



Changes in Fourth-Grade Descriptive Writing Following Chain Writing Instruction in an Islamic Elementary School

Annisa Fitriani^{1*}, Khairunnisa¹, Aris Sugianto²,

¹Universitas Islam Negeri Banjarmasin, Indonesia

²Western Sydney University, Australia

*Corresponding author: annisa@uin-antasri.ac.id

Abstract

Developing coherent descriptive paragraphs requires elementary pupils to coordinate idea generation, sentence organization, and writing conventions, yet short-term changes following sequential collaborative writing in Islamic primary classrooms remain insufficiently documented. This study aims to examine fourth-grade pupils' descriptive writing performance before and after chain writing instruction in an Indonesian Islamic elementary school. The study employed a quantitative pre-experimental one-group pretest–posttest design involving 34 purposively selected pupils at MI Plus Darul Ilmi, Banjarbaru. Across three instructional meetings, pupils engaged in sequential collaborative drafting and guided review; writing performance was assessed before and after instruction using a researcher-developed rubric. Descriptive statistics, the Wilcoxon signed-rank test, and individual normalized-gain analysis were supplemented by illustrative classroom writing samples. The findings revealed an increase in mean scores from 60.59 (SD = 7.46) to 81.62 (SD = 5.99), with all 34 paired scores increasing ($Z = -5.350$, $p < .001$; mean N-gain = 0.54, moderate). Selected writing examples illustrated clearer paragraph organization, but continued mechanical errors. However, the absence of a comparison group and incomplete scoring-reliability information prevent causal attribution. The study contributes context-specific evidence of observed writing changes, with implications for integrating collaborative drafting, explicit revision, and stronger future assessments.

ARTICLE HISTORY

Received:

25 May 2025

Accepted:

29 June 2026

Published:

30 September 2026

KEYWORDS

chain writing;
collaborative writing;
descriptive writing;
elementary literacy;
writing performance



1. Introduction

Descriptive writing enables primary pupils to transform observations into connected explanations of people, objects, places, or events. Producing a coherent paragraph requires more than generating isolated statements: young writers must develop relevant details, order ideas, choose vocabulary, and monitor spelling and punctuation. A systematic review of elementary writing productivity found that performance varies with task genre and time allocation, while a meta-analysis of writing interventions identified differences in outcomes by instructional focus and learner characteristics (Martin & Dockrell, 2024; Kim et



al., 2021). These findings support close attention to both the writing process and the conditions under which writing is assessed.

In Indonesian primary settings, studies have documented difficulties in paragraph formation and descriptive writing, including limited elaboration, weak connections between sentences, and inaccurate written conventions (Dewi et al., 2023; Nurazizah & Darmayanti, 2024). The local classroom concern at MI Plus Darul Ilmi, Banjarbaru, was similar: initial instructional observations identified restricted idea development, loosely related sentences, and inconsistent capitalization or punctuation. Those observations motivated classroom instruction; they were not a representative survey of Indonesian madrasahs. Picture-based activities and other writing supports have been investigated in elementary education, but their value depends on the text task and instructional process (Hasan, 2022).

Chain writing, sometimes described as estafet writing, is a sequential collaborative procedure in which pupils read the emerging shared text and contribute in turn. The turn-taking structure potentially makes continuity visible: each contribution has to relate to what has already been written. Indonesian studies have examined chain writing in primary descriptive writing, primary narrative writing, and other contexts (Diwansyah et al., 2022; Putri et al., 2022; Harmi et al., 2025; Nabilah et al., 2023). Their varied participants, language settings, genres, and designs preclude treating favorable results as a universal method-specific effect. A senior high school comparison reported by Kartikasari et al. (2023) further illustrates the need to avoid assuming uniformly positive outcomes across contexts.

Collaborative composition provides a plausible pedagogical rationale without guaranteeing that all group members participate equally. A review by Svenlin and Sørhaug (2023) describes diverse configurations of collaborative writing, while Salo et al. (2023) documented distinct participatory roles among elementary pupils. A case study of primary writers further indicates that planning and revision can unfold through interaction around a shared text (López-Flamarique et al., 2023). Studies of dialogic writing suggest that explicit teacher support and purposeful feedback conversations matter for substantive revisions (Bouwer & van der Veen, 2024; Bouwer, van Braak, et al., 2024). Thus, sequential co-construction may provide opportunities to discuss sentence connection and textual coherence, but the present study did not directly measure interaction quality or participation by each group member.

Writing quality also encompasses features that are not automatically improved by collaboration. Transcription and spelling are associated with early composition performance (Malpique et al., 2025), and research on writing feedback distinguishes substantive revision from corrections of surface errors (Scherer et al., 2024). Evidence on young English-as-a-foreign-language writers

and explicit instruction in second-language settings offers useful design considerations, although it cannot be transferred uncritically to Indonesian-language composition (Milla & García-Mayo, 2024; Trang & Barrot, 2024). The separate consideration of organization and conventions is therefore important even when an overall writing score is used.

The present study addresses a bounded empirical question: what short-term changes in descriptive writing scores and selected classroom texts were observed in one fourth-grade Islamic elementary classroom following chain writing instruction? Specifically, it describes pretest–posttest performance distributions, tests the direction of paired score differences, summarizes individual normalized gain, and illustrates features of pupils’ writing across three meetings. The contribution is a transparent account of within-class change rather than a causal test of chain writing against an alternative approach.

