



Classroom climate, educational resources, and school sector-setting configuration in pupils' Kohlberg-inspired moral justifications in Morocco: A multilevel analysis

Mustapha Fahli¹, Jamal Ksiksou², Lhoussaine Maskour³ and Abdel Fettah Zahidi⁴

¹Laboratory of Society, Communication, and Rhetoric of Discourse, Faculty of Letters and Human Sciences, Tetouan, Abdelmalek Essaâdi University, Morocco (ORCID: 0009-0004-5452-8369)

²Higher Institute of Nursing Professions and Health Techniques (ISPITS), Tetouan, Morocco (ORCID: 0000-0002-7608-4576)

³Sociology and Psychology Laboratory, Faculty of Letters and Human Sciences Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco (ORCID: 0000-0002-9966-3474)

⁴Ibn Zohr University, Laboratory of Science and Technology Research (LRST), ESEF, Agadir, Morocco (ORCID: 0009-0002-7697-2458)

Corresponding Author: Jamal Ksiksou, jamal.ksiksou@usmba.ac.ma

Submitted: 27 July 2026 Revised: 26 September 2026 Accepted: 1 October 2026 Published online: 12 October 2026

Abstract

This study examined associations between observed classroom climate, pupils' educational resources, school sector-setting configuration, and selections of Kohlberg-inspired moral justifications, in Meknès province, Morocco. The cross-sectional observational sample comprised 337 pupils nested in 24 classrooms across 14 primary schools. Classroom climate was assessed during three two-hour observation sessions per classroom using French-language CLASS K-3 materials for Grades 1–3 and Upper Elementary materials for Grades 4–6. Moral justification was measured through ten contextualized fixed-choice situations and analyzed as an exploratory composite. Primary multilevel models nested pupils within classrooms, and small-sample sensitivity analyses addressed dependence at both classroom and school levels. More favorable classroom climate and greater individual educational resources were associated with higher scores on the exploratory moral-justification index. The climate association remained statistically distinguishable from zero under classroom-clustered inference but was only marginal under the more conservative school-clustered sensitivity analysis. The climate-by-resources interaction was positive under classroom-clustered inference but was not statistically distinguishable from zero under school-clustered sensitivity analyses. A climate-by-grade-band/CLASS-version interaction showed the same pattern of attenuation under school clustering. Because CLASS version was determined by grade band, developmental and version-related moderation cannot be separated. School-configuration and separate sector-setting contrasts remained imprecisely estimated. The findings support further longitudinal, multi-observer research on classroom interaction as a socially differentiated educational opportunity structure. They do not identify causal mechanisms and should not be generalized beyond the sampled schools.

Keywords: Classroom climate; Educational resources; Moral judgment; Multilevel analysis; Primary education; Sector-setting configuration

1. Introduction

Schools are not merely sites of knowledge transmission. Through everyday patterns of participation, rule enforcement, disagreement, and justification, classrooms also organize pupils' experience of authority, reciprocity, and collective life. In the Durkheimian tradition, moral education brings together discipline, attachment to the group, and autonomy of the will (Durkheim, 1925/2012). The classroom can thus be understood as a setting in which norms are not only stated but enacted, negotiated, and interpreted through ordinary interaction.

Research on school climate has established consistent links with engagement, safety, learning, and a range of non-cognitive outcomes (Cohen et al., 2009; Thapa et al., 2013; Wang & Degol, 2016; Wang et al., 2020; Zynuddin et al., 2023). Less is known, however, about how the observable quality of classroom interactions relates to moral judgment—that is, to pupils' selection of reasons grounded not only in anticipated punishment or immediate interest, but also in reciprocity, shared rules, and interpersonal expectations.

Recent work confirms the value of a more differentiated approach to climate, one attentive to age, level of analysis, and mode of measurement. Aldridge and Blackstock (2024) validated an instrument specifically designed for primary-school pupils, while Aldridge and McLure (2024) showed that certain dimensions of climate are linked to the construction of moral identity, with

patterns that vary across schools. Together, these findings argue for distinguishing perceived school climate from observed classroom climate, while still recognizing their joint contribution to pupils' socialization.

The relationship between classroom interaction and moral judgment is consistent with the developmental perspectives of Piaget and Kohlberg. Piaget emphasized cooperation, decentration, and reciprocal relations among peers in the movement from heteronomous toward more autonomous forms of moral judgment, whereas Kohlberg highlighted role-taking, exposure to differing perspectives, and reasoned moral discussion as conditions that can support increasingly advanced forms of moral reasoning (Kohlberg, 1981; Piaget, 1932/1965; Power et al., 1989). Contemporary educational research similarly shows that structured discussion, argumentation, and engagement with multiple perspectives can strengthen students' argumentation skills. For example, Nugroho et al. (2025) found improved university students' argumentation skills in an inquiry-based intervention that included debate and the evaluation of opposing positions. The pedagogical environment, however, does not produce such development mechanically, nor do pupils enter classroom discussion on equal footing. Students differ in the linguistic, cultural, and material resources available to them outside school. Family socioeconomic status has been associated with differences in children's vocabulary development and reading development, while large-scale international evidence shows substantial socioeconomic gaps in student performance (Dolean et al., 2019; Hoff, 2003; Organisation for Economic Co-operation and Development [OECD], 2023a). These inequalities matter for classroom interaction because the ability to articulate a position, interpret written material, formulate an argument, and participate confidently in reasoned discussion depends partly on resources and competencies that are themselves socially patterned. Classroom dialogue may therefore create opportunities for moral reflection, but students' capacity to take advantage of those opportunities is shaped by unequal conditions that precede and extend beyond the classroom.

Evidence from contexts beyond the classroom also suggests that children's moral judgments are sensitive to the social environments in which they develop. Ball et al. (2017), for example, found that neighborhood risk and maternal disciplinary practices were jointly associated with preschoolers' judgments of moral transgressions, with some neighborhood effects varying according to the nature of maternal discipline. Socioeconomic context may also matter, although its influence appears to be context-dependent rather than uniform. Peretz-Lange et al. (2022) found that both objective and subjective socioeconomic status were associated with children's fairness judgments of novel resource distributions, but in different ways. A recent meta-analysis of 11 studies found no robust overall association between socioeconomic status and children's and adolescents' distributive-justice preferences, with high heterogeneity; its narrative synthesis suggested that these associations may depend on context (Wilke et al., 2026). These findings support the broader proposition that moral judgment is embedded in children's social ecology rather than determined solely by individual characteristics.

The classroom constitutes another proximal setting in which such judgments may be shaped. Classroom discourse is structured through teacher–class, teacher–student, and student–student interaction patterns, which influence the opportunities pupils have to explain, challenge, and respond to ideas (Çelik Demirci & Baki, 2023). It is therefore plausible that variation in classroom interaction may also be associated with the forms of justification pupils select in morally salient situations. At the same time, such classroom processes occur within broader socioeconomic and institutional contexts. Multilevel educational research demonstrates the importance of distinguishing between pupil-level and school-level conditions, including socioeconomic background and home educational resources (Coşkun & Karadağ, 2023). This distinction is especially relevant in Morocco, where an analysis of Service Delivery Indicators data from a nationally representative sample of 300 public and private primary schools found that school efficiency varied by school type, with rural community schools generally less efficient (Ibourk & El Aynaoui, 2026). Against this background, the literature identified for the present study appears

fragmented across levels of analysis. Research has examined neighborhood and family conditions in relation to moral judgment, socioeconomic influences on fairness judgments, classroom interaction, and student- and school-level inequalities, but these strands have rarely been integrated within a single model. In the literature located for this review, no study was identified that jointly examined observed classroom interaction, pupils' educational resources, and sector-setting configuration within a multilevel analysis of moral judgment in Moroccan primary schools. The present study addresses this gap while treating sector-setting contrasts as descriptive contextual configurations rather than as causal effects.

A sociological reading thus invites us to consider classroom properties together with pupils' social origins. At the statistical level, a climate-by-resources interaction may be compatible with two contrasting patterns. A weaker resource gradient at higher climate scores would be compatible with compensation, whereas a steeper gradient would be compatible with cumulative advantage. This comparison is informed by the literature on socially differentiated educational opportunities (Bourdieu, 1986; Bourdieu & Passeron, 1970; Coleman, 1988; DiPrete & Eirich, 2006), but it does not identify a mechanism of resource conversion.

Within this framework, school sector and setting are more than a backdrop. The observed sample contains private and public schools in urban and peri-urban settings, together with rural public schools, and these cells differ in social composition and pedagogical conditions. A raw configuration gap may capture differences in family resources, school organization, access, or interactional quality, and not necessarily any effect of location or sector as such. The issue is less one of mechanically opposing urban and rural than of understanding how unequally distributed schooling conditions are bundled across sector and place.

In Morocco, the issue resonates with institutional orientations that assign schools an explicit role in values education. The 2015–2030 Strategic Vision and framework Law 51-17 tie this mission to demands for equity and quality (Conseil supérieur de l'éducation, de la formation et de la recherche scientifique [CSEFRS], 2015, 2017). Within this outlook, moral education and the organization of school life are conceived as levers for advancing the emancipatory dimension of schooling.

The Moroccan context nonetheless remains sparsely documented by empirical research specifically devoted to school climate. In two public primary schools, Alladatin and Barjit (2023) show that pupils' perceptions combine relational, material, and territorial dimensions. PISA 2022 data likewise reveal a fairly widespread sense of belonging among 15-year-old Moroccan pupils, alongside difficulties with discipline, safety, and exposure to bullying (OECD, 2023b). The national report on gender equality in and through education similarly lists school climate and environment among the dimensions to be examined through the lens of social and gender inequality (CSEFRS, 2024).

Moroccan research on values education has focused mainly on curricula and prescribed materials. Ben El Bou (2020) examines the place of values in the Islamic-education curriculum, while Oussikoum (2023) shows that images in primary-school textbooks contribute to shaping ethical, civic, and social representations, while at times also perpetuating stereotypes. The present study shifts attention from curricular prescription to the everyday interactions in which norms are actually experienced, discussed, and justified.

The study was conducted in Meknès province across seven municipalities—three urban, two peri-urban, and two rural—selected to represent heterogeneous school contexts. The study makes four methodological contributions. First, moral-justification responses are recoded separately for each vignette so that numerical response position is not conflated with theoretical level. Second, the analysis explicitly models the nesting of pupils within classrooms. Third, pupils' educational resources are decomposed into within-class individual resources and classroom-mean composition. Fourth, the climate-by-resources interaction is interpreted against two competing associational patterns: compensation and cumulative advantage.

Five questions guide our analysis: How much of the variation in the exploratory moral-justification index lies between classrooms? Does climate remain associated with this index once

social and school characteristics are controlled for? Do individual educational resources and classroom composition produce distinct associations? Does the climate-moral-judgment relationship vary according to pupils' resources? And, finally, do descriptive school sector-setting differences persist once these measured characteristics are taken into account?

Accordingly, classroom climate is examined as a potential educational opportunity structure. The framing is theoretical, and the empirical question is whether the association between observed classroom climate and the vignette-based moral-justification index varies with pupils' educational resources after adjustment for the measured covariates.

2. Theoretical Framework

2.1. Classroom Climate as an Interactional Order

Classroom climate is understood here as the relatively stable configuration of relationships, expectations, rules, and possibilities for participation that organizes pupils' everyday experience. It cannot be reduced to a collective mood, nor to the sum of a teacher's individual qualities. Rather, it constitutes an interactional order: certain behaviors are encouraged, others prohibited or discouraged; some forms of speech are invited and elaborated, whereas others are cut short; and rules may be imposed as bare constraints or articulated as shared principles. This conceptualization aligns with the Teaching through Interactions framework, which places teacher-pupil interactions at the heart of the learning experience and distinguishes emotional support, classroom organization, and instructional support as related but distinct domains (Hamre et al., 2013).

These three domains point to distinct but potentially normatively consequential processes. A classroom in which disagreement can be voiced without fear and rules are applied clearly and fairly provides safer ground for pupils to speak, question, and justify their views. Recent evidence similarly distinguishes supportive authority—characterized by clarity, fairness, responsiveness, and respectful correction—from controlling authority marked by pressure, public criticism, and restricted dialogue; students' accounts connected these forms of authority to the perceived safety or risk of participation (Al-Makhalid, 2026). This distinction is reinforced by pre-service teachers' accounts indicating that humiliation, mockery, ignoring, and excessive criticism can undermine pupils' motivation, self-esteem, and classroom participation (Sancak Aydın & Nalbant, 2026). On such relational ground, open-ended questions, elaborated feedback, and instructional dialogue (Hamre et al., 2013; Pianta et al., 2012) are more likely to generate sustained argumentation because tentative or awkward answers are less likely to be experienced as socially risky. The classroom can then become, in the course of ordinary activity, a micro-society in which norms are not merely applied but can also be articulated, questioned, and discussed.

Retrospective accounts of schooling further illuminate how fairness and support are experienced through everyday teacher-pupil relationships. In Kahveci's (2023) study, university students' recollections grouped positive teacher conduct around patient and respectful communication, fair treatment, and individual support and trust. Participants described such experiences as strengthening self-confidence and making it easier to express themselves, whereas discrimination, humiliation, and oppressive conduct were associated with diminished self-confidence, motivation, and respect for teachers. The balance of praise and reprimand offers a complementary perspective on the same interactional environment. Drawing on the available literature, Kobayashi (2023) concludes that current evidence does not justify a single positive-to-negative interaction ratio applicable across classrooms. Instead, praise should generally occur more often than reprimand, while teachers remain responsive to pupils' behavior and the conditions of the particular class.

The conceptualization adopted here does not assume that every positive climate is moral in nature. A classroom can be calm and efficient while leaving little room for autonomy, participation, or normative discussion. Work on democratic classroom climate similarly treats participation, teacher relations, decision-making, justice, equality, and respect as substantive features of the classroom environment rather than equating a desirable climate with orderliness

alone (Akar & Uluçınar, 2023). The argument advanced here therefore concerns the combination of a safe relational environment, predictable and fair organization, and cognitively demanding interaction. If such interactional conditions matter for pupils' moral justification, variation in moral-justification scores should be observable not only between pupils but also between classrooms, and more favorable classroom climates should be associated with higher scores. These expectations lead to the first two hypotheses:

H1. A substantial share of the variation in scores on the exploratory ten-situation moral-justification index (JM10) lies between classrooms.

H2. A more favorable classroom climate is associated with higher JM10 scores, independently of individual and family characteristics.

2.2. Moral Judgment between Development and Socialization

Moral judgment is understood here as the capacity to select and rank justifications when facing situations of transgression, conflict, or responsibility. In the Piagetian and Kohlbergian traditions, development is described as a transformation in the structure of reasoning, from heteronomous or consequence-centered orientations toward forms of judgment increasingly organized around reciprocity, shared expectations, social order, and, at more advanced levels, generalizable principles (Kohlberg, 1984; Piaget, 1932/1965). The social-domain tradition complicates any single linear hierarchy by showing that children differentiate moral concerns from social conventions and other domains of social experience, and that the criteria brought to judgment vary with the kind of situation and norm at stake (Nucci, 2001; Turiel, 1983).

Post-Kohlbergian extensions support a less rigid reading of reasoning levels. Rest et al. (1999) reformulate cognitive structures in terms of moral schemas rather than sharply bounded stages, permitting a less categorical interpretation of differences in moral reasoning. Gibbs (2019) similarly situates justice reasoning within a broader account that includes social perspective-taking, empathy, moral motivation, and prosocial and antisocial conduct. Cultural context further matters for how obligation, autonomy, authority, and internalization are understood. As described by Halstead (2007), Islamic moral education emphasizes the duties set out in the shari'ah and in Islamic teaching, rejects moral autonomy as a goal of moral education, and treats the internalization of Islamic values as central to moral formation. The broader significance of culturally embedded normative frameworks is also visible in value-based education in Bhutan, where moral responsibility, discipline, and appropriate conduct are articulated through specific cultural traditions (Dorji et al., 2025). Taken together, these perspectives justify reading pupils' responses to the vignettes as situated orientations of justification rather than as an exhaustive ranking of their morality.

This dual character has direct implications for measurement. The vignettes should not be assumed a priori to function as interchangeable indicators of a perfectly unidimensional trait; they also constitute a deliberately varied sample of moral situations (Rest et al., 1999; Turiel, 1983). A global score may nevertheless summarize a general orientation if the empirical structure of the responses supports such an interpretation. Internal consistency and factor structure should therefore be examined, together with sensitivity analyses assessing whether the overall result changes materially when atypical situations are excluded. Situation-level analyses are likewise necessary to determine whether an apparent general pattern is driven by a limited subset of scenarios.

From a sociological standpoint, moral judgment is therefore not treated as an isolated property of the individual. The forms of justification available to pupils are also shaped by opportunities to participate, explain their positions, encounter alternative viewpoints, and experience authority as fair and intelligible rather than merely controlling. Democratic classroom perspectives foreground participation, respect, equality and freedom, teacher relationships, and opportunities for decision-making as elements of pupils' normative educational environment (Akar & Uluçınar, 2023). Related evidence from retrospective accounts of schooling distinguishes supportive authority characterized by clarity, fairness, responsiveness, and respectful correction from controlling

authority marked by pressure, public criticism, and restricted dialogue; these contrasting experiences were linked to the perceived safety or risk of expressing one's academic voice (Al-Makhalid, 2026). Work on moral education in schools likewise emphasizes the role of teachers, disciplinary practices, and the school environment in processes of moral-value internalization (Durasa et al., 2024). These studies do not establish that such classroom conditions directly produce more advanced moral judgment. They do, however, make the socialization mechanism examined here theoretically plausible: pupils encounter some settings in which reasons are solicited, explanations are valued, and rules can be understood as more than sanctions. Accordingly, JM10 is treated throughout the analyses as an exploratory index of situated moral justification rather than as an exhaustive measure of pupils' moral development.

2.3. Educational Resources and Cumulative Advantage

The household educational resources considered in this study—number of books, internet access, and the presence of a desk or workspace—are treated as a limited proxy for study-related resources associated with socioeconomic position rather than as a comprehensive measure of socioeconomic status. Parental educational attainment is retained separately as an indicator of family educational and cultural background. Comparable multilevel research distinguishes household educational resources from school-level socioeconomic composition (Coşkun & Karadağ, 2023).

The resource measure is decomposed into the pupil's deviation from the classroom mean and the classroom mean itself, separating within- from between-classroom associations (Enders & Tofighi, 2007). The climate-by-resources interaction tests whether the association between resources and JM10 varies with classroom climate. Because the data are cross-sectional and observational, this interaction is interpreted as a pattern of differential association rather than as evidence of a causal resource-conversion process or cumulative advantage as a temporal process (DiPrete & Eirich, 2006). This framework leads to one resource hypothesis and two competing moderation hypotheses:

H3. Individual educational resources and a classroom's average resource composition are positively associated with JM10 scores.

H4a. Compensatory hypothesis: The positive association between classroom climate and JM10 scores is stronger among less-resourced pupils.

H4b. Competing cumulative-advantage hypothesis: The positive association between classroom climate and JM10 scores is stronger among more highly resourced pupils.

2.4. School Sector, Setting, and Contextual Configuration

The source data contain five observed sector-setting cells: private urban, private peri-urban, public urban, public peri-urban, and public rural. For M3, the two private cells are pooled into one reference category because the private peri-urban cell contains only two classrooms and no private rural schools were sampled. The resulting four-level variable is therefore interpreted as a descriptive sector-setting configuration rather than as a pure sector or territorial effect.

Configuration coefficients are treated as exploratory contextual contrasts. Their attenuation after adjustment is not interpreted as mediation (Rijnhart et al., 2021). Separate sector and setting main effects are reported only as a sensitivity analysis, while sector-by-setting interactions are not estimated because of the incomplete and highly uneven cross-classification. Sector and setting are used to situate classroom climate and household resources within broader inequalities in educational opportunities and school resources rather than as causal exposures in themselves (Çimen et al., 2024; Parry & Kumi, 2026). Any change in the configuration coefficients after adjustment is interpreted descriptively rather than as evidence of mediation. Given the sparse and unbalanced contextual structure, the corresponding hypothesis is deliberately exploratory:

H5. Descriptive differences among the observed school sector-setting configurations are attenuated after adjustment for educational resources and classroom climate. Given the small, unbalanced contextual sample and the sparse private peri-urban cell, this hypothesis is treated as

exploratory.

3. Method

3.1. Research Design

The study used an observational, non-experimental, multilevel design. Pupils were sampled within classrooms and classrooms within schools, whereas the primary statistical model included two analytical levels: pupils at Level 1 and classrooms at Level 2. School remained part of the sampling hierarchy. A three-level model with pupils nested in classrooms and classrooms nested in schools was estimated as a diagnostic sensitivity analysis (Section 6; Supplementary Table S10), but limited within-school replication and an unstable variance decomposition precluded its use for primary inference. Pupil-level characteristics—educational resources, moral judgment, and sociodemographic variables—were measured cross-sectionally in a single wave, whereas classroom climate was based on three repeated observation sessions per classroom. Because classroom climate is an emergent property of ordinary teacher-pupil interaction, no experimental manipulation was attempted. The primary two-level framework separates pupil-level covariates from classroom-level associations, consistent with methodological guidance for hierarchical data (Maas & Hox, 2005; Snijders & Bosker, 2012), while school-level dependence is examined through dedicated sensitivity analyses.

3.2. Participants

Eligible pupils were enrolled in grades 1 through 6 of primary school in public or private schools in Meknès province, had parents or legal guardians who had given consent for the individual components of the study (the moral-judgment questionnaire, educational resources, sociodemographic characteristics), and had themselves given assent.

The analytic sample comprises 337 pupils across all six primary grades, grouped into 24 classes in 14 schools in Meknès province. It includes 170 boys (50.4%) and 167 girls (49.6%); mean age is 9.12 years ($SD = 2.02$; range: 6–13 years). The observed sector-setting cells are 87 pupils in private urban schools (4 classes, 2 schools), 47 in private peri-urban schools (2 classes, 2 schools), 79 in public urban schools (6 classes, 4 schools), 60 in public peri-urban schools (6 classes, 3 schools), and 64 in public rural schools (6 classes, 3 schools). The two private cells are pooled only for the four-level M3 configuration contrast.

Sampling followed a multistage stratified procedure. The design targeted two classrooms in each of 14 participating schools and exactly four classrooms at each grade level from Grade 1 through Grade 6. This yielded 12 classrooms in Grades 1–3, observed with CLASS K–3, and 12 classrooms in Grades 4–6, observed with CLASS Upper Elementary. All 14 schools were retained. Ten schools contributed both planned classrooms (20 classrooms), whereas four schools contributed one classroom each because the second targeted classroom did not meet the requirement that parental consent be obtained for every pupil before classroom access and questionnaire administration. Accordingly, 24 of 28 targeted classrooms were retained (85.7%). The four non-retained classrooms are not classified as parental refusals: the complete-consent condition was simply not met within the authorized fieldwork period. The analytic dataset does not contain the individual recruitment denominators required to calculate a pupil-level participation rate, and those denominators could not be reconstructed reliably from the available de-identified research files. We therefore report class-level retention and acknowledge the absence of a pupil-level participation denominator as a limitation.

Parental consent was collected in accordance with the field procedure approved by the ethics committee. Oral consent was used primarily in rural and peri-urban settings, where some parents faced literacy-related barriers or preferred to provide explicit oral agreement rather than sign a written form. Written consent was obtained from a smaller number of parents. The absence of a signed form did not constitute refusal when explicit oral consent had been provided. Pupil assent was obtained only after parental consent and before questionnaire administration. Institutional authorization was also obtained from the participating schools and parent associations. Consent

records are confidential and are not included in the anonymized dataset or supplementary files.

3.3. Data Collection and Measures

Data collection drew on two distinct instruments and procedures, presented in turn in this section: observation of classroom climate (3.3.1) and administration of the moral-judgment questionnaire (3.3.2), supplemented by the collection of educational resources and control variables (3.3.3).

Classroom observations were conducted by the primary observer, a French-language pedagogical inspector classified as Hors Grade within the Moroccan public-service framework and a graduate of the Centre de Formation des Inspecteurs. The observer has 14 years of experience as a primary-school teacher and a further 14 years in pedagogical inspection, the latter devoted to classroom observation, professional support, and evaluation of teaching practices. The observer also completed a five-day continuing-education course focused on CLASS administration, delivered within the inspectorate's professional-development program by a university specialist in psychometric assessment. This training was not certification issued by the CLASS developer. Professional experience and training support rater qualification but cannot substitute for standardized procedures or empirical evidence of agreement. Research on rater effects likewise shows that rater characteristics, preparation, and shared frames of reference can influence the consistency and accuracy of judgments (Myford & Wolfe, 2003; Woehr & Huffcutt, 1994). The potential role effect associated with the observer's professional position is addressed under Limitations.

The initial schedule nonetheless had to be adjusted on several occasions without changing the planned number of sessions per classroom. In rural schools, access sometimes required accommodation of teachers' transport constraints. Exceptional rainfall during the field period also required rescheduling part of the second observation phase. This operational adjustment did not change the observation protocol.

3.3.1. Classroom-climate measure

Classroom climate was assessed using the Classroom Assessment Scoring System (CLASS®), which focuses on the quality of teacher-pupil interactions. The French-language CLASS materials served as the scoring reference throughout the study. For Grades 1–3, observations followed the K–3 framework, organized around Emotional Support, Classroom Organization, and Instructional Support; for Grades 4–6, observations followed the Upper Elementary framework appropriate to that age range (Pianta et al., 2008; Pianta et al., 2012). Scores were recorded on the standard 1–7 CLASS metric. The study does not reproduce the proprietary behavioral indicators, scoring anchors, or examples contained in the CLASS manuals. External observation avoids reliance on teachers' accounts of their own practice, which may differ from pupils' experiences. In a Croatian survey involving 951 pupils and 280 teachers, Žižanović (2025) found that teachers generally assessed the conditions and practices associated with active learning more favorably than pupils did. This divergence illustrates the importance of distinguishing between perspectives on classroom processes, without itself establishing the superiority of observational measures.

The two CLASS versions were assigned a priori by grade band rather than combined into a newly created instrument. Version-specific manifestations were interpreted within the corresponding age-appropriate framework, while the analytical climate summaries were harmonized at the domain level for cross-grade modeling. Because grade band deterministically selected CLASS version, comparisons across Grades 1–3 and Grades 4–6 are descriptive and cannot establish measurement invariance or separate developmental differences from version-related differences.

The primary observer rated every classroom during three separate sessions. The provincial education authority authorized a second inspector to observe only four classrooms, once in each classroom. This restriction was imposed by the institutional conditions of field access and prevented the broader double-observation component initially planned. The four classrooms observed by both inspectors were used solely for a limited calibration exercise.

Before the double-rating exercise, the operational definitions and rating criteria were harmonized so that both observers applied a common interpretive frame, thereby reducing variation attributable to coding differences rather than to the classroom processes themselves (Hallgren, 2012; Kolbe & Burnett, 1991). For this purpose, an auxiliary bilingual harmonization aid was prepared from the French-language CLASS materials. Direct translation, blind back-translation, committee review, and expert appraisal were used only to clarify the operational wording shared by the two observers. The resulting material was a calibration aid, not an Arabic version of CLASS administered to pupils or teachers, and the French-language CLASS materials remained the scoring reference. All 44 analytic indicators retained their original dimension assignments. Because the adapted Upper Elementary set covered the broader range of dimensions and largely encompassed those of the K-3 framework, the expert review was conducted on that set. Five university professor-researchers rated the relevance of the harmonized wording: S-CVI/Ave = .95 and S-CVI/UA = .73; 32 indicators had I-CVI = 1.00 and 12 had I-CVI = .80. Under Lynn's (1986) strict criterion for a five-expert panel, the latter values fall short of unanimity, and we report them as such. Their modified kappa was .763, and mean modified kappa across all indicators was .935 (Polit & Beck, 2006). These indices document expert agreement on the auxiliary harmonization material. They are not presented as evidence that a new translated CLASS instrument has been validated or that measurement equivalence across languages or grade-band versions has been established.

The primary observer is a native Arabic speaker and professionally proficient in French, allowing classroom interactions conducted in Arabic and/or French to be interpreted directly while applying the French-language CLASS materials. Previous CLASS applications in Arabic-speaking settings provide contextual precedent for observational use in the region (Verma et al., 2022; von Suchodoletz et al., 2020), but they do not establish cultural or measurement equivalence for the present sample.

Three broad domain scores were computed: emotional support, classroom organization, and instructional support. The overall index corresponds to their mean and was standardized across the 24 classes. Pupil engagement was reserved for robustness analysis. Engagement is not a dimension of the published CLASS K-3 framework, which comprises ten dimensions within three domains. To obtain a common set of eleven dimensions across grades, Engagement was also rated in Grades 1-3 classrooms, applying the definition used in the adapted Upper Elementary framework. Some of these classrooms included over-age pupils who had repeated a grade. Because this rating is non-standard for the K-3 version, Engagement was excluded from the climate index used in all models and served only in robustness analyses. Excluding its four indicators left the reliability and calibration results essentially unchanged (overall $\alpha = .988$, 95% CI [.982, .991]; ICC(2,1) = .986 across 40 dimension-by-class pairs).

Internal consistency was estimated across the 24 unique classrooms rather than across the 337 pupil rows that repeat the same contextual scores. Alpha coefficients for the eleven CLASS dimensions ranged from .867 to .973; corresponding broad-component values were .971 for Emotional Support, .960 for Classroom Organization, .967 for Instructional Support, and .897 for Engagement. The overall index yielded $\alpha = .988$; a nonparametric classroom bootstrap gave a 95% CI [.983, .991] (5,000 replications; seed 20260803). Such a high coefficient may reflect genuine covariation, redundancy, a single-observer halo, or some combination of these processes and should not be interpreted as evidence of inter-observer reliability. In the four double-rated classrooms, the 44 dimension-by-class pairs yielded ICC(2,1) = .987 for absolute agreement. A class-cluster bootstrap that resampled classroom identifiers with replacement while retaining each classroom's eleven dimension observations produced a 95% CI [.953, .998] (20,000 ordered resamples; seed 20260916). Because four source classrooms generate only 35 distinct classroom-count compositions, the interval should be read as an indicative, dependence-adjusted sensitivity interval. Complementary descriptive indices were Pearson's $r = .987$, mean absolute difference = 0.074 point, exact agreement = 77.3%, agreement within 0.25 point = 93.2%, and agreement within 0.50 point = 100%. Consistent with reporting guidance for reliability and agreement studies, these

findings are treated as limited calibration evidence between the two observers, not as an estimate of inter-observer reliability generalizable to all 24 classrooms (Koo & Li, 2016; Kottner et al., 2011).

Descriptive grade-band checks were based on 12 K–3 classrooms and 12 Upper Elementary classrooms. Overall alpha was .988 and .989, respectively; dimension-level alpha ranges were .867–.973 and .804–.972, and mean inter-dimension correlations were .842 and .861. The cosine similarity between the two 11×11 inter-dimension correlation matrices was .994. These statistics describe coarse similarity only. With 12 classrooms per group and CLASS version determined by grade band, the design cannot establish configural, metric, or scalar invariance or distinguish developmental differences from version-related differences.

3.3.2. Moral-judgment measure

Moral judgment was measured with a contextualized scale of moral reasoning for the primary-school setting. The instrument draws on the first four levels of the Kohlbergian framework, without being equated with the Moral Judgment Interview or the Moral Competence Test. It aims to identify the level of justification a pupil selects when facing a concrete situation, not to assign the child a definitive structural stage. The ten vignettes concern cheating, mockery, exclusion, conflict, resource allocation, rule-compliance, mutual help, acknowledging one's mistake, contestation, and responsibility.

Each vignette is followed by four predetermined justifications corresponding to obedience and avoidance of punishment, instrumental exchange, interpersonal expectations, and the maintenance of collective rules. Postconventional levels were not offered, given the conceptual and verbal complexity their justifications would pose for primary-school pupils. A response indicates the reasoning a pupil favors in that particular situation and says little about the pupil's morality as a whole.

Kohlbergian theory has been the subject of extensive cross-cultural research. This body of work lends partial support to a developmental progression while also raising questions about the cultural validity of the content, the frequency of higher levels, and the socially accepted forms of justification (Snarey, 1985). The dilemmas were accordingly reformulated from situations familiar to children attending school in Morocco, to improve their ecological relevance without claiming equivalence with standardized instruments. Research on children's morality also draws on approaches beyond the cognitive-developmental tradition. In their study of an indigenous-language storytelling intervention with pupils with mild intellectual disability, Jacob et al. (2024) reported using an adapted Moral Foundations Questionnaire to assess moral values. The distinction is relevant here: their measure addresses the moral concerns pupils endorse, whereas the present study examines how pupils justify their judgments.

The mapping between response options and the four theoretical levels was reviewed by one expert who had taken part in the Arabic adaptation of the Moral Competence Test. No classroom pretest was conducted because the provincial authority did not authorize an additional field phase. Minor wording clarifications were made during instrument development to improve the comprehensibility of selected situations and options; they did not alter the number of situations, the four-option response format, or the level assigned in advance to each option. A complete item-level revision log was not retained. The exact Arabic questionnaire used in the field is treated as the authoritative administered instrument and is supplied with its coding key. Because no second qualified specialist independently cross-validated the option-to-level mapping, this procedure represents content review, not formal evidence of construct validity or measurement invariance.

To reduce comprehension difficulties, the vignettes were phrased in accessible language and anchored in plausible school or family experiences. This contextualization does not establish either the universality of the stages or equivalence with Kohlberg's standard scoring system. It aims instead at an ecologically relevant measure of justification orientations.

Because the material order of the four response options varied across vignettes, the initial numeric codes did not correspond to a constant level. Each item was recoded separately from the correspondence manual, after which a mean score from 1 to 4 was computed. Higher values

indicate more frequent recourse to interpersonal or normative justifications; lower values point instead to sanction or immediate self-interest. The four codes are ordinal and bounded, so their arithmetic mean is treated as an approximately continuous summary for the linear models. The situational sensitivity analysis addresses this assumption only partially.

Across the 337 protocols, JM10 shows a Cronbach's alpha of .853 (nonparametric pupil-bootstrap 95% CI [.829, .873]; 5,000 replications, seed 20260803), an ordinal alpha of .886, and a one-factor ordinal omega of .908. Ordinal alpha and omega are computed from the polychoric inter-item correlation matrix; omega uses a one-factor principal-axis common-factor solution. Their convergence supports use of the ten-situation mean as a broad exploratory composite for association analyses, but does not establish a validated unidimensional scale (Gadernann et al., 2012; McDonald, 1999).

Conventional factorability diagnostics calculated from the Pearson inter-item correlation matrix gave KMO = .770 and Bartlett's test of sphericity, $\chi^2(45) = 3472.23$, $p < .001$. Ordinal reliability coefficients, eigenvalue inspection, and the diagnostic factor analysis were based on the polychoric matrix. Exploratory results indicated that strict unidimensionality should not be assumed. JM10 is best read as a broad vignette-based moral-justification index and does not diagnose a single Kohlbergian stage.

The contestation and mutual-help situations show weaker item-total correlations. A post hoc sensitivity score excluding them (JM8) yields alpha = .917 (bootstrap 95% CI [.900, .932]), ordinal alpha = .940, and one-factor ordinal omega = .942, using the same computational definitions as for JM10. In the JM8 sensitivity analysis, the classroom-climate main association remains positive (classroom-clustered CR2 $p = .0043$; wild-bootstrap $p = .0103$), but the climate-by-resources interaction is imprecisely estimated (CR2 $p = .1173$; wild-bootstrap $p = .1283$). JM10 is retained as the primary exploratory outcome to preserve all ten contexts, while JM8 demonstrates that the interaction is sensitive to the outcome-scoring choice.

Finally, the 3,370 situational responses were analyzed separately, distinguishing levels 3–4 from levels 1–2 and controlling for the situation. Inference was clustered by classroom, which also nests responses from the same pupils, but the limited number of classrooms remains a constraint. This sensitivity analysis reduces reliance on the equidistance assumption without creating additional independent contextual units.

The questionnaire was administered by the other teacher assigned to the class, not by the teacher whose practice had been observed with CLASS. Moroccan primary classes are generally taught by one Arabic-medium and one French-medium teacher, and the non-observed member of that pair administered the questionnaire. The teacher read aloud each situation and then each of its four options, providing neutral clarification when needed before pupils selected a response. The option sequence was fixed for all pupils, differed across situations, and had been reordered once before field administration so that option position did not follow a uniform developmental progression. It was not randomized by pupil or class. A member of the research team attended every administration, briefed the administering teacher, monitored comprehension, and ensured that neither the developmental level assigned to an option nor a preferred moral answer was explained or suggested. Pupils answered individually; larger classes were divided into two groups to reduce copying. The paper form contained a pseudonymous pupil code, class, age, and date, but no pupil name. This supported linkage and quality control without direct personal identification. Oral support was intended to reduce, not eliminate, variation attributable to reading comprehension; no independent measure of literacy or social desirability was collected.

3.3.3. Educational resources, approximate socioeconomic status, and control variables

The individual index combines three standardized components: internet access, the presence of a desk or workspace, and the logarithm of the number of books at home. These variables represent study-related material and cultural resources. They are a limited proxy and do not measure household income, occupation, wealth, language proficiency, or family support. The classroom mean of this index forms the contextual-composition measure. Individual standing is obtained by

subtracting this mean from the pupil's score and standardizing the result, in line with recommendations on centering in multilevel models (Enders & Tofighi, 2007).

Family cultural capital is approximated by the average educational attainment of the father and mother. Sex is coded 1 for boys and 0 for girls, and grade level is standardized. Age was not entered alongside grade level because of their strong correlation in the sample ($r = .92$). Grade level was retained because it more directly captures cumulative exposure to the primary curriculum.

Two classroom-level variables were used only in robustness analyses. *Effectif_Classe* denotes the number of pupils represented in the analytic dataset for that class, not the school's administrative class enrollment. Teacher experience (*teacher_exp*) denotes completed years of service as a teacher since recruitment. It was obtained from the official pedagogical-organization records of each school at the start of the 2025–2026 school year and linked by class identifier to the teacher whose practice was observed with CLASS. The records report complete years; the observed teachers were permanent civil servants and no interruption of service was recorded. This variable measures seniority, not qualification, instructional quality, or CLASS competence.

3.4. Analytic Strategy

We designed a five-step analysis. The psychometric properties of JM10 are examined first, using Cronbach's alpha with a nonparametric pupil-bootstrap confidence interval, ordinal alpha, one-factor ordinal omega, item-total correlations, the KMO measure, Bartlett's test, and diagnostic eigenvalue inspection followed by a two-factor principal-axis extraction. KMO and Bartlett are conventional diagnostics based on the Pearson inter-item matrix; ordinal alpha, ordinal omega, eigenvalues, and the diagnostic factor analysis use the polychoric matrix. For classroom climate, internal consistency was computed on a single row per class, so as not to artificially inflate the coefficients by repeating the same contextual scores.

Random-intercept linear models were then estimated with pupils at level 1 and classrooms at level 2 (Snijders & Bosker, 2012). M0 decomposes the variance in JM10 scores between the individual and contextual levels. M1 introduces sex, grade level, family cultural capital, and the individual educational-resource proxy centered within class. M2 adds classroom climate and classroom-mean resource composition. M3 extends M2 with the four-level school configuration: private schools pooled across urban and peri-urban settings as the reference, versus urban public, peri-urban public, and rural public. M4 is a separate model extending M2 with the interaction between classroom climate and individual resources.

Because only 14 schools were sampled and four contribute a single classroom, a three-level model was not suitable for primary inference. It was nevertheless estimated as a diagnostic check. It converged with a boundary warning and non-positive-definite Hessian and is reported in Section 6 and Supplementary Table S10. The classroom remains the primary contextual unit, consistent with the study's focus on interactions observed within classrooms.

Comparisons between M0, M1, and M2, and then between M2 and each of its extensions M3 and M4, were carried out by maximum likelihood so as to allow likelihood-ratio tests. With 24 classes, inference on contextual predictors remains fragile. We therefore supplement the likelihood-based models with classroom-clustered bias-reduced linearization standard errors (CR2) and coefficient-specific Satterthwaite degrees of freedom. Restricted wild-cluster-bootstrap tests with Rademacher weights, 9,999 replications, and fixed seed 20260803 are added for focal contextual and configuration terms when pertinent. The same CR2-Satterthwaite rule is applied to configuration, class-size, teacher-experience, factor-based, and sector-setting-decomposition sensitivity models; results based only on the naive $G - 1 = 23$ degrees of freedom are not used for inference. These auxiliary ordinary least-squares (OLS) and logistic analyses are sensitivity checks rather than independent replications (Bell & McCaffrey, 2002; Cameron et al., 2008; Imbens & Kolesár, 2016; Maas & Hox, 2005; McNeish, 2017; Pustejovsky & Tipton, 2018).

Analyses were conducted in Python 3.13.5 using pandas, NumPy, SciPy, and statsmodels; the supplied analysis script records the computational specification and reproduces the JM10/JM8 reliability and factorability diagnostics reported in the manuscript. The canonical SAV was

independently checked in IBM SPSS Statistics 26, reproducing the study counts, CLASS internal-consistency coefficients, conventional mixed-model point estimates and likelihood diagnostics within expected rounding, the separate sector-setting model, the class-size and teacher-experience sensitivity patterns, and the grade-band/CLASS-version diagnostic. CR2-Satterthwaite, restricted wild-cluster-bootstrap, and polychoric analyses were obtained from the supplied Python workflow rather than native SPSS procedures. No imputation was required: all 337 pupils entered the multilevel models and all 3,370 responses entered the situational analysis. Ordinal alpha and one-factor ordinal omega were estimated from the polychoric correlation matrix; the diagnostic factor analysis likewise used the polychoric matrix.

Recent work on small level-2 samples shows that the quality of inference depends on estimation method, model specification, and the corrections applied. Yamamoto and Miyazaki (2025) in particular highlight the value of restricted maximum likelihood for estimating variance components in certain configurations with few groups. In the present study, maximum likelihood was nonetheless retained to compare nested models. Interpretive caution rests instead on the convergence of clustered standard errors, the small-sample corrections, and the sensitivity analyses.

The intraclass correlation coefficient was derived from M0. Marginal and conditional R^2 summarize, respectively, variance accounted for by fixed effects and by the model as a whole, following Nakagawa and Schielzeth (2013) and the extension by Nakagawa et al. (2017). The four-level school configuration was introduced in M3. Because the raw sample contains both urban and peri-urban private schools, a separate sensitivity model estimated additive main effects for public sector (vs private), peri-urban setting (vs urban), and rural setting (vs urban); the unbalanced design does not support a sector-by-setting interaction. Focal classroom-level tests were supplemented by CR2-Satterthwaite and restricted wild-cluster-bootstrap inference clustered first by classroom (24 clusters) and, as a more conservative sensitivity analysis for dependence shared by classrooms within the same school, by school (14 clusters). A formal climate $\times$ grade-band/CLASS-version interaction was estimated with the M4 covariate set. Because grade band deterministically selects CLASS version, this term is interpreted only as a combined grade-band/version sensitivity test. Finally, a logistic analysis of the 3,370 situational responses distinguished levels 3–4 from levels 1–2, controlled for situation and the same covariates as M2, and used inference robust to clustering by classroom (Huang et al., 2023).

Coefficients are reported with 95% confidence intervals. P-values are interpreted jointly with effect sizes, the consistency of specifications, and the limits of the design. Given the observational design, results are framed as associations.

4. Results

4.1. Social and Sector-setting Description

Across the full sample, 58.2% of pupils have internet access and 51.0% have a desk or workspace. The median number of books is 10 (interquartile range: 2–31). Combined in the educational-resources index, these three indicators also describe a material and cultural dimension of socioeconomic position. The mean JM10 exploratory-index score is 2.182 (standard deviation, $SD = 0.691$), and the mean overall climate score is 4.831 out of 7. Descriptively, the private urban cell has the highest mean resources, climate, and JM10 score, while the public rural cell has the lowest mean resources and JM10 score. Table 1 reports all five observed sector-setting cells, so that the private peri-urban schools remain visible instead of disappearing inside a single label. The descriptive gaps do not support attributing causality to sector or setting, since the observed cells combine social composition, institutional organization, and spatial context. They justify modeling the four-level pooled configuration only as an exploratory contextual contrast. Figure 1 visualizes the five observed raw sector-setting cells.

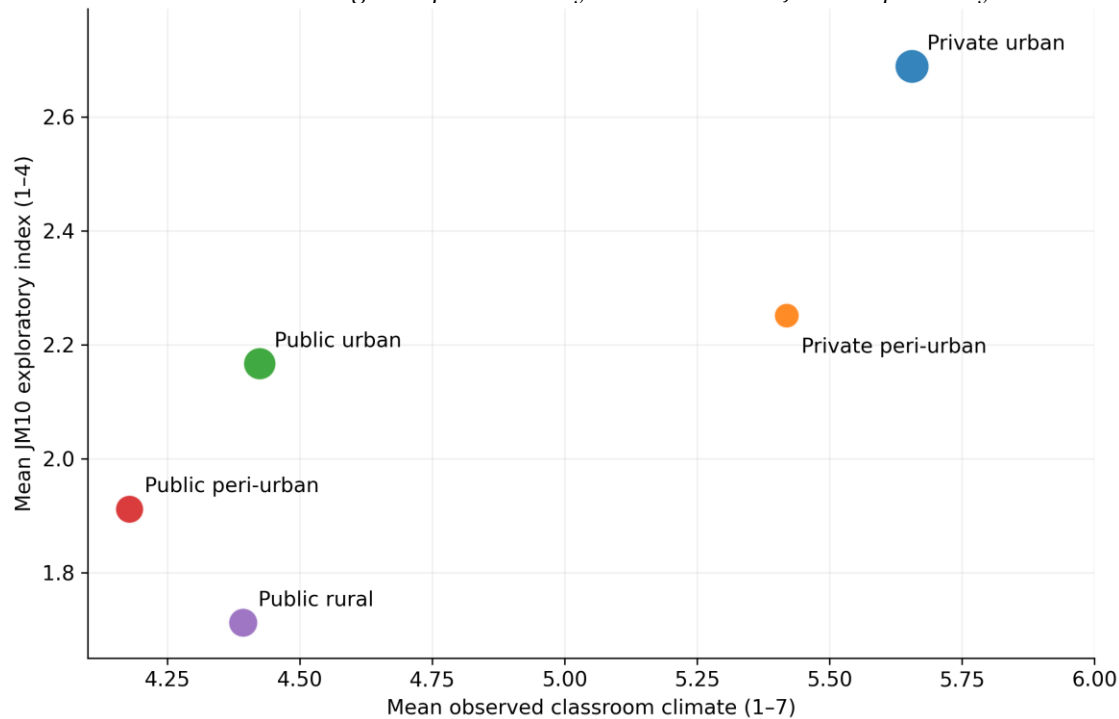
Table 1
Characteristics by observed school sector-setting cell

Sector-setting cell	Pupils	Classrooms	Schools	JM10 exploratory index	Climate	Educational resources	Cultural capital
Private urban	87	4	2	2.689	5.655	0.626	4.316
Private peri-urban	47	2	2	2.251	5.419	0.444	3.957
Public urban	79	6	4	2.167	4.424	0.150	2.823
Public peri-urban	60	6	3	1.912	4.178	−0.396	2.117
Public rural	64	6	3	1.712	4.393	−0.990	1.359

Note. JM10 is the exploratory mean of the ten moral-justification situations, measured on a 1-to-4 scale; classroom climate is measured on a 1-to-7 scale. The educational-resources index is a standardized composite of household educational resources, a partial proxy for socioeconomic status (see Section 2.3); cultural capital corresponds to parents' average educational attainment.

Figure 1

Observed school sector-setting cells positioned by climate and the JM10 exploratory index



Note. Point size is proportional to the number of pupils.

4.2. Psychometric Properties

The ten situations yielded $\alpha = .853$ (95% CI [.829, .873]), ordinal $\alpha = .886$, and one-factor ordinal $\omega = .908$. Conventional Pearson-matrix diagnostics gave $KMO = .770$ and Bartlett's $\chi^2(45) = 3472.23$, $p < .001$. These coefficients support reporting a broad composite for exploratory association analyses, but they do not establish strict unidimensionality. The first two eigenvalues of the full polychoric correlation matrix were 5.712 and 1.308. The subsequent polychoric exploratory factor analysis produced an improper Heywood solution for the Mockery item, a second factor containing only two items, and near-zero communalities for Contestation and Mutual help. Accordingly, the ten-situation mean (JM10) is described as an exploratory composite index, not a validated unidimensional scale or a diagnostic Kohlbergian stage. The full diagnostic loadings and factor-based regressions are reported only in Supplementary Tables S4-S5. Their interactions receive little support under CR2-Satterthwaite inference (Factor 1: $b = 0.110$, $p = .099$; Factor 2: $b = 0.034$, $p = .545$) and are not evidence that the interaction is consistently present across facets.

JM8, obtained after excluding the two weak situations, is a post hoc sensitivity score rather than an independently validated or preregistered measure. It is used solely to test whether the main conclusions depend on Contestation and Mutual help. Its climate main-effect result is similar in direction to JM10, whereas the evidence for the climate-by-resources interaction is weak under classroom-clustered CR2 or wild-bootstrap inference (Supplementary Table S12). Neither JM8 nor the factor-specific analyses establishes measurement invariance across ages or grades.

All CLASS dimension coefficients exceeded .86. The overall $\alpha = .988$ indicates extremely high internal consistency, but such a value may also reflect substantial overlap among indicators or shared-observer effects. Domain-specific sensitivity analyses were used to assess whether the principal association depended disproportionately on a single component. Table 2 summarizes reliability estimates for the exploratory moral-justification composites and the CLASS measures.

Table 2

Reliability of the main measures

Measure	Items	Alpha	95% CI	Ordinal alpha	Omega
JM10 exploratory composite	10	.853	[.829, .873]	.886	.908
JM8 post hoc sensitivity score	8	.917	[.900, .932]	.940	.942
Emotional support	16	.971*	—	—	—
Classroom organization	12	.960*	—	—	—
Instructional support	12	.967*	—	—	—
Overall climate	44	.988*	[.983, .991]	—	—

Note. * Coefficients computed on the 24 unique classrooms. The three domain coefficients are derived from the corresponding item sets. The overall coefficient covers all 44 indicators, including the four Engagement indicators; excluding them yields $\alpha = .988$ (95% CI [.982, .991]). The climate index used in the models is the mean of the three domain scores.

4.3. Between-class Variance and Main Effects

The null model estimates a between-class variance of 0.155 and a residual variance of 0.304. The intraclass correlation coefficient is .338: 33.8% of the variation in JM10 scores lies between classrooms. H1 is supported with respect to this exploratory index, and the hierarchical structure cannot be ignored.

Introducing individual characteristics in M1 substantially improves fit relative to M0, $\chi^2(4) = 332.69$, $p < .001$. Grade level is positively associated with JM10 scores ($b = 0.254$, $p < .001$), as are educational resources relative to classmates ($b = 0.426$, $p < .001$). Parental cultural capital shows no independent association; the coefficient for boys is negative but close to the conventional threshold.

Adding classroom climate and classroom-mean resources in M2 further improves fit, $\chi^2(2) = 51.56$, $p < .001$. A one-standard-deviation increase in climate is associated with a 0.141-point increase in JM10 (95% CI [0.086, 0.196], $p < .001$). Classroom-mean resource composition is also positive ($b = 0.249$, 95% CI [0.183, 0.314], $p < .001$), as are within-class individual resources ($b = 0.430$, $p < .001$) and grade level ($b = 0.201$, $p < .001$). Fixed effects explain 75.9% of the total variance; the residual ICC falls to about 5.2%. H2 is supported in the primary mixed model and classroom-clustered checks but is sensitive to the more conservative school-clustered wild bootstrap reported in Section 4.5; H3 is supported for the exploratory JM10 outcome. Table 3 presents the nested maximum-likelihood multilevel models.

4.4. Interaction between Climate and Resources

M4 tests the competing compensation and cumulative-advantage hypotheses. The maximum-likelihood mixed model yielded a positive interaction ($b = 0.075$). In the auxiliary classroom-clustered OLS specification, $b = 0.076$ (SE_CR2 = 0.031), Satterthwaite $df = 7.6$, $p = .0417$, with restricted wild-bootstrap $p = .043$. When the same auxiliary specification was clustered at the 14-school level, the point estimate was unchanged but uncertainty increased (SE_CR2 = 0.036, $df = 6.08$, $p = .0756$; wild-bootstrap $p = .097$, 95% CI [−0.017, 0.169]). The interaction is positive but sensitive to clustering level and is not considered robust evidence of cumulative advantage.

Table 3
Multilevel models predicting the exploratory JM10 composite

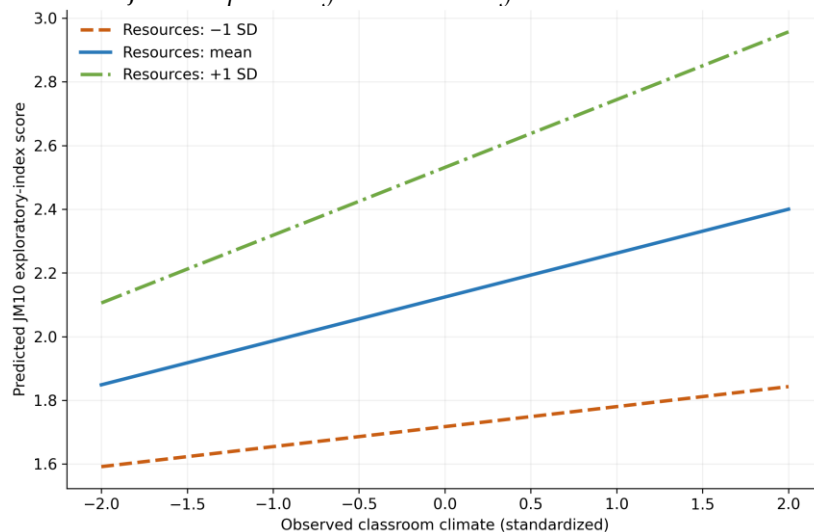
Predictor	M1 individual	M2 contextual	M3 school configuration	M4 interaction
Boy	−0.070† (0.037)	−0.072* (0.036)	−0.074* (0.036)	−0.058 (0.036)
Grade level	0.254*** (0.068)	0.201*** (0.026)	0.209*** (0.032)	0.203*** (0.025)
Cultural capital	0.021 (0.033)	−0.004 (0.030)	−0.010 (0.030)	0.014 (0.029)
Individual educational resources	0.426*** (0.019)	0.430*** (0.019)	0.431*** (0.019)	0.407*** (0.019)
Classroom climate	—	0.141*** (0.028)	0.120*** (0.030)	0.138*** (0.027)
Classroom composition in educational resources	—	0.249*** (0.033)	0.222† (0.133)	0.237*** (0.033)
Climate × resources	—	—	—	0.075*** (0.018)
Urban public	—	—	−0.112 (0.117)	—
Peri-urban public	—	—	−0.147 (0.228)	—
Rural public	—	—	−0.097 (0.347)	—
Akaike information criterion (AIC)	285.642	238.087	241.480	223.319
Marginal R ²	.548	.759	.763	.771
Conditional R ²	.759	.772	.773	.783
N (pupils)	337	337	337	337
N (classrooms)	24	24	24	24

Note. Unstandardized coefficients; standard errors in parentheses. Model-based standard errors are taken from the full maximum-likelihood Hessian; the supplied code cross-checks the same point estimates and variance components with a direct Cholesky implementation. The small-sample-corrected results in Table 4 guide interpretation of the focal and configuration coefficients. M3 reference category: private schools pooled across urban and peri-urban settings. † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$. M0: between-class variance = 0.155, residual variance = 0.304. Marginal and conditional R² were computed following Nakagawa and Schielzeth (2013), as updated by Nakagawa et al. (2017). The markedly larger standard error for classroom resource composition in M3 (0.133 versus 0.033 in M2 and M4) reflects collinearity between class composition and the configuration indicators; it does not indicate a data-entry error.

Based on the model-based and robust coefficients, the estimated slope of climate is positive at low, mean, and high levels of individual resources, but is weaker at the lower resource level. Because the interaction is sensitive to school-level clustering and also weakens in the JM8 outcome sensitivity analysis, H4b receives only tentative support. Figure 2 displays the corresponding model-based predictions; it does not depict a causal effect of changing climate.

Figure 2

Predicted JM10 exploratory-index score by classroom climate and educational resources



Note. Predictions use the fixed portion of M4 with other standardized covariates held at zero and sex fixed at the reference category (female). Curves illustrate the fitted statistical interaction and do not represent causal effects.

4.5. School Configuration and Robustness Analyses

Adding the four-level school configuration in M3 did not improve fit relative to M2, $\chi^2(3) = 2.61$, $p = .456$. Relative to private schools pooled across urban and peri-urban settings, the mixed-model estimates for urban public ($b = -0.112$), peri-urban public ($b = -0.147$), and rural public ($b = -0.097$) were imprecise, with intervals that included zero. These estimates should not be interpreted as evidence of equivalence between configurations.

The structurally parallel auxiliary OLS specification was examined with classroom-clustered CR2-Satterthwaite inference. The configuration contrasts were -0.092 for urban public ($SE_CR2 = 0.110$, $df = 7.63$, $p = .426$), -0.116 for peri-urban public ($SE_CR2 = 0.196$, $df = 7.28$, $p = .573$), and -0.046 for rural public ($SE_CR2 = 0.307$, $df = 7.06$, $p = .885$); restricted wild-bootstrap tests led to the same substantive conclusion. The M3 classroom resource-composition coefficient was also stress-tested because its model-based standard error increases sharply when configuration indicators are added: the auxiliary estimate was $b = 0.232$, $SE_CR2 = 0.129$, $df = 7.75$, $p = .112$; wild-bootstrap $p = .116$, 95% CI $[-0.069, 0.533]$.

The adjusted configuration coefficients were imprecise, which does not imply that sector-setting differences disappear or are mediated. The raw data contain private schools in both urban and peri-urban settings, so sector and setting are not perfectly confounded. A separate main-effects decomposition (public sector; peri-urban and rural settings relative to urban) was estimated as a sensitivity analysis. It did not improve fit relative to M2, $\chi^2(3) = 5.64$, $p = .130$, and none of the three classroom-clustered CR2/wild-bootstrap contrasts reached conventional significance (Supplementary Table S11). Because there are no rural private schools and only two peri-urban private classrooms, this unbalanced decomposition cannot support a sector-by-setting interaction or causal attribution. H5 is therefore not confirmed.

The positive climate coefficient also appeared when the overall climate index was replaced, one at a time, by standardized emotional support, classroom organization, instructional support, and engagement components in otherwise M2-type maximum-likelihood models ($b = 0.133$ – 0.146). The JM8 sensitivity analysis likewise retained a positive climate association, while its climate-by-resources interaction was imprecisely estimated under classroom-clustered CR2 or wild-bootstrap inference. A situation-level logistic analysis of the 3,370 responses, with inference robust to the 24 classroom clusters (Huang et al., 2023), yielded an odds ratio (OR) of 1.463 (95% CI $[1.164, 1.841]$) for classroom climate. These alternative climate-summary and outcome-scoring checks are reported in Supplementary Table S12; they reuse the same sample and are not independent replications of the class-level result.

These sensitivity results show that inferential conclusions depend on the contextual clustering level. For the M2 climate coefficient, school-clustered CR2-Satterthwaite inference gave $b = 0.132$, $SE_CR2 = 0.048$, $df = 7.48$, $p = .0265$, whereas the 14-school wild bootstrap gave $p = .0548$, 95% CI $[-0.003, 0.266]$. For the M4 climate-by-resources interaction, school-clustered p values were .0756 (CR2-Satterthwaite) and .0973 (wild bootstrap). The formal climate $\times$ grade-band/CLASS-version interaction was $b = 0.130$: classroom-clustered $p = .0293$ and wild $p = .0470$, but school-clustered $p = .0569$ and wild $p = .0712$. These checks are sensitivity analyses rather than independent replications, and the grade-band term cannot distinguish developmental moderation from CLASS-version moderation.

Under classroom clustering, the climate main effect and climate-by-resources interaction are statistically distinguishable from zero for the primary JM10 outcome using the prespecified small-sample procedures. Under 14-school clustering, the climate coefficient remains distinguishable from zero by CR2-Satterthwaite ($p = .0265$) but not by the restricted wild bootstrap at the .05 threshold ($p = .0548$), while the climate-by-resources interaction is not statistically distinguishable from zero by either school-clustered procedure. The JM8 outcome sensitivity is more conservative still for the interaction (classroom-clustered CR2 $p = .1173$; wild-bootstrap $p = .1283$). The three four-level configuration contrasts and the M3 classroom-composition coefficient are imprecise. A Holm sensitivity adjustment across the three prespecified focal JM10 association tests (individual

Table 4

Small-sample-corrected sensitivity inference for focal and contextual coefficients

<i>Term</i>	<i>b</i>	<i>CR2 SE</i>	<i>Satterthwaite df</i>	<i>CR2 p</i>	<i>Wild-bootstrap p [95% CI]</i>
M2: classroom climate	0.132	0.037	10.3	.0052	.0096 [0.035, 0.228]
M4: climate × resources	0.076	0.031	7.6	.0417	.0427 [0.003, 0.150]
M3: urban public vs pooled private	−0.092	0.110	7.63	.4258	.3776 [−0.320, 0.136]
M3: peri-urban public vs pooled private	−0.116	0.196	7.28	.5726	.5163 [−0.508, 0.276]
M3: rural public vs pooled private	−0.046	0.307	7.06	.8849	.8751 [−0.638, 0.546]
M3 class resource composition (24 classrooms)	0.232	0.129	7.75	.112	.116 [−0.069, 0.533]
M2 climate (14 schools)	0.132	0.048	7.48	.0265	.0548 [−0.003, 0.266]
M4 climate × resources (14 schools)	0.076	0.036	6.08	.0756	.0973 [−0.017, 0.169]
M4 + climate × grade band/version (24 classrooms)	0.130	0.051	9.99	.0293	.0470 [0.003, 0.257]
M4 + climate × grade band/version (14 schools)	0.130	0.057	6.77	.0569	.0712 [−0.014, 0.274]

Note. Auxiliary OLS specifications use CR2 bias-reduced linearized standard errors with coefficient-specific Satterthwaite degrees of freedom. Restricted wild-cluster bootstrap uses Rademacher weights, 9,999 replications, seed 20260803. Classroom-clustered rows use 24 clusters; school-clustered rows use 14 clusters. The M3 configuration reference is private schools pooled across urban and peri-urban settings. The grade-band interaction retains the M4 covariate set and adds Upper Elementary (Grades 4–6) and climate × Upper Elementary. Because grade determines CLASS version, this term cannot separate development from version effects. Bootstrap confidence intervals are bootstrap-t intervals.

resources, climate, and climate × resources) gives adjusted *p*-values < .001, .0105, and .0417, respectively; this calculation does not address dependence among the broader set of robustness specifications. More than 15 distinct non-primary robustness or sensitivity specifications are reported across the main text and supplement; they are presented to reveal sensitivity to modeling, measurement, and clustering choices, and none is treated as a separate confirmatory hypothesis. An exploratory M4 specification allowed the individual-resource slope to vary across classrooms. The estimated slope variance was 0.0153 and the interaction estimate was 0.087. With only 24 classrooms, these variance components and associated tests are imprecise. The model is presented as a sensitivity analysis and does not show that the resource association genuinely differs across all classrooms.

5. Discussion

5.1. Classroom Climate as an Institutional Resource for Moral Socialization

Taken together, these findings are consistent with viewing classroom climate as an institutional resource for moral socialization, while stopping short of attributing causal effects to it. Pupils in classrooms characterized by warmer, better-organized, and more instructionally supportive interactions tended to obtain higher scores on the exploratory moral-justification index. The first major finding concerns the extent of between-class structuring. Before any predictors were introduced, approximately one third of the variance in JM10 lay between classrooms. This proportion does not show that classrooms cause these differences; it establishes that pupils' vignette-based responses are substantially clustered within learning contexts and therefore supports H1 by identifying the classroom as a relevant level of contextual variation. The classroom-climate main effect provides qualified support for H2: it was statistically distinguishable from zero under classroom-clustered inference, whereas the 14-school clustered wild-bootstrap

check yielded only marginal evidence. The stability of the direction alongside variation in statistical precision warrants a cautious interpretation of the climate association rather than a strong confirmatory claim.

Several theoretical pathways could be compatible with this association, but none can be identified with the present cross-sectional data. A safe relational climate might reduce the perceived cost of disagreement; consistently applied rules might make norms more predictable; and instructional support might provide linguistic opportunities to formulate and defend reasons. Related evidence shows that fair, respectful, and supportive teacher conduct is associated with greater trust, confidence, and willingness to engage (Kahveci, 2023), while supportive teacher authority characterized by clarity, fairness, responsiveness, and respectful correction has been linked to a safer context for expressing academic voice (Al-Makhalid, 2026). These studies do not establish a mechanism linking classroom climate to moral justification, but they make the proposed relational and discursive pathways plausible. More broadly, qualitative evidence from performance-driven schooling indicates that institutional expectations and recognition practices can shape how pupils interpret their educational experiences and construct academic identities (Danışman & Bilgiç, 2026). Such pathways remain hypotheses for longitudinal or intervention research rather than mechanisms demonstrated by the present study.

This interpretation echoes, though it does not replicate, the findings of Aldridge and McLure (2024), who found significant relationships between dimensions of school climate and students' moral identity, while also showing that the relevant climate dimensions differed across the two schools studied. Their study concerns perceived school climate and moral identity, whereas the present study examines observed classroom interaction and vignette-based moral justification; the convergence therefore lies at the broader contextual level rather than in the specific constructs or measures. Together, the findings reinforce the value of examining moral development in relation to interactional and institutional environments while using designs capable of testing developmental mechanisms.

5.2. Heterogeneity in the Educational-resource Gradient

The second major finding concerns the climate-by-resources interaction. Under classroom-level inference, its positive estimate indicates that the educational-resource gradient becomes steeper as observed classroom climate improves, whereas the evidence weakens when uncertainty is clustered at the 14-school level. The result therefore provides qualified support for H4b, the competing cumulative-advantage hypothesis, rather than H4a, the compensatory hypothesis. More precisely, it is consistent with an advantage-amplifying pattern rather than evidence of cumulative advantage as a developmental process. The interpretation remains associational and tentative because unmeasured pupil, family, teacher, or school characteristics could contribute to the observed moderation. Comparable multilevel research has shown that school characteristics can strengthen as well as weaken socioeconomic gradients, underscoring that ostensibly favorable educational contexts need not operate compensatorily (Gustafsson et al., 2018).

One possible explanation concerns the distribution of teacher attention, encouragement, and opportunities for participation. In Kahveci's (2023) retrospective accounts, discrimination based on both academic achievement and socioeconomic status was reported: some teachers were described as showing greater interest in higher-achieving pupils and as being more supportive or flexible toward pupils from financially advantaged or socially respected families. Such accounts provide a plausible illustration of how an apparently supportive classroom environment might still be experienced or used unevenly across pupils. They do not, however, demonstrate that differential teacher treatment explains the interaction observed here. The present study did not measure differential treatment, prior attainment, or comparable pupil-specific exposure to teacher support, so these possibilities remain questions for longitudinal or observational research with richer measures of classroom processes.

The broader literature also cautions against treating socioeconomic advantage as having a uniform relation with moral reasoning. Li et al. (2025), studying 4,122 Chinese adolescents, found

distinct associations for objective and subjective socioeconomic status across different dimensions of moral reasoning rather than a single uniformly positive SES effect. More recent evidence similarly found only limited direct associations between objective-subjective SES configurations and overall prosocial moral reasoning (Li et al., 2026). These studies differ from the present research in age, measures, cultural setting, and analytic design, and therefore contextualize rather than directly corroborate the present climate-by-resources interaction.

The interaction should consequently not be read as showing that favorable classroom climates disadvantage less-resourced pupils. The estimated climate association remains positive among less-resourced pupils; the difference lies in its magnitude, with the association being stronger among pupils with greater educational resources. Nor does the present resource index capture the family educational environment comprehensively. Family processes such as parental involvement constitute a distinct dimension: a meta-analysis of 119 studies reported a small positive association between parental involvement and student achievement (Danışman, 2017). Although that evidence concerns academic achievement rather than moral justification, it reinforces the point that material study resources represent only one component of pupils' family educational circumstances. The present finding is therefore best interpreted as preliminary evidence that the association between a favorable classroom climate and moral justification may depend on the educational resources pupils bring to that environment, rather than as evidence that classroom climate itself produces cumulative inequality.

5.3. School Sector and Setting as a Distribution of Conditions

The observed sector-setting cells show sizeable descriptive gaps, but the adjusted configuration contrasts are too imprecise to determine whether meaningful residual differences remain. The attenuation of these contrasts is directionally consistent with H5, but the hypothesis cannot be regarded as confirmed given the uncertainty of the adjusted estimates. A more cautious interpretation is that some of the unadjusted differences overlap statistically with educational resources and classroom climate. Sector and setting are therefore better understood here as configurations of educational conditions rather than as single causal exposures. This interpretation is consistent with evidence that public and private schools may share substantial resource and implementation constraints rather than differ uniformly by sector (Baidoo, 2026), while resource shortages and administrative barriers can vary across urban, semi-urban, and rural school contexts (Zagkotas, 2026). School infrastructure itself can also shape the conditions under which participation and learning are sustained (Njau et al., 2026).

The sensitivity analysis separating sector and setting does not materially change this conclusion. Although their main effects are statistically estimable, the design remains strongly unbalanced: no rural private schools were sampled and only two peri-urban private classrooms are represented. The estimates therefore cannot isolate a general territorial or sector effect, and attenuation after adjustment should not be interpreted as mediation. Related evidence also suggests that school administrators report contextual constraints related to physical facilities, transportation, and local administrative conditions (Saritiken, 2023). More broadly, these findings caution against reducing contextual differences to location alone. Future studies should use more balanced sampling across sector-setting cells and directly measure accessibility, infrastructure, staffing, and local institutional conditions before stronger territorial or sectoral interpretations are attempted.

5.4. Contribution to the Measurement of Moral Judgment

In school-based measurement of moral judgment, the study provides a practical lesson: a relatively short set of closed-response vignettes can capture meaningful variation in pupils' justification orientations, provided that the scoring structure is examined rather than assumed. This is consistent with previous work showing that vignette-based, structured-response formats can provide a practical means of assessing moral reasoning in secondary-school populations, while requiring explicit evidence concerning their internal structure and scoring properties (Lim &

Chapman, 2022a, 2022b).

The two-factor structure and the weak correlations for contestation and mutual help nonetheless point to residual heterogeneity. This is not necessarily surprising for vignette-based measurement, because moral scenarios can differ systematically in the features and domains they elicit, making situation-level examination important alongside aggregate scoring (Clifford et al., 2015). The overall JM10 score is therefore defensible only as a broad exploratory index; it should not be presented as a validated, exhaustive, or strictly unidimensional measure of moral judgment. More generally, factor structure, reliability, and other internal diagnostics provide only part of the evidence required for construct validation (Boateng et al., 2018).

The JM8 analysis is especially informative because the climate main association remains positive, whereas the climate-by-resources interaction is no longer distinguishable from zero under classroom-clustered CR2/wild-bootstrap inference. Thus, the substantive conclusion concerning classroom climate appears more robust to the alternative scoring specification than the moderation claim. At the same time, the change in inferential conclusions illustrates why measurement sensitivity and uncertainty estimation should be considered jointly. Small-sample corrections such as CR2 are specifically intended to improve cluster-robust inference when the number of independent clusters is limited (Pustejovsky & Tipton, 2018). The exploratory factor results, outcome-sensitivity analyses, and small-cluster corrections should therefore be understood as complementary diagnostics: they strengthen the transparency of the analysis, but none independently validates either the JM10 instrument or the substantive hypotheses.

5.5. Implications for Policy and Practice

These findings suggest treating values education not simply as the addition of prescribed content but as a dimension of pupils' everyday experience of schooling. Values are also enacted through teachers' instructional decisions, responses to disagreement, and management of classroom interaction, making classroom practice itself an important site of value formation (Aktaş, 2023). Teacher education and professional development could therefore give explicit attention to leading reasoned discussions, making the grounds for rules intelligible, managing disagreement without humiliation, fostering pupil participation, and using feedback that invites explanation. Such an emphasis is consistent with evidence that dialogue and interaction are central components of many approaches to moral education (Schuitema et al., 2008), while humiliating, mocking, ignoring, and excessively critical teacher responses were perceived by pre-service teachers to undermine pupils' motivation, self-esteem, and participation (Sancak Aydın & Nalbant, 2026).

Focused professional development may offer one means of supporting such practices. In a qualitative study involving eight early-childhood teachers in Türkiye, Akyol (2025) found that, following eight training sessions over four weeks, teachers reported greater involvement of children in classroom decision-making and more opportunities for choice. More broadly, practicum-based teacher education may provide opportunities for prospective teachers to reconsider classroom interaction, respond to diverse student needs, and develop more reflective pedagogical decision-making. In an eight-week qualitative study with 21 pre-service mathematics teachers, generative-AI-supported practicum experiences were reported to help candidates recognize students' academic and emotional needs and justify their instructional choices. The findings of Akyol (2025) likewise concern teachers' reported participation practices in early-childhood settings and do not establish whether comparable training would influence the moral justifications offered by primary-school pupils. The present findings also do not justify assuming that improving classroom climate would causally improve moral justification. Because the climate-by-resources interaction is sensitive both to the level of clustering and to the definition of the outcome, any intervention derived from these findings should be evaluated prospectively not only for its average effect but also for whether its benefits are distributed similarly across pupils with different levels of educational resources.

