fikroh
Jurnal Pemikiran dan Pendidikan Islam

Vol. 19, No. 3 September 2026
p-ISSN : 2087-7501
e-ISSN : 2715-4459

Beyond Restraint: Mapping Akhlakul Karimah and Its Implications for Moral Education in Madrasah Aliyah

Ahmad Manshur[1], Giati Anisah[2*], Zumrotul Fauziah[3]
[1,2,3,] Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Indonesia
*e-mail: giati@unugiri.ac.id

Article Information
Received: July 26 , 2026
Revised: September 27, 2026
Accepted: November 03, 2026

Keywords
Akhlakul Karimah; Akidah Akhlak; Madrasah Aliyah; Empathy; Academic Honesty

ABSTRACT

Moral education in madrasahs is commonly assessed through a single global score, so the specific behaviours in which students are strong or weak remain unmapped. This study used a cross-sectional descriptive-comparative survey to map the akhlakul karimah (noble character) profile reported by Madrasah Aliyah students across four domains of Islamic ethics: devotion to parents (birr al-wālidayn), care for orphans, honesty and the refusal of corruption, and discipline in prayer. A 23-item Likert-type questionnaire was completed by 126 students; every record met the pre-specified data-quality criteria and all were analysed using descriptive statistics, the Friedman test, the Kruskal–Wallis test, and three sensitivity analyses. Students reported a high level of akhlakul karimah (M = 4.14; 82.9%), and the four domains did not differ significantly, $\chi^2(3) = 7.71$, $p = .052$, $W = .02$. The variation lay at indicator level: restraint from prohibited acts and ritual worship were endorsed most strongly, whereas empathic help for orphans (M = 3.59) and moral judgement when a parental order conflicts with religious teaching (M = 3.48) were the weakest; grade X students reported higher scores than grades XI and XII ($\varepsilon^2 = .18$). Theoretically, the findings locate the gap between moral knowledge and moral action in a specific region of moral life: restraint is internalised earlier and more uniformly than active care and practical wisdom, so aggregate moral scores systematically conceal this asymmetry. Practically, Akidah Akhlak learning and its assessment should move from prohibition-centred content toward empathic experience, dilemma-based reasoning, and indicator-level evaluation. The findings rest on self-reports from a single madrasah and on an instrument of limited reliability.

INTRODUCTION

Shaping moral character (akhlak) is the central aim of Islamic Religious Education (Pendidikan Agama Islam, PAI). In Islamic education, success is not measured only by how much knowledge learners have mastered; it is judged equally by the maturity of the conduct that grows out of that knowledge (Sahin, 2018; Tambak, Yasin, et al., 2026). In madrasahs this aim is formalised through the subject of Akidah Akhlak (Islamic creed and morals), one of the PAI subjects that

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

distinguish madrasahs from general schools, and it is pursued within an Indonesian Islamic schooling system that is itself under pressure to define the values it transmits (Nasir & Rijal, 2021; Zuhdi, 2018). The Madrasah Aliyah (MA), the Islamic senior secondary school, matters particularly here because its students are in late adolescence, an age at which religious experience strongly shapes identity, values, and social behaviour (Hardy et al., 2019).

The urgency of moral education becomes clear when integrity in Indonesian schools is examined. The 2024 Education Integrity Assessment Survey (Survei Penilaian Integritas Pendidikan) involved 449,865 respondents from 36,888 educational institutions. It recorded an Education Integrity Index of 69.50, placing it at the "corrective" level, and found that cheating still occurs in 78% of schools and 98% of higher education institutions (Komisi Pemberantasan Korupsi, 2025). Other studies support this picture. Senior high school students do not regard many forms of academic dishonesty as violations (Fadila, 2022), and academic dishonesty has been linked to the way learners reason about moral problems (Ampuni et al., 2020). Meta-analytic evidence shows that academic dishonesty rises with performance pressure and achievement motivation (Krou et al., 2021) and with the belief that one's peers are cheating (Zhao et al., 2022), which makes it a problem of moral climate rather than of individual knowledge alone. Many studies accordingly call for anti-corruption education to be integrated into Islamic education through instruction, institutional culture, or the hidden curriculum (Abdurrahman et al., 2021; Firman et al., 2021; Suyadi et al., 2021).

The four behavioural domains examined in this study were not chosen arbitrarily. Each represents a distinct field of obligation in Islamic ethics, and together they span both the vertical and the horizontal dimensions of akhlak. The first, devotion to parents (birr al-wālidayn), is the domain of ethics toward one's parents: the Qur'an joins the command to worship God with the command not to speak harshly to parents (QS. Al-Isrā' [17]: 23–24), while setting an explicit limit on that obedience (QS. Luqmān [31]: 15). The second, care for orphans, is the domain of social ethics, in which the Qur'an both forbids consuming the property of orphans unjustly (QS. An-Nisā' [4]: 10) and demands active concern, treating the repulsing of the orphan as a denial of religion (QS. Al-Mā'ūn [107]: 1–3; QS. Aḍ-Ḍuḥā [93]: 9). The third, honesty and the refusal of corruption, is the domain of moral integrity, grounded in the prohibition of ghulūl, the betrayal of entrusted property (QS. Āli 'Imrān [3]: 161). The fourth, discipline and care in prayer, is the domain of servitude and spiritual discipline, the ritual practice through which madrasahs routinely cultivate self-regulation and religious character. What makes this set analytically useful is that each domain carries both a prohibition and a positive demand, so the same four domains can show whether students' moral conduct is equally developed on both sides. The domains also matter empirically: religiosity is positively related to students' prosocial behaviour (Safrilsyah et al., 2021), yet religion reaches adolescents through socialisation, relationships, and experience rather than automatically (Hardy et al., 2019), and prosocial conduct is itself multidimensional, so that helping in one context does not predict helping in another (Carlo & Padilla-Walker, 2020).

There is considerable research on the moral conduct of MA students, yet the findings remain inconsistent. Lubis (2022) found that the school environment, teacher competence, and educational quality jointly explained 49.4% of MA students' moral conduct, whereas Muttaqin et al. (2023) reported that Akidah Akhlak learning had no significant effect on students' akhlakul karimah ($r = .08$). Qualitative studies describe various ways of fostering morals, such as role modelling, advice, and habituation (Tambak, Yasin, et al., 2026), while meta-analytic work on character education outside the Islamic education literature indicates modest but reliable gains in student behaviour (Jeynes, 2019). The persistent distance between intensive moral instruction and conduct that has not fully changed is known in moral education research as the gap between moral knowledge and moral action (Darnell et al., 2019; Kristjánsson et al., 2021).

