the resulting synthesis is expected to be a reference for educators and managers of Islamic educational institutions in designing and implementing a more comprehensive, fair, and meaningful evaluation, so that it is truly capable of measuring the success of true learning.

## **RESEARCH METHOD**

This study utilizes a qualitative literature review approach. This approach was chosen because it is

appropriate for exploratory and interpretative research objectives, namely to gain an in-depth understanding of the concepts, models, and instruments of authentic assessment in Islamic Religious Education. Qualitative literature study allows researchers to delve into existing academic texts, analyses arguments, identify patterns, and build conceptual syntheses without collecting primary data (Creswell & Poth, 2016). Through this approach, various perspectives and findings from previous studies were critically reviewed to produce a comprehensive and nuanced understanding of the challenges and opportunities for developing evaluations that focus on the affective and spiritual dimensions.

The main method applied in this analysis was thematic synthesis. This method provides a systematic framework for integrating findings from various qualitative and conceptual literature. According to Thomas and Harden (2008), thematic synthesis is carried out in three sequential stages. The first stage involves comprehensive coding of the relevant literature content, in which text passages related to the problem formulation are coded. The second stage involves organizing these codes into interrelated groups to form descriptive themes. Descriptive themes are still very close to the original data from the literature reviewed. The third stage is the analytical stage, where descriptive themes are further developed into analytical themes that go beyond the content of the primary literature to produce new interpretations and answer research questions.

The application of the thematic synthesis method in this study was carried out using a rigorous procedure. Literature searches were conducted in academic databases such as Google Scholar, ERIC, JSTOR, and ScienceDirect using a combination of keywords: "authentic assessment Islamic education", "affective assessment PAI", "spiritual evaluation", "character assessment instrument", and "authentic evaluation PAI". The main inclusion criteria were publications that directly discussed learning evaluation, particularly the affective/spiritual aspects, and were relevant to Islamic education or general education that could be adapted. Each document that met the criteria was then read critically, and important parts were noted and coded. The analysis process was iterative, continuously comparing and revising codes and themes. To ensure the validity of the analysis, Bowen's (2009) qualitative document analysis framework was used as a reference to maintain transparency and intellectual discipline during the coding and

interpretation process. Thus, the resulting synthesis was not only descriptive but also provided meaningful interpretative contributions to the development of the field of PAI evaluation studies.

## RESULT AND DISCUSSION

### Characteristics and Principles of Authentic Assessment Can Be Applied to Measure Affective and Spiritual Dimensions in Islamic Religious Education Learning

In religious education assessment, authentic evaluation has become an increasingly relevant approach to capture the deeper dimensions of learning. The application of authentic assessment to measure the affective and spiritual dimensions in Islamic Religious Education requires a deep understanding of its main characteristics, the first of which is its suitability for real-world tasks or performance. This principle requires that assessment not be conducted in a vacuum, separate from daily religious practices. From the teacher's perspective, this assessment includes understanding and teaching to assess, address, and evaluate spiritual care, while from the student's perspective, it includes perceptions, concerns, and also how to appreciate and provide spiritual care (Sarrión-Bravo, 2022). In PAI, this means that evaluation instruments must be designed to observe and document how students apply Islamic values in authentic life situations. For example, rather than simply asking for the definition of congregational prayer, authentic evaluation can take the form of an observation project and student participation in congregational prayer at the school mosque or home environment, accompanied by a reflection journal about the experiences and challenges encountered. This approach allows teachers to assess not only procedural knowledge, but also the attitude, consistency, and spiritual appreciation of the worship. Assessment becomes more meaningful because it is directly linked to the actual learning objectives, namely to familiarize oneself with and appreciate worship (Mueller, 2016). This characteristic emphasizes that evaluation must reflect living and real religious practices.

In addition to its connection to the real world, authentic evaluation also requires a more comprehensive view of student development. The second crucial characteristic is an emphasis on both process and product. Authentic evaluation looks not only at the final result, but also at how students achieved that result. For affective dimensions such as developing patience or gratitude, the internal process and efforts to be consistent are far more important than static final achievements.

Understanding individual processes, including emotional regulation, is a key factor in overcoming challenges in learning and developing the full potential of learners (Noviyanti & Darmawan, 2023). Instruments such as development portfolios are particularly appropriate here. A learner may be asked to collect evidence in a portfolio that shows their development journey, such as notes on efforts to control anger, photos of them participating in social activities, or short video recordings of their sunnah worship practices. The assessment of this portfolio will focus on progress, self-reflection, and commitment to self-improvement over time. This method is in line with the Islamic concepts of *mujahidin an-nafs* (struggle against desires) and *muhasabah* (self-evaluation), so that evaluation becomes part of the character education process itself (McMillan, 2010). This approach places evaluation as a means of spiritual guidance, not merely a tool for measuring results.