6. Limitations

First, the design is observational. Repeated classroom observations reduce dependence of the climate measure on any single session, but the exploratory moral-justification index was measured only once. Causal change cannot be inferred. Longitudinal research incorporating baseline moral judgment and repeated follow-up of classroom climate would be required to examine temporal change more directly.

Second, the contextual sample comprises only 24 classrooms and 14 schools. The three-level diagnostic model (pupils within classrooms within schools) converged, but the maximum-likelihood solution generated a boundary warning and a non-positive-definite Hessian. The Python solution, using Powell optimization, yielded the same boundary solution verified in IBM SPSS Statistics 26. The boundary warning is retained. Four of the 14 schools contribute only one sampled classroom, so the level-3 decomposition is too fragile for primary inference. The two-level classroom model is retained as primary, while 14-school CR2-Satterthwaite and wild-bootstrap analyses explicitly test sensitivity to shared school dependence. These school-clustered checks weaken the focal climate-by-resources interaction, which is no longer statistically distinguishable from zero once dependence among classrooms in the same school is taken into account. They also place the classroom-climate main effect's own school-clustered wild-bootstrap *p* value just above .05, meaning its significance under classroom clustering does not fully survive this more conservative check either. The manuscript does not describe either the main effect or the interaction as uniformly robust across clustering levels.

Third, one pedagogical inspector rated all 24 classrooms. The primary observer had 14 years of primary-school teaching experience and 14 years of pedagogical-inspection experience at the time of the study. Both inspectors involved in the four-class calibration had also completed the same five-day CLASS-focused continuing-education training within the inspectorate's professional-development program, and neither is described as developer-certified. The observed agreement was high, but the institutionally restricted four-class calibration remains too small to establish full-sample inter-observer reliability. The 44 dimension-by-class observations are nested within those four classrooms, and the bootstrap interval is necessarily coarse because only 35 distinct classroom-count compositions are possible. The calibration findings are not generalized to the remaining classrooms. The exceptionally high overall α may also reflect genuine covariance, redundancy, a halo effect, or a combination of these processes. Replication with several independent observers across a larger classroom sample is needed.

Fourth, K-3 and Upper Elementary versions were necessarily assigned by grade band. A formal climate $\times$ grade-band/CLASS-version term was added: it was positive under classroom clustering but attenuated under school clustering. Because version is perfectly determined by grade band, this analysis cannot distinguish developmental moderation from version-related differences and does not establish measurement invariance. In addition, Engagement was rated in Grades 1-3 classrooms although it is not part of the K-3 framework. Results involving this dimension, including the eleven-dimension comparisons, should be read with caution.

Fifth, the outcome is a closed-response, Kohlberg-inspired exploratory composite rather than a validated structural-stage diagnosis. The two-factor extraction was improper, one factor had only two items, and two situations had nearly null communalities. JM10 is consequently an exploratory composite and JM8 a post hoc sensitivity analysis. The absence of independent measures of reading comprehension, social desirability, baseline moral judgment, prior achievement, language proficiency, and cognitive ability leaves residual confounding and measurement uncertainty. In addition, response-option order was fixed within each situation for all pupils rather than randomized by pupil or class. Although the sequence differed across situations and was reordered before fielding to avoid a uniform developmental progression by position, response-position effects cannot be excluded.

Sixth, the resource index captures three study-related household resources but neither income, occupation, wealth, language practices, nor family support. Effectif_Classe represents analytic participants, not administrative class size, and teacher experience represents seniority rather than teaching qualification.

Seventh, class-level retention was 85.7%, but a pupil-level participation rate cannot be reconstructed from the analytic dataset or the available de-identified research files. Requiring complete parental consent at classroom level may have selected classrooms in which family contact was more successful. The school sector-setting cells are also unbalanced: private schools occur in urban and peri-urban settings, no rural private school was sampled, and the peri-urban private cell contains only two classrooms. Accordingly, both the pooled four-level configuration and the separate sector-setting sensitivity model are descriptive and cannot identify causal sector or territorial effects. The findings are restricted to the sampled schools in Meknès province and should not be generalized to Morocco as a whole, a caution consistent with the national territorial heterogeneity documented by Ibourk and Raoui (2025).

7. Conclusion

In this provincial sample, the exploratory moral-justification index varied substantially between classrooms and was positively associated with observed classroom climate and individual educational resources after adjustment for the measured covariates. The direction of the climate association was stable, but its inferential strength depended on whether uncertainty was clustered by 24 classrooms or by 14 schools.

The estimated climate-by-resources interaction was positive under classroom-clustered inference but was not statistically distinguishable from zero under school-clustered sensitivity analyses. It is best read as tentative evidence compatible with cumulative advantage, well short of a robust demonstration of differential resource conversion. A separate climate $\times$ grade-band/CLASS-version interaction showed the same sensitivity and cannot distinguish developmental moderation from instrument-version effects.

The study contributes on two fronts. On the substantive side, it situates observed classroom interaction within an unequal distribution of opportunities for moral socialization, examining it together with household educational resources and the sector-setting configuration of schooling within a single provincial multilevel design. On the methodological side, it shows how what we conclude about such settings depends on the level at which uncertainty is modeled and on the conventions by which moral justification is scored. The estimates are compatible with socially differentiated associations, but they stop short of identifying the mechanisms that produce them. Policies concerned with values education may therefore attend simultaneously to pedagogical dialogue and unequal educational resources, while prospective intervention studies are required before conclusions can be drawn about effectiveness or distributional consequences.

Acknowledgments: The authors thank the participating pupils, their families, and the staff of the participating schools for their cooperation throughout the research process. They also thank the teachers who administered the questionnaire, the second inspector who took part in the calibration observations, the five university professor-researchers who reviewed the observation harmonization material, and the expert who reviewed the mapping of the moral-justification response options. The use of AI-assisted tools during manuscript preparation is disclosed in the AI Usage and Assistance section.

AI usage and assistance: During manuscript preparation and revision, the authors used generative AI tools (ChatGPT, OpenAI; Claude, Anthropic) as editorial and technical support for language refinement, structural clarity, consistency checks, and limited assistance in checking code. All research questions, design decisions, data collection, analytical choices, statistical verification, interpretation of results, and conclusions remained the sole responsibility of the authors. All AI-assisted output was critically reviewed and verified against the data and original sources before inclusion. No AI tool was used to generate, simulate, or alter research data. The authors take full responsibility for the final manuscript.

Author contributions: Mustapha Fahli: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. Jamal Ksiksou: Methodology,

Validation, Resources, Data curation, Writing – review & editing, Visualization, Supervision, Project administration. Lhoussaine Maskour: Methodology, Validation, Resources, Writing – review & editing, Visualization, Supervision, Project administration. Abdel Fettah Zahidi: Methodology, Validation, Resources, Writing – review & editing, Visualization, Supervision, Project administration. All authors read and approved the final manuscript.

Data availability: The de-identified variable dictionary, analysis code, statistical syntax, instruments, coding documentation, aggregate results, and supplementary tables are provided as supplementary material. Individual-level data concern minors and are not deposited publicly. A de-identified analysis dataset may be considered for controlled access upon reasonable request, subject to the ethics approval, institutional authorization, and applicable confidentiality requirements.

Declaration of interest: The authors declare that they have no conflict of interest regarding the publication of this paper.

Ethical statement: The study was approved by the Research Committee of the Higher Institute of Nursing and Health Technical Professions (Protocol No. 34-2025, 1st July 2025). The approved field procedure permitted explicit oral parental consent, supplemented by written consent where parents were willing and able to sign. This procedure accommodated literacy-related barriers and parents who agreed to participation but declined a written signature. The absence of a signature was not treated as refusal when explicit oral agreement had been obtained. Pupil assent was obtained after parental consent and before administration. Institutional access was authorized by the relevant educational authorities and participating schools.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- Akar, C., & Uluçınar, U. (2023). A scale development study on democratic attitude among third and fourth grade students. *Journal of Pedagogical Research*, 7(4), 203–216. <https://doi.org/10.33902/JPR.202320926>
- Aktaş, F. N. (2023). A comparative study of teachers' values alignment strategies in the classroom: Face-to-face versus distance education. *Journal of Pedagogical Research*, 7(4), 339–355. <https://doi.org/10.33902/JPR.202323210>
- Akyol, T. (2025). The effect of participatory pedagogy and children's participation training on early childhood teachers' views on children's participation. *Journal of Pedagogical Research*, 9(4), 102–118. <https://doi.org/10.33902/JPR.202533781>
- Al-Makhalid, K. A. M. (2026). Supportive and controlling teacher authority, academic voice, and self-regulated learning among undergraduates at a Saudi public university. *Journal of Pedagogical Sociology and Psychology*, 8(2), e46143. <https://doi.org/10.33902/JPSP.202646143>
- Aldridge, J. M., & Blackstock, M. J. (2024). Assessing students' perceptions of school climate in primary schools. *Learning Environments Research*, 27, 579–602. <https://doi.org/10.1007/s10984-024-09492-2>
- Aldridge, J. M., & McLure, F. (2024). Investigating the influence of the school climate of church-based schools on students' moral identity development and hope for the future. *Learning Environments Research*, 27(3), 971–993. <https://doi.org/10.1007/s10984-024-09515-y>
- Alladatin, J., & Barjit, R. (2023). Analyse des perceptions des élèves sur le climat scolaire : étude de cas de deux écoles au Maroc [An analysis of pupils' perceptions of school climate: A case study of two schools in Morocco]. *Revue de l'École Supérieure de l'Éducation et de la Formation*, 1(1), 44–53. <https://revue-rise.uiz.ac.ma/index.php/rise/article/view/5>
- Baidoo, J. (2026). Public versus private: A comparative evaluation of Ghanaian teachers' perspectives of implementing inclusive education. *International Journal of Didactical Studies*, 7(1), 36684. <https://doi.org/10.33902/ijods.202636684>
- Ball, C. L., Smetana, J. G., Sturge-Apple, M. L., Suor, J. H., & Skibo, M. A. (2017). Moral development in context: Associations of neighborhood and maternal discipline with preschoolers' moral judgments. *Developmental Psychology*, 53(10), 1881–1894. <https://doi.org/10.1037/dev0000378>
- Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage

- samples. *Survey Methodology*, 28(2), 169–181.
- Ben El Bou, I. (2020). Didactique des valeurs dans le nouveau curriculum d'éducation islamique à l'école marocaine [The didactics of values in the new Islamic-education curriculum in Moroccan schools]. *Massalek de L'éducation et de la Formation*, 3(1), 20904. <https://doi.org/10.48403/IMIST.PRSM/massalek-v3i1.20904>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood Press.
- Bourdieu, P., & Passeron, J.-C. (1970). *La reproduction. Éléments pour une théorie du système d'enseignement* [Reproduction in education, society and culture]. Minuit.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427. <https://doi.org/10.1162/rest.90.3.414>
- Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, 47(4), 1178–1198. <https://doi.org/10.3758/s13428-014-0551-2>
- Cohen, J., McCabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1), 180–213.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120. <https://doi.org/10.1086/228943>
- Conseil supérieur de l'éducation, de la formation et de la recherche scientifique. (2015). *Vision stratégique de la réforme 2015–2030 : Pour une école de l'équité, de la qualité et de la promotion* [Strategic vision for reform 2015–2030: Toward a school of equity, quality, and advancement]. <https://www.csefrs.ma/fr/kits-de-presse/presentation-de-la-vision-strategique-de-la-reforme-2015-2030/?lang=fr>
- Conseil supérieur de l'éducation, de la formation et de la recherche scientifique. (2017). *L'éducation aux valeurs dans le système national d'éducation, de formation et de recherche scientifique* [Values education within the national education, training, and scientific research system]. <https://www.csefrs.ma/publications/%D8%A7%D9%84%D8%AA%D8%B1%D8%A8%D9%8A%D8%A9-%D8%B9%D9%84%D9%89-%D8%A7%D9%84%D9%82%D9%8A%D9%85/>
- Conseil supérieur de l'éducation, de la formation et de la recherche scientifique. (2024). *Égalité hommes-femmes dans et à travers l'éducation* [Gender equality in and through education]. <https://www.csefrs.ma/publications/egalite-hommes-femmes-dans-et-a-travers-leducation/?lang=fr>
- Coşkun, B., & Karadağ, E. (2023). The effect of student and school characteristics on TIMSS 2015 science and mathematics achievement: The case of Türkiye. *Journal of Pedagogical Research*, 7(1), 203–227. <https://doi.org/10.33902/JPR.202318875>
- Çelik Demirci, S., & Baki, A. (2023). Characterizing mathematical discourse according to teacher and student interactions: The core of mathematical discourse. *Journal of Pedagogical Research*, 7(4), 144–164. <https://doi.org/10.33902/JPR.202321852>
- Çimen, İ., Yücel, C., & Karadağ, E. (2024). Modelling the effectiveness of schools based on their equality of opportunities. *Journal of Pedagogical Research*, 8(2), 1–25. <https://doi.org/10.33902/JPR.202426078>
- Danişman, Ş. (2017). The effect of parent involvement on student achievement. In E. Karadağ (Ed.), *The factors effecting student achievement: Meta-analysis of empirical studies* (pp. 271–290). Springer. https://doi.org/10.1007/978-3-319-56083-0_17
- Danişman, Ş. (2026). Generative AI supported practicum experiences contribute to the pedagogical reasoning and professional identity development of preservice mathematics teachers. *E-Learning and Digital Media*. Advance online publication. <https://doi.org/10.1177/20427530261465688>
- Danişman, Ş., & Bilgiç, E. (2026). The fragile “math person”: Identities and existential experiences of mathematically gifted students in performance-driven schooling. *Gifted Education International*, 42(2), 283–301. <https://doi.org/10.1177/02614294261417905>
- DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, 32, 271–297. <https://doi.org/10.1146/annurev.soc.32.061604.123127>
- Dolean, D., Melby-Lervåg, M., Tincas, I., Damsa, C., & Lervåg, A. (2019). Achievement gap: Socioeconomic status affects reading development beyond language and cognition in children facing poverty. *Learning and Instruction*, 63, 101218. <https://doi.org/10.1016/j.learninstruc.2019.101218>
- Dorji, S., Lhamo, J., Tshering, N., Jamtsho, N., & Tenzin, K. (2025). Integrating Bhutanese values into ICT

- education: Perspectives of ICT teachers in Bhutan. *Sustainable and Pedagogical Technologies*, 1(1), e40960. <https://doi.org/10.33902/SPT.202540960>
- Durasa, H., Fitriani, S. N., Erlin, E., Kusuma, K. C. A., & Sriartha, I. P. (2024). Emile Durkheim's moral perspective on reducing bullying and violence in schools: A phenomenological study. *International Journal of Didactical Studies*, 5(2), 27615. <https://doi.org/10.33902/IJODS.202427615>
- Durkheim, É. (2012). *L'éducation morale* [Moral education]. Presses universitaires de France. (Original work published 1925).
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. <https://doi.org/10.1037/1082-989X.12.2.121>
- Gadernann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3), 1–13.
- Gibbs, J. C. (2019). *Moral development and reality: Beyond the theories of Kohlberg, Hoffman, and Haidt* (4th ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780190878214.001.0001>
- Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8: Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*, 57, 16–30. <https://doi.org/10.1016/j.stueduc.2016.09.004>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Halstead, J. M. (2007). Islamic values: A distinctive framework for moral education? *Journal of Moral Education*, 36(3), 283–296. <https://doi.org/10.1080/03057240701643056>
- Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, 113(4), 461–487. <https://doi.org/10.1086/669616>
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74(5), 1368–1378. <https://doi.org/10.1111/1467-8624.00612>
- Huang, F. L., Zhang, B., & Li, X. (2023). Using robust standard errors for the analysis of binary outcomes with a small number of clusters. *Journal of Research on Educational Effectiveness*, 16(2), 213–245. <https://doi.org/10.1080/19345747.2022.2100301>
- Ibourk, A., & El Aynaoui, K. (2026). Educational effectiveness in rural areas: What SDIs teach us about multigrade classes. *International Journal of Educational Development*, 120, 103482. <https://doi.org/10.1016/j.ijedudev.2025.103482>
- Ibourk, A., & Raoui, S. (2025). Territorial obstacles causing early school dropout in Morocco: Multivariate spatial analysis. *Heliyon*, 11(1), e41386. <https://doi.org/10.1016/j.heliyon.2024.e41386>
- Imbens, G. W., & Kolesár, M. (2016). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*, 98(4), 701–712. https://doi.org/10.1162/REST_a_00552
- Jacob, U. S., Busari, I. O., Pillay, J., & Oyefeso, E. O. (2024). Moral values and social skills of pupils with mild intellectual disability: Does indigenous language storytelling and gender play any role? *Journal of Pedagogical Sociology and Psychology*, 6(2), 17–29. <https://doi.org/10.33902/jpsp.202424850>
- Kahveci, H. (2023). The positive and negative effects of teacher attitudes and behaviors on student progress. *Journal of Pedagogical Research*, 7(1), 290–306. <https://doi.org/10.33902/JPR.202319128>
- Kobayashi, F. (2023). Recommendable positive-to-negative interaction ratio for teachers. *International Journal of Didactical Studies*, 4(1), 18046. <https://doi.org/10.33902/ijods.202318046>
- Kohlberg, L. (1981). *Essays on moral development (Vol. 1): The philosophy of moral development*. Harper & Row.
- Kohlberg, L. (1984). *The psychology of moral development: The nature and validity of moral stages* (Essays on moral development: Vol. 2). Harper & Row.
- Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research*, 18(2), 243–250. <https://doi.org/10.1086/209256>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), 96–106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>

- Li, X., Cao, T., Hu, R., Huang, K., & Guo, C. (2025). The impact of socioeconomic status on adolescent moral reasoning: Exploring a dual-pathway cognitive model. *Behavioral Sciences*, 15(10), 1347. <https://doi.org/10.3390/bs15101347>
- Li, X., Zhang, J., Zhao, X., & Guo, C. (2026). Objective-subjective socioeconomic status configurations, self-acceptance, and adolescent prosocial moral reasoning: A response surface analysis. *Behavioral Sciences*, 16(8), Article 21582440221085271. <https://doi.org/10.3390/bs16081367>
- Lim, L., & Chapman, E. (2022a). Development and preliminary validation of the Moral Reasoning Questionnaire for secondary school students. *SAGE Open*, 12(1), Article 18344909221087418. <https://doi.org/10.1177/21582440221085271>
- Lim, L., & Chapman, E. (2022b). Validation of the Moral Reasoning Questionnaire against Rasch measurement theory. *Journal of Pacific Rim Psychology*, 16. <https://doi.org/10.1177/18344909221087418>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92. <https://doi.org/10.1027/1614-2241.1.3.86>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.
- McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research*, 52(5), 661–670. <https://doi.org/10.1080/00273171.2017.1344538>
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386–422.
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14, 20170213. <https://doi.org/10.1098/rsif.2017.0213>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Njau, S. T., Onyango, G., & Itegi, F. (2026). Sustainable school infrastructure for learning continuity: Health, hygiene, and safety conditions of public secondary school buildings in Kenya. *Sustainable Education and Digital Transformation*, 1(1), e43788. <https://doi.org/10.33902/SED.T.202643788>
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge University Press.
- Nugroho, A. A., Sajidan, S., Suranto, S., & Masykuri, M. (2025). Enhancing students' argumentation skills through socio-scientific real-world inquiry: A quasi-experimental study in biological education. *Journal of Pedagogical Research*, 9(1), 337–362. <https://doi.org/10.33902/JPR.202531979>
- Organisation for Economic Co-operation and Development. (2023a). *PISA 2022 results (Volume I): The state of learning and equity in education*. <https://doi.org/10.1787/53f23881-en>
- Organisation for Economic Co-operation and Development. (2023b). *PISA 2022 results (Volumes I and II): Country note – Morocco*. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/morocco_10dfcb74-en.html
- Oussikoum, A. (2023). L'image dans les manuels scolaires au Maroc : moyen pédagogique ou source de stéréotypes [Images in Moroccan school textbooks: A pedagogical tool or a source of stereotypes]. *Massalek de l'éducation et de la formation*, 6, 131–145. <https://doi.org/10.48403/IMIST.PRSM/massalek-v6i0.37895>
- Parry, N. M., & Kumi, E. O. (2026). Participation in and implementation of continuous professional development learning into classroom practice: Challenges faced by teachers. *International Journal of Didactical Studies*, 7(2), e40662. <https://doi.org/10.33902/ijods.202640662>
- Peretz-Lange, R., Harvey, T., & Blake, P. R. (2022). Socioeconomic status predicts children's moral judgments of novel resource distributions. *Developmental Science*, 25(4), e13230. <https://doi.org/10.1111/desc.13230>
- Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). Free Press. (Original work published 1932).
- Pianta, R. C., Hamre, B. K., & Mintz, S. (2012). *Classroom assessment scoring system: Upper elementary manual*. Teachstone.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual: K-3*. Paul H. Brookes.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what is being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>

- Power, F. C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. Columbia University Press.
- Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4), 672–683. <https://doi.org/10.1080/07350015.2016.1247004>
- Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. Psychology Press.
- Rijnhart, J. J. M., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W. R., & Heymans, M. W. (2021). Mediation analysis methods used in observational research: A scoping review and recommendations. *BMC Medical Research Methodology*, 21, 226. <https://doi.org/10.1186/s12874-021-01426-3>
- Sancak Aydın, G., & Nalbant, A. (2026). Teacher-perpetrated emotional violence in the classroom: A study based on pre-service teachers' perceptions. *Journal of Pedagogical Sociology and Psychology*, 8(1), e42066. <https://doi.org/10.33902/JPSP.202642066>
- Sartiken, H. (2023). Evaluation of the problems experienced in multicultural schools from the perspective of school administrators. *Journal of Pedagogical Sociology and Psychology*, 5(3), 209–220. <https://doi.org/10.33902/JPSP.202323508>
- Schuitema, J., ten Dam, G., & Veugelers, W. (2008). Teaching strategies for moral education: A review. *Journal of Curriculum Studies*, 40(1), 69–89. <https://doi.org/10.1080/00220270701294210>
- Snarey, J. R. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin*, 97(2), 202–232. <https://doi.org/10.1037/0033-2909.97.2.202>
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83(3), 357–385. <https://doi.org/10.3102/0034654313483907>
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.
- Verma, P., Hearn, K., Zahran, R., & Alowais, A. (2022). The quality of private early childhood education and care centers: A Ras Al Khaimah-based case study. *Gulf Education and Social Policy Review*, 3(1), 50–78. <https://doi.org/10.18502/gespr.v3i1.11495>
- von Suchodoletz, A., Barza, L., & Larsen, R. A. A. (2020). Examination of teacher-child interactions in early childhood education programmes in the United Arab Emirates. *International Journal of Early Years Education*, 28(1), 6–21. <https://doi.org/10.1080/09669760.2019.1594720>
- Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, 28(2), 315–352. <https://doi.org/10.1007/s10648-015-9319-1>
- Wang, M.-T., Degol, J. L., Amemiya, J. L., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, 57, 100912. <https://doi.org/10.1016/j.dr.2020.100912>
- Wilke, J., Bauer, A. M., Gahtan, J., Bandaru, Y. S., & Bäker, N. (2026). Who deserves what – and why? A systematic review and meta-analysis on distributive justice preferences and low socioeconomic backgrounds in children and adolescents. *Current Research in Ecological and Social Psychology*, 10, 100277. <https://doi.org/10.1016/j.cresp.2026.100277>
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67(3), 189–205. <https://doi.org/10.1111/j.2044-8325.1994.tb00562.x>
- Yamamoto, Y., & Miyazaki, Y. (2025). Small sample methods in multilevel analysis. *The Journal of Experimental Education*, 93(1), 186–217. <https://doi.org/10.1080/00220973.2024.2329084>
- Zagkotas, V. (2026). Invisible spaces, visible opportunities: The use of school common areas in teaching practices in Greek schools. *International Journal of Didactical Studies*, 7(2), e39047. <https://doi.org/10.33902/ijods.202639047>
- Žižanović, S. (2025). Determinants of active learning from the perspective of students and teachers in Croatia. *Journal of Pedagogical Research*, 9(3), 291–311. <https://doi.org/10.33902/JPR.202533518>
- Zynuddin, S. N., Kenayathulla, H. B., & Sumintono, B. (2023). The relationship between school climate and students' non-cognitive skills: A systematic literature review. *Heliyon*, 9(4), e14773. <https://doi.org/10.1016/j.heliyon.2023.e14773>