



Quo Vadis Collaborative Learning In Primary Schools: A Systematic Review of Trends, Epistemological Shifts, and Research Gaps

¹Amri Saputra, ²Dini Cahyati, ³Shukri Bin Suparti

¹Sunan Kalijaga State Islamic University, Indonesia; ²Tanggamus College of Education, Indonesia; ³Sahabat II National School, Malaysia

e-mail: 125204011017@student.uin-suka.ac.id

Abstract

Research on collaborative learning in primary schools has grown substantially over the past decade, but its epistemological depth and global representation remain uneven. This article maps the developmental trajectory of collaborative learning research in primary schools from 2010 to 2026 and identifies the research gaps that remain open. Based on a systematic review of 198 Scopus-indexed articles following the PRISMA protocol, the analysis identifies five developmental phases that reflect shifts in epistemological focus rather than chronology alone: the exploration of dialogic technology (2010–2014), the structuring of cooperative models (2015–2019), pandemic-driven disruption and digitalization (2020–2022), the integration of social-emotional and artificial intelligence approaches (2023–2024), and an emerging orientation toward inclusivity and sustainability (2025–2026). The analysis further identifies three clusters of structural gaps: theoretical (the dominance of Western frameworks and the scarcity of locally grounded models), methodological (the predominance of short-term designs and limited analysis of interaction processes), and practical (the absence of cost-effectiveness studies and limited research in marginalized school contexts). In the Indonesian context, the most salient gap is the absence of a collaborative learning model that systematically incorporates the value of gotong royong (communal mutual assistance) as a pedagogical foundation. This gap suggests a concrete direction for future theory-building and empirical validation in Southeast Asia.

Keywords: Collaborative Learning, Systematic Literature Review, Research Gaps, Primary School, Gotong Royong



Licensed under [Creative Commons Attribution-ShareAlike 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/).

*Copyright (c) 2026 Amri Saputra, Dini Cahyati, Shukri Bin Suparti

Introduction

Collaborative learning has become one of the most actively researched approaches in 21st-century primary education. Evidence from 198 Scopus-indexed articles published between 2010 and 2026 consistently links collaborative learning to gains in academic achievement, social-emotional skills, and learning motivation among primary school students ([Karmina et al., 2024](#); [León-García et al., 2026](#); [Talkhan et al.,](#)

2025). Publication volume in this field has grown by more than 79% over the past six years. Such rapid growth raises a methodological concern common to fast-expanding fields: an increase in the number of studies does not necessarily correspond to an increase in conceptual or epistemological depth.

This concern motivates the central question of the present review: where has research on collaborative learning in primary education actually progressed, and where has it not? Many recent studies continue to test established models in familiar contexts using conventional designs. As a result, the literature has grown rich in volume, yet several mechanisms, boundary conditions, and developmental opportunities of collaborative learning remain insufficiently understood (Hardianti et al., 2025; Webb et al., 2021).

Several systematic reviews and meta-analyses have examined collaborative or cooperative learning in primary education, typically with a focus on intervention effectiveness and student outcomes. For example, Talkhan et al. (2025) conducted a meta-analysis of the effects of cooperative learning on mathematics achievement, while other reviews have similarly emphasized effect sizes within specific subject areas. Although these studies provide important evidence regarding the effectiveness of collaborative learning, they offer limited insight into how the field itself has evolved over time. Questions concerning the development of dominant research themes, methodological orientations, geographical distribution, and underexplored areas of inquiry remain insufficiently addressed. Consequently, a broader perspective is needed to understand the trajectory of knowledge production in this field.

The present review addresses this need by adopting a developmental and epistemological lens. Rather than treating the literature as a collection of interventions to be evaluated, it examines the literature as a body of knowledge to be mapped and interpreted. Using a corpus of 198 Scopus-indexed articles published between 2010 and 2026, the review identifies publication trends, evolving research themes, methodological tendencies, and emerging research opportunities. Three features distinguish this review from previous syntheses. First, it traces how dominant research questions and methodological orientations have evolved across five identifiable periods between 2010 and 2026, rather than aggregating effect sizes across studies. Second, it identifies structural gaps across theoretical, methodological, and practical dimensions. Third, it examines the geographic and conceptual representativeness of the field, leading to the

identification of a potentially important but underexplored research opportunity: the development of a collaborative learning model grounded in the Indonesian value of gotong royong. This combination of developmental mapping, structural gap analysis, and attention to local conceptual representation appears to have received limited attention in existing review studies of collaborative learning in primary education.

This gap carries particular significance in Indonesia, where the ongoing transition to the Kurikulum Merdeka (Merdeka Curriculum) emphasizes project-based learning and the value of gotong royong within the Profil Pelajar Pancasila (Pancasila Student Profile). Most conceptual frameworks used in collaborative learning research remain rooted in Western scholarly traditions. Although several scholars have suggested that gotong royong may share important conceptual affinities with the principle of positive interdependence in collaborative learning, this relationship has not yet been systematically theorized, operationalized, or empirically validated as the basis of a collaborative learning model in primary education (Muin et al., 2024; Susanti et al., 2023). Therefore, understanding the global development of collaborative learning research is important not only for identifying international research trends but also for uncovering opportunities to develop contextually relevant models that reflect local educational values and realities.

Research Method

This study employs a systematic literature review (SLR) with an analytical-critical orientation. The primary aim is not to synthesize intervention effectiveness, but to map the structure and developmental trajectory of knowledge on collaborative learning in primary schools. The review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 protocol as the framework for study selection and reporting (Liberati et al., 2009; Page et al., 2021).

