Jurnal Pustaka Indonesia (JPI)
2023, Vol. 3, No. 3, 247–257
https://doi.org/10.62159/jpi.v3i3.2849

OPEN ACCESS Research Article

Implementing the Think–Pair–Square (TPSq) Cooperative Learning Model to Improve the Mathematics Learning Outcomes at Islamic Junior High School, Pekanbaru

Tuti Syafrianti[1], Windayani[2], Muhammad Zuhri[3], Hamdan Thahir[4], Rahma Aulia[5], Ayu Kenanga[6], Nada Sukmana[7]
[1-7]Islamic Education, Institut Agama Islam Tafaqquh Fiddin Dumai, Dumai, Indonesia

Abstract
Low student participation and limited peer interaction can restrict mathematical understanding, particularly when instruction is dominated by teacher explanation. This classroom action research investigated whether the Think-Pair-Square (TPSq) cooperative learning technique could improve the learning process and mathematics achievement of 32 Grade 8 students (12 boys and 20 girls) at SMP Islam YLPI Pekanbaru. The intervention was implemented in two cycles, each comprising planning, action, observation, and reflection, with three instructional meetings followed by one cycle achievement test. Students worked in eight heterogeneous groups of four through individual thinking, paired explanation, and four-member comparison and consolidation. Data were obtained from open observation sheets and essay-based achievement tests and were analyzed descriptively using teacher-student activity patterns, individual development scores, group recognition, attainment of the minimum mastery criterion of 68, indicator-level mastery, mean scores, and score-frequency distributions. The findings showed progressively stronger implementation fidelity and student engagement across the two cycles. Classical mastery increased from 62.50% at baseline to 68.75% after Cycle I and 81.25% after Cycle II, while mean achievement rose from 78.50 to 79.13 and 83.16, respectively. Five of the eight groups attained the highest recognition category in Cycle II, compared with none in Cycle I. Indicator analysis showed strong mastery of defining and identifying prisms and pyramids and calculating surface area and volume, although only 50.00% mastered proportional changes in volume when edge lengths changed. These results indicate that TPSq can support mathematics achievement by combining individual accountability, peer explanation, and group-level consolidation. The practical implication is that teachers should use indicator-level evidence to refine prompts, provide targeted scaffolding for complex proportional reasoning, and align group rewards with individual improvement rather than group products alone.

ARTICLE HISTORY
Received : 27 Agustus 2023
Revised : 10 September 2023
Accepted : 20 November 2023

KEYWORDS
Classroom Action Research; Cooperative Learning; Mathematics Achievement; Prisms And Pyramids; Think-Pair-Square.

CORRESPONDING AUTHOR
***Tuti Syafrianti,** Islamic Education, Institut Agama Islam Tafaqquh Fiddin Dumai, Dumai, Indonesia. Email: tutisyafrianti@iaitfdumai.ac.id

Introduction

Matematika Mathematics education is expected to develop conceptual understanding, procedural fluency, reasoning, communication, and the capacity to solve unfamiliar problems. These outcomes are difficult to achieve when students mainly receive explanations and reproduce procedures without sustained opportunities to articulate, test, and revise their ideas. Evidence across science, technology, engineering, and mathematics indicates that active participation produces stronger achievement than uninterrupted exposition, while interactive engagement can also reduce achievement disparities when participation is meaningfully structured (Chi & Wylie, 2014; Freeman et al., 2014; Howe & Abedin, 2013; Theobald et al., 2020). In mathematics classrooms, discussion is especially important because students need to connect verbal, symbolic, diagrammatic, and spatial representations and justify why a procedure or relationship is valid rather than merely state an answer (Carbonneau et al., 2013; Mix et al., 2016; Rittle-Johnson & Star, 2007; Uttal et al., 2013).

