

The Measurement Problem in Values-Based Curriculum Management: A Systematic Review of Affective Assessment Instruments in Islamic Educational Institutions

Parino¹, Agus Subagya², Mahmudi³

¹²³ Universitas Bakti Indonesia, Parinoubi1973@gmail.com¹, agus.subagya212@gmail.com²,
mahmudi.bali@gmail.com³

Article history

Submitted: 2026/08/01; Revised: 2026/08/05; Accepted: 2026/08/07

Abstract

The Indonesian Ministry of Religious Affairs introduced the Love-Based Curriculum (Kurikulum Berbasis Cinta, KBC) through Ministerial Decree No. 1503 of 2025, requiring every madrasah to organise intracurricular, cocurricular, and extracurricular activity around five affective commitments known as Panca Cinta. Scholarship has described how madrasah leaders plan and enact this curriculum, yet the evaluative function of curriculum management has attracted far less attention. This review examines how affective outcomes are assessed within values-based curriculum management in Islamic educational institutions, which instruments are reported, what psychometric evidence supports them, and what obstacles prevent measurement. Following the PRISMA 2020 statement, the review searched Scopus, ERIC, DOAJ, Garuda, and Google Scholar for studies published between 2015 and 2026, screened 316 unique records, and retained 30 studies for synthesis. Screening agreement between two independent reviewers was substantial. Thematic synthesis produced four findings. First, affective targets are stated at the level of dispositions rather than behaviours, which leaves them unobservable. Second, assessment practice depends almost entirely on teacher observation, anecdotal records, and documentary evidence collected at classroom level; institutional and supervisory indicators are absent from the corpus. Third, no included study reported construct validity evidence or a reliability coefficient for an affective instrument. Fourth, the barriers reported by practitioners are managerial rather than pedagogical, concerning indicator specification, reporting alignment, and supervisory capacity. The review proposes a three-tier assessment architecture linking classroom evidence, institutional indicators, and supervisory verification, and translates Panca Cinta into low-inference observable indicators. Measurement is positioned as a control function of curriculum management rather than a technical accessory to instruction

Keywords

affective assessment; curriculum management; Islamic education; Love-Based Curriculum; systematic review



© 202x by the authors. This is an open-access publication under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY SA) license, <https://creativecommons.org/licenses/by-sa/4.0/>.

INTRODUCTION

Indonesian madrasahs entered the second semester of the 2025/2026 academic year under a curriculum framework that asks them to organise schooling around affection. Ministerial Decree No. 1503 of 2025 revised the earlier decree on curriculum implementation in Islamic schools and established two governing approaches: deep learning and the Love-Based Curriculum. The latter rests on five commitments, collectively called *Panca Cinta*, that run from love of God and the Prophet through love of knowledge, love of the environment, love of self and other people, and love of country. Madrasahs are expected to express these commitments across subject teaching, cocurricular projects, extracurricular activity, and institutional climate, without adding instructional hours.

The policy has generated a rapid scholarly response. Within eighteen months of the decree, Indonesian journals published conceptual analyses of the philosophical foundations of the curriculum (Qamariah & Anwar, 2025), studies of teacher readiness (Nada & Listiana, 2025), subject-level implementation accounts in Islamic creed and ethics (Nst & Al-Husna, 2025) and in mathematics (Ahmawati et al., 2025), and arguments connecting the curriculum to religious moderation (Aslinda & Amrizon, 2025; Khairani et al., 2025). A smaller set of studies has approached the curriculum from within educational management. A study of Islamic primary schools in Palembang examined planning, implementation, and evaluation, and reported that planning proceeded collaboratively by aligning instructional documents with ministry formats while retaining existing practice, and that implementation depended on persuasive communication and teacher modelling (Basori et al., 2025). Other work has considered how the curriculum might be integrated with the deep learning approach mandated by the same decree (Syafaruddin et al., 2025).

What these accounts share is a common shape. They describe planning in some detail, describe implementation in more detail, and then treat evaluation briefly, often by reporting that assessment is conducted through observation and teacher judgement. The asymmetry is not accidental. Evaluation is the point at which an affective curriculum meets an assessment system built for cognitive attainment, and the literature has so far documented the difficulty rather than resolved it. This matters for reasons that belong to management rather than pedagogy. In the classical account of the management cycle, controlling depends on comparing performance against a standard. Where no standard has been operationalised, the controlling function cannot be exercised, and the remaining functions lose their corrective feedback. A curriculum whose central outcomes cannot be measured is

therefore not merely difficult to assess; it is difficult to manage. Supervision becomes exhortation, quality assurance becomes document verification, and improvement planning proceeds without evidence about whether the previous cycle achieved anything.

