

Reconstructing Assessment in Islamic Religious Education: A Multicultural Framework for Diversity Awareness and Social Cohesion

Nurul Latifatul Inayati^{1*}, Musa Asy'arie², Sekar Ayu Aryani³, and Muthoifin⁴

^{1,4} Universitas Muhammadiyah Surakarta, Indonesia

^{2,3} Universitas Islam Negeri Sunan Kalijaga, Indonesia

*nl122@ums.ac.id

ARTICLE INFO

Article history

Received April 5, 2026

Revised May 22, 2026

Accepted June 20, 2026

Keywords: RE Evaluation;
Multiculturalism; CIPP Model;
Religious Moderation; Social
Harmony

ABSTRACT

Current assessment practices in Islamic Religious Education (IRE) remain largely dominated by cognitive and textual measures, limiting their capacity to evaluate students' attitudes, behaviours, and social engagement in multicultural contexts. This study aims to analyse the limitations of conventional IRE assessment and reconstruct a multicultural-based evaluation framework that promotes diversity awareness and social cohesion. Using a Systematic Literature Review (SLR) guided by the PRISMA protocol, this study examined 51 Scopus-indexed articles published between 2020 and 2025. The findings reveal that conventional assessment models tend to be superficial because they emphasise conceptual understanding while neglecting affective, behavioural, and social dimensions. The review also shows that adaptive multicultural learning strategies, including contextual, dialogic, collaborative, and authentic approaches, are more effective in cultivating tolerance and inclusive attitudes. This study proposes a reconstructed CIPP-based assessment framework encompassing Context, Input, Process, and Product evaluation. The framework contributes to IRE evaluation theory by shifting assessment from academic measurement toward transformative evaluation of value internalisation, inclusive practice, and social harmony.

1. INTRODUCTION

Education is a critical institutional mechanism for sustaining peaceful coexistence in societies characterised by cultural, ethnic, linguistic, and religious plurality. This function is particularly significant in Indonesia, a nation comprising more than 300 ethnic groups and at least 718 regional languages, whose diversity represents both a substantial form of social capital and a persistent governance challenge. Although Indonesia's Religious Harmony Index increased from 76.02 in 2023 to 76.47 in 2024, the score remains below the "very high" category, suggesting that interreligious tolerance, equality, and cooperation require continued institutional reinforcement. Within this context, schools occupy a strategic position because they shape not only students' knowledge but also the values, dispositions, and social competencies through which diversity is interpreted and negotiated (Aydin et al., 2026). The educational urgency of this issue is reinforced by quantitative evidence indicating that inclusive attitudes have not been consistently internalised among Indonesian students (Tur & Yildiz, 2026). A national survey involving 2,358 upper-secondary students across 34 provinces found that 51.1% agreed that the chairperson of the student council should be Muslim, while 78.5% believed that the governmental system recognised by Islam was based on the caliphate. At the higher-education level, another national survey involving 2,866 students from more than 90 universities reported that 30.16% of respondents demonstrated low or very low levels of religious tolerance. These findings do not imply that Indonesian students are uniformly intolerant, but they reveal the persistence of exclusivist religious orientations that may

hinder equitable participation, recognition of difference, and constructive intergroup relationships (Ward et al., 2025). Islamic Religious Education (IRE) is therefore strategically positioned to address these challenges because it is expected to connect religious knowledge with ethical conduct, social responsibility, religious moderation, and respect for human dignity (Al Mubarak et al., 2023). In a multicultural society, IRE should enable students to understand Islamic teachings while simultaneously developing the capacity to engage constructively with people from different religious, ethnic, and cultural backgrounds. Nevertheless, the implementation of IRE frequently remains dominated by textual transmission, doctrinal explanation, and the reproduction of normative knowledge, with comparatively limited opportunities for dialogue, collaborative problem-solving, intercultural encounters, and critical reflection on social diversity (Farouq et al., 2025). This condition creates a persistent discrepancy between the transformative aspirations of religious education and the actual learning experiences through which students are expected to internalise multicultural values.

Recent scholarship has increasingly moved beyond exclusively doctrinal conceptions of IRE by examining the integration of multicultural values, religious moderation, experiential learning, and social engagement into curricula and classroom practices. Empirical studies suggest that multicultural Islamic education can strengthen tolerance, solidarity, and social cohesion when religious teachings are connected to dialogue, case-based learning, community participation, inclusive school cultures, and direct experiences of diversity (Nigar et al., 2025). Parallel developments in educational assessment have emphasised authentic, formative, and diagnostic approaches as alternatives to conventional examinations, particularly because these approaches can capture complex learning processes that cannot be adequately represented through written tests alone. Despite these advances, the existing body of research remains fragmented in several important respects. Studies on multicultural IRE have predominantly concentrated on curriculum content, teaching strategies, teacher roles, institutional practices, and the effects of instructional interventions on tolerance or religious moderation, while comparatively limited attention has been devoted to the architecture of assessment itself. Research on authentic and diagnostic assessment in Islamic education has generally addressed learning quality at a broad level but has rarely operationalised diversity awareness and social cohesion as measurable, multidimensional learning outcomes. Moreover, existing assessment practices tend to privilege cognitive achievement and summative examinations, whereas students' attitudes, intercultural dispositions, dialogical competencies, collaborative behaviours, and participation in diverse social settings remain insufficiently represented. The central problem addressed in this study is the structural misalignment between the multicultural objectives of Islamic Religious Education and the assessment systems used to determine whether those objectives have been achieved. IRE is expected to cultivate diversity awareness, tolerance, mutual recognition, cooperation, and social cohesion, yet its assessment practices continue to emphasise the recall and reproduction of religious knowledge (Santaliestra et al., 2024). As a result, conventional evaluation provides insufficient evidence regarding whether students can interpret religious differences ethically, communicate across social boundaries, respond to prejudice, cooperate with members of different groups, and translate religious values into inclusive social behaviour. This misalignment limits teachers' ability to diagnose students' developmental needs, provide meaningful formative feedback, and evaluate the broader social contribution of religious education. Addressing this problem requires an assessment framework that treats multicultural competence as a multidimensional construct encompassing cognitive understanding, affective orientation, ethical judgement, interpersonal competence, and observable social practice.