In order for qualitative assessments to be accountable, systematic and structured instruments are required. The third principle is the importance of clear and transparent assessment criteria, which are often manifested in the form of rubrics. Rubrics set quality standards for various levels of achievement in a task or performance. In assessing spiritual dimensions such as the sincerity of prayer or the sincerity of charity, which are in fact very subjective, rubrics serve as a tool that makes assessments more objective and measurable. The use of fair and clear rubrics is also a form of implementing the Islamic human capital framework, which emphasizes fairness and accountability in managing and accessing human potential (Darmawan, 2021). A rubric for assessing reflective reports on fasting, for example, may include criteria such as the depth of analysis of the temptations faced, honesty in revealing weaknesses, and the quality of improvement plans for the next Ramadan. Each criterion is described in descriptors for different levels of achievement, for example, from "beginning to develop" to "very good". By using rubrics, teachers have consistent guidelines, and students understand the expectations that must be met. This transparency reduces the impression of subjectivity and increases fairness in the assessment of the affective-spiritual domain (Airasian & Russell, 2001). The clarity of these criteria strengthens the legitimacy of evaluation in a personal and spiritual domain.

In practice, authentic evaluation does not stand alone but is integrated into the overall learning process. Authentic evaluation is carried out not only to measure students' level of ability in the field of

knowledge, but also to assess their mentality and abilities (Achmad & Prastowo, 2022). In addition, authentic evaluation emphasizes continuous assessment integrated with learning. This characteristic is particularly relevant because the formation of attitudes and spirituality is a long process that cannot be measured with a single test. Assessment must be carried out formatively throughout the semester or academic year, through various methods such as systematic observation, anecdotal notes, and direct dialogue. PAI teachers can use participant observation techniques to record changes in students' social behavior in the school environment, such as how they resolve conflicts or help friends. These notes are then discussed in confidential and constructive personal guidance sessions (*mau'izhah*). Thus, evaluation is no longer a specter, but rather a tool for mutual guidance and reflection that supports student growth. This integration ensures that the measurement of affective-spiritual dimensions is truly part of the educational process, not just an administrative ritual at the end of the period (Kusaeri, 2014). This characteristic reinforces the function of evaluation as a continuous coaching process.

Changes in the approach to evaluation have a direct impact on the redefinition of the role of teachers. The application of these principles requires a shift in the role of teachers from examiners to facilitators and learning colleagues. Teachers are required to have keen observation skills, empathetic communication skills, and sensitivity to detect subtle changes in students' attitudes. Teachers also need to spend more time interacting individually, analyzing evidence of progress, and providing constructive feedback. This role is in line with the concept of teachers in Islam as *marabi*, educators who not only teach but also nurture and touch hearts. In authentic assessment, a *marabi* will use assessment data to understand the spiritual condition of students and provide targeted guidance, rather than labelling or ranking them (Hidayat et al., 2018). This transformation of roles makes assessment part of a meaningful pedagogical relationship.

Theoretically, the authentic assessment approach has a strong foundation in educational assessment studies. From an evaluation theory perspective, this approach is in line with the CIPP (Context, Input, Process, Product) model developed by Stufflebeam, particularly in the process and product components. Authentic evaluation allows teachers to continuously evaluate the affective-spiritual learning process (such as participation in value discussions, involvement in charity projects)

and its products (observable behavioral changes). Data from process evaluation is used to make immediate improvements, while product data is used to assess the overall success of the programmed. Thus, authentic evaluation serves a dual purpose: as a formative diagnostic tool and as a more comprehensive summative tool. This framework provides a strong theoretical basis for developing a coherent PAI evaluation system oriented towards continuous improvement. This alignment confirms that authentic evaluation has scientific and practical legitimacy.

The implementation of this approach also requires policy and institutional system support. The managerial implications of applying these characteristics are significant. First, educational institutions need to develop evaluation policies that formally recognize and allocate time and resources for authentic assessment. Without this structural support, teachers will find it difficult to devote time to in-depth observation and individual feedback. Second, it is necessary to develop a bank of standard instruments and rubrics for various affective-spiritual aspects of PAI that can be adapted by teachers. Providing practical examples and portfolio templates will greatly assist teachers in getting started. Third, the learning outcome reporting system must be revised to accommodate and communicate achievements in the affective-spiritual domain to parents in a meaningful way, not only in the form of numbers, but also in the form of narrative descriptions of the development of students' attitudes and character. These steps ensure the sustainability and consistency of authentic assessment implementation.

