

# Who assesses morality, and how? A systematic review of research gaps in Akidah Akhlak assessment



Andi Kurnia Muin<sup>a</sup>  | Wasis<sup>a</sup>  | Rivo Panji Yudha<sup>a</sup> 

<sup>a</sup>Postgraduate School, Doctoral Program in Educational Research and Evaluation, State University of Surabaya, East Java, Indonesia.

**Abstract** Moral and character assessment has gained increasing attention in educational research, yet systematic examination of who assesses morality and how assessment is operationalized remains limited. This systematic review addresses these questions with particular attention to Akidah Akhlak assessment marginalization within Islamic education. Following PRISMA 2024 guidelines, we systematically searched Scopus for peer-reviewed publications (2015-2025) addressing moral, character, and values-based assessment. From 443 retrieved records, 187 articles underwent detailed thematic synthesis examining assessor configurations, methodological approaches, and research gaps. Bibliometric analysis characterized publication trends and disciplinary patterns. The review revealed fivefold growth in moral assessment publications (18 in 2015 to 90 in 2025), occurring predominantly within secular Western frameworks. Only one study (0.2%) explicitly addressed Akidah Akhlak assessment—where Akidah refers to Islamic creed and belief system, and Akhlak refers to Islamic moral character and ethics. Methodologically, quantitative survey-based approaches dominated (30.0%), with minimal authentic assessment methods (rubrics 0.5%, portfolios 0.5%). Critically, 98.6% of studies failed to specify assessor identity, revealing fundamental ambiguity regarding who evaluates moral learning. Multirater configurations were entirely absent. Akidah Akhlak marginalization reflects structural dynamics including database indexing biases and disciplinary concentration favoring psychology over religious education. Current practices demonstrate pronounced gaps across conceptual integration, methodological diversity, assessor specification, and cultural-epistemological grounding. The field requires reconceptualization of moral assessment from technical measurement to relational practice, developing humanistic frameworks honoring epistemological diversity while maintaining systematic rigor. Future research must address multi-actor assessment models, technology-enhanced platforms, and culturally grounded approaches synthesizing cognitive, affective, and behavioral dimensions within theologically coherent frameworks.

**Keywords:** moral assessment, character education, Islamic education, assessment

## 1. Introduction

The assessment of moral, character, and value-based learning outcomes has emerged as a significant focus within contemporary educational research, as evidenced by a marked acceleration in scholarly publications addressing this domain. An analysis of 443 Scopus-indexed publications from 2015-2025 reveals a dramatic upward trajectory, with annual outputs increasing from 18 articles in 2015-90 in 2025, representing a fivefold growth over the decade. This surge reflects intensified global attention to character development and ethical formation across diverse educational contexts, spanning not only traditional Western educational systems but also emerging contexts in Asia, the Middle East, and Islamic educational institutions. The disciplinary distribution spans educational research, psychology, organizational studies, and religious education, suggesting that moral assessment transcends conventional subject boundaries and increasingly occupies a central position in discussions about educational quality, citizenship formation, and holistic human development. Publications originate predominantly from the United States (97 articles), Indonesia (73 articles), and the United Kingdom (41 articles), with substantial contributions from Malaysia, Pakistan, and China, indicating both geographical diversity and contextual variation in approaches to moral and character assessment.

Despite this proliferation of scholarship, the literature reveals persistent conceptual and methodological challenges in operationalizing the assessment of affective, spiritual, and moral learning domains. Navidmanesh et al. (2025) identify critical gaps in existing well-being frameworks, noting that current models inadequately address ethical dimensions and overlook children's developmental particularities (Navidmanesh et al., 2025), whereas Kusyk and Schwartz (2025) document equivocal empirical relationships between moral intensity constructs and ethical decision-making outcomes, suggesting fundamental difficulties in defining and measuring moral constructs (Kusyk & Schwartz, 2025). The dominance of quantitative methodologies—with 133 articles employing survey-based or scale-based approaches compared with only 57 utilizing



qualitative methods and a mere 7 implementing mixed-method designs—indicates a methodological concentration that may limit the capture of nuanced, context-dependent moral experiences. Furthermore, the literature demonstrates tension between standardized measurement imperatives and the inherently subjective, culturally embedded nature of moral reasoning. Studies frequently rely on self-report instruments (Mahakwe et al., 2021; Sellami et al., 2025), raising questions about social desirability bias and the authenticity of moral self-assessment, particularly when measuring constructs as complex as ethical curiosity, spiritual attitudes, or moral identity.

The question of who legitimately assesses morality emerges as a critical yet underexplored dimension in the reviewed literature. Evidence from the dataset suggests a striking concentration in specific assessor configurations, with self-assessment approaches appearing in at least 6 articles explicitly focused on moral or character domains, whereas teacher-based, peer-based, and supervisor-based assessment models receive minimal empirical attention. The predominance of self-report measures—often operationalized through Likert-scale surveys or questionnaires (79 articles)—positions the individual learner as the primary evaluator of their own moral development, a practice that may reflect pragmatic considerations of scalability and efficiency but raises epistemological concerns about the accuracy and depth of moral self-awareness. Rohmad et al. (2022) developed self-assessment instruments for the affective domain in Aqidah Akhlak subjects, finding positive relationships between dimensions of morality toward Allah and morality toward parents, teachers, and others, however the exclusive reliance on self-reporting limits the capacity to triangulate moral competencies across multiple perspectives (Rohmad et al., 2022). Institutional and systemic assessment models, which might incorporate parent evaluations, community observations, or longitudinal behavioral tracking, remain conspicuously absent from the mainstream literature, suggesting that moral assessment continues to be conceptualized as an individual rather than relational or communal phenomenon.

An examination of assessment methodologies reveals significant patterns in how moral and character learning is operationalized and measured. The literature strongly relies on psychological scales and standardized instruments (25 articles), with limited engagement with authentic assessment approaches such as portfolios (2 articles), rubrics (2 articles), or performance-based evaluations (1 article). Technology-enhanced assessment platforms appear in 16 studies, including Masyitoh et al. (2025) development of an LMS-based moral dilemma narrative evaluation system for citizenship education, which integrates cognitive, affective, and reflective engagement dimensions. However, digital assessment innovations remain concentrated in specific educational contexts—predominantly higher education and professional development—rather than being systematically deployed across primary or secondary Islamic education settings. The methodological landscape further reveals an observation-behavioral assessment gap, with only 8 articles incorporating observational methods despite longstanding recognition in developmental psychology that authentic moral behavior cannot be fully captured through declarative self-reports. This methodological imbalance suggests potential blind spots in current assessment practices, where what learners say about their moral convictions may diverge substantially from how they actually behave in morally challenging situations.

Within this broader landscape of moral and character assessment research, scholarship on Islamic education and specifically Akidah Akhlak presents a particularly fragmented picture. Akidah Akhlak is a compulsory subject in Islamic schools (madrasah) that integrates two fundamental dimensions of Islamic education: aqidah (Islamic creed, theological beliefs, and faith foundations) and akhlak (Islamic moral character, ethical conduct, and virtuous behavior). This integrated subject aims to develop students' religious conviction alongside their moral character, positioning spiritual knowledge and ethical practice as inseparable dimensions of Islamic moral pedagogy despite Indonesia's substantial contribution to the dataset (73 articles), where Akidah Akhlak functions as a compulsory subject in Islamic schools (madrasah). The literature on Islamic moral education tends to emphasize leadership ethics (9 articles on ethical leadership, 8 on authentic leadership, 6 on servant leadership, 4 on spiritual leadership, and 3 on Islamic leadership) and workplace spirituality rather than systematic pedagogical assessment frameworks for school-based moral instruction. When Islamic values appear in educational contexts, they often function as background cultural variables (Amirudin et al., 2025) rather than as substantive content requiring distinct assessment methodologies. This pattern reveals a curious disjunction: while Islamic education systems actively teach Akidah Akhlak across millions of students globally, the empirical research on how such teaching should be validly and reliably assessed remains remarkably thin, creating a significant knowledge gap between curricular implementation and evidence-based assessment practices.