Accordingly, this study aims to identify the limitations of conventional assessment in multicultural IRE, synthesise assessment strategies and indicators relevant to diversity awareness

and social cohesion, and reconstruct an integrated assessment framework based on the Context, Input, Process, and Product dimensions. The study employs a systematic literature review guided by the PRISMA framework and analyses 51 scholarly articles indexed in the Scopus database. Through this process, the study examines how multicultural learning outcomes have been conceptualised, assessed, and implemented, and subsequently organises the available evidence into a comprehensive framework for evaluating cognitive, affective, behavioural, pedagogical, and contextual dimensions of IRE. This research is urgent because the effectiveness of multicultural religious education cannot be established solely from the presence of tolerance-related content in curricula or from students' ability to provide normatively acceptable answers in examinations. Without valid and context-sensitive assessment, schools may report successful learning even when students have not developed the dispositions and competencies required to live constructively with difference. Theoretically, this study advances the literature by connecting multicultural Islamic education, assessment theory, and programme evaluation within a unified analytical framework. Practically, the reconstructed framework may assist teachers, curriculum developers, school leaders, and policymakers in designing assessment systems that evaluate not only what students know about Islam and diversity but also how they interpret differences, make ethical decisions, interact with others, and contribute to social cohesion.

2. METHODS

This study employed a Systematic Literature Review (SLR) method following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol. The articles included in the review were published between 2020 and 2025. This approach was selected because it enables the systematic, transparent, and replicable synthesis of findings from multiple studies, thereby producing conclusions with a high level of reliability and credibility.

2.1 Data Sources and Search Strategy

The primary source of data was the Scopus database because it provides broad international coverage of peer-reviewed literature and enables systematic filtering by publication year, document type, subject area, and keywords. The search covered publications from 2020 to 2025 in order to capture recent developments in multicultural education, religious moderation, and assessment practices in Islamic Religious Education. The search strategy was developed by combining three clusters of keywords: (1) Islamic Religious Education, represented by terms such as "Islamic religious education," "Islamic education," and "PAI"; (2) multicultural and social values, represented by "multicultural," "religious moderation," "tolerance," "diversity," "pluralism," "social harmony," and "social cohesion"; and (3) assessment and curriculum terms, represented by "assessment," "evaluation," "learning evaluation," "curriculum," and "education evaluation." The search string was structured using Boolean operators and applied to the title, abstract, and keyword fields. A representative search string was: TITLE-ABS-KEY(("Islamic religious education" OR "Islamic education" OR PAI) AND (multicultural OR "religious moderation" OR tolerance OR diversity OR pluralism OR "social harmony" OR "social cohesion") AND (assessment OR evaluation OR "learning evaluation" OR curriculum)).

2.2 Article Selection Process (PRISMA Flow)

The article selection process was conducted through four stages in accordance with the PRISMA protocol, as illustrated in Figure 1.

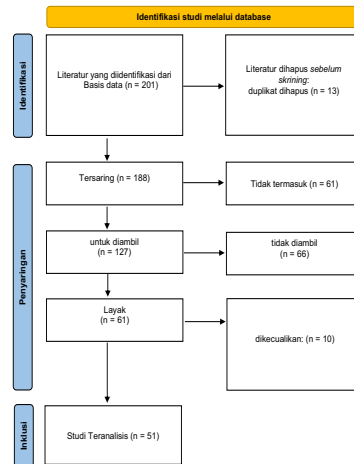


Figure 1. Process (PRISMA Flow)

During the identification stage, 201 records were retrieved from Scopus. After removing 13 duplicate records, 188 articles proceeded to title and abstract screening. At the screening stage, 61 records were excluded because their focus did not align with multicultural IRE assessment, religious moderation in education, or pedagogical evaluation. The remaining 127 records were sought for full-text retrieval; however, 66 reports could not be retrieved or did not provide sufficient full-text access. Thus, 61 full-text articles were assessed for eligibility. After full-text assessment, 10 articles were excluded because they did not meet one or more eligibility criteria, such as limited pedagogical relevance, absence of assessment or curriculum implications, or insufficient connection to multicultural values. Finally, 51 articles were included in the qualitative synthesis. This staged selection procedure was used to ensure that the final dataset was not merely thematically related to Islamic education in general, but specifically relevant to the reconstruction of assessment in Islamic Religious Education within multicultural contexts. The PRISMA flow was also used as an audit trail to make the decision-making process transparent and reproducible.