In addition to its benefits, authentic assessment also requires a high degree of ethical sensitivity. The implementation of authentic assessment also poses ethical challenges that need to be considered. Assessment of personal and spiritual domains must be carried out with great care, maintaining confidentiality and avoiding stigmatization. The aim is to guide, not to judge. This vigilance is important so that evaluation practices do not inadvertently reinforce social stereotypes that can shape opportunities and inequalities, including in educational settings (Sajjapong et al., 2022). Therefore, teachers must be equipped with an understanding of assessment ethics and empathetic communication techniques. Ultimately, the application of authentic assessment characteristics and principles in PAI is an effort to respect the complexity and sanctity of religious education objectives. By measuring what is truly important,

namely changes in attitude, spiritual depth, and moral improvement, the assessment system can become a true partner in realizing the vision of Islamic Religious Education that is transformative and has a real impact on the lives of students. The importance of interventions and support systems that break the cycle of similar obstacles is also found in other studies, for example, in efforts to break the cycle of poverty that limits children's access to quality education (Hariani & Masnawati, 2022). This entire explanation confirms that authentic evaluation serves as an instrument for the continuous development of faith and character.

### **Types and Forms of Valid and Reliable Evaluation Instruments that have been Developed to Assess the Attitudes, Worship Practices, and Moral Changes of Students**

In the evaluation of Islamic Religious Education learning, the selection of assessment instruments plays a crucial role. The process of gathering information by teachers regarding the development and achievement of learning carried out by students is done through various instruments that are able to express, prove, or show accurately that the learning objectives have been fully mastered and achieved (Ubaidillah et al., 2021). Various types of evaluation instruments have been developed and adapted to assess the affective and spiritual dimensions in Islamic education, with efforts to meet validity and reliability standards. One of the most widely used and developed instruments is a psychological scale based on Likert's theory. This scale is designed to measure the intensity or frequency of an attitude, belief, or religious practice. In Islamic education, examples include the Muslim Attitude Towards Religion Scale (MATRS) developed by Wilde and Joseph (1997) or adaptations of religiosity scales such as those derived from Glock and Stark's dimensions of religiosity. These scales generally contain statements such as "Praying five times a day is a top priority in my daily life" or "I feel calm after reading the Qur'an", with answer options ranging from "Strongly Disagree" to "Strongly Agree". The validity of the content is maintained through reviews by experts in the fields of Islamic psychology and religious education, while reliability is tested using methods such as Cronbach's Alpha. The advantage of this instrument lies in its ease of administration and statistical analysis, allowing it to be used for large-scale surveys. However, its main weakness is its vulnerability to social desirability bias, where students may choose answers that are considered normatively good rather than those that reflect the

actual conditions (Hill & Hood, 1999). This condition indicates that the use of attitude scales needs to be understood critically so that the assessment results are not misinterpreted.

In response to the limitations of self-report instruments, other approaches have been developed to strengthen the objectivity of assessment. To overcome the weaknesses of self-report scales and capture more objective data, systematic observation instruments have become an important option. These instruments require trained teachers or observers to document specific behaviors of students based on predetermined operational indicators. An example is the Checklist for Observing Islamic Behaviors or an observation rubric for assessing worship practices such as congregational prayer. The indicators can cover physical aspects (punctuality, perfection of movements) and attempts to approach spiritual aspects (devotion as seen from calmness). Observations are conducted periodically and in various situations to obtain a more comprehensive picture and reduce subjectivity. The validity of this instrument is highly dependent on the clarity of the definition of the behavior being observed and the training of the observer. Inter-rater reliability needs to be tested to ensure consistency in assessment. Techniques such as time sampling or event sampling can be applied to make observations more structured and measurable. The advantage of observation is its ability to capture actual behavior, but it requires a significant investment of time and commitment from teachers (Mueller, 2016). This emphasizes that systematic observation requires resource readiness in order for the results to be accountable.

In addition to observation, assessments that emphasize process and reflection are also gaining attention. An authentic assessment tool that is becoming increasingly popular is portfolio assessment. A portfolio in PAI is a collection of evidence of work and reflections by students that is systematically compiled to show the development of attitudes, worship practices, and morals over a period of time. Its contents can be very diverse, such as photos of participation in charity activities, video recordings of presentations on the values of a verse, weekly spiritual reflection journals, memorization certificates, or the results of interviews with parents about behavioral changes at home. The effectiveness of this assessment is also influenced by the learning environment and student independence at home, which have been shown to affect learning motivation and engagement in the educational process (Irawan & Darmawan, 2023).