These strands of research leave three gaps, and each carries a consequence that the present study is designed to address. The first is empirical. Moral conduct in madrasahs is almost always reported as a single composite score, or else studied through the methods used to form it, so the field knows how high students score but not where within their conduct that score comes from. The consequence is not merely imprecision: a high aggregate can be produced by strong compliance with prohibitions while active care and moral judgement remain undeveloped, and the aggregate itself conceals that possibility. Mapping specific indicators across several domains simultaneously is therefore not a descriptive refinement but the only way to establish whether the knowledge–action gap is general or concentrated in particular kinds of moral demand (Darnell et al., 2019). The second gap is methodological. Surveys of religious and moral behaviour rest on self-report, which is especially vulnerable to socially desirable responding (Bergen & Labonté, 2020; Larson, 2019) and to careless or non-independent responding when questionnaires are completed within class groups (Curran, 2016; Ward & Meade, 2023). Screening for data quality, item-level examination, and effect sizes are nevertheless rarely reported in this literature, so a reader cannot judge how much of a reported moral profile is conduct and how much is measurement. The consequence is a body of findings that cannot be compared with one another or accumulated. The third gap is contextual. Very little is known about how moral profiles differ between grade levels within one madrasah, although this is precisely the information an Akidah Akhlak teacher needs in order to decide what to emphasise in each year. Taken together, the three gaps mean that the field cannot at present say which aspects of students' moral conduct its instruction is actually failing to reach.

This study pursues three objectives. The first is to determine the level of akhlakul karimah reported by MA students, overall and in each of the four domains. The second is to identify which specific indicators are the strongest and which the weakest, within and across those domains. The third is to test whether reported akhlakul karimah differs between the four domains and between grade levels. The novelty of the study lies in combining indicator-level mapping with explicit data-quality screening and sensitivity analysis, so that the findings state not only how favourable students' reported conduct is, but where it is uneven and how far that unevenness survives changes in the analytic decisions. The results are intended to give Akidah Akhlak teachers a basis for selecting better-targeted learning strategies, and madrasah administrators a basis for designing moral formation programmes.

METHODS

This study used a quantitative approach with a cross-sectional descriptive-comparative survey design (Creswell & Creswell, 2018). A cross-sectional descriptive-comparative survey was chosen because the aim was to map the level of akhlakul karimah reported by students at a single point in time and to compare it across domains and across grade levels, rather than to test the effect of an intervention. The same design term is used consistently in the abstract, in this section, and in the discussion of limitations. The study was conducted at a Madrasah Aliyah in Bojonegoro Regency, East Java, Indonesia. Participants were recruited by purposive sampling, with three inclusion criteria: active enrolment in grade X, XI, or XII; current enrolment in the subject Akidah Akhlak; and voluntary willingness to complete the questionnaire. A total of 126 students returned a complete questionnaire: 26 in grade X (20.6%), 70 in grade XI (55.6%), and 30 in grade XII (23.8%). The unequal distribution across grades reflects differences in how many students in each grade responded rather than a planned quota, and it is treated as a limitation and examined in the sensitivity analyses rather than corrected by weighting.

Data were collected using an akhlakul karimah questionnaire developed by the researchers. The questionnaire consisted of 23 statements distributed across the four domains (Table 1) and was answered on a five-point Likert scale: Strongly Disagree (1), Disagree (2), Undecided/Neutral (3), Agree (4), and Strongly Agree (5). Four negatively worded statements, namely items 2, 10, 13, and 23, were reverse-scored so that a higher score always indicates more favourable conduct; because reversed items are known to behave differently from positively keyed ones, they were examined separately in the item analysis (Suárez-Alvarez et al., 2018; Vigil-Colet et al., 2020). The instructions emphasised that there were no right or wrong answers, that responses would not affect report-card grades, and that respondents' identities would be kept confidential; these instructions were intended to reduce socially desirable responding (Bergen & Labonté, 2020; Larson, 2019).

Table 1. Blueprint of the akhlakul karimah questionnaire

Domain	**Indicator**	**Item numbers**	**Negative items**
A. Morals toward parents	Courteous speech, limits of obedience, apologising, helping with housework, not shouting, praying for parents	1–6	2
B. Morals toward orphans	Not taking their rights, fairness, empathy and helping, not using their property, safeguarding entrusted property, refusing invitations	7–12	10
C. Honesty and anti-corruption	Returning excess money, not taking school property, refusing to cheat, honest reporting of assignments, not manipulating data, reminding peers	13–18	13
D. Discipline and care in prayer	Praying on time, feeling sinful when neglecting prayer, praying in congregation, not delaying, reminding peers	19–23	23

Note. Negative items were reverse-scored (Strongly Agree = 1 to Strongly Disagree = 5).

The questionnaire was administered online through Google Forms without collecting respondents' names, and students were asked to complete it independently. Because administration was not individually supervised and students of the same class responded within the same period, the possibility that some students completed the questionnaire in one another's presence cannot be excluded. Rather than assume this away, the possibility was examined empirically in the data screening and quantified in the sensitivity analyses reported below.

Three data-quality criteria were specified before analysis, following recommendations on careless and invalid responding (Curran, 2016; Ward & Meade, 2023). A record was excluded only if it was incomplete on any of the 23 items, if it showed invariant responding by giving the same option on all 23 items, or if it was submitted outside the administration window. No record met any exclusion criterion, so all 126 records were retained for the main analysis. Two further patterns were recorded as indicators of possible non-independence rather than as grounds for exclusion. First, records reproducing another record's answers on all 23 items were flagged as exact duplicates, and the plausibility of such patterns arising by chance was assessed with a permutation baseline in which each item column was independently resampled 200 times. Second, the number of differing items was computed for every pair of respondents, and pairs differing on two items or fewer were grouped into clusters of near-identical patterns. Both features were carried forward into the sensitivity analyses rather than used to delete data.

Item quality was examined using Pearson item–total correlations, with a critical value of r = .175 (df = 124, α = .05, two-tailed). Reliability was estimated with Cronbach's alpha for the total scale and for each domain, using .70 as a reference threshold while recognising that alpha is a lower-bound estimate sensitive to scale length and content heterogeneity (Taber, 2018). Descriptive statistics comprised frequency distributions, means (M), standard deviations (SD), medians (Mdn), and the percentage of the maximum score, that is, the mean divided by the maximum score of 5. Means were categorised using an interval of 0.80: 1.00–1.80 very low, 1.81–2.60 low, 2.61–3.40 moderate, 3.41–4.20 high, and 4.21–5.00 very high.