Data Source and Search Strategy

The data source is the Scopus database, selected for its broad coverage and indexing quality standards (Mongeon & Paul-Hus, 2016). The search combined the **keywords** collaborative learning, cooperative learning, primary education, elementary school, and primary school in the title, abstract, and keyword fields, covering publications from 2010 to 2026

Eligibility Criteria

Eligibility criteria were defined a priori and applied in four sequential stages, following PRISMA reporting standards. Table 1 presents the inclusion and exclusion criteria for each stage, together with the number of records excluded and the documented reason for exclusion.

Table 1 Inclusion and Exclusion Criteria

Stage	Criterion	Inclusion	Exclusion	Records excluded (n)
1	Document type	Journal article	Conference paper, book chapter, review, editorial, other non-journal document	293
2	Language	English	Non-English full text	82
3	Source type and publication status	Journal source; final/in-press publication status	Non-journal source; article-in-press or not yet finalized	12
4	Accessibility	Open access	Subscription/paywalled access	187

Applying these criteria sequentially reduced the initial pool of 772 records to a final corpus of 198 articles, with 574 records excluded across the four stages for the documented reasons summarized in Table M1. The full selection process is presented as a PRISMA flow diagram in Figure 1.

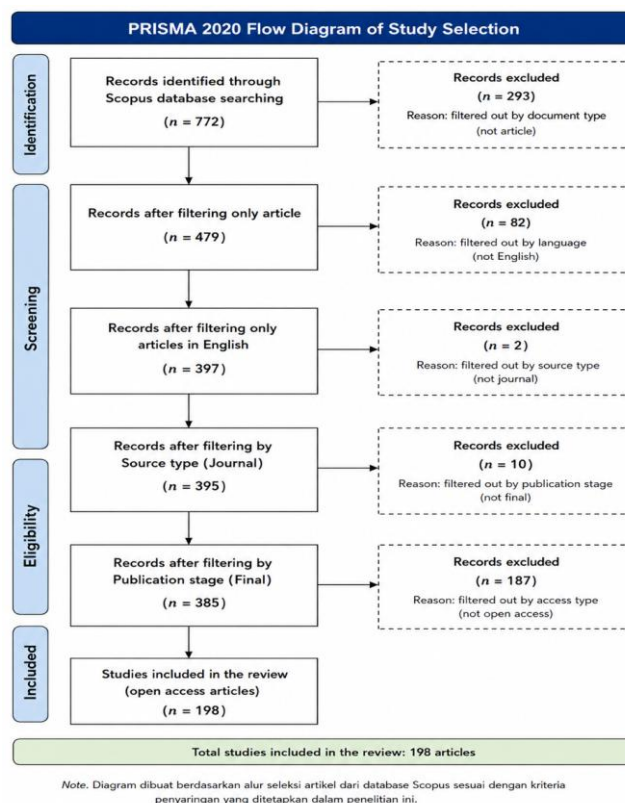


Figure 1 PRISMA Flowchart of Systematic Literature Selection

Coding Procedure and Inter-Rater Reliability

Each eligible article was coded according to five categories: bibliographic characteristics, research design, geographic context, thematic focus, and epistemological position within the field's developmental trajectory. Screening, filtering, and bibliometric coding were carried out and organized using R (RStudio), which supported the management of the 772-record search output through each elimination stage and the subsequent tabulation of coded variables.

Coding was conducted by a single reviewer (the first author). Consequently, inter-rater reliability statistics (e.g., Cohen's Kappa) were not computed, since such measures require at least two independent coders. We acknowledge this as a methodological limitation: single-coder review increases the risk of undetected coding inconsistency or subjective bias in category assignment, particularly for more interpretive categories such as epistemological position and gap classification. To partially mitigate this risk, coding categories were defined with explicit operational criteria prior to data extraction, and a random subset of coded entries was re-checked by the same coder after an interval to assess intra-rater consistency. This limitation, and the recommendation that future reviews in this area employ multiple coders with formal inter-rater agreement statistics, is also restated in the Discussion and Limitations sections.

Quality Appraisal

Beyond eligibility screening, each of the 198 included articles was further evaluated using a quality appraisal checklist developed for this review, rather than a published standardized instrument (e.g., CASP or MMAT). The checklist assessed each article against criteria commonly used in SLR quality appraisal: clarity of research aims, appropriateness of research design for the stated aims, adequacy of sample description, and clarity of reported findings. Articles were not excluded from the corpus on the basis of quality appraisal scores; instead, the appraisal results were used as an analytical input to interpret the strength of evidence underlying the developmental-phase and gap-cluster claims discussed in the Findings and Discussion section.

Risk of Selection Bias

We note at the outset that the open-access-only restriction applied at the fourth screening stage, while necessary for ensuring full-text accessibility for coding,

introduces a risk of selection bias. Open-access publication is not random with respect to research context: it can be correlated with institutional funding, publisher policy, and geographic origin, and 187 of the 574 excluded records (the largest single exclusion category) were removed for this reason alone. This restriction may therefore have disproportionately excluded studies from under-resourced contexts or non-open-access regional journals, including Indonesian-language journals. The implications of this restriction for the corpus's geographic representativeness are discussed further in the Discussion and Limitations sections.

Analytical Approach

Analysis proceeded in two layers. The first layer was descriptive bibliometric analysis, mapping the distribution of publications by time, geographic origin, methodological design, and thematic domain. The second layer was gap analysis, using a critical evaluative framework to identify underrepresented areas, unexamined assumptions, and unexploited research opportunities (Booth et al., 2016). The two layers were integrated through narrative synthesis to produce a knowledge map intended to be both descriptive and forward-looking, indicating directions for future research.

Results and Discussion

A Note on Analytical Procedure

Before presenting the findings, it is necessary to clarify how the analytical categories used throughout this section were derived, since the developmental phases, gap clusters, and the prominence given to the gotong royong model are interpretive products of the coding process described in the Research Method section, not categories given in advance by the data.