The main strength of portfolios lies in their

ability to document the process and progress, not just the final results. Its validity is contextual and authentic because it is directly related to the student's work. The reliability of the assessment is enhanced by the use of clear analytical rubrics to assess various components of the portfolio. However, portfolio assessment requires teachers to be able to provide in-depth feedback and facilitate student reflection. The management and storage of physical or digital documents is also a logistical consideration that needs to be addressed (Arifin, 2011). With these characteristics, portfolios place students as active subjects in the evaluation process.

Practice-oriented assessment is also an important part of PAI evaluation. In addition, performance-based assessment techniques are highly relevant for assessing worship practices. This instrument evaluates students' ability to demonstrate a skill or procedure in a situation that resembles real-life conditions. Examples include assessing the practice of prayer, ablution, reciting the Qur'an with tartil, or presentations on solving ethical problems based on Islamic principles. Assessment usually uses a detailed scoring rubric that covers aspects such as procedure, understanding, and quality of presentation. Content validity is maintained by aligning performance tasks directly with learning objectives. Reliability is achieved through teacher training in the use of rubrics and, where possible, the involvement of more than one assessor. The advantage of performance assessment is its high suitability for the applied learning objectives of PAI. However, its implementation takes longer than written tests, especially for classes with a large number of students (McMillan, 2010). This situation requires careful planning to ensure that performance assessment remains effective and proportionate.

This situation requires careful planning to ensure that performance assessment remains effective and proportionate. It is important to note that some authentic assessment methods and tools, such as rubrics, portfolios, or self- and peer-assessment, do not always fully represent authentic activities and can be categorized as both authentic and traditional assessments. Assessment activities should reflect students' ability to use their knowledge outside the academic environment (Sokhanvar et al., 2021). To assess more subtle aspects of character and behavioral change, instruments such as anecdotal notes and self-reflection journals are of significant value. Anecdotal notes are objective descriptions of important events or behaviors observed by teachers, recorded immediately after the event. These notes focus on behavioral facts, not interpretations.

Meanwhile, reflection journals provide space for students to explore their thoughts, feelings, struggles, and spiritual progress privately. This process is in line with the multidisciplinary educational goal of fostering social competence, which requires space for self-expression and deep reflection (Hariyani et al., 2021). The validity of these instruments lies in the depth and richness of the qualitative data they produce. Its reliability is more about consistency and discipline in documentation. Both instruments are invaluable because they can capture spontaneous learning moments and internal developments that are invisible to other instruments. The challenge is that the data generated is highly personal and subjective, requiring trust and a good relationship between teachers and students. The analysis is also more qualitative and interpretative (Kusaeri, 2014). This characteristic shows that qualitative instruments require ethical sensitivity and professionalism from teachers.

To broaden the perspective of assessment, a multi-source approach has been developed. In an effort to improve validity, several approaches have begun to develop 360-degree assessment instruments that involve various sources of assessment. In this model, the assessment of students' attitudes and character comes not only from teachers, but also from peers, self-assessment, and even parents. For example, to assess attitudes of responsibility and cooperation, peer assessment forms can be used after completing group projects. Feedback from parents about behavioral changes at home is also very valuable data. This approach can provide a more comprehensive picture and reduce bias from a single source. Its validity is increased because the construct is measured from various perspectives. However, good coordination and socialization are necessary so that all parties understand the objectives and appropriate assessment methods. The integration of data from various sources can then become rich material for dialogue between teachers, students, and parents to plan the next steps for development (Fitzpatrick et al., 2012). This approach positions assessment as a continuous collaborative process.

A theoretical foundation is needed to integrate the various instruments used. The evaluation theory underlying this diversity of instruments is the concept of triangulation. Triangulation in evaluation refers to the use of various methods and data sources to examine the same phenomenon, in this case the affective-spiritual development of students. By combining data from attitude scales, observations, portfolios, and reflections, teachers can build a

stronger, more valid, and reliable understanding of student achievement. Data from one method can complement or confirm findings from another method. For example, a high score on the honesty attitude scale supported by anecdotal notes about honest behavior in difficult situations will provide greater confidence than relying on only one source of data. Therefore, the best recommendation in developing an authentic PAI evaluation system is not to choose one "best" instrument, but to design a system that combines several instruments complementarily to obtain a complete picture. This framework emphasizes the importance of integrating methods in order to gain a comprehensive understanding of student development.

