

Assessing Islamic Character in Elementary School Students: A Multi-Source Educational Management Framework

Rustam¹⁾, Ilham²⁾, Syarifuddin³⁾, Ade S. Anhar⁴⁾, Muhammad Iqbal⁵⁾

^{1,2,3,4)} Universitas Muhammadiyah Bima, Indonesia

⁵⁾ Universitas Pendidikan Mandalika Mataram, Indonesia

Corresponding author email: muhammadiqbal@undikma.ac.id

Received: 15-02-2026

Revised: 26-05-2026

Accepted: 15-07-2026

Info Artikel

Keywords:

Islamic Character
Assessment, Educational
Management, Character
Education, Multi-
Informant Assessment,
Elementary Education

Abstract

Islamic character assessment in elementary schools presents challenges that are at once psychometric and managerial. Curriculum mandates have grown stronger, from the National Education System Law to the Merdeka Curriculum and its Pancasila Student Profile, yet school practice remains impression-based and weakly evidenced. The gap addressed here is the absence of a framework that joins assessment methods, bias mitigation, and educational management into a design that a school can actually run. This conceptual-analytic article synthesises literature on educational assessment, affective measurement, psychometrics, and Islamic educational management to develop a multi-source assessment framework for the primary level. Islamic character is operationalised into three dimensions, namely *ubudiyah*, personal *akhlak*, and social *akhlak*. The analysis shows that assessment methods are systematically complementary rather than hierarchically ordered, and that most threats to validity originate in organisational conditions, not in the instrument. The proposed model triangulates evidence across teachers, parents, and students, and is delivered through three implementation packages calibrated to differing levels of school capacity. The article contributes a design in which managerial feasibility is treated as a precondition of validity rather than a concession against it. Its limitation is that the framework has not been empirically tested; subsequent work should establish content validity, inter-rater reliability, and the managerial conditions under which the packages are sustained.

Kata Kunci:

Penilaian Karakter Islami,
Manajemen Pendidikan,
Pendidikan Karakter,
Penilaian Multi-Informan,
Pendidikan Dasar

Abstrak

Penilaian karakter Islami di sekolah dasar menghadirkan tantangan yang sekaligus bersifat psikometris dan manajerial. Mandat kurikulum semakin kuat, mulai dari Undang-Undang Sistem Pendidikan Nasional hingga Kurikulum Merdeka dan Profil Pelajar Pancasila, namun praktik di sekolah masih berbasis kesan dan lemah bukti. Kesenjangan yang dijawab oleh artikel ini adalah belum tersedianya kerangka yang menyatukan metode penilaian, mitigasi bias, dan manajemen pendidikan ke dalam desain yang benar-benar dapat dijalankan di sekolah. Artikel konseptual-analitik ini menyintesis literatur asesmen pendidikan, pengukuran afektif, psikometri, dan manajemen pendidikan Islam untuk mengembangkan kerangka penilaian multi-sumber di jenjang dasar. Karakter Islami dioperasionalkan ke dalam tiga dimensi, yaitu *ubudiyah*, *akhlak* pribadi, dan *akhlak* sosial. Analisis menunjukkan bahwa metode-metode penilaian bersifat saling melengkapi secara sistematis, bukan tersusun secara hierarkis, dan bahwa sebagian besar ancaman terhadap validitas bersumber pada kondisi organisasional, bukan pada instrumennya. Model yang diusulkan mentriangulasi bukti dari guru, orang tua, dan siswa, serta disampaikan melalui tiga paket implementasi yang disesuaikan dengan kapasitas sekolah. Kontribusi artikel ini terletak pada desain yang memperlakukan kelayakan manajerial sebagai prasyarat validitas, bukan sebagai kompromi terhadapnya. Keterbatasannya, kerangka ini belum diuji secara empiris; penelitian lanjutan perlu menetapkan validitas isi, reliabilitas antar-rater, serta kondisi manajerial yang menopang keberlanjutan paket-paket tersebut.

INTRODUCTION

National education policy in Indonesia places character formation at the centre of its normative aims (Kartiko et al., 2025). The National Education System Law (Law No. 20/2003) defines the purpose of education as the development of learners who are faithful, pious, and of noble character, and subsequent assessment regulation (*Permendikbud No. 23 Tahun 2016*, 2016), made effective assessment a mandatory component of holistic evaluation. The Merdeka Curriculum reinforces this position through the Pancasila Student Profile, whose first dimension is faith, piety, and noble character (Ihwanah et al., 2024; Kartiko et al., 2026). For *madrasah* and Islamic primary schools, the Ministry of Religious Affairs extends the same expectation into quality assurance, requiring that character formation be evidenced rather than asserted. Character assessment is thus, by regulation, a managerial obligation and not merely a pedagogical preference.

The consequences of a weak assessment system are visible in practice. Schools continue to report peer bullying, intolerance between student groups, declining *attitude* toward teachers, and a gap between what children can recite about religious values and how they behave (Nurhidayati et al., 2024). (Lickona, 1999) gives this gap a name: *moral* knowing, *moral* feeling, and *moral* action come apart, and character education fails at the seam. No assessment confined to a single form of evidence can detect that separation, since each form reaches only one of the three. Schools generally treat such incidents as matters of discipline. They are better read as symptoms of a feedback failure. Attitude scores are awarded as Good or Very Good with nothing behind them, and guidance is triggered by whichever incident happens to be visible (Ulumudin & Fujianita, 2019). Working from such data, a school leader cannot tell a struggling child from a quiet one. Character education programmes may therefore run for years, while nobody can say whether they are working.