Normality was tested with the Shapiro–Wilk test. Because neither the domain scores nor the total score were normally distributed, differences between domains were tested with the Friedman test followed by Wilcoxon signed-rank tests, and differences between grade levels with the Kruskal–Wallis test followed by Mann–Whitney tests; both sets of post hoc tests used the Holm correction. Effect sizes were reported as Kendall's W, the rank-biserial correlation (rrb), and epsilon squared (ε^2), with Welch's ANOVA and η^2 as a parametric comparison (Field, 2018; Hair et al., 2019). Three sensitivity analyses were specified in advance. The first retained only the first occurrence of each exact response pattern. The second retained one representative record from each cluster of near-identical patterns. The third re-estimated the main comparisons on the reduced set of items that all satisfied the item–total criterion. Analyses were carried out in Microsoft Excel and in Python using NumPy and SciPy (Harris et al., 2020; Virtanen et al., 2020).

RESULTS & DISCUSSION

This section reports, in turn, the quality of the data and the profile of the respondents, the quality of the instrument, the overall level of akhlakul karimah, the indicator-level profile, differences between domains, differences between grade levels, and three sensitivity analyses. All scores are on a 1–5 scale, and the four negatively worded items (2, 10, 13, and 23) were reverse-scored so that higher scores always indicate more favourable self-reported conduct.

Data quality and respondent profile

All 126 returned questionnaires were complete, none showed invariant responding across the 23 items, and none was submitted outside the administration window; the longest run of identical consecutive responses was 16 of 23 items, observed in four records. No record therefore met an exclusion criterion, and all 126 were analysed. The analysed sample comprised 26 students in grade X (20.6%), 70 in grade XI (55.6%), and 30 in grade XII (23.8%). Two patterns of possible non-independence were nevertheless recorded. First, 36 records reproduced, item for item, the complete 23-item answer pattern of another record, leaving 90 distinct patterns. To judge how unlikely such agreement is, the 126 records were resampled 200 times by independently permuting each item column, a procedure that preserves every item's response distribution while removing any dependence between respondents; no exact duplicate pair arose in any of the 200 resamples (M = 0.00, maximum = 0). Second, grouping records that differed on no more than two of the 23 items produced 40 clusters: 23 records stood alone, the remaining 103 fell into 17 clusters of which the largest contained 18 records, and every cluster lay entirely within a single grade. These records were retained in the main analysis, and their influence is quantified in the sensitivity analyses below.

Instrument validity and reliability

Item quality was examined with Pearson item–total correlations. With N = 126, the critical value at α = .05 (two-tailed, df = 124) is r = .175. Twenty of the 23 items exceeded this value, with

coefficients ranging from .234 (item 3) to .728 (item 8) (Table 4). Three items failed the criterion: item 2 (r = −.179), item 5 (r = .034), and item 19 (r = −.026). Item 2 is a negatively worded item whose correlation is not merely low but negative, which is consistent with evidence that reversed items are liable to be misread (Suárez-Alvarez et al., 2018; Vigil-Colet et al., 2020); items 5 and 19 attracted responses concentrated in a single category and therefore varied too little to correlate with the total.

The internal consistency of the 23-item scale was α = .655, below the conventional threshold of .70. Domain-level coefficients were lower still (Table 2), particularly for morals toward parents (α = .158) and for prayer (α = .509). When the three failing items were removed and the item–total correlations recalculated, item 12 also fell below the critical value; removing it produced a 19-item set in which every item met the criterion (r = .194 to .796), with α = .767. The analyses that follow are reported for the full 23-item instrument, which preserves the content coverage of all four domains, while the 19-item set serves as a sensitivity check on every conclusion drawn from composite scores.

Table 2. Internal consistency of the instrument (N = 126)

Scale	Number of items	Cronbach's α
A. Morals toward parents	6	.158
B. Morals toward orphans	6	.533
C. Honesty and anti-corruption	6	.715
D. Discipline and care in prayer	5	.509
Total scale	23	.655
Valid item set (items 1, 3, 4, 6–11, 13–18, 20–23)	19	.767

Note. The reduced set excludes items 2, 5, 12, and 19, which fell below the critical item–total correlation (*r* = .175) in successive estimation rounds.

Overall level of akhlakul karimah

Students reported a high level of akhlakul karimah (M = 4.14, SD = 0.29), equivalent to 82.9% of the maximum score (Table 3). All four domains fell in the high category, with means between 4.11 and 4.18: prayer (M = 4.18, SD = 0.41), orphans (M = 4.16, SD = 0.53), parents (M = 4.13, SD = 0.32), and honesty and anti-corruption (M = 4.11, SD = 0.59). The honesty domain showed the widest dispersion and the parents domain the narrowest, meaning that students' responses were least uniform on honesty. Shapiro–Wilk tests indicated that neither the total score nor any domain score was normally distributed (all p < .001), so nonparametric tests were used for all inferential comparisons. Throughout this section the wording "students reported" is used deliberately: the data are students' reports of their own conduct, not observations of that conduct.

Table 3. Descriptive statistics for the four domains (N = 126)

Domain	Items	M	SD	Mdn	% of maximum score	Category	Mean rank
A. Morals toward parents	1–6	4.13	0.32	4.17	82.5	High	2.33
B. Morals toward orphans	7–12	4.16	0.53	4.08	83.2	High	2.37
C. Honesty and anti-corruption	13–18	4.11	0.59	4.17	82.2	High	2.69

Domain	Items	M	SD	Mdn	% of maximum score	Category	Mean rank
D. Discipline and care in prayer	19–23	4.18	0.41	4.20	83.5	High	2.62
Overall	1–23	4.14	0.29	4.09	82.9	High	–

Note. Categories: 1.00–1.80 very low; 1.81–2.60 low; 2.61–3.40 moderate; 3.41–4.20 high; 4.21–5.00 very high. Mean ranks are from the Friedman test.

Indicator-level profile

Figure 1 and Table 4 show that the high overall level conceals considerable variation among indicators. Twelve indicators fell in the very high category. The strongest were praying for one's parents after every prayer (item 6; *M* = 4.67), safeguarding entrusted orphan property (item 11; *M* = 4.55), treating orphan and non-orphan peers fairly (item 8; *M* = 4.47), not using orphans' property for personal gain (item 10; *M* = 4.43), and reminding peers who are about to cheat (item 18; *M* = 4.35). Items 1, 6, and 11 were endorsed by every respondent.

At the other end of the scale the weakest indicators were not distributed at random. The lowest was moral judgement when a parental order contradicts religious teaching (item 2; *M* = 3.48, *SD* = 1.31), on which 29.4% of students gave answers indicating that they would still obey. Next came compassion and the urge to help orphans in need (item 9; *M* = 3.59, *SD* = 1.45), where responses were sharply divided, 34.9% disagreeing and 61.9% agreeing, and never taking the rights of orphans (item 7; *M* = 3.68, *SD* = 1.28), where 19.8% of responses fell at the two lowest scores. Refusing to cheat when an opportunity arises (item 15; *M* = 3.90) attracted the largest share of neutral responses in the honesty domain (15.1%), and apologising to parents after a mistake (item 3; *M* = 4.14) the largest share in the parents domain (14.3%).