The diversity of evaluation instruments has consequences at the institutional level. The managerial implications of this diversity of instruments require educational institutions to take several strategic steps. First, an assessment development team needs to be formed to select, adapt, or develop instruments that are in line with the vision, mission, and specific objectives of the institution. Second, investment in teacher training must focus on technical mastery of the use of various instruments, including the preparation of rubrics, valid observation techniques, and qualitative data analysis. Third, the school management information system needs to be supported by a platform that enables the collection, storage, and analysis of diverse authentic assessment data, both quantitative (scores) and qualitative (narratives, recordings). Strategies to improve teacher performance through motivation and work discipline are an important foundation for ensuring the effective and sustainable implementation of these various evaluation instruments (Sulipah & Mardikaningsih, 2023). With the support of policies, teacher capacity, and adequate infrastructure, the application of these valid and reliable evaluation instruments can transform the landscape of PAI assessment to be more fair, meaningful, and truly reflective of the success of education in shaping individuals with character and spirituality. Thus, it is emphasized that the success of PAI evaluation is highly dependent on the synergy between policy, educator competence, and a sustainable support system.

### **Supporting and Inhibiting Factors in the Implementation of Authentic Assessment Models**

The discussion of authentic assessment in Islamic Education needs to begin with the context of institutional policy and leadership. The implementation of authentic assessment models in

Islamic Education learning faces a number of interrelated supporting and inhibiting factors, starting at the institutional policy and leadership level. The most decisive supporting factor is the strong commitment of educational leaders, such as school principals or Islamic boarding school leaders. The effectiveness of educational change is highly dependent on the ability of school principals to mobilize all resources to improve quality (Ismaya et al., 2023). When leaders have a clear vision of the importance of comprehensive assessment and make authentic assessment a programmed priority, the allocation of resources will follow. This support can be realized in the form of explicit curriculum policies, the provision of budgets for teacher training and instrument development, and the creation of a school climate that encourages pedagogical innovation. Transformative leadership is able to build collective awareness among teachers about the urgency of a paradigm shift in assessment, from simply measuring memorization to measuring changes in attitudes and behavior (Fitzpatrick et al., 2012). Without support from top management, change initiatives initiated by a handful of teachers will find it difficult to survive and spread into school culture. Discussions about authentic assessment in PAI need to begin with institutional policy and leadership.

The emphasis on the role of leadership becomes even more relevant when viewed from the dynamics of education policy. The implementation of authentic assessment models in PAI learning faces a number of interrelated supporting and inhibiting factors, starting at the level of institutional policy and leadership. The most decisive supporting factor is the strong commitment of educational leaders, such as school principals or Islamic boarding school leaders. When leaders have a clear vision of the importance of comprehensive assessment and make authentic assessment a programmed priority, resource allocation will follow. This support can be realized in the form of explicit curriculum policies, the provision of budgets for teacher training and instrument development, and the creation of a school climate that encourages pedagogical innovation. Transformative leadership is able to build collective awareness among teachers about the urgency of shifting the assessment paradigm, from simply measuring memorization to measuring changes in attitudes and behavior (Fitzpatrick et al., 2012). Without support from top management, change initiatives initiated by a handful of teachers will find it difficult to survive and spread into school culture. This description emphasizes that

consistency in policy and leadership vision determines the sustainability of evaluation innovation.

Beyond supporting factors, various structural constraints also influence the success of implementation. On the other hand, factors that hinder the success of character education include the negative influence of technology, limited supervision by madrasah institutions, and weak cooperation between madrasahs and parents (Syarnubi et al., 2021). The main inhibiting factors often stem from macro education policies that still emphasize standardization and accountability based on numbers. The national examination system or school examinations that are oriented towards multiple-choice tests for Islamic Education subjects create enormous pressure for teachers and students. Teachers feel trapped in a dilemma: on the one hand, they understand the importance of affective assessment, but on the other hand, they are required to prepare students for exams whose format ignores this aspect. As a result, teachers tend to teach and assess for exams (teaching to the test), which hinders the exploration and application of more time-consuming authentic evaluation methods. This policy creates a misalignment between the ideal objectives of PAI education and the nationally implemented evaluation system, thus becoming a structural disincentive to implementing change (Kusaeri, 2014). Pressure from parents, who often focus on numerical scores and academic rankings, further reinforces this obstacle. Policy barriers such as these limit teachers' scope for developing meaningful assessments.

In addition to policy, individual teacher factors are important determinants of practice in the field. At the individual teacher level, crucial supporting factors are professional competence and intrinsic motivation. Teachers who have a deep understanding of the concept of authentic evaluation, technical skills in developing rubrics, conducting systematic observations, and providing constructive feedback will be more confident in implementing it. Teachers' intrinsic motivation, which stems from the belief that their duty is to shape the character of their students, is a strong driver for overcoming technical complexities. Teachers who see themselves as *marabi* (educators) will be more motivated to find ways to assess the inner development of their students than teachers who see themselves only as instructors of material. Early successful experiences, such as seeing the positive impact of portfolio feedback on student learning motivation, can further strengthen teachers' commitment to continue developing authentic

assessment practices (McMillan, 2010). Thus, developing teachers' capacity and enthusiasm is a key investment. This factor emphasizes that changes in evaluation are highly dependent on teachers' readiness and professional orientation.