Cooperative learning offers a theoretically grounded response to this instructional challenge. Social interdependence theory proposes that learning improves when students perceive that individual and group success are positively linked, while each member remains accountable for contributing to a shared goal. Effective cooperative learning therefore requires more than placing students in groups; it requires positive interdependence, individual accountability, promotive interaction, interpersonal skills, and group processing (Gillies, 2016; Johnson & Johnson, 2009; Kyndt et al., 2013; Slavin, 2014). A recent second-order meta-analysis likewise found a moderate overall effect of cooperative learning on academic achievement, higher-order thinking, and affective outcomes, reinforcing the need to examine not only whether group work is used but how its interactional architecture is designed (Güngör et al., 2026; Kyndt et al., 2013; Slavin, 2014).

The mathematics-specific evidence is broadly consistent with this theoretical position. Meta-analyses have reported positive effects of cooperative learning on mathematics achievement and attitudes across different grade levels and national contexts, although effect magnitudes vary according to implementation quality, task design, and comparison conditions (Capar & Tarim, 2015; Ridwan et al., 2022; Turgut & Turgut, 2018). More focused studies have shown that Think-Pair-Share structures, including variants combined with problem-based learning or higher-order-thinking questions, can improve mathematics achievement, critical thinking, and self-efficacy (Hardiyanto & Santoso, 2018; Zaki et al., 2024). Comparable benefits have also been reported in other STEM domains, suggesting that structured peer explanation can support learning when content is cognitively demanding and students are required to make their reasoning visible (Springer et al., 1999; Warfa, 2016; Zaki et al., 2024). The broader peer-learning literature also reports positive academic effects across ability levels and school phases, particularly when group size, task structure, and tutor-tutee responsibilities are clearly specified (Bowman-Perrott et al., 2013; Lou et al., 1996; Rohrbeck et al., 2003; Topping, 2005).

Think-Pair-Square (TPSq) is a structured extension of paired discussion. Students first formulate an individual response, then explain and compare their reasoning with a partner, and finally merge two pairs into a four-member group to evaluate alternatives and construct a shared solution. The sequence protects individual thinking time before social influence occurs, increases the number of students who must verbalize an idea, and creates a second comparison phase in which misconceptions and incomplete strategies can be challenged. Research on peer instruction, classroom dialogue, and comparison-based mathematics learning indicates that learning is strengthened when students explain answers, confront contrasting reasoning, and receive prompts that require justification rather than simple agreement (Crouch & Mazur, 2001; Murphy et al., 2009; Rittle-Johnson & Star, 2009; Smith et al., 2009; Star & Rittle-Johnson, 2009; Webb, 2009). From the Interactive-Constructive-Active-Passive perspective, TPSq is potentially powerful because it moves students from individual constructive activity to interactive co-construction, provided that the task and teacher prompts elicit substantive reasoning (Chi & Wylie, 2014; Howe & Abedin, 2013; Webb, 2009).

Nevertheless, group work does not automatically produce productive learning. Collaboration can be undermined by free riding, status differences, friendship patterns, weak collaborative skills, or excessive coordination demands (Barron, 2003; Janssen et al., 2010; Kirschner et al., 2009; Le et al., 2018). Reviews of cooperative learning emphasize that teachers must explicitly establish norms, monitor explanations, distribute participation, and connect group recognition to individual accountability (Dzemidzic Kristiansen et al., 2019; Gillies, 2016; Stanczak et al., 2022). Heterogeneous grouping and structured roles may also promote social inclusion and stronger peer relationships, but these benefits depend on sustained implementation rather than a single isolated lesson (Klang et al., 2020; Roseth et al., 2008; Van Ryzin & Roseth, 2018, 2019). Teacher learning is therefore central: implementation quality improves when teachers experience, analyze, and iteratively refine cooperative structures rather than treating them as fixed activity formats (Keramati & Gillies, 2024; Le et al., 2018; Yang, 2023).