2.3 Inclusion and Exclusion Criteria

Eligibility criteria were established before the full-text review to ensure consistency in article selection. The criteria covered publication year, source, thematic relevance, pedagogical relevance, assessment orientation, document accessibility, and language. Articles were included only when they contributed evidence, concepts, or implications that could inform the reconstruction of multicultural IRE assessment. The detailed inclusion and exclusion criteria are presented in Table 1.

Table 1. Inclusion and Exclusion Criteria

Criterion	Inclusion Criteria	Exclusion Criteria
Publication period	Articles published between 2020 and 2025.	Articles published outside the designated period.
Database and indexation	Articles indexed in Scopus and retrievable through the database search.	Non-indexed sources or records outside the selected database.
Thematic relevance	Studies discussing Islamic Religious Education, Islamic education, religious moderation, multicultural education, tolerance, diversity awareness, or social cohesion in educational contexts.	Studies focusing only on theological, doctrinal, or textual debates without educational or multicultural relevance.
Pedagogical relevance	Studies addressing learning processes, curriculum development, assessment, evaluation, teacher practice, student attitudes, or educational policy implications.	Studies with no connection to teaching, learning, curriculum, or educational evaluation.
Assessment	Studies containing explicit or implicit	Studies that discuss multicultural

orientation	implications for assessment, learning evaluation, competency measurement, value internalization, or student behavior.	values but provide no evaluative, curricular, or pedagogical implications.
Document type	Peer-reviewed journal articles, review articles, and scholarly studies relevant to the research focus.	Editorials, book reviews, short notes, opinion pieces, or documents lacking scholarly analysis.
Full-text availability	Articles with accessible full text for detailed review and coding.	Records whose full text could not be retrieved or assessed.
Language	Indonesian- and English-language articles, with no initial language restriction during database searching.	Articles that could not be reliably interpreted or verified during full-text review.

2.4 Data Analysis

Data were analyzed using narrative synthesis and thematic analysis. Narrative synthesis was used to explain patterns across studies, while thematic analysis was used to identify key dimensions for reconstructing multicultural IRE assessment. The analysis proceeded through four steps: (1) familiarization with the included articles; (2) identification of themes related to conventional assessment limitations, multicultural learning strategies, and value internalization; (3) organization of themes into the CIPP categories of context, input, process, and product; and (4) formulation of an assessment framework that integrates cognitive, affective, psychomotor, and social domains.

3. RESULTS AND DISCUSSION

3.1 Critique of Conventional Evaluation Models

The synthesis of the reviewed literature indicates that conventional assessment models in Islamic Religious Education (IRE) are no longer sufficient to address the educational demands of multicultural societies. Although conventional evaluation has contributed to measuring students' mastery of religious concepts, doctrines, and textual knowledge, its scope remains limited when IRE is expected to cultivate diversity awareness, religious moderation, inclusive attitudes, and social cohesion (Setiawan et al., 2026). The main limitation lies not only in the use of written examinations, but also in the underlying evaluative paradigm that treats religious learning primarily as the acquisition of correct knowledge rather than as a transformative process involving cognition, affect, behaviour, and social interaction. Therefore, the critique of conventional IRE assessment must be directed at its epistemological orientation, methodological instruments, and practical consequences for multicultural education (Rostandi et al., 2025). First, conventional IRE evaluation is dominated by a cognitive-textual orientation. In many educational settings, assessment continues to emphasise students' ability to recall religious concepts, define normative values, explain doctrinal principles, or reproduce textbook-based answers. Such practices may be administratively efficient and relatively easy to standardise, but they provide only a partial picture of students' religious learning (Azkiya et al., 2025). The central problem is that multicultural competence cannot be reduced to conceptual familiarity with tolerance, moderation, or respect for diversity. A student may be able to explain the meaning of religious moderation in an examination while still demonstrating prejudice, exclusionary behaviour, or reluctance to engage constructively with people from different backgrounds. This gap between knowing and being reveals a serious validity problem in conventional assessment: the instrument measures what students remember, but not necessarily what they internalise or practise (Wakano et al., 2025). This critique is consistent with Chotimah, Qudsy, and Yusuf's finding that the implementation of religious moderation in Islamic educational management often remains confined to the textual realm. Their study shows that religious moderation may be formally present in policy documents and instructional materials, yet its implementation does not always reach the level of lived practice (Chotimah et al., 2025). In the

context of assessment, this means that evaluation may produce high academic scores without providing evidence that students have developed the capacity to recognise difference, manage disagreement, or act inclusively in social life. Consequently, conventional assessment risks creating an illusion of achievement: students appear successful because they perform well in tests, while the deeper goals of multicultural IRE remain unexamined. Second, conventional evaluation tends to separate religious knowledge from social behaviour. This separation is problematic because the objectives of IRE, especially in multicultural contexts, are inherently integrative. IRE does not merely aim to transmit religious information; it also seeks to shape ethical awareness, moral disposition, and responsible participation in society (Arifin et al., 2023). When assessment focuses narrowly on cognitive achievement, the affective, psychomotor, and social dimensions of religious learning become secondary or invisible. As a result, teachers may lack reliable evidence about whether students are developing empathy, openness, fairness, mutual respect, and willingness to cooperate across differences. In this sense, conventional assessment narrows the meaning of educational success and weakens the transformative mission of IRE. The problem becomes more serious when multicultural values are assessed only through declarative statements. For example, students may be asked to define tolerance, mention examples of religious moderation, or identify verses and hadith related to social harmony. These tasks are useful at the level of basic understanding, but they are insufficient for evaluating how students respond to actual diversity (Muzayanah & Ghony, 2025). Multicultural awareness requires the ability to interpret social complexity, recognise bias, listen to different perspectives, negotiate disagreement, and avoid discriminatory conduct. These capacities cannot be adequately measured through written tests alone. Therefore, an assessment model that prioritises textual answers over observable conduct fails to capture the multidimensional nature of multicultural learning.