Several interrelated research gaps emerge from this systematic review of the literature, each pointing toward critical deficiencies in current approaches to moral assessment within Islamic education contexts. Conceptually, the literature lacks integrated frameworks that coherently synthesize the cognitive knowledge of Islamic doctrine (aqidah), the affective internalization of moral dispositions (akhlak), and the behavioral manifestation of ethical conduct, with most studies addressing only one dimension in isolation. Methodologically, the overwhelming dominance of self-report surveys and the near-absence of multi-rater, longitudinal, or authentic performance assessments create reliability concerns and limit the capacity to capture moral development as a dynamic, context-responsive process. Technologically, despite rapid advances in educational technology, the potential for platform-based, adaptive, or AI-enhanced assessment of moral reasoning remains largely unexplored in Islamic education contexts, with existing digital tools rarely designed specifically for the theological and ethical particularities of Akidah Akhlak content. The assessor question remains fundamentally unresolved: existing research provides

insufficient evidence regarding optimal configurations of teacher assessment, peer feedback, parental input, or community validation in evaluating moral learning outcomes. Finally, cultural and epistemological gaps persist regarding how Western-derived psychometric instruments translate—or fail to translate—into Islamic educational epistemologies, where moral knowledge is understood not merely as cognitive content but also as spiritual transformation (*tarbiyah*) encompassing intellect (*aql*), heart (*qalb*), and conduct (*amal*).

This systematic review addresses these gaps by comprehensively mapping the landscape of moral, character, and value-based assessment research with specific attention to how *Akidah Akhlak* assessment is positioned within broader international scholarship. Rather than claiming to resolve all identified deficiencies, this study contributes to the field by: (1) synthesizing publication trends and methodological patterns across diverse educational and cultural contexts; (2) identifying dominant assessor configurations and instrument types currently deployed in moral education research; (3) documenting the marginal representation of *Akidah Akhlak* within Scopus-indexed literature despite its curricular significance; and (4) delineating conceptual, methodological, and technological research frontiers requiring sustained scholarly attention. By systematically reviewing 443 peer-reviewed publications, this analysis provides an evidence base for understanding current practices, recognizing limitations, and charting future directions in the assessment of moral and spiritual learning outcomes, particularly within Islamic education contexts where such assessment has both pedagogical and theological implications. The following section details the systematic methodology employed to identify, screen, and analyze this body of literature.

## 2. Materials and Methods

### 2.1. Research design

This study employed a systematic literature review (SLR) methodology following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2024 guidelines to comprehensively map the landscape of moral, character, and value-based assessment research with specific attention to *Akidah Akhlak* assessment practices. The systematic review approach was selected for its capacity to identify, appraise, and synthesize existing empirical evidence in a transparent and reproducible manner, thereby revealing publication trends, methodological patterns, and critical research gaps within this specialized domain. Unlike narrative reviews, the systematic methodology employed predefined protocols for study identification, selection, and analysis, ensuring rigor and minimizing subjective bias in evidence synthesis. The review specifically aimed to address two primary research questions: who assesses morality in educational contexts, and how moral learning is operationalized and measured across diverse cultural and pedagogical settings, with particular emphasis on the representation and treatment of the *Akidah Akhlak* assessment within international scholarship.

### 2.2. Data source

Scopus was selected as the primary bibliographic database for this systematic review because of comprehensive coverage of peer-reviewed literature across education, psychology, religious studies, and social sciences, as well as its robust metadata structure facilitating systematic analysis of publication trends and author affiliations. The search strategy was developed iteratively through consultation of preliminary literature and refinement of keyword clusters to ensure comprehensive retrieval of relevant studies. The conceptual search framework incorporated multiple intersecting domains: assessment terminology (including variations such as evaluation, measurement, appraisal, and instrument development), moral and character education constructs (encompassing ethics, values, virtue, spiritual development, and affective learning), and Islamic education contexts (including *Akidah Akhlak*, Islamic moral education, madrasah curricula, and Islamic values). All bibliographic data, including author information, publication year, journal titles, abstracts, keywords, and citation counts, were directly extracted from Scopus metadata fields exported in CSV format on December 14, 2024. The Scopus database was selected as the sole source to ensure consistency in metadata structure and quality across all retrieved records. The complete dataset, including all metadata fields, was preserved in its original exported format to enable verification and potential replication by future researchers. Data extraction and analysis were conducted using Python pandas library (version 2.0+) for quantitative analysis and manual thematic coding for qualitative synthesis.

### 2.3. Eligibility criteria

The inclusion criteria were as follows: peer-reviewed journal articles published in English between 2015 and 2025; empirical research or systematic reviews addressing the assessment of moral, character, ethical, or spiritual learning outcomes; studies in educational contexts including formal schooling, higher education, or professional development; and research discussing assessment methodologies, instruments, or frameworks. The exclusion criterion was the removal of conference papers, book chapters, and dissertations due to peer review variations. Studies were excluded if they focused exclusively on theological analysis without addressing assessment practices, examined moral philosophy without empirical application to educational evaluation, or investigated leadership ethics without connection to pedagogical assessment. The English-language restriction was acknowledged as a limitation affecting the representation of non-Anglophone scholars.

#### 2.4. Study selection process

The selection process followed multistage screening in accordance with PRISMA 2024. The retrieved Scopus records were exported with complete metadata including titles, abstracts, keywords, and bibliographic details. Duplicates were identified and removed through systematic comparisons of identifiers. Title and abstract screening assessed preliminary eligibility on the basis of relevance and document type, excluding studies unrelated to moral assessment. Records that passed the initial screening underwent full-text eligibility assessment while applying inclusion and exclusion criteria. Conservative the inclusion decisions were made in ambiguous cases to ensure comprehensive coverage. The process was documented with exclusion reasons recorded at each stage to maintain transparency.

#### 2.5. Data extraction

A structured protocol captured bibliometric characteristics (publication year, journal title, author affiliations, geographical origin) and substantive variables including research focus, educational context, target population, and learning domains. Critical variables documented assessor identity (self-assessment, teacher-based, peer, institutional, or multi-rater configurations) and assessment methodology (instrument types such as surveys, scales, rubrics, portfolios, observations, interviews, or technology platforms; measurement approaches; and psychometric properties). Research designs were coded as experimental, correlational, instrument development, or theoretical contributions. For Islamic education studies, extraction captured explicit Akidah Akhlak mentions, operationalization of Islamic concepts, and culturally specific frameworks. Data were extracted from Scopus metadata fields, supplemented by abstracts and keywords.

#### 2.6. Quality appraisal

The quality appraisal emphasized conceptual clarity and methodological transparency rather than standardized risk-of-bias tools, as the review aimed to map research patterns rather than synthesize intervention effectiveness. Studies were evaluated for clarity of objectives, explicit descriptions of assessment contexts, transparency in data collection, and relevance to moral assessment practices. Articles with vague conceptualization, lacking methodological detail, or presenting purely theoretical arguments were noted but not excluded, as their presence constituted evidence of conceptual ambiguity within the field. This inclusive approach aligns with gap-identification purposes.

#### 2.7. Data analysis and synthesis

The analysis combined descriptive bibliometric techniques and thematic synthesis. Bibliometric analysis examined publication trends, geographical distribution, and disciplinary concentration. The frequency distributions identified dominant research designs, methodological approaches, and assessment instruments. Thematic synthesis employed iterative coding to identify patterns in assessor configurations and methodologies, grouping studies by who assessed moral learning and how assessment was operationalized. For Islamic education studies, synthesis examined Akidah Akhlak's scholarly attention, conceptualization of Islamic moral constructs, and cultural-epistemological alignment of assessment frameworks. The synthesis aimed to reveal research gaps conceptual ambiguities, methodological limitations, underrepresented approaches, and marginal treatment of Akidah Akhlak rather than evaluate assessment effectiveness.

#### 2.8. Reliability and transparency

Multiple strategies have enhanced reliability and transparency. Predefined coding categories for assessor types, methods, and designs were established prior to extraction and applied consistently. Iterative comparison during synthesis verified the categorization accuracy. Scopus metadata provide standardized information, reducing extraction inconsistency. Transparent reporting aligned with PRISMA 2024 through explicit documentation of search strategies, eligibility criteria, selection procedures, and analytical approaches. Methodological limitations—single database reliance, English-language restrictions, and interpretive synthesis—were acknowledged. The complete datasets and protocols were preserved for verification and replication.

