



Exploring Changes In Historical Literacy Among Grade IV Elementary Students Following Local Wisdom–Integrated History Learning: A Small-Scale Pilot Study

Ridho Sugihandika^{1*}, M. Ikhsan Riski¹, Joni Prima¹, Dwi Arifah Natisya¹, Jalilla¹,
Widya Aprilia¹, Nurridha Aulia¹
Samudra University¹, Langsa, Indonesia
Email: handika116600@gmail.com^{*1}

Article History

Submitted: 2026-07-27
Revised: 2026-08-01
Accepted: 2026-08-12
Published: 2026-08-15

Keywords:

local wisdom; historical literacy; history learning; primary school; pilot study

Abstract

This pilot study explores changes in historical literacy among grade IV elementary school students following the integration of local wisdom in history learning. Indonesia's relatively low reading literacy performance (PISA 2022) provides a broader literacy context; however, historical literacy represents a more specific disciplinary competency requiring targeted investigation. This study employed a one-group pretest–posttest pre-experimental design with eight grade IV students in one elementary school, using total sampling. The instrument was a 15-item essay-based historical literacy test contextualized with local wisdom, validated by three experts (Aiken's $V = 0.86$) and tested for reliability (test–retest $r = 0.74$). Two trained raters scored the essays using an analytical rubric, with inter-rater reliability established prior to analysis. The six-session intervention integrated local traditions, folklore, and customary practices relevant to the grade IV curriculum. Prior to inferential testing, the normality of difference scores was examined using the Shapiro–Wilk test and Q–Q plots. Paired-sample analysis (two-tailed) indicated a statistically significant increase from pretest ($M = 41.00$, $SD = 4.57$) to posttest ($M = 66.25$, $SD = 6.02$), with a mean normalized gain of 0.432 (medium category). A sensitivity analysis using the Wilcoxon signed-rank test corroborated this pattern. Given the absence of a control group and the very small sample, these findings should be interpreted as preliminary evidence that local wisdom integration may support historical literacy development. Further research with control groups and larger samples is needed before causal or effectiveness claims can be made. The practical implication suggests that teachers might consider local resources as a potential context for history learning, pending stronger empirical validation.



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution ShareAlike 4.0 International License (CC-BY-SA-4.0)

INTRODUCTION

Historical literacy is a fundamental competency that enables individuals to understand past events, evaluate historical sources, and make evidence-based interpretations of historical phenomena. It encompasses disciplinary skills such as chronological reasoning, sourcing, contextualization, and corroboration (Wineburg, 2001). In the Indonesian context, concerns about students' literacy capabilities have been highlighted by international assessments. The Organisation for Economic Co-operation and Development (OECD) through the Programme for International Student Assessment (PISA) 2022 reported Indonesia's reading literacy score as 359, far below the OECD average of 476 (OECD, 2023). While this reflects general reading comprehension challenges, historical literacy constitutes a distinct disciplinary competency involving the interpretation of evidence, evaluation of sources, and construction of historical arguments (Wineburg, 2001). Nevertheless, low general literacy may constrain students' capacity to engage with historical texts, underscoring the need for domain-specific interventions beginning at the elementary level.

In the Indonesian elementary school curriculum (Kurikulum Merdeka), history is not taught as a standalone subject but is integrated within the subject Ilmu Pengetahuan Alam dan Sosial (IPAS) a combined natural and social science course. Consequently, historical content at the grade IV level is embedded within broader IPAS themes such as "Kehidupan Masyarakat di Lingkungan Sekitar" and "Perkembangan Teknologi dan Kebudayaan." The term "history learning" in this study therefore refers specifically to the history-related components within the IPAS framework, and the intervention was designed to align with the IPAS learning objectives for grade IV rather than a separate history syllabus. The lesson plans were developed as part of the IPAS instructional design, and the historical literacy test items were contextualized within the IPAS content scope for this grade level.

At the national level, the Minimum Competency Assessment (Asesmen Kompetensi Minimum, AKM) indicates that Indonesian students' literacy skills require significant improvement. Many elementary school students struggle to understand historical concepts contextually and to apply them to explain events in their immediate environment. Research by Sriwati (2023) identified historical misconceptions among Indonesian students, including confusion between the 1965 September 30 Movement (G30S) and the 1948 Madiun PKI Rebellion, with some students even writing that Sukarno was a descendant of the deity Vishnu. While this evidence reflects broader challenges in history education at various levels, it highlights the urgent need for foundational historical reasoning skills beginning at the elementary level. Such findings suggest that abstract, memorization-heavy approaches may fail to develop genuine historical understanding.