Third, conventional IRE evaluation is often summative rather than developmental. Assessment is commonly conducted at the end of a chapter, unit, semester, or academic year, mainly to determine grades. This practice provides limited information about the process through which students gradually develop values, attitudes, and social competencies. Diversity awareness and inclusive behaviour are not formed instantly; they emerge through repeated interaction, reflection, feedback, and correction (Mustapa et al., 2025). A summative model cannot adequately detect whether students' attitudes are changing, whether classroom interaction is becoming more inclusive, or whether prejudicial assumptions are being challenged over time. Fourth, conventional evaluation makes limited use of authentic assessment. Written examinations, multiple-choice tests, and short-answer questions may be appropriate for measuring factual and conceptual knowledge, but they are inadequate for assessing ethical judgement and social practice. Multicultural IRE requires assessment instruments that allow students to demonstrate their competencies in meaningful contexts (Supriyatno & Ubabuddin, 2019). These may include behavioural observation, reflective journals, portfolios, peer assessment, collaborative projects, case analysis, community-based tasks, and records of participation in intergroup or intercultural activities. Without authentic assessment, teachers cannot adequately evaluate whether students are able to apply Islamic values in concrete situations involving diversity, conflict, cooperation, or social responsibility (Supriyanto, Amrin, 2022). Fifth, conventional evaluation often neglects the institutional and sociocultural context of schools. This is a significant weakness because multicultural IRE is always embedded in specific social realities. Schools differ in terms of student diversity, community relations, institutional culture, leadership commitment, curriculum orientation, and exposure to intergroup tension. A uniform assessment model that ignores these contextual differences may fail to identify the actual challenges faced by students and teachers (Takunas et al., 2024). For example, an assessment instrument designed for a religiously homogeneous school may not capture the complexities faced by students in a plural environment where daily interaction with different

religious or cultural groups is common. Sixth, conventional assessment is vulnerable to problems of construct validity and evidence sufficiency. Concepts such as tolerance, moderation, empathy, inclusion, and social cohesion are complex constructs. They cannot be measured reliably through a single test or a single source of evidence. A student's written response may indicate conceptual understanding, but it does not necessarily prove affective acceptance or behavioural consistency. Likewise, self-reported attitudes may be influenced by social desirability, especially when students know that tolerance is an expected moral answer (Azkiya et al., 2025). Teacher observation may provide richer contextual insight, but it may also be influenced by subjective judgement. Therefore, credible multicultural assessment requires triangulation across multiple sources, including tests, observations, portfolios, peer feedback, reflective writing, project work, and school records.

3.2 Adaptive Multicultural Learning Strategies

The literature synthesis indicates that the reconstruction of assessment in Islamic Religious Education (IRE) cannot be separated from the transformation of learning strategies. If conventional assessment is limited because it measures religious knowledge without adequately capturing value internalisation and social conduct, then adaptive multicultural learning strategies become the pedagogical foundation through which diversity awareness and social cohesion can be meaningfully developed (Muzayanah & Ghony, 2025). In this study, adaptive multicultural learning refers to instructional practices that respond to students' sociocultural realities, integrate Islamic ethical values with lived experiences of diversity, and create opportunities for students to practise tolerance, empathy, dialogue, and cooperation in authentic contexts. The first strategy is the contextualisation of IRE through local wisdom. The reviewed studies suggest that multicultural values become more pedagogically meaningful when they are not presented merely as abstract moral concepts but are connected to cultural practices familiar to students and communities. Rostandi, Busro, and Wasik's study on *Menyama Braya* in Bali, for example, demonstrates how local wisdom can be integrated into pesantren life through curriculum, language use, interfaith social interaction, and community-based cooperation (Rostandi et al., 2025). Their findings show that tolerance is strengthened not simply through classroom explanation but through everyday social practices such as *gotong royong*, interaction with Hindu teachers, and engagement with the surrounding community. This illustrates that local wisdom can function as a cultural bridge between Islamic values and plural social realities. The second strategy is dialogic and collaborative learning. Multicultural competence cannot be adequately formed through one-directional instruction because tolerance, empathy, and social cohesion require students to encounter different perspectives, articulate disagreement ethically, and negotiate meaning with others. Takunas et al.'s study of multicultural Islamic Religious Education learning at SMKN 1 Poso demonstrates that religious harmony was developed through both formal classroom learning and informal activities, including the "Religious Saturday" programme, which involved cross-cultural, socio-religious, and humanitarian collaboration (Takunas et al., 2024). This finding is important because it shows that multicultural learning does not occur only when diversity is discussed as content; it occurs when students participate in structured interaction across social and religious boundaries. Dialogic learning is therefore not merely a teaching technique but an ethical-pedagogical condition for multicultural IRE. Through dialogue, students learn to listen, clarify assumptions, recognise prejudice, and develop moral imagination toward those who are different (Mustapa et al., 2025). Collaborative learning extends this process by requiring students to work together on shared tasks, thereby transforming diversity from an object of discussion into a lived experience of cooperation. Nevertheless, dialogue and collaboration must be carefully facilitated. Without clear ethical norms, teacher sensitivity, and inclusive classroom structures, dialogue may become superficial, dominated by majority voices, or reduced to symbolic participation (Wakano et al., 2025). Likewise, group work cannot automatically be considered multicultural if students remain socially

segregated, if minority voices are marginalised, or if collaboration merely reproduces existing peer hierarchies.