Table 4. Item statistics and item–total correlations (N = 126)

No	Indicator	Score 1–2 (%)	Score 3 (%)	Score 4–5 (%)	M	SD	r(it)	Category
1	Speaking gently and politely to parents	0.0	0.0	100.0	4.16	0.37	.408	High
2	Not obeying parental orders that contradict religion (R)	29.4	6.3	64.3	3.48	1.31	−.179*	High
3	Apologising to parents when proven wrong	0.0	14.3	85.7	4.14	0.64	.234	High
4	Helping with housework without being asked	0.0	11.9	88.1	4.21	0.64	.471	Very high
5	Not shouting or raising one's voice at parents	0.0	14.3	85.7	4.10	0.62	.034*	High
6	Praying for both parents after every prayer	0.0	0.0	100.0	4.67	0.47	.523	Very high
7	Never taking the rights of orphans	19.8	0.0	80.2	3.68	1.28	.563	High
8	Treating orphan and non-orphan peers fairly	0.0	6.3	93.7	4.47	0.62	.728	Very high

No	Indicator	Score 1–2 (%)	Score 3 (%)	Score 4–5 (%)	M	SD	r(it)	Category
9	Feeling compassion and moved to help orphans in need	34.9	3.2	61.9	3.59	1.45	.497	High
10	Not using orphans' property for personal gain (R)	3.2	0.0	96.8	4.43	0.66	.505	Very high
11	Safeguarding entrusted orphan property	0.0	0.0	100.0	4.55	0.50	.510	Very high
12	Refusing peers' invitations to take orphans' property	11.1	0.0	88.9	4.25	0.93	.268	Very high
13	Returning excess shopping money (R)	11.1	0.0	88.9	3.94	0.99	.412	High
14	Not taking school property without permission	7.1	0.0	92.9	4.29	1.03	.591	Very high
15	Refusing to cheat even when given the chance	6.3	15.1	78.6	3.90	0.98	.461	High
16	Reporting assignment results honestly	0.0	3.2	96.8	4.32	0.53	.396	Very high
17	Not manipulating data for personal gain	12.7	0.0	87.3	3.88	1.18	.401	High
18	Reminding peers who are about to cheat	0.8	0.0	99.2	4.35	0.57	.236	Very high
19	Performing obligatory prayers on time	0.0	7.9	92.1	4.25	0.59	−.026*	Very high
20	Feeling sinful when missing prayer without a valid reason	6.3	0.0	93.7	4.00	0.88	.526	High
21	Striving to pray in congregation	0.8	4.8	94.4	4.25	0.58	.317	Very high
22	Not delaying prayer until its time has passed	0.0	9.5	90.5	4.25	0.62	.261	Very high
23	Reminding peers who neglect prayer (R)	6.3	5.6	88.1	4.13	0.81	.249	High

Note. (R) = reverse-scored item; scores are shown after reversal. *r*(it) = Pearson item–total correlation; * = below the critical value ($r = .175$, $df = 124$, $\alpha = .05$).

Within the parents domain, the indicators of supplication and courteous speech (items 1 and 6) were endorsed almost uniformly, whereas the two indicators that require moral judgement or the admission of wrongdoing (items 2 and 3) were the weakest. Within the orphans domain, the four indicators concerning restraint from, or protection of, orphans' property (items 8, 10, 11, and 12) all fell in the very high category ($M = 4.25$ to 4.55), in sharp contrast with the two indicators that require active concern or the forgoing of a benefit (items 9 and 7).

The honesty and anti-corruption domain was comparatively even ($M = 3.88$ to 4.35). Its lowest indicators were refusing to cheat when an opportunity arises (item 15; $M = 3.90$) and not manipulating data for personal gain (item 17; $M = 3.88$), while reminding peers who appear about to cheat (item

18; *M* = 4.35) fell in the very high category, so the weaker indicators cannot simply be reduced to those that involve other people. The prayer domain was the most homogeneous (*M* = 4.00 to 4.25). Responses clustered at the "Agree" level: 74.6% of students chose it for item 20 on feeling sinful when neglecting prayer, and only 19.0% chose the highest score. This concentration explains both the low variability of the domain and the failure of item 19 to meet the item–total criterion.

Across domains a consistent pattern emerges. Indicators describing restraint from prohibited acts and the performance of ritual worship were endorsed most strongly, whereas those requiring empathic action (item 9) or moral judgement when parental authority conflicts with religious norms (item 2) were the weakest. A second pattern concerns item wording. The three positively keyed items phrased with a negation ("I never…", "I would not…"; items 7, 14, and 17) each attracted a cluster of responses at the lowest score (14.3%, 7.1%, and 12.7%), which may reflect misreading of the negated sentence rather than the conduct itself.