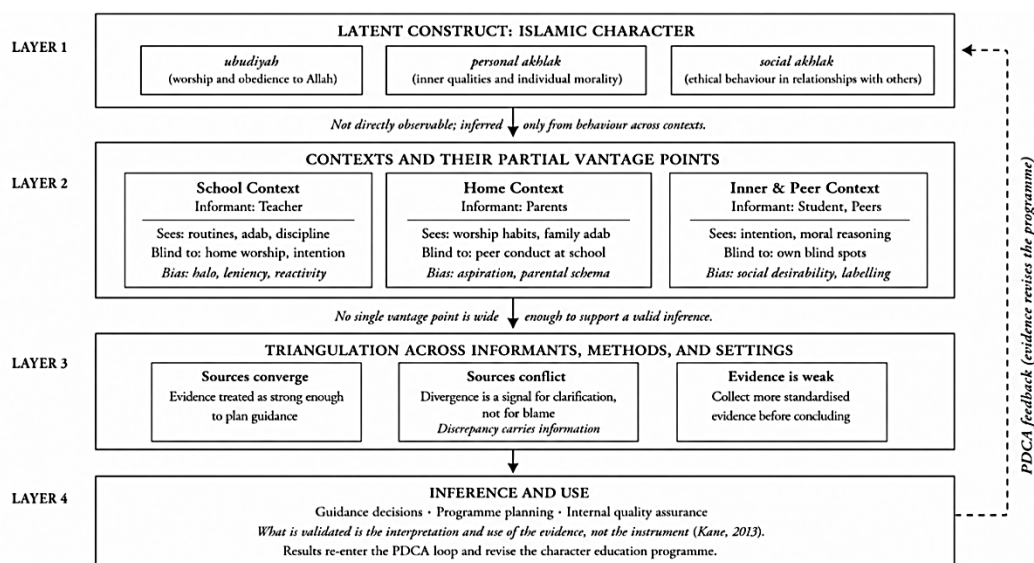
Three bodies of scholarship have addressed parts of this problem. Research on Islamic character education has examined habituation strategies, role modelling, and school culture as mechanisms of formation (Cahyanto et al., 2024; Fauza et al., 2023; Hariyono et al., 2023). Assessment scholarship has advanced the measurement side, establishing the principles of instructionally useful classroom assessment (Popham, 2017), the properties of multi-informant designs, and the psychometric hazards of affective measurement in children (Anthony et al., 2020; De Los Reyes et al., 2015). A third stream, situated in Islamic educational management, has established the planning, organising, and supervisory functions through which *madrasah* and Islamic schools are governed (Machali Imam & Hamid Noor, 2017), and has traced the alignment between assessment policy and classroom practice, finding that affective and psychomotor assessment remain limited, inconsistent, and weakly instrumented despite regulatory mandates (Amri, 2024; Setiadi, 2016; Sukenti et al., 2021).

Policy analyses diagnose the gap but supply no design that closes it. What is missing is a framework linking the dimensions of Islamic character to specific evidence types, mapping the biases each evidence type carries, and situating the whole within the managerial functions that sustain it. The difficulty runs deeper still. (Jukes et al., 2021). examined the forty-three assessment frameworks catalogued in the Harvard EASEL taxonomy and found that none includes obedience among its constructs, although the communities they studied ranked it among the qualities they most wanted children to develop (Jukes et al., 2021). Instruments built elsewhere do not merely fit the Islamic character imperfectly. In places, they lack the construct altogether, and an instrument that never asks cannot be made valid by better administration (Kurniawati & Mawardi, 2021).

Teachers are asked for a comprehensive assessment without any reduction in burden (Zebua & Astuti, 2024).

This article addresses that gap by developing a multi-source framework for assessing Islamic character in elementary schools, read explicitly through the lens of educational management. It pursues three objectives: to map the available assessment methods together with the evidence each can and cannot provide; to construct a cross-method bias map with corresponding mitigation strategies; and to propose a multi-source model with tiered implementation packages calibrated to the managerial capacity of ordinary schools.

Figure 1. The Logic of Multi-Source Islamic Character Assessment



Source: Author Data Collection, 2026

The framework rests on a single proposition, stated here so that it may be tested and, if necessary, refuted. Islamic character is a latent construct (Kane, 2013) locates the object of validation exactly: what is validated is the interpretation and use made of scores, not the instrument that produced them. The question facing a school is thus whether its evidence supports the inference it draws about a child. Triangulation is a coordination problem before it is a psychometric one, which places the answer in the hands of management rather than measurement alone. Figure 1 traces the logic that follows, from the latent construct through the partial vantage points of each informant to a use that can be validated. Its managerial layer draws on Terry's classical functions (Wawan & Wasliman, 2026), the PDCA improvement cycle, Total Quality Management in education (Iyer, 2018), and the Indonesian literature on Islamic educational management (Machali Imam & Hamid Noor, 2017).

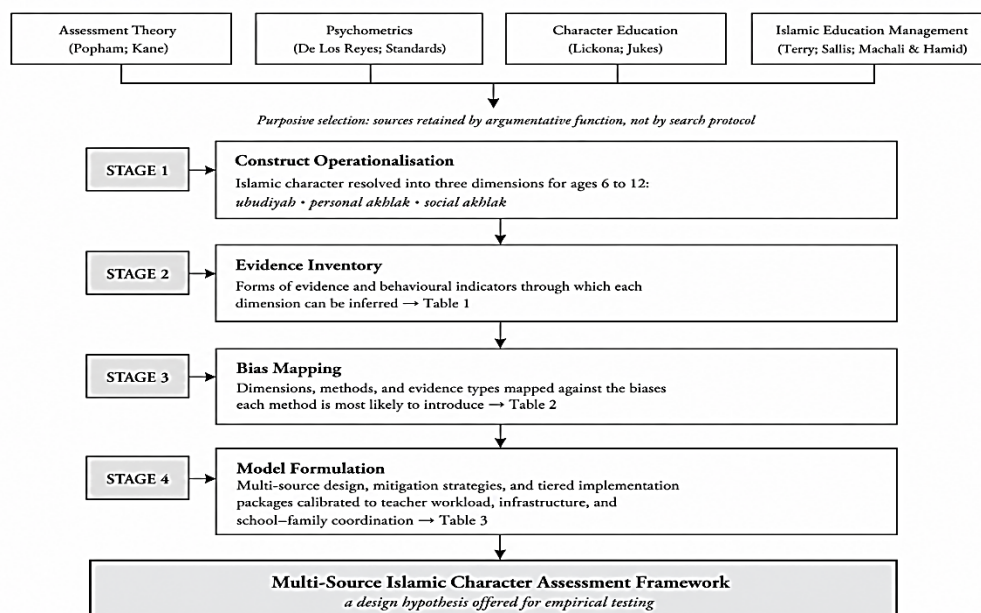
METHOD

This article is a conceptual study and does not report empirical data. The purpose is therefore not to estimate an effect or to summarise empirical findings exhaustively, but to establish the internal logic and practical utility of the resulting framework (Hair, Jr et al., 2021) Because the aim is conceptual integration n, sources were selected purposively rather than through systematic search protocols (Hair et al., 2021). No claim of comprehensiveness is made, and no screening statistics are reported, since these would imply a review logic that this article does not adopt. Sources were retained when they contributed one of four functions to the argument: defining the

construct of Islamic character at the primary level or theorising the managerial conditions under which assessment systems are sustained. Literature published between 2016 and 2025 was prioritised, while older foundational works were retained where they remain the standard reference for a concept (Mushtaq et al., 2025).