The literature also suggests why a single total score can obscure theoretically different outcomes. Writing productivity refers to how much text is produced under specific task conditions, whereas coherence concerns how sentences and details relate to a controlling topic. Neither quantity alone guarantees precise vocabulary or conventional spelling. Martin and Dockrell (2024) show that even productivity measures are sensitive to task characteristics. Studies of children’s early transcription processes and dialogic revision likewise identify different learning demands within composition (Malpique et al., 2025; Bouwer & van der Veen, 2024). Consequently, the present use of a composite writing-performance rubric can document a score pattern but cannot determine which underlying skill changed most without domain-specific results.

A further instructional distinction is between cooperating to produce one text and individually being able to compose a comparable text. Chain writing gives pupils access to the developing ideas of classmates, which may reduce the immediate burden of generating every sentence alone. It also changes the audience for each contribution: a pupil has to consider the existing draft rather than write an unrelated sentence. Such an arrangement is compatible with collaborative planning and revision documented in primary classrooms (López-Flamarique et al., 2023). Yet group text quality, frequency of verbal participation, and an individual pupil’s independently assessed skill are three separate outcomes. The present classroom record documents reported individual score distributions and selected shared-task products, but not a full account connecting these levels of evidence.

This distinction helps refine the study’s research gap. The manuscript does not claim that chain writing is new or that its use has been absent from Indonesian schools. Earlier local investigations have already considered chain-writing approaches in fourth-grade descriptive writing (Diwansyah et al., 2022; Harmi et al., 2025). What remains useful here is a carefully bounded description of how an

entire intact fourth-grade madrasah class performed across two assessment occasions, considered alongside the visible characteristics of selected classroom texts. That contribution is modest but potentially informative for elementary teachers who need to interpret changes in several dimensions of writing without claiming an experimental advantage.

2. Methods

The study used a quantitative pre-experimental one-group pretest–posttest design. It was conducted at MI Plus Darul Ilmi in Banjarbaru, Indonesia. The fourth-grade population comprised 72 pupils. Class IV A was selected purposively because preliminary classroom work had identified challenges in idea development, sentence organization, and writing conventions. All 34 pupils in this intact class were included in the paired comparison. The design permits examination of changes across assessment occasions but cannot eliminate maturation, repeated testing, teacher attention, or other time-related explanations. Observations and examples of written work were used as contextual illustrations, not as an independent qualitative sample.

The use of a purposively selected class defines the scope of the case. The pre-existing class was selected because writing difficulties had already been noticed in local classroom work; this was a practical educational decision rather than random sampling. Such selection also means that score changes cannot be assumed to characterize all fourth-grade learners at the same school, let alone all Indonesian Islamic elementary schools. The recorded population of 72 fourth graders should not be confused with the analytic sample of 34 pupils. No sex-, age-, ability-, or other subgroup breakdown was supplied, and subgroup comparisons are therefore not attempted. Similarly, there is no evidence that an otherwise equivalent class completed an alternative instructional approach during the same period.

The instructional sequence comprised three meetings for a chain-writing activity. Students prepared ideas, contributed successive sentences to shared descriptive texts, reread the developing drafts, and revisited organization and writing conventions with teacher guidance. The original record does not report session length, group composition, the full task prompts, or the delay between pretest and posttest. These omissions limit implementation fidelity checks and exact replication and are not filled by assumption in this report.

A researcher-developed writing-performance rubric addressed five dimensions: idea content, specificity of object description, text organization, style and diction, and spelling and punctuation. A lecturer specializing in Indonesian language education and a classroom teacher reviewed the rubric's indicators and clarity; revisions were reportedly made before use. Scores were reported on a 0–

100 scale. Domain weights, full performance descriptors, independent scoring records, and inter-rater reliability were not available in the manuscript. Accordingly, the appraisal supports reported content-oriented expert review, not a claim of established scoring reliability. Contemporary assessment scholarship reinforces the importance of distinguishing the reliability and validity of writing-score interpretations (Bouwer, Lesterhuis, et al., 2024).

Table 1.
Writing assessment domains and intended performance evidence

Domain	Performance focus
Idea content	Relevance and elaboration of ideas within the descriptive topic
Object description	Specificity and appropriateness of details describing the focal subject
Text organization	Logical sequence, sentence connection, and paragraph cohesion
Style and diction	Vocabulary variety and appropriate sentence expression
Spelling and punctuation	Use of orthographic conventions and punctuation affects clarity.

For descriptive interpretation, the original manuscript classified scores as very limited (0–39), limited (40–59), moderate (60–74), capable (75–84), and highly capable (85–100). These categories were local reporting bands and should not be treated as externally validated proficiency benchmarks. The source record does not establish whether every final score was based solely on an individually produced assessment text or how collaborative products were excluded from assessment; this distinction requires author documentation.

For conceptual clarity, each assessment dimension corresponds to a distinct observable aspect of written work. Idea content concerns topic relevance and elaboration; object description concerns the selection of distinguishing details; organization concerns the relationship among sentences and the continuity of a paragraph; diction concerns communicative appropriateness; and written conventions concern spelling, capitalization, and punctuation. This conceptual separation follows the more general literature on differences among composition processes, text quality, and transcription (Kim et al., 2021; Malpique et al., 2025). Nonetheless, because the numerical weights and item-level scoring rules are unavailable, the analytic unit for the present study is the reported overall score, not five independently validated construct scores.

