MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924
Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i3.12525

Improving University Students' Argumentative Text Writing Skills through the Collaborative Learning Model

Jafar[1], Rosary Iriany[2], Andi Muhdar[3]
[1]Universitas Negeri Makassar, Indonesia; [2]Universitas Pancasakti Makassar, Indonesia;
[3]Universitas Lamappapoleonro, Indonesia
jafar@unm.ac.id; rosaryiriany2401@gmail.com

Submitted:	Revised:	Accepted:	Published:
Sep 20, 2026	Oct 19, 2026	Oct 30, 2026	Nov 4, 2026

Abstract

Although collaborative learning has been widely examined in language education, evidence concerning argumentative writing in compulsory Indonesian language courses for non-language majors remains limited. This study examined changes in university students' argumentative writing performance following a collaborative learning intervention. A quantitative approach with a pre-experimental one-group pretest–posttest design was employed, involving 38 first-year Mathematics Education students selected through cluster random sampling from three classes. Data were collected using argumentative writing tests administered before and after the intervention. Scores were analyzed using descriptive statistics, the Shapiro–Wilk test, a paired-samples t test, normalized gain (N-gain), and Cohen's d_z. The mean score increased from 60.66 (SD = 7.67) to 88.00 (SD = 5.64), with a statistically significant difference, $t(37) = 18.41$, $p < .001$, and a large within-group effect size ($d_z = 2.99$). The mean N-gain was 0.685, classified as moderate, while the proportion of students meeting the mastery criterion increased from 5.3% to 97.4%. These findings indicate improved argumentative writing performance following the intervention and extend evidence on collaborative writing to a national-language general education context. The results are consistent with social-constructivist accounts emphasizing structured peer interaction in argument construction and offer practical guidance for lecturers designing collaborative writing tasks. Further research should employ controlled designs and examine individual dimensions

Mikailalsys is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

of argumentative writing to assess the intervention's contribution more rigorously.

Keywords: Argumentative Writing; Collaborative Learning; Indonesian Language Course; Peer Interaction; Higher Education

INTRODUCTION

Argumentative writing is one of the core academic literacies that university students are expected to master. Composing an argumentative text requires students to take a clear position on an issue, support that position with relevant evidence, connect claims and evidence through sound reasoning, and anticipate opposing views. Toulmin's (2003) model describes these demands in terms of claims, data, warrants, backing, qualifiers, and rebuttals, and it remains a widely used framework for judging the quality of written arguments. Because argumentation underlies academic essays, reports, and research articles, the ability to write argumentative texts is closely linked to students' success in higher education and to their later participation in public and professional discourse.

Empirical studies nevertheless show that many university students find argumentative writing difficult. Kleemola et al. (2022), analysing essays written by 196 first-year students at two Finnish universities, found that taking a clear and consistent position was a major challenge for novice students. Ghanbari and Salari (2022) reported that undergraduate essays stated claims (56%) more often than they supplied supporting data (41%), and that counterarguments and rebuttals were rare. In the Indonesian context, Ananda et al. (2024) found that students struggled to generate and organize ideas, formulate thesis statements, handle counterclaims, and integrate sources. These problems are not limited to students: Nakrowi and Mulyati (2021) showed that the argument quality of published articles in Indonesian language education was generally low and dominated by a simple claim–data structure. Together, these findings suggest that argumentative writing needs deliberate instructional support at the university level.

In Indonesia, the Indonesian language (Bahasa Indonesia) course is compulsory for undergraduate programmes under Law Number 12 of 2012 on Higher Education (Republik Indonesia, 2012), and academic writing, including argumentative text, is one of its central learning outcomes. For students in non-language programmes such as mathematics

education, this course is often the main formal opportunity to develop academic writing in the national language. The present study argues that when argumentative writing is taught mainly as an individual, product-oriented task, students have few opportunities to test their claims against the views of others, even though argumentation is inherently dialogic. Learning designs that make students articulate, justify, question, and revise arguments together are therefore likely to be better suited to developing argumentative writing than designs that rely on individual drafting alone.