The synthesis proceeded in four stages. The construct of Islamic character was first operationalised into three dimensions relevant to children aged 6 to 12 (*ubudiyah*, personal *akhlak*, and social *akhlak*). The forms of evidence and behavioural indicators commonly used to infer these dimensions were then inventoried. Dimensions, methods, and evidence types were subsequently mapped against the biases each method is most likely to introduce. (Hair & Alamer, 2022; Sarstedt et al., 2023).. The resulting map was finally used to formulate a multi-source model, including mitigation strategies and tiered implementation packages calibrated to managerial realities such as teacher workload, school infrastructure, and coordination with parents. Where psychometric and managerial considerations pulled in different directions, the tension was recorded and examined, on the view that both belong to a single argument about validity (Markus & Borsboom, 2024).

Figure 2. Stages of Conceptual Synthesis



Source: Authors Data Collection, 2026

RESULTS AND DISCUSSION

Results

Conceptual Framework of Islamic Character in Elementary Students

Islamic character is defined here as moral, spiritual, and social qualities rooted in Islamic values and expressed consistently in a learner's thoughts, attitudes, and behaviour in relation to Allah SWT, the self, other people, and the natural environment (Komalasari & Yakubu, 2023). For children aged 6 to 12, this construct should be treated as character in formation, strengthened through habituation, role modelling, and sustained guidance. It should not be judged as a fixed personal trait (Arief, 2023).

Drawing on PAI and character education literature, Islamic character can be organised into three dimensions. *Ubudiyah* or worship reflects the vertical relationship with Allah SWT, with indicators such as praying before and after activities, orderly participation in school religious

routines, and proper etiquette during *wudu* and worship (Marlia, 2022). Personal *akhlak* includes honesty, for example, admitting mistakes and not cheating; discipline, such as punctuality and orderliness; and responsibility or trustworthiness (*amanah*), shown in completing tasks and safeguarding entrusted items (Jazil & Firdaus, 2024). Social *akhlak* or *adab* covers politeness and respect, such as greeting and courteous speech, together with empathy, care, and helpfulness toward peers (Abdullah et al., 2025; Pika, 2026).

Developmental principles are essential for validity. Elementary children tend to reason concretely (Piaget) and display pre-conventional to early conventional moral reasoning (Fahiroh et al., 2026; Hasan et al., 2026; Rizqiani et al., 2025). Self-report by young children is therefore vulnerable to social desirability and limited introspection. Character assessment should rely more heavily on adult observations supported by concrete behavioural language and appropriate rater training (De Los Reyes et al., 2015). These developmental considerations directly shape the method selection and bias-mitigation strategies presented below.

Mapping Islamic Character Assessment Methods

The predominance of cognitive assessment in PAI practice has been documented. Rustam et al. (2025) found that PAI teachers generally rely on written tests and semester examinations as primary instruments, while affective assessment is often manifested merely as scores in report cards without a systematic observation process. This pattern is inconsistent with the principles of instructionally useful assessment (Popham, 2017), which require affective assessment to function as a continuous monitoring instrument for the internalisation of Islamic values. In practice, it has become an administrative formality.

Structured observation is a foundational method that captures the frequency and quality of observable behaviour in authentic contexts (Al-Hendawi et al., 2025). Checklists help standardise indicators, while rubrics specify performance levels with behavioural descriptors. Its strengths include closeness to real behaviour and reduced reliance on inference. Its limitations include subjectivity when the criteria are vague and reactivity when children know they are being observed (McKown, 2017). Effective use requires scheduling, rater training, and standardised instruments provided by the school.

Teacher rating scales are practical and leverage teachers' continuous interaction with students (Südkamp et al., 2012). Evidence from multi-informant designs indicates that teacher proxy-reports track direct assessment of children closely enough to serve as a dependable indicator of competence (Bonifacci et al., 2023), which supports their central place in the model proposed here. They are, however, susceptible to halo effects and to leniency or severity bias. Behaviour-anchored rubrics and periodic calibration are recommended to reduce these risks (Cash et al., 2012; Matosas-López et al., 2019). Parent ratings complement teacher data by capturing behaviour at home, particularly worship routines and family *adab*, but they require school-family coordination (Nurhidayati et al., 2024). They also run systematically higher than teacher ratings, a pattern attributed to parental aspiration (Asadi, 2026). Such ratings may also register the parent's own schema about the child as much as the child's observed conduct (Alismail & Aladsani, 2026).

Student self-reports can access internal experience but are limited for elementary students because of social desirability (Baxter et al., 2004). Items should be concrete and situational, and self-reports should function as supplementary evidence (Setiawati et al., 2024). Moral vignettes and situational judgment tasks present dilemmas that elicit moral reasoning and can reduce purely

normative responding (Kruepke et al., 2018; Whetzel et al., 2020), but they require careful readability testing given varied literacy levels.

Portfolios document character development over time and provide longitudinal evidence useful for managerial decisions about guidance programme effectiveness (Earl, 2013; Sackstein, 2024; Yılmaz et al., 2023). A light portfolio design of one page per week is more realistic for elementary contexts. Peer assessment may offer social perspectives but carries ethical risks such as labelling and bullying; where applied, it should be anonymous, aggregated, and focused only on positive behaviours. In summary, each method provides distinct but partial evidence of Islamic character. No single method is sufficient. This complementarity is what motivates a multi-source design. Table 1 provides a systematic comparison.

Table 1. Comparison of Islamic Character Assessment Methods for Elementary School Students

Method	Type of Evidence Generated	Informant	Suitable Dimension	Strengths	Limitations	Notes
Structured Observation	Frequency of behaviour	Teacher	Social <i>akhlak</i> , adab, discipline	Close to actual behaviour	Subjective without clear criteria	Requires operational indicators
Performance Rubric	Quality levels (1–4)	Teacher, mentor	Quality of <i>akhlak/adab</i>	Reduces bias	Requires rater training	Provide behavioural examples for each level
Teacher Rating Scale	Tendency scores	Homeroom /PAI teacher	Across school situations	Practical	Halo effect	Limit the number of items, use a brief scale
Parent Rating Scale	Home behaviour scores	Parents	Home worship, family <i>adab</i>	Captures home context	Social desirability bias	Use a concise and concrete format
Student Self-Report	Self-perception	Student	Attitudes, reflection	Access to internal aspects	High social desirability	Use concrete language; supplementary evidence
Vignettes/Scenarios	Response choices	Student	Moral reasoning	Reduces purely normative answers	Requires readability testing	4–8 short scenarios
Portfolio	Process evidence	Teacher + student	Consistency of habituation	Captures process & continuity	Administrative burden	One page per week
Peer Assessment	Social perception	Peers	Social <i>akhlak</i>	Reflects social dynamics	Risk labelling of	Anonymous, aggregated, focus on positive behaviour

Source: Authors' own Construction, Synthesised from the Literature Reviewed in this Section.