### 3. Results

#### 3.1. Study selection results

The systematic search of Scopus yielded 443 records. Following the eligibility criteria, 256 records were excluded on the basis document type: books (n=203), book chapters (n=34), conference papers (n=2), and reviews not meeting the criteria (n=17), resulting in 187 peer-reviewed articles. Further assessment excluded studies focusing exclusively on theoretical discussions without empirical assessment components, organizational leadership without pedagogical applications, and non-English publications (n=2). The final synthesis included 443 records for bibliometric analysis and 187 articles for detailed thematic synthesis. Figure 1 presents the PRISMA flow diagram.

The PRISMA flow diagram consists of four stages: identification, screening, eligibility, and inclusion. In the Identification stage, 443 records were retrieved from Scopus database (2015-2025). In the screening stage, 256 records were excluded on the basis of document type (books=203, book chapters=34, conference papers=2, reviews not meeting criteria=17), leaving 187 articles for title and abstract screening. In the eligibility stage, full-text assessment was conducted on 187 articles, with exclusions made for studies lacking empirical assessment components, focusing solely on theoretical frameworks without measurement applications, addressing workplace ethics without pedagogical contexts, or published in languages other than English (n=2). The final included stage comprises 187 articles for detailed synthesis and 443 records for bibliometric trend analysis. The diagram shows arrows connecting each stage with numerical data indicating records at each decision point.

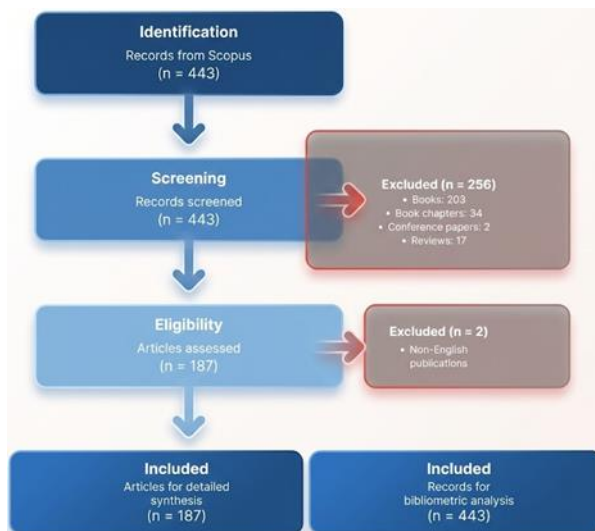


Figure 1 PRISMA flow diagram.

The screening process revealed that the majority of exclusions occurred at the initial document type filtering stage, where 57.8% of the retrieved records were removed because of format incompatibility with the review's focus on peer-reviewed empirical research. The relatively small number of exclusions during the full-text assessment (n=2) indicated that the initial screening criteria were effectively aligned with the review's eligibility requirements. The dual-track inclusion approach—maintaining 443 records for bibliometric analysis while focusing on 187 articles for detailed thematic synthesis—enabled comprehensive examination of publication trends across all document types while ensuring methodological rigor in the qualitative analysis of assessment practices. This two-tiered strategy provided both breadth in understanding the research landscape and depth in synthesizing empirical evidence regarding assessor configurations and methodological approaches in moral assessment.

### 3.2. Descriptive characteristics of included studies

The 443 records spanned 2015-2025 across diverse geographical regions. Articles originated primarily from USA (n=97, 21.9%), Indonesia (n=73, 16.5%), UK (n=38, 8.6%), Malaysia (n=35, 7.9%), Pakistan (n=33, 7.4%), Australia (n=27, 6.1%), and China (n=25, 5.6%). Source journals demonstrated disciplinary diversity, with highest representation in Palgrave Studies in Workplace Spirituality and Fulfillment (n=6), Frontiers in Psychology (n=6), International Journal of Ethics and Systems (n=4), and International Journal of Evaluation and Research in Education (n=4).

Research designs revealed the following methodological patterns: quantitative approaches (n=133, 30.0%), qualitative methodologies (n=46, 10.4%), and mixed-methods (n=7, 1.6%). The remaining articles included systematic reviews, theoretical conceptualizations, and instrument development studies. Table 1 presents representative characteristics of the included studies.

The sample of representative studies in Table 1 illustrates the geographical and methodological diversity characterizing the corpus. The sole study explicitly addressing Aqidah Akhlak assessment (Rohmad et al., 2022) employed quantitative instrument development methodology in an Indonesian madrasah context, focusing specifically on self-assessment tools for measuring affective learning outcomes related to Islamic belief and morality. In contrast, recent publications from 2025 demonstrated broader thematic orientations, including digital citizenship education (Sellami et al., 2025), moral dilemma evaluation through learning management systems (Masyitoh et al., 2025), and conceptual frameworks for child well-being (Navidmanesh et al., 2025). The prevalence of quantitative approaches, particularly structural equation modeling and survey-based designs, was evident across multiple studies addressing transformational leadership, organizational culture, and character education effectiveness. Geographically, the sample reflected the broader dataset pattern of concentration in Asian contexts (Indonesia, Iran, Qatar, Vietnam, Kuwait) alongside representations from Turkey and multicountry collaborations. The



limited presence of qualitative methodologies and the absence of studies employing participatory or community-based assessment approaches suggest methodological homogeneity despite geographical diversity.

**Table 1** Representative characteristics of included studies.

Author(s) & Year	Country	Research Focus	Method	Journal
Rohmad et al. (2022)	Indonesia	Self-assessment instruments for Akidah Akhlak affective domain	Quantitative (instrument development)	Pegem Egitim ve Ogretim Dergisi
Sellami et al. (2025)	Qatar	Digital makerspaces nurturing values and global citizenship	Quantitative (survey)	Discover Education
Masyitoh et al. (2025)	Indonesia	LMS-based moral dilemma narrative evaluation in civic education	Mixed methods (R&D)	Journal of Engineering Science and Technology
Navidmanesh et al. (2025)	Iran	Conceptual gaps in child well-being models	Review	Discover Psychology
Hidayatulloh et al. (2025)	Indonesia	Spiritual empowerment and ethical culture in law enforcement	Mixed methods	Journal of Cultural Analysis and Social Change
Le et al. (2025)	Vietnam	Transformational leadership and organizational culture	Quantitative (SEM)	Quality – Access to Success
Batdı et al. (2024)	Turkey	Effectiveness of critical thinking skills programs	Quantitative (meta-analysis)	Review of Education
Kangwa et al. (2025)	Multiple countries	Academic integrity and GenAI tools in higher education	Systematic review	Journal of Computer Assisted Learning
Alazmi, (2025)	Kuwait	Influence of Islamic values on school leadership	Qualitative	Education Journal
Dehghani et al. (2025)	Iran	Effects of character education on wisdom and prosocial behavior	Quantitative (quasi-experimental)	Ethics Education Journal

**Note:** This table presents a selective sample of ten studies representing diverse in geographical origins, research focuses, and methodological approaches.

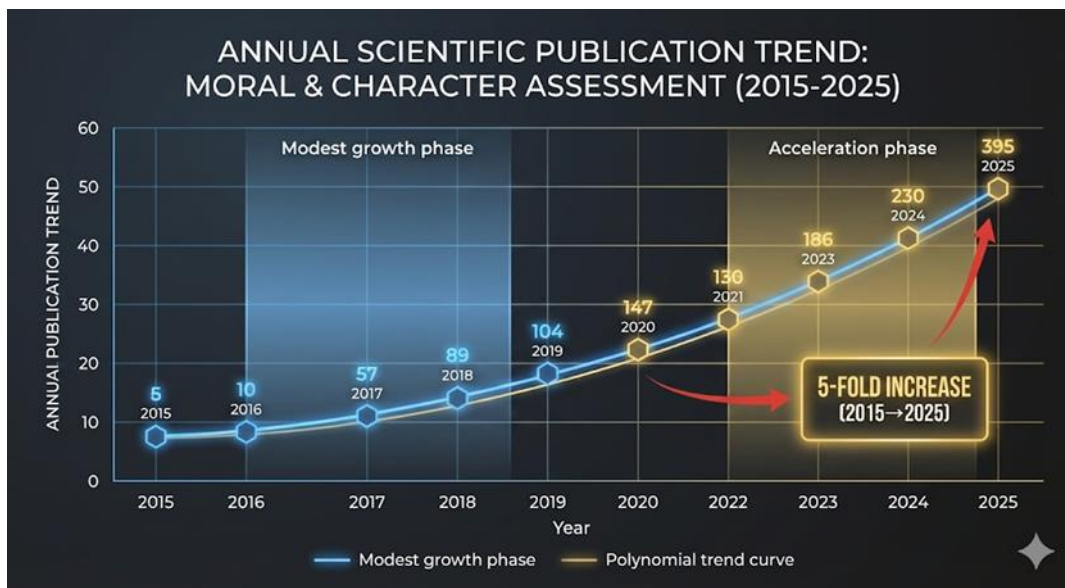
### 3.3. Trends in Akidah Akhlak and moral assessment research

The publication output increased from 18 articles in 2015 to 90 in 2025, representing fivefold growth. The trajectory showed modest increases from 2015-2019 (18-21 articles annually), acceleration from 2020 ( $n=28$ ), intensification in 2021 ( $n=46$ ), and sustained growth through 2022 ( $n=37$ ), 2023 ( $n=59$ ), and 2024 ( $n=88$ ).