Collaborative learning offers such a design. Social interdependence theory holds that when individuals share a common goal and depend on one another's contributions, they tend to exert more effort, use higher-level reasoning, and achieve more than when they work individually or competitively (Johnson & Johnson, 2009). In higher education, Scager et al. (2016) showed that positive interdependence emerges when students work on challenging, open tasks that genuinely require the contributions of every member. From a social-constructivist perspective, peer interaction allows learners to perform within their zone of proximal development, where they can accomplish with others what they cannot yet accomplish alone (Vygotsky, 1978). Collaborative cognitive load theory adds that complex tasks, such as composing an argument, can be handled more effectively when the cognitive load is distributed across the working memories of several group members (Kirschner et al., 2018). Meta-analytic evidence supports these theoretical expectations: collaboration in computer-supported settings improves knowledge gain and skill acquisition (Chen et al., 2018), collaborative writing benefits second language learners' writing (Elabdali, 2021), and collaborative learning is associated with greater engagement and learning performance among university students (Qureshi et al., 2023).

Several studies have examined collaborative approaches to argumentation and writing. Shinta and Filia (2020) found that collaborative learning increased the number of argument components and the argument level produced by 21 Indonesian undergraduates in a one-group pretest–posttest study. Parlindungan et al. (2023) reported significant gains in literacy and argumentation among 30 Indonesian university students after collaborative reasoning was implemented in online classes. Landrieu et al. (2024) showed that combining explicit instruction with collaborative writing improved secondary school students' argumentative writing and writing self-efficacy, while Wang et al. (2025) demonstrated that metacognitive instruction enhanced the effectiveness of collaborative writing for English as a foreign language (EFL) learners. Related work has shown that supported online peer

feedback improves the quality of argumentative essays (Noroozi et al., 2023), that a flipped classroom combined with collaborative learning enhances the writing skills of Indonesian university students (Kheryadi et al., 2024), and that project-based learning improves Bahasa Indonesia writing skills (Kurniati et al., 2024). Recent reviews also emphasize that the benefits of collaborative writing depend on how collaboration is designed and supported (Zheng et al., 2026), and the arrival of generative artificial intelligence has opened new forms of collaboration in argumentative writing classrooms (Su et al., 2023).

Despite this growing body of work, a clear research gap remains. Most studies on collaborative argumentative writing have been conducted in EFL classrooms or secondary schools, or have focused on online peer feedback and technology-mediated collaboration. Few studies have examined the collaborative learning model for argumentative text writing in Indonesian, the national language, within the compulsory general education course taken by students of non-language programmes. In addition, many classroom studies report only statistical significance, without indicating how large the improvement is relative to the room for improvement or in standardized terms. The novelty of this study therefore lies in (a) its context, namely first-year mathematics education students learning to write argumentative texts in the compulsory Indonesian language course, and (b) its reporting of the magnitude of improvement through normalized gain (Hake, 1998) and a standardized effect size (Lakens, 2013), in addition to hypothesis testing. Theoretically, the study is grounded in social interdependence theory, the social-constructivist view of learning, and Toulmin's model of argument.

Based on this rationale, this study aimed to examine the improvement in university students' ability to compose argumentative texts after learning through the collaborative learning model. Specifically, it sought to describe students' argumentative writing scores before and after the intervention, to test whether the difference between pretest and posttest scores was statistically significant, and to determine the magnitude of the improvement through N-Gain and effect size. The hypothesis tested was that the mean posttest score would be significantly higher than the mean pretest score.

METHODS

Type of Research

This study employed a quantitative approach. The quantitative approach was chosen because the study aimed to measure students' argumentative writing ability numerically and to test statistically whether this ability improved after the collaborative learning model was implemented. The study was classified as pre-experimental research, as the intervention was applied to a single intact group without a comparison group.

Research Design

The study used a one-group pretest–posttest design (Creswell & Creswell, 2018), which can be expressed as $O_1 - X - O_2$. O_1 is the pretest administered before the intervention, X is the instruction using the collaborative learning model, and O_2 is the posttest administered after the intervention. This design allowed the researcher to compare each student's argumentative writing ability before and after the intervention, which matched the aim of examining improvement within the same group of students.