The prism and pyramid unit provides a demanding context in which to examine TPSq. Students must coordinate properties of three-dimensional objects, nets, surface area, volume, and proportional change. Spatial training and concrete or visual representations can improve mathematics performance, but representations are most effective when learners connect them to underlying relations and receive sufficient guidance (Carbonneau et al., 2013; Mix et al., 2016; Uttal et al., 2013). Research on worked examples and collaborative problem solving further indicates that peer interaction is not uniformly beneficial: collaboration can support elaboration and comparison, yet it may impose additional cognitive load when tasks are poorly sequenced or when students lack prior knowledge (Kirschner et al., 2009; Retnowati et al., 2010, 2017). The gradual Think-Pair-Square sequence is therefore pedagogically relevant because it allows individual processing before increasingly complex social coordination.

The present study was conducted in Grade 8 at SMP Islam YLPI Pekanbaru, where preliminary observation indicated that instruction was largely teacher centered, student questioning was limited, and several students were disengaged during explanation. School records also showed that only 11 of 32 students had met the minimum mastery criterion on a preceding algebra topic. Although earlier studies and meta-analyses establish that cooperative learning and Think-Pair-Share can improve aggregate achievement, the literature gives less attention to how a pair-to-square structure develops across iterative classroom-action cycles and how improvement appears simultaneously in individual development scores, group recognition, classical mastery, indicator-level mastery, mean achievement, and score distributions (Capar & Tarim, 2015; Hardiyanto & Santoso, 2018; Ridwan et al., 2022; Turgut & Turgut, 2018; Zaki et al., 2024). This multi-level, cycle-based account constitutes the study's specific contribution. Accordingly, the study aimed to determine whether the Think-Pair-Square cooperative learning technique could improve the mathematics learning process and achievement of Grade 8 students on prisms and pyramids and to identify which learning indicators remained difficult after two action cycles.

Method

This study employed classroom action research designed to improve an ongoing instructional practice through repeated cycles of planning, action, observation, and reflection. The research was conducted in two cycles during the 2011/2012 academic year in Grade VIII1 at SMP Islam YLPI Pekanbaru. The participants were 32 students, comprising 12 boys and 20 girls. For the cooperative intervention, students were assigned to eight heterogeneous groups of four. Group composition was based on baseline achievement, with approximately 25% high-achieving, 50% medium-achieving, and 25% lower-achieving students distributed across groups. The original class order was ranked from the highest to the lowest baseline score and then used to create academically heterogeneous groups.

The instructional topic was prisms and pyramids. Each cycle consisted of three instructional meetings and one achievement-test meeting. The learning package included a syllabus, lesson plans, and student worksheets. During each lesson, the teacher introduced the learning objective and problem context, after which students completed the Think-Pair-Square sequence. In the think phase, each student independently interpreted the problem and drafted a solution. In the pair phase, students compared answers with one partner, explained procedures, and revised errors. In the square phase, two pairs joined to evaluate alternative strategies, agree on a justified group response, and prepare the result for whole-class discussion. Individual learning remained accountable because cycle tests were completed independently, while group recognition was based on the development of members' scores rather than on a single collective product.

Data were collected through open classroom-observation sheets and essay-based mathematics achievement tests. The observation sheets documented the alignment between planned and enacted instruction, teacher classroom management, time allocation, student attention, pair interaction, four-member discussion, and task completion. The achievement tests were administered after Cycle I and Cycle II. The baseline score, Cycle I test, and Cycle II test were used to calculate individual development scores, group-recognition categories, the number and percentage of students meeting the school minimum mastery criterion of 68, mean achievement, indicator-level mastery, and score-frequency distributions. The archival manuscript did not report psychometric evidence for the tests or formal inter-observer reliability; therefore, no validity or reliability coefficients are claimed in this revision.

Quantitative data were analyzed descriptively because the purpose of the action research was to document within-class instructional improvement rather than estimate a population-level causal effect. The action was considered successful when teacher and student activities became more consistent with the planned cooperative sequence and when the number and percentage of students meeting the mastery criterion increased from baseline to Cycle I and Cycle II. Indicator-level analysis was used to identify concepts that required additional instructional support. Qualitative observation notes were reviewed after each meeting to identify implementation problems and determine revisions for the next lesson or cycle. To preserve the evidentiary status of the original dataset, no unreported inferential tests, effect sizes, confidence intervals, ethics approval, or consent procedures were reconstructed. The absence of those details is addressed in the limitations.