Bias Map in Islamic Character Assessment

Method selection must consider bias risk, because bias distorts inferences about students' character and weakens accountability. For school leaders, a bias map helps align assessment design with the requirements of validity and fairness. Social desirability bias is especially salient in religious character assessment, producing inflated scores with low variance (Nazari et al., 2024). Mitigation includes situational item formats, triangulation with observation, and clear messaging that there are no right answers (Setiawati et al., 2024). Halo effects occur when one prominent attribute, such as academic achievement, contaminates judgments about character (Sanrey et al., 2021; Schmidt et al., 2023). Mitigation includes behaviour-anchored rubrics, separating academic from character judgments, and using more than one rater (Maier et al., 2016).

Leniency and severity bias emerge when raters are systematically too generous or too strict, creating inequity across classes; mitigation includes detailed scoring guidance and inter-rater

calibration (Matosas-López et al., 2019). Reactivity, or Hawthorne effects, can be reduced through repeated observation during routine activities (Harvey, 2018). Context bias reflects differences between behaviour at school and at home, and its mitigation requires a multi-informant approach (De Los Reyes et al., 2015). Ethical issues such as labelling and privacy should be addressed through a formative and developmental orientation and through confidentiality. Together, these nine bias types constitute a risk map that informs both instrument design and implementation protocol. Table 2 presents the full map with mitigation strategies.

Table 2. Bias Map of Islamic Character Assessment and Mitigation Strategies

Type of Bias	Common Source	Impact	Mitigation Strategy
Social Desirability	Student self-report, parent rating	Inflated and low-variance scores	Use situational items, triangulation, and emphasise no right or wrong answers
Halo Effect	The teacher's general impression	Scores influenced by academic achievement/personality	Behaviour-anchored rubrics; separate academic and character assessment
Leniency/Severity	Different standards among raters	Inequity across classes	Scoring guidelines, behavioural examples, rater calibration
Recency Effect	Most recent events remembered	Fluctuating scores	Period-based assessment, systematic observation records
Reactivity	Students are aware of being observed	Unnatural behaviour	Repeated observations during routine activities
Context Bias	Different behaviour at school vs home	Partial conclusions	Multi-informant approach (teacher + parents)
Indicator Ambiguity	Non-operationalised items	Inconsistent interpretation	Clear operational definitions, behavioural examples
Labelling/Stigma	Punitive-oriented assessment	Psychological impact	Formative approach, confidentiality, developmental language
Cultural/Family Bias	Different family backgrounds	Inequity	Reasonable indicators; distinguish opportunity from willingness

Source: Authors' own Construction, Synthesised from the Literature Reviewed

The Recommended Multi-Source Assessment Model

The bias map indicates that reliance on a single source heightens the risk of inaccurate or unfair decisions. A multi-source approach integrates multi-informant data from teachers, parents, and students with multi-evidence inputs from observation, ratings, vignettes, and portfolios (Achenbach, 2006; De Los Reyes et al., 2015). The model is guided by five design principles: the dimensions of Islamic character as the foundation from which indicators, and therefore methods, are derived; evidence triangulation as the basis of valid inference; quality and bias control as a condition on all evidence rather than a correction applied to it; proportionality of teacher workload; and continuity of formative, non-punitive assessment.

Managerially, coordination among informants is a core organising responsibility of school leaders. Selecting an implementation package is a strategic decision that should match human resources, budget, and infrastructure. Assessment results should then feed into guidance programming and programme evaluation as a form of evidence-based decision making (Lee et al., 2024). A practical triangulation rule can be applied without complex score aggregation. Where two or more sources converge on a pattern, the evidence is treated as strong enough to provide guidance. Where sources conflict, the divergence is treated as a signal for clarification. It is never grounds for blaming the student. Where evidence is weak or impression-based, more standardised evidence is collected before any conclusion is drawn.

Between evidence collection and evidence use sits a layer of quality and bias control. The school owns this layer, not the individual teacher. Four provisions constitute it. Indicators and rubrics must be behaviour-anchored and authored centrally, so that raters are not each inventing

their own standard. Raters must be calibrated against one another, and calibration requires scheduled meeting time, not goodwill. The mitigation strategies mapped in Table 2 must be enacted method by method, since a general commitment to fairness changes nothing at the point of scoring. And the whole must be governed by confidentiality and a formative orientation, so that assessment does not become a mechanism of labelling. Treating these four as a gate through which all evidence passes, rather than as a correction applied to results afterwards, is what distinguishes a multi-source system from a merely multi-instrument one. What remains is to make the whole affordable.

Three-tiered packages are proposed. The Minimal Package comprises weekly rubric-based teacher observation using sampling, together with concise monthly parent ratings, adding roughly ten to fifteen minutes of documentation per week. The Moderate Package adds a light portfolio of one weekly reflection sheet, guided by teachers and adapted to children's literacy levels. The Comprehensive Package further adds moral vignettes two or three times per year to capture moral reasoning alongside routine behavioural observation. Table 3 presents the integrated operational matrix.

Table 3. Operational Matrix of the Multi-Source Assessment Recommendation

Islamic Character Dimension	Operational Indicators	Primary Assessment Method	Informant	Type of Evidence Generated	Quality and Bias-Mitigation Notes
Social <i>Akhlak</i> (<i>Adab</i> toward teachers & peers)	Greeting, polite speech, orderly queuing	Structured observation + performance rubric	Teacher	Behaviour frequency + rubric score (1–4) + descriptive notes	Behaviour-anchored rubric; minimum two observations per period
Personal <i>Akhlak</i> (Honesty)	Admitting mistakes, not cheating	Structured observation + vignette/situational task	Teacher + student	Observation notes + vignette-based moral reasoning score	Decision-based vignette; triangulated with observed behaviour
Personal <i>Akhlak</i> (Discipline & responsibility)	Punctuality, task completion	Checklist + teacher rating scale	Teacher	Checklist score + rating scale score	Clear operational definitions to reduce rater ambiguity
Personal <i>Akhlak</i> (Trustworthiness/ <i>Amanah</i>)	Safeguarding entrusted items, fulfilling assigned duties	Performance rubric	Teacher	Rubric score + supporting behavioural evidence	Equal opportunity for behavioural demonstration
Social <i>Akhlak</i> (Empathy & helpfulness)	Helping peers, sharing	Structured observation + peer assessment	Teacher + peers	Observation score + aggregated peer data	Peer data anonymous, aggregated, and positively framed
<i>Ubudiyah</i> (Worship habits)	Prayers, proper wudu etiquette	Parent rating scale + structured observation	Parents + teacher	Home rating score + school observation score	Concrete and context-specific items; cross-context triangulation
Personal <i>Akhlak</i> (Gratitude & positive attitude)	Expressing gratitude, avoiding excessive complaints	Student self-report + structured observation	Student + teacher	Self-report score + observation notes	Self-report treated as supplementary evidence
Consistency of Habituation (Cross-Dimensional)	Sustaining positive behaviour over time	Light portfolio	Teacher + student + parents	Portfolio artifacts + reflective notes	One-page format per week to ensure feasibility