The classroom writing examples and the test scores should also be kept analytically separate. A writing sample can illustrate a possible feature of an instructional process, but does not, by itself, estimate the prevalence of that feature in a class. For example, a text with additional descriptive details may display more elaboration than another photographed text without implying that the same pupil acquired a lasting skill. As a result, the examples are presented as contextual illustrations and their descriptive characteristics are summarized

cautiously, without counting them as a second dataset or claiming formal triangulation across independently coded sources.

Data were summarized using means, standard deviations, variance, score ranges, and proficiency-band counts. Shapiro–Wilk significance values were reported separately for pretest and posttest score distributions. A two-sided Wilcoxon signed-rank test evaluated paired score changes, and the sign of *Z* was interpreted according to software coding rather than as the direction of improvement. The study also reported an average of student-level normalized gains, using $(\text{posttest} - \text{pretest}) / (100 - \text{pretest})$ for a pupil whose pretest score was below 100. Low, moderate, and high gain categories were defined as below 0.30, 0.30–0.70, and above 0.70. The reported analyses used SPSS Statistics 26. Individual paired scores, the raw scoring sheet, and SPSS output were not included in the available file; thus, the analysis below accurately reproduces reported statistics but does not represent an independent reanalysis.

Three methodological distinctions guide statistical interpretation. First, the means and score bands describe the distribution of the observed cohort and do not measure a causal mechanism. Second, the Wilcoxon signed-rank test summarizes the direction and ranks of paired differences rather than the magnitude of a treatment effect relative to a concurrent comparator. Third, normalized gain describes the proportion of available score improvement realized by each pupil but can be affected by baseline scores and is not interchangeable with a causal effect size.

3. Results

3.1 Changes in writing-score distributions

The mean descriptive writing score increased from 60.59 (SD = 7.46) at pretest to 81.62 (SD = 5.99) at posttest, an observed increase of 21.03 points. The minimum score increased from 35 to 70, and the maximum from 70 to 90. Score distributions shifted toward the two higher locally defined proficiency categories (Table 2). The band counts are cross-sectional summaries of each occasion, not a matrix of person-level transitions.

Table 2.

Descriptive writing performance before and after chain writing (*n* = 34)

Statistic or score band	Pretest	Posttest
Mean (SD)	60.59 (7.46)	81.62 (5.99)
Variance	55.70	35.94
Minimum–maximum	35–70	70–90

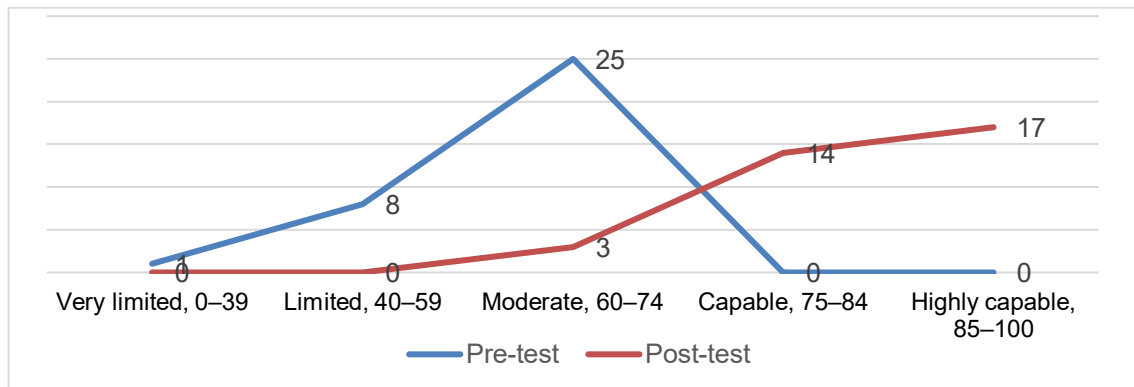


Fig. 1. Changes in writing performance before and after chain writing

Before instruction, 25 pupils (73.5%) were in the moderate category, eight (23.5%) in the limited category, and one (2.9%) in the very limited category. At posttest, 17 pupils (50.0%) were in the highly capable category, 14 (41.2%) in the capable category, and three (8.8%) remained in the moderate category. These findings document higher scores under the study rubric; they do not establish why the scores changed.

The distributional shift in Table 2 is noteworthy because no pupil was in the two highest score bands before instruction, whereas 31 of 34 pupils appeared in those bands at posttest. This is a valid statement about the marginal distributions under the reported local rubric. It does not disclose exactly which initial band each individual pupil occupied or how the changes varied across the five rubric domains. A paired cross-tabulation, anonymized score list, and domain-level means would be needed to examine whether lower-scoring pupils changed differently from initially stronger writers or whether organization changed more than writing conventions. The current record supports only an overall class-level descriptive comparison.

3.2 Paired-score differences and normalized gain

Shapiro–Wilk tests reported $p = .001$ for pretest scores and $p = .009$ for posttest scores, indicating statistically detectable departures from normality in each score distribution. These are not tests of the normality of paired differences. According to the reported Wilcoxon signed-rank output, all 34 pupils had positive score differences, with no negative differences and no ties. The positive mean rank was 17.50, and the sum of positive ranks was 595.00. The paired test returned $Z = -5.350$, two-sided $p < .001$. The negative sign follows the software test convention and does not mean scores decreased.