The five developmental phases presented below were not defined by splitting the 2010–2026 range into equal time bins. Phase boundaries were identified by examining, within each year-band of coded articles, which research questions, methodological approaches, and conceptual framings were most frequently represented, and then checking whether a noticeable shift in these three indicators coincided with a documented field-level event (for example, the COVID-19 pandemic, or the rollout of the Kurikulum Merdeka policy). A boundary between two phases was drawn where a shift in dominant research question or methodological orientation co-occurred with such an event, rather than at an arbitrary five-year mark. We acknowledge

that this procedure remains interpretive: it was performed by a single reviewer (as noted in the Research Method section) and was not independently cross-checked by a second coder. Readers should therefore treat the five-phase narrative as an analytical lens proposed by the authors for organizing the corpus, rather than as a classification that the 198 articles themselves report.

Similarly, the three gap clusters (theoretical, methodological, practical) were derived by grouping the limitations explicitly reported in the discussion sections of the coded articles, together with patterns of absence identified by the authors (e.g., the consistent absence of cost-effectiveness analyses). A limitation is classified as theoretical when it concerns the conceptual framework used to define or measure collaboration; methodological when it concerns research design, duration, or instrumentation; and practical when it concerns implementation feasibility or accessibility. This classification was also performed by a single reviewer and should be read as an interpretive synthesis rather than a coding scheme validated through formal inter-rater agreement.

Anatomy of the Corpus: Who Researches What and How

Before mapping the direction of this field, it is useful to first describe who has participated in this body of research and in what capacity. The distribution of the 198 analyzed articles, summarized in Table 2, reveals structural patterns that are not visible when reading studies individually but become meaningful when read collectively.

Table 2 Anatomy of the Collaborative Learning Research Corpus in Primary Schools (n = 198)

Characteristic	Category	n	(%)	Analytical Notes
Publication Period	2010–2014	15	7.6	Early phase: exploration of dialogic technology and CSCL
	2015–2019	26	13.1	Structuring cooperative models; dominance of quasi-experimental designs
	2020–2022	72	36.4	Acceleration: pandemic disruption and forced digitalization
	2023–2026	85	42.9	Maturation: inclusivity, AI, and educational equity
First Author Origin	Spain	59	29.8	Strong PE research tradition and primary school teacher education
	Indonesia	17	8.6	Increasing trend post-Merdeka Belajar; strategic momentum
	United States	16	8.1	Focus on CSCL, metacognition, and working memory
	Netherlands	11	5.6	Excellence in group dynamics and shared regulation research (SSRL)
	Others (>35 countries)	95	48.0	Broad global representation; but asymmetric Europe–Americas
Research Design	Quantitative/Experimental	98	49.5	Dominance of quasi-experimental designs; limited controls
	Qualitative	42	21.2	Focus on processes and teacher-student perspectives
	Mixed Methods	35	17.7	Captures CL complexity in a balanced manner
	SLR/Meta-Analysis	12	6.1	Opportunities for systematic synthesis remain wide open
	R&D/Development	11	5.6	Model and technology innovation beginning to emerge
Subject Area	Mathematics	38	19.2	Most frequently studied domain; Jigsaw and STAD dominant
	Language/Literacy	28	14.1	Peer tutoring as the dominant strategy
	Technology/Programming	25	12.6	Fastest growth 2023–2026; AI and metaverse CL
	Science	22	11.1	Strong inquiry-collaborative learning
	Cross-domain/General	85	43.0	High potential for cross-curricular CL integration

The most striking pattern in the temporal distribution is the explosion of publications in the past six years, with more than 79% of the 198 articles published between 2020 and 2026. This concentration corresponds with the documented disruption caused by the COVID-19 pandemic, which prompted a wave of research seeking to document, explain, and respond to that disruption (Du et al., 2022; Erbil & Kocabaş, 2020).

Geographically, Spain accounts for the largest share of first-author origin (29.8%), reflecting an established research tradition in collaborative physical education and primary teacher training. Indonesia's contribution (8.6%) is comparatively smaller in absolute terms, but notable given its alignment with the Kurikulum Merdeka policy momentum (Barokah et al., 2025; Nurdianah, 2025). Although more than 35 countries are represented, over 65% of the coded articles originate from Europe and North America (see Table 1, rows for Spain, United States, and Netherlands, combined with the residual distribution within the 'Others' category). This concentration indicates that the field's apparent global scope remains geographically uneven (Park & Nuntrakune, 2013).

From a methodological perspective, the dominance of quantitative and quasi-experimental designs (49.5%) reflects an emphasis on producing measurable evidence of effectiveness. The substantial share of mixed-methods studies (17.7%) suggests a parallel recognition that the collaboration process cannot be fully captured through outcome measures alone. SLR/meta-analysis and longitudinal designs remain comparatively rare (6.1%), indicating that designs best suited to capturing complexity are still underrepresented relative to the field's stated ambitions (Vu et al., 2024).

Five Developmental Phases: A Trajectory Narrative

As explained in the analytical procedure note above, the five phases below are an interpretive periodization based on shifts in dominant research questions, methodological orientation, and conceptual framing, cross-checked against documented field-level events. Table 3 presents a structured summary of these five phases, including their epistemological shifts, central research questions, representative findings, drivers of change, and phase-specific limitations. The narrative that follows interprets the patterns in Table 3, drawing on representative studies as illustrations rather than as exhaustive evidence for each phase.