The third strategy is strengthening teachers as multicultural agents. In multicultural IRE, teachers are not only transmitters of religious knowledge but also interpreters of diversity, facilitators of dialogue, role models of inclusive conduct, and evaluators of students' ethical-social development. Setiawan et al. identify five interconnected roles of *ustaz* in cultivating moderate Islamic values in pesantren: conservator, innovator, transmitter, transformer, and organizer (Setiawan et al., 2026). These roles show that teachers contribute to multicultural formation not only through instruction but also through institutional practice, modelling, programme design, and the organisation of character-strengthening activities (Arifin et al., 2023; Azkiya et al., 2025). The fourth strategy is the use of contextual and multicultural learning resources. Learning resources are not neutral instruments; they shape what forms of religious identity, cultural difference, and social belonging are made visible to students. Azkiya, Hasanuddin, and Hayati's development of a multicultural value-based e-module demonstrates that digital learning materials can support multicultural education when they are systematically designed, validated, and tested for practicality. Their study reports an overall expert validation score of 90.99%, educator practicality of 91.67%, and student practicality of 97.99%, indicating that well-designed multicultural resources can support classroom implementation (Azkiya et al., 2025). Even so, the use of e-modules, historical materials, or local cultural content should not be assumed to be inherently multicultural. A learning resource becomes multicultural only when it represents diversity accurately, avoids stereotyping, encourages critical reflection, and provides opportunities for students to apply values in meaningful contexts. For example, an e-module that merely inserts images of cultural diversity without inviting students to analyse inequality, prejudice, or social responsibility may remain superficial (Suharso et al., 2020). Similarly, historical heritage can become a powerful learning resource only when teachers guide students to interpret it as evidence of cultural encounter, religious coexistence, and shared civic life.

3.3 Reconstruction of Evaluation Components Using the CIPP Model

The reconstruction of Islamic Religious Education (IRE) assessment in multicultural contexts requires an evaluative framework that is capable of examining not only students' final learning outcomes, but also the contextual, institutional, pedagogical, and social conditions that shape those outcomes (Arifin et al., 2023). For this reason, the CIPP model—Context, Input, Process, and Product—provides a relevant framework for transforming IRE assessment from a narrow measurement of cognitive achievement into a comprehensive evaluation of diversity awareness, value internalisation, inclusive behaviour, and social cohesion. Unlike conventional assessment models that often position evaluation as a terminal activity conducted after instruction, the CIPP model understands evaluation as a continuous, decision-oriented, and improvement-based process. In this study, the CIPP model is not adopted merely as an administrative programme evaluation tool, but is reconstructed as a multicultural assessment framework. This reconstruction is necessary because diversity awareness and social cohesion cannot be validly assessed through written examinations alone. Students' attitudes toward difference, their ability to manage disagreement, their willingness to cooperate across cultural or religious boundaries, and their everyday inclusive practices are shaped by multiple factors: the diversity climate of the school, the readiness of teachers, the representativeness of learning materials, the quality of classroom interaction, and the consistency of institutional support (Setiawan et al., 2026). Therefore, a valid assessment of multicultural IRE must examine the entire educational ecology in which students' religious understanding and social dispositions are formed. The Context component functions as the diagnostic foundation of the reconstructed model. It identifies the sociocultural, institutional, and religious conditions within which IRE learning takes place. In conventional evaluation, context is

often treated as a background variable or ignored altogether (Cantika et al., 2026). This is problematic because multicultural competence is never formed in a vacuum. Students' understanding of tolerance, moderation, justice, and social harmony is influenced by the diversity of their peer environment, the inclusiveness of school policies, the representation of religious and cultural identities in the curriculum, and the broader community relations surrounding the school. Context evaluation should therefore go beyond demographic mapping. It should not merely record students' ethnic, cultural, or religious backgrounds, but should examine how diversity is experienced, regulated, represented, and negotiated within the school environment (Rostandi et al., 2025).