One promising approach to address this issue is the integration of local wisdom in history learning. In this study, local wisdom is defined operationally as the knowledge, values, beliefs, and customary practices transmitted within a local community, including oral traditions, folklore, customary norms, and community narratives. This is distinguished from local history, which refers to documented events, artifacts, and narratives specific to a geographical area. While related, local wisdom emphasizes living cultural practices and community knowledge, whereas local history emphasizes documented past events. The pedagogical rationale is that local wisdom provides culturally meaningful contexts that encourage engagement with local historical sources, thereby supporting chronological reasoning, source evaluation, and evidence-based interpretation.

Previous studies offer relevant groundwork. Mursidi and Alit (2025) found that history learning based on local wisdom can increase students' historical awareness and critical thinking skills. Sariyatun and Marpelina (2024) reported that integrating local history into the curriculum helps students build national identity. However, most of these studies employ qualitative designs, and quantitative empirical evidence on the direct relationship between local wisdom integration and elementary students' historical literacy remains limited. A systematic review by Fikri et al. (2023) concluded that while integrating local history into Indonesian history education is promising, more quantitative studies are needed to validate its effects empirically. Similarly, Rispan and Sudrajat (2020) highlighted the need for stronger quantitative approaches in examining the relationship between local wisdom inheritance and student learning outcomes.

A review of prior literature reveals specific gaps. First, the majority of studies on local wisdom in history learning are conducted at the secondary or university level, while quantitative investigations at the elementary school level remain rare. Second, existing research tends to focus on the development of learning products such as modules or teaching materials rather than on empirical testing of variable relationships. Third, quantitative empirical testing of the relationship between local wisdom integration and elementary students' historical literacy is scarce, particularly at the primary school level. This pilot study is positioned as a preliminary investigation to explore whether this relationship warrants further testing with stronger, quasi-experimental designs.

The novelty of this study lies in its exploratory examination of how local wisdom integration relates to three specific dimensions of historical literacy: chronological understanding, source analysis, and evidence interpretation among grade IV students. This population–construct–context combination has not been quantitatively investigated in the Indonesian literature. Given the small sample, this study is explicitly positioned as a pilot investigation intended to generate hypotheses for future confirmatory research rather than to establish definitive causal claims.

The three dimensions of historical literacy targeted in this study (1) chronological understanding, (2) evaluation and analysis of historical sources, and (3) interpretation of historical data and evidence

are derived from Wineburg's (2001) framework of historical thinking. At the grade IV level (typically ages 9–10), students are developing concrete operational reasoning, enabling them to sequence events, distinguish between fact and opinion in simple sources, and draw basic inferences from pictorial and textual evidence. Therefore, the instrument was adapted to match this developmental stage by using locally contextualized narratives and visual sources rather than abstract historiographical texts.

The intervention draws on contextual learning theory, which posits that knowledge construction becomes more meaningful when new concepts are linked to familiar contexts (Purwoto & Kurniawan, 2025). When historical concepts are presented through local wisdom that is familiar to students, they may activate prior knowledge and integrate new learning into existing cognitive schemas. However, a critical historical literacy perspective also requires that students learn to interrogate sources including local wisdom narratives by asking who created an account, for what purpose, and what other evidence supports or contradicts it (Wineburg, 2001).

Given the exploratory nature of this pilot study, the following research questions were formulated without a directional causal hypothesis: (1) To what extent do grade IV students' historical literacy scores change following the integration of local wisdom in history learning? (2) What is the magnitude of score improvement, as measured by normalized gain, following the six-session intervention?

METHOD

Research Design

This study employed a pre-experimental one-group pretest–posttest design as a small-scale pilot investigation. This design was chosen due to field constraints involving a single class with eight students, making the formation of a control group unfeasible. Consequently, the study does not claim causal effectiveness but rather explores preliminary patterns that warrant further testing with stronger designs. The design is symbolized as $O_1 X O_2$, where O_1 is the pretest, X is the six-session treatment integrating local wisdom, and O_2 is the posttest (Marsden & Torgerson, 2012). Although this design has well-known limitations in controlling internal validity threats such as maturation, history, and testing effects, it remains relevant for generating initial evidence under naturalistic conditions (Bierer et al., 2025).