Result And Discussion

Implementation Fidelity and Classroom Process across Cycles

Observation data indicated that the TPSq sequence was not fully established at the beginning of Cycle I. During the first meeting, the teacher had difficulty maintaining a calm transition into heterogeneous groups, instructions were not consistently audible, time allocation was uneven, and several students completed unrelated activities rather than working seriously with their partners. By the second meeting, classroom control and time management improved, students moved into groups more efficiently, and pair discussions became more focused. In Cycle II, the teacher implemented the sequence more consistently, distributed worksheets without substantial disruption, monitored pair and square discussions more actively, and completed the planned stages within the available time. Students also showed greater familiarity with the routine, stronger task persistence, and more sustained cooperation. The qualitative progression suggests that the learning gains reported below occurred alongside improved implementation fidelity rather than as an isolated test-score change.

Individual Development Scores and Group Recognition

Individual development scores in Cycle I were calculated from the difference between the baseline score and the Cycle I test, whereas Cycle II development scores were calculated from the difference between the Cycle I and Cycle II tests. Table 1 shows a shift from the lowest development categories toward scores of 20 and 30. The number of students receiving a development score of 5 declined from 10 to 4, while the number receiving a score of 20 increased from 7 to 12 and the number receiving a score of 30 increased from 10 to 13. Thus, 25 students (78.13%) were in the two highest development categories in Cycle II, compared with 17 students (53.13%) in Cycle I.

Table 1. Student Development Scores in Cycles I and II

Development score	Cycle I, n	Cycle I, %	Cycle II, n	Cycle II, %
5	10	31.25	4	12.50
10	5	15.63	3	9.38
20	7	21.88	12	37.50
30	10	31.25	13	40.63

The development scores were aggregated to determine group recognition. As shown in Table 2, no group attained the Super category in Cycle I. In Cycle II, Groups I, II, V, VI, and VII received Super recognition; Groups IV and VIII received Great recognition; and Group III received Good recognition. Although Group III declined from Great to Good, the class-level distribution changed substantially because five of eight groups moved into the highest category. The pattern indicates that improvement was distributed across several groups rather than concentrated in a single high-performing team.

Table 2. Group Recognition in Cycles I and II

Group	Cycle I		Cycle II	
	Group score	Recognition	Group score	Recognition
I	18.75	Great	25.00	Super
II	16.25	Good	25.00	Super
III	22.50	Great	16.25	Good
IV	10.00	Good	17.50	Great
V	17.50	Great	23.75	Super
VI	18.75	Great	23.75	Super
VII	16.25	Good	22.50	Super
VIII	15.00	Good	18.75	Great

Achievement of the Minimum Mastery Criterion

Table 3 presents the progression in mastery of the school minimum criterion of 68. At baseline, 20 of 32 students (62.50%) met the criterion. The number increased to 22 students (68.75%) after Cycle I and 26 students (81.25%) after Cycle II. Relative to baseline, the final cycle added six students to the mastery group and reduced the number below mastery from 12 to 6. The largest increase occurred between Cycle I and Cycle II, after classroom management, pacing, and student familiarity with the cooperative routine had improved

Table 3. Attainment of the Minimum Mastery Criterion

	Baseline	Cycle I test	Cycle II test
Number of students who met the Minimum Mastery Criterion (KKM)	20 students	22 students	26 students
Percentage of Students Who Met the Minimum Mastery Criterion (KKM)	62.50	68.75	81.25

Indicator-Level Mastery

The Cycle I test assessed definitions, elements, and nets of prisms and pyramids. As reported in Table 4, 27 students (84.38%) mastered the definition indicator and the identification of elements. Mastery was lower for constructing nets, with 22 students (68.75%) meeting the criterion. Observation of student work indicated that errors in the third indicator frequently involved incorrectly arranging or reversing faces rather than a complete absence of knowledge about nets.

