SITTAH: Journal of Primary Education
P-ISSN: 2745-4479 E-ISSN: 2745-4487
Journal Homepage: https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/sittah
Journal e-mail: jurnalsittah@iainkediri.ac.id
Page: 89-108

Improving Narrative Writing Literacy Using a Picture Series Based on Local Stories

Dovila Johansz,[1] Fransheine Rumtutuly,[2*]
[1,2]Primary School Education Program (PGSD), PDSKU Southwest Maluku Regency, Pattimura University, Indonesia
[1]dovilajohansz@gmail.com, [2]rumtutulyfransheine21@gmail.com

*Corresponding Author
Received: 2026-07-06; Approved: 2026-10-04; Published: 2026-10-30

Abstract

Narrative writing is a core literacy skill, yet coastal schools often rely on materials that do not reflect students' surroundings. This study examined changes in narrative writing literacy among third-grade students following instruction using a picture series based on local Tiakur children's stories. A pre-experimental one-group pretest–posttest design involved all 20 students at Public Elementary School (*Sekolah Dasar Negeri*/SDN) Tiakur in Southwest Maluku across six 60-minute sessions, including a pretest and a posttest. An analytic rubric of four indicators, with very high inter-rater reliability (ICC = 0.978), scored the writing. Data were analyzed with descriptive statistics, N-Gain, a paired-samples t-test, and a Wilcoxon signed-rank test. This study found that the mean score increased from 47.50 to 81.88, with a moderate N-Gain mean of 0.65 ($t(19) = 13.30$, $p < 0.001$), and a consistent Wilcoxon result. All four indicators improved significantly, with the greatest improvement in content coherence and the least in vocabulary choice. Overall, the picture series was associated with higher narrative writing scores, suggesting that teachers can adapt local stories as a writing scaffold. It offers a culturally grounded medium for a setting rarely studied in writing research. Without a control group, causality cannot be established; therefore, future studies should include comparison groups and longer interventions.

Keywords: Coastal Community, Elementary School, Local Children's Stories, Narrative Writing Literacy, Picture-series Media.

Abstrak

Menulis cerita merupakan keterampilan literasi inti, tetapi sekolah pesisir sering memakai bahan yang tidak mencerminkan lingkungan siswa. Penelitian ini mengkaji perubahan literasi menulis cerita siswa kelas III setelah belajar dengan media gambar seri berbasis cerita anak lokal Tiakur. Desain pre-eksperimen *one-group pretest* dan *posttest* melibatkan seluruh 20 siswa SD Negeri Tiakur, Maluku Barat Daya, dalam enam pertemuan berdurasi 60 menit, termasuk *pretest* dan *posttest*. Rubrik analitik empat indikator dengan reliabilitas antar penilai sangat tinggi (ICC = 0,978) menilai tulisan siswa. Data dianalisis dengan statistik deskriptif, N-Gain, *paired sample t-test*, dan uji Wilcoxon *signed-rank*. Penelitian ini menemukan bahwa skor rata-rata naik dari 47,50 menjadi 81,88 dengan N-Gain rata-rata 0,65 kategori sedang, $t(19) = 13,30$, $p < 0,001$, dan hasil Wilcoxon yang konsisten. Keempat indikator meningkat signifikan, terbesar pada keterpaduan isi dan terkecil pada pilihan kosakata. Secara keseluruhan, gambar seri berbasis cerita lokal berkaitan dengan skor menulis cerita yang lebih tinggi, sehingga guru dapat mengadaptasi cerita lokal sebagai penopang menulis. Penelitian ini menawarkan media visual berbasis budaya untuk konteks yang jarang dikaji dalam penelitian menulis. Tanpa kelompok kontrol, sebab akibat belum dapat disimpulkan,

Copyright © 2026 by the authors.
This work is licensed under the terms and conditions of the Creative Commons Attribution (CC BY NC SA) license (https://creativecommons.org/licenses/by-nc-sa/4.0/).

sehingga penelitian lanjutan perlu memakai kelompok pembanding dan intervensi yang lebih panjang.

Kata Kunci: Cerita Anak Lokal, Literasi Menulis Cerita, Masyarakat Pesisir, Media Gambar Seri, Sekolah Dasar.

INTRODUCTION

Writing is a fundamental competence that underpins learning across school subjects, including science, social studies, and mathematics. Elementary school is where writing habits begin to form (Graham, 2022; Graham et al., 2020). Writing and reading reinforce each other, so teaching them together is beneficial (Graham et al., 2018, 2019). However, writing is demanding because it involves idea generation, language, transcription, and self-regulation simultaneously (Kim & Schatschneider, 2017). This complexity slows the development of writing skills, so how to teach them remains a concern in many countries. For example, writing motives and writing achievement have been examined together among elementary school students from diverse language backgrounds (Camping et al., 2023). Such research is relevant to multilingual classrooms, such as those in eastern Indonesia. A national survey in Norway has also examined how teachers in Grades 1 to 3 teach writing, as well as their preparation and beliefs (Graham et al., 2021).

An intervention study in Norway tested an approach that increases meaningful writing activities for beginning writers (Skar et al., 2024). A meta-analysis of writing interventions was conducted in Turkey because most of the evidence on interventions comes from the United States and Europe (Graham et al., 2022). Literacy outcomes in Indonesia point to a similar need. In PISA 2022, Indonesian students' mean reading score was 359, compared with an OECD average of 476 (OECD, 2023). Within this literacy framework, narrative writing deserves particular attention in the early grades. Stories are the first genre that children encounter, and explicit instruction in narrative skills strengthens young children's narrative abilities (Pesco & Gagné, 2017). Narrative writing requires students to generate ideas, sequence events, choose appropriate words, and maintain textual coherence. Many elementary school students struggle with these demands (Harahap et al., 2022).