However, teachers' readiness often comes face to face with the reality of their workload. The biggest obstacles at the teacher level are often high workloads and time constraints. The implementation of authentic assessments such as continuous observation, portfolio assessment, or individual conferences does require a much greater allocation of time than simply correcting multiple-choice answer sheets. PAI teachers who may teach many classes with dozens of students in each class will find it difficult to find the time to assess each student's work in depth. In addition, many teachers are not familiar with qualitative data analysis and descriptive narrative writing as a form of assessment report. These limitations often cause teachers to revert to conventional assessment methods that are more practical and time-efficient, even though they are aware of their limitations. Without a reduction in teaching load or the addition of adequate incentives, fatigue and time constraints will become real physical barriers to the sustainability of authentic assessment practices (Arifin, 2011). This situation shows that structural support is urgently needed to maintain the sustainability of authentic assessment practices.

Beyond individual teachers, the professional environment also provides significant supporting resources. Potential external supporting factors include support from professional communities and partnerships with universities. The existence of active Teacher Working Groups (KKG) or Subject Teacher Working Groups (MGMP) for Islamic Education can serve as a forum for sharing experiences, problems, and solutions related to authentic assessment. In forums such as these, teachers can jointly develop instruments, agree on rubrics, and provide feedback to one another. Partnerships with education or *tarbiyah* faculties from local universities can also provide a fresh injection of theoretical and methodological knowledge. Master's or doctoral students can be involved in classroom action research or instrument development, resulting in practical, evidence-based products for teachers. This kind of support network overcomes the professional isolation often experienced by teachers and provides intellectual resources to overcome the technical challenges of implementation (Brookhart, 2013). The existence of a professional community strengthens teachers' capacity

to deal with the complexities of authentic assessment.

Conversely, external support is not always automatically available. External barriers arise from a lack of understanding and support from parents and the community. Parents who are accustomed to traditional assessment systems may question the validity and fairness of assessments that appear subjective, such as observations or portfolios. They may consider numerical test scores to be more "real" and comparable than narrative descriptions of their child's behavioral development. If communication between the school and parents is not effective, resistance from parents may arise and put additional pressure on teachers to revert to old methods. Therefore, intensive and transparent communication with parents about the philosophy, objectives, and mechanisms of authentic assessment is a step that should not be overlooked. Schools need to demonstrate how these new methods provide a complete and more useful picture of their children's development than a mere string of numbers (Mueller, 2016). Without the understanding and support of parents, efforts to reform evaluation in schools may lose one of their important pillars. Parental involvement is a determining factor in the social legitimacy of authentic evaluation.

Analysis of these factors can be deepened through the theoretical framework of educational change. From the perspective of educational change theory, the implementation of authentic assessment can be analyzed using the Concerns-Based Adoption Model (CBAM) framework developed by Hall and Hord. This theory explains that in adopting an innovation, teachers experience various stages of concern, ranging from self-concern (e.g., "Can I do it?") then concern for the task (e.g., "how much time will it take?"), to finally concern for the impact (e.g., "how will this affect the students?"). Barriers often arise in the early stages when teachers are still anxious about the consequences of innovation for themselves. Supporting factors, such as good training and colleague support, are needed to help teachers through these stages. Successful implementation requires sensitive monitoring of teachers' concerns and the provision of appropriate support at each stage, rather than simply instructions or pressure from above (Hall & Hord, 2006). This humane and gradual approach is more likely to result in sustainable change than a rigid instructional approach. This framework helps to understand that teacher resistance is a natural part of the change process.

These findings have direct implications for school management. The managerial implications of

analyzing these factors are clear and demand a comprehensive approach. First, school-level policies must be designed to create alignment between goals, learning, and evaluation, and to provide teachers with time to carry out authentic assessment. Second, continuous professional development programmed must focus on improving teachers' assessment competencies, accompanied by intensive coaching, rather than merely one-way training. Third, the learning outcome reporting system must be reformed to communicate affective-spiritual achievements meaningfully to all stakeholders, including parents. Fourth, a school culture that values process and development, not just final results in the form of numerical scores, needs to be built through recognition and appreciation of non-academic achievements. These steps show that authentic evaluation requires consistent and structured managerial support.

This entire discussion emphasizes that authentic evaluation is a systemic change agenda. The successful implementation of the authentic assessment model in PAI does not depend on a single factor, but on the synergistic interaction between visionary leadership, supportive policies, adequate teacher competence, support from colleagues and parents, and a conducive school culture. Overcoming obstacles and strengthening supporting factors requires patience, long-term commitment, and collaboration from all parties. By strategically managing these factors, transforming PAI learning evaluation from a cognitive-centered approach to an authentic one is not impossible, but rather a necessity to realize religious education that is truly meaningful and transformative for students' lives. This conclusion emphasizes that the success of authentic evaluation is the result of collective work oriented towards long-term change.