It is important to note that in the Indonesian elementary curriculum (Kurikulum Merdeka), history is taught as part of the integrated subject Ilmu Pengetahuan Alam dan Sosial (IPAS) rather than as a discrete course. The intervention was therefore designed to fit within the IPAS learning objectives for grade IV, specifically the unit on "Kehidupan Masyarakat di Lingkungan Sekitar." The six sessions were structured as IPAS lessons, and the assessment instrument was aligned with the IPAS competency standards for historical thinking at this level.

Setting and Participants

The study was conducted at a public elementary school in a rural district of Aceh Province, Indonesia. The school serves an Acehnese Muslim community with rich oral traditions, customary practices (adat), and historical consciousness shaped by both the pre-Islamic maritime heritage and the later Islamic sultanate period. The population comprised all grade IV students enrolled in the 2024/2025 academic year. The sample consisted of the entire population ($N = 8$ students), obtained through total sampling. Total sampling was used because the number of students was very limited, and no additional selection process was applied. The results are therefore specific to this single class and should not be generalized to the broader population of elementary school students.

Ethical Considerations

This study involving minor participants adhered to the ethical principles of voluntary participation, parental informed consent, student assent, and data confidentiality. Written permission was obtained from the school principal prior to data collection. Informed consent was secured from parents or guardians, and verbal assent was obtained from each student after explaining the study in age-appropriate language. Students were informed that their participation would not affect their grades and that they could withdraw at any time. Participant data were anonymized using student codes (S1–S8), and all test materials were stored securely with access limited to the research team. The study followed the ethical standards stipulated in the Regulation of the Minister of Education, Culture, Research, and Technology of the Republic of Indonesia No. 54 of 2015 concerning Standards of Higher Education.

Instrument

The data collection instrument was a historical literacy essay test consisting of 15 items. The test was designed based on a historical literacy framework comprising three competencies: (1) understanding past events chronologically, (2) evaluating and analyzing historical sources, and (3) interpreting historical data and evidence (Wineburg, 2001). All items were contextualized with local wisdom relevant to the grade IV curriculum.

Table 1. Instrument Blueprint for Historical Literacy Test

Dimension	Indicator	Item No.	Maximum Score
Chronological understanding	Sequencing local historical events; identifying temporal relationships	1–5	30
Source analysis	Distinguishing fact, opinion, and fiction in local narratives; identifying source origin and purpose	6–10	35
Interpretation of evidence	Drawing conclusions from multiple local sources; making simple historical inferences	11–15	35
Total		15	100

The maximum possible score was 100. Raw scores from the 15 essay items were converted to a 0–100 scale proportionally. Each item was scored using an analytical rubric with criteria for each dimension, ranging from 0 to the item-specific maximum. The sum of item scores was then converted to the 0–100 scale using the formula: $\text{Converted Score} = (\text{Raw Total} / 100) \times 100$.

Validity. Content validity was established by three expert validators: two history education lecturers and one classroom teacher with at least 10 years of teaching experience. Validation used the Aiken's V coefficient, yielding a value of 0.86, which exceeds the minimum acceptable threshold and indicates very high content validity (Tadeko et al., 2025).

Reliability. A test–retest reliability check was conducted on an equivalent group of students (not the study subjects) with a two-week interval, yielding a coefficient of 0.74, indicating acceptable reliability (Rusilowati et al., 2016). In addition, a limited trial was conducted to analyze item difficulty and discriminating power, resulting in 13% easy items, 67% medium items, and 20% difficult items.

Scoring Reliability. Essay responses were scored by two trained raters using the analytical rubric. Both raters were blinded to whether responses came from the pretest or posttest to minimize scoring bias. Inter-rater reliability was calculated using Cohen's kappa, yielding $\kappa = 0.85$, indicating substantial agreement. Discrepancies were resolved through discussion prior to finalizing scores.