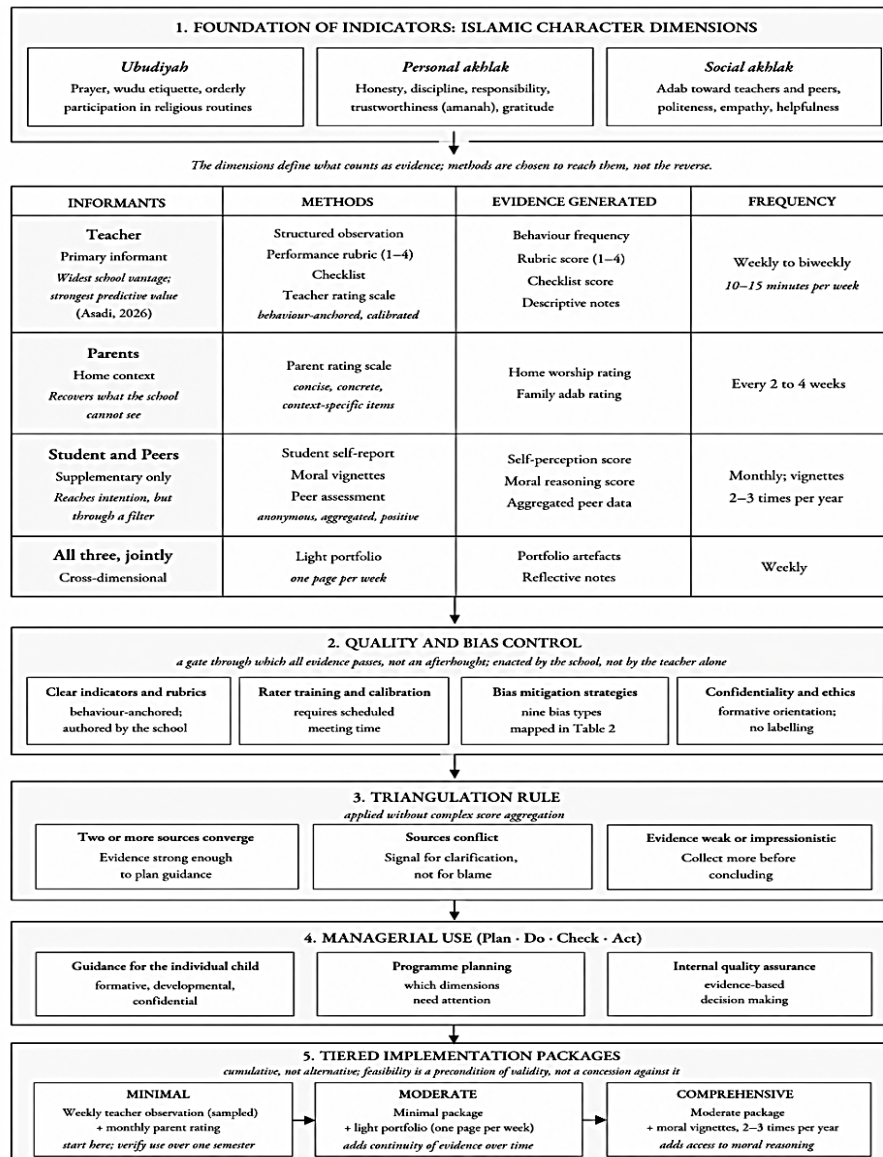
Source: Authors' own Construction, Integrating Tables 1 and 2

Figure 3 sets out the model in operation as a single chain. It begins with the three dimensions of Islamic character, which define what counts as evidence and therefore govern the choice of

method. Evidence then flows from informants and methods through the layer of quality and bias control, into the triangulation rule, and finally into managerial use for guidance, programme planning, and internal quality assurance, with results re-entering the PDCA loop to revise both the programme and the indicators themselves.

The teacher is the primary source, parents supply the home context that the school cannot see, and student self-reports and peer data serve only as supplementary evidence. The three implementation packages are cumulative. Each builds on the one before it, so that a school begins with the Minimal Package and expands only once sustained use has been verified.

Figure 3. Operational Multi-Source Islamic Character Assessment Model



Source: Authors' own Construction, Operationalising in Tables 1,2 and 3.

DISCUSSION

Complementarity, Not Superiority: Why No Method Can Stand Alone

The method comparison in Table 1 is often read as a ranking exercise, as though the task were to identify the single most valid instrument for character assessment. The evidence points the other way. Structured observation stays close to behaviour but cannot reach intention, and it purchases authenticity at the cost of reactivity (Fu et al., 2018; Nunes et al., 2021). Self-report

reaches intention, but in children of primary age, it reaches it through a filter of social desirability and limited introspection (Baxter et al., 2004). Parent ratings recover the home context that teachers cannot see, yet they carry the family's own investment in the child's moral image (Nurhidayati et al., 2024). Each method is strong where another is weak, and the pattern holds across every pairing.

This confirms (De Los Reyes et al., 2015), who argue that discrepancies between informants in child assessment are not measurement error to be averaged away but substantive information about context. It also qualifies a common assumption in the Indonesian literature. Studies that seek to improve character assessment by developing a better single instrument (Kurniawati & Mawardi, 2021; Zebua & Astuti, 2024) treat the problem as one of instrument quality. The present analysis suggests the constraint lies upstream. Validation concerns the interpretation drawn from scores. It follows that no instrument administered from one vantage point can support a valid inference about a construct that is by definition expressed across vantage points, however well that instrument is built (Kane, 2013).

(Asadi, 2026) supplies direct evidence from a Muslim-majority, Arabic-speaking setting. Across four grade levels, parents rated children's social and emotional skills highest, children rated themselves in between, and teachers rated them lowest, with agreement among informants only moderate and the gap widening as children grew older. Only the teachers' ratings predicted academic outcomes to a meaningful degree. Two implications follow for the present framework. Informants are not interchangeable and should not be weighted equally. The discrepancy between them carries information in its own right, since a child rated highly at home and poorly at school is telling the school something about context that neither rating alone conveys. The triangulation rule proposed here accordingly treats divergence as a prompt for inquiry.

The argument so far assumes the instrument at least asks the right question. That assumption does not always hold. (Jukes et al., 2021) report that where a construct valued by a community is absent from the instrument, no amount of triangulation will recover it, because the evidence was never sought in the first place. *Ubudiyah*, *adab*, and *amanah* occupy this position in relation to the mainstream assessment literature. This is why the framework proposed here begins by operationalising the construct into behavioural indicators, and only then asks which instruments can reach them.