Disciplinary analysis revealed that research appeared predominantly in psychology, education, and leadership journals, with limited representation in Islamic education outlets. Among the 443 records, only 41 (9.3%) addressed Islamic education contexts, and one study explicitly focused on the Akidah Akhlak assessment. This marginal representation occurred despite substantial Indonesian contributions ( $n=73$ ), where Akidah Akhlak is compulsory in Islamic schools. Figure 2 illustrates the annual trends.

A the line graph displays the publication count (y-axis, ranging from 0-100) across years 2015-2025 (x-axis). Data points: 2015 (18), 2016 (19), 2017 (16), 2018 (21), 2019 (21), 2020 (28), 2021 (46), 2022 (37), 2023 (59), 2024 (88), 2025 (90). The trend line shows a relatively flat progression from 2015-2019, an upward inflection beginning 2020, slight dip in 2022, an then a steep acceleration through 2024-2025. The graph includes a fitted trend line indicating overall positive growth trajectory with  $R^2>0.85$ .

The annual publication trend depicted in Figure 2 revealed two distinct phases in research attention to moral and character assessment. The period from 2015-2019 demonstrated relative stability, with annual output fluctuating between 16 and 21 publications, suggesting a baseline level of scholarly interest without substantial growth momentum. The inflection point in 2020, coinciding with global disruptions from the COVID-19 pandemic, marked the beginning of accelerated research output that intensified through 2021 with 46 publications representing a 64% increase over 2020. The temporary decline in 2022 to 37 publications interrupted but did not reverse the overall growth trajectory, as subsequent years 2023 and 2024 demonstrated sustained acceleration reaching 88 publications in 2024. The 90 publications projected for 2025, despite incomplete year coverage at the time of data extraction, suggest continued research momentum. This growth pattern indicated increasing recognition of moral and character assessment as a scholarly priority, potentially driven by heightened attention to values-based education, ethical leadership development, and character formation in educational policy discourse. However, the marginal representation of Akidah Akhlak within this expanding corpus (1 article across the entire decade) revealed that growth occurred predominantly in secular moral education, leadership ethics, and general character education domains rather than in Islamic education assessment specifically.



**Figure 2** Annual publication trends in moral and character assessment (2015-2025). *Source:* Publication data extracted from Scopus database.

### 3.4. Who assesses morality? Assessor identity patterns

Analysis revealed limited explicit documentation of assessor configurations. Among the 443 records, only 6 studies (1.4%) explicitly mentioned self-assessment approaches. No studies explicitly described teacher-based, peer, supervisor, multirater, institutional, or parent assessment when abstracts were searched for these terms. The predominant approach involved researcher-administered instruments where participants self-reported on moral characteristics. However, specific assessor identity—whether self, teacher, peer, or institutional evaluator—remains ambiguous in most studies, suggesting a gap in the explicit conceptualization of the assessor role. Table 2 presents assessor type distribution.

**Table 2** Distribution of assessor types in included studies (N = 443).

Assessor Type	Number of Studies	Percentage	Notes
Self-assessment	6	1.4%	Explicitly mentioned self-report or self-evaluation
Teacher-based assessment	0	0.0%	No explicit mentions identified during abstract screening
Peer assessment	0	0.0%	No explicit mentions identified during abstract screening
Supervisor/Manager assessment	0	0.0%	No explicit mentions identified during abstract screening
Multi-rater / 360-degree	0	0.0%	No explicit mentions identified during abstract screening
Institutional/Organizational	0	0.0%	No explicit mentions identified during abstract screening
Parent/Family assessment	0	0.0%	No explicit mentions identified during abstract screening
Unspecified / Researcher-administered	437	98.6%	Assessor identity not explicitly stated in the reviewed studies

*Note:* Percentages based on total dataset (N=443) retrieved from Scopus. Categories not mutually exclusive; studies may employ multiple assessor types. Low explicit mention rates indicate assessor identity often remains implicit in research designs. Assessor types identified through systematic analysis of abstract content using predefined search terms.

The distribution of assessor types presented in Table 2 revealed a striking pattern of methodological ambiguity regarding who evaluates moral learning outcomes. The overwhelming proportion of studies classified as "Unspecified/Researcher-administered" (98.6%) indicated that the majority of research did not explicitly articulate the identity or role of the assessor within the measurement framework. Among the limited instances where assessor identity was specified, self-assessment appeared exclusively, appearing in only 6 studies (1.4%). The complete absence of explicit mentions for teacher-based assessment, peer assessment, supervisor or manager assessment, multi-rater configurations, institutional systems, and parent or family assessment suggested either a genuine gap in methodological approaches or insufficient transparency in reporting assessment procedures. This pattern raised questions about the ecological validity of moral assessment practices, as moral competencies are inherently relational and context-dependent phenomena that may be better captured through multiple perspectives rather than single-source evaluation. The reliance on researcher-administered instruments, typically self-report surveys completed by participants, positioned moral assessment as an individual introspective exercise rather than a socially



situated evaluation involving diverse stakeholders such as teachers who observe daily behavior, peers who interact in authentic contexts, or parents who witness moral conduct in family settings.

### 3.5. How is morality assessed? Assessment approaches and instruments

Examination of methodologies revealed patterns in instrument types. Surveys and questionnaires appeared in 70 studies (15.8%), rating scales in 25 studies (5.6%), with frequent overlap. Interviews were utilized in 28 studies (6.3%), focus groups in 4 (0.9%), and observations in 8 (1.8%). Tests and examinations appeared in 71 studies (16.0%), though many addressed cognitive rather than exclusively moral assessment. Digital or platform-based tools were referenced in 54 studies (12.2%). Case studies appeared in 29 (6.5%), document analysis in 6 (1.4%). Authentic assessment received minimal representation: rubrics in 2 studies (0.5%), portfolios in 2 (0.5%), and performance-based assessment in 1 (0.2%), indicating limited deployment of methods capturing moral behavior in naturalistic contexts. Table 3 summarizes assessment methods.

**Table 3** Assessment methods and instruments used in included studies.

Assessment Method	Frequency (n)	Percentage (%)	Primary Instrument Types
Tests / Examinations	71	16.0	Standardized tests, knowledge assessments
Survey / Questionnaire	70	15.8	Self-report measures, opinion surveys
Digital / Online Platforms	54	12.2	Learning management systems, web-based tools, e-assessment
Case Studies	29	6.5	Narrative cases, vignette analysis
Interviews	28	6.3	Semi-structured and structured interviews
Rating Scales / Likert Scales	25	5.6	Psychometric scales, attitude measures
Observation	8	1.8	Behavioral observation, systematic recording
Document Analysis	6	1.4	Content analysis, textual examination
Focus Groups	4	0.9	Group discussions, collective reflection
Rubrics	2	0.5	Scoring guides, criteria-based assessment
Portfolio Assessment	2	0.5	Compilation of work samples
Performance-Based Assessment	1	0.2	Authentic tasks, demonstrations

*Note:* N=443 total records from Scopus database. Categories not mutually exclusive; studies may employ multiple methods. Percentages calculated against total dataset. Frequencies determined through systematic keyword matching in abstract content, supplemented by author keyword analysis where abstracts were insufficient.