Test Forms. To mitigate the risk of "teaching to the test," the posttest used a parallel equivalent form with different item contexts but comparable difficulty and dimension coverage. While both tests were contextualized with local wisdom, the specific local narratives and sources differed between the pretest and posttest (e.g., pretest used folklore from one gampong, whereas posttest used customary practices from a neighboring gampong). Additionally, transfer items assessing historical literacy with unfamiliar local contexts were included to distinguish genuine skill acquisition from content familiarity.

Intervention (Treatment)

The treatment consisted of six sessions of history learning that integrated local wisdom. Each session lasted 2×35 minutes. Learning materials were associated with local wisdom contexts such as local traditions, historical relics in the surrounding environment, folklore, and customary practices relevant to the grade IV history curriculum.

Epistemological Treatment of Local Wisdom Sources

The intervention deliberately treated different forms of local wisdom as distinct categories of historical sources, each requiring specific evaluative approaches. This epistemological differentiation was designed to align with the sourcing and corroboration dimensions of historical literacy (Wineburg, 2001). The four types of local wisdom used were categorized as follows:

1. Oral traditions and folklore (e.g., origin myths, folktales, hikayat) were positioned as subjective narrative sources. Students were taught to ask: Who told this story? When was it recorded?

Whose perspective does it represent? These sources were treated as valuable but requiring corroboration.

2. Material relics and historical artifacts (e.g., colonial-era cannons, gravestones, traditional shipbuilding tools) were treated as material primary sources. Students practiced physical description, dating hypotheses, and inference from physical evidence.
3. Customary practices and rituals (e.g., peusijuk ceremonies, agricultural calendars, meunasah communal activities) were analyzed as living traditions that have changed over time. Students examined how practices today might differ from historical descriptions and what this reveals about continuity and change.
4. Written local documents (e.g., village records, family genealogies, naskah kuno manuscripts) were introduced as documentary sources when available, allowing comparison between oral and written accounts of the same event.

This classification was explicitly taught to students during Meeting 1 and reinforced in subsequent sessions. By treating local wisdom not as a monolithic "truth" but as a set of source types with varying reliability and perspective, the intervention aimed to foster genuine historical reasoning rather than mere cultural memorization.

Table 2. Intervention Design: Local Wisdom Integration Across Six Sessions

Meeting	Local Wisdom Content	Historical Literacy Target	Learning Activity
1	Origin myth of the gampong	Chronological sequencing	Students construct a timeline from an oral narrative (hikayat) related to the founding of their village
2	Historical relics in the vicinity (e.g., colonial-era gravestones, old mosque foundations)	Source identification and description	Field observation and documentation of artifacts; students record physical characteristics and hypothesize origins
3	Local customary law (adat)	Source evaluation	Comparing oral transmission and written records of the customary law; identifying discrepancies
4	Acehnese folktale (hikayat)	Distinguishing fact, opinion, and fiction	Students categorize elements of the folktale into historical facts, moral teachings, and fictional embellishments
5	Traditional ceremony peusijuk	Chronological reasoning and contextualization	Sequencing the ritual procedures historically and connecting them to agricultural or social calendars
6	Community elder interview (transcribed excerpts)	Synthesis and interpretation	Integrating oral testimony with written records to construct a short narrative about the local resistance against colonial forces

Before treatment, students completed the pretest (O_1). After six sessions, they completed the posttest (O_2) with the equivalent instrument. Data collection was carried out by the researcher with guidance from the classroom teacher to ensure that students worked independently.

Data Analysis

Data were analyzed using descriptive and inferential statistics with IBM SPSS Statistics version 26. Descriptive analysis included means, standard deviations, minimum and maximum scores, and percentages to describe pretest and posttest performance.

Assumption checking. Prior to parametric testing, the normality of difference scores (posttest minus pretest) was examined using the Shapiro–Wilk test and visual inspection of Q–Q plots. The Shapiro–Wilk test yielded $W = 0.831$, $p = 0.060$, indicating that the difference scores did not significantly deviate from normality at the conventional $\alpha = 0.05$ level. Visual inspection of Q–Q plots further supported the approximate normality of the difference distribution.

Inferential analysis. A paired-sample t-test was conducted using a two-tailed significance level of $\alpha = 0.05$. Although a directional increase was anticipated based on contextual learning theory, a two-tailed test was employed to provide more conservative and widely accepted estimates. As a sensitivity analysis, the Wilcoxon signed-rank test (nonparametric alternative) was also conducted to verify whether the conclusion remained robust regardless of distributional assumptions.