Table 3.
Reported paired writing-score statistics

Measure	Reported result	Interpretation
Shapiro–Wilk, pretest	$p = .001$	Distribution at the pretest occasion
Shapiro–Wilk, posttest	$p = .009$	Distribution at the posttest occasion
Wilcoxon rank counts	34 positive; 0 negative; 0 ties	Each pupil had a higher reported posttest score
Positive mean/sum ranks	17.50 / 595.00	Internally consistent with 34 positive differences
Wilcoxon signed-rank test	$Z = -5.350$; $p < .001$	Within-class difference between paired assessment occasions
Standardized statistic $ Z /\sqrt{n}$	0.92	Magnitude of observed rank-based statistic; not a causal effect

The manuscript reported a mean of pupil-level normalized gains of 0.54 (moderate), with observed individual gains ranging from 0.33 to 0.71. Computing a normalized change from the two rounded class means yields approximately 0.53; the mean of individual gain ratios need not equal that class-mean ratio. Individual gain frequencies, a paired-score dataset, and confidence intervals were not available for independent verification.

The rank-based result adds information beyond the mean difference. With 34 positive paired ranks and no reported ties, there is evidence that the direction of score change was consistent within the assessed class; this is not merely the consequence of a few extremely high posttest scores raising the average. The standardized statistic calculated from the reported Z and sample size, $|-5.350|/\sqrt{34}$, is approximately 0.92. This quantity is a standardized form of the reported rank-test statistic and should not be called the causal effect of chain writing. In the absence of the pupil-level paired scores, it also cannot substitute for a confidence interval describing uncertainty around the mean or median within-class change.

3.3 Illustrations from classroom writing

Three selected text samples from the instructional meetings offer contextual illustrations of differences in written products, summarized in Table 4. The first shows short, loosely connected descriptive statements and visible punctuation difficulties; the second shows more extended writing and some progression toward a continuous paragraph; and the third shows a more sustained sequence of statements with additional details, while mechanical errors remain. The examples concern different writing topics, and the available record does not establish whether they were produced by the same pupil. They must therefore not be interpreted as a matched longitudinal sequence or independent evidence

of across-class growth. The original sample photographs contain handwritten names; these images are not reproduced here to avoid disclosing potential pupil identifiers.

Table 4.

Features reported in selected writing examples

Meeting	Illustrative textual feature	Continuing limitation
First	Short statements, repetition, and fragmented paragraph structure	Weak linkage and punctuation
Second	More elaborate description and paragraph-like organization	Uneven transitions and capitalization
Third	More extended sequencing and additional descriptive details	Some spelling and punctuation errors

Moreover, classroom writing samples provided descriptive evidence across three instructional meetings. At the first meeting, some students wrote disconnected sentences, repeated words, and made punctuation errors. By the second meeting, paragraph structure and detail were more evident, although sentence connections and capitalization remained inconsistent. By the third meeting, the examples documented more developed descriptions, more coherent paragraph organization, and fewer visible writing-convention errors. The examples are illustrative classroom observations; separate scores for each rubric dimension were not reported.

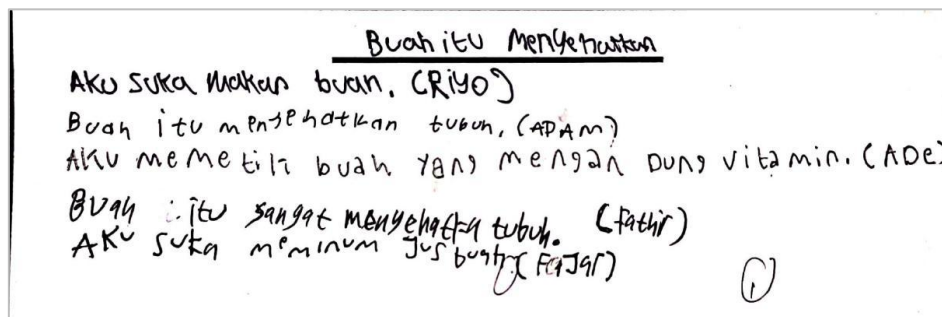


Figure 1. Example of students' descriptive writing at the first meeting.

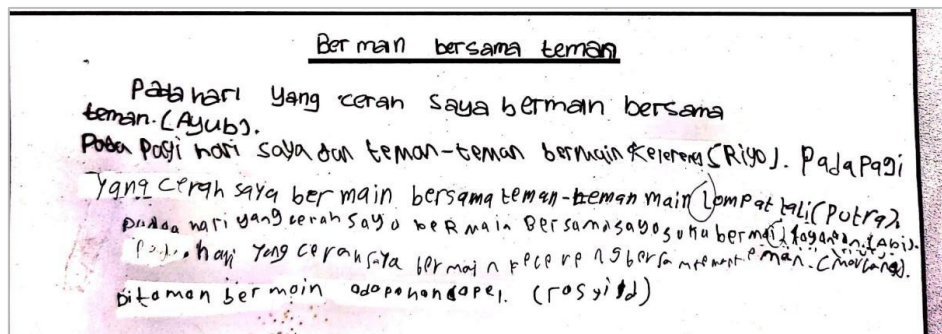


Figure 2. Example of students' descriptive writing at the second meeting.

