

Source-Based Inquiry Learning to Improve University Students' Historical Literacy in South Asian History

Guntur Arie Wibowo^{1,*}, Riyadi², Hendra Kurniawan³, Ipong Jazimah⁴, & Ahmad Sharani⁵

¹Elementary School Teacher Education, Universitas Negeri Surabaya, Indonesia.

²History Education Study Program, Universitas Negeri Surabaya, Indonesia.

³History Education Study Program, Universitas Sanata Dharma, Indonesia.

⁴History Education Study Program, Universitas Muhammadiyah Purwokerto, Indonesia

⁵Science, Technology, Engineering, and Mathematics, Institut Pendidikan Guru Kampus Darulaman, Malaysia

*Corresponding email: gunturwibowo@unesa.ac.id

Received: 03 July 2026

Accepted: 17 September 2026

Published: 07 October 2026

Abstract: This study examined the effectiveness of source-based inquiry learning in improving university students' historical literacy in a South Asian History course. Historical literacy was conceptualized through students' abilities in sourcing, contextualization, corroboration, and evidence-based historical argumentation. A quantitative quasi-experimental pretest-posttest control group design involved 58 fourth-semester history education students. The experimental group comprised 29 students who received source-based inquiry learning, while the control group received conventional instruction. Data were collected using a source-based historical literacy essay test and analyzed through descriptive statistics, an independent-samples t-test, analysis of covariance, and normalized gain analysis. The experimental group achieved a higher posttest mean than the control group (78.93 vs. 64.90) and a greater improvement (28.15 vs. 14.48 points). The between-group difference was statistically significant, $t(56) = 6.79$, $p < 0.001$. Analysis of covariance confirmed a significant treatment effect after controlling for pretest scores, $F(1, 55) = 110.82$, $p < 0.001$, with a large effect size (0.668). The experimental group obtained a moderate normalized gain of 0.58, compared with 0.30 in the control group, which lay at the lower boundary of the low category. Sourcing and contextualization showed the greatest improvement. Source-based inquiry learning effectively strengthens historical literacy and disciplinary reasoning. Its integration into university history education can support more analytical, contextual, critical, and evidence-centered learning.

Keywords: source-based inquiry learning, historical literacy, history education, primary historical sources, historical thinking.

Article's DOI: <https://doi.org/10.23960/jpp.v16i4.pp1954-1974>

■ INTRODUCTION

History is no longer understood simply as a body of facts that students are expected to remember. Contemporary history education now pays greater attention to how historical knowledge is produced, questioned, and justified. This change has brought historical thinking, historical consciousness, and historical literacy to the center of the learning process. Rather than accepting accounts of the past as fixed narratives, students are expected to examine a source's origin, assess its credibility, reconstruct the context

in which it was produced, compare competing accounts, and build interpretations they can defend with evidence. Recent scholarship reflects this disciplinary orientation, showing that historical thinking and historical literacy are increasingly treated as essential for developing critical, reflective, and evidence-based understandings of the past (Gibson et al., 2025). This need is particularly pressing in an information environment where historical claims circulate rapidly, and public memory, political interpretation, and simplified narratives often

compete (Ana Isabel & Barca, 2024; Barreiro & Perez-Manjarrez, 2025; Tribukait, 2024).

For students in history education programs, these competencies carry an additional significance. They are not only studying history; they are also being prepared to teach it. Their ability to judge the validity of evidence, recognize how narratives are constructed, and explain why historical interpretations differ will eventually shape the way history is presented in school classrooms. In this sense, historical literacy involves more than reading historical information. It includes sourcing, contextualization, corroboration, and constructing arguments supported by historical evidence (Kim et al., 2023). Historical understanding deepens when students can assess historical significance, engage with different perspectives, and connect individual pieces of evidence to larger historical narratives (Åhsberg, 2024; Smets, 2024). Historical empathy adds another layer because learners must interpret past actions within the political, social, and cultural conditions of their time rather than judge them only through contemporary values (Harnes, 2025; Karn, 2024). For prospective teachers, then, historical literacy is closely related to both academic competence and professional readiness.

However, the way history is taught does not always reflect these disciplinary expectations. Informative and teacher-centered instruction remains common. In an observational study of 28 classes taught by 14 prospective teachers, classroom discourse focused more on transmitting information than on sustained historical thinking (Miralles-Sánchez et al., 2025). Under such conditions, students may learn what happened without having sufficient opportunities to ask how historical claims are established. They have fewer occasions to interrogate documents, compare perspectives, challenge interpretations, or justify conclusions from evidence. A comparable difficulty appears in Indonesia. Historical literacy at the junior secondary level

still encounters conceptual, strategic, and source-related constraints (Lionar et al., 2025). Attempts to improve it have often emphasized assessment, instructional media, or individual learning projects, while historical literacy itself has not consistently functioned as the organizing principle of classroom practice (Lionar et al., 2025; T. Wijaya et al., 2025). The same issue becomes increasingly important at the university level, where digital historical literacy cannot simply be expected to emerge from students' routine exposure to digital information; it requires deliberate instructional design (A. M. Wijaya et al., 2026).