Effect size. Effect sizes were calculated using Cohen's d_z (standardized mean difference for paired data, using the standard deviation of difference scores as the denominator) and Hedges' g with correction for small-sample bias. The formulas used were:

- Cohen's $d_z = M_{\text{diff}} / SD_{\text{diff}}$
- Hedges' $g = \text{Cohen's } d_z \times [1 - (3 / (4 \times df - 1))]$

Given the extremely small sample ($n = 8$), these effect size estimates are highly unstable and should not be generalized beyond this specific context. The very large values primarily reflect low variability in difference scores rather than universally massive treatment effects. Confidence intervals (95%) for effect sizes were computed to acknowledge this uncertainty.

Normalized gain (N-gain). N-gain was calculated using Hake's (1998) formula:

$$g = (\text{posttest} - \text{pretest}) / (\text{maximum score} - \text{pretest})$$

While N-gain is commonly used in physics education research, its application in this study serves as a standardized index of score improvement relative to remaining possible gain, facilitating comparison with future studies on historical literacy interventions. The interpretation refers to Hake's criteria: $g < 0.3$ (low), $0.3 \leq g < 0.7$ (medium), and $g \geq 0.7$ (high). The average N-gain of 0.432 (medium category) may reflect factors such as treatment duration, instrument difficulty, baseline student ability, and the conceptual complexity of grade IV historical material, rather than sample size per se.

RESULTS AND DISCUSSION

Results

Historical literacy pretest and posttest data from the eight grade IV students are presented in Table 3. The pretest score showed a mean of 41.00 ($SD = 4.57$) with a range of 35 to 48. Without a validated cutoff for historical literacy proficiency levels at this grade, this score is reported descriptively rather than assigned a categorical label. After the six-session intervention, the posttest score increased to a mean of 66.25 ($SD = 6.02$) with a range of 58 to 75. All students showed an increase in scores from pretest to posttest, with a mean difference of 25.25 points.

Table 3. Summary of Descriptive Statistics of Pretest and Posttest Scores

Statistics	Pretest	Posttest
N	8	8
Mean	41.00	66.25
Standard Deviation	4.57	6.02
Minimum	35	58
Maximum	48	75

Individual student scores and N-gain values are presented in Table 4. All students obtained N-gain scores in the medium category ($0.3 \leq g < 0.7$), with a class average of 0.432. Student S3 obtained the highest N-gain (0.545), and Student S4 obtained the lowest (0.354). No student fell into the low ($g < 0.3$) or high ($g \geq 0.7$) categories.

Table 4. Individual Pretest, Posttest, and N-Gain Scores

Student	Pretest	Posttest	Difference	N-Gain	Category
S1	42	68	26	0.448	Medium
S2	38	62	24	0.387	Medium
S3	45	75	30	0.545	Medium
S4	35	58	23	0.354	Medium

S5	48	72	24	0.462	Medium
S6	40	65	25	0.417	Medium
S7	36	60	24	0.375	Medium
S8	44	70	26	0.464	Medium
Mean	41.00	66.25	25.25	0.432	Medium

Note. Maximum possible score = 100. N-gain = (posttest – pretest) / (100 – pretest).

Dimension-Specific Analysis. To examine whether the observed improvement was uniform across the three dimensions of historical literacy, descriptive statistics were calculated for each subscale. The results are presented in Table 5.

Table 5. Descriptive Statistics and N-Gain by Historical Literacy Dimension

Dimension	Pretest M (SD)	Posttest M (SD)	Mean Difference	N-Gain	Category
Chronological understanding	12.50 (1.77)	20.75 (2.12)	8.25	0.477	Medium
Source analysis	14.25 (1.83)	22.75 (2.71)	8.50	0.415	Medium
Evidence interpretation	14.25 (1.04)	22.75 (1.28)	8.50	0.411	Medium

The largest improvement was observed in the chronological understanding dimension, with an N-gain of 0.477. This may be attributable to the concrete nature of chronological sequencing tasks, which align well with the concrete operational reasoning typical of grade IV students. Conversely, the evidence interpretation dimension showed the smallest gain (N-gain = 0.411), possibly because evaluating source reliability and drawing inferences from multiple sources requires more abstract critical thinking that may exceed the developmental stage of 9–10-year-olds, or because the intervention provided fewer explicit opportunities for source criticism compared to narrative construction.