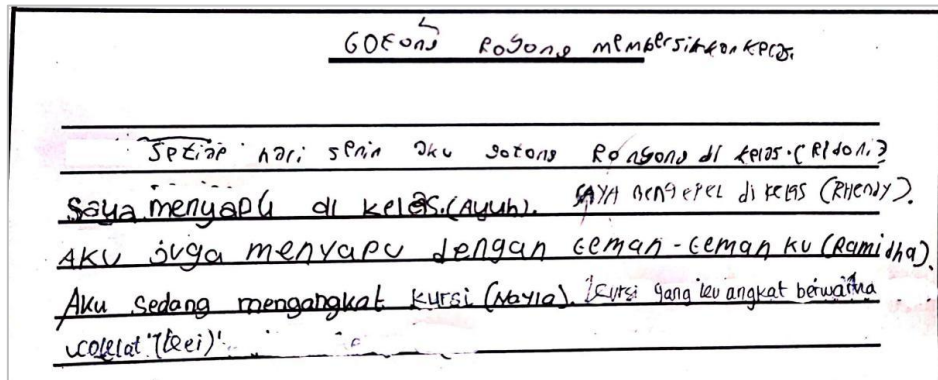


Figure 3. Example of students' descriptive writing at the third meeting.

The three writing samples provide contextual information that a single score does not convey. In the first example, sentences were placed separately and did not always form a continuous description; repetition and punctuation problems were apparent. The second example illustrates the emergence of paragraph-like organization, along with ongoing weaknesses in sentence connections and capitalization. The third example shows more detailed descriptive language and greater attention to sentence ordering, although some lexical and punctuation problems persisted. Therefore, these are selected classroom examples and should not be interpreted as a systematic qualitative sample representing every participant. Their function is to illustrate observable characteristics of the writing process, while the quantitative assessment provides the study's principal outcome measure.

4. Discussion

4.1 Interpreting the observed within-class change

The central empirical finding is a consistent increase in scores across the 34 paired observations, accompanied by a descriptive shift toward higher rubric bands. The difference of 21.03 points and the Wilcoxon result provide evidence that this class performed differently on the post-instruction writing assessment than at baseline. The finding is compatible with the possibility that sequential collaborative drafting supported pupils during the short instructional sequence, but the design does not distinguish instruction-specific effects from repeated task exposure, maturation, changes in teacher support, or other classroom events. Prior primary-grade meta-analytic evidence shows that writing outcomes depend on instructional focus and pupils' characteristics (Kim et al., 2021); the present result should be positioned as local, short-term evidence rather than generalized efficacy.

There is a practical rationale for chain writing as a visible form of shared composing. Reading preceding sentences and extending them may draw attention to topic continuity and elaboration. Studies of primary collaborative-

writing tasks and dialogic writing describe how planning, peer contributions, and discussion can be organized around a shared text (López-Flamarique et al., 2023; Bouwer & van der Veen, 2024). However, this study neither coded peer dialogue nor traced which pupil contributed to each revision. It would be unwarranted to infer that every member developed the same underlying planning or revision strategies.

Interpreting a rapid increase across three meetings requires particular care. Short intervals can make classroom routines easier to learn, increase familiarity with assessment expectations, and allow pupils to reproduce structures explicitly modeled by a teacher. Such changes can matter educationally even when they are not durable or specific to the chain-writing technique. Repeated assessment can also interact with these processes: knowing what a descriptive paragraph should contain on the second occasion may improve performance independently of the turn-taking rule. The present study has no delayed writing task and no comparator that received the same prompts, teacher feedback, and writing time without chain writing. Thus, a balanced conclusion emphasizes the change actually measured and treats longer-term learning and method-specific contribution as future research questions.

4.2 Shared text production and participation

Sequential turn-taking makes participation structurally possible but does not guarantee equal substantive participation. Observational research found that elementary pupils adopt different content-, literacy-, and process-related roles in collaborative writing (Salo et al., 2023), while the L1 scoping review by Svenlin and Sørhaug (2023) documents considerable variation in collaborative-writing organization. Future accounts of chain writing should report group composition, turn allocation, teacher prompting, who revises whose text, and evidence of the resulting decisions. Such records would provide a stronger basis for explaining why an instructional sequence coincides with writing-score change.

Teacher involvement is important because interaction around a shared draft may remain superficial without guidance. Dialogic feedback research emphasizes the quality of conversation and the connection between feedback and meaningful revision (Bouwer, van Braak, et al., 2024). Primary peer-tutoring and peer-assessment literature likewise underscores the importance of structured support rather than collaboration as an isolated label (Corcelles-Seuba et al., 2025; Stecca et al., 2022). The present study did not independently evaluate these processes; they are design considerations for more rigorous follow-up work.

This issue has implications for the social organization of classroom instruction. A shared text may distribute cognitive work productively when pupils offer ideas, examine wording, and evaluate how one sentence follows another. Yet the sequence might also permit a pupil to contribute a short phrase without

understanding why it fits. Observational evidence that distinguishes roles in collaborative writing makes these different interpretations plausible (Salo et al., 2023). Future classroom research could distinguish authoring, suggesting, questioning, checking, and revising turns, and relate them to subsequent individually completed writing tasks. Such evidence would allow researchers to examine whether the collaborative structure provides a temporary task scaffold or supports more transferable individual decisions.