Table 3 Trajectory of Collaborative Learning Research Development in Primary Schools (2010–2026)

Period	Epistemological Shift	Central Question	Representative Findings	Drivers of Shift	Phase Limitations
2010–2014	Exploration: Technology as dialogic mediator	Can technology (wikis, IWBs) support shared knowledge construction?	Technology opens space for meaningful dialogue, but is highly dependent on teacher design (Pifarré & Kleine Staarman, 2011; Warwick et al., 2010)	Early classroom technology proliferation; Web 2.0 optimism in education	Exploratory-descriptive; minimal control; documenting 'what' not 'why'
2015–2019	Accumulation: Model structuring and pursuit of effectiveness evidence	Which model is most effective for 21st-century competencies?	Jigsaw and STAD are consistent, but teacher facilitation is more decisive than the model itself (Fernández-Río & Suarez, 2016; Viilo et al., 2018)	Demand for evidence-based practice; 'what works' movement in global education policy	Uniformity in outcome measurement; absence of process analysis; quasi-experimental bias
2020–2022	Disruption: Fundamental redefinition of 'collaboration' in a crisis era	How does CL survive and adapt when physical presence is not possible?	E-CL is flexible, but the pandemic exposed long-hidden digital access inequalities (Du et al., 2022; Erbil & Kocabaş, 2020)	COVID-19 pandemic forced redefinition of 'co-presence' from physical to digital; wave of emergency pedagogical innovation	Digital-ready sample bias; short-term effects; equity neglected in design
2023–2024	Maturity: CL as a holistic academic-social-digital ecosystem	How do AI and emotional intelligence shape the future collaborative ecosystem?	CL transcends academic strategy; SEL, AI, and metaverse open new dimensions (Chen & Huang, 2024; Karmina et al., 2024; León-García et al., 2026)	Post-pandemic resurgence of SEL; advances in generative AI and immersive educational environments	Small samples; no large-scale AI-CL RCTs; AI-CL process mechanisms unmapped
2025–2026	Orientation: Inclusivity, sustainability, and educational justice as the core agenda	Who has not been reached by CL and why? How can CL become a catalyst for equitable transformation?	CL is meaningful for all students only when designed with contextual sensitivity (Liu et al., 2026; Paavilainen et al., 2026; Zhao et al., 2026)	SDG 4 agenda and global educational equity; recognition that 'ideal' models are not universally applicable	Longitudinal designs still very rare; locally-ecological models still neglected

The first phase (2010–2014) reflects enthusiasm about technology's potential as a mediator of collaboration. Pifarré & Kleine Staarman (2011) found that wikis can create meaningful dialogue spaces, while Minelli-de-Oliveira et al., (2014) found that although digital textbooks supported active learning, implementation still faced obstacles, particularly teacher readiness. This divergence is illustrative, not conclusive: it indicates that technology alone does not guarantee meaningful collaboration, a pattern that we interpret as the conceptual driver of the shift toward the next phase.

The second phase (2015–2019) shows a shift from questions about technology toward questions about model effectiveness. The Jigsaw and STAD models received considerable attention. However, several studies in this period found that teacher facilitation quality and interaction dynamics were more decisive than the choice of model itself (Veldman et al., 2020; Viilo et al., 2018), a finding we interpret as exposing a methodological blind spot rather than a fully resolved debate.

The third phase (2020–2022) coincides with the COVID-19 pandemic. The pandemic prompted a redefinition of collaboration from physical co-presence to digital interaction, producing a surge of research on e-collaborative learning and Computer-Supported Collaborative Learning (CSCL) (Erbil & Kocabaş, 2020). This period also surfaced digital access inequalities that earlier, technology-optimistic studies had not foregrounded.

The fourth phase (2023–2024) is marked by the integration of social-emotional learning (SEL) and emerging studies on AI and the metaverse as collaborative environments. León-García et al. (2026) found that social skills, group processing, and individual accountability significantly predicted prosocial behavior ($\beta = 0.194, 0.095$, and 0.091 , respectively), while Chen & Huang, (2024) examined Jigsaw within a metaverse environment for programming education. We interpret these studies as jointly indicating a broadening of how collaboration is conceptualized, from an academic strategy toward a multidimensional developmental process.

The fifth phase (2025–2026), still ongoing, shows an increasing orientation toward inclusivity and equity. The study by Paavilainen et al., (2026) examined shared regulation in multigrade classrooms, Zhao et al., (2026) on EdTech interventions in Sierra Leone, and Liu et al., (2026) examined self-regulated cooperative programming. Read together, these studies suggest that models developed under resource-rich conditions do not necessarily generalize to more diverse contexts (Barnes-Story et al., 2025).

The Gap Map: An Argument for Three Gap Clusters

If the developmental-phase narrative (summarized in Table 3) describes where research has gone, the gap map describes territories that remain underexamined. As noted in the analytical procedure note above, the three clusters below (theoretical, methodological, practical) are an interpretive grouping of limitations reported within the coded articles, combined with patterns of absence identified by the authors. Table 4 presents a comprehensive mapping of these gaps, including proposed research directions and their specific relevance to the Indonesian educational context.