This problem is especially visible in courses built around contested and multivocal histories. South Asian History is one such course. Colonialism, nationalism, the India–Pakistan partition, identity conflict, religion, migration, and broader socio-political transformations are difficult to understand through lectures and secondary readings alone. These topics involve different actors, conflicting memories, competing political interests, and sources created for very different purposes. A colonial report, a political speech, a newspaper article, a map, or a memoir may refer to the same event while presenting sharply different versions. Working directly with such materials makes one point difficult to ignore: historical knowledge is constructed through interpretation. Evidence can be incomplete, partisan, or even contradictory. Without systematic experience in dealing with these problems, students may remember key events and chronology but still struggle to explain why historians disagree or why different narratives emerge from the same historical episode.

An instructional response therefore needs to do more than make students active in a general sense. It should engage them in the practices through which historical knowledge is actually examined. Inquiry learning offers such a structure. Students begin with a problem or historical question, locate and examine relevant evidence,

consider the origin and purpose of sources, compare accounts, and develop interpretations from what they find. Evidence from higher education supports this approach. A quantitative study involving 240 prospective history teachers reported stronger academic outcomes for students taught through inquiry-based learning than for those taught through traditional lectures (Tirado-Olivares et al., 2021). More importantly for the present study, inquiry corresponds closely with the major dimensions of historical literacy. Examining authorship and purpose develops sourcing; reconstructing surrounding conditions supports contextualization; comparing documents develops corroboration; and constructing conclusions from evidence strengthens historical argumentation. When historical sources become the basis of the inquiry process, the connection between pedagogy and disciplinary practice becomes particularly direct.

Previous studies point in the same direction, although they approach the issue from different angles. Primary source-based inquiry has been associated with stronger historical literacy and critical thinking because students must verify and interpret evidence rather than receive an established account (Kumalasari et al., 2025). Using primary sources also gives learners authentic opportunities to examine traces of the past and determine their evidentiary value (T. Wijaya et al., 2025). Project-based learning has likewise been reported to support historical literacy by involving students in investigation, source use, and the production of historically grounded outcomes (Lionar et al., 2025). What these findings share is not merely the use of a particular method. They suggest that students learn history differently when investigation and evidence are placed at the center of the learning process. However, documents or projects alone are not sufficient. Unless students are guided to question, contextualize, compare, and use evidence in argumentation, an activity may remain

procedural without necessarily developing historical reasoning.

This distinction reveals an important gap in the existing literature. A review of Indonesian studies on historical literacy published between 2019 and 2024 shows that research has largely concentrated at junior and senior secondary school levels, while studies in higher education remain relatively limited (Syah et al., 2025). The thematic focus is also uneven. Much previous research has focused on Indonesian history or world history in a broad sense; South Asian History has received far less attention. At the same time, researchers have often examined media, assessment, project-based learning, primary sources, and inquiry as separate instructional components. Less clear is the specific effect of combining inquiry procedures with systematic historical source analysis on university students' historical literacy in a regional history course. This question matters because South Asian History requires students to negotiate competing narratives, reconstruct political and cultural contexts, and make sense of evidence produced by actors whose interests often conflicted.

Against this background, the present study examines the effectiveness of source-based inquiry learning in improving the historical literacy of students enrolled in a South Asian History course. It compares students' historical literacy before and after instruction and evaluates whether students who engage in source-based inquiry achieve stronger outcomes than those who receive conventional instruction. The study focuses specifically on sourcing, contextualization, corroboration, and evidence-based historical argumentation. It brings inquiry learning and systematic source analysis together in a university-level regional history course, extending previous historical literacy research beyond predominantly school-based and general-history settings. The study therefore tests the hypothesis that students

who participate in source-based inquiry learning will demonstrate significantly higher historical literacy than students who receive conventional instruction. Conceptually, the findings are expected to clarify the relationship between inquiry learning and historical literacy. Practically, they may provide a basis for designing history instruction for prospective teachers that is more investigative, analytical, and grounded in historical evidence.

The study therefore tests the hypothesis that students who participate in source-based inquiry learning will demonstrate significantly higher historical literacy than students who receive conventional instruction. To examine this hypothesis, the following research questions were addressed: (1) Is there a significant difference in posttest historical literacy scores between students who receive source-based inquiry learning and those who receive conventional instruction after controlling for pretest scores? (2) To what extent does source-based inquiry learning improve each dimension of historical literacy (sourcing, contextualization, corroboration, and historical argumentation) relative to conventional instruction? (3) What is the magnitude of the treatment effect on students' overall historical literacy and its sub-skills?