Bias as a Managerial Problem, Not Only a Psychometric One

The nine bias types in Table 2 are conventionally treated as psychometric defects, addressed through instrument design. Examining their sources tells a different story. Leniency and severity arise because raters have never been calibrated against one another, which is a coordination failure. Recency effects arise because no systematic record exists between assessment periods, which is a documentation failure. Context bias arises because the school has no channel through which home behaviour reaches it, which is a communication failure. Only social desirability and indicator ambiguity are meaningfully addressed at the level of the item itself (Nazari et al., 2024; Setiawati et al., 2024).

The mitigation column, therefore, reads, on inspection, as a list of managerial actions. Behaviour-anchored rubrics require the school to author and distribute them. Inter-rater calibration requires a scheduled meeting time (Cash et al., 2012; Matosas-López et al., 2019). Multi-informant triangulation requires a functioning school-family relationship. None of these can be enacted by a teacher working alone, which helps explain why the affective assessment gap

documented (Ulumudin & Fujianita, 2019), and corroborated in an independent *madrasah* sample by (Sukenti et al., 2021), has proved so persistent despite two decades of regulation. The problem has been treated as a matter of teacher practice when it is largely a matter of school organisation.

Managerial Feasibility as a Condition of Validity

Positioning the model within educational management theory clarifies what the assessment system is for. In Terry's classical scheme of planning, organising, actuating, and controlling, character assessment is usually assigned to Controlling (Abdulkareem & Adekunle, 2025). The Indonesian literature on Islamic educational management echoes this placement, treating supervision as the terminal function of the managerial cycle (Nurjanah & Hermawan, 2025; Samsuri et al., 2025). The multi-source design extends it into the other three functions. It informs planning by identifying which dimensions require programmatic attention. It constitutes Organising by specifying who reports what and when. It engages by giving parents a defined role. Read through Total Quality Management in education, the model operationalises stakeholder involvement, evidence-based decision making, and continuous improvement (Ah-Teck & Starr, 2014; Mtitu, 2025). The last of these runs as a Plan-Do-Check-Act loop: triangulated data feed the revision of guidance, and the revised programme is then reassessed (Ansari & Rajan, 2026; Rugbani & Nqabisa, 2026).

The tiered packages address the tension that has defeated previous designs. Comprehensive assessment systems fail not because teachers reject them but because they exceed available capacity, at which point compliance becomes fictional, and the data become worthless (Zebua & Astuti, 2024). Feasibility is thus not a concession made against validity. It is a precondition of validity, because an unimplementable system yields no evidence at all, and an interpretation unsupported by evidence cannot be validated on any account of validity (Kane, 2013). The Minimal Package is designed to be sustainable at roughly ten to fifteen minutes of documentation per week. A system of modest scope that a school actually sustains will yield more usable evidence than an ambitious one that exists only on paper.

Table 4. Summary of Analysis: Findings, Contributions, and Recommendations

Research Focus	Key Finding	Contribution to Literature	Recommendation for Schools
Mapping of assessment methods	Methods are systematically complementary; each is strong where another is weak, so no single method supports a valid inference.	Reframes method selection from a ranking problem to a coverage problem; extends De Los Reyes et al. (2015) and Asadi (2026) to Islamic character.	Combine at least two informant types before making any guidance decision.
Bias Map	Seven of the nine bias types originate in coordination, documentation, or communication failures rather than in the instrument.	Relocates bias mitigation from psychometrics to school management; explains the persistence of the affective assessment gap.	Treat rater calibration and parent orientation as scheduled managerial activities, not optional extras.
Multi-Source Model	Managerial feasibility is a precondition of validity rather than a trade-off against it.	Integrates assessment design with POAC and TQM/PDCA; proposes tiered packages calibrated to real capacity.	Begin with the Minimal Package, verify sustained use over one semester, then expand.

Source: Authors' Data Collection, 2026

For principals and *madrasah* leaders, the framework offers a structured basis for integrating Islamic character assessment into internal quality assurance: selecting an appropriate package, allocating time and budget for assessor training, building school-family coordination mechanisms, and using triangulated data to refine character education programmes. The tiered design is

intended to enable gradual capacity building, so that a school may begin with the Minimal Package, evaluate feasibility, and expand as readiness grows. For researchers, three agendas follow. Psychometrically validated instruments are needed for Indonesian elementary contexts, covering all three dimensions with adequate construct validity and inter-rater reliability evidence.

Comparative studies of the implementation packages would provide empirical guidance for policy. Investigations of managerial moderators, including leadership style, organisational culture, and the quality of professional development, would strengthen the evidence base for systemic implementation. At the policy level, Islamic character assessment should be integrated with the attitude assessment systems of the Merdeka Curriculum under a formative and developmental orientation, and supported with adequate regulatory legitimacy and resources. The Ministry of Religious Affairs could incorporate multi-source assessment design as a quality indicator within madrasah accreditation, thereby giving schools a reason to move beyond impression-based practice.

CONCLUSION

This article set out to develop an accountable and implementable framework for assessing Islamic character in elementary schools, addressing three foci. On methods, the analysis found the available instruments to be systematically complementary rather than hierarchically ordered: structured observation, rubrics, rating scales, vignettes, and portfolios each supply partial evidence that no single instrument can supply alone. On bias, it found that most threats to validity originate in the organisational conditions surrounding the instrument. Rater calibration, systematic record-keeping, and school-family communication are managerial provisions before they are psychometric ones. On the model, it found managerial feasibility to function as a precondition of validity, since an assessment system exceeding a school's capacity produces compliance without evidence. The framework that results operationalises Islamic character into three dimensions, links each dimension to specific evidence types and informants, and offers three implementation packages calibrated to differing levels of school capacity.

The contribution of this article lies in joining two literatures that have developed separately, showing that what presents as a measurement problem in Islamic character assessment is also, and perhaps primarily, a problem of school management. It situates that argument within the wider cross-cultural critique of assessment frameworks developed in Western settings, which have been shown to omit outright the constructs that non-Western communities rank among the most important. Its limitation is equally clear. The framework is conceptual and has not been empirically tested, and its dimensions, rubrics, and packages remain propositions awaiting validation. Four lines of work follow. Content validity should be established through expert panels drawing on both PAI and assessment expertise. Inter-rater reliability should be estimated across teacher pairs in SD and MI settings alike. Convergence and divergence between school and home ratings should be examined as a direct test of the triangulation rule.