The sequence also creates a need for classroom management that is only indirectly addressed by the current record. In a group writing task, the teacher must decide how pupils rotate, how long each contribution lasts, how disagreements are resolved, and when the whole text is reread. These choices determine whether the assignment is merely an exercise in taking turns or an occasion for mathematical-like argument about language decisions: why a detail is relevant, where a sentence belongs, and what the reader needs to know. Dialogic-writing research shows that supportive conversation routines can be deliberately taught (Bouwer, van Braak, et al., 2024), whereas peer-tutoring research emphasizes intentional task organization (Corcelles-Seuba et al., 2025). The source article does not provide a procedural record detailed enough to identify which of these practices occurred.

4.3 Coherence and writing conventions require complementary supports

The selected writing examples suggest that connected sentence production and orthographic accuracy should be considered separately. A pupil may write a more extended description while continuing to make spelling, capitalization, or punctuation errors. Evidence on transcription skills among young writers illustrates why these lower-level processes remain consequential for composition (Malpique et al., 2025), and a study of Year 4 word-processing and typing instruction further highlights the specificity of transcription-focused learning targets (Gahshan & Weintraub, 2025). Neither study establishes the mechanism operating in the present classroom.

A practical response would combine sequential drafting with explicit checks of descriptive detail, sentence linkage, and selected writing conventions. Feedback should target more than punctuation alone: the critical issue is whether each added sentence advances the intended description. Research on dialogic writing, writing feedback, and peer revision provides a rationale for linking talk and revision without assuming automatic gains (Bouwer & van der Veen, 2024; Scherer et al., 2024; Stuulen et al., 2024). Such practices are proposed implications rather than separately evaluated intervention components in this study.

In particular, a higher overall score should not be equated with the elimination of writing difficulties. The samples summarized in Table 4 display continuing punctuation and spelling problems even when prose appears more

connected. Orthographic control and sentence-level revision may require different teaching moves from topic planning or detail generation. Research on transcription skills and feedback is consistent with this distinction, while evidence from language-learning settings suggests that the timing of explicit correction can influence particular accuracy outcomes (Malpique et al., 2025; Milla & García-Mayo, 2024; Trang & Barrot, 2024). These sources provide reasons to evaluate mechanics separately; they do not justify a claim that the present intervention improved coherence more than mechanics, because dimension-specific paired scores were not supplied.

A classroom-level design implication is a writing cycle with different purposes at different stages. Early stages may invite pupils to propose observable characteristics of an object, place, or activity and choose a clear descriptive focus. Sequential drafting can then direct attention to whether each new sentence extends the same description and whether it introduces a relevant detail. A rereading stage can ask pupils to identify repetition and missing connections, while a separate editing stage can focus on agreed conventions such as sentence boundaries and capitalization. The sequence would need to be modeled explicitly and adapted to fourth-graders' readiness. In this manuscript, it is proposed as a teaching implication drawn from the available results and related scholarship, not described as a rigorously documented or isolated sequence of treatment components (Bouwer & van der Veen, 2024; Scherer et al., 2024).

4.4 Contextual comparison with chain-writing research

The present findings resemble reports of positive writing outcomes after chain-writing instruction in Indonesian settings, including elementary descriptive tasks (Diwansyah et al., 2022; Harmi et al., 2025) and elementary narrative writing (Putri et al., 2022). Other studies have examined descriptive writing in different ages and language contexts (Cahyaningtyas et al., 2024; Nabilah et al., 2023). Yet cross-study resemblance cannot establish a common mechanism: task genres, language proficiency, instructional time, scoring procedures, and comparison conditions differ. Kartikasari et al. (2023) provide a contrasting secondary-school result that cautions against a universal effectiveness claim.

The wider elementary writing-intervention literature demonstrates the value of explicit comparison groups and attention to revision outcomes. Studies of fourth-grade peer feedback (Sugiarti et al., 2026) and comprehensive writing interventions (Carmona et al., 2026) illustrate different designs and constructs that cannot be conflated with chain writing. These comparisons motivate more specific future questions about whether the sequential contribution structure adds value beyond other collaborative or teacher-guided writing methods.

Transferability must also be considered at the level of writing genre. Descriptive writing emphasizes selecting features and maintaining a coherent focus on a person, place, object, or experience. Narrative writing often requires

event sequence, complication, and resolution, while instruction in a second language introduces additional vocabulary and grammar constraints. Chain-writing investigations carried out in these different circumstances may share a turn-taking procedure while measuring different outcomes (Putri et al., 2022; Cahyaningtyas et al., 2024; Nabilah et al., 2023). Even within descriptive writing, comparing student texts composed with different prompts is challenging. Consequently, the present evidence is best interpreted in relation to the actual local assessment and does not establish whether gains would appear with unfamiliar genres or delayed independent tasks.

The distinction between perceived improvement in group products and measured individual development is particularly consequential for future experimental work. A future comparison might equalize writing time, teacher instruction, review opportunities, and assessment prompts between chain writing and another collaborative format, varying only the ordered contribution structure. Repeated individually written texts could then be scored anonymously by trained raters. That design would permit comparisons of total writing quality and the separate outcomes of content, organization, and conventions. More than one class in each condition would also be necessary to limit the confounding of a teaching method with the characteristics of a particular classroom. The current results justify investigating these alternatives, but do not establish the magnitude of their effects.