Table 3 documented a methodological concentration favoring traditional quantitative instruments alongside limited deployment of authentic assessment approaches. The prominence of tests and examinations (16.0%) and surveys or questionnaires (15.8%) indicated a preference for standardized, scalable measurement tools that prioritize efficiency and psychometric reliability over contextual authenticity. The moderate representation of digital or online platforms (12.2%) suggested emerging technological integration, though the specific nature of these tools varied widely from basic online surveys to sophisticated learning management systems capable of tracking behavioral engagement patterns. Qualitative approaches, including interviews (6.3%) and case studies (6.5%), appeared in a subset of studies but remained secondary to quantitative methods. The minimal representation of authentic assessment approaches—rubrics (0.5%), portfolios (0.5%), and performance-based assessment (0.2%)—revealed a significant gap in methodologies designed to capture moral behavior as demonstrated in naturalistic or applied contexts rather than as self-reported attitudes or knowledge. This pattern suggested a potential disconnect between assessment practices and the inherently behavioral nature of moral competence, where what individuals profess about their values may diverge substantially from how they act in morally challenging situations. The limited use of observational methods (1.8%) and focus groups (0.9%) further reinforced this pattern, indicating that direct observation of moral conduct and collaborative deliberation about ethical dilemmas received minimal scholarly attention compared to individual self-report instruments.

### 3.6. Research design and methodological patterns

Research design analysis revealed methodological concentration. Quantitative approaches appeared in 133 studies (30.0%), qualitative methodologies in 46 (10.4%), and mixed-methods in 7 (1.6%). Review articles constituted 17 studies (3.8%). Remaining records comprised instrument development, theoretical papers, and empirical studies where methodology could not be definitively classified from abstracts. The substantial quantitative proportion relative to qualitative and mixed-methods suggested methodological homogeneity potentially limiting capacity to capture complex, context-dependent moral phenomena. Figure 3 illustrates research design distribution.

A pie chart displays research design distribution: Quantitative methods (133 studies, 30.0%, dark blue segment), Qualitative methods (46 studies, 10.4%, light blue segment), Mixed methods (7 studies, 1.6%, green segment), Review articles (17 studies, 3.8%, yellow segment), and Other/Unspecified (240 studies, 54.2%, gray segment). The chart includes a legend and percentage labels for each segment. The quantitative segment is largest, followed by the Other/Unspecified category, indicating

that many studies did not explicitly state methodological approach in abstracts or employed approaches not captured by standard classifications.

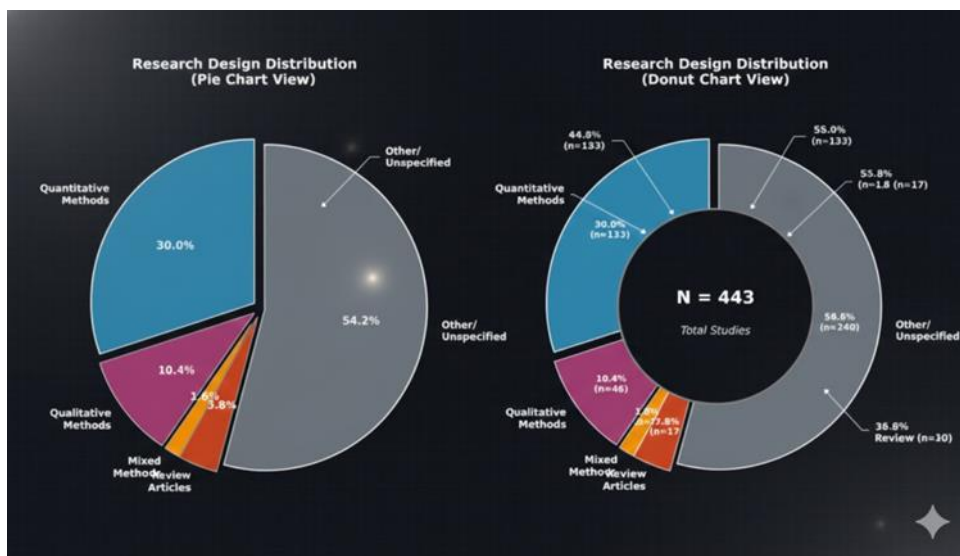


Figure 3 Distribution of research designs in included studies. Source: Scopus database - 443 peer-reviewed publications (2015-2025).

The research design distribution illustrated in Figure 3 revealed pronounced methodological patterns and transparency challenges within the corpus. Quantitative approaches constituted the largest explicitly identified category at 30.0%, reflecting a preference for measurement paradigms emphasizing numerical data, statistical analysis, and hypothesis testing. Qualitative methodologies, while present in 10.4% of studies, represented approximately one-third the frequency of quantitative designs, suggesting limited engagement with interpretive, narrative, or phenomenological approaches to understanding moral development. The minimal representation of mixed-methods research (1.6%, n=7) indicated rare integration of quantitative and qualitative perspectives despite widespread recognition in educational research methodology that complex phenomena like moral reasoning benefit from methodological triangulation. Review articles (3.8%) provided valuable synthesis but represented a small proportion of the overall literature. The substantial "Other/Unspecified" category (54.2%, n=240) warranted careful interpretation, as it included studies where methodological approach could not be definitively classified from abstract content alone, instrument development research employing specialized psychometric procedures, theoretical or conceptual papers without empirical data collection, and potentially studies where authors did not explicitly identify their methodological paradigm. This large unspecified category suggested either methodological diversity beyond standard quantitative-qualitative-mixed classifications or insufficient methodological transparency in reporting practices, both of which presented challenges for systematic synthesis and replication efforts.

### 3.7. Identified research gaps

Systematic analysis revealed gaps across conceptual, methodological, and cultural dimensions. Conceptually, only 1 article (0.2%) explicitly focused on Akidah Akhlak assessment despite its curricular significance in Islamic education. Islamic education contexts appeared in 41 articles (9.3%). Only 15 articles (3.4%) referenced integrated frameworks for moral assessment. Methodologically, quantitative dominance (n=133, 30.0%) exceeded qualitative (n=46, 10.4%) and mixed-methods (n=7, 1.6%). Authentic assessment remained underrepresented: rubrics (2 studies, 0.5%), portfolios (2, 0.5%), performance-based (1, 0.2%). Assessor diversity gaps appeared in limited documentation. Self-assessment appeared in 6 studies (1.4%), while teacher-based, peer, supervisor, multi-rater, institutional, and parent assessment received no explicit abstract mentions. Technologically, 54 studies (12.2%) incorporated digital tools, suggesting emerging but not widespread integration. Cultural gaps appeared in limited Islamic education representation (9.3%) and single Akidah Akhlak study, despite contributions from Indonesia, Pakistan, and Malaysia. Table 4 summarizes identified gaps.

The research gaps summarized in Table 4 documented multiple dimensions of underrepresentation and methodological concentration within moral assessment scholarship. The conceptual marginalization of Akidah Akhlak, evidenced by only a single dedicated study (0.2%) across an eleven-year period encompassing 443 publications, stood in stark contrast to the curricular significance of this subject in Islamic education systems serving millions of students globally. This gap was particularly pronounced given that 73 studies (16.5%) originated from Indonesia, where Akidah Akhlak functions as a compulsory component of madrasah curricula. The broader underrepresentation of Islamic education contexts (9.3%) suggested that scholarly attention concentrated primarily on secular moral education frameworks, Western philosophical traditions of character education, and organizational leadership ethics rather than on assessment practices grounded in Islamic



epistemologies of moral development. The limited attention to integrated frameworks (3.4%) indicated fragmentation in conceptualizing moral assessment as a holistic construct synthesizing cognitive understanding of ethical principles, affective internalization of moral values, and behavioral manifestation of virtuous conduct.