Table 4 Research Gap Map and Future Scholarly Agenda

Gap Dimension	Current Research Condition	Proposed Research Direction	Gap Classification	Relevance for Indonesian Context
THEORETICAL GAPS: Limitations of Conceptual Frameworks				
Dominance of Western Theory	Dominant CL models derive from Western theory (Johnson & Johnson, 1994) ; local values not yet systematically integrated into CL design	Development of CL models grounded in local wisdom (gotong royong); validation of CL model adaptations in Indonesian primary school contexts	Theoretical Gap	Original novelty: CL–Gotong Royong model as Indonesia's scholarly contribution to global literature (Hardianti et al., 2025 ; Khanif, 2023)
Scarcity of Validated Local Models	Indonesian researchers often claim the relevance of gotong royong without operationalizing it in empirically validated designs	Construct validation studies of gotong royong in CL; development of culturally sensitive instruments for Indonesia's collectivist context	Theoretical Gap	Opportunity for genuinely distinct theoretical contribution from Western traditions; relevant to Merdeka Curriculum (Naparan & Alinsug, 2021)
METHODOLOGICAL GAPS: Research Design Limitations				
Short-term Studies (Temporal Gap)	Majority of studies <6 months; no cumulative data on CL impact across academic years; effectiveness claims risk overestimation	Longitudinal designs of 2–5 years; simultaneous tracking of social and academic impact; cumulative impact instruments	Methodological Gap	Studies on CL impact on character and achievement of Indonesian primary students across grade levels; relevant for evaluating Merdeka Curriculum impact (Vu et al., 2024)
Scarcity of Internal Process Analysis (The Black Box of Collaboration)	Most studies only measure input-output; mechanism of 'how collaboration produces understanding' remains unmapped	Analysis of turn-taking, discussion patterns, SSRL based on video recording; process-oriented approaches	Methodological Gap	Action research based on CL process analysis in Indonesian primary schools; microgenetic studies of group interaction dynamics (Hogenkamp et al., 2021 ; Webb et al., 2021)
Collective Regulation (SSRL)	SSRL rarely emerges spontaneously; SSRL instruments not yet cross-culturally validated; only studies from Northern Europe have examined it in depth	Development of cross-cultural SSRL instruments; interventions to build group regulation capacity	Methodological Gap	Validation of SSRL instruments in Indonesia's collectivist cultural context; relevant to characteristics of gotong royong as social regulation (Paavilainen et al., 2026)

PRACTICAL GAPS: Limitations of Relevance and Accessibility				
Cost-Effectiveness of Implementation	Almost no studies systematically analyze the cost-benefit ratio of CL implementation	Cost-effectiveness analysis; comparison of CL models in limited-budget contexts; evidence-based policy recommendations	Practical Gap	Cost analysis of CL implementation in public vs. private schools and remote-area schools in Indonesia; budget arguments for CL policy at Kemendikbudristek (Zhao et al., 2026)
CL in Marginal School Contexts	Majority of studies originate from schools in developed countries with adequate infrastructure; remote contexts, LMICs, and multigrade classrooms are marginalized	Locally-ecological research; adaptation of CL models for diverse sociocultural contexts; interventions in remote areas	Practical Gap	Development of contextual CL models for Indonesian primary schools in remote and underprivileged areas; relevant to educational equity policy (Barnes-Story et al., 2025 ; Park & Nuntrakune, 2013)
Differentiation and Cognitive Inclusion	Minimal research for students with special needs, low ability, and inclusive contexts; working memory as a moderator not yet systematically accounted for	CL + differentiated instruction; adaptive cognitive scaffolding; CL for students with ADHD, giftedness, and learning difficulties	Practical Gap	Inclusive CL in special and inclusive schools in Indonesia under Merdeka Curriculum; relevant to equitable education policy (Buchs et al., 2021 ; Du et al., 2022)

The theoretical gap cluster centers on the dominance of Western conceptual frameworks. Given that more than 65% of the coded articles originate from Europe and North America (see the caveat above regarding Table 2), there is a risk that locally produced findings are generalized as universal knowledge. Park & Nuntrakune, (2013) raised this concern in the Asian context; our reading of the corpus suggests it remains largely unaddressed. We note, further, that Indonesian researchers in the corpus frequently invoke gotong royong as conceptually relevant to collaborative learning, but, based on our coding, none of the Indonesian-authored articles in this corpus operationalized it as a validated learning design (Susanti et al., 2023).

The methodological gap cluster reflects three interrelated patterns: the predominance of short-duration studies, limited analysis of internal interaction processes, and limited cross-cultural validation of shared regulation (SSRL) instruments. Most coded studies measured outcomes shortly after interventions lasting a few weeks, which constrains claims about whether collaborative skills are durably internalized (Vu et al., 2024). Furthermore, Webb et al. (2021) showed that no single interaction pattern is universally most effective, yet most coded studies treated group interaction as an unexamined mechanism (Hogenkamp et al., 2021).

The practical gap cluster reflects the near-absence of cost-effectiveness studies and the underrepresentation of marginal school contexts in the corpus. Zhao et al., (2026) indicate that technology-based interventions can produce positive impact in low- and middle-income countries, but without accompanying cost data, such findings are difficult to translate into budget-level policy recommendations (Barnes-Story et al., 2025).

Toward a Collaborative Learning Model Based on Gotong Royong

This subsection proposes a conceptual agenda rather than reporting a finding derived from the 198-article corpus. None of the coded articles tested or validated a gotong royong-based collaborative learning model; the corpus instead motivates this agenda by demonstrating, through the theoretical gap cluster above, the absence of such a model in the existing literature. We present the comparison and development agenda below as the authors' proposed research direction, grounded in conceptual analysis and existing literature on gotong royong and cooperative learning theory, not as an empirical result.

Why Gotong Royong Differs from Western Cooperative Learning (Proposed Conceptual Comparison)

Gotong royong is not a literal translation of cooperative learning. It refers to a social practice with distinct values and logic. In the Johnson & Johnson, (1994) framework that dominates the literature, positive interdependence is treated as a procedural construct, designed through task structure, role assignment, and reward systems. Gotong royong, by contrast, is grounded in communal interdependence that is already part of everyday social life rather than externally engineered ([Susanti et al., 2023](#)). The comparison in Table 5 is the authors' conceptual proposal, intended to motivate the development agenda that follows; it is not a result derived from coding the 198 articles.