4.5 Methodological constraints and future research

Several limitations are decisive. First, one purposively selected class without a comparison group cannot support a method-specific causal claim. Second, the same class was assessed twice in a brief teaching period, leaving possible practice effects and changes in assessment familiarity uncontrolled. Third, the rubric's weights and detailed performance anchors, scorer configuration, and inter-rater agreement are not available, preventing an independent appraisal of measurement precision. Writing-assessment research explicitly shows why reliable comparative judgments and transparent rating procedures are important (Bouwer, Lesterhuis, et al., 2024). Fourth, individual paired scores and raw gain values are unavailable, so the Wilcoxon Z and mean pupil-level N-gain cannot be independently reconstructed. Finally, the three selected samples represent different topics and cannot substantiate within-pupil textual development.

Future studies should compare chain writing with a credible alternative across multiple classes, document lesson timing and group composition, use comparable individual writing prompts, preserve anonymized drafts and revision histories, and establish blinded or independently checked scoring. Reporting rubric-level changes in content, organization, diction, and conventions would clarify which dimensions shift. Video or transcript analysis of pupil dialogue could

examine collaborative mechanisms, while delayed tests would help distinguish short-term task familiarity from retained skill. These extensions would permit more defensible judgments about mechanisms, transferability, and durability.

Transparent documentation would strengthen both the scientific and pedagogical usefulness of a subsequent study. Alongside the final pretest and posttest scores, researchers could archive anonymized writing prompts, grading rubrics, examples at different score levels, rater-training procedures, and student-level change distributions. For collaborative lessons, a simple record of participation roles, draft changes, and teacher feedback could connect performance statistics to plausible instructional processes. This would help distinguish whether improvements are concentrated in idea development, paragraph cohesion, mechanics, or combinations of those domains. Without such documentation, a substantial average gain risks being overinterpreted as a complete account of pupils' writing development. Therefore, the present article treats measurement transparency as part of its substantive contribution rather than a merely administrative limitation.

5. Conclusion

Following three meetings of chain writing instruction, fourth-grade pupils in one Indonesian Islamic elementary classroom obtained higher descriptive writing assessment scores than at pretest. The mean increased from 60.59 to 81.62; all 34 paired scores increased. The reported Wilcoxon difference was statistically significant ($Z = -5.350$, $p < .001$), and the mean student-level normalized gain was 0.54 (moderate). Selected classroom texts provided additional illustrations of paragraph elaboration and continuing difficulties in written conventions. The findings document consistent short-term within-class change, but do not prove that chain writing alone caused the improvement.

The study contributes a context-specific account of a sequential collaborative drafting routine linked to measured changes in writing performance and highlights the need to connect text organization with explicit revision of language conventions. Elementary teachers may consider such a routine as a structured opportunity for participation and discussion, while stronger claims about comparative effectiveness require controlled designs, transparent scoring, participant-level data, and direct observation of collaborative processes.

Declarations

All authors have reviewed and approved the final manuscript and declare no competing interests. Research permission was obtained from MI Plus Darul Ilmi Banjarbaru, Indonesia, with informed consent from parents or guardians and age-appropriate assent from participating students. Ethical principles, voluntary

participation, and participant confidentiality were maintained throughout the study. Grammarly was used to support the refinement of academic language during manuscript preparation. No research data, statistical results, or findings were generated or altered using artificial intelligence. All authors verified the manuscript and accept full responsibility for its accuracy, research integrity, and scholarly content.

References

- Bouwer, R., & van der Veen, C. (2024). Write, talk, and rewrite: The effectiveness of a dialogic writing intervention in upper elementary education. *Reading and Writing*, 37(6), 1435–1456. <https://doi.org/10.1007/s11145-023-10474-8>
- Bouwer, R., Lesterhuis, M., De Smedt, F., Van Keer, H., & De Maeyer, S. (2024). Comparative approaches to the assessment of writing: Reliability and validity of benchmark rating and comparative judgment. *Journal of Writing Research*, 15(3), 497–518. <https://doi.org/10.17239/jowr-2024.15.03.03>
- Bouwer, R., van Braak, M., & van der Veen, C. (2024). Dialogic writing in the upper grades of primary school: How to support peer feedback conversations that promote meaningful revisions. *Learning and Instruction*, 93, Article 101965. <https://doi.org/10.1016/j.learninstruc.2024.101965>
- Cahyaningtyas, A. N., Riyadi, S., & Kariadi, M. T. (2024). Using the chain writing method to improve the students' skills in writing descriptive texts. *Jurnal Vokasia*, 4(1), 1–15. <https://doi.org/10.20884/1.vokasia.2024.4.1.12141>
- Carmona, M., De los Santos-Roig, M., Mata, S., & Serrano, F. (2026). Effectiveness of a comprehensive game-based intervention for writing skills. *Reading and Writing*, 39, 121–139. <https://doi.org/10.1007/s11145-025-10636-w>
- Corcelles-Seuba, M., Ribosa, J., & Calaforra-Fabuel, P. J. (2025). Peer tutoring to improve writing competence in primary education. *Revista Española de Pedagogía*, 83(291), 363–381. <https://doi.org/10.22550/2174-0909.4156>
- Dewi, F. D. K., Slamet, S. Y., & Kurniawan, S. B. (2023). Analisis kesalahan paragraf dalam menulis karangan deskripsi peserta didik kelas VI pada pembelajaran bahasa Indonesia. *Didaktika Dwija Indria*, 11(4). <https://doi.org/10.20961/ddi.v11i4.76726>
- Diwansyah, A., Prasetyo, T., & Laeli, S. (2022). Pengaruh metode menulis berantai (estafet writing) terhadap keterampilan menulis deskripsi di kelas IV SDN Caringin. *Al-Kaff: Jurnal Sosial Humaniora*, 5(2), 103–112.
- Gahshan, N., & Weintraub, N. (2025). Effect of an instructional program for word processing and efficient typing on Year 4 students' composition. *Reading and Writing*, 38, 2661–2683. <https://doi.org/10.1007/s11145-024-10613-9>

- Harmi, L. V., Kharisma, A., & Tulljanah, R. (2025). Pengaruh metode pembelajaran chain writing terhadap keterampilan menulis deskripsi siswa kelas IV SDN 04 Tarandam. *EDU RESEARCH*, 6(4), 67–74. <https://doi.org/10.47827/jer.v6i4.1566>
- Hasan, H. (2022). Peran media gambar berseri terhadap kemampuan menulis karangan siswa sekolah dasar. *Ainara Journal: Jurnal Penelitian dan PKM Bidang Ilmu Pendidikan*, 3(2), 111–117. <https://doi.org/10.54371/ainj.v2i4.99>
- Kartikasari, I., Sudar, S., Basuki, B., Ngafif, A., & Rokhayati, T. (2023). The effect of the use of the estafet writing to improve students' writing of descriptive text at a senior high school. *Scripta: English Department Journal*, 10(1), 67–75. <https://doi.org/10.37729/scripta.v10i1.2671>
- Kim, Y.-S. G., Yang, D., Reyes, M., & Connor, C. (2021). Writing instruction improves students' writing skills differentially depending on focal instruction and children: A meta-analysis for primary grade students. *Educational Research Review*, 34, Article 100408. <https://doi.org/10.1016/j.edurev.2021.100408>
- López-Flamarique, M., Aznárez-Mauleón, M., & García-del-Real, I. (2023). How do primary students plan and review a text? A case study of a collaborative writing task. *Didáctica. Lengua y Literatura*, 35, 113–125. <https://doi.org/10.5209/dill.81810>
- Malpique, A. A., Asil, M., Pino-Pasternak, D., Ledger, S., & Teo, T. (2025). The contributions of transcription skills to paper-based and computer-based text composing in the early years. *Reading and Writing*, 38(4), 911–945. <https://doi.org/10.1007/s11145-024-10543-6>
- Martin, C., & Dockrell, J. E. (2024). Writing productivity development in elementary school: A systematic review. *Assessing Writing*, 60, Article 100834. <https://doi.org/10.1016/j.asw.2024.100834>
- Milla, R., & García-Mayo, M. del P. (2024). Collaborative writing, written corrective feedback, and motivation among child EFL learners. *Porta Linguarum*, 41, 27–43. <https://doi.org/10.30827/portalin.vi41.23971>
- Nabilah, R., As'ari, A., & Rahmawati, E. (2023). Effect of chain writing method on students' writing ability of descriptive text. *English Education, Linguistics, and Literature Journal*, 2(2), 95–102. <https://doi.org/10.32678/ell.v2i2.6175>
- Nurazizah, S. A., & Darmayanti, M. (2024). Keterampilan menulis siswa sekolah dasar: Systematic literature review dan bibliometric analysis. *Aksara*, 36(2), 337–358. <https://doi.org/10.29255/aksara.v36i2.4236>
- Putri, N. E., Marta, R., & Surya, Y. F. (2022). Peningkatan keterampilan menulis karangan narasi menggunakan metode chain writing di sekolah dasar. *MODELING: Jurnal Program Studi PGMI*, 9(4), 83–88. <https://doi.org/10.36835/modeling.v9i4.1254>

- Salo, A.-E., Routarinne, S., Juvonen, R., & Kaasinen, A. (2023). Participatory roles adopted by elementary students when writing collaboratively in environmental and social studies classrooms. *Journal of Writing Research*, 15(1), 73–103. <https://doi.org/10.17239/jowr-2023.15.01.04>
- Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction*, 93, Article 101961. <https://doi.org/10.1016/j.learninstruc.2024.101961>
- Stecca, E., Grion, V., Zaggia, C., & Restiglian, E. (2022). The practice of peer assessment in primary school: A systematic literature review. *Italian Journal of Educational Research*, 28, 85–95. <https://doi.org/10.7346/sird-012022-p85>
- Stuulen, J., Bouwer, R., & van den Bergh, H. (2024). Effects of a comparative feedback method on how students use peer feedback for revision. *L1 Educational Studies in Language and Literature*, 24(1), 1–21. <https://doi.org/10.21248/l1esll.2024.24.1.671>
- Sugiarti, D. H., Syihabuddin, S., Sastromiharjo, A., & Kurniawan, K. (2026). Peer feedback and writing revision in primary education: A quasi-experimental study of fourth-grade narrative writing. *PrimaryEdu: Journal of Primary Education*, 10(1), 15–27. <https://doi.org/10.22460/pej.v10i1.6928>
- Svenlin, M., & Sørhaug, J. O. (2023). Collaborative writing in L1 school contexts: A scoping review. *Scandinavian Journal of Educational Research*, 67(6), 980–996. <https://doi.org/10.1080/00313831.2022.2115128>
- Trang, N. H., & Barrot, J. S. (2024). Pre-writing and post-writing explicit instruction: Differential effects on L2 learners' writing accuracy. *RELC Journal*, 55(3), 688–704. <https://doi.org/10.1177/00336882231184270>