Normality and Sensitivity Analysis. The Shapiro–Wilk test on difference scores yielded $W = 0.831$, $p = 0.060$, supporting the assumption of normality. Q–Q plots showed approximate linearity with minor deviations at the tails, consistent with the small sample size. As a sensitivity analysis, the Wilcoxon signed-rank test confirmed a statistically significant increase, $Z = 2.521$, $p = 0.012$, consistent with the parametric result.

The paired-sample t-test (two-tailed) indicated a statistically significant difference between pretest and posttest scores, $t(7) = 32.646$, $p < 0.001$. The mean difference was 25.25, with a 95% confidence interval [23.42; 27.08], which does not include zero. The effect size calculated using Cohen's d_z yielded 11.542, and Hedges' g yielded 10.260, with 95% CIs of [10.71; 12.38] and [9.52; 11.00] respectively. These values appear extreme and reflect the very small standard deviation of difference scores ($SD = 2.19$) combined with the small sample. They should not be interpreted as indicating a universally massive effect but rather as unstable estimates requiring replication in larger samples.

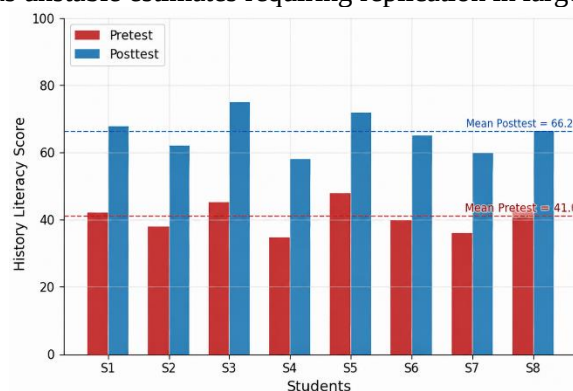


Figure 1. Comparison of Pretest and Posttest Scores of Historical Literacy for Grade IV Students

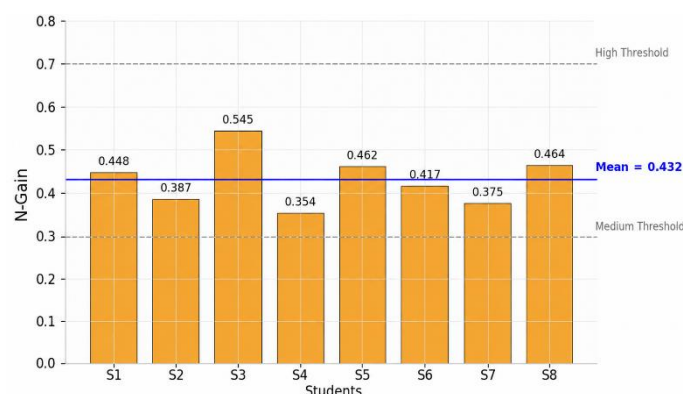


Figure 2. Individual N-Gain Scores of Grade IV Students

Discussion

This pilot study found a statistically significant increase in historical literacy scores following a six-session intervention integrating local wisdom into history learning among eight grade IV students. However, given the one-group design and very small sample, this increase cannot be attributed solely to the intervention. Alternative explanations including testing effects, student maturation, familiarity with the assessment format, and concurrent learning experiences remain plausible (Marsden & Torgerson, 2012). The findings are therefore best understood as preliminary evidence that local wisdom integration may be associated with score improvements, rather than as confirmation of causal effectiveness.

The observed improvement aligns with the theoretical proposition that contextualizing abstract historical concepts through familiar local wisdom can facilitate knowledge construction (Purwoto & Kurniawan, 2025). When students encounter historical concepts through origin myths, local relics, and customary practices embedded in their daily environment, they may activate prior knowledge and integrate new information into existing cognitive schemas. This interpretive mechanism is consistent with contextual learning theory. However, the present design does not isolate this mechanism from other threats to internal validity.