The collaborative learning model was implemented over six meetings of 100 minutes each, held between the pretest and the posttest. The 38 students were organized into eight heterogeneous groups (six groups of five and two groups of four) based on their pretest scores, so that each group contained students with high, moderate, and low initial ability. Each meeting followed the same sequence of stages. First, the lecturer stated the learning objectives and explained the focus of the meeting, such as the structure of argumentative texts, the formulation of a thesis, the use of evidence, or the presentation of counterarguments. Second, each group analysed a model argumentative text to identify its thesis, arguments, evidence, and reaffirmation. Third, group members discussed an assigned issue, negotiated a shared position, gathered supporting evidence from credible sources, and anticipated opposing views. Fourth, each group drafted an argumentative text together, with every member responsible for at least one part of the text. Fifth, groups presented their drafts and gave one another feedback using a checklist based on the structure of argumentative texts. Sixth, each group revised its text based on peer and lecturer feedback, and the meeting closed with individual reflection on what had been learned. Throughout the process, the lecturer acted as a facilitator who monitored group work and provided guidance when groups encountered difficulties.

Participants and Sampling Technique

The population comprised all first-year students of the Mathematics Education study programme, 2026 cohort, who were enrolled in the Indonesian language course and distributed across three classes at Universitas Negeri Makassar, Indonesia. Because the students were already organized into intact classes, cluster random sampling was used: one of the three classes was randomly selected by lottery as the research sample. The selected class consisted of 38 students, all of whom took both the pretest and the posttest. The study was conducted in the odd semester of the 2026/2027 academic year. Before the study began, students were informed of its purpose and procedures, participation was voluntary, and the results had no negative consequences for their course grades. To protect participants' privacy, students were identified only by numerical codes (S1–S38).

Instruments and Data Collection

Data were collected using an argumentative writing test. In both the pretest and the posttest, students were asked to write an argumentative text individually, of approximately 400 to 600 words, within 90 minutes. The pretest prompt concerned the use of generative artificial intelligence by university students in completing academic assignments, and the posttest prompt concerned the obligation for university students to take part in community service; the two issues were chosen to be comparable in familiarity and complexity. The texts were scored using an analytic rubric consisting of five aspects: thesis or position statement (20 points), arguments and supporting evidence (30 points), reasoning and coherence (20 points), reaffirmation or conclusion (10 points), and language use, including diction, sentence structure, and spelling conventions (20 points), yielding a total score from 0 to 100. The rubric was developed from the learning outcomes of the course and from Toulmin's (2003) model of argument, and each aspect was accompanied by descriptors for each score level to support consistent scoring. Before use, the test prompts and the rubric were reviewed by lecturers of Indonesian language education to ensure that they covered the intended aspects of argumentative writing (content validity) and were clearly worded. The pretest was administered one week before the first collaborative learning session, and the posttest was administered one week after the final session.

Data Analysis Technique

Data were analysed in four stages using Python 3 with the SciPy statistical library. First, descriptive statistics (mean, standard deviation, median, variance, minimum, maximum,

and range) were calculated for the pretest and posttest scores, and the scores were classified into five ability categories as shown in Table 1. A score of 75 was used as the individual mastery criterion. Second, the normality of the pretest scores, posttest scores, and paired differences was tested using the Shapiro–Wilk test at a significance level of .05. Third, because the normality assumption was met, the hypothesis was tested using the paired-samples *t*-test at $\alpha = .05$.