Research on visual support in writing continues to grow, with findings scattered across diverse designs, contexts, and outcomes. A meta-analysis showed that writing instruction improves early-grade students' writing, with the size of the effect depending on the instructional focus and children's initial writing skills (Kim et al., 2021). Picture-based

media are commonly justified by the cognitive theory of multimedia learning, which holds that people learn more deeply from words and pictures together than from words alone (Mayer, 2024). Another review reported favorable learning outcomes from story-based multimedia learning (Wu & Chen, 2020). Studies in Indonesian elementary classrooms point in the same direction: sequential pictures help students develop ideas, organize events, and maintain story sequence (Astuti & Rambe, 2024; Lingga & Rambe, 2026; Pujiyanti et al., 2026; Sofiyatun et al., 2026; Sugiharti & Wulandari, 2017). Visual images have also supported idea generation and creativity among university students in Indonesia (Listyani, 2019). Studies of preschool and kindergarten children have reported that attention to picture book illustrations is associated with better narrative ability (Davis et al., 2024) and that children can use visual elements such as color, framing, and line meaningfully in their visual narrative compositions (Pantaleo, 2025).

Comparing these studies reveals three patterns. First, the reviewed studies generally use generic images and treat their cultural content as incidental, so it remains unclear whether familiar local content adds value. Second, results are generally reported as overall scores, which makes it difficult to see which writing indicators changed. Third, the evidence on writing interventions is concentrated in the United States and Europe, whereas evidence from other contexts is only beginning to accumulate (Graham et al., 2022). Coastal communities appear to be underrepresented in this literature, even though international efforts to measure ocean literacy among young people indicate growing attention to children's understanding of coastal and marine environments (Fauville et al., 2019). Studies that combine sequential picture media, fishing-community stories, and rubric-based assessment of narrative writing in elementary schools in eastern Indonesia remain scarce. This gap is both contextual and analytical. Contextually, little is known about picture series that depict the lives of fishing communities and convey care for the coastal environment. Analytically, existing studies rarely link score gains to specific writing indicators.

SD Negeri Tiakur is located in a coastal community in Southwest Maluku Regency, where daily life is closely tied to fishing. Narrative writing instruction at this school has relied on generic examples that do not reflect students' real experiences. A preliminary measurement indicated that students needed support. Before the intervention, the mean narrative writing score was 47.50 on a scale of 25 to 100, and the highest score was only 62.50. Observation notes showed that most students wrote only a few simple sentences without a clear plot. In response, a picture-series medium based on local children's stories

was developed. Its five pictures depict a morning scene at Tiakur Beach, children helping their parents, a beach littered with trash, a communal clean-up of the beach, and a clean and pleasant coastline. The series was designed as a visual stimulus for writing and to convey the values of *gotong royong* (cooperation) and care for the coastal environment. Learning materials drawn from students' own surroundings draw on their experiences and cultural knowledge. This approach has been associated with stronger engagement and learning in research on culturally relevant education (Aronson & Laughter, 2016), and place-based education regards the local environment and community as a source of learning (Semken et al., 2017).

This study tests a picture series based on local Tiakur stories and examines changes in students' writing across four rubric indicators. To the best of the authors' knowledge, the novelty of this study lies in combining local content, sequential pictures, and indicator-based analysis. The study aims to examine changes in the narrative writing literacy of third-grade students at SD Negeri Tiakur after learning with a picture-series medium based on local Tiakur children's stories. Two research questions guide the study. First, how do students' narrative writing scores change, and is the change statistically significant? Second, how do students' scores change on each indicator, namely idea development, plot, vocabulary choice, and content coherence? The Results and Discussion sections follow these two questions, and the findings are interpreted through the lenses of multimedia learning and scaffolding.

RESEARCH METHODS

This study used a pre-experimental one-group pretest–posttest design that compared the same group of students before and after the intervention. This design cannot rule out the influence of maturation, external events, or testing effects, so the results are interpreted as associations rather than causal evidence (Knapp, 2016). The design was chosen because the school has only one third-grade class. The study was conducted at SD Negeri Tiakur, Southwest Maluku Regency, from 14 February to 15 March 2026, in six 60-minute sessions, including the pretest and posttest, taught by Seli Romode, S.Pd. This one-month span aligned with the regular class schedule and was sufficient to examine initial changes following the use of five pictures, but it still left room for maturation or practice to influence the results (Knapp, 2016).

SD Negeri Tiakur was selected purposively for three reasons. The school serves a fishing community whose daily life corresponds to the content of the medium, its narrative

writing instruction has relied on generic examples, and the pretest results confirmed that students needed support. The research subjects were all 20 third-grade students, comprising boys and girls with varying levels of writing literacy. Because the population was limited, all students were included in the sample using total sampling. All students took part in the pretest, the treatment, and the posttest, so there were no missing data.

The treatment was a picture-series medium based on local Tiakur children's stories, consisting of five sequential color pictures printed on A4 paper. The pictures depict a morning scene at Tiakur Beach as a fishing settlement, children helping their parents on the beach, a beach littered with trash, a communal clean-up of the beach, and a clean beach. The medium was validated by a subject-matter expert, a language expert, and a media expert before use. Learning was teacher-guided and followed three steps. The teacher presented the pictures in sequence, and the students observed them; the students discussed the content of the pictures and the connections among events; and each student then wrote a story following the sequence of events. In this way, the pictures served as scaffolding for planning the story before students wrote independently.