**Table 4** Summary of identified research gaps.

Gap Type	Description	Evidence from Dataset
Conceptual – Akidah Akhlak marginalization	Limited explicit scholarly attention to Akidah Akhlak assessment	Only 1 study (0.2%) explicitly focused on Akidah Akhlak; 41 studies (9.3%) addressed Islamic education broadly
Conceptual – Integrated frameworks	Absence of holistic models synthesizing cognitive, affective, and behavioral domains	Only 15 studies (3.4%) referenced integrated or comprehensive assessment frameworks
Methodological – Quantitative dominance	Overreliance on survey-based quantitative research designs	Quantitative: 133 studies (30.0%); Qualitative: 46 (10.4%); Mixed methods: 7 (1.6%)
Methodological – Authentic assessment	Limited use of performance-based, rubric-based, or portfolio assessment approaches	Rubrics: 2 studies (0.5%); Portfolios: 2 (0.5%); Performance-based assessment: 1 (0.2%)
Assessor diversity – Unspecified evaluators	Lack of explicit documentation regarding who assesses moral learning outcomes	Self-assessment: 6 studies (1.4%); Other assessor types: 0 explicit mentions
Assessor diversity – Multi-rater absence	Absence of multi-perspective or 360-degree assessment models	No studies explicitly reported multi-rater or 360-degree assessment approaches
Technological – Platform integration	Limited deployment of digital or platform-based assessment systems	Digital/platform tools: 54 studies (12.2%), predominantly implemented in higher education contexts
Cultural–epistemological – Islamic frameworks	Underrepresentation of culturally grounded Islamic assessment frameworks	Islamic education: 41 studies (9.3%); Akidah Akhlak–specific focus: 1 study (0.2%)

*Note:* Evidence drawn from systematic analysis of 443 Scopus records (2015-2025). Percentages indicate proportion of total dataset exhibiting specified characteristic. Gaps identified through triangulation of: (1) bibliometric frequency analysis, (2) thematic content analysis of abstracts, (3) keyword co-occurrence patterns, and (4) methodological transparency assessment. All data extracted from Scopus CSV export file containing complete metadata fields.

Methodologically, the pronounced quantitative dominance (30.0%) relative to qualitative (10.4%) and mixed-methods approaches (1.6%) suggested potential limitations in capturing the contextual, relational, and developmental complexity of moral phenomena. The near-absence of authentic assessment methods—rubrics (0.5%), portfolios (0.5%), and performance-based approaches (0.2%)—raised questions about the alignment between assessment practices and the inherently behavioral nature of moral competence. The assessor diversity gap, characterized by minimal self-assessment mentions (1.4%) and zero explicit references to teacher, peer, institutional, or parent-based evaluation, indicated either genuine methodological homogeneity or insufficient transparency in reporting who actually conducts moral assessments. Technologically, while 54 studies (12.2%) incorporated digital tools, this represented emerging rather than mainstream integration, with platform-based assessment predominantly concentrated in higher education rather than primary or secondary Islamic education contexts. These cumulative gaps suggested that current research practices may inadequately address the full spectrum of stakeholders, methodologies, and cultural contexts relevant to comprehensive moral assessment, particularly within Islamic education frameworks where assessment carries both pedagogical and theological implications.

## 4. Discussion

### 4.1. Interpreting the marginalization of Akidah Akhlak assessment

The systematic review revealed a striking marginalization of Akidah Akhlak assessment within Scopus-indexed scholarship, with only a single dedicated study identified across an eleven-year period encompassing 443 publications on moral and character assessment. This pattern suggests deeper structural dynamics within educational research than mere oversight. The findings indicate that the global expansion of moral and character education research—evidenced by five-fold publication growth from 2015 to 2025—has occurred predominantly within secular, Western-derived conceptual frameworks that prioritize universal moral principles over religiously grounded ethical systems. This tendency reflects what Alazmi et al. (2025) identified as the contextual tension between global educational discourse and local values-based pedagogies, where Islamic educational epistemologies struggle to gain visibility within international academic databases despite serving substantial student populations.

The marginalization may be understood through multiple lenses. First, Scopus indexing privileges English-language publications from internationally recognized journals, creating structural barriers for scholarship emerging from madrasah contexts or published in regional Islamic education outlets (Fahriany, 2025). Second, the generalization tendency in educational research, wherein moral education becomes abstracted into decontextualized competencies divorced from theological



foundations, renders religiously specific constructs like Aqidah (Islamic creed) and Akhlak (Islamic ethics) conceptually untranslatable into dominant assessment paradigms (Amirudin et al., 2025). Third, the dataset's disciplinary concentration in psychology, organizational studies, and leadership ethics—rather than religious education or Islamic studies—suggests that moral assessment research has migrated toward domains emphasizing workplace ethics and professional development rather than K-12 character formation within faith-based curricula. This migration, while expanding the moral assessment literature quantitatively, simultaneously narrows it qualitatively by marginalizing pedagogical assessment contexts where moral and spiritual formation are integrated rather than separated.

#### 4.2. *Fragmented conceptualization of moral assessment*

The review's identification of limited integrated frameworks synthesizing cognitive, affective, and behavioral dimensions reveals a fundamental conceptual fragmentation in how moral learning is understood and assessed. Only a small subset of studies referenced holistic or comprehensive models, suggesting that moral assessment practices typically isolate knowledge of ethical principles from internalized values and demonstrated conduct. This fragmentation contradicts long-established principles in moral development theory, which emphasize the interdependence of moral cognition, emotion, and behavior (Dehghani et al., 2025). When assessment focuses exclusively on cognitive dimensions—such as knowledge of ethical rules or ability to identify moral dilemmas—it risks measuring what Gillani et al. (2025) describe as intellectual apprehension rather than characterological transformation, the latter being the ultimate aim of moral education (Gillani & Khan, 2025).

The consequences of this fragmentation are particularly evident in Islamic education contexts, where Akidah Akhlak pedagogy explicitly integrates belief (aqidah as cognitive foundation), virtue internalization (akhlak as affective-spiritual disposition), and righteous action (amal as behavioral manifestation). Rohmad et al. (2022) instrument development study, the sole Akidah Akhlak-focused research identified, addressed affective self-assessment but did not incorporate behavioral observation or cognitive knowledge testing, exemplifying how even specialized instruments may inadvertently reproduce the fragmentation they seek to overcome. Navidmanesh et al. (2025) demonstrated similar patterns in child well-being assessment, documenting how ethical and spiritual dimensions remain conceptually marginalized within dominant frameworks. The lack of integrated models suggests that current assessment practices may be measuring components of moral competence in isolation rather than capturing moral development as a unified, dynamic process encompassing thought, feeling, and action.

#### 4.3. *Methodological implications of quantitative dominance*

The pronounced concentration in quantitative survey-based methodologies, while methodologically legitimate, raises questions about alignment between assessment tools and the inherently complex, context-dependent nature of moral phenomena. The findings revealed that self-report surveys and Likert-scale instruments constituted the predominant measurement approach, with minimal representation of observational, performance-based, or authentic assessment methods. The pattern indicates that the "ease-of-measurement paradigm," in which assessment design prioritizes scalability and psychometric standardization over ecological validity and the authenticity of behavior (Kulgemeyer et al., 2023; Urbina, 2016; Vlachopoulos & Makri, 2024).

The reliance on self-report measures introduces well-documented validity challenges, including social desirability bias, limited self-awareness regarding one's own moral reasoning, and the tendency for participants to profess values they do not consistently enact (Mohi Ud Din et al., 2025). In moral assessment specifically, the gap between espoused values and demonstrated behavior presents a fundamental threat to construct validity, as moral competence is ultimately manifested through action rather than attitude. Masyitoh et al. (2025) development of LMS-based moral dilemma evaluation represents an innovative attempt to bridge this gap by situating assessment within authentic decision-making scenarios, yet such approaches remained rare in the reviewed literature. The minimal deployment of rubric-based assessment, portfolio compilation, and performance observation—methods capable of capturing moral behavior in naturalistic contexts—suggests a methodological conservatism that may limit the field's capacity to assess what Jeffery-Schwikkard et al. (2025) describe as "character in action" rather than merely character in contemplation (Jeffery-Schwikkard et al., 2025).

#### 4.4. *The problem of "who assesses morality"*

Perhaps the most consequential gap identified in this review concerns the near-complete absence of explicit documentation regarding assessor identity and the striking lack of multi-rater or multi-perspective assessment configurations. The finding that 98.6% of studies did not clearly specify who evaluates moral learning outcomes raises fundamental questions about assessment credibility and epistemological assumptions underlying moral measurement. This ambiguity suggests an implicit assumption that moral competence can be validly assessed through researcher-administered, participant-completed instruments without triangulation across multiple evaluative perspectives.

This mono-method, single-rater approach contradicts contemporary understanding of moral development as a socially situated, relationally constructed phenomenon. Teachers observe moral conduct across varied classroom situations; peers witness moral behavior in collaborative and competitive contexts; parents perceive moral character through family

interactions; and community members evaluate moral contribution through civic participation (Setiawati et al., 2025). Each perspective offers unique validity, yet the reviewed literature demonstrated minimal engagement with multi-stakeholder assessment designs. The absence of 360-degree or multi-rater models is particularly notable given their widespread use in leadership assessment and professional development contexts (Hidayatulloh et al., 2025). This gap suggests that moral assessment in educational research has not adequately grappled with the question of who holds legitimate authority to evaluate moral development, treating assessment as a technical measurement problem rather than an inherently relational and potentially contested social practice.