The Ministry of Religious Affairs has assigned approximately 4,700 madrasah supervisors to safeguard implementation of the curriculum, which raises a practical question about what those supervisors are expected to examine. The problem is not unique to Indonesia. Research on character education has long observed that value formation resists the instruments used for academic attainment, and a systematic review of character education strategies in primary and secondary schools identified evaluation systems for character as an unresolved difficulty alongside teacher preparation (Hadi et al., 2025). Studies of humanistic learning management in multicultural madrasahs describe managerial arrangements in careful detail while treating outcome evidence as a matter of professional impression (Nuryadin et al., 2024). Work on Islamic education studies more broadly has argued that the field tends to assert normative aims without submitting them to empirical scrutiny (Sahin, 2018).

At the same time, the wider policy environment has increased the stakes attached to measurement. Curriculum management in Islamic institutions has been examined comparatively against general education (Yulia et al., 2025), and integration of Islamic values into curriculum design has been reviewed conceptually (Kholil, 2024), but neither line of work has addressed instrumentation. The result is a literature that argues persuasively for values-based curriculum and says little about how anyone would know whether it worked.

Several reviews have already covered the conceptual terrain. A systematic library review of love-based curriculum literature published from 2020 onward drew on Google Scholar, ERIC, JSTOR, and DOAJ, and was positioned as a comprehensive conceptual analysis of the initiative (Qamariah & Anwar, 2025). A second review examined implementation literature in Islamic primary schools in a single city (Maharani et al., 2025), and a third read the curriculum as a strategy for religious moderation (Khairani et al., 2025). None of these reviews took assessment instruments as the object of analysis, and none appraised the methodological quality of the studies they synthesised. The present review is positioned in that space.

Theoretical Framing

The review adopts curriculum management as its analytical frame rather than assessment theory. Curriculum management is conventionally described through four

functions: planning, organising, implementing, and evaluating. The fourth function is not a terminal reporting activity but the mechanism through which the first three are corrected. Evaluation supplies the comparison between intended and actual performance from which planning decisions in the following cycle are derived. Where that comparison is unavailable, the cycle becomes open rather than closed, and each iteration reproduces the assumptions of the previous one.

Two features of this frame matter for the analysis that follows. The first concerns level. Evaluation in a school system operates at more than one level: teachers evaluate learners, leaders evaluate programmes, and supervisors evaluate institutions. An instrument that functions at one level does not necessarily serve another, and information generated at classroom level must be transformed before it can inform decisions at institutional or system level. The second concerns consequence. Measurement changes what is measured, and the strength of that effect depends on what follows from the result. Assessment designed for improvement and assessment designed for accountability impose different requirements on instrument design, and conflating them tends to degrade both.

Guided by this frame, the review addresses four questions. First, which affective outcome domains are specified in studies of values-based curriculum management in Islamic educational institutions, and at what level of behavioural specificity? Second, which assessment instruments and procedures are reported, and at which level of the management system are they located? Third, what evidence of validity and reliability is reported for those instruments? Fourth, which barriers to affective measurement are identified, and what design principles for instrumentation can be derived from them?

The review makes three contributions. It shifts attention from the conceptual layer of values-based curriculum, which is now well covered, to the measurement layer, which is not. It appraises the methodological quality of the evidence base rather than summarising claims. And it converts the findings into an assessment architecture that distributes measurement responsibility across classroom, institutional, and supervisory levels, together with a translation of Panca Cinta into indicators that can be observed rather than inferred.

METHODS

The review followed the reporting requirements of the PRISMA 2020 statement (Page et al., 2021). A protocol specifying research questions, eligibility criteria, search strategy, appraisal procedure, and synthesis method was prepared before searching

and was not amended during the review. The protocol was not registered in a public registry, which is noted as a limitation.

Eligibility Criteria Because the review addresses a phenomenon rather than an intervention effect, eligibility was framed using the SPIDER tool. The sample comprised Islamic educational institutions, including *raudhatul athfal*, *madrasah ibtidaiyah*, *madrasah tsanawiyah*, *madrasah aliyah*, Islamic boarding schools, and Islamic religious education programmes in general schools. The phenomenon of interest was the management, implementation, or evaluation of a curriculum organised around values, character, humanistic principles, or affective commitments, including but not limited to the Love-Based Curriculum. Eligible designs comprised empirical studies of any type and conceptual or review studies published in peer-reviewed journals. The evaluation component covered any reported procedure, instrument, indicator, or judgement process used to determine whether affective outcomes had been achieved. Qualitative, quantitative, mixed methods, and conceptual research types were all admitted.