## CONCLUSION

Based on a systematic review of the literature, it can be concluded that the development of an authentic assessment model for Islamic Religious Education (IRE) learning is both a necessity and a complex challenge. First, the characteristics of authentic assessment, such as relevance to real-world tasks, emphasis on process, use of transparent rubrics, and continuous assessment, provide a strong operational framework for measuring affective and spiritual dimensions in a more meaningful way than conventional tests. Second, there are various instruments that can be developed with consideration of validity and reliability, ranging from psychometric scales, systematic observation,

development portfolios, performance assessments, to anecdotal notes and reflection journals, all of which complement each other in the principle of data triangulation. Thirdly, the successful implementation of this model is greatly influenced by the interaction of supporting factors such as visionary leadership, teacher competence, and professional community support, with inhibiting factors including teacher workload, pressure from number-oriented policies, and a lack of understanding from parents and the wider system.

The findings of this study have important implications for policy, school management, and teacher professional development. At the policy level, adjustments to national and local evaluation standards are needed to provide formal space and recognition for authentic assessment in PAI, reducing the misalignment between educational goals and the assessment system. For school management, the implications require strategic planning that includes the allocation of time and budget resources for teacher training, the development of a bank of instruments, and the reform of a reporting system capable of communicating affective-spiritual achievements in a descriptive and meaningful manner. For teachers and teacher education institutions, these findings emphasize the need for continuous capacity building in the field of authentic assessment, not only as a technical skill but also as an integral part of the educational philosophy of character and spirituality building.

Based on a series of analyses, several suggestions are put forward. First, for further research, it is recommended to conduct research and development (R&D) to produce prototype models and instruments for authentic PAI evaluation specific to various levels of education (MI/MTs/MA), equipped with implementation guidelines and empirical validity-reliability test results. Second, for the Ministry of Religious Affairs and the Education Office, the main recommendation is to initiate a pilot programmed or demonstration of the application of authentic PAI evaluation in a number of partner schools/madrasahs, accompanied by intensive mentoring and formative evaluation to develop policies that are more evidence-based. Third, for PAI teacher professional organizations (such as KGPAI/ASPAI) and LPTK, it is recommended to make authentic assessment literacy a core subject in ongoing activities such as workshops, seminars, and teacher training curricula, thereby building a community of practitioners who support and innovate with one another.

## REFERENCES

- Achmad, G. H., & Prastowo, A. (2022). Authentic Assessment Techniques on Cognitive Aspects in Islamic Religious Education Learning at Elementary School Level. *Jurnal Ilmiah Sekolah Dasar*, 6(1), 75-84.
- Airasian, P. W., & Russell, M. K. (2001). *Classroom Assessment: Concepts and Applications* (6th ed.). McGraw-Hill, New York.
- Al Mursyidi, B. M. & D. Darmawan. (2023). The Influence of Academic Success of Islamic Religious Education and Social Media Involvement on Student Morality. *Al-Fikru: Jurnal Ilmiah*, 17(2), 321-331.
- Anwar, R. N. (2021). Evaluasi Pembelajaran Daring Pendidikan Agama Islam di Perguruan Tinggi Umum. *Proceeding Umsurabaya: Conference of Elementary Studies*.
- Arifin, Z. (2011). *Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur*. Remaja Rosdakarya, Bandung.
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27-40.
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD, Alexandria.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th ed.). Sage Publications, Thousand Oaks.
- Darmawan, D. (2021). Islamic Human Capital Implementation as Effective Framework in Islamic Human Capital Management. *Journal of Social Science Studies*, 1(1), 239-244.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). *Program Evaluation: Alternative Approaches and Practical Guidelines* (4th ed.). Pearson, Boston.
- Hall, G. E., & Hord, S. M. (2006). *Implementing Change: Patterns, Principles, and Potholes* (4th ed.). Pearson, Boston.
- Hariani, M., Aliyah, N. D., & Issalillah, F. (2021). Legal Guarantee of Children's Rights in Education and Health. *Journal of Social Science Studies*, 1(2), 177-180.
- Hariani, M., & Mardikaningsih, R. (2022). The Social Education Role in Shaping Students' Global Awareness in Higher Education. *Journal of Social Science Studies*, 2(1), 55-60.
- Hariani, M., & Masnawati, E. (2022). Assessing the Impact of Poverty on Children's Access to and Quality of Education: Implications for Interventions and the Breaking of the Poverty Cycle. *Journal of Social Science Studies*, 2(2), 1-6.
- Hariyadi, A., Jenuri, J., Darmawan, D., Suwarma, D.