Table 5 Comparison of Western Cooperative Learning and the CL–Gotong Royong Model

Dimension of Comparison	Western Cooperative Learning (Johnson & Johnson, 1994)	CL–Gotong Royong Model (Proposed)
Fundamental Value Basis	Positive interdependence as a technical-procedural construct; rooted in an individualist tradition compromised by the need for cooperation	Interdependence rooted in communal values; gotong royong as a social ethos already internalized in Indonesian community life
Goal Orientation	Achievement of individual academic goals through group cooperation; success measured by individual outcomes	Collective progress as an intrinsic value; the process of togetherness is valued equally with results; success is shared success
Role of Social Relations	Social relations as an instrument for achieving academic goals; social skills formally trained	Social relations as a foundation before and beyond academic goals; gotong royong builds an organic learning community
Cultural Context	Designed in Western individualist cultural context; requires adaptation for collectivist contexts	Aligned with Indonesian collectivist values; potential for implementation resistance at the grassroots level is lower
Curricular Relevance	Generally relevant, but requires adjustment to the Gotong Royong dimension of the Pancasila Student Profile	Directly aligned with the Gotong Royong dimension; supports P5 and the Pancasila Student Profile ideals; compatible with the spirit of Merdeka Belajar
Evidence Status	Extensively validated in global literature; meta-analyses with strong effect sizes (Talkhan et al., 2025): Hedges' $g = 0.71$ – 0.79)	Conceptually promising; requires systematic empirical validation through Indonesian contextual research – this is the primary research opportunity

The most salient proposed difference concerns goal orientation. In Western cooperative learning, cooperation functions instrumentally toward individual academic

outcomes. In gotong royong, collective progress is treated as an end in itself. We propose that this difference, if operationalized in learning design, could plausibly generate collaborative dynamics not yet documented in the Western-theory-based literature reviewed here ([Jakavonytė-Staškuvienė et al., 2021](#); [Naparan & Alinsug, 2021](#)); this remains a hypothesis to be tested, not a documented outcome.

Conceptual Strengths and Relevance to the Merdeka Curriculum

We propose three reasons why gotong royong is a conceptually promising basis for future model development, while emphasizing that these are arguments for why the agenda is worth pursuing, not evidence that the model already works. First, a gotong royong-based model may face lower implementation resistance, since the underlying value is already familiar to students and families rather than introduced as an unfamiliar framework ([Susanti et al., 2023](#)). Second, it aligns conceptually with the Gotong Royong dimension of the Profil Pelajar Pancasila and the Proyek Penguatan Profil Pelajar Pancasila (P5); the corpus shows a rising number of Indonesian studies engaging with Kurikulum Merdeka ([Hardianti et al., 2025](#); [Karmina et al., 2024](#)), which we read as contextual support for the proposal's relevance, not as direct evidence for the model itself. Third, existing findings on the social-emotional dimension of collaboration provide indirect support for the plausibility of the proposal: León-García et al. (2026) found that social skills, group processing, and individual accountability significantly predict prosocial behavior, and [Ito et al., \(2022\)](#) found that participatory and cooperative experiences in primary school correlate with prosocial dispositions persisting into adulthood. These studies were conducted on Western cooperative learning models, not on gotong royong; we cite them as a plausibility argument for why a culturally congruent collaborative model is worth developing, not as validation of the proposed model itself.

Proposed Development Agenda for the CL-Gotong Royong Model

We propose that developing an empirically validated CL-Gotong Royong model would require three sequential stages, none of which have yet been undertaken within the scope of this review.

Stage 1 (conceptual framework development) would operationalize gotong royong principles, such as voluntariness, communal care, shared burdens, and the experience of togetherness, into observable learning-design elements. This stage would

require qualitative research (interviews, classroom observation, focus groups) to document how teachers and students naturally express gotong royong values in academic settings ([Susanti et al., 2023](#)).

Stage 2 (instrument development and validation) would build measurement tools for gotong royong-based collaboration quality, addressing the limitation that existing SSRL instruments were developed in different cultural contexts and may not capture collectivist dimensions of collaboration ([Paavilainen et al., 2026](#)).

Stage 3 (validity testing) would involve quasi-experimental studies with stronger controls, followed by longitudinal designs assessing whether effects persist over time. We propose that, to be recognized within the broader CL literature, the model's effect sizes would need to be benchmarked against existing meta-analytic estimates for related models, such as the Hedges' $g = 0.71$ to 0.79 range reported by [Talkhan et al., \(2025\)](#) for Jigsaw and STAD.

If pursued, this agenda would extend beyond local relevance, contributing to a broader literature that is beginning to recognize the importance of cultural sensitivity in collaborative learning research ([Barnes-Story et al., 2025](#); [Park & Nuntrakune, 2013](#)). We present this as a direction for future research, contingent on the empirical validation outlined above, rather than as a claim already substantiated by this review.

Implications

The mapping of trends and gaps in this study has direct implications for three communities: researchers, policymakers, and education practitioners.

For the research community, the most important implication is the need to move from confirmatory toward generative research. Too many studies begin with the question 'is model X more effective than conventional learning?'. A question that has been answered too many times with similar results. What is more urgently needed are questions that open new territory: how exactly does group interaction produce deeper understanding? Under what conditions does collaborative learning actually harm certain students? How does collective regulation form and how can it be facilitated? Questions like these require more complex designs, longer time frames, and the courage to produce ambiguous findings ([Hogenkamp et al., 2021](#); [Webb et al., 2021](#)),

For policymakers in Indonesia, the most pressing implication pertains to Kurikulum Merdeka. The spirit of project-based learning and strengthening of the Profil

Pelajar Pancasila through the Gotong Royong dimension creates a policy context highly conducive to collaborative learning. However, the finding that teacher competence, not model choice, is the strongest predictor of successful implementation ([Jakavonytė-Staškuvienė et al., 2021](#); [Veldman et al., 2020](#)) must be translated into concrete and sustained budget priorities for teacher professional development. Without this investment, the risk of reducing collaborative learning to meaningless group work routines becomes very high.

For Indonesian researchers in particular, this gap map reveals an opportunity that has never been greater. A contribution of this kind (a collaborative learning model rooted in gotong royong values, validated across diverse Indonesian primary school contexts from Sabang to Merauke) is not only locally relevant but also potentially provides a conceptual alternative not yet available in the Western frameworks that dominate the literature ([Hardianti et al., 2025](#)).