Records were excluded if they addressed cognitive attainment only, if the educational setting was not Islamic, if the publication was an opinion piece, news item, or conference abstract without full text, or if no assessment procedure of any kind was described. Studies published between January 2015 and June 2026 were eligible, in Indonesian or English.

Five sources were searched: Scopus, ERIC, DOAJ, Garuda, and Google Scholar. Google Scholar results were limited to the first 200 records sorted by relevance and were used to identify literature not indexed elsewhere, which is a substantial concern for Indonesian educational research. The English search string was ("love-based curriculum" OR "curriculum of love" OR "values-based curriculum" OR "character education" OR "humanistic education") AND ("Islamic education" OR *madrasah* OR *pesantren*) AND (management OR governance OR evaluation OR assessment OR supervision). The Indonesian string was ("kurikulum berbasis cinta" OR "kurikulum cinta" OR "pendidikan nilai" OR "pendidikan karakter") AND (*madrasah* OR "pendidikan Islam") AND (manajemen OR evaluasi OR penilaian OR asesmen OR supervisi). Reference lists of all included studies were screened manually, and forward citation searching was performed on the three most frequently cited records in the corpus.

Selection Process

Records were exported to a reference manager and deduplicated. Two reviewers independently screened titles and abstracts against the eligibility criteria,

then independently assessed full texts. Disagreements were resolved by discussion, and a third reviewer was available for unresolved cases, which did not arise. Inter-rater agreement at the full-text stage was calculated using Cohen's kappa and interpreted against established benchmarks (Landis & Koch, 1977). The search identified 412 records. Removal of 96 duplicates left 316 records for title and abstract screening, of which 241 were excluded. Seventy-five full texts were assessed, and 45 were excluded: 19 addressed cognitive outcomes only, 14 described no assessment procedure, 7 were not set in Islamic educational institutions, and 5 were not peer-reviewed. Thirty studies were included in the synthesis. Agreement between reviewers was substantial ($\kappa = 0.82$).

Data Extraction, Appraisal, and Synthesis

A structured form captured bibliographic details, setting and educational level, study design and sample, the curriculum or values framework in use, stated affective outcome domains, assessment procedures and instruments, any psychometric evidence, the organisational level at which assessment occurred, and reported barriers. One reviewer extracted data and a second verified every entry against the source text. Methodological quality was appraised with the Mixed Methods Appraisal Tool. Conceptual and review studies, which fall outside the scope of that tool, were appraised separately for transparency of method, explicit search description, and traceability of claims to sources. Appraisal results informed the interpretation of findings but did not determine inclusion, since exclusion on quality grounds would have removed most of the recent literature on a curriculum introduced only in 2025. Findings were synthesised thematically (Thomas & Harden, 2008), with coding conducted according to an established procedure for thematic analysis (Braun & Clarke, 2006). Extracted text was coded line by line, codes were grouped into descriptive themes closely tied to the source material, and analytical themes were then generated to address the review questions. Because the corpus contained no comparable effect estimates, statistical pooling was neither possible nor attempted.

RESULTS AND DISCUSSION

Characteristics of the Included Studies

The 30 included studies were published between 2016 and 2026, with 23 appearing in 2025 or 2026. This distribution reflects the recency of the policy rather than any restriction in the search. Twenty-six studies were conducted in Indonesia, and four were located outside Indonesia and retained because they addressed values-based curriculum in Islamic education more generally (Aderibigbe et al., 2023; Gleibermann, 2016; Raihani, 2018; Sahin, 2018).

By design, 11 studies were qualitative case studies, 9 were conceptual analyses, 6 were literature reviews of varying rigour, 3 were descriptive surveys, and 1 was a mixed methods study. No experimental or quasi-experimental design appeared in the corpus. Table 1 summarises the characteristics of the included studies.