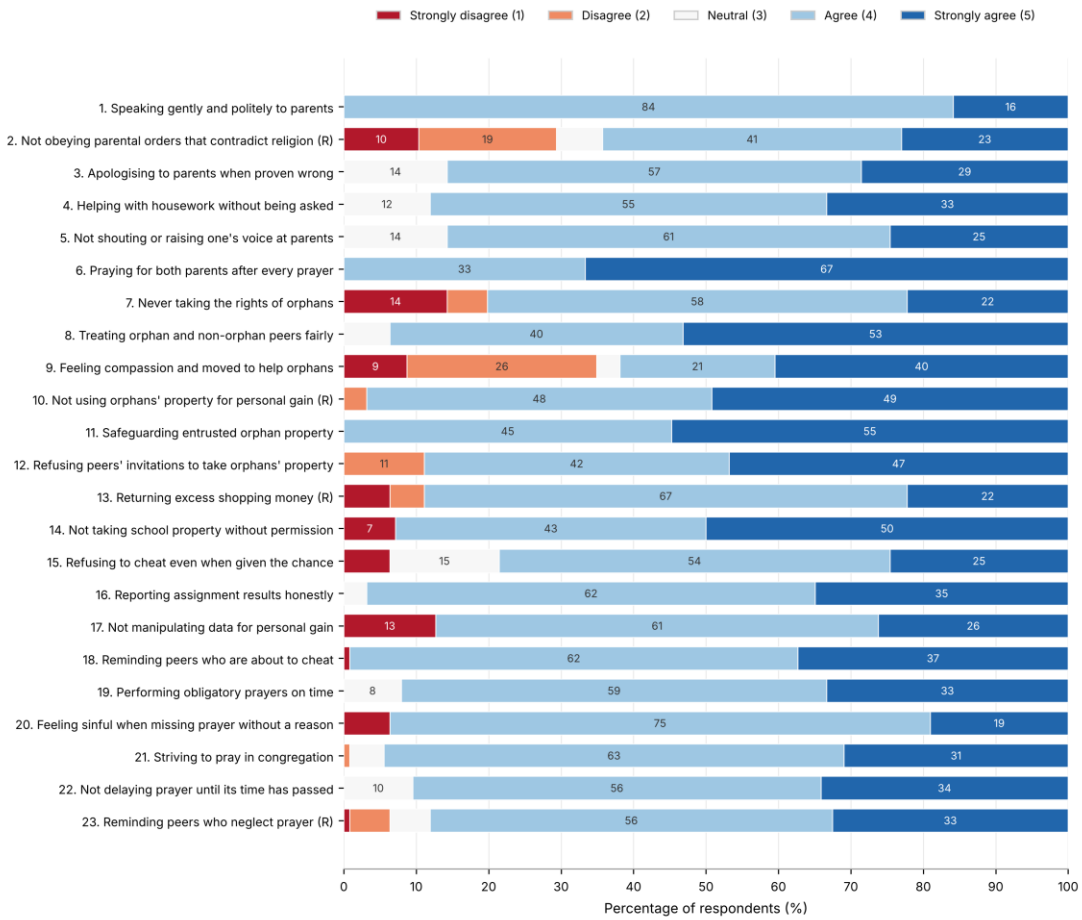


Figure 1. Distribution of response scores across the 23 indicators (N = 126); (R) = reverse-scored item

Differences between domains

The Friedman test did not show a statistically significant difference among the four domains, $\chi^2(3) = 7.71$, $p = .052$, and the effect was negligible (Kendall's W = .02). Wilcoxon signed-rank tests with Holm correction confirmed that no pair of domains differed (all corrected $p \geq .686$; the largest effect was rrb = .17 for parents versus honesty). The differences between domain means were 0.02 to 0.07 points on the five-point scale. At the level of whole domains, therefore, students' reported akhlakul karimah was essentially flat. This null result is itself informative, because it means that the unevenness documented in the previous section disappears once scores are aggregated into domains.

Differences between grade levels

Total scores did differ across grade levels, $H(2) = 22.76$, $p < .001$, with a large effect ($\varepsilon^2 = .18$). Welch's ANOVA led to the same conclusion, $F(2, 55.86) = 17.75$, $p < .001$, $\eta^2 = .18$. Grade X students reported the highest level ($M = 4.38$, $SD = 0.22$), while grades XI ($M = 4.07$, $SD = 0.26$) and XII ($M = 4.12$, $SD = 0.32$) were similar (Table 5; Figure 2). Mann–Whitney tests with Holm correction confirmed that grade X scored higher than grade XI ($p < .001$, rrb = .65) and grade XII ($p = .006$, rrb = .46), whereas grades XI and XII did not differ ($p = .492$, rrb = −.09).

Grade differences appeared in all four domains but ran in different directions. The largest was in the orphans domain, $H(2) = 45.54$, $p < .001$, $\varepsilon^2 = .36$, with grade X ($M = 4.74$) well above grades XII ($M = 4.19$) and XI ($M = 3.93$). In the honesty domain the scores declined steadily across grades, $H(2) = 30.61$, $p < .001$, $\varepsilon^2 = .25$, from grade X ($M = 4.50$) through grade XI ($M = 4.18$) to grade XII ($M = 3.62$). The pattern was reversed in the prayer domain, where grade XII was highest, $H(2) = 13.73$, $p = .001$, $\varepsilon^2 = .11$, and in the parents domain, where grade XII was also highest, $H(2) = 15.68$, $p < .001$, $\varepsilon^2 = .13$. At indicator level the grade gap was most visible for empathy toward orphans (item 9): 92.3% of grade X students agreed or strongly agreed, compared with 80.0% in grade XII and only 42.9% in grade XI, where 51.4% disagreed.

Table 5. Domain scores by grade level and Kruskal–Wallis tests (N = 126)

Domain	Grade X (n = 26) M (SD)	Grade XI (n = 70) M (SD)	Grade XII (n = 30) M (SD)	H(2)	p	ε^2
A. Morals toward parents	4.18 (0.20)	4.04 (0.32)	4.29 (0.36)	15.68	< .001	.13
B. Morals toward orphans	4.74 (0.31)	3.93 (0.49)	4.19 (0.37)	45.54	< .001	.36
C. Honesty and anti-corruption	4.50 (0.41)	4.18 (0.21)	3.62 (0.92)	30.61	< .001	.25
D. Discipline and care in prayer	4.02 (0.37)	4.13 (0.39)	4.43 (0.40)	13.73	.001	.11
Overall	4.38 (0.22)	4.07 (0.26)	4.12 (0.32)	22.76	< .001	.18

Note. ε^2 = epsilon-squared effect size for the Kruskal–Wallis test.

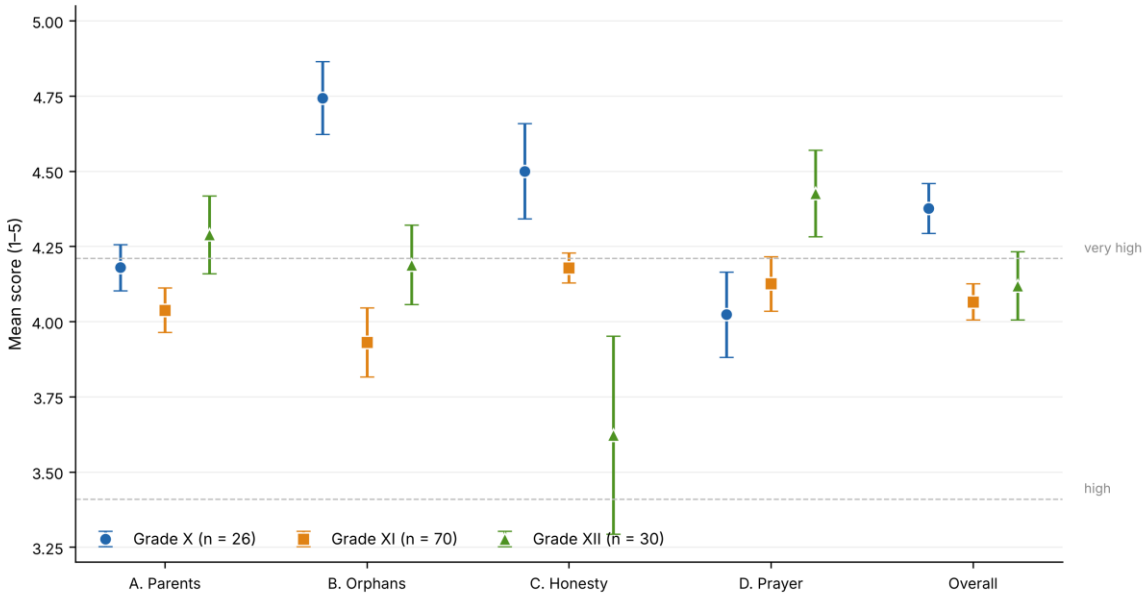


Figure 2. Mean domain and overall scores by grade level with 95% confidence intervals (N = 126)

Sensitivity analyses

The first sensitivity analysis retained only the first occurrence of each exact 23-item response pattern, reducing the sample to 90 records (20 in grade X, 58 in grade XI, 12 in grade XII). The overall mean was essentially unchanged (M = 4.10, SD = 0.26), and the grade-level difference remained significant and became larger, $H(2) = 27.51$, $p < .001$, $\varepsilon^2 = .31$. The comparison between domains, however, changed: the Friedman test became significant, $\chi^2(3) = 17.58$, $p < .001$, $W = .07$, and the honesty domain scored higher than the parents ($p = .015$, rrb = .38), orphans ($p = .020$, rrb = .35), and prayer ($p = .005$, rrb = .42) domains. The ranking of domains is therefore sensitive to which records are retained and should not be interpreted substantively; the indicator-level pattern and the grade-level difference are not sensitive in this way.