Table 1. Categories of Argumentative Writing Ability

No.	Score Interval	Category
1	90–100	Very high
2	80–89	High
3	65–79	Moderate
4	55–64	Low
5	0–54	Very low

Fourth, the magnitude of improvement was determined using the normalized gain (N-Gain) proposed by Hake (1998), calculated with Equation (1), where S_{post} is the posttest score, S_{pre} is the pretest score, and S_{max} is the maximum possible score (100). N-Gain values were interpreted as high ($g \geq 0.70$), moderate ($0.30 \leq g < 0.70$), or low ($g < 0.30$).

$$N\text{-}Gain = \frac{S_{post} - S_{pre}}{S_{max} - S_{pre}} \quad (1)$$

In addition, the standardized effect size for the paired design was calculated as Cohen's d_z using Equation (2), where M_D is the mean of the paired differences and SD_D is the standard deviation of those differences (Lakens, 2013). Values of 0.20, 0.50, and 0.80 were interpreted as small, medium, and large effects, respectively.

$$d_z = \frac{M_D}{SD_D} \quad (2)$$

RESULTS

Descriptive Statistics of Argumentative Writing Scores

The descriptive statistics of students' argumentative writing scores before and after the implementation of the collaborative learning model are presented in Table 2.

Table 2. Descriptive Statistics of Pretest and Posttest Scores (N = 38)

Statistic	Pretest	Posttest
Mean	60.66	88.00
Standard deviation	7.67	5.64
Median	59.00	87.50
Variance	58.88	31.78
Minimum	48	74
Maximum	76	99
Range	28	25

As shown in Table 2, the mean score increased from 60.66 (*SD* = 7.67) in the pretest to 88.00 (*SD* = 5.64) in the posttest, a mean increase of 27.34 points. The median rose from 59.00 to 87.50, the minimum score from 48 to 74, and the maximum score from 76 to 99. The standard deviation, variance, and range were all smaller in the posttest, indicating that students' scores became more homogeneous after the intervention. Every student obtained a higher posttest score than pretest score, with individual increases ranging from 10 to 46 points (median = 28.5). The comparison of the main descriptive statistics is visualized in Figure 1.

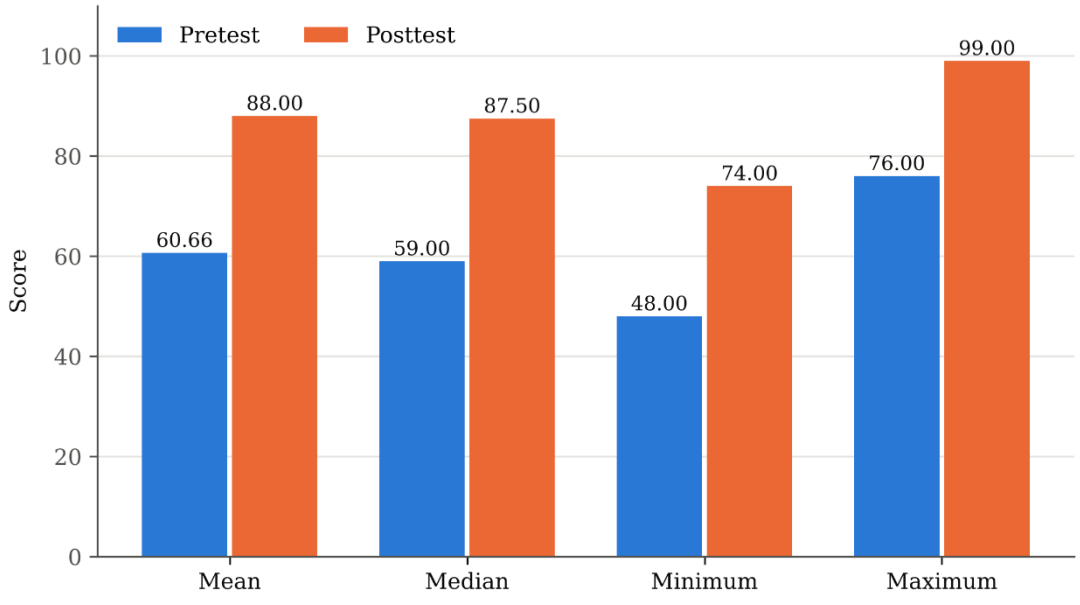


Figure 1. Descriptive Statistics of Students' Argumentative Writing Scores in the Pretest and Posttest

Distribution of Ability Categories and Mastery

The distribution of students across ability categories is presented in Table 3. In the pretest, all students were in the moderate category or below: 9 students (23.7%) were in the very low category, 15 students (39.5%) in the low category, and 14 students (36.8%) in the moderate category. In the posttest, no student remained in the low or very low categories; 19 students (50.0%) reached the high category, 17 students (44.7%) the very high category, and 2 students (5.3%) remained in the moderate category.