Table 4. Indicator-Level Mastery on the Cycle I Test

No.	Learning indicator	Mastery, %
1	Define prisms and pyramids	84.38
2	Identify the elements of prisms and pyramids	84.38
3	Construct nets of prisms and pyramids	68.75

The Cycle II test addressed surface area, volume, and changes in volume. Table 5 shows that 30 students (93.75%) mastered the calculation of prism surface area and volume, while 27 students (84.38%) mastered the corresponding calculations for pyramids. The most difficult indicator was determining how the volume of a prism or pyramid changes when edge lengths change: only 16 students (50.00%) met the criterion. Student responses suggested difficulty interpreting the multiplicative relationship between linear scale and volume and maintaining accuracy across multi-step calculations.

Table 5. Indicator-Level Mastery on the Cycle II Test

No.	Learning indicator	Students mastering, n	Mastery, %
1	Calculate the surface area and volume of a prism	30	93.75
2	Calculate the surface area and volume of a pyramid	27	84.38
3	Determine changes in prism and pyramid volume when edge lengths change	16	50.00

Mean Achievement and Score Distribution

The mean score increased modestly from 78.50 at baseline to 79.13 after Cycle I and then more clearly to 83.16 after Cycle II (Table 6). The total increase from baseline to Cycle II was 4.66 points. The delayed acceleration is consistent with the observation data: students and the teacher required time to establish the cooperative routine before the strongest improvement became visible.

Table 6. Mean Scores at Baseline, Cycle I, and Cycle II

Statistic	Baseline	Cycle I test	Cycle II test
Mean score	78.50	79.13	83.16

The frequency distribution provides a more detailed view of change than the mean alone. As shown in Table 7, the number of students in the lowest interval (59-65) declined from 9 at baseline to 3 in Cycle II, whereas the number in the highest interval (94-100) increased from 5 to 13. The middle intervals remained populated, indicating that improvement was not uniform; nevertheless, the overall distribution shifted upward, with fewer very low scores and a markedly larger high-achievement group.

Table 7. Frequency Distribution of Mathematics Achievement Scores

Score interval	Baseline	Cycle I	Cycle II	Category
	Number of Students	Number of Students	Number of Students	
59-65	9	8	3	Low
66-72	4	4	6	Low
73-79	0	4	6	Moderate
80-86	7	2	3	Moderate
87-93	7	5	1	High
94-100	5	9	13	High

Discussion

The results show a coherent, although descriptive, pattern of instructional improvement. Classical mastery increased by 18.75 percentage points from baseline to Cycle II, the mean rose by 4.66 points, the lowest score interval contracted, and five groups attained the highest recognition category. These changes align with meta-analytic evidence that cooperative learning has a positive effect on mathematics achievement and with recent mathematics studies reporting stronger performance under Think-Pair-Share-based instruction than under conventional teaching (Capar & Tarim, 2015; Ridwan et al., 2022; Turgut & Turgut, 2018; Zaki et al., 2024). The finding also agrees with Hardiyanto and Santoso (2018), who found a TPS setting effective for achievement, critical thinking, and self-efficacy. The current study extends these comparisons by showing how progress appeared across several classroom indicators over two cycles rather than only in a post-intervention group mean.

A plausible mechanism is the sequencing of individual accountability and progressively wider explanation. In the think phase, students had to generate an initial representation or procedure; in the pair phase, they articulated and checked that reasoning; and in the square phase, they encountered additional alternatives and negotiated a shared response. This interpretation is consistent with evidence that peer discussion improves answers when students explain conceptual reasoning, not merely vote on an option (Crouch & Mazur, 2001; Smith et al., 2009). It is also compatible with the ICAP framework and with research showing that comparison of solution methods can promote conceptual knowledge and procedural flexibility (Chi & Wylie, 2014; Rittle-Johnson & Star, 2007, 2009; Star & Rittle-Johnson, 2009). The increase from Cycle I to Cycle II suggests that these mechanisms became more available after students learned the interaction routine and the teacher improved monitoring and pacing.