The second sensitivity analysis retained one representative record from each cluster of near-identical patterns (n = 40; 8 in grade X, 26 in grade XI, 6 in grade XII). The overall mean was again stable (M = 4.10, SD = 0.27) and the grade-level difference remained significant, $H(2) = 9.12$, $p = .010$, $\varepsilon^2 = .23$, although in this much smaller sample only the grade X–XI contrast survived the Holm correction ($p = .008$, rrb = .72). Because every cluster lay within a single grade, this analysis is deliberately conservative: it removes genuine respondents along with any redundant ones. Its main implication is that the number of mutually independent response patterns in these data is appreciably smaller than 126, a point taken up among the limitations.

The third sensitivity analysis re-estimated the main comparisons on the 19-item set in which every item met the item–total criterion (M = 4.17, SD = 0.37). The grade-level difference remained significant, $H(2) = 16.63$, $p < .001$, $\varepsilon^2 = .13$, and the comparison between domains remained non-significant, $\chi^2(3) = 6.24$, $p = .101$, $W = .02$. Across the three analyses the two robust findings are the unevenness at indicator level and the higher scores of grade X; the ranking of the four domains is not robust.

DISCUSSION

This study pursued three objectives, and the discussion is organised around them. The first was to establish the level of akhlakul karimah reported by MA students; the second, to identify the strongest and the weakest indicators; the third, to test for differences between domains and between grade levels. The three subsections below take these objectives in turn. The implications that follow from them, together with the limitations of the study and suggestions for further research, are set out after the conclusion.

The level of akhlakul karimah and the caution required in reading it

Students reported a level equivalent to 82.9% of the maximum. This is consistent with the general picture that the madrasah environment supports moral formation (Lubis, 2022), but it has to be read with care. The honesty and anti-corruption domain obtained a high mean even though national data show that cheating remains common in the great majority of Indonesian schools (Komisi Pemberantasan Korupsi, 2025) and that secondary school students tend not to regard many forms of academic dishonesty as violations (Fadila, 2022). The contrast suggests that high scores reflect students' normative knowledge of how they ought to behave more closely than the actual frequency of the behaviour. Presenting oneself in line with norms is a known risk of self-report, and it is strongest for conduct carrying moral and religious value (Bergen & Labonté, 2020; Bispo Júnior, 2022; Larson, 2019). For this reason the findings are stated throughout as what students reported, not as what they were observed to do.

The response pattern in the prayer domain supports this reading. Three-quarters of the students marked "Agree" when asked whether they felt sinful after missing a prayer, and fewer than one in five gave the highest rating; the domain's items varied so little that one of them failed the item–total criterion. Uniform responses of this kind are what one would expect where a practice is a daily institutional routine: students know the expected answer without expressing a stronger personal commitment. Muttaqin et al. (2023) found that Akidah Akhlak learning was unrelated to MA students' akhlakul karimah, a reminder that high moral scores do not by themselves indicate that learning has succeeded. The most informative result of this study is therefore not the high overall mean but the relative differences between indicators, which are more robust to biases that shift all items in the same direction.

The strongest and weakest indicators: restraint, empathy, judgement, and temptation

Restraint and empathy. The sharpest contrast in the data lies within the orphans domain. Four indicators concerning restraint from, or protection of, orphans' property fell in the very high category, whereas compassion and the urge to help were among the three weakest indicators in the whole instrument, with responses almost evenly divided. The same shape recurs elsewhere: clear prohibitions, such as not taking school property or not using orphans' property, were widely endorsed, while actions requiring emotional involvement and personal initiative were endorsed considerably less. Prohibitions are easy to understand and carry sanctions; empathic action requires sensitivity to another person's need and a real opportunity to act.

Developmental research helps explain the asymmetry. In a systematic review, Silke et al. (2018) found that adolescents' empathy and prosocial behaviour are shaped by a wide range of social and psychological factors and do not follow from knowing the norm, and Carlo and Padilla-Walker (2020) argue that prosocial behaviour is multidimensional, so that helping in one context does not predict helping in another. Religiosity is positively related to students' prosocial behaviour (Safrilsyah et al., 2021), but religion reaches adolescents through socialisation, relationships, and experience rather than automatically (Hardy et al., 2019). Islamic teaching itself makes the demand for active care explicit: QS. Al-Mā'ūn [107]: 1–3 describes the one who repulses the orphan and does not urge the feeding of the poor as denying religion, and QS. Aḍ-Ḍuḥā [93]: 9 forbids treating the orphan harshly. Islamic ethics therefore does not consist merely in the absence of harmful acts; it requires active concern for others, and it is on precisely this second requirement that the students in this study reported least. The pattern locates the gap between moral knowledge and moral action (Darnell et al., 2019) in a specific region of moral life rather than treating it as a general deficit.

Moral judgement in value conflicts. The weakest indicator of all concerned situations in which a parental order contradicts religious teaching: close to three students in ten indicated that they would still obey. Two readings are plausible and are not mutually exclusive. Substantively, students appear to have internalised the duty of devotion to parents strongly but not its limit, which QS. Luqmān [31]: 15 states explicitly, namely that obedience to parents does not extend to what contradicts the teaching of God. Weighing two obligations that appear to conflict is a matter of practical wisdom rather than of rule recall, and practical wisdom develops through practice in judging concrete cases (Darnell et al., 2019; Kristjánsson et al., 2021). Methodologically, this is a negatively worded item whose item–total correlation was not merely low but negative, so some students may have misread the direction of the statement (Suárez-Alvarez et al., 2018; Vigil-Colet et al., 2020). Both readings point in the same practical direction: students need explicit, guided practice in judging value conflicts.