Narrative writing literacy was measured using a performance assessment in which each student wrote a story individually in class. Stories were scored with an analytic rubric of four indicators. Idea development refers to the appropriateness and richness of ideas; plot, to the presence and order of the opening, main events, and ending; vocabulary choice, to the variety and precision of words; and content coherence, to the logical connections between events and between sentences. Each indicator was scored on a scale of 1 to 4, with higher scores indicating better performance. The final score was obtained by summing the scores of the four indicators, dividing by 16, and multiplying by 100, yielding final scores ranging from 25 to 100. A rubric was chosen because it makes assessment criteria explicit and has been associated with higher academic achievement (Panadero et al., 2023; Panadero & Jonsson, 2020). An expert in elementary education and an expert in Indonesian language validated the rubric's content. The same rubric and procedures were used in the pretest and posttest, so the pretest served as a baseline performance assessment.

Data collection took place in three stages. In the pretest, students wrote a story about beach activities without the picture series to establish their baseline narrative writing literacy. In the treatment stage, students learned with the picture series through the three steps described above. In the posttest, students wrote a story in response to the prompt "Write a story from these five pictures in the correct order." The pretest and posttest were part of the six sessions, each lasting 60 minutes. Observation notes recorded students'

engagement and the characteristics of their writing during lessons, and photographs documented the study's implementation.

Three experts validated the feasibility of the medium before use: an Indonesian-language subject-matter expert (SR, Master's degree, Pattimura University), a language expert (CK, Master's degree, Pattimura University), and a media expert (BT, Master's degree, Pattimura University). Each expert rated the medium using an instrument appropriate to their field on a scale of 1 to 5, where 1 meant very unfeasible, and 5 meant very feasible. The instrument contained eight items for the subject-matter and language experts and seven items for the media expert, including clarity of the plot, closeness to the local context, explanation of local vocabulary, characters' expressions and gestures, and durability of the material. The feasibility percentage was calculated by dividing the total score by the maximum score, namely 40 for the subject-matter and language experts and 35 for the media expert, and multiplying by 100. The results were interpreted using the following criteria: 81–100% very feasible, 61–80% feasible, 41–60% fairly feasible, 21–40% less feasible, and 20% or below not feasible (Riduwan, 2022). The validators' suggestions served as the basis for revising the medium.

The inter-rater reliability of the rubric was examined through a pilot scoring of 20 narrative writing samples from third-grade students at SD Inpres Herley, a school outside the research subjects with comparable characteristics. Two third-grade teachers from SD Kristen Kaiwatu and SD Inpres Herley independently scored the writing using the same rubric, and total scores were obtained by summing the scores for the four indicators. These pilot data were separate from the pretest and posttest scores and were used only to assess the rubric's reliability. Agreement was analyzed using the Intraclass Correlation Coefficient (ICC) with a two-way mixed-effects model, an absolute-agreement definition, and a single-measures type, because the pilot evaluated absolute agreement between two specific raters and did not generalize the results to a broader population of raters. The choice of model, definition, and reporting format for the ICC, along with the 95% confidence intervals, followed established guidelines (Koo & Li, 2016).

The analysis was conducted on the total score (scale of 25 to 100) and on each indicator score (scale of 1 to 4). Performance at the two measurement points was summarized using the mean, minimum, maximum, and standard deviation. Observation notes and documentation were read against the four rubric indicators and then compared with the test results to triangulate the findings, which, together with expert validation, were intended to strengthen the credibility of the results. Score improvement was calculated for

each student using N-Gain (Hake, 1998) with the following formula, where the maximum total score is 100, and the maximum score for each indicator is 4.

$$\text{N} - \text{Gain} = \frac{Skor\ Posttest - Skor\ Pretest}{Skor\ Maksimal - Skor\ Pretes}$$

N-Gain adjusts the raw gain for each student's room for improvement, so it remains appropriate to use as long as it is read together with the distribution of scores (Coletta & Steinert, 2020). Its values were grouped into high ($g \geq 0.7$), moderate ($0.3 \leq g < 0.7$), and low ($g < 0.3$) categories (Hake, 1998). The null hypothesis (H_0) stated that there is no difference between the mean pretest and posttest scores, whereas the alternative hypothesis (H_a) stated that there is a difference. The normality of the pretest scores, posttest scores, and score differences was examined using the Kolmogorov–Smirnov test with Lilliefors correction and the Shapiro–Wilk test. Changes in the total score were tested using a paired-samples *t*-test at $\alpha = 0.05$ and confirmed using the Wilcoxon signed-rank test because the scores were derived from a rubric with a discrete scale. Changes in each indicator, which was measured on an ordinal scale, were tested using the Wilcoxon signed-rank test with a Bonferroni correction at $\alpha = 0.0125$. Cohen's d_z was calculated as t divided by the square root of n (Lakens, 2013), and the Wilcoxon effect size as $r = |Z|$ divided by the square root of N. Effect sizes were interpreted cautiously because designs without a comparison group tend to produce larger effects than rigorous studies (Kraft, 2020).

RESULTS AND DISCUSSION

Results

The results are presented in three parts. The first part addresses the first research question through the total score, namely descriptive statistics, assumption checks, N-Gain categories, and significance tests. The second part addresses the second research question through the score on each rubric indicator and the observation notes. The third part reports the medium's validity and the rubric's reliability from the pilot test as a basis for assessing instrument quality.

A Significant Increase in Total Score with a Moderate Level of Improvement

To gain an overview of students' narrative writing skill attainment before and after the implementation of the instructional intervention, pretests and posttests were administered. The assessment covered four indicators of narrative writing proficiency, which were then aggregated into a total score and converted to a scale of 25–100. Comparing these two sets of results provides an initial insight into the changes in student

achievement following the instruction. Descriptively, the data include the total scores for each measurement, along with their corresponding descriptive statistics. A summary of the descriptive statistics for the pretest and posttest total scores is presented in Table 1.