The improvement in group recognition provides additional evidence that the intervention did not depend solely on a few high achievers. Five heterogeneous groups reached the Super category in Cycle II, and 78.13% of students were in the two highest individual-development categories. This pattern is compatible with social interdependence theory, which predicts stronger achievement when group rewards are linked to individual contribution and improvement (Johnson & Johnson, 2009; Slavin, 2014). It also parallels research indicating that cooperative goal structures can support both achievement and positive peer relations among early adolescents (Roseth et al., 2008; Van Ryzin & Roseth, 2018, 2019). However, Group III's decline demonstrates that heterogeneity and rewards do not guarantee continuous growth for every team. Teacher attention to participation balance, explanation quality, and status dynamics remains necessary, as emphasized by Barron (2003), Le et al. (2018), and Webb (2009).

The indicator-level findings reveal that TPSq was more successful for definitions, identification of geometric elements, nets, and direct surface-area or volume calculations than for proportional changes in volume. This distinction is pedagogically important. The final indicator requires students to coordinate spatial visualization with the multiplicative effect of linear scaling on a three-dimensional measure, a relationship that is conceptually more demanding than substituting values into a familiar formula. Research shows that spatial skills are related to mathematical performance and can be trained, but representations and manipulatives are effective only when learners explicitly connect perceptual features to mathematical structure (Carbonneau et al., 2013; Mix et al., 2016; Uttal et al., 2013). The 50.00% mastery rate therefore indicates a need for additional cube-based models, dynamic scaling examples, comparison of correct and incorrect solutions, and prompts that ask students to predict and justify the exponent relationship before calculating.

The findings should not be interpreted as evidence that collaboration is always superior to individual learning. Retnowati et al. (2010, 2017) showed that collaborative work can be less effective than individual study when worked examples already provide substantial guidance and group coordination adds unnecessary cognitive load. Similarly, Kirschner et al. (2009) and Janssen et al. (2010) emphasize that the benefits of collaboration depend on task complexity, prior knowledge, and coordination demands. In the present study, the initial individual think phase may have reduced this risk by allowing students to construct a preliminary representation before negotiating with peers. The result supports a contingent interpretation: TPSq is most defensible when tasks require explanation and comparison, students have enough prerequisite knowledge to contribute, and the teacher actively scaffolds transitions between phases.

The study's novelty lies in its integrated cycle-based evidence chain. Rather than reporting only a final mean or mastery percentage, it connects changes in teacher-student activity, individual development scores, group awards, classical mastery, indicator-level attainment, average achievement, and score distribution. This configuration provides a practical diagnostic model for classroom improvement: teachers can determine not only whether average performance rises, but which concepts remain difficult, whether low performers are moving, and whether gains are distributed across groups. Theoretically, the findings illustrate how social interdependence and interactive knowledge construction can operate within a pair-to-square progression. Practically, the results imply that teachers should preserve individual thinking time, require explanation during pairing, use four-member comparison to surface alternative strategies, link group recognition to individual growth, and use indicator-level results to design reteaching. Wider adoption should be accompanied by professional learning that helps teachers establish cooperative norms and manage common obstacles such as unequal participation and free riding (Dzemidzic Kristiansen et al., 2019; Keramati & Gillies, 2024; Klang et al., 2020; Stanczak et al., 2022; Yang, 2023).

Several limitations constrain the strength and generalizability of the conclusions. The study involved one intact class of 32 students in a single school, used no control group or random assignment, and relied on descriptive comparisons; consequently, maturation, repeated testing, teacher adaptation, and other contextual factors cannot be separated from the intervention effect. The instructional period was short, no delayed retention test was administered, and the findings concern one geometry unit, limiting transfer to other mathematical topics. The original manuscript did not report test validity, reliability, inter-observer agreement, formal ethics approval, consent procedures, or detailed observation scores. The data were also collected in the 2011/2012 academic year, which limits temporal relevance to current curriculum conditions. Future research should use validated instruments, independent observers, comparison groups or stronger single-case designs, delayed assessments, and process measures such as coded student explanations. Larger multisite studies should also examine whether the pair-to-square sequence is equally effective for students with different prior-knowledge levels and whether targeted visual scaffolds improve proportional-volume reasoning.