Honesty under temptation. Within the honesty domain, the lowest indicators were refusing to cheat when an opportunity arises and not manipulating data for personal gain, whereas reminding a peer who is about to cheat was among the strongest in the whole instrument. Students thus found it

easier to uphold the honesty norm for others than for themselves under temptation. This is consistent with meta-analytic evidence that academic dishonesty is driven less by ignorance of the norm than by motivational pressure to perform (Krou et al., 2021) and by the belief that peers are cheating (Zhao et al., 2022), and with Indonesian evidence linking academic dishonesty to how learners reason about moral problems (Ampuni et al., 2020) and to pressure for high grades, inadequate preparation, and permissive peers (Rosari et al., 2026). Anti-corruption education in madrasahs therefore cannot be delivered as subject content alone; it has to be carried by an honest assessment culture and by the institution's hidden curriculum (Abdurrahman et al., 2021; Firman et al., 2021; Suyadi et al., 2021).

Differences between domains and between grade levels

At the level of whole domains, the four domains of akhlakul karimah did not differ, and the sensitivity analyses showed that whatever ordering does appear depends on which records and which items are retained. This null result deserves emphasis rather than apology, because it means that a domain-level score, the form in which moral conduct is usually reported in this literature, carries almost no information about where students are strong or weak; the information lies one level down, at the indicator. Grade X students reported the highest scores, most clearly in the orphans and honesty domains, while grades XI and XII were similar overall although they differed by domain. This should be interpreted cautiously. It cannot be read in simple developmental terms, since longitudinal evidence shows that prosocial behaviour in adolescence does not follow a uniform downward trajectory and differs by gender (Memmott-Elison et al., 2020; Van der Graaff et al., 2018). A more likely contributor is the context in which the questionnaire was completed: near-identical response patterns occurred within the same class and never across classes, so part of the grade difference may reflect the atmosphere or manner of completion in each class rather than differences in conduct. The difference survived both record-level sensitivity analyses, but in the smallest sample only the grade X–XI contrast remained significant after correction, and the number of grade XII respondents was modest. The grade difference is therefore better treated as a hypothesis for future testing than as evidence that moral conduct declines over the years of schooling. For teachers it remains useful as an early signal that empathy and honesty deserve particular attention in grades XI and XII.

CONCLUSION

This study aimed to identify the strengths and weaknesses in MA students' reported moral conduct rather than merely determine their overall level of akhlakul karimah. The findings revealed a high overall level of moral conduct, with no statistically significant differences among the four domains. However, indicator-level analysis exposed substantial variations: students demonstrated stronger tendencies to refrain from prohibited acts and perform ritual worship, but weaker engagement in empathic assistance toward orphans and in evaluating parental obedience when parental instructions conflict with religious teachings. Grade X students reported higher scores than those in grades XI and XII, although this difference was partly associated with the classroom context during questionnaire administration. These findings suggest that moral education in madrasahs should not be evaluated solely through compliance with religious prohibitions and ritual obligations, but also through the development of empathy, social responsibility, and practical wisdom in navigating competing moral obligations.

Theoretically, this study extends discussions of the relationship between moral knowledge and moral action by demonstrating that the gap is not distributed equally across moral domains. Rather, it becomes more apparent when moral expectations shift from avoiding prohibited behaviour to engaging in active care, and from following established rules to making context-sensitive moral judgments. The findings also highlight the limitations of composite moral scores, which may conceal

meaningful differences between specific behavioural indicators. Accordingly, research on akhlak and character education should incorporate indicator-level assessments alongside overall scores. Practically, Akidah Akhlak learning should move beyond a predominantly prohibition-oriented approach by incorporating experiential and service learning to strengthen empathy, particularly toward vulnerable groups. Case-based learning and moral dilemma discussions can also help students evaluate competing obligations, including the boundaries of obedience to parents. Furthermore, role modelling and habituation should extend beyond ritual practices to concrete social engagement, while moral assessment should combine indicator-level mapping with teacher observations, behavioural records, and peer assessments. Reflection on worship practices is also necessary to connect religious routines with honesty, compassion, and social responsibility.

Several limitations should be considered when interpreting these findings. First, reliance on self-reported data without behavioural observation leaves the possibility of socially desirable responses. Second, the instrument demonstrated relatively modest reliability ($\alpha = .655$), with three items failing the item–total criterion and irregular responses to negatively worded statements. Third, although all responses met the predetermined data-quality criteria, repeated answer patterns and near-identical response clusters within individual classes raise concerns about response independence. Finally, the cross-sectional design, single-madrasah setting, and unequal grade-level sample sizes limit generalisability and prevent causal interpretations of grade differences. Future studies should therefore improve instrument construction through pilot testing, the avoidance of negatively worded items, and confirmatory factor analysis. Supervised and time-stamped questionnaire administration may also strengthen response verification. Multi-madrasah and mixed-methods designs are recommended to examine whether the identified restraint–empathy gap extends across different educational contexts and to compare reported conduct with observed behaviour. Longitudinal and experimental research would further clarify developmental differences across grade levels and assess the effectiveness of empathy-oriented and dilemma-based approaches in Akidah Akhlak learning.

ACKNOWLEDGMENT

We thank Universitas Nahdlatul Ulama Sunan Giri for funding this research. We also thank the student respondents at the madrasah in Bojonegoro Regency.

REFERENCES

Abdurrahman, M., A., & Usamah. (2021). The internalization of anti-corruption values as hidden curriculum in Gontor educational system. *Nadwa: Jurnal Pendidikan Islam*, *15*(2), 191–216.

Ampuni, S., Kautsari, N., Maharani, M., Kuswardani, S., & Buwono, S. B. S. (2020). Academic dishonesty in Indonesian college students: An investigation from a moral psychology perspective. *Journal of Academic Ethics*, *18*(4), 395–417. https://doi.org/10.1007/s10805-019-09352-2

Bergen, N., & Labonté, R. (2020). "Everything is perfect, and we have no problems": Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research*, *30*(5), 783–792. https://doi.org/10.1177/1049732319889354

Bispo Júnior, J. P. (2022). Social desirability bias in qualitative health research. *Revista de Saúde Pública*, *56*, 101. https://doi.org/10.11606/s1518-8787.2022056004164

Carlo, G., & Padilla-Walker, L. M. (2020). Adolescents' prosocial behaviors through a multidimensional and multicultural lens. *Child Development Perspectives*, *14*(4), 265–272. https://doi.org/10.1111/cdep.12391

Chan, H. H.-K., Kwong, H. Y. C., Shu, G. L. F., Ting, C. Y., & Lai, F. H.-Y. (2021). Effects of experiential learning programmes on adolescent prosocial behaviour, empathy, and subjective

well-being: A systematic review and meta-analysis. *Frontiers in Psychology*, *12*, 709699. https://doi.org/10.3389/fpsyg.2021.709699

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006

Darnell, C., Gulliford, L., Kristjánsson, K., & Paris, P. (2019). Phronesis and the knowledge-action gap in moral psychology and moral education: A new synthesis? *Human Development*, *62*(3), 101–129. https://doi.org/10.1159/000496136

Fadila, S. A. P. (2022). Students' perception on academic dishonesty in a senior high school. *Journal of English and Education*, *8*(1), 30–36. https://doi.org/10.20885/jee.v8i1.23293

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE Publications.