Table 3. Distribution of Students across Ability Categories

Category	Score Interval	Pretest f	Pretest %	Posttest f	Posttest %
Very high	90–100	0	0.0	17	44.7
High	80–89	0	0.0	19	50.0
Moderate	65–79	14	36.8	2	5.3
Low	55–64	15	39.5	0	0.0
Very low	0–54	9	23.7	0	0.0
Total		38	100.0	38	100.0

Using the mastery criterion of 75, only 2 students (5.3%) achieved mastery in the pretest, whereas 37 students (97.4%) did so in the posttest.

Assumption Testing and Hypothesis Testing

The results of the Shapiro–Wilk normality test are presented in Table 4. The significance values for the pretest scores (p = .153), the posttest scores (p = .718), and the paired differences (p = .506) were all greater than .05, indicating that the data were normally distributed. The assumption for the paired-samples t-test was therefore met.

Table 4. Results of the Shapiro–Wilk Normality Test

Data	Statistic (W)	df	Sig. (p)	Decision
Pretest	0.957	38	.153	Normal
Posttest	0.980	38	.718	Normal
Difference (post – pre)	0.974	38	.506	Normal

The results of the paired-samples t-test are presented in Table 5. The mean difference between posttest and pretest scores was 27.34 points (SD = 9.16), with a 95% confidence interval of 24.33 to 30.35. The difference was statistically significant, $t(37) = 18.41$, $p < .001$. The null hypothesis was therefore rejected, meaning that students' argumentative writing scores after learning through the collaborative learning model were significantly higher than their scores before the intervention.

Table 5. Results of the Paired-Samples t-Test

Pair	Mean Diff.	SD	95% CI	t	df	Sig. (2-tailed)
Posttest – Pretest	27.34	9.16	24.33 – 30.35	18.41	37	< .001

Magnitude of Improvement

The magnitude of improvement was analysed using N-Gain and effect size, as summarized in Table 6. The mean N-Gain was 0.685 (SD = 0.159), which falls into the moderate category, with individual values ranging from 0.357 to 0.970. Eighteen students (47.4%) obtained a high N-Gain, 20 students (52.6%) obtained a moderate N-Gain, and no student obtained a low N-Gain. The effect size was d_z = 2.99, far exceeding the 0.80 threshold for a large effect.

Table 6. N-Gain Distribution and Effect Size

Indicator	Value	Category
Mean N-Gain (SD)	0.685 (0.159)	Moderate
Minimum – maximum N-Gain	0.357 – 0.970	–
Students with high N-Gain (g ≥ 0.70)	18 (47.4%)	–
Students with moderate N-Gain (0.30 ≤ g < 0.70)	20 (52.6%)	–
Students with low N-Gain (g < 0.30)	0 (0.0%)	–
Effect size (Cohen's d_z)	2.99	Large

Divergent Findings and Anomalies

Although all students improved, the size of the improvement was not uniform. One student (S5) scored 74 in the posttest and thus remained just below the mastery criterion, despite increasing by 17 points from the pretest. The smallest N-Gain values were obtained by S38 (pretest 72, posttest 82, g = 0.357), S6 (pretest 67, posttest 79, g = 0.364), and S5 (pretest 57, posttest 74, g = 0.395), all of which were in the moderate category. Conversely, the largest N-Gain values were obtained by S31 (pretest 67, posttest 99, g = 0.970) and S23 (pretest 59, posttest 97, g = 0.927).