- M., & Pramono, S. A. (2023). Building of the Pancasila Character with Religious Harmony in the Globalization Era. *AL-ISHLAH: Jurnal Pendidikan*, 15(2), 2126-2133.
- Hidayat, T., Rizal, A. S., & Fahrudin, F. (2018). Pendidikan dalam Perspektif Islam dan Peranannya dalam Membina Kepribadian Islami. *Jurnal Mudarrisuna: Media Kajian Pendidikan Agama Islam*, 8(2), 218-244.
- Hill, P. C., & Hood, R. W. (Eds.). (1999). *Measures of Religiosity*. Religious Education Press, Birmingham.
- Irawan, A. I. & D. Darmawan. (2023). Examining How Family Environment and Learning Independence Affect Student Learning Motivation at MI Babussalam Krian Sidoarjo. *Jurnal Cahaya Mandalika*, 4(2), 1324-1333.
- Ismaya, B., S. Sutrisno, D. Darmawan, J. Jahroni, & N. Kholis. (2023). Strategy for Leadership: How Principals of Successful Schools Improve Education Quality. *Al-Tanzim: Jurnal Manajemen Pendidikan Islam*, 7(1), 247-259.
- Kusaeri. (2014). *Acuan dan Teknik Penilaian Proses dan Hasil Belajar dalam Kurikulum 2013*. Ar-Ruzz Media, Yogyakarta.
- Majid, A. B. A., Shofiyah, R., Masnawati, E., Masfufah, M., & Safira, M. E. (2023). The Society's Response to the Intentions of Studying the Islamic Religious Education Program. *International Journal of Service Science, Management, Engineering, and Technology*, 4(3), 7-12.
- McMillan, J. H. (2010). *Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction (6th ed.)*. Pearson Education Canada, Toronto.
- Mueller, J. (2016, 1 October). *Authentic Assessment Toolbox*. <http://jfmuller.faculty.noctrl.edu/toolbox/index.htm>
- Muntu, N. (2022). Evaluasi Pembelajaran PAI dan Budi Pekerti pada MTs dan SMP di Kota Bitung. *Journal of Islamic Education Policy*, 7(2), 130-158.
- Noviyanti, D. V. & D. Darmawan. (2023). The Role of Academic Tutoring and Emotion Regulation in Mitigating Learning Boredom Among Students at SMA Khairunnas Gunung Anyar Surabaya. *Jurnal Cahaya Mandalika*, 4(2), 1334-1342.
- Sajjapong, T., Darmawan, D., & Marsal, A. P. (2022). The Role of Social Stereotypes in Shaping Opportunities and Inequalities in Society: Their Impact on Education, Employment, and Intergroup Interactions. *Bulletin of Science, Technology and Society*, 1(1), 44-49.
- Saputra, A. (2022). Strategi Evaluasi Pembelajaran Pendidikan Agama Islam Pada SMP. *Jurnal Genta Mulia*, 13(2), 73-83.
- Sarrión-Bravo, J. A., González-Aguña, A., Abengózar-Muela, R., Renghea, A., Fernández-Batalla, M., Santamaría-García, J. M., & Ruiz-Moral, R. (2022, October). Competence in Spiritual and Emotional Care: Learning Outcomes for the Evaluation of Nursing Students. *In Healthcare*, 10(10), 2062.
- Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of Authentic Assessment for Improving the Learning Experience and Employability Skills of Higher Education Students: A Systematic Literature Review. *Studies in Educational Evaluation*, 70, 101030.
- Sulipah, S., & Mardikaningsih, R. (2023). Strategies to Improve Teacher Performance through Motivation and Work Discipline. *Bulletin of Science, Technology and Society*, 2(3), 37-43.
- Syarnubi, S., Mansir, F., Purnomo, M. E., Harto, K., & Hawi, A. (2021). Implementing Character Education in Madrasah. *Jurnal Pendidikan Islam*, 7(1), 77-94.
- Thomas, J., & Harden, A. (2008). Methods for the Thematic Synthesis of Qualitative Research in Systematic Reviews. *BMC Medical Research Methodology*, 8(45), 1-10.
- Ubaidillah, A. F., Yenuri, A. A., & Fitriyani, F. (2021). Multiculturalism Of Authentic Assessment in Islamic Religious Education in Indonesia. *Turkish Online Journal of Qualitative Inquiry*, 12(6), 5536-5546.
- Wilde, A., & Joseph, S. (1997). Religiosity and Personality in a Moslem Context. *Personality and Individual Differences*, 23(5), 899-900.