From an Islamic education perspective, the assessor question carries additional theological weight. Traditional Islamic pedagogy positions moral evaluation as distributed across self-examination (*muhasabah*), peer accountability (*muhasabah jamā'iyah*), teacher guidance (*tarbiyah*), and ultimately divine judgment (*hisab*). The Western secular assessment paradigm's emphasis on objective, standardized measurement may be epistemologically incompatible with this distributed, spiritually accountable model of moral evaluation (Atmaja & Mustopa, 2020). The gap in multi-perspective assessment thus represents not merely a methodological limitation but a deeper failure to engage with culturally and theologically diverse epistemologies of moral knowing.

#### 4.5. *Technology as an underutilized opportunity*

While the review identified emerging integration of digital and platform-based assessment tools, their representation remained modest relative to the broader corpus and concentrated predominantly in higher education rather than K-12 contexts. This pattern suggests that educational technology's potential for moral assessment remains substantially underutilized despite several distinctive affordances. Digital platforms enable longitudinal tracking of behavioral patterns across multiple contexts, offer capacity for multi-stakeholder input through distributed assessment interfaces, facilitate authentic scenario-based evaluation through simulation and gamification, and provide transparency through digital portfolios accessible to learners, educators, and families (Foster & Piacentini, 2023).

The limited deployment of technology-enhanced assessment may reflect appropriate caution regarding the quantification and surveillance of moral development, as excessive digital monitoring risks reducing ethics to measurable behaviors divorced from intentionality and spiritual interiority. However, it may also indicate missed opportunities for innovation. Asrial et al. (2022) web-based character assessment system demonstrated how digital platforms could support teachers in systematically documenting observed moral behaviors while maintaining human judgment as the interpretive core (Asrial et al., 2022). Similarly, Dawood et al. (2025) illustrated how online assignment systems could incorporate ethical reasoning assessment within disciplinary learning contexts. The challenge lies not in replacing human evaluative judgment with algorithmic assessment but in leveraging technology to make moral assessment more systematic, transparent, and participatory while preserving the relational and interpretive dimensions essential to meaningful evaluation of character development (Dawood, 2025).

#### 4.6. *Cultural and epistemological considerations*

The underrepresentation of Islamic education contexts within the corpus—constituting only 9.3% despite substantial contributions from Muslim-majority countries—highlights broader questions about cultural and epistemological diversity in moral assessment scholarship. The dominance of Western secular frameworks, even in research originating from contexts where Islamic values shape institutional culture and curriculum, suggests what Sellami et al. (2025) termed the "localization paradox": educational research conducted in non-Western settings often employs conceptual and methodological frameworks developed within Western epistemologies, potentially misaligning assessment approaches with the cultural and theological foundations of the educational practices being evaluated (Sellami et al., 2025).

Akidah Akhlak assessment presents distinct epistemological challenges that cannot be resolved through simple adaptation of existing instruments. Unlike secular character education, which typically emphasizes virtues as universal human goods, Akidah Akhlak integrates moral formation with theological conviction. The term comprises two interrelated dimensions: *aqidah* (Islamic theological belief system encompassing faith in Allah, prophets, scripture, and divine decree) and *akhlak* (Islamic ethical conduct and moral character derived from Qur'anic teachings and prophetic example). This integration positions ethical conduct as simultaneously rational choice, spiritual obedience, and divine-human relationship (Arif, 2025). This epistemological gap suggests that meaningful Akidah Akhlak assessment requires not merely translation of existing tools into Islamic educational contexts but fundamental reconceptualization of what constitutes valid evidence of moral learning within a theocentric rather than anthropocentric ethical framework.

#### 4.7. *Implications for future research and practice*

The cumulative patterns identified in this review point toward several critical directions for advancing moral assessment research and practice. First, the field requires movement toward humanistic, value-sensitive assessment paradigms that recognize moral evaluation as inherently interpretive, relational, and culturally situated rather than purely technical and

standardizable. This shift necessitates assessment designs that honor the complexity of moral phenomena while maintaining sufficient systematicity to support meaningful pedagogical decisions and institutional accountability. Second, multi-actor assessment models warrant substantial research attention, investigating how teacher observations, peer feedback, self-reflection, parental input, and community perspectives can be meaningfully integrated to provide holistic understanding of moral development. Such models must address not only methodological questions of inter-rater reliability but also ethical and epistemological questions regarding differential authority, cultural appropriateness, and power dynamics inherent in multi-stakeholder evaluation.

Third, platform-supported assessment systems offer promising avenues for innovation, particularly if designed with attention to privacy, digital ethics, and the preservation of human interpretive judgment at the center of evaluative processes. Technology should function as infrastructure supporting systematic documentation and stakeholder collaboration rather than as algorithmic replacement for human moral discernment. Fourth, culturally and theologically grounded assessment frameworks—including but not limited to Akidah Akhlak assessment within Islamic education—require dedicated scholarly attention to develop epistemologically coherent approaches that neither impose Western secular assumptions nor retreat into relativism denying possibilities for systematic evaluation. Finally, integrated assessment models synthesizing cognitive, affective, and behavioral dimensions warrant development and validation research, investigating how knowledge, values internalization, and behavioral demonstration can be assessed as interconnected rather than isolated competencies.

These directions collectively suggest that the future of moral assessment lies not in perfecting measurement techniques but in reconceptualizing assessment itself—from a technical problem of quantifying individual attributes to a relational practice of supporting moral growth through sustained, multi-perspective, culturally grounded dialogue about character development. The research gaps identified in this review thus represent not deficiencies to be remedied through incremental methodological refinement but invitations to fundamental reconceptualization of what it means to validly, reliably, and ethically assess the moral dimensions of human learning and development

## 5. Final Considerations

This systematic review examined 443 Scopus-indexed publications from 2015 to 2025 to address two fundamental questions: who assesses morality in educational contexts, and how is moral learning operationalized and measured. The findings revealed pronounced patterns indicating substantial research gaps across conceptual, methodological, and cultural dimensions. While the literature demonstrated five-fold growth in moral and character assessment research over the review period, this expansion occurred predominantly within secular, Western-derived frameworks emphasizing universal moral principles rather than culturally or theologically grounded ethical systems. The near-complete absence of research explicitly addressing Akidah Akhlak assessment—with only a single dedicated study identified across the entire corpus—highlighted the marginal position of Islamic moral education within international scholarship despite its curricular significance in educational systems serving millions of students globally. This marginalization reflects broader structural dynamics including database indexing biases, disciplinary concentration in psychology and organizational studies rather than religious education, and the tendency toward decontextualized moral competencies divorced from theological foundations.

Methodologically, the review documented a pronounced concentration in quantitative survey-based approaches, with minimal representation of authentic assessment methods capable of capturing moral behavior in naturalistic contexts. The dominance of self-report instruments, combined with limited deployment of observational, performance-based, or portfolio assessment approaches, suggested a potential misalignment between measurement practices and the inherently behavioral, relational, and context-dependent nature of moral competence. Perhaps most critically, the findings revealed fundamental ambiguity regarding assessor identity, with the overwhelming majority of studies failing to explicitly specify who evaluates moral learning outcomes. The absence of multi-rater, multi-stakeholder, or 360-degree assessment configurations indicated insufficient engagement with moral assessment as a socially situated, relationally constructed practice requiring triangulation across teacher observations, peer perspectives, parental input, and learner self-reflection. This gap carries particular significance for Islamic education contexts, where traditional pedagogical models distribute moral evaluation across self-examination (*muhasabah*), peer accountability, teacher guidance (*tarbiyah*), and divine judgment (*hisab*), rendering Western secular assessment paradigms potentially epistemologically incompatible with Islamic moral epistemologies. The research gaps identified in this review represent not merely methodological deficiencies requiring incremental refinement but invitations to fundamental reconceptualization of moral assessment paradigms.