Table 1. Descriptive Statistics of Total Pretest and Posttest Narrative Writing Scores

Statistic	Pretest	Posttest
N	20	20
Mean	47.50	81.88
Standard deviation	8.70	9.91
Minimum	37.50	68.75
Maximum	62.50	100.00

Table 1 shows that the class mean rose by 34.38 points, from 47.50 on the pretest to 81.88 on the posttest. The lowest posttest score (68.75) was higher than the highest pretest score (62.50), so the two distributions did not overlap. Every student improved, with individual differences ranging from 18.75 to 56.25 points. The distribution of scores widened slightly. The standard deviation rose from 8.70 to 9.91, and the score range widened from 25.00 to 31.25 points, indicating that the magnitude of the gain varied among students. All pretest scores fell between 37.50 and 62.50, so there was still considerable room for improvement. The normality assumption was checked for the pretest scores, posttest scores, and score differences because the paired-samples *t*-test requires normally distributed differences. Table 2 reports the results.

Table 2. Results of the Kolmogorov–Smirnov and Shapiro–Wilk Normality Tests

Data group	Kolmogorov–Smirnov (Lilliefors), *p*	Shapiro–Wilk, *p*	Remarks
Pretest	> 0.200	0.316	Normal
Posttest	> 0.200	0.224	Normal
Score difference	0.142	0.437	Normal

All data groups were normally distributed according to both tests. For the pretest scores, the Kolmogorov–Smirnov *p*-value was greater than 0.200, and the Shapiro–Wilk *p*-value was 0.316. For the posttest scores, the Kolmogorov–Smirnov *p*-value was also greater than 0.200, and the Shapiro–Wilk *p*-value was 0.224. The score differences were also normally distributed ($p = 0.142$ on the Kolmogorov–Smirnov test and $p = 0.437$ on the Shapiro–Wilk test). Because all *p*-values were greater than 0.05, the assumption of normality of the differences for the paired-samples *t*-test was met. The Wilcoxon signed-rank test is still reported as confirmation because the scores were derived from a rubric with a discrete scale. With only 20 students, the normality tests have limited power. Table 3 shows the distribution of students by N-Gain category.

Table 3. Distribution of Students' Narrative Writing Scores by N-Gain Category

N-Gain category	Range of values (g)	f	Percentage
High	$g \geq 0.7$	7	35%
Moderate	$0.3 \leq g < 0.7$	13	65%
Low	$g < 0.3$	0	0%
Total		20	100%

Note. Mean N-Gain = 0.65 (SD = 0.18, range 0.38 to 1.00), which falls in the moderate category.

Table 3 shows that every student improved to at least the moderate category. Seven students (35%) reached the high category, 13 students (65%) reached the moderate category, and no student fell in the low category. The mean N-Gain of 0.65 places the class in the moderate category according to the criteria in the Research Methods section and close to the threshold of the high category (0.70). Two points can be noted. First, the absence of low gains indicates that improvement occurred among students with different starting points. Second, students on average closed nearly two-thirds of the gap between their initial ability and the maximum score, so there is still room for improvement. The significance of the change in total score was tested using a paired-samples t-test and confirmed with the Wilcoxon signed-rank test. Table 4 reports the results together with the effect sizes.

Table 4. Results of the Significance Tests for Total Pretest and Posttest Scores

Test	Statistic	df	p	Effect size
Paired-samples t-test	$t = 13.30$	19	< 0.001	$d_z = 2.97$
Wilcoxon signed-rank	$Z = -3.95$	–	< 0.001	$r = 0.88$

Note. Mean difference = 34.38 points (SD = 11.56), with a 95% confidence interval from 28.96 to 39.79. Cohen's d_z was calculated as t divided by the square root of n, and r as $|Z|$ divided by the square root of N.

As Table 4 shows, the difference between the pretest and posttest was statistically significant, $t(19) = 13.30$, $p < 0.001$, with a standardized effect size of $d_z = 2.97$. The Wilcoxon test led to the same conclusion ($Z = -3.95$, $p < 0.001$, $r = 0.88$), and all 20 students had higher posttest scores. Because the significance value was below $\alpha = 0.05$, the null hypothesis (H_0) was rejected, and the alternative hypothesis (H_a) was accepted. The 95% confidence interval for the mean difference ranged from 28.96 to 39.79 points, so even its lower bound indicated a large gain. This effect size is very large by conventional benchmarks. However, effects from single-group designs tend to be inflated because there is no comparison group, and effects in rigorous educational studies are typically much smaller (Kraft, 2020). These results show a consistent difference in scores after the intervention. They do not show that the medium alone caused all of the change.

Changes in the Four Narrative Writing Indicators

To gain a more detailed understanding of the changes in students' narrative writing skills, a further analysis was conducted on each indicator within the assessment rubric. This analysis aimed to identify which aspects of writing skills changed following the implementation of the instructional intervention, while also examining variation in improvement across indicators. Assessment scores for each indicator were based on an ordinal scale of 1–4; therefore, changes in achievement were analyzed by comparing pretest and posttest results. Differences were assessed using the Wilcoxon signed-rank test. Given that four separate tests were performed, one for each indicator, the significance level was adjusted using the Bonferroni correction, resulting in an α value of 0.0125 (0.05/4). The details regarding scores, N-Gain, and the statistical test results for the four narrative writing skill indicators are presented in Table 5.