A critical consideration is whether students developed genuine historical reasoning or merely recalled local narratives. Historical literacy requires not only engagement with sources but also the ability to question them (Wineburg, 2001). The intervention exposed students to local folklore and elder testimony, yet an essential dimension of historical thinking source criticism was not explicitly emphasized. Future iterations should incorporate activities where students interrogate local wisdom narratives by asking: Who created this account? When? For what purpose? What other evidence supports or contradicts it? Positioning local wisdom as an object of historical inquiry rather than merely a vehicle for content delivery would strengthen both the intervention and its alignment with disciplinary standards.

An untapped pedagogical opportunity in this intervention was the explicit incorporation of cognitive conflict between folklore and other forms of evidence. Historical literacy, at its core, requires students to navigate competing accounts and evaluate which evidence is more reliable (Wineburg, 2001). Local wisdom particularly oral traditions and folklore often contains narratives that may diverge from archaeological, documentary, or scientific evidence. For example, oral tradition in the students' gampong claims the settlement was founded in the 17th century by a Sufi saint (syekh) who miraculously cleared the forest overnight and established the first mosque. However, archaeological surveys in the broader region have uncovered pre-Islamic Hindu-Buddhist stone inscriptions (batu bersurat) and 9th-century Chinese ceramic shards, indicating continuous human settlement centuries before the arrival of Islam in Aceh. Rather than treating the oral tradition as an unquestioned repository of truth, the intervention could have more explicitly positioned such contradictions as learning opportunities for corroboration. Students could have been asked: Does the folklore match the artifact? If not, why might they differ? Which source is more reliable for answering what question? This approach would have more closely aligned the intervention with the disciplinary practice of historical inquiry, where sources

are interrogated rather than accepted. Future iterations of this study should deliberately design activities that pit oral tradition against material or documentary evidence, thereby scaffolding students' ability to evaluate reliability, detect bias, and construct evidence-based arguments skills that lie at the heart of historical literacy.

Furthermore, the use of local wisdom may have increased engagement, but it also raises the possibility that posttest gains partly reflect cultural familiarity rather than transferable historical literacy skills. The parallel test forms and transfer items were included to mitigate this risk, yet the small sample precludes robust conclusions about whether gains represent domain-general skill acquisition or context-specific recall. Future studies should include a greater proportion of transfer items with entirely unfamiliar local contexts to disentangle these possibilities.

The average N-gain of 0.432 (medium category) suggests that, within the constraints of this pilot, students improved but did not reach ceiling levels. This magnitude may reflect the treatment duration, the cognitive demands of essay-based assessment for grade IV students, baseline ability levels, and the conceptual complexity of the targeted historical literacy dimensions. It is methodologically inappropriate to attribute the N-gain magnitude to the small sample size; rather, instructional and measurement factors offer more plausible explanatory accounts.

Limitations. The primary limitations are the pre-experimental design and the very small sample ($n = 8$). The absence of a control group makes it impossible to rule out history, maturation, and testing effects. The small sample compromises the precision and stability of effect size estimates and severely limits generalizability, rather than simply reducing statistical power. With an extremely large observed effect size, statistical significance was achieved despite the small sample, but this does not imply that the effect is robust or replicable (Bierer et al., 2025). Additionally, the intervention was delivered by a single teacher in one school, raising questions about implementation fidelity and contextual specificity. The six-session duration may also have been insufficient to produce more substantial changes in a complex competency such as historical literacy.

Implications for Practice and Research. Pending stronger evidence, teachers might consider local wisdom as one of several potential tools for contextualizing history learning. The approach requires minimal financial resources because it leverages existing community knowledge. However, curriculum developers should await confirmatory quasi-experimental evidence before mandating large-scale integration. For researchers, the priority is to move beyond one-group designs. Quasi-experimental studies with control groups, larger and more diverse samples, and longer interventions are needed to test whether local wisdom integration causally improves historical literacy. Longitudinal designs could illuminate developmental trajectories, while mixed-methods approaches could reveal the cognitive and affective mechanisms underlying any observed gains. Explicit inclusion of source criticism activities within the local wisdom framework represents a particularly promising avenue for strengthening both pedagogical practice and research validity.

CONCLUSION

This pilot study provides preliminary evidence that grade IV students' historical literacy scores increased following a six-session intervention integrating local wisdom into history learning. The observed improvement, reflected in higher posttest scores and a medium normalized gain, suggests that local-context integration may support historical literacy development in young learners. However, the one-group pre-experimental design, absence of a control group, and very small sample ($n = 8$) preclude any causal inference. The results are best understood as generating hypotheses for future research rather than confirming intervention effectiveness.