A further pattern concerned the relationship between initial ability and improvement. Pretest scores were strongly and negatively correlated with raw score gains (r = −.79, p < .001) and weakly but significantly negatively correlated with N-Gain (r = −.32, p = .048), whereas the correlation between pretest and posttest scores was negligible and not significant (r = .08, p = .639). In other words, students who started with lower scores tended to improve more, and students' posttest standing was largely unrelated to their initial standing.

DISCUSSION

Analysis and Interpretation of Findings

This study found that students' argumentative writing scores improved significantly after learning through the collaborative learning model, with a mean increase of 27.34 points, a moderate mean N-Gain of 0.685, and a large effect size. These results answer the research objective and support the hypothesis that posttest scores would be significantly higher than pretest scores. Before the intervention, most students (63.2%) were in the low or very low categories and only 5.3% met the mastery criterion, which reflects the difficulties with argumentative writing reported in earlier studies. After the intervention, 94.7% of students were in the high or very high categories and 97.4% met the mastery criterion. The reduction in standard deviation and range further indicates that the collaborative learning model not only raised the average performance but also narrowed the gap between students.

These improvements can be interpreted in terms of the processes that collaborative learning makes possible. Writing an argumentative text together requires group members to agree on a position, to justify it to one another, and to decide which evidence is strong enough to include. In this process, implicit reasoning becomes explicit, weak claims are questioned by peers, and possible counterarguments are raised before the text is written. These activities correspond to the elements of Toulmin's (2003) model that novice writers most often neglect, namely the warrant that links evidence to a claim and the rebuttal that addresses opposing views. The finding that students with lower initial scores tended to gain more is consistent with the idea that peer interaction allows less proficient students to work within their zone of proximal development with the support of more capable peers (Vygotsky, 1978). At the same time, the negligible correlation between pretest and posttest scores indicates that, after the intervention, students' performance was no longer determined mainly by where they started. This interpretation should be treated with caution, however, because larger gains among initially low-scoring students may also partly reflect regression toward the mean, which cannot be ruled out in a single-group design.

Comparison with Theories and Previous Literature

The findings are consistent with social interdependence theory, which predicts that shared goals and mutual dependence promote higher-level reasoning and achievement (Johnson & Johnson, 2009). Argumentative writing tasks are well suited to creating positive interdependence because they are challenging and open-ended, conditions that Scager et al.

(2016) identified as conducive to genuine collaboration in higher education. The findings also fit collaborative cognitive load theory: composing an argument requires students to manage content, structure, evidence, and language simultaneously, and distributing this load among group members can make the task more manageable (Kirschner et al., 2018).

The results are in line with previous empirical studies. Shinta and Filia (2020) and Parlindungan et al. (2023) both used one-group pretest–posttest designs with Indonesian university students and reported significant improvements in argumentation after collaborative learning and collaborative reasoning, respectively. The present study extends these findings from English as a foreign language to argumentative writing in Indonesian and from language or communication programmes to mathematics education students. The results are also consistent with Landrieu et al. (2024), who found that collaborative writing combined with explicit instruction improved argumentative writing, and with Kheryadi et al. (2024) and Kurniati et al. (2024), who reported improvements in the writing skills of Indonesian university students through collaborative and project-based approaches. At a broader level, the positive effect agrees with meta-analytic evidence that collaboration enhances skill acquisition (Chen et al., 2018) and that collaborative writing benefits learners' writing development (Elabdali, 2021).

The effect size in this study (d_z = 2.99) is, however, much larger than the average effects reported in meta-analyses, such as the effect of 0.64 on skill acquisition in computer-supported collaborative learning (Chen et al., 2018). This difference should not be read as evidence that the collaborative learning model is several times more effective than reported elsewhere. First, d_z is calculated from the standard deviation of paired differences, which is typically smaller than the standard deviation of scores, so it is not directly comparable to between-group effect sizes (Lakens, 2013). Second, a one-group design without a comparison group cannot separate the effect of the intervention from other influences such as practice with the test format, the regular content of the course, and students' general adjustment to university writing. Third, the moderate mean N-Gain (0.685), just below the high category, offers a more conservative picture of the improvement relative to the room available for growth. Wang et al. (2025) and Zheng et al. (2026) further suggest that the benefits of collaborative writing depend on how collaboration is supported, for example through metacognitive guidance, which may explain why some students in this study, such as S5, S6, and S38, gained less than their peers.

Research Implications

Theoretically, the findings support social-constructivist and social interdependence accounts of argumentative writing by showing that structured peer interaction is associated with substantial improvement in students' argumentative texts in a national-language general education course. The findings also show the value of reporting N-Gain and effect sizes together with significance tests, as these indicators give a more complete picture of the size and practical meaning of an instructional effect. Practically, lecturers of the compulsory Indonesian language course, particularly in non-language programmes, can use the collaborative learning model to teach argumentative writing by designing open, challenging tasks that require groups to negotiate positions, gather evidence, and anticipate counterarguments. Because some students improved less than others, lecturers should also monitor group interaction and provide additional scaffolding, such as guiding questions on warrants and rebuttals or structured peer feedback (Noroozi et al., 2023), for students who show smaller gains. As generative artificial intelligence becomes part of students' writing practices (Su et al., 2023), collaborative tasks can also be designed to encourage students to evaluate machine-generated arguments critically rather than adopt them uncritically.

Research Limitations and Directions for Future Research

This study has several limitations. First, the one-group pretest–posttest design does not include a control group, so the improvement cannot be attributed with certainty to the collaborative learning model alone, and threats such as testing effects, maturation, and regression toward the mean cannot be excluded. Second, the sample consisted of a single class of 38 mathematics education students at one university, which limits the generalizability of the findings to other programmes and institutions. Third, the analysis was based on total scores, so it does not show which aspects of argumentative writing, such as thesis formulation, evidence, reasoning, or rebuttal, improved most. Fourth, the posttest was administered immediately after the intervention, so the retention of the improvement is unknown. Future studies are encouraged to use quasi-experimental or experimental designs with comparison groups, to involve students from several programmes and institutions, to analyse improvement for each aspect of the argumentative writing rubric, to administer delayed posttests, and to examine the quality of group interaction through observation or discourse analysis.

CONCLUSION

This study concludes that the collaborative learning model improved the ability of first-year mathematics education students to compose argumentative texts in the Indonesian language course. Students' performance shifted from predominantly low and very low categories before the intervention to predominantly high and very high categories after it, almost all students reached the mastery criterion, and the improvement was statistically significant, of moderate normalized magnitude, and large in standardized terms. Students who started with lower scores tended to benefit the most, although the extent of improvement varied among individuals.

The study contributes to knowledge by extending evidence on collaborative argumentative writing, which has so far come mainly from EFL and secondary school contexts, to argumentative writing in Indonesian within a compulsory general education course for non-language majors. It also demonstrates the value of reporting normalized gain and effect size alongside significance tests in classroom-based studies of writing instruction. These contributions should be understood within the scope of a single-group design in one class.

Future research is recommended to test the collaborative learning model using quasi-experimental or experimental designs with comparison groups and larger, more diverse samples; to examine improvement in each component of argumentative writing; to assess the durability of the improvement through delayed posttests; and to investigate how specific forms of support, such as metacognitive guidance and structured peer feedback, shape the quality of collaboration and the gains of students who benefit less from group work.

REFERENCES

Ananda, C. R., Mardiah, R., & Fajaryani, N. (2024). Writing argumentative essays: Jambi EFL students' challenges and strategies. *Indonesian Journal of English Language Teaching and Applied Linguistics, 9*(2), 339–357. https://doi.org/10.21093/ijeltal.v9i2.1722

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research, 88*(6), 799–843. https://doi.org/10.3102/0034654318791584

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications. https://www.sagepub.com/shop/buy-a-book/research-design-5-255675

Elabdali, R. (2021). Are two heads really better than one? A meta-analysis of the L2 learning benefits of collaborative writing. *Journal of Second Language Writing, 52*, Article 100788. https://doi.org/10.1016/j.jslw.2020.100788

Ghanbari, N., & Salari, M. (2022). Problematizing argumentative writing in an Iranian EFL undergraduate context. *Frontiers in Psychology, 13*, Article 862400. https://doi.org/10.3389/fpsyg.2022.862400

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

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

Kheryadi, Suseno, M., & Lustyantie, N. (2024). Flipped classroom with collaborative learning approach in enhancing writing skills of Indonesian university students. *International Journal of Evaluation and Research in Education, 13*(1), 407–413. https://doi.org/10.11591/ijere.v13i1.25269

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning, 13*(2), 213–233. https://doi.org/10.1007/s11412-018-9277-y

Kleemola, K., Hyytinen, H., & Toom, A. (2022). The challenge of position-taking in novice higher education students' argumentative writing. *Frontiers in Education, 7*, Article 885987. https://doi.org/10.3389/feduc.2022.885987

Kurniati, L., Syaputri, W., Efendi, Y., Tussolekha, R., & Saputry, D. (2024). Improving Bahasa Indonesia writing skill: A project-based learning experimental study. *JPPI (Jurnal Penelitian Pendidikan Indonesia), 10*(3), 775–782. https://doi.org/10.29210/020244090

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Landrieu, Y., De Smedt, F., Van Keer, H., & De Wever, B. (2024). Argumentation in collaboration: The impact of explicit instruction and collaborative writing on secondary school students' argumentative writing. *Reading and Writing, 37*(6), 1407–1434. https://doi.org/10.1007/s11145-023-10439-x

Nakrowi, Z. S., & Mulyati, Y. (2021). Evaluasi Kualitas Argumen pada Artikel Jurnal [Evaluation of argument quality in journal articles]. *LITERA, 20*(1), 90–109. https://doi.org/10.21831/ltr.v20i1.34287

Noroozi, O., Banihashem, S. K., Biemans, H. J. A., Smits, M., Vervoort, M. T. W., & Verbaan, C.-L. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies, 28*(10), 12757–12784. https://doi.org/10.1007/s10639-023-11683-y

Parlindungan, F., Mursyidin, Kurniasih, Rahmatillah, R., Nuthihar, R., & Oussou, S. (2023). The effect of collaborative reasoning on Indonesian university students' literacy and argumentation skills. *JEELS (Journal of English Education and Linguistics Studies), 10*(1), 97–118. https://doi.org/10.30762/jeels.v10i1.788

Qureshi, M. A., Khaskheli, A., Qureshi, J. A., Raza, S. A., & Yousufi, S. Q. (2023). Factors affecting students' learning performance through collaborative learning and engagement. *Interactive Learning Environments, 31*(4), 2371–2391. https://doi.org/10.1080/10494820.2021.1884886

Republik Indonesia. (2012). *Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi* [Law of the Republic of Indonesia Number 12 of 2012 on Higher Education]. https://peraturan.bpk.go.id/Details/39063/uu-no-12-tahun-2012

Scager, K., Boonstra, J., Peeters, T., Vulperhorst, J., & Wiegant, F. (2016). Collaborative learning in higher education: Evoking positive interdependence. *CBE—Life Sciences Education, 15*(4), Article ar69. https://doi.org/10.1187/cbe.16-07-0219

Shinta, D. K., & Filia. (2020). Improving students' arguments through collaborative learning. *Indonesian Journal of Applied Linguistics, 10*(2), 349–358. https://doi.org/10.17509/ijal.v10i2.28602

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, Article 100752. https://doi.org/10.1016/j.asw.2023.100752

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press. https://www.cambridge.org/core/books/the-uses-of-argument/26CF801BC12004587B66778297D5567C

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, K., Zhang, L. J., & Cooper, M. (2025). Metacognitive instruction for improving the effectiveness of collaborative writing for EFL learners' writing development. *The Asia-Pacific Education Researcher, 34*, 661–673. https://doi.org/10.1007/s40299-024-00886-7

Zheng, Y., Cao, Z., & Yu, S. (2026). Research into practice: Collaborative writing in L2 contexts. *Language Teaching*. Advance online publication. https://doi.org/10.1017/S0261444826101232