REFERENCES

- Abdulkareem, F. O., & Adekunle, O. B. (2025). Effect of Strategic Human Resource Management Practices on Human Capital Flight in Selected Hospitals in the Federal Capital Territory, Abuja. *African Journal of Management and Business Research*, 19(1), 334–356. <https://doi.org/10.62154/ajmbr.2025.019.01026>

- Abdullah, S. F. S., Ishak, A. H., & Wahid, F. N. A. (2025). Children's Etiquette Towards Parents Based on Kitab Al- Adab Al-Mufrad By Al-Bukhari: Etika Anak Terhadap Ibu Bapa Berdasarkan Kitab Al-Adab Al-Mufrad Karya Al-Bukhari. *Sains Insani*, 10(1), 51–61.
- Ah-Teck, J. C., & Starr, K. E. (2014). Total Quality Management in Mauritian education and principals' decision-making for school improvement: 'Driven' or 'informed' by data? *Journal of Educational Administration*, 52(6), 833–849. <https://doi.org/10.1108/JEA-06-2012-0075>
- Amri, U. (2024). Authentic Assessment of Attitudes in Islamic Religious Education Subjects in the Formation of Students' Religious Character at SMAN 7 Sijunjung. *Ruhama: Islamic Education Journal*, 7(1), 1–10. <https://doi.org/10.31869/ruhama.v7i1.5417>
- Ansari, M. S., & Rajan, U. C. (2026). *Cybersecurity Risk Management in IT Projects: A Systematic Review and Classification of Risks, Best Practices, and Framework Effectiveness (2020-2025)* (SSRN Scholarly Paper No. 6036034). Social Science Research Network. <https://doi.org/10.2139/ssrn.6036034>
- Anthony, C. J., Elliott, S. N., DiPerna, J. C., & Lei, P.-W. (2020). Multirater assessment of young children's social and emotional learning via the SSIS SEL Brief Scales – Preschool Forms. *Early Childhood Research Quarterly*, 53, 625–637. <https://doi.org/10.1016/j.ecresq.2020.07.006>
- Arief, M. M. (2023). Reshaping Habituation Program to Optimize Children Self-Development in Elementary. *Journal of Media and Pedagogical Practices*, 1(01), 1–11.
- Asadi, I. (2026). Evaluating social–emotional skills among Arabic-speaking students: A multi-informant study of students, parents and teachers. *British Journal of Educational Psychology*, 96(2), 674–688. <https://doi.org/10.1111/bjep.70044>
- Baxter, S. D., Smith, A. F., Litaker, M. S., Baglio, M. L., Guinn, C. H., & Shaffer, N. M. (2004). Children's Social Desirability and Dietary Reports. *Journal of Nutrition Education and Behavior*, 36(2), 84–89. [https://doi.org/10.1016/S1499-4046\(06\)60138-3](https://doi.org/10.1016/S1499-4046(06)60138-3)
- Cahyanto, B., Indana, F. M., Fiveronica, I., Salamah, E. R., & Garbacz, N. A. (2024). Integration of Religious Character in School Culture: An Investigation of Character Development Practices in Islamic Elementary School. *ELEMENTARY: Islamic Teacher Journal*, 12(1), 49–74. <https://doi.org/10.21043/elementary.v12i1.23309>
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. <https://doi.org/10.1037/a0038498>
- Fahiroh, S., Syarnubi, S., Fadil, A., Nasrin, M. S., & Firmansyah, F. (2026). Instilling Friendly Character Through Greeting Activities: The Pedagogical Role of Islamic Education Teachers in Public Junior High Schools. *Tadrib: Jurnal Pendidikan Agama Islam*, 12(1), 15–29. <https://doi.org/10.19109/tadrib.v12i1.33879>
- Fauza, W., Akil, & Sitika, A. J. (2023). Habituation Strategy in Forming Student Character at Al-Fajar Islamic High School, Bekasi City. *Al-Afkar, Journal For Islamic Studies*, 6(4), 450–459. <https://doi.org/10.31943/afkarjournal.v6i4.786>
- Fu, Y., Liu, X., Wang, Y., & Chao, R.-F. (2018). How experiential consumption moderates the effects of souvenir authenticity on behavioral intention through perceived value. *Tourism Management*, 69, 356–367. <https://doi.org/10.1016/j.tourman.2018.06.023>

- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). An Introduction to Structural Equation Modeling. In J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, & S. Ray (Eds), *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-80519-7_1
- Hair, Jr, J. F., M. Hult, G. T., Ringle, C. M., & Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications, Inc. <https://collegepublishing.sagepub.com/products/a-primer-on-partial-least-squares-structural-equation-modeling-pls-sem-3-270548>
- Hariyono, Sm, I., Hidayatulloh, M. A., & Huda, M. (2023). School Culture-Based Internalization of Nationalism and Religious Characters in Islamic Elementary School. *MUDARRISA: Jurnal Kajian Pendidikan Islam*, 15(2), 135–157. <https://doi.org/10.18326/mudarrisa.v15i2.384>
- Hasan, H. M., Djollong, A. F., Naim, M., Hamid, S., & Hamang, N. (2026). Religious Values-Based Role Playing for Character Development in Early Childhood Education: A Case Study at Aisyiyah Bustanul Athfal Kindergarten, Enrekang Regency. *Journal of Education and Learning Sciences*, 6(1), 573–595. <https://doi.org/10.56404/jels.v6i1.340>
- Ihwanah, A., Ismail, F., Astrid, A., Maryamah, & Sukirman. (2024). The Problem of Integrating Pancasila Student Profile with Islamic Religious Education Material Through the Implementation of Content Differentiation Learning in Elementary School. *JIP Jurnal Ilmiah PGMI*, 10(2), 62–75. <https://doi.org/10.19109/jip.v10i2.25371>
- Iyer, V. G. (2018). *Total Quality Management (TQM) or Continuous Improvement System (CIS) in Education Sector and Its Implementation Framework towards Sustainable International Development*. 546–555. <https://doi.org/10.2991/csece-18.2018.119>
- Jazil, T., & Firdaus, A. (2024). The Concept and Practice of Integrity, Honesty, and Fairness in Business: Al-Itqan, Islamic Business Value. *Abdelhak Boudjelida Papers & Citations*, 24, 75–102. <https://doi.org/10.52305/OJLG4208>
- Jukes, M. C. H., Mgonda, N. L., Tibenda, J. J., Gabrieli, P., Jeremiah, G., Betts, K. L., Williams, J., & Bub, K. L. (2021). Building an assessment of community-defined social-emotional competencies from the ground up in Tanzania. *Child Development*, 92(6), e1095–e1109. <https://doi.org/10.1111/cdev.13673>
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedem.12000>
- Kartiko, A., Arif, M., Rokhman, M., Ma'arif, M. A., & Aprilianto, A. (2025). Legal Review of Inclusive Education Policy: A Systematic Literature Review 2015-2025. *International Journal of Law and Society*, 4(1), 22–46. <https://doi.org/10.59683/ijls.v4i1.152>
- Kartiko, A., Rokhman, M., Maarif, M. A., Sirojuddin, A., & Kayadibi, S. (2026). Bridging Policy and Practice: Implementation Gaps in Law Number 14 of 2005 on Teachers and Lecturers during the Merdeka Curriculum Era. *International Journal of Law and Society*, 5(1), 210–222. <https://doi.org/10.59683/ijls.v5i1.628>
- Komalasari, M., & Yakubu, A. B. (2023). Implementation of Student Character Formation Through Islamic Religious Education. *At-Tadzkir: Islamic Education Journal*, 2(1), 52–64. <https://doi.org/10.59373/attadzkir.v2i1.16>