Table 1. Characteristics of the Included Studies (N = 30)

Characteristic	Category	n
Year of publication	2016–2019	4
	2020–2024	3
	2025–2026	23
Study design	Qualitative case study	11
	Conceptual analysis	9
	Literature review	6
	Descriptive survey	3
	Mixed methods	1
Educational level	Islamic primary school	12
	Islamic secondary school	7
	Islamic religious education in general schools	4
	Islamic boarding school	3
	Not specified	4
Setting	Indonesia	26
	Outside Indonesia	4
Values framework	Love-Based Curriculum	17
	Character education	6
	Humanistic Islamic education	5
	Multicultural or inclusive education	2

Quality appraisal returned a mixed picture. Among the empirical studies, sampling rationale and data collection procedures were generally reported adequately, but only 4 of 21 described a strategy for addressing researcher influence, and only 6 reported any form of triangulation beyond combining interviews with documents. Among the conceptual and review studies, 5 of 9 described no search procedure at all, and 3 made claims about implementation without citing empirical sources. Only two review studies described a systematic procedure and named their databases (Hadi et al., 2025; Qamariah & Anwar, 2025).

Affective Outcome Domains and Their Specificity

Coding of stated outcomes produced five domains that align closely with the *Panca Cinta* structure, although studies published before 2025 used different

vocabulary for overlapping constructs.

The first domain concerns religious devotion and the relationship between the learner and God, described variously as sincerity, God-consciousness, and spiritual maturity; one study framed it as religious maturity and treated it as the intended outcome of humanistic learning management in secondary schools (Salamuddin et al., 2025). The second domain concerns disposition toward knowledge, expressed as curiosity, persistence, and what several studies called love of learning. The third concerns environmental consciousness, articulated in the ecotheological language adopted in the policy documents (Koto et al., 2025). The fourth, and by far the most frequently reported, concerns disposition toward self and other people, including empathy, compassion, tolerance, and non-violence, and is located at the centre of the contribution of the curriculum to religious moderation (Aslinda & Amrizon, 2025; Danil et al., 2025). The fifth concerns civic and national attachment, which appeared least often and was usually mentioned without elaboration.

The more consequential finding concerns the level at which these outcomes were specified. Across the corpus, outcomes were stated as dispositions rather than as behaviours. Studies reported that learners should become empathetic, should love knowledge, or should demonstrate care for the environment. Very few translated these into what a learner would be seen doing. Of the 30 studies, 6 offered any behavioural description, and in each case the description appeared as an illustrative example within a narrative rather than as a specified indicator. The closest approach connected affective aims to specific lesson activities in mathematics planning (Ahmawati et al., 2025), but the connection ran from value to activity rather than from value to observable evidence.

This is the first substantive result of the review. The corpus specifies constructs at a level of abstraction that no instrument can operationalise. Empathy as a disposition is a latent variable; empathy as a set of observable classroom behaviours is measurable. The literature has consistently chosen the former.

Assessment Instruments and Their Organisational Location

Assessment procedures reported across the corpus clustered into six types. Teacher observation was the dominant procedure, reported in 22 studies, and in every case it was unstructured or minimally structured; no study reproduced an observation schedule or reported the categories teachers were asked to observe. Anecdotal and narrative records appeared in 9 studies, typically as teacher notes on individual learners, and one study described such records within an evaluation process that also

drew on interviews, participant observation, focus groups, and document analysis, although these were research procedures rather than institutional assessment instruments (Basori et al., 2025).

Portfolio and product evidence appeared in 7 studies, mainly in connection with cocurricular projects, and the link between the artefact and the affective claim was rarely explained. Learner self-report appeared in 5 studies in the form of Likert-type statements about attitudes, but no study reproduced the items, reported the number of items, or described how items had been developed. Documentary verification appeared in 8 studies and consisted of checking that lesson plans, syllabi, and school documents referenced the required values; an examination of teacher readiness found documentary alignment to be the most frequently completed requirement and the least informative about practice (Nada & Listiana, 2025). Peer and parental report appeared in 2 studies and in both cases was described as informal. Table 2 summarises these procedures.

Table 2. Assessment Procedures Reported, by Frequency, Structure, and Organisational Level

Procedure	n	Structure reported	Organisational level
Teacher observation	22	None or minimal	Classroom
Anecdotal or narrative records	9	None	Classroom
Documentary verification	8	Checklist of required references	Classroom and institutional
Portfolio and product evidence	7	Task-specific criteria only	Classroom
Learner self-report	5	Likert items, not reproduced	Classroom
Peer or parental report	2	None	Classroom
Institutional affective indicator	0	Not applicable	Institutional
Supervisory outcome instrument	0	Not applicable	System

Locating these procedures within the management system produced the second substantive result. Twenty-eight of the 30 studies placed assessment entirely at classroom level, in the hands of individual teachers exercising professional judgement. Two studies described institution-level review, and in both cases the review examined documents rather than outcomes. No study in the corpus described an instrument used by madrasah supervisors to verify affective outcomes, and no study described an

institutional indicator that would allow the affective climate of one madrasah to be compared with another or with its own performance in a previous period.

The consequence is that the evaluation function of curriculum management is delegated downward to the point of instruction and does not return upward as management information. Studies of curriculum management in Islamic institutions have described planning and organising as collective and leader-driven activities (Aliyah et al., 2024; Yulia et al., 2025), yet the evidence base for evaluating affective outcomes stops at the classroom door.

Psychometric Evidence

The third result can be stated briefly. No included study reported a reliability coefficient for an affective instrument. No study reported evidence of construct validity through factor analysis, item response modelling, or any other procedure. Three studies referred to expert review of instruments, which constitutes a claim to content validity, but none described the number of experts, their qualifications, the review criteria, or any agreement statistic. No study reported measurement invariance across school types, learner age, or gender, and no study reported test-retest stability.

Among the quantitative and mixed methods studies, sample sizes ranged from 32 to 214, which would in principle support exploratory analysis of a short instrument. None conducted such an analysis, so the absence is not explained by sample constraints. Two studies acknowledged the problem directly, reporting that teachers regarded the absence of assessment guidance as the most difficult aspect of preparing for the curriculum (Nada & Listiana, 2025) and identifying character evaluation systems as a persistent unresolved challenge across the character education literature (Hadi et al., 2025).

Barriers to Affective Measurement

Coding of barriers produced five themes. Construct abstraction was reported in 18 studies; participants and authors described difficulty in deciding what would count as evidence that a learner had developed a given value, and framed the difficulty as conceptual rather than technical. Teachers did not report lacking instruments so much as lacking agreement about what the instrument would measure.

The absence of operational indicators in policy documents was reported in 12 studies. Guidance documents specify the values and describe integration strategies but stop short of indicators, and where indicators were mentioned they restated the value in different words. Reporting system misalignment was reported in 9 studies.

Madrasah reporting infrastructure is organised around subject-level attainment, so affective judgements have no destination; teachers either omit them or convert them into a general behavioural remark, and two studies noted that this conversion destroys whatever information the original judgement carried.

Teacher assessment capacity was reported in 8 studies. Training on the curriculum concentrated on philosophy and instructional method, while assessment received little attention and was treated as a matter of attitude rather than technique. Social desirability and performativity was reported in 5 studies: where learners know that a value is being assessed, self-report and observed behaviour both shift, and one study observed that this risk grows when affective judgements carry consequences for learners or for institutional reputation.

Four of these five barriers are managerial. They concern specification, documentation infrastructure, professional development, and the incentive structure surrounding measurement. Only the last is a technical measurement problem in the narrow sense. This finding reframes affective assessment as a design problem for curriculum management rather than a limitation of assessment science.

Measurement as the Control Function of Curriculum Management

The pattern across the corpus is consistent. Values-based curriculum in Islamic educational institutions is planned collectively, implemented through modelling and persuasion, and evaluated through professional impression. The first two elements are documented in detail; the third is documented as a difficulty. Read through the management cycle, this describes a system operating without its control function.

The practical consequence is visible in supervision. Where supervisors are charged with safeguarding curriculum quality but have no outcome indicators, supervision reduces to verifying that documents exist and that the required vocabulary appears in them. Studies in the corpus describe exactly this pattern, and documentary alignment has been found to be the requirement most reliably satisfied and least informative about practice (Nada & Listiana, 2025). Document compliance becomes a proxy for value formation, a substitution the literature has noticed without naming. A second consequence concerns improvement planning. Madrasahs are expected to use evaluation results to revise curriculum documents in the following cycle, but where evaluation produces narrative impressions rather than comparable indicators, revision cannot be evidence-based, and the annual planning cycle repeats the assumptions of the previous year.

The corpus attributes measurement difficulty to the abstract character of the

values themselves, and this explanation is incomplete. Affective constructs of comparable abstraction are routinely measured in educational psychology, and the wider literature on values education contains instruments for empathy, prosocial behaviour, and environmental attitude. The obstacle is that the policy framework and the research literature have both stopped at the level of the construct. *Panca Cinta* names five commitments, and neither the guidance documents nor the studies reviewed here decompose those commitments into behavioural referents. Without decomposition there is nothing for an instrument to sample. This is a specification failure rather than a measurement failure, and it belongs to the actors who design curriculum documents rather than to the teachers asked to assess.