Table 5. The Four Narrative Writing Indicators

Indicator	Pretest *M* (*SD*)	Posttest *M* (*SD*)	Difference	N-Gain	Category	*Z*	*r*
Idea development	2.05 (0.69)	3.50 (0.51)	+1.45	0.72	High	−3.80	0.85
Plot	1.85 (0.59)	3.15 (0.49)	+1.30	0.60	Moderate	−3.96	0.89
Vocabulary choice	1.90 (0.64)	3.10 (0.64)	+1.20	0.53	Moderate	−3.62	0.81
Content coherence	1.80 (0.41)	3.35 (0.49)	+1.55	0.69	Moderate	−4.02	0.90

Note. M is the mean, and *SD* is the standard deviation. Difference is the mean posttest score minus the pretest score. N-Gain was calculated for each student, with a maximum score of 4; all p-values < 0.001.

All four indicators improved significantly after the treatment (all $p < 0.001$, below the Bonferroni threshold of 0.0125). Mean scores rose by 1.20 to 1.55 points on the 4-point scale. The largest gains occurred in content coherence (+1.55) and idea development (+1.45), followed by plot (+1.30) and vocabulary choice (+1.20). According to N-Gain, idea development reached the high category (0.72), whereas content coherence (0.69), plot (0.60), and vocabulary choice (0.53) fell in the moderate category. On the pretest, no student scored 4 on any indicator, and four to five students scored 1 on each indicator. On the posttest, no student scored 1, and 10 students scored 4 on idea development, 7 on content coherence, 5 on vocabulary choice, and 4 on plot. All students improved in content coherence; 19 in plot, 18 in idea development, and 16 in vocabulary choice. No student declined on any indicator.

The score gains were accompanied by changes in how students constructed their stories. Observation notes and documentation were examined against the four rubric indicators. Figure 1 and Figure 2 show the two main stages of the lesson: observing and discussing the pictures, and then writing the story.



Figure 1. Observation and Discussion of a Picture-series Media Based on Local Tiakur Stories

Figure 1 shows the first stage, in which the class observed the pictures together under the teacher's guidance. Students sat facing the teacher, who presented the printed picture series at the front of the classroom. This arrangement made the five pictures a shared reference that every student could see and discuss before writing. It likely supported the development of ideas and plot indicators because students first agreed on the characters, the setting of the events, and what happened first and last. Observation notes recorded visibly heightened enthusiasm at this stage, especially when students discussed the fisher-child character, the atmosphere at Tiakur Beach, and the communal clean-up of the beach.



Figure 2. Writing Stories Based on Local Tiakur Stories

Figure 2 shows the second stage, in which students wrote at shared tables with a sheet of paper in front of each student while the teacher supervised the class. The transition

from group discussion to individual writing is important because the posttest writing reflects what each student could do once shared support was available. On the pretest, most students produced only a few simple sentences without a clear plot. After observing and discussing the picture series, they composed stories with an opening, main events, and an ending that depicted the beach becoming clean again. Their vocabulary became more varied, and the logical connections between events became clearer.

Students who had initially written only three to four sentences were able to develop stories of eight to ten sentences in a coherent sequence. The observation notes were consistent with the indicator scores. Idea development was evident in longer stories that drew on the pictures and on students' knowledge of the beach. Plot changed from disconnected sentences to stories with an opening, main events, and an ending. Vocabulary choice became more varied, and coherence improved as events were logically connected. The observations were qualitative and were therefore used to interpret, not to test, the score changes in Table 5.

Validity of the Medium and Reliability of the Rubric in the Pilot Test

All three validators rated the picture-series medium as feasible. The subject-matter expert gave a score of 32 out of 40 (80.00%), the language expert 31 out of 40 (77.50%), and the media expert 26 out of 35 (74.29%). Overall, a score of 89 out of 115 (77.39%) placed the medium in the feasible category. In the subject-matter expert's assessment, all items received a score of 4, including those on closeness to the local context and clarity of the story plot. A score of 3 appeared on three items, namely the characters' expressions and gestures and the durability of the material in the media expert's assessment, and the explanation of local vocabulary in the glossary in the language expert's assessment. These results indicate that the medium is feasible for use. Feasibility is not the same as evidence of effectiveness, which was tested through changes in student scores.

The reliability of the rubric was examined in a pilot scoring of 20 narrative writing samples from third-grade students at SD Inpres Herley by two third-grade teachers working independently. An analysis using the Intraclass Correlation Coefficient (ICC) with a two-way mixed-effects model, an absolute agreement definition, and a single-measures type yielded an ICC of 0.978 with a 95% confidence interval of 0.942 to 0.991, $F(19, 19) = 98.725$, $p < 0.001$. These results indicate very high inter-rater reliability in the pilot test. Table 6 summarizes the results of the analysis.

Table 6. Results of the Inter-rater Reliability Test Based on the ICC in the Rubric Pilot Test

Statistic	Result
Number of pilot writing samples	20
Number of raters	2
ICC model	Two-way mixed-effects
Agreement definition	Absolute agreement
Measurement type	Single measures
ICC	0.978
95% confidence interval	0.942 to 0.991
F	98.725
df_1	19
df_2	19
p value	< 0.001
Mean total score, Rater 1 (*SD*)	12.45 (1.76)
Mean total score, Rater 2 (*SD*)	12.30 (1.89)
Identical total scores	85% (17 of 20 samples)
Identical scores per indicator	96.25% (77 of 80 ratings)

Note. ICC = Intraclass Correlation Coefficient. *SD* = standard deviation. The data came from the rubric pilot test and are separate from the pretest and posttest scores.

The mean total score was 12.45 (*SD* = 1.76) for Rater 1 and 12.30 (*SD* = 1.89) for Rater 2. The two raters' total scores were identical for 17 of the 20 writing samples (85%), and the remaining 3 samples differed by 1 point. Agreement on identical scores per indicator reached 96.25%, namely 77 of 80 ratings. These findings show that the two raters gave very similar scores in the pilot test. The results apply to the two raters involved and to the pilot data.