Conclusion

This review mapped the developmental trajectory of collaborative learning research in primary schools between 2010 and 2026, based on 198 Scopus-indexed articles. The corpus shows a field that has grown sharply in volume, with more than 79% of articles published since 2020, while remaining concentrated in a small number of countries and dominated by short-duration, quantitative designs. Read together with the gap analysis, this pattern converges on three structural gaps: a theoretical gap in the form of limited representation of non-Western conceptual frameworks, a methodological gap in the form of short study durations and limited process-level analysis, and a practical gap in the form of missing cost-effectiveness evidence and underrepresented marginal school contexts. We interpret the chronological shifts in dominant research questions across this corpus as marking five epistemological phases, and we read the geographic and theoretical concentration documented here as pointing to a specific opportunity: a collaborative learning model grounded in gotong royong. We present both the periodization and the gotong royong model as our interpretation and proposed research agenda, not as findings directly established by the 198 coded articles, since no article in this corpus tested or validated such a model.

This conclusion should be read alongside several limitations. The screening process excluded all non-English-language articles, a significant source of selection bias

that systematically excludes Indonesian-language research by Indonesian researchers, including work directly relevant to the gotong royong agenda proposed here. The restriction to open-access publications, the largest single exclusion category in this review, may have further excluded studies from under-resourced contexts, since open-access status is not randomly distributed across institutions or regions. All coding was also performed by a single reviewer without independent verification, increasing the risk of undetected classification inconsistency in the more interpretive categories, and the proposed CL-Gotong Royong model itself remains conceptual, its plausibility resting on indirect evidence from Western cooperative learning studies rather than direct empirical testing. Future work on this agenda should accordingly proceed concretely: observing how gotong royong principles are expressed in Indonesian classrooms to build an operational learning-design framework, developing and validating an instrument to measure gotong royong-based collaboration, and testing the resulting model over at least one academic year against an established model such as Jigsaw or STAD, with effect sizes reported on terms comparable to existing benchmarks. Pursued this way, the gotong royong agenda would move from a proposed direction to a tested contribution, addressing the gap this review identifies rather than the one it claims to have already filled.

References

- Barnes-Story, A. E., Robinette, J. J., Wawire, B. A., & Jere, B. (2025). Teaching and managing large classes: A global scoping review of methods with primary and secondary learners. *International Journal of Educational Development*, 119, Artikel 103450. <https://doi.org/10.1016/j.ijedudev.2025.103450>
- Barokah, A. I., Yuliana, L., & Raharja, S. (2025). Developing a community-based academic supervision model: A strategic reform for enhancing professional learning in Indonesian primary schools. *European Journal of Sustainable Development Research*, 9(4), Artikel em0282. <https://doi.org/10.29333/ejosdr/16862>
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2nd ed.). Sage.
- Buchs, C., Dumesnil, A., Chanal, J., & Butera, F. (2021). Dual effects of partner's competence: Resource interdependence in cooperative learning at elementary school. *Education Sciences*, 11(5), Artikel 210. <https://doi.org/10.3390/educsci11050210>
- Chen, C.-M., & Huang, M.-Y. (2024). Enhancing programming learning performance through a Jigsaw collaborative learning method in a metaverse virtual space. *International Journal of STEM Education*, 11(1), Artikel 12.

<https://doi.org/10.1186/s40594-024-00495-2>

- Du, X., Chen, C., & Lin, H. (2022). The impact of working memory capacity on collaborative learning in elementary school students. *Frontiers in Psychology*, 13, Artikel 1027523. <https://doi.org/10.3389/fpsyg.2022.1027523>
- Erbil, D. G., & Kocabaş, A. (2020). Flipping the 4th grade social studies course in a cooperative way: Effects on academic achievement and motivation. *Studies in Educational Evaluation*, 66, Artikel 100878. <https://doi.org/10.1016/j.stueduc.2020.100878>
- Fernández-Río, J., & Suarez, C. (2016). Feasibility and students' preliminary views on parkour in a group of primary school children. *Physical Education and Sport Pedagogy*, 21(3), 281–294. <https://doi.org/10.1080/17408989.2014.946008>
- Hardianti, N., Sudarnika, E., & Lukman, D. W. (2025). Implementation of collaborative learning to improve knowledge about rabies among students at elementary school in Dompu, West Nusa Tenggara Province, Indonesia. *International Journal of Agriculture and Biosciences*, 14(2), 301–307. <https://doi.org/10.47278/journal.ijab/2025.004>
- Hogenkamp, L., Van Dijk, A. M., & Eysink, T. H. S. (2021). Analyzing socially shared regulation of learning during cooperative learning and the role of equal contribution: A grounded theory approach. *Education Sciences*, 11(9), Artikel 512. <https://doi.org/10.3390/educsci11090512>
- Ito, T., Kubota, K., & Ohtake, F. (2022). Long-term consequences of the hidden curriculum on social preferences. *Japanese Economic Review*, 73(2), 269–297. <https://doi.org/10.1007/s42973-019-00033-8>
- Jakavonytė-Staškuvienė, D., Žemgulienė, A., & Sakadolskis, E. (2021). Cooperative learning issues in elementary education: A Lithuanian case study. *Journal of Education Culture and Society*, 12(1), 445–468. <https://doi.org/10.15503/jecs2021.1.445.468>
- Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive, and individualistic learning* (4th ed.). Allyn and Bacon.
- Karmina, S., Dyson, B., & Setyowati, L. (2024). Teachers' perspectives on implementing cooperative learning to promote social and emotional learning. *Cakrawala Pendidikan*, 43(2), 470–479. <https://doi.org/10.21831/cp.v43i2.68447>
- Khanif, A. (2023). Islamic religious education learning strategy for Alpha generation: A case study at Darul Qur'an School Elementary School Semarang City. *Ascarya: Journal of Islamic Science, Culture and Social Studies*, 3(1), 36–45. <https://doi.org/10.53754/iscs.v3i1.461>
- León-García, L.-F., Mendo-Lázaro, S., López-Ramos, V.-M., & León-del-Barco, B. (2026). Cooperative classrooms and prosocial behavior in primary education. *Frontiers in Psychology*, 17, Artikel 1725692. <https://doi.org/10.3389/fpsyg.2026.1725692>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*,