A second reason is visible in the studies that connect the curriculum to religious moderation (Aslinda & Amrizon, 2025; Danil et al., 2025; Khairani et al., 2025). Values in this framework are held to be intrinsic and non-instrumental, so measuring them can appear to convert a moral commitment into a performance target, which practitioners in several studies described as contrary to the spirit of the curriculum. The concern is legitimate and has a long history in character education. It does not follow that measurement should be abandoned, only that its consequences should be designed carefully.

A Three-Tier Assessment Architecture

The findings support a distribution of assessment responsibility across three organisational levels rather than concentration at classroom level. At the classroom evidence layer, teachers collect low-inference behavioural evidence during ordinary instruction using structured observation categories rather than open impression. The evidence is descriptive and is not converted into a grade; its function is to inform the instructional decisions of the teacher and to feed the next tier in aggregated, non-identifying form.

At the institutional indicator layer, the madrasah maintains a small set of climate and behaviour indicators derived from aggregated classroom evidence and from independent sources, such as incident records, participation rates in cocurricular activity with environmental or social content, and short anonymous learner and parent surveys administered twice a year. These indicators describe the institution rather than individual learners, which removes most of the performativity risk identified in the corpus.

At the supervisory verification layer, supervisors examine the integrity of the process rather than the level of the outcome. Their instrument asks whether classroom

categories are specified and used, whether institutional indicators are collected and reviewed, whether the review produced documented decisions, and whether those decisions appear in the following planning cycle. This gives supervisors an object of inspection other than documents.

The architecture responds directly to the barriers identified above. The first tier addresses construct abstraction by requiring behavioural categories. The second addresses reporting misalignment by creating an institutional destination for affective information. The third addresses supervisory capacity by defining a verifiable object. Separating individual evidence from institutional indicators addresses the performativity concern.

Applying the architecture requires behavioural referents for each commitment. Love of God and the Prophet may be evidenced through voluntary participation in congregational worship without supervision, consistency between stated commitment and conduct when unobserved, and reference to religious reasoning when explaining ethical choices. Love of knowledge may be evidenced through persistence on tasks after initial failure, self-initiated extension of a task beyond its requirements, and questions that seek explanation rather than confirmation. Love of the environment may be evidenced through unprompted corrective action in shared spaces and participation in maintenance activity that carries no reward. Love of self and other people may be evidenced through assistance offered to lower-status peers, restraint under provocation, accurate reporting of one's own error, and intervention in peer conflict without adult prompting. Love of country may be evidenced through participation in civic activity beyond ceremonial requirement and through accurate, non-derogatory description of groups other than one's own. Table 3 maps these referents to evidence sources at each tier.

Table 3. Mapping of Panca Cinta Commitments to Low-Inference Indicators and Evidence Sources

Commitment	Behavioural referent	Classroom evidence source	Institutional indicator
Love of God and the Prophet	Voluntary worship without supervision; religious reasoning in ethical explanation	Structured observation during unsupervised periods	Proportion of unsupervised congregational attendance
Love of knowledge	Persistence after failure; self-initiated task extension; explanatory questioning	Task-completion records; classified question log	Rate of voluntary participation in academic enrichment
Love of the	Unprompted	Observation	Waste and

environment	corrective action in shared spaces; unrewarded maintenance activity	schedule for shared spaces	energy indicators; project participation rate
Love of self and other people	Assistance to lower-status peers; restraint under provocation; self-reported error	Structured peer-interaction observation	Incident records; anonymous climate survey
Love of country	Non-ceremonial civic participation; accurate description of other groups	Discourse sample from civic tasks	Community engagement records

These referents are illustrative rather than definitive. Their value lies in demonstrating that the constructs can be brought to a level at which agreement between observers becomes possible, which is the precondition for any reliability estimate.

CONCLUSION

This review examined how affective outcomes are assessed within values-based curriculum management in Islamic educational institutions, drawing on 30 studies published between 2016 and 2026. The evidence supports four conclusions. Affective outcomes are specified as dispositions rather than behaviours, which places them beyond the reach of any instrument. Assessment practice consists almost entirely of unstructured teacher observation located at classroom level, with no instrument operating at institutional or supervisory level. No study in the corpus reported construct validity evidence or a reliability coefficient. The barriers practitioners describe are predominantly managerial, concerning indicator specification, reporting infrastructure, professional development, and the consequences attached to measurement.

Taken together, these findings identify measurement as the weakest element of a curriculum reform that is otherwise well documented. The review proposes a three-tier architecture that assigns behavioural evidence to teachers, institutional indicators to school leaders, and process verification to supervisors, and it offers behavioural referents for each of the five commitments in the Panca Cinta framework. The proposal is a starting point for instrument development rather than a finished solution. Until such instruments exist and have been validated, the evaluative function of curriculum management in Indonesian madrasahs will continue to rest on impressions that cannot be aggregated, compared, or acted upon.

REFERENCES

- Aderibigbe, S. A., Idriz, M., Alzouebi, K., AlOthman, H., Hamdi, W. B., & Companioni, A. A. (2023). Fostering tolerance and respect for diversity through the fundamentals of Islamic education. *Religions*, 14(2), 212. <https://doi.org/10.3390/rel14020212>

- Ahmawati, F. D., Zainuri, A., & Arni, A. (2025). Perencanaan pembelajaran matematika kurikulum berbasis cinta pada SMK Kesehatan Athallah. *Katalis Pendidikan: Jurnal Ilmu Pendidikan dan Matematika*, 2(4), 1–7. <https://doi.org/10.62383/katalis.v2i4.2389>
- Aliyah, N., Thabrani, A. M., Amal, B. K., & Samosir, S. L. (2024). Based Islamic education curriculum management. *Al-Hayat: Journal of Islamic Education*, 8(3), 1158–1172.
- Aslinda, A., & Amrizon, A. (2025). Curriculum of love and religious moderation in madrasah learning: Strategy for building a generation with humanistic and tolerant character. *Indonesian Journal of Teaching and Learning*, 4(2), 99–107. <https://doi.org/10.56855/intel.v4i2.1477>
- Aulia Gusli, R., Iswantir, M., Akhyar, M., & Lestari, K. M. (2024). Inovasi kurikulum pendidikan Islam era 4.0 di MTsN 1 Pariaman. *Idarah Tarbawiyah: Journal of Management in Islamic Education*, 5(1), 77–88. <https://doi.org/10.32832/itjmie.v5i2.16401>
- Basori, R., Zainuri, A., & Mahendra, A. (2025). Implementation and management of a love-based curriculum in Madrasah Ibtidaiyah of Palembang. *Journal of Educational Sciences*, 9(5), 3731–3743. <https://doi.org/10.31258/jes.9.5.p.3731-3743>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Danil, H., Syafaruddin, B., & Sarda, M. (2025). Insertion of a love-based curriculum in multicultural Islamic religious education. *ETDC: Indonesian Journal of Research and Educational Review*, 4(4), 1323–1335. <https://doi.org/10.51574/ijrer.v4i4.3745>
- Gleibermann, E. (2016). A curriculum of love. *Tikkun*, 31(4), 54–57. <https://doi.org/10.1215/08879982-3676912>
- Hadi, Y., Kholis, N., Remanita, Y., & Harta, L. I. (2025). Systematic literature review on character education strategies in primary and secondary schools. *Journal of Educational Research and Practice*, 3(2), 321–340. <https://doi.org/10.70376/jerp.v3i2.389>
- Idris, S., & Adawiah, R. (2024). Implementasi Kurikulum Merdeka pada Madrasah Ibtidaiyah di Indonesia. *Tawshiyah: Jurnal Sosial Keagamaan dan Pendidikan Islam*, 19(1), 30–43. <https://doi.org/10.32923/taw.v19i1.4512>
- Ilyas, E. L., Setyaningrum, N., & Sumarni, S. (2022). A model of inclusive education curriculum in Islamic education institutions: A case study in Banten Province, Indonesia. *Jurnal Pendidikan Agama Islam*, 19(2). <https://doi.org/10.14421/jpai.2022.192-02>
- Khairani, V., Fitriani, F., Ekowati, E., Daulay, I. R., Darmawan, D., Anggraini, V., & Aslami, S. (2025). Kurikulum cinta sebagai strategi moderasi beragama dalam dunia pendidikan: Tinjauan literatur. *Studia Sosia Religia*, 8(2), 75–83. <https://doi.org/10.51900/ssr.v8i2.25008>
- Kholil, M. (2024). Integrasi nilai-nilai Islam dalam kurikulum pendidikan: Studi literatur pada perspektif pendidikan berbasis tauhid. *Al-Mikraj: Jurnal Studi*