Discussion

This study produced three main findings. First, students' total narrative writing scores rose significantly from 47.50 to 81.88, and every posttest score was higher than every pretest score. Second, the improvement was moderate according to N-Gain (0.65), with every student reaching the moderate or high category. Third, all four indicators improved significantly, most in content coherence and idea development and least in vocabulary choice, and observations showed more complete stories. This section explains these findings through multimedia learning theory and scaffolding. It also compares the findings with previous studies and states the study's contribution.

Visual media can influence narrative writing ability through several interrelated mechanisms. Pictures and words are processed through separate but interconnected channels. Meaningful learning occurs when learners connect the two (Mayer, 2024). A picture series gives students a visual form of the story that must be translated into words. This process encourages active processing in both channels. Writing also places demands

on limited working memory. Instructional support that externalizes part of that load frees capacity for other tasks (Sweller et al., 2019). Sequential pictures carry the plot, so students do not have to hold the story sequence in memory while searching for words and sentences.

The pictures can also serve as scaffolding, that is, temporary support that helps learners perform beyond what they could achieve on their own (Wood et al., 1976). A meta-analysis of computer-based scaffolding in STEM education showed benefits for cognitive outcomes (Belland et al., 2017). In narrative writing, this support targets text generation, which is a core component of writing quality (Kim & Schatschneider, 2017). Local content adds another condition, because a familiar setting can activate prior knowledge that students use as material for ideas and vocabulary. This study did not test these mechanisms directly. Among early-grade children, reading pictures involves connecting picture content to life experiences, and the coherence and cohesion of oral narratives increase across grade levels (Zhang et al., 2024).

The significant score improvement is consistent with previous studies that have reported that picture series support story writing in elementary school classrooms (Astuti & Rambe, 2024; Lingga & Rambe, 2026; Pujiyanti et al., 2026; Sofiyatun et al., 2026) and that visual images support idea generation and creativity (Listyani, 2019). The results also accord with a review that found story-based multimedia learning beneficial (Wu & Chen, 2020) and with meta-analytic evidence that writing instruction can improve early-grade students' writing (Kim et al., 2021). This study differs from earlier studies in the content of its medium. Earlier studies generally relied on generic picture series.

This study used pictures drawn from the daily life of a fishing community, so students could write about places and activities they knew. Its contribution is to provide preliminary indications that locally grounded picture series can be used in coastal schools in eastern Indonesia, a context rarely reached by international writing research (Graham et al., 2022). The very large standardized effect size should not be taken at face value because a single-group design cannot separate the influence of the medium from maturation, practice, or testing effects (Knapp, 2016; Kraft, 2020). The very high reliability of the rubric in the pilot test supports consistent scoring but does not eliminate the limitations of this design.

Because narrative writing literacy was assessed through four indicators, the changes are read indicator by indicator (Table 5). Idea development and plot relate to the generation and organization of text. Idea development showed the highest relative increase (N-Gain =

0.72), whereas plot rose by 1.30 points (N-Gain = 0.60). Sequential pictures give students a ready-made story framework. That is consistent with evidence that explicit support for story structure strengthens children's narrative skills (Pesco & Gagné, 2017). Oral narrative ability in kindergarten is related to the quality of written stories in first grade (Pinto et al., 2016). Narrative instruction that uses pictures and icons to mark story elements and is then gradually faded has been reported to improve the quality of kindergarten children's narrative writing, with effects lasting three to four weeks (Kirby et al., 2021). Students moved from a few disconnected sentences to stories with an opening, main events, and an ending. This structural change is what the plot indicator is intended to capture.

Vocabulary choice and content coherence relate to language and cohesion. Students discussed the content of each picture before writing, so they could name what they saw and put events into words. Content coherence recorded the largest gain (+1.55 points), experienced by all students, consistent with the clearer connections between events on the posttest. Vocabulary choice rose the least (+1.20 points, with an N-Gain of 0.53) and did not change for four students. One possibility is that the pictures supply the sequence of events directly, whereas new vocabulary requires language input that pictures do not provide. This study did not test that explanation. This observation is consistent with the view that writing involves coordinating several component skills. Stronger idea generation and organization can improve overall text quality (Kim & Schatschneider, 2017). The contribution of this study is to link change to specific rubric indicators both quantitatively and qualitatively. The pattern across indicators is descriptive because it comes from a single group without a comparison and from a rubric with a four-level scale.

The mean N-Gain in the moderate category indicates a fairly large but not complete improvement. Three possible explanations are outlined below, and this study did not test them. The first is the duration of the treatment. One month with six 60-minute sessions may not have been enough to consolidate a skill as complex as narrative writing. The effect of duration on the size of the improvement cannot be assessed because duration was not varied. Studies of picture series in fifth-grade elementary classes can serve as a comparison for future research (Pujiyanti et al., 2026; Sofiyatun et al., 2026). The second explanation is variation in starting points. N-Gain expresses gain relative to room for improvement and is therefore sensitive to initial ability. Its use remains defensible as long as it is interpreted together with the distribution of scores (Coletta & Steinert, 2020; Hake, 1998). A meta-analysis has also shown that the effects of instruction can differ between children with weaker writing skills and those with stronger writing skills (Kim et al., 2021). For this

reason, Table 3 reports the distribution of gains and not only the mean. No student remained in the low category, so improvement appears to have occurred across different levels of initial ability.