- 6(7), Artikel e1000100. <https://doi.org/10.1371/journal.pmed.1000100>
- Liu, C., Cui, Z., & Weng, X. (2026). Effects of a self-regulated cooperative programming approach on elementary students' programming knowledge, metacognitive awareness, and self-efficacy. *Journal of Computer Assisted Learning*, 42(1), Artikel e70167. <https://doi.org/10.1002/jcal.70167>
- Minelli-de-Oliveira, J., Camacho, M., & Gisbert, M. (2014). Exploring student and teacher perception of e-textbooks in a primary school. *Comunicar*, 21(42), 87–95. <https://doi.org/10.3916/C42-2014-08>
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Muin, A., Maisaroh, S., Ramadhan, W., & Sulaimon, J. T. (2024). Effectiveness of the P5 Program in the Merdeka Curriculum to increase the creativity and independence of Class V elementary school students. *Al Ibtida: Jurnal Pendidikan Guru MI*, 11(1), 1–16. <https://doi.org/10.24235/al.ibtida.snj.v11i1.17954>
- Naparan, G. B., & Alinsug, V. G. (2021). Classroom strategies of multigrade teachers. *Social Sciences & Humanities Open*, 3(1), Artikel 100109. <https://doi.org/10.1016/j.ssaho.2021.100109>
- Nurdianah, L. (2025). Implementasi modul ajar berbasis multiple intelligences pendekatan collaborative learning dalam Kurikulum Merdeka. *Jurnal Pendidikan Dasar Flobamorata (JPDF)*, 6(3), 434–441. <https://doi.org/10.51494/jpdf.v6i3.2285>
- Paavilainen, T., López-Pernas, S., Väisänen, S., Kontkanen, S., & Hirsto, L. (2026). Implementing learning design with learning analytics to scaffold the regulation of collaboration in primary education. *Smart Learning Environments*, 13(1), Artikel 4. <https://doi.org/10.1186/s40561-026-00453-1>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, Artikel n160. <https://doi.org/10.1136/bmj.n160>
- Park, J. Y., & Nuntrakune, T. (2013). A conceptual framework for the cultural integration of cooperative learning: A Thai primary mathematics education perspective. *Eurasia Journal of Mathematics, Science and Technology Education*, 9(3), 247–258. <https://doi.org/10.12973/eurasia.2013.933a>
- Pifarré, M., & Kleine Staarman, J. (2011). Wiki-supported collaborative learning in primary education: How a dialogic space is created for thinking together. *International Journal of Computer-Supported Collaborative Learning*, 6(2), 187–205. <https://doi.org/10.1007/s11412-011-9116-x>
- Susanti, A., Darmansyah, A., Tyas, D. N., Hidayat, R., Syahputri, D. O., Wulandari, S., & Rahmasari, A. (2023). The implementation of project for strengthening the profile of Pancasila students in the Independent Curriculum for elementary school students. *IJECA (International Journal of Education and Curriculum Application)*, 6(2), 113–122. <https://doi.org/10.31764/ijeca.v6i2.15474>

- Talkhan, E., Alhubaidah, S., Muthanna, A., & Qadhi, S. (2025). The effect of cooperative learning toward mathematics achievement of primary students: A systematic review using meta-analysis. *Social Sciences & Humanities Open*, 12, Artikel 102247. <https://doi.org/10.1016/j.ssaho.2025.102247>
- Veldman, M. A., Van Kuijk, M. F., Doolaard, S., & Bosker, R. J. (2020). The proof of the pudding is in the eating? Implementation of cooperative learning: Differences in teachers' attitudes and beliefs. *Teachers and Teaching: Theory and Practice*, 26(1), 103–117. <https://doi.org/10.1080/13540602.2020.1740197>
- Viilo, M., Seitamaa-Hakkarainen, P., & Hakkarainen, K. (2018). Long-term teacher orchestration of technology-mediated collaborative inquiry. *Scandinavian Journal of Educational Research*, 62(3), 407–432. <https://doi.org/10.1080/00313831.2016.1258665>
- Vu, K., Dao, V., DeJaeghere, J., & Glewwe, P. (2024). Collaborative learning in the Vietnam Escuela Nueva Model and students' learning behaviors: A mixed methods longitudinal study. *International Journal of Educational Development*, 106, Artikel 103017. <https://doi.org/10.1016/j.ijedudev.2024.103017>
- Warwick, P., Mercer, N., Kershner, R., & Staarman, J. K. (2010). In the mind and in the technology: The vicarious presence of the teacher in pupil's learning of science in collaborative group activity at the interactive whiteboard. *Computers & Education*, 55(1), 350–362. <https://doi.org/10.1016/j.compedu.2010.02.001>
- Webb, N. M., Ing, M., Burnheimer, E., Johnson, N. C., Franke, M. L., & Zimmerman, J. (2021). Is there a right way? Productive patterns of interaction during collaborative problem solving. *Education Sciences*, 11(5), Artikel 214. <https://doi.org/10.3390/educsci11050214>
- Zhao, A., Lurvink, A.-F., Klune, C., Hinks, J., Beoku-Betts, I., Choragudi, U., & Iyer, G. (2026). Assessing different implementation modalities of an EdTech intervention for lower-primary learners: A difference-in-difference study in Sierra Leone. *British Journal of Educational Technology*, 57(3), 775–796. <https://doi.org/10.1111/bjet.70056>