Conclusion

The two-cycle classroom action research indicates that the Think-Pair-Square cooperative learning technique was associated with an improved mathematics learning process and higher achievement among Grade 8 students studying prisms and pyramids. Mastery increased from 62.50% at baseline to 68.75% after Cycle I and 81.25% after Cycle II, the mean score rose from 78.50 to 83.16, the proportion of students in the highest score interval increased, and five groups reached the Super recognition category in the final cycle. The strongest outcomes were found for defining geometric solids, identifying their elements, and calculating surface area and volume, whereas proportional changes in volume remained difficult. The findings support the use of a structured progression from individual thinking to paired explanation and four-member comparison, with group recognition tied to individual improvement. For practice, TPSq should be combined with explicit prompts, visual or concrete representations, and indicator-based reteaching for complex spatial and proportional concepts. Because the evidence comes from one small, non-comparative, historically dated classroom dataset, the conclusions should be treated as context-specific evidence of instructional improvement rather than a causal estimate of effectiveness.

Referensi

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39–55. https://doi.org/10.1080/02796015.2013.12087490

Capar, G., & Tarim, K. (2015). Efficacy of the cooperative learning method on mathematics achievement and attitude: A meta-analysis research. *Educational Sciences: Theory & Practice, 15*(2), 553–559. https://doi.org/10.12738/estp.2015.2.2098

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400. https://doi.org/10.1037/a0031084

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970–977. https://doi.org/10.1119/1.1374249

Dzemidzic Kristiansen, S., Burner, T., & Johnsen, B. H. (2019). Face-to-face promotive interaction leading to successful cooperative learning: A review study. *Cogent Education, 6*(1), Article 1674067. https://doi.org/10.1080/2331186X.2019.1674067

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gillies, R. M. (2016). Cooperative learning: Review of research and practice. *Australian Journal of Teacher Education, 41*(3), 39–54. https://doi.org/10.14221/ajte.2016v41n3.3

Güngör, F., Altunbaşak, İ., Aydın, A. T., Erdem, C., & Kaya, M. (2026). A second-order meta-analysis on the effects of cooperative learning on students' academic achievement, higher-order thinking, and affective

behaviors. *Current Psychology, 45*, Article 552. https://doi.org/10.1007/s12144-025-08943-0
Hardiyanto, W., & Santoso, R. H. (2018). Efektivitas PBL setting TTW dan TPS ditinjau dari prestasi belajar, berpikir kritis, dan *self-efficacy* siswa. *Jurnal Riset Pendidikan Matematika, 5*(1), 116–126. https://doi.org/10.21831/jrpm.v5i1.11127
Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education, 43*(3), 325–356. https://doi.org/10.1080/0305764X.2013.786024
Janssen, J., Kirschner, F., Erkens, G., Kirschner, P. A., & Paas, F. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational Psychology Review, 22*(2), 139–154. https://doi.org/10.1007/s10648-010-9131-x
Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057
Keramati, M. R., & Gillies, R. M. (2024). Teaching cooperative learning through cooperative learning environment: A qualitative follow-up of an experimental study. *Interactive Learning Environments, 32*(3), 879–891. https://doi.org/10.1080/10494820.2022.2100429
Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review, 21*(1), 31–42. https://doi.org/10.1007/s10648-008-9095-2
Klang, N., Olsson, I., Wilder, J., Lindqvist, G., Fohlin, N., & Nilholm, C. (2020). A cooperative learning intervention to promote social inclusion in heterogeneous classrooms. *Frontiers in Psychology, 11*, Article 586489. https://doi.org/10.3389/fpsyg.2020.586489
Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning: Do recent studies falsify or verify earlier findings? *Educational Research Review, 10*, 133–149. https://doi.org/10.1016/j.edurev.2013.02.002
Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education, 48*(1), 103–122. https://doi.org/10.1080/0305764X.2016.1259389
Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423–458. https://doi.org/10.3102/00346543066004423
Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General, 145*(9), 1206–1227. https://doi.org/10.1037/xge0000182
Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' high-level comprehension of text: A meta-analysis. *Journal of Educational Psychology, 101*(3), 740–764. https://doi.org/10.1037/a0015576
Retnowati, E., Ayres, P., & Sweller, J. (2010). Worked example effects in individual and group work settings. *Educational Psychology, 30*(3), 349–367. https://doi.org/10.1080/01443411003659960
Retnowati, E., Ayres, P., & Sweller, J. (2017). Can collaborative learning improve the effectiveness of worked examples in learning mathematics? *Journal of Educational Psychology, 109*(5), 666–679. https://doi.org/10.1037/edu0000167
Ridwan, M. R., Hadi, S., & Jailani, J. (2022). A meta-analysis study on the effectiveness of a cooperative learning model on vocational high school students' mathematics learning outcomes. *Participatory Educational Research, 9*(4), 396–421. https://doi.org/10.17275/per.22.97.9.4
Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561
Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. https://doi.org/10.1037/a0014224
Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95*(2), 240–257. https://doi.org/10.1037/0022-0663.95.2.240
Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer

relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223–246. https://doi.org/10.1037/0033-2909.134.2.223

Slavin, R. E. (2014). Cooperative learning and academic achievement: Why does groupwork work? *Anales de Psicología, 30*(3), 785–791. https://doi.org/10.6018/analesps.30.3.201201

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*(5910), 122–124. https://doi.org/10.1126/science.1165919

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51. https://doi.org/10.3102/00346543069001021

Stanczak, A., Darnon, C., Robert, A., Demolliens, M., Sanrey, C., Bressoux, P., Huguet, P., Buchs, C., & Butera, F. (2022). Do jigsaw classrooms improve learning outcomes? Five experiments and an internal meta-analysis. *Journal of Educational Psychology, 114*(6), 1461–1476. https://doi.org/10.1037/edu0000730

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology, 102*(4), 408–426. https://doi.org/10.1016/j.jecp.2008.11.004

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., II, Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., . . . Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences, 117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645. https://doi.org/10.1080/01443410500345172

Turgut, S., & Turgut, İ. G. (2018). The effects of cooperative learning on mathematics achievement in Turkey: A meta-analysis study. *International Journal of Instruction, 11*(3), 663–680. https://doi.org/10.12973/iji.2018.11345a

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402. https://doi.org/10.1037/a0028446

Van Ryzin, M. J., & Roseth, C. J. (2018). Cooperative learning in middle school: A means to improve peer relations and reduce victimization, bullying, and related outcomes. *Journal of Educational Psychology, 110*(8), 1192–1201. https://doi.org/10.1037/edu0000265

Van Ryzin, M. J., & Roseth, C. J. (2019). Effects of cooperative learning on peer relations, empathy, and bullying in middle school. *Aggressive Behavior, 45*(6), 643–651. https://doi.org/10.1002/ab.21858

Warfa, A.-R. M. (2016). Using cooperative learning to teach chemistry: A meta-analytic review. *Journal of Chemical Education, 93*(2), 248–255. https://doi.org/10.1021/acs.jchemed.5b00608

Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology, 79*(1), 1–28. https://doi.org/10.1348/000709908X380772

Yang, X. (2023). A historical review of collaborative learning and cooperative learning. *TechTrends, 67*(5), 718–728. https://doi.org/10.1007/s11528-022-00823-9

Zaki, A., Sahid, S., Nurhaliza, R., Naufal, M. A., Huda, M., & Hassan, M. N. (2024). Enhancing mathematical achievement through the Think-Pair-Share cooperative learning model with higher-order thinking skills questions. *Jurnal Riset Pendidikan Matematika, 11*(2), 106–117. https://doi.org/10.21831/jrpm.v11i2.76963