The third explanation relates to the medium itself. Five pictures are sufficient to build a simple plot but may offer limited cues for students who need more help with richer vocabulary and more complex sentences. Adding pictures or presenting the picture series in an interactive digital format could extend its benefits, as suggested by a review of digital storytelling and a study of digital media for writing in elementary schools (Marlina et al., 2024; Wu & Chen, 2020). These three conditions, namely duration, variation in starting points, and number of pictures, can be tested directly in future studies with a comparison group and a longer treatment.

The picture series also carried themes of *gotong royong* and care for the coastal environment. The students' stories ended with the beach clean again. Culturally relevant education has been associated with stronger engagement because it connects learning with students' knowledge and community (Aronson & Laughter, 2016). Place-based approaches regard the local environment as a legitimate source of learning (Semken et al., 2017). Local wisdom in children's storybooks has also been discussed as a means of character education (Suryani et al., 2024). Interest in how children understand marine and coastal environments is also growing internationally (Fauville et al., 2019). A children's storybook on aquatic environments used with third- and fourth-grade students in Portugal, including in a town with a fishing tradition, was associated with increased knowledge of threats to aquatic ecosystems (Aurélio et al., 2021). That study measured environmental knowledge, not writing skills. Within this framework, the Tiakur picture series combines literacy goals with a local environmental message. The students' stories contained this theme, but this study did not measure environmental values or attitudes and therefore cannot claim that either was strengthened. Future studies could assess these outcomes directly.

These findings have practical and theoretical implications. For teachers, this study offers a simple procedure that other coastal schools can adapt. The procedure involves collecting local stories, turning them into a short series of pictures, and guiding students through observation, discussion, and individual writing. For curriculum developers, this study can support the use of local environmental and cultural content as a resource for literacy learning rather than as an add-on. Theoretically, the results are consistent with multimedia learning and scaffolding perspectives as applied to locally grounded picture series in elementary schools in eastern Indonesia. These findings add to evidence that has

largely come from other regions (Graham et al., 2022). The study makes three contributions: a locally grounded picture series with an environmental theme, an indicator-based examination of changes in narrative writing, and data from a coastal school that is rarely represented in writing research.

CONCLUSION

This study shows that the narrative writing literacy of third-grade students at SD Negeri Tiakur improved significantly after learning with a picture-series medium based on local Tiakur children's stories. The improvement was in the moderate category and was experienced by all students. At the indicator level, all four indicators improved significantly, with the greatest gains in content coherence and idea development and the least in vocabulary choice. Observations showed more complete and coherent stories, and the pilot scoring showed very high inter-rater reliability for the rubric.

These findings imply that teachers in coastal schools can use local stories and sequential pictures as scaffolding that connects literacy learning with students' environment and with the values of *gotong royong* and care for the coast. The conclusions are limited by a single group at a single school, the absence of a comparison group, the short intervention period, the four-level scale of the indicator rubric, and inter-rater reliability that was examined on pilot data and not on the pretest and posttest scores. Future research should use a comparison group, a longer intervention, a delayed posttest, a larger and more diverse sample, and inter-rater reliability analyses of the pretest, posttest, and individual indicator scores. A digital version of the medium could also be tested to assess the generalizability of the results.

REFERENCES

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research*, *86*(1), 163–206. https://doi.org/10.3102/0034654315582066

Astuti, N. W., & Rambe, R. N. (2024). Pengaruh media gambar berseri terhadap kemampuan menulis siswa kelas rendah. *Aulad: Journal on Early Childhood*, *7*(2), 554–562.

Aurélio, L., França, S., Sequeira, V., Boaventura, D., Correia, M. J., Pinto, B., Amoroso, S., Feio, M. J., Brito, C., & Chainho, P. (2021). Tell a story to save a river: Assessing the impact of using a children's book in the classroom as a tool to promote environmental awareness. *Frontiers in Marine Science*, *8*, 699122.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-

analysis. *Review of Educational Research*, *87*(2), 309–344. https://doi.org/10.3102/0034654316670999

Camping, A., Graham, S., & Harris, K. R. (2023). Writing motives and writing achievement of elementary school students from diverse language backgrounds. *Journal of Educational Psychology*, *115*(7), 1028–1043. https://doi.org/10.1037/edu0000796

Coletta, V. P., & Steinert, J. J. (2020). Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. *Physical Review Physics Education Research*, *16*(1), Article 010108. https://doi.org/10.1103/PhysRevPhysEducRes.16.010108

Davis, T. D., Schrodt, K., & Lee, S. (2024). An Exploration of the Impact of Quality Illustrations in Children's Picture Books on Preschool Student Narrative Ability. *Reading Psychology*, *45*(7), 639–661. https://doi.org/10.1080/02702711.2024.2351480

Fauville, G., Strang, C., Cannady, M. A., & Chen, Y.-F. (2019). Development of the International Ocean Literacy Survey: Measuring knowledge across the world. *Environmental Education Research*, *25*(2), 238–263. https://doi.org/10.1080/13504622.2018.1440381

Graham, S. (2022). A walk through the landscape of writing: Insights from a program of writing research. *Educational Psychologist*, *57*(2), 55–72. https://doi.org/10.1080/00461520.2021.1951734

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The effects of writing on learning in science, social studies, and mathematics: A meta-analysis. *Review of Educational Research*, *90*(2), 179–226. https://doi.org/10.3102/0034654320914744

Graham, S., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2018). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*, *53*(3), 279–304. https://doi.org/10.1002/rrq.194

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkaoui, K., McKeown, D., & Talbot, S. (2019). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research*, *89*(2), 243–284. https://doi.org/10.3102/0034654318814500

Graham, S., Skar, G. B., & Falk, D. Y. (2021). Teaching writing in the primary grades in Norway: A national survey. *Reading and Writing*, *34*(3), 529–563. https://doi.org/10.1007/s11145-020-10080-y

Graham, S., Tavsanli, Ö. F., & Kaldirim, A. (2022). Improving writing skills of students in Turkey: A meta-analysis of writing interventions. *Educational Psychology Review*, *34*(2), 889–934. https://doi.org/10.1007/s10648-021-09639-0

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.