- Islam dan Humaniora, 5(1), 2206–2214. <https://doi.org/10.37680/almikraj.v5i01.6929>
- Koto, M. K., Hasibuan, E. K., Sandi, R. R., Siregar, A. S. B., & Darlis, A. (2025). Pendidikan Islam dan kurikulum cinta. *Nusantara: Jurnal Ilmu Pengetahuan Sosial*, 12(8), 3278–3287. <https://doi.org/10.31604/jips.v12i8.2025.3278-3287>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Maharani, A., Nugraha, L., Zainuri, A., & Hamzah, A. (2025). Implementasi kurikulum berbasis cinta pada Madrasah Ibtidaiyah di Kota Palembang: Sebuah studi literatur. *El-Banar: Jurnal Pendidikan dan Pengajaran*, 8(2), 100–111. <https://doi.org/10.54125/elbanar.v8i2.884>
- Mukhsin, M., Sahudi, S., Hidayatulloh, Y., Nasir, M., & Alfani, I. H. D. (2025). Love-based curriculum in Indonesian madrasahs: A conceptual analysis toward inclusive-humanistic Islamic education. *Syamil: Journal of Islamic Education*, 13(3), 315–336.
- Nada, Z. Q., & Listiana, H. (2025). Analisis kesiapan guru madrasah dalam insersi kurikulum cinta. *Entita: Jurnal Pendidikan Ilmu Pengetahuan Sosial dan Ilmu-Ilmu Sosial*, 385–400. <https://doi.org/10.19105/ejpis.v1i.19188>
- Nst, P. T., & Al-Husna, K. I. (2025). Implementasi kurikulum berbasis cinta pada pembelajaran akidah akhlak di Madrasah Aliyah Negeri 1 Mandailing Natal. *Al-Aulia: Jurnal Pendidikan dan Ilmu-Ilmu Keislaman*, 11(2), 202–211. <https://doi.org/10.46963/aulia.v11i2.3462>
- Nuryadin, N., Salamah, S., Nashrulloh, G. K. M. R., & Norlaila, N. (2024). Humanistic-based learning management: Harmonizing multiculturalism in building an inclusive learning environment. *Al-Tanzim: Jurnal Manajemen Pendidikan Islam*, 8(1), 159–173. <https://doi.org/10.33650/al-tanzim.v8i1.7067>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Qamariah, Z., & Anwar, K. (2025). Analisis konseptual kurikulum cinta: Pendekatan humanistik dan implikasinya terhadap pendidikan Islam. *Nusantara: Jurnal Pendidikan Indonesia*, 5(2), 426–442. <https://doi.org/10.62491/njpi.2025.v5i2-13>
- Raihani, R. (2018). Education for multicultural citizens in Indonesia: Policies and practices. *Compare: A Journal of Comparative and International Education*, 48(6), 992–1009. <https://doi.org/10.1080/03057925.2017.1399250>
- Rumina, R. (2017). Islam sebagai paradigma pendidikan moral universal. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, 2(3), 114–124. <https://doi.org/10.31538/ndh.v2i3.28>
- Sahin, A. (2018). Critical issues in Islamic education studies: Rethinking Islamic and Western liberal secular values of education. *Religions*, 9(11), 335. <https://doi.org/10.3390/rel9110335>

- Salamuddin, A., Rahmat, M., Firmansyah, M. I., & Suresman, E. (2025). Humanistic approach to Islamic education learning management in shaping religious maturity among high school students: An exploratory study. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, 10(2), 411–429.
- Syafaruddin, B., Sarda, M., Mawaddah, N., & Agustinanda, B. (2025). Integration of love-based curriculum and deep learning in madrasahs: A holistic approach to Islamic education. *ETDC: Indonesian Journal of Research and Educational Review*, 5(1), 681–691. <https://doi.org/10.51574/ijrer.v5i1.4367>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. <https://doi.org/10.1186/1471-2288-8-45>
- Wawan, W., & Wasliman, I. (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>
- Yulia, F., Arizal, J., Wijaya, C., Hidayat, R., & Azad, I. (2025). Comparative study of curriculum management models between Islamic educational institutions and general educational institutions. *Jurnal Tadzkiyyah: Pendidikan Islam*, 16(2), 353–368.