- Kurniawati, D., & Mawardi, M. (2021). Pengembangan Instrumen Penilaian Sikap Gotong Royong dalam Pembelajaran Tematik di Sekolah Dasar. *EDUKATIF: JURNAL ILMU PENDIDIKAN*, 3(3), 640–648. <https://doi.org/10.31004/edukatif.v3i3.387>
- Lickona, T. (1999). Character Education: The Cultivation of Virtue. In *Instructional-design Theories and Models*. Routledge.
- Machali Imam, I., & Hamid Noor, N. (2017). *Pengantar Manajemen Pendidikan Islam; Perencanaan, Pengorganisasian, dan Pengawasan dalam Pengelolaan Pendidikan Islam*. Program Studi Manajemen Pendidikan Islam Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta. <https://digilib.uin-suka.ac.id/id/eprint/38824/>
- Markus, K. A., & Borsboom, D. (2024). *Frontiers of Test Validity Theory: Measurement, Causation, and Meaning* (2nd edn). Routledge. <https://doi.org/10.4324/9781003398219>
- Marlia, A. (2022). Characteristics of School Culture in Inculcating Religious Values at SD Pusri Palembang. *Borneo Educational Journal (Borju)*, 5(1), 78–88. <https://doi.org/10.24903/bej.v5i1.1189>
- Mtitu, E. A. (2025). Exploring factors for the introduction and implementation of quality assurance systems in selected higher education institutions in Tanzania. *Cogent Education*, 12(1), 2455767. <https://doi.org/10.1080/2331186X.2025.2455767>
- Mushtaq, S., Mohsin, M., & Mushtaq, M. M. (2025). A Systematic Literature Review on the Implementation and Challenges of Zero Trust Architecture Across Domains. *Sensors*, 25(19), 6118. <https://doi.org/10.3390/s25196118>
- Nazari, S., Leite, W. L., & Huggins-Manley, A. C. (2024). Enhancing the Detection of Social Desirability Bias Using Machine Learning: A Novel Application of Person-Fit Indices. *Educational and Psychological Measurement*, 84(6), 1107–1137. <https://doi.org/10.1177/00131644241255109>
- Nunes, J. C., Ordanini, A., & Giambastiani, G. (2021). The Concept of Authenticity: What It Means to Consumers. *Journal of Marketing*, 85(4), 1–20. <https://doi.org/10.1177/0022242921997081>
- Nurhidayati, N., Anjarini, T., & Ratnaningsih, A. (2024). Implementation of School-Based Management to Optimize Character Education in Primary School | Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme. *Scaffolding: Jurnal Pendidikan Islam Dan Multikulturalisme*, 6(3), 739–753. <https://doi.org/10.37680/scaffolding.v6i3.8488>
- Nurjanah, S., & Hermawan, A. (2025). Principles of Islamic Education Management: The Perspective of the Qur'an and Hadith in Building Quality Education. *Jurnal Manajemen Pendidikan Islam Darussalam*, 7(1), 61–78. <https://doi.org/10.30739/jmpid.v7i1.3619>
- Permendikbud No. 23 Tahun 2016. (2016). Database Peraturan | JDIH BPK. <http://peraturan.bpk.go.id/Details/224434/permendikbud-no-23-tahun-2016>
- Pika, P. (2026). Morals Towards Friends Based on Islamic Teachings. *Al-Murabbi Jurnal Pendidikan Islam*, 4(1), 81–88. <https://doi.org/10.62086/al-murabbi.v4i1.1162>
- Popham, W. J. (2017). Disposed to Measure Dispositions. *The Teacher Educator*, 52(3), 284–289. <https://doi.org/10.1080/08878730.2017.1316606>
- Rizqiani, I. S., Tobroni, T., & Faridi, F. (2025). Spiritual and mental harmonization: Religious education through the lens of psychology. *Al-Misbah (Jurnal Islamic Studies)*, 13(1), 11–21. <https://doi.org/10.26555/almisbah.v13i1.12634>

- Rugbani, A., & Nqabisa, S. (2026). Curriculum alignment with international graduate attribute standards in engineering education: The CPUT experience. *The Curriculum Journal*, n/a(n/a). <https://doi.org/10.1002/curj.70074>
- Samsuri, S., Sukardi, A., & Basri, H. (2025). Holistic Monitoring Model Based on Islamic Theology in the Education Management Control System. *Nidhomul Haq : Jurnal Manajemen Pendidikan Islam*, 10(3), 796–812. <https://doi.org/10.31538/ndhq.v10i3.272>
- Setiadi, H. (2016). Pelaksanaan penilaian pada Kurikulum 2013. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 166–178. <https://doi.org/10.21831/pep.v20i2.7173>
- Setiawati, F. A., Widyastuti, T., Fathiyah, K. N., & Nabila, T. S. (2024). Minimizing Social Desirability in Questionnaires of Non-Cognitive Measurements. *European Journal of Psychology and Educational Research*, 7(1), 33–43. <https://doi.org/10.12973/ejper.7.1.33>
- Sukenti, D., Tambak, S., & Siregar, E. (2021). Learning Assessment for Madrasah Teacher: Strengthening Islamic Psychosocial and Emotional Intelligence. *AL-ISHLAH: Jurnal Pendidikan*, 13(1), 725–740. <https://doi.org/10.35445/alishlah.v13i1.552>
- Ulumudin, I., & Fujianita, S. (2019). The implementation of attitude assessment in Curriculum 2013 at elementary schools. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 23(1), 46–56. <https://doi.org/10.21831/pep.v23i1.23391>
- Wawan, W., & Wasliman, E. D. (2026). Systematic Management of Morality Development in Strengthening Elementary School Character Education. *Journal of Innovation and Research in Primary Education*, 5(1), 553–565. <https://doi.org/10.56916/jirpe.v5i1.2778>
- Zebua, I. W. S., & Astuti, S. (2024). Pengembangan Instrumen Penilaian Sikap untuk Mengungkap Sikap Disiplin Siswa SD. *Jurnal Basicedu*, 8(5), 3877–3887. <https://doi.org/10.31004/basicedu.v8i5.8691>