The Input component examines the readiness and adequacy of the resources required to implement multicultural IRE. These inputs include teacher competence, curriculum design, learning materials, assessment instruments, institutional support, and learning resources that reflect diversity (AN et al., 2024). In conventional assessment, these elements are often separated from student evaluation. As a result, weak student outcomes may be attributed solely to students' lack of understanding or poor attitudes, without examining whether the school has provided sufficient opportunities for multicultural learning (Prasetyo et al., 2023). From a critical perspective, input evaluation is essential because assessment validity depends on instructional validity. Students cannot reasonably be expected to demonstrate multicultural competence if they have not been exposed to inclusive materials, dialogic pedagogy, authentic learning tasks, and teachers who are prepared to facilitate sensitive discussions on diversity. The Process component represents the most important corrective to conventional IRE assessment because it examines how multicultural values are actually taught, practised, negotiated, and reinforced during learning. Conventional evaluation often focuses on final examination results, while the learning process through which students develop tolerance, empathy, and inclusive behaviour remains largely invisible (Azkiya et al., 2025). This creates a methodological gap: multicultural outcomes are expected, but the pedagogical experiences that produce those outcomes are not systematically evaluated (Husna et al., 2026). In a reconstructed framework, process evaluation should examine the quality of classroom interaction, the use of dialogic pedagogy, opportunities for students to express different perspectives, the ethical management of disagreement, the inclusiveness of group activities, and the extent to which learning encourages critical reflection on prejudice, exclusion, and social responsibility. The Product component assesses the extent to which multicultural IRE produces meaningful outcomes in students' knowledge, attitudes, behaviour, and social relations. In conventional assessment, product evaluation is often reduced to academic achievement, such as examination scores or mastery of religious concepts. In a multicultural framework, however, product evaluation must be broadened to include four interrelated domains: multicultural understanding, affective openness, inclusive behaviour, and contribution to social cohesion. Multicultural understanding refers to students' ability to interpret religious and cultural diversity in a contextual, fair, and non-reductive way (Takunas et al., 2024). Affective openness refers to empathy, respect, willingness to listen, and ethical sensitivity toward others. Product evaluation should employ authentic and multi-source instruments, including performance-based tasks, collaborative projects, behavioural portfolios, peer assessment, structured observation, reflective essays, attitude scales, teacher reports, counselling records, and documentation of student participation in intergroup activities (Setiawan et al., 2026). Written tests may still be used, but only as one component of a broader assessment system. They are useful for measuring conceptual understanding, but insufficient for validating whether students demonstrate inclusive behaviour in real or simulated social contexts. The main challenge in product evaluation is construct validity.

3.4 An Integrated CIPP-Based Conceptual Framework for Multicultural IRE Assessment

The synthesis of the reviewed studies indicates that the principal weakness of conventional Islamic Religious Education assessment lies not merely in the limited range of instruments employed, but in the absence of an integrated evaluative logic connecting educational context, programme readiness, instructional processes, and socially meaningful outcomes (Nizam et al., 2025). Existing assessment practices frequently isolate student achievement from the institutional and sociocultural conditions through which such achievement is produced. Consequently, low levels of tolerance, weak intercultural interaction, or exclusionary behaviour may be attributed solely to individual students, even when these outcomes are partly shaped by inadequate teacher preparation, monocultural learning materials, restrictive school policies, or learning environments that provide few opportunities for meaningful engagement across differences.



Figure 2. An integrated CIPP-based framework for multicultural IRE assessment

As presented in Figure 2, the proposed framework integrates the four CIPP components within a cyclical and mutually dependent architecture. Context, Input, Process, and Product are not treated as discrete administrative categories, but as analytically connected dimensions that explain how multicultural educational outcomes are generated, constrained, and improved (Demirci-Ünal & Öztürk, 2025). The framework begins with the multicultural and socioreligious environment because assessment cannot be regarded as culturally neutral or institutionally detached. Student diversity, school inclusion policies, institutional culture, and community-level tensions influence what should be assessed, whose experiences are recognised, and which forms of participation are considered legitimate. Context evaluation therefore establishes the normative and empirical baseline against which all subsequent assessment decisions must be judged. The Context component directly informs Input evaluation. Teacher competence, curricular content, learning materials, assessment instruments, and institutional support should not be treated as generic resources whose adequacy can be determined independently of the identified sociocultural needs (Margatama et al., 2024). Their quality must be assessed according to their capacity to respond to particular forms of diversity, inequality, prejudice, and interaction. For example, the availability of IRE textbooks or teaching modules is insufficient if those materials present diversity only as an abstract moral principle while avoiding contested social realities. Likewise, teacher mastery of religious content cannot be regarded as adequate professional readiness if teachers lack the intercultural sensitivity, dialogical competence, and evaluative literacy required to facilitate and assess learning in diverse classrooms (Yu et al., 2025). The framework therefore conceptualises Input as a condition of possibility rather than a procedural checklist. Weak inputs are likely to compromise the validity of subsequent assessments because students cannot reasonably be expected to demonstrate multicultural competencies that have not been adequately taught, modelled, practised, or institutionally supported. This point is important because conventional evaluation often interprets low student performance as evidence of individual deficiency without

examining whether the instructional conditions were sufficient to produce the expected outcomes (Charoensilp, 2024). The proposed model corrects this asymmetry by requiring evaluators to assess whether teachers, curricula, learning resources, and school structures provide credible opportunities for the development of diversity awareness and inclusive conduct. In this sense, assessment validity depends not only on instrument quality but also on the quality of the educational opportunities preceding measurement.

The Process component constitutes the central mediating mechanism within the framework. It is through dialogic learning, collaborative inquiry, formative assessment, reflection, and sustained intercultural engagement that contextual needs and educational inputs are translated into student development (Williams et al., 2022). This process orientation challenges the assumption that multicultural values can be internalised through declarative instruction alone. Students may be able to reproduce definitions of tolerance, moderation, or social harmony while remaining unable to negotiate disagreement, recognise implicit bias, challenge exclusion, or cooperate across social boundaries. Within this model, formative assessment has a particularly important function because diversity awareness and inclusive behaviour develop gradually and unevenly. End-of-semester examinations provide limited insight into how students respond to conflict, revise prejudicial assumptions, or transfer ethical understanding into social action. Continuous assessment enables teachers to identify contradictions between students' stated values and observable conduct, as well as changes that occur across different situations and over time. Reflective journals, dialogue records, peer feedback, collaborative tasks, and structured observations can therefore provide evidence of developmental trajectories that conventional tests cannot capture (Aktan, 2025). The Product component extends evaluation beyond cognitive achievement to encompass multicultural understanding, openness and empathy, inclusive behaviour, and social cohesion. These outcomes are presented in Figure 2 as connected but not interchangeable (Çapçı & Durmuşoğlu, 2022). Multicultural understanding refers primarily to students' ability to interpret religious and cultural plurality in a contextual and non-reductive manner. Openness and empathy concern affective and ethical orientations toward others, while inclusive behaviour refers to observable practices of cooperation, recognition, and non-discrimination. Social cohesion represents the broader relational and institutional outcome emerging when inclusive dispositions are translated into sustained patterns of trust, participation, and constructive interaction.

4. CONCLUSION

This study demonstrates that the current Islamic Religious Education (IRE) evaluation model, which predominantly focuses on cognitive aspects, has not been able to adequately assess behavioral change and social harmony as the primary objectives of multicultural education. To address these limitations, this study reconstructs a CIPP-based (Context, Input, Process, Product) evaluation model that enables a more comprehensive assessment of the learning context, readiness, processes, and outcomes. The main theoretical contribution of this article lies in proposing an expanded evaluative function, shifting IRE evaluation from merely measuring learning outcomes toward a transformational evaluation system grounded in social cohesion. A synthesis of 51 articles reveals that the integration of local wisdom, dialogic-collaborative learning approaches, and the strengthening of teachers' roles are effective in fostering tolerant and inclusive attitudes when supported by authentic and multidimensional evaluation practices. These findings have important implications for teachers, school principals, and policymakers in developing a more holistic IRE evaluation system. However, this study is limited to data sources indexed in Scopus. Therefore, future research should expand the database coverage and conduct empirical validation of the proposed model through direct implementation in educational settings.

5. REFERENCES

- Aktan, O. (2025). Attitude towards refugee students 'the case of Duzce province.' *Globalisation, Societies and Education*, 23(3), 738–751. <https://doi.org/10.1080/14767724.2023.2225453>
- Al Mubarak, F. U., Rohmani, A. F., & Shobahiya, M. (2023). Transformasi Pendidikan Agama Islam: Tinjauan Aspek Psikologis, Filosofis, dan Sosiologis Materi PAI di SMP/MTs. *Iseedu: Journal of Islamic Educational Thoughts and Practices*, 7(2), 244–256.
- AN, A. N., . M., . M., & . W. (2024). Bibliometric Analysis of Islamic Education and Character Development in Religious Education Practices in Indonesia. *Pakistan Journal of Life and Social Sciences (PJLSS)*, 22(2). <https://doi.org/10.57239/PJLSS-2024-22.2.0086>
- Arifin, S., Aryani, S. A., & Prayitno, H. J. (2023). Improving The Professional Teacher Competence Through Clinical Supervision Based on Multicultural Values in Pesantren. *Nazhruna: Jurnal Pendidikan Islam*, 6(3), 386–402. <https://doi.org/10.31538/nzh.v6i3.4037>
- Aydin, S., Raihan, F., Corliss, T., & Alwyn, S. (2026). Governance-Embedded Character Education: Institutionalizing Adab in Islamic School-Based Management. *Islamic Pedagogy Evaluation Journal*, 1(1 SE-Articles), 1–16.
- Azkiya, H., WS, H., & Hayati, Y. (2025). E-Module Based on Multicultural Values: Development Strategy for Islamic Primary Education. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, 5(4 SE-Articles), 421–434. <https://doi.org/10.31538/munaddhomah.v5i4.1511>
- Cantika, S. N., Choiruddin, A. A., Anwar, A., & Faqihuddin, M. (2026). Reconceptualizing Learning Evaluation through Spiritual Psychology: Tawazun and Muhasabah Perspectives. *Islamic Pedagogy Evaluation Journal*, 1(1 SE-Articles), 33–46.
- Çapçı, S., & Durmuşoğlu, M. C. (2022). Investigating Preschool Teachers' Perceptions of Multicultural Competence. *Egitim ve Bilim*, 47(211), 27–45. <https://doi.org/10.15390/EB.2022.11046>
- Charoensilp, P. (2024). Intercultural Sensitivity as a Factor in Perceived Culturally Responsive Teaching of Teachers in Northern Thailand. *REFlections*, 31(1), 90–117. <https://doi.org/10.61508/refl.v31i1.271669>
- Chotimah, C., Qudsy, S. Z., & Yusuf, M. (2025). Superficial implementation of religious moderation in Islamic educational management. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2024.2442235>
- Demirci-Ünal, Z., & Öztürk, G. (2025). Is Need-Based Action Possible for Each Child?: Preschool Teachers' Differentiated Instruction Within Multicultural Classrooms. *Early Childhood Education Journal*, 53(7), 2699–2718. <https://doi.org/10.1007/s10643-024-01753-2>
- Farouq, Y., Khalil, M., & Yahya, T. I. D. (2025). Deconstructing Islamic Religious Education (IRE) Materials to Foster Multicultural Awareness Among Millennial Students. *Multicultural Islamic Education Review*, 3(2), 189–190. <https://doi.org/10.23917/mier.v3i2.10181>
- Husna, S. F. N., Prasanti, R. R., Amaliah, N., Husniyah, F. N., & Qameer, N. (2026). Assessing Digital Technology Integration in Islamic Religious Education: A CIPP Model Evaluation. *Islamic Pedagogy Evaluation Journal*, 1(1 SE-Articles), 17–32.
- Margatama, P., Ramdane, T., & Wildan, M. (2024). *Analyzing the Practice of Critical Thinking Skills for Islamic Education Students in the International Campus*. 12(2), 235–246. <https://doi.org/10.22219/progresiva.v12i02.29520>
- Mustapa, A., Hidayah, N., Chusniyah, T., Hanafi, Y., & Chaiwitornwanich, A. (2025). Designing Multicultural-Based Islamic Education to Counteract Student Radicalization in Indonesia. *Journal for Deradicalization*, 43, 93–139.
- Muzayanah, U. H., & Ghony, D. (2025). The Harmonization of multicultural Islamic education values in diversity: A case from Indonesia. *Edelweiss Applied Science and Technology*, 9(3), 1212–1226. <https://doi.org/10.55214/25768484.v9i3.5468>
- Nigar, N., Dovchin, S., Wilson, R., & Kostogriz, A. (2025). Hybrid professionalism: multilingual cosmopolitan identities of immigrant English teachers. *Journal of Multilingual and Multicultural Development*. <https://doi.org/10.1080/01434632.2025.2494644>
- Nizam, A. K., Khairan, M., Corliss, T., & Alwyn, S. (2025). Building Multicultural Understanding Through Student Competence: A Conceptual Framework. *Multicultural Islamic Education Review*, 3(2), 133–148. <https://doi.org/10.23917/mier.v3i2.12490>

- Prasetyo, W. H., Sunardjoko, B., Muhibbin, A., Mahadir Naidu, N. B., & Muthali'in, A. (2023). Promoting Digital Citizenship among Student-Teachers: The Role of Project-Based Learning in Improving Appropriate Online Behaviors. *Participatory Educational Research*, 10(1), 389–407. <https://doi.org/10.17275/per.23.21.10.1>
- Rostandi, U., Busro, B., & Wasik, A. (2025). From tradition to tolerance: Menyama Braya as a model for multicultural education in Islamic boarding schools. *Acta Scientiarum. Education*, 48, e75776. <https://doi.org/10.4025/actascieduc.v48i1.75776>
- Santaliestra, J. R. S., Costa, R. S., Maldonado, J. J. V., & Martínez-Domingo, J. A. (2024). A systematic review and meta-analysis of the evidence for mobile learning as developer of skills in a second language. *Porta Linguarum*, 2024(11), 109–125. <https://doi.org/10.30827/portalin.viXI.30143>
- Setiawan, A. I., Zaenal Abidin, Y., Rustandi, R., Sarbini, A., & Aziz, R. (2026). Transforming Religious Education Through Inclusivity: How Indonesian Pesantren Cultivate Moderate Islamic Values and Da'wah Practices. *Nazhruna: Jurnal Pendidikan Islam*, 9(1 SE-Articles), 70–92. <https://doi.org/10.31538/nzh.v9i1.406>
- Suharso, R., Setyowati, D. L., & Utomo, C. B. (2020). Teaching multiculturalism based on Islamic historical relics in Northern Java. *International Journal of Innovation, Creativity and Change*, 11(10), 445–463.
- Supriyanto, Amrin, S. (2022). The Role of Islamic Religious Education Teachers in Implementing Multicultural Education Based on Values of Local Wisdom in State Junior High School 15 Surakarta. *IMProvement*, 9(1), 65–81. <https://doi.org/https://doi.org/10.21009/Improvement.091.07>
- Supriyatno, T., & Ubabuddin, U. (2019). Internalization of multicultural values in learning islamic education. *Library Philosophy and Practice*, 2019, 1–7.
- Takunas, R., Mashuri, S., Tahang Basire, J. H., Dulumina, G. B., & Mohi, S. M. (2024). Multicultural Islamic Religious Education Learning to Build Religious Harmony. *Nazhruna: Jurnal Pendidikan Islam*, 7(3), 590–607. <https://doi.org/10.31538/nzh.v7i3.18>
- Tur, S., & Yildiz, A. (2026). Primary School Teachers' Experiences Regarding Children's Rights and Multicultural Education. *International Online Journal of Education and Teaching*, 13(1), 223–232.
- Wakano, A., Ernas, S., Sai'dah, U., Lapele, F., & Prihono, E. W. (2025). Beyond Rhetoric And Resistance: Institutional Pathways Of Multiculturalism In Islamic Higher Education. *Jurnal Pendidikan Islam*, 11(2), 293–306. <https://doi.org/10.15575/jpi.v11i2.50392>
- Ward, F., Cho, J., Jung, J. K., Tognocchi, C., Beadles, T., & Cheon, J. (2025). ChatGPT for the intellectual soul: a Deweyan perspective on AI-based multicultural classroom praxis. *Multicultural Education Review*, 17(1), 1–18. <https://doi.org/10.1080/2005615X.2025.2467759>
- Williams, B. K., Smith, N., & Williams, N. N. (2022). Removing Cultural Boundaries and Transforming Our World with Multicultural Texts. *Journal of Negro Education*, 91(4), 530–540.
- Yu, S., Gil, M., Lyu, S.-R., & Son, K. (2025). Multicultural Awareness and Educational Continuity Among Early Childhood and Elementary School Teachers in South Korea. *International Journal of Early Childhood*. <https://doi.org/10.1007/s13158-025-00431-7>