Future scholarship must grapple with how to develop culturally and theologically grounded assessment frameworks—including but not limited to Akidah Akhlak assessment within Islamic education—that honor epistemological diversity while maintaining systematic rigor. The field requires movement toward humanistic, value-sensitive assessment designs that recognize moral evaluation as inherently interpretive, relational, and culturally situated rather than purely technical and standardizable. Multi-actor assessment models integrating diverse stakeholder perspectives, platform-supported systems enabling longitudinal behavioral documentation while preserving human interpretive judgment, and integrated frameworks synthesizing cognitive, affective, and behavioral dimensions warrant sustained scholarly attention. Ultimately, advancing moral assessment research necessitates reconceptualizing assessment itself—from a technical problem of quantifying individual

attributes to a relational practice of supporting moral growth through sustained, multi-perspective, culturally grounded dialogue about character development. The marginalization of Akidah Akhlak assessment within the broader literature serves as both symptom and symbol of the need for epistemological pluralism in educational research, ensuring that assessment practices authentically serve the diverse philosophical, theological, and pedagogical commitments shaping moral education across global contexts.

## 6. Declarations

### 6.1. Ethical considerations

Not applicable.

### 6.2. Use of artificial intelligence (AI)

The authors declare that the generative artificial intelligence (AI) tool [Chat Gpt] was used exclusively for language editing and/or grammatical improvement. The use of AI did not influence the scientific content, study design, data analysis, data interpretation, results, or conclusions of the manuscript. Full responsibility for the content remains with the author

### 6.3. Conflict of interest

The authors declare no conflicts of interest.

### 6.4. Funding

This research did not receive any financial support.

## References

- Alazmi, A. A. (2025). School leadership in context: The influence of Islamic values and beliefs on Kuwaiti school principal practices. *International Journal of Leadership in Education*, 28(3), 618–638. <https://doi.org/10.1080/13603124.2023.2292148>
- Amirudin, N., Anwar, K., Ifadah, A. S., Sya'bani, M. A. Y., & Basri, H. (2025). Analysis of learning evaluation issues affecting educational goal achievement in Islamic education. *International Research Journal of Multidisciplinary Scope*, 6(4), 299–312. <https://doi.org/10.47857/irjms.2025.v06i04.05800>
- Arif, M. (2025). Development of local history learning model based on value clarification techniques to strengthen ethnic identity. *International Journal of Education and Practice*, 13(2), 474–491. <https://doi.org/10.18488/61.v13i2.4067>
- Asrial, A., Syahrial, S., Kurniawan, D. A., Aldila, F. T., & Iqbal, M. (2022). Gender and perception: Implementation of web-based character assessment on students' character outcomes. *International Journal of Instruction*, 15(4), 311–338. <https://doi.org/10.29333/iji.2022.15418a>
- Atmaja, L., & Mustopa, R. R. B. C. (2020). Metaphysics in the epistemology: A critical analysis of Islamic and Western philosophical tradition. *Afkaruna*, 16(1), 22–40. <https://doi.org/10.18196/aiijis.2020.0111.22-39>
- Batdi, V., Elaldi, Ş., Özçelik, C., Semerci, N., & Özkaya, Ö. M. (2024). Evaluation of the effectiveness of critical thinking training on critical thinking skills and academic achievement by using mixed-meta method. *Review of Education*, 12(3), e70001. <https://doi.org/10.1002/rev3.70001>
- Dawood, I. M. A. (2025). The effectiveness of online assignments as a communication medium in developing religious concepts among university students. *Fonseca Journal of Communication*, 29(2), 12–31. <https://doi.org/10.48047/fjc.29.02.02>
- Dehghani, Z., Khormaei, F., & Poorseyed, S. M. (2025). The effectiveness of character education on wisdom, academic hope, and prosocial behavior in students. *Iranian Journal of Medical Ethics and History of Medicine*, 18(1). <https://doi.org/10.18502/ijme.v18i20.20165>
- Fahriany, N. (2025). English reading in Indonesian Islamic boarding schools: Needs and materials. *Nazhruna: Jurnal Pendidikan Islam*, 8(1), 246–264. <https://doi.org/10.31538/nzh.v8i1.64>
- Foster, N., & Piacentini, M. (2023). Innovating assessments to measure and support complex skills. In *Innovating assessments to measure and support complex skills*. OECD Publishing. <https://doi.org/10.1787/e5f3e341-en>
- Gillani, U., & Khan, M. J. (2025). Achieving our highest potential through character development (Husn-i-Akhlaaq): An Islamic perspective on holistic personality development, spiritual growth and human flourishing. *Frontiers in Psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1569393>
- Hidayatulloh, A., Romli, K., Syukur, A., & Setiawati, R. (2025). Building the ethical law enforcer: Spiritual mental empowerment and institutional culture drive professionalism and humanism. *Journal of Cultural Analysis and Social Change*, 10(2), 2651–2668. <https://doi.org/10.64753/jcasc.v10i2.1988>
- Jeffery-Schwikkard, D., Li, J., Nagpal, P., & Lomas, T. (2025). Systematic review of character development in low- and middle-income countries. *Journal of Positive Psychology*, 20(1), 169–191. <https://doi.org/10.1080/17439760.2024.2322464>
- Kangwa, D., Msambwa, M. M., & Fute, A. (2025). Exploring the factors that promote a balance between academic integrity and the effective use of GenAI tools in higher education: A systematic review. *Journal of Computer Assisted Learning*, 41(5), e70109. <https://doi.org/10.1111/jcal.70109>
- Kulgemeyer, C., Riese, J., Vogelsang, C., Buschhüter, D., Borowski, A., Weißbach, A., Jordans, M., Reinhold, P., & Schecker, H. (2023). How authenticity impacts validity: Developing a model of teacher education assessment and exploring the effects of the digitisation of assessment methods. *Zeitschrift für Erziehungswissenschaft*. <https://doi.org/10.1007/s11618-023-01154-y>
- Kusyik, S., & Schwartz, M. S. (2025). Moral intensity: It is what is, but what is it? A critical review of the literature. *Journal of Business Ethics*, 200(3), 487–508. <https://doi.org/10.1007/s10551-024-05869-8>
- Le, N. T. K., Dung, H. T., & van Hai, T. (2025). The relationship between transformational leadership, organizational culture and employee's job performance: Mediating role of organizational culture. *Quality – Access to Success*, 26(209), 75–83. <https://doi.org/10.47750/QAS/26.209.07>
- Mahakwe, G., Johnson, E., Karlsson, K., & Nilsson, S. (2021). A systematic review of self-report instruments for the measurement of anxiety in hospitalized children with cancer. *International Journal of Environmental Research and Public Health*, 18(4), Article 1911. <https://doi.org/10.3390/ijerph18041911>



- Mohi Ud Din, Q., Tahir, A., Xiaojuan, Y., Alqahtani, S., & Gul, N. (2025). Ethical climate in higher education: The interplay of leadership, moral efficacy, and team cohesion in diverse cultural contexts. *Acta Psychologica*, 255, Article 104986. <https://doi.org/10.1016/j.actpsy.2025.104986>
- Navidmanesh, S., Kadkhodae, M., Ghadirinokabadi, L., & Emamifar, A. (2025). Resolving conceptual and practical gaps in current child well-being models through the holistic well-being model for children. *Discover Psychology*, 5(1). <https://doi.org/10.1007/s44202-025-00382-z>
- Rohmad, R., Dharin, A., & Azis, D. K. (2022). Developing self-assessment instruments of affective domain on belief and morality (Aqidah Akhlak) subject in Madrasah Tsanawiyah. *Pegem Eğitim ve Öğretim Dergisi*, 13(1), 193–201. <https://doi.org/10.47750/pegegog.13.01.21>
- Sellami, I., Amin, H., Ozturk, O., Zaman, A., Sever, S. D., & Tok, E. (2025). Digital, localised and human-centred design makerspaces: Nurturing skills, values and global citizenship for sustainability. *Discover Education*, 4(1). <https://doi.org/10.1007/s44217-025-00413-w>
- Setiawati, E., Fajarwati, A., Koenarso, D. A. P., & Herlida Sari, A. (2025). A framework for justice and integrity in digital character education. In *Proceedings of the ICCIT 2025*. IEEE. <https://doi.org/10.1109/ICCIT65724.2025.11167260>
- Urbina, S. (2016). *Essentials of psychological testing* (4th ed.). Wiley.
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2024.101425>