Core finding: Students showed a statistically significant score increase across the aggregated historical literacy measure following the intervention, though the mechanism whether treatment effect, testing, maturation, or a combination of factors remains undetermined.

Meaning: Local wisdom holds potential as a culturally meaningful context for elementary history instruction, but its unique contribution above and beyond standard instruction has not been established.

Strength of evidence: Weak, due to the pre-experimental design and very small sample.

Limitations: No control group; $n = 8$; single school and teacher; short intervention duration; potential teaching-to-the-test risk; unstable effect size estimates.

Future research: Quasi-experimental designs with control groups, larger and more diverse samples, longer interventions, and explicit inclusion of source criticism activities are needed to test whether local wisdom integration causally improves historical literacy and to understand the cognitive mechanisms underlying any observed gains. Dimension-specific analyses should also be reported to identify which aspects of historical literacy are most responsive to the intervention.

REFERENCES

- Bierer, S. B., Beck Dallaghan, G., Borges, N. J., Brondfield, S., Fung, C. C., Huggett, K. N., Teal, C. R., Thammasitboon, S., & Colbert, C. Y. (2025). Moving beyond simplistic research design in health professions education: What a one-group pretest–posttest design will not prove. *MedEdPORTAL*, 21, 11527. https://doi.org/10.15766/mep_2374-8265.11527
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Fikri, A., Syahza, A., & Putra, Z. (2023). Systematic review of integration of local history in history education in Indonesia based on learning technology. *Al-Ishlah Jurnal Pendidikan*, 15(2), 1434–1443. <https://doi.org/10.35445/alishlah.v15i2.2364>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, 38(5), 583–616. <https://doi.org/10.1080/03054985.2012.731208>
- Mursidi, A., & Alit, D. M. (2025). Transformation of history education based on local wisdom to foster historical awareness in Generation Z. *Jurnal Iqra': Kajian Ilmu Pendidikan*, 10(3), 1–20. <https://doi.org/10.25217/ji.v10i3.6525>
- Organisation for Economic Co-operation and Development. (2023). *PISA 2022 results: Indonesia – Country note*. OECD Publishing. <https://www.oecd.org/pisa/countrynotes/IDN%20PISA%202022%20Country%20Note%20English.pdf>
- Purwoto, S., & Kurniawan, G. F. (2025). Integration of digital storytelling in project-based learning to develop students historical literacy: Literature review. *Jurnal Pendidikan IPS*, 15(3). <https://doi.org/10.37630/jpi.v15i3.3236>
- Rispan, R., & Sudrajat, A. (2020). Pewarisan nilai-nilai kearifan lokal Kalosara dalam pembelajaran sejarah untuk membangun karakter siswa. *HISTORIA: Jurnal Program Studi Pendidikan Sejarah*, 8(1), 61. <https://doi.org/10.24127/hj.v8i1.2254>
- Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an instrument of scientific literacy assessment on the cycle theme. *International Journal of Environmental and Science Education*, 11(12), 5718–5727.
- Sariyatun, S., & Marpelina, L. (2024). Strengthening identity through the integration of local history in the learning curriculum. *HISTORIA: Jurnal Program Studi Pendidikan Sejarah*, 12(1), 297. <https://doi.org/10.24127/hj.v12i1.9382>
- Sriwati. (2023). Dilema miskonsepsi dalam buku teks sejarah Indonesia. *Edukatif: Jurnal Ilmu Pendidikan*, 5(2), 768–775. <https://doi.org/10.31004/edukatif.v5i2.4801>
- Tadeko, N., Rosana, D., & Purwastuti, L. A. (2025). Weighted-scoring scientific literacy instrument: Development and validation of a contextualized assessment for junior high school students. *European Public & Social Innovation Review*, 11, 1–21. <https://doi.org/10.31637/epsir-2026-2149>
- Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Temple University Press.
- Yasa, I. W. P., Syahrin, A. A., Arta, K. S., & Pageh, I. M. (2025). Revitalizing local culture-based history learning: Opportunities and challenges toward transformative deep learning practices. *SHS Web of Conferences*, 221, 01007. <https://doi.org/10.1051/shsconf/202522101007>