Harahap, D. G. S., Nasution, F., Nst, E. S., & Sormin, S. A. (2022). Analisis kemampuan literasi siswa sekolah dasar. *Jurnal Basicedu*, *6*(2), 2089–2098.

Kim, Y.-S. G., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW).

Journal of Educational Psychology, *109*(1), 35–50. https://doi.org/10.1037/edu0000129

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

Kirby, M. S., Spencer, T. D., & Chen, Y.-J. I. (2021). Oral Narrative Instruction Improves Kindergarten Writing. *Reading \& Writing Quarterly*, *37*(6), 574–591. https://doi.org/10.1080/10573569.2021.1879696

Knapp, T. R. (2016). Why is the one-group pretest–posttest design still used? *Clinical Nursing Research*, *25*(5), 467–472. https://doi.org/10.1177/1054773816666280

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863.

Lingga, M. F., & Rambe, A. H. (2026). Pengembangan Media Gambar Seri Berbasis Puzzle Untuk Meningkatkan Kemampuan Menulis Narasi Siswa Kelas IV MI: Development of a Puzzle-Based Picture Series to Improve Narrative Writing Skills Among Fourth-Grade Students at Madrasah Ibtidaiyah. *Attadrib: Jurnal Pendidikan Guru Madrasah Ibtidaiyah*, *9*(1), 29–38.

Listyani. (2019). The use of a visual image to promote narrative writing ability and creativity. *Eurasian Journal of Educational Research*, *80*, 193–224. https://doi.org/10.14689/ejer.2019.80.10

Marlina, L., Taufik, T., & Hasan, H. (2024). Pengaruh Media Pembelajaran Berbasis Digital Terhadap Kemampuan Menulis Karangan Siswa Sekolah Dasar. *Edu Sociata (Jurnal Pendidikan Sosiologi)*, *7*(2), 134–143.

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*, Article 8. https://doi.org/10.1007/s10648-023-09842-1

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*, Article 100329. https://doi.org/10.1016/j.edurev.2020.100329

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review*, *35*, Article 113. https://doi.org/10.1007/s10648-023-09823-4

Pantaleo, S. (2025). The meaning-making in kindergarten children's visual narrative compositions. *Journal of Early Childhood Literacy*, *25*(2), 336–366. https://doi.org/10.1177/14687984231161114

Pesco, D., & Gagné, A. (2017). Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. *Early Education and Development*, *28*(7), 773–793. https://doi.org/10.1080/10409289.2015.1060800

Pinto, G., Tarchi, C., & Bigozzi, L. (2016). Development in narrative competences from oral to written stories in five- to seven-year-old children. *Early Childhood Research Quarterly*, *36*, 1–10. https://doi.org/https://doi.org/10.1016/j.ecresq.2015.12.001

Pujiyanti, P., Sitorus, F. R., & Ningsih, W. (2026). Implementasi Media Gambar Berseri untuk Peningkatan Kemampuan Menulis Karangan Narasi Siswa Kelas V SD Negeri Lae Saga. *Lingua Franca: Jurnal Bahasa, Sastra, Dan Pengajarannya*, *10*(1), 11–21.

Riduwan, M. B. A. (2022). *Skala pengukuran variabel-variabel penelitian*.

Semken, S., Ward, E. G., Moosavi, S., & Chinn, P. W. U. (2017). Place-based education in geoscience: Theory, research, practice, and assessment. *Journal of Geoscience Education*, *65*(4), 542–562. https://doi.org/10.5408/17-276.1

Skar, G. B., Graham, S., Huebner, A., Kvistad, A. H., Johansen, M. B., & Aasen, A. J. (2024). A longitudinal intervention study of the effects of increasing amount of meaningful writing across grades 1 and 2. *Reading and Writing*, *37*(6), 1345–1373. https://doi.org/10.1007/s11145-023-10460-0

Sofiyatun, S., Wuryani, M. T., & Faida, M. (2026). Implementasi Media Gambar Seri sebagai Sarana Literasi Kreatif dalam Keterampilan Menulis pada Siswa Kelas V Sekolah Dasar. *Jurnal Wawasan Pendidikan*, *6*(1), 253–266.

Sugiharti, R. E., & Wulandari, M. (2017). Peningkatan keterampilan menulis karangan narasi siswa kelas Va dengan menggunakan media gambar seri pada mata pelajaran bahasa Indonesia di SDIT Ad-Damawiyah Cibitung. *Pedagogik: Jurnal Pendidikan Guru Sekolah Dasar*, *5*(2), 1–12.

Suryani, A., Julianti, A., Budiman, A., & Wiyaringtyas, A. R. (2024). Peran kearifan lokal dalam buku cerita untuk pendidikan karakter anak. *Deskomvis: Jurnal Ilmiah Desain Komunikasi Visual, Seni Rupa Dan Media*, *5*(1), 20–27.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100.

Wu, J., & Chen, D.-T. V. (2020). A systematic review of educational digital storytelling. *Computers & Education*, *147*, Article 103786. https://doi.org/10.1016/j.compedu.2019.103786

Zhang, R., Brown, S., & Lysaker, J. (2024). A Cross-Case Study of Children's Oral Narratives and Image Reading of a Wordless Picturebook. *Reading \& Writing Quarterly*, *40*(3), 191–210. https://doi.org/10.1080/10573569.2023.2184737