Firman, F., Kaso, N., Arifuddin, A., Mirnawati, M., Ilham, D., & Karim, A. R. (2021). Anti-corruption education model in Islamic universities. *Al-Ishlah: Jurnal Pendidikan*, *13*(3), 2146–2158. https://doi.org/10.35445/alishlah.v13i3.843

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Han, H., Choi, Y.-J., Dawson, K. J., & Jeong, C. (2018). Moral growth mindset is associated with change in voluntary service engagement. *PLoS ONE*, *13*(8), e0202327. https://doi.org/10.1371/journal.pone.0202327

Hardy, S. A., Nelson, J. M., Moore, J. P., & King, P. E. (2019). Processes of religious and spiritual influence in adolescence: A systematic review of 30 years of research. *Journal of Research on Adolescence*, *29*(2), 254–275. https://doi.org/10.1111/jora.12486

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Jeynes, W. H. (2019). A meta-analysis on the relationship between character education and student achievement and behavioral outcomes. *Education and Urban Society*, *51*(1), 33–71. https://doi.org/10.1177/0013124517747681

Komisi Pemberantasan Korupsi. (2025). *Temuan hasil SPI Pendidikan 2024: Menyontek dan plagiarisme masih merebak di sekolah dan kampus*. https://www.kpk.go.id/id/ruang-informasi/berita/temuan-hasil-spi-pendidikan-2024-menyontek-dan-plagiarisme-masih-merebak-di-sekolah-dan-kampus

Kristjánsson, K., Fowers, B., Darnell, C., & Pollard, D. (2021). Phronesis (practical wisdom) as a type of contextual integrative thinking. *Review of General Psychology*, *25*(3), 239–257. https://doi.org/10.1177/10892680211023063

Krou, M. R., Fong, C. J., & Hoff, M. A. (2021). Achievement motivation and academic dishonesty: A meta-analytic investigation. *Educational Psychology Review*, *33*(2), 427–458. https://doi.org/10.1007/s10648-020-09557-7

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research*, *61*(5), 534–547. https://doi.org/10.1177/1470785318805305

Lubis, N. S. (2022). Pembentukan akhlak siswa di madrasah: Kontribusi lingkungan sekolah, kompetensi guru, dan mutu pendidikan. *Jurnal Pendidikan Agama Islam Al-Thariqah*, *7*(1), 137–156. https://doi.org/10.25299/al-thariqah.2022.vol7(1).8847

Memmott-Elison, M. K., Holmgren, H. G., Padilla-Walker, L. M., & Hawkins, A. J. (2020). Associations between prosocial behavior, externalizing behaviors, and internalizing symptoms during adolescence: A meta-analysis. *Journal of Adolescence*, *80*, 98–114. https://doi.org/10.1016/j.adolescence.2020.01.012

Muttaqin, M. R., Fauzan, U., & Malihah, N. (2023). Pengaruh pembelajaran akidah akhlak terhadap akhlakul karimah siswa Madrasah Aliyah Miftahul Ulum Anggana. *Tawazun: Jurnal Pendidikan Islam*, *16*(2), 363–376. https://doi.org/10.32832/tawazun.v16i2.14233

Nasir, M., & Rijal, M. K. (2021). Keeping the middle path: Mainstreaming religious moderation through Islamic higher education institutions in Indonesia. *Indonesian Journal of Islam and Muslim Societies*, *11*(2), 213–241. https://doi.org/10.18326/ijims.v11i2.213-241

Rosari, R., Chariri, A., & Utomo, D. C. (2026). Academic dishonesty among university students: Gender, semester differences, and influencing factors. *Emerging Science Journal*, *10*(2), 956–973. https://doi.org/10.28991/ESJ-2026-010-02-019

Safrilsyah, S., Yusoff, M. Z. M., Othman, M. K., & Ibrahim, I. (2021). Religiosity and prosocial behavior on student of Islamic and public schools at Banda Aceh, Indonesia. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah*, *6*(1), 1–8. https://doi.org/10.24042/tadris.v6i1.6295

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

Salam, M., Awang Iskandar, D. N., Ibrahim, D. H. A., & Farooq, M. S. (2019). Service learning in higher education: A systematic literature review. *Asia Pacific Education Review*, *20*(4), 573–593. https://doi.org/10.1007/s12564-019-09580-6

Silke, C., Brady, B., Boylan, C., & Dolan, P. (2018). Factors influencing the development of empathy and pro-social behaviour among adolescents: A systematic review. *Children and Youth Services Review*, *94*, 421–436. https://doi.org/10.1016/j.childyouth.2018.07.027

Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

Suyadi, N., Z., & Asmorojati, A. W. (2021). The insertion of anti-corruption education into Islamic education learning based on neuroscience. *International Journal of Evaluation and Research in Education*, *10*(4), 1417–1425.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tambak, S., & Sukenti, D. (2024). Case-based learning method in learning: Is it effective to improve teaching skills of madrasa teachers in Indonesia? *Journal of Learning for Development*, *11*(1), 151–164. https://doi.org/10.56059/jl4d.v11i1.763

Tambak, S., Yasin, M. D. H. bin F., Lubis, F. S., & Tambak, F. A. M. (2026). Negotiating authority and innovation: Project-based learning and the changing professional landscape of madrasah teachers. *Jurnal Pendidikan Agama Islam Al-Thariqah*, *11*(1), 21–39. https://doi.org/10.25299/althariqah.2026.vol11(1).27544

Van der Graaff, J., Carlo, G., Crocetti, E., Koot, H. M., & Branje, S. (2018). Prosocial behavior in adolescence: Gender differences in development and links with empathy. *Journal of Youth and Adolescence, 47*(5), 1086–1099. https://doi.org/10.1007/s10964-017-0786-1

Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*, *32*(1), 108–114. https://doi.org/10.7334/psicothema2019.286

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology*, *74*, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Zhao, L., Mao, H., Compton, B. J., Peng, J., Fu, G., Fang, F., Heyman, G. D., & Lee, K. (2022). Academic dishonesty and its relations to peer cheating and culture: A meta-analysis of the perceived peer cheating effect. *Educational Research Review*, *36*, 100455. https://doi.org/10.1016/j.edurev.2022.100455

Zuhdi, M. (2018). Challenging moderate Muslims: Indonesia's Muslim schools in the midst of religious conservatism. *Religions, 9*(10), 310. https://doi.org/10.3390/rel9100310