■ METHOD

Research Approach and Design

This study used a quantitative, quasi-experimental pretest–posttest control group design to investigate whether source-based inquiry learning could improve university students' historical literacy. The study was conducted in a regular university classroom setting rather than under fully controlled experimental conditions. Two existing classes, Unit 01 and Unit 02, were involved. Because these classes had already been formed as part of the university's academic organization, assigning individual students randomly to different groups was not feasible

without disrupting the normal learning arrangement.

For this reason, a quasi-experimental design was considered the most appropriate option. This design allows researchers to examine the effect of an instructional intervention while retaining naturally occurring classroom groups, a common situation in educational research (Ong et al., 2021; Puspito et al., 2021).

Unit 01 was designated as the experimental group and received source-based inquiry learning, whereas Unit 02 served as the control group and followed conventional instruction. Both groups completed the same historical literacy test twice: once before the intervention and again after the instructional period. The pretest captured students' initial historical literacy, while the posttest identified changes after instruction.

The independent variable was source-based inquiry learning, and the dependent variable was students' historical literacy. Pretest scores were also used as baseline information when comparing the two groups. This was important because, with intact classes, some initial differences may remain even when the groups have broadly similar academic characteristics. The design therefore allowed the study to examine both improvement within each group and differences in achievement between the experimental and control groups after the intervention.

This study followed ethical principles for educational research involving human participants. Since the participants were students enrolled in the researcher's course, careful consideration was given to minimizing the potential influence of the instructor–student relationship and ensuring voluntary participation.

Before data collection began, the students were informed about the study objectives, the learning procedures, the types of data collected, and the intended use of the research findings. The students were also informed that their participation was voluntary and that their decision to participate

would not affect their academic grades, course evaluation, or learning opportunities. Verbal informed consent was obtained from the students after they had received sufficient information about the research.

The study was conducted as part of regular classroom activities, and the source-based inquiry learning intervention was implemented to ensure that all students received equivalent educational benefits. To protect participants' privacy and confidentiality, students' identities were anonymized using identification codes during data analysis and reporting. The collected data were used solely for academic research purposes, and the published findings included no personally identifiable information.

Population and Sample

The population consisted of fourth-semester students enrolled in the South Asian History course in the History Education Study Program, Faculty of Teacher Training and Education, Universitas Samudra. At the time of the study, the students were organized into two parallel classes, Unit 01 and Unit 02. Each class contained 29 students, giving a total of 58 participants. Unit 01, with 29 students, served as the experimental group, while the remaining 29 students in Unit 02 served as the control group.

Participants were selected using purposive sampling. The choice of classes was guided primarily by academic and administrative considerations because the existing class structure did not allow random redistribution of students. Even so, the two classes shared several relevant characteristics. They were in the same semester, studied the same course, followed comparable academic schedules, and experienced broadly similar instructional conditions. Their initial characteristics were considered relatively comparable in terms of age distribution (range = 19-22 years, $M = 20.4$, $SD = 0.82$), gender composition (experimental group: 18 female, 11

male; control group: 17 female, 12 male), and previous academic performance (cumulative GPA range = 2.80-3.65). Statistically, the two groups' pretest scores did not differ significantly, $t(56) = 0.18$, $p = 0.857$, with a mean difference of only 0.36 points, supporting baseline comparability between the two classes.

The study was conducted within the formal academic setting of the History Education Study Program at Universitas Samudra through an academic research arrangement with the study program and the lecturer responsible for the South Asian History course. The researcher was not a member of Universitas Samudra's permanent faculty; instead, the intervention was implemented with permission and cooperation from the relevant academic personnel at the institution. This arrangement allowed the intervention to occur within the regular course structure while maintaining the institutional procedures governing the research.

Using intact classes offered a practical advantage: the intervention could proceed without changing students' course registration or disrupting normal university procedures. More importantly, it preserved the authentic classroom environment in which the instructional model was intended to operate. The comparison between the two classes then examined whether differences in historical literacy were associated with the treatment rather than simply with participation in the South Asian History course.

Intervention Procedures

The intervention extended over 12 weeks and included six instructional meetings. Each meeting lasted 2×100 minutes. This duration allowed students in the experimental class to work through the inquiry process without reducing source analysis to a brief classroom exercise. Between scheduled meetings, students continued their regular course-related learning activities. In the experimental group, students used these

intervals to review historical sources, complete source-analysis tasks, and prepare materials related to the inquiry activities introduced during the preceding meetings. In the control group, students continued with regular textbook-based readings, assignments, and independent study. Thus, the intervals between meetings were part of the ongoing course activities rather than periods without learning.

Learning in the experimental group began with the orientation of a historical problem. Students then formulated historical questions, identified relevant sources, analyzed and evaluated those sources, compared evidence, developed historical interpretations, and presented the results of their inquiry. The sequence was designed so that students did not simply read historical documents. They had to ask what the documents could reveal, how far they could be trusted, and how one source compared with another.

The materials were deliberately drawn from different types of historical evidence and perspectives related to South Asian history. They included Jawaharlal Nehru's *Tryst with Destiny* speech, *The Times* newspaper edition of 15 August 1947, Mohandas Gandhi's memoirs, the Indian Independence Act of 1947, and historical maps of South Asian kingdoms. Students used these materials to examine themes such as political independence, colonial legacies, national identity, and territorial change.

During source analysis, students considered authorship, purpose, historical context, relevance, credibility, and possible bias. As a result, students did not treat a source as neutral information. Students were also expected to compare evidence across documents, identify agreements or contradictions, and support their interpretations with specific historical evidence.

The control group studied during the same course period but followed conventional instruction. Teaching mainly consisted of lectures,

question-and-answer sessions, and assignments based on textbook readings. Students in this group encountered similar historical themes, but they did not work through the systematic stages of primary-source analysis and historical inquiry used in the experimental class. The main distinction between the two conditions, therefore, lay in the way students engaged with historical knowledge: the experimental group investigated historical questions through sources and evidence, whereas the control group received instruction primarily through conventional classroom explanation and reading activities.

Because both groups were enrolled in the same study program, interaction between students outside the classroom could not be completely restricted. To minimize potential treatment contamination, the experimental and control classes were taught separately, and the experimental group did not receive the specific inquiry worksheets, source-analysis tasks, or instructional procedures used in the experimental group during the intervention period. Students in each class were directed to complete the learning activities assigned to their respective instructional condition. Nevertheless, informal communication among students outside scheduled class activities remained possible and could not be fully controlled by the researcher. This possibility was therefore considered a limitation of the quasi-experimental setting.

Data and Research Instrument

The study relied on primary data collected directly from participants through a source-based historical literacy test. Data were obtained twice. The first measurement was administered as a pretest before the intervention, and the second as a posttest after the instructional period.

An essay format was selected because the abilities examined in this study extend beyond factual recall. Historical literacy requires students to work with information contained in sources,

judge the reliability and relevance of evidence, reconstruct historical context, compare different accounts, and develop interpretations that can be defended with evidence. These processes are difficult to capture adequately through questions that require only recognition or recall.

The test blueprint covered four dimensions of historical literacy: sourcing, contextualization, corroboration, and historical argumentation. Each dimension was weighted proportionally to maintain balance across the instrument. Sourcing measured students' ability to identify a historical source's origin, author, purpose, and characteristics. Contextualization focused on whether students could interpret evidence in relation to the temporal, political, social, and cultural circumstances in which it emerged. Corroboration examined their ability to compare evidence from several sources and recognize consistencies or contradictions. Historical argumentation, meanwhile, assessed how well students developed interpretations, used evidence, and constructed historically defensible conclusions.

Before using the instrument in the main study, the researchers evaluated its validity and reliability. Content validity was assessed using Aiken's *V*, which produced a coefficient of 0.85. This result indicated good correspondence between the test items and the intended indicators of historical literacy. This correspondence was necessary because interpreting students' scores depended on whether the items adequately represented the construct being measured (Fadkhurosi, 2023; Hidayati et al., 2023).

Reliability was examined through an instrument trial involving students outside the research sample. The analysis yielded a Cronbach's alpha coefficient of 0.87, indicating high internal consistency. Taken together, the validity and reliability results provided sufficient evidence that the instrument could measure students' historical literacy in the main study.

Because the historical literacy test used an essay format with rubric-based scoring across four dimensions, we also assessed inter-rater reliability. Two independent raters, both history education lecturers with expertise in historical literacy assessment, scored a subsample of 15 pretest and 15 posttest essays using the four-dimension rubric. Before scoring, raters completed a calibration session to align their understanding of each rubric criterion. Inter-rater agreement was calculated using Cohen's Kappa for categorical consistency and the Intraclass Correlation Coefficient (ICC, two-way random, absolute agreement, single measure) for score reliability. The analysis yielded a Cohen's Kappa of 0.84 and an ICC of 0.89 (95% CI [0.81, 0.94]), indicating strong inter-rater reliability. Discrepancies were resolved through discussion until consensus was reached.

Data Analysis

The analysis proceeded in two broad stages. First, descriptive statistics examined patterns in students' scores before and after the intervention. The analysis included the mean, standard deviation, minimum score, maximum score, and percentage of improvement for both groups. These statistics provided an initial picture of historical literacy in the experimental and control classes and showed how scores changed over the instructional period.

Inferential analysis was conducted after examining the relevant statistical assumptions. Normality was tested using the Shapiro–Wilk test, while Levene's test assessed equality of variance between groups. Because ANCOVA was also used, the homogeneity of regression slopes was examined to determine whether the relationship between pretest and posttest scores was sufficiently comparable across the two groups.

An independent-samples *t*-test was used to compare the posttest historical literacy scores of the experimental and control groups. This

analysis provided a direct comparison of achievement after instruction. Posttest differences alone, however, could still be influenced by students' initial scores. For that reason, ANCOVA was also performed, with pretest scores entered as a covariate. In a pretest–posttest control group design, this procedure allows posttest differences to be examined after statistically accounting for variation in baseline achievement and can therefore provide a more precise estimate of the treatment effect (Nayeri et al., 2024).

Changes in learning outcomes were further examined using normalized gain analysis. The N-gain score represented the relative improvement from pretest to posttest and was interpreted according to the high, moderate, and low categories proposed by Hake (1998). This analysis complemented the significance tests by showing the extent of improvement achieved by students during the intervention.

All statistical analyses were performed using statistical software at a significance level of 0.05. A probability value below 0.05 was interpreted as evidence for rejecting the null hypothesis. Accordingly, we evaluated the effectiveness of source-based inquiry learning by assessing whether the experimental group showed significantly stronger historical literacy outcomes than the group receiving conventional instruction.

■ RESULT AND DISCUSSION

Participants and Analytical Sequence

The results are presented in six tables. Table 1 provides descriptive statistics for pretest and posttest scores. Table 2 reports the results of normality and homogeneity tests. Table 3 presents the independent-samples t-test, while Table 4 summarizes the ANCOVA results. Table 5 displays the N-gain analysis for each historical literacy sub-skill, and Table 6 presents the overall mean N-gain by group. This structure allows readers to trace the analytical sequence from assumption testing through effect-size estimation.

The analysis used data from 58 fourth-semester students in the History Education Study Program, Faculty of Teacher Training and Education, Universitas Samudra. The students were drawn from two parallel classes of equal size. Unit 01, consisting of 29 students, formed the experimental group and participated in source-based inquiry learning. The other 29 students in Unit 02 served as the control group and received conventional instruction.

The results were examined in several stages. Descriptive statistics were first used to establish the initial pattern of scores and the changes observed after instruction. Assumption tests were then conducted before proceeding to between-group comparisons, ANCOVA, and normalized gain analysis. Together, these analyses determined whether the two groups started from comparable conditions, how much their historical literacy changed, and whether the observed differences were associated with the instructional treatment.

Descriptive Statistics of Pretest and Posttest Scores

The descriptive results are presented in Table 1. Before the intervention, the experimental group obtained a mean historical literacy score of 50.78 (SD = 7.79), while the control group recorded a mean of 50.42 (SD = 7.34). The difference was only 0.36 points, and an independent-samples t-test confirmed that the pretest scores of the two groups did not differ significantly, $t(56) = 0.18$, $p = 0.857$. At both the descriptive and inferential levels, therefore, the two groups began with equivalent historical literacy scores.

A clearer difference emerged after the intervention. The experimental group reached a posttest mean of 78.93 (SD = 6.84), compared with 64.90 (SD = 8.78) in the control group. This represents an increase of 28.15 points for students who experienced source-based inquiry learning. The corresponding increase in the control

group was 14.48 points. In other words, the experimental group improved by 13.67 points more than the control group.

The pattern is noteworthy not only because the experimental group achieved a higher mean score. Its posttest scores were also less

dispersed, with a standard deviation of 6.84 compared with 8.78 in the control group. Descriptively, this suggests that the improvement in the experimental class was relatively consistent across students rather than being driven only by a few high performers.

Table 1. Descriptive statistics of historical literacy pretest and posttest scores

Group	N	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Mean N-gain
Experimental	29	50.78	7.79	78.93	6.84	0.58
Control	29	50.42	7.34	64.90	8.78	0.30

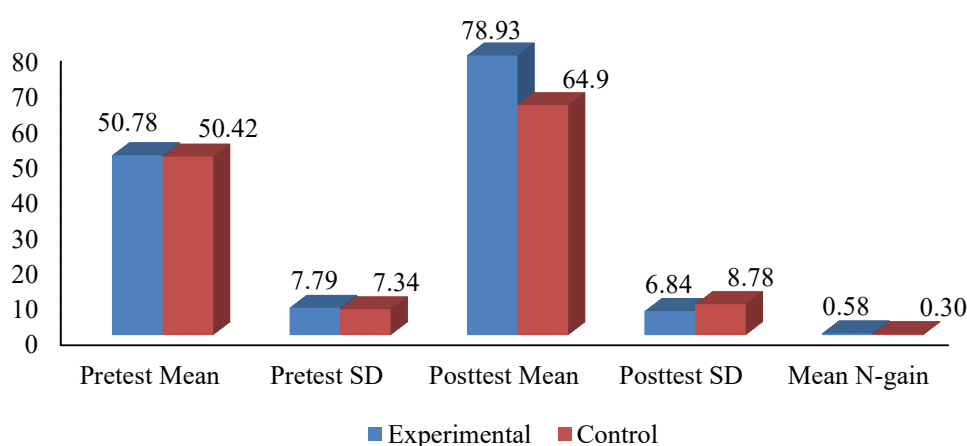


Figure 1. Comparison of pretest and posttest historical literacy scores by groups

Normality and Homogeneity Tests

Before the group differences were tested statistically, the distributional assumptions required for parametric analysis were examined. The Shapiro-Wilk test was conducted separately for the pretest and posttest scores in each group.

As shown in Table 2, the significance values were 0.751 for the experimental pretest, 0.989 for the control pretest, 0.688 for the experimental posttest, and 0.871 for the control posttest. None of these values fell below 0.05. There was therefore no statistical evidence that any of the four score distributions departed significantly from normality.

Levene's Test produced a significance value of 0.417. Because this value was above the 0.05

criterion, the posttest variances of the experimental and control groups were considered homogeneous. However, the homogeneity-of-regression-slopes assumption required for conventional ANCOVA was not satisfied. The pretest-by-group interaction was statistically significant ($p = 0.023$), with the estimated regression slopes being 1.02 for the experimental group and 0.63 for the control group, indicating that the relationship between pretest and posttest scores differed between the two instructional groups. Therefore, conventional ANCOVA was not used as the sole basis for assessing the treatment effect. To address this violation, a conditional effect analysis was conducted using PROCESS Macro Model 1 (Hayes, 2022), with

group as the focal predictor, posttest as the outcome variable, and pretest as the moderator. The Group $\times$ Pretest interaction was statistically significant ($b = 0.3964$, $p = 0.0227$), confirming that the effect of instructional group varied according to students' initial pretest scores. Nevertheless, the conditional effect of group remained statistically significant at low pretest

scores (effect = -16.97 , $p < 0.001$), medium pretest scores (effect = -13.78 , $p < 0.001$), and high pretest scores (effect = -10.75 , $p < 0.001$). Thus, although the magnitude of the group difference varied across pretest levels, we consistently observed a statistically significant group difference at all three examined pretest levels.

Table 2. Results of the normality and homogeneity tests

Test	Group/Data	Sig.	Interpretation
Shapiro–Wilk	Experimental pretest	0.751	Normal
Shapiro–Wilk	Control pretest	0.989	Normal
Shapiro–Wilk	Experimental posttest	0.688	Normal
Shapiro–Wilk	Control posttest	0.871	Normal
Levene's Test	Both groups	0.417	Homogeneous

Hypothesis Testing

The independent-samples t-test showed a clear difference between the two groups after instruction. The analysis yielded $t(56) = 6.79$, $p < 0.001$, with a mean difference of 14.04 points

and a large effect size (Cohen's $d = 1.78$). Thus, students in the experimental group achieved significantly higher posttest historical literacy scores than those in the control group.

Table 3. Independent-Samples t-Test of historical literacy posttest scores

Variable	t	df	Sig. (2-tailed)	Mean Difference	Cohen's d
Historical literacy	6.79	56	0.000	14.04	1.78

The result was consistent with the t-test. After controlling for pretest performance, the instructional model remained significantly associated with posttest historical literacy, $F(1,$

$55) = 110.82$, $p < 0.001$. The difference observed at the end of the intervention therefore remained statistically significant even after adjustment for students' initial scores.

Table 4. Summary of independent-samples t-Test and ANCOVA results

Test	Statistic	df	Sig.	Effect Size
Independent-samples t-test	$t = 6.79$	56	$p < 0.001$	Cohen's $d = 1.78$
ANCOVA	$F = 110.82$	(1, 55)	$p < 0.001$	Partial eta squared = 0.668

The two inferential analyses point in the same direction. The independent-samples t-test identified a significant posttest difference between the experimental and control groups, while ANCOVA showed that this difference remained after adjustment for pretest scores. Viewed alongside the descriptive increase of 28.15 points in the experimental group, compared with 14.48

points in the control group, the findings indicate that students who participated in source-based inquiry learning achieved stronger historical literacy outcomes than students who received conventional instruction.

To clarify the magnitude of the adjusted group difference, estimated marginal means (EMMs) were computed from the ANCOVA

model. After controlling for pretest scores, the experimental group obtained an adjusted posttest mean of 78.79 (SE = 0.92), while the control group obtained an adjusted mean of 65.04 (SE = 0.92). The adjusted mean difference was 13.75 points (95% CI [11.13, 16.36]), confirming that the experimental group maintained a substantially higher level of historical literacy even after statistical adjustment for baseline performance.

Improvement in Historical Literacy Sub-skills and N-gain

To examine where the improvement occurred, the analysis was broken down into four components of historical literacy: sourcing, contextualization, corroboration, and historical argumentation. The pattern in Table 5 is consistent across all four sub-skills. In every case, the experimental group recorded a higher N-gain than the control group.

Table 5. N-gain analysis of historical literacy sub-skills

Historical Literacy Component	Experimental N-gain	Control N-gain	Difference	Category Experimental	Category Control
Sourcing	0.66	0.35	0.31	Moderate	Moderate
Contextualization	0.62	0.29	0.33	Moderate	Low
Corroboration	0.54	0.28	0.26	Moderate	Low
Historical argumentation	0.50	0.27	0.23	Moderate	Low
Overall Historical Literacy	0.58	0.30	0.28	Moderate	Low

The N-gain analysis presented in Table 5 shows a consistent pattern across all four historical literacy sub-skills. The experimental group recorded higher N-gain values than the control group in sourcing, contextualization, corroboration, and historical argumentation. The largest difference was in contextualization, with an N-gain of 0.62 in the experimental group compared with 0.29 in the control group, for a difference of 0.33. Sourcing also showed a substantial difference, with N-gain values of 0.66 and 0.35 for the experimental and control groups, respectively. Meanwhile, corroboration and historical argumentation showed differences of 0.26 and 0.23, respectively. Overall, the experimental group achieved a mean N-gain of 0.58, categorized as moderate, whereas the control group obtained a mean N-gain of 0.30. These findings indicate that source-based inquiry learning produced greater improvements in historical literacy than conventional instruction,

both overall and across each of the four historical literacy sub-skills.

Figure 2. Comparison of pretest and posttest scores across four historical literacy sub-skills. Bars represent group means, with error bars indicating standard deviations. N-gain values are displayed above each pair of bars. The experimental group (source-based inquiry learning) is shown in red (posttest) and light blue (pretest); the control group (conventional instruction) is shown in dark blue (posttest) and grey (pretest).

Taken together, the pattern is consistent across the analyses reported above. The experimental group recorded a larger increase in mean score, a significantly higher posttest result, a significant adjusted treatment effect in the ANCOVA, and higher N-gain values across all four historical literacy sub-skills. Overall, the evidence supports the conclusion that source-based inquiry learning was more effective than

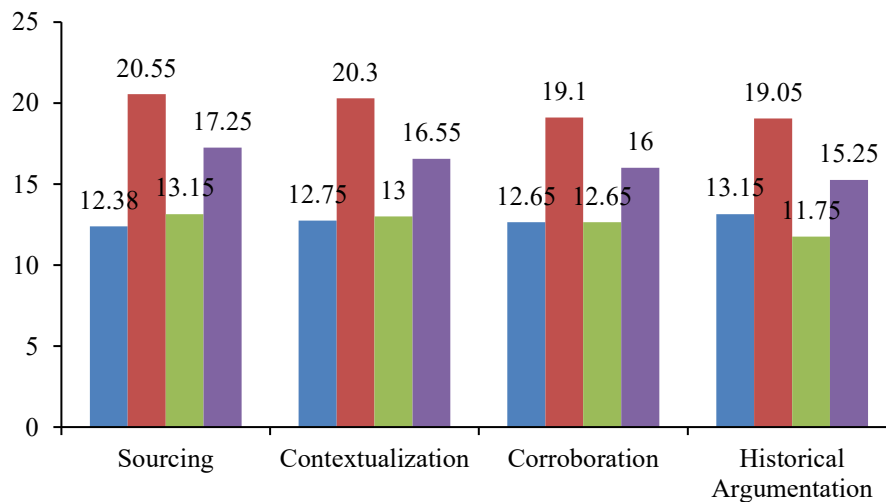


Figure 2. Comparison of pretest and posttest scores across four historical literacy sub-skills

conventional instruction in improving students' historical literacy.

Interpretation of the Main Findings

The findings point to a clear advantage for students who learned through source-based inquiry. At the end of the intervention, the experimental group achieved a mean historical literacy score of 78.93, compared with 64.90 in the control group. More informative, however, is the pattern of change from the beginning of the study. The experimental group improved by 28.15 points, almost twice the 14.48-point increase observed in the control group. The two groups began at nearly the same level, but their scores diverged noticeably after instruction.

The inferential analyses support the same pattern. The posttest difference between the two groups was statistically significant, $t(56) = 6.79$, $p < 0.001$, with a mean difference of 14.04 points and a large effect (Cohen's $d = 1.78$). Because intact classes were used, some variation in initial ability could still have influenced the final scores even though the pretest means were statistically equivalent. ANCOVA was therefore important. After pretest performance was statistically controlled, the instructional model remained significantly associated with posttest historical

literacy, $F(1, 55) = 110.82$, $p < 0.001$. The partial eta squared value of 0.668 further indicates a large adjusted group effect. In practical terms, 66.8% of the variance in posttest scores was explained by the instructional condition after accounting for baseline differences.

This result is meaningful because historical literacy is not simply a matter of remembering more historical information. It involves learning how historical claims are constructed and how evidence supports or challenges an interpretation. Students need to consider where a source comes from, place it in context, compare it with other evidence, and determine what conclusions can reasonably be drawn from it (Kim Wilson et al., 2023). The experimental group's stronger performance therefore suggests that repeated engagement with historical sources gave students more opportunities to practice these disciplinary processes than conventional instruction did.

From this perspective, the findings fit the broader shift in history education away from transmitting established narratives and toward learning that emphasizes interpretation, evidence, and reasoned judgment (Gibson et al., 2025). Source-based inquiry appears particularly compatible with this orientation because students are not positioned only as recipients of historical

explanations. They are required to work through some of the same questions that structure historical investigation: Who produced this evidence? Under what circumstances? For what purpose? How does it compare with other accounts? Moreover, what interpretation can be defended based on the available evidence?

Cognitive Processes in Source-Based Inquiry Learning

The sub-skill results offer a more detailed picture of where the observed improvement occurred. Although these N-gain values are descriptive rather than separate tests of statistical significance, the same general pattern appeared across sourcing, contextualization, corroboration, and historical argumentation. The largest absolute gain in the experimental group was found in sourcing (N-gain = 0.66), followed closely by contextualization (0.62), while the largest between-group disparity appeared in contextualization (0.33), followed by sourcing (0.31).

The strong sourcing result is understandable given the activities students repeatedly encountered during the intervention. When examining materials such as Jawaharlal Nehru's *Tryst with Destiny*, The Times edition of 15 August 1947, Gandhi's memoirs, the Indian Independence Act of 1947, and historical maps, students were asked to consider authorship, date of production, intended audience, purpose, and possible bias. The document itself was not treated as self-explanatory evidence. Students had to ask who produced it and why before deciding how far they could use its content to interpret the past.

Repeated exposure to this kind of questioning may explain why sourcing improved the most. Instead of approaching documents as transparent records of events, students were encouraged to recognize them as products of particular political, social, and historical circumstances. This interpretation aligns with Kim

Wilson et al. (2023), which identifies the deconstruction and evaluation of source material as an important element in developing historical literacy. The present findings suggest that the same principle can be applied in a university-level South Asian History course where historical documents become the starting point for investigation rather than supplementary material at the end of a lecture.

Contextualization showed a similarly strong pattern, with an N-gain of 0.62 in the experimental group and 0.29 in the control group. During inquiry activities, students were not asked to identify what a document said. They also had to consider the political, social, economic, and temporal conditions surrounding its production. This was particularly relevant when studying colonial rule, nationalism, partition, identity, and religious conflict, where the meaning of a statement or decision can change substantially when removed from its historical setting.

Such contextual work may also reduce the tendency to interpret the past only through present-day assumptions. Historical actors operated within circumstances, constraints, and categories that were not necessarily the same as those familiar to contemporary students. Reconstructing those circumstances can therefore support a more historically grounded understanding of action and decision-making. This interpretation aligns with studies that emphasize contextual understanding, historical significance, perspective, and empathy as key elements of meaningful historical learning (Åhsberg, 2024; Harnes, 2025; Karn, 2024; Smets, 2024).

A comparable pattern was found for corroboration. The experimental group reached an N-gain of 0.54, while the control group recorded 0.28. This skill was practiced when students compared different accounts of similar events rather than relying on one source as the definitive version of the past. The partition of India provided a particularly suitable context for this activity. British newspaper reports, political

speeches, legal documents, memoirs, and maps did not necessarily emphasize the same actors, causes, or consequences.

Working across these materials exposed students to both agreement and contradiction. One source could emphasize political independence, another administrative arrangements, and another the human consequences of partition. Such comparisons required students to identify omissions, differences in emphasis, and competing interpretations. In this sense, corroboration did more than confirm factual information. It encouraged students to recognize that historical accounts are shaped by perspective and that a defensible interpretation often requires evidence from more than one source. This is especially relevant to South Asian History, where many events continue to be remembered and interpreted differently across political, national, religious, and cultural contexts.

Historical argumentation showed the lowest N-gain among the four sub-skills, although the experimental group still reached 0.5 compared with 0.27 in the control group. This pattern can be understood through cognitive load theory and the developmental trajectory of historical reasoning. Sourcing, contextualization, and corroboration operate primarily at the analysis and evaluation levels of Bloom's revised taxonomy but can be scaffolded through structured worksheets and guided questioning (Reisman, 2012; Wineburg, 2010). Historical argumentation, by contrast, requires students to synthesize analyzed evidence into an original claim. This process imposes significant element interactivity (Sweller, 2011): learners must simultaneously hold multiple pieces of evidence in working memory, assess their relative weight, reconcile contradictions, and formulate a defensible thesis.

From a scaffolding perspective, the six inquiry meetings in this study provided repeated

opportunities for modeled argumentation (the instructor demonstrated how claims are supported by evidence) and shared argumentation (students constructed arguments collaboratively before writing independently). However, the transition from document analysis to evidence-based claim formation represents a known bottleneck in historical thinking development (VanSledright, 2015). Students in this study were able to identify relevant evidence and recognize contradictions, but constructing a historically defensible argument required them to manage epistemic uncertainty the recognition that multiple interpretations are possible and that evidence can be read in competing ways (Goldman et al., 2016). This uncertainty creates additional cognitive load that may explain why argumentation improved more modestly than the other three dimensions, even within the inquiry condition.

The fact that the experimental group still achieved twice the gain of the control group suggests that the inquiry sequence's final stage, where students presented and defended interpretations, functioned as a form of cognitive apprenticeship (Collins et al., 1989). By making their reasoning visible to peers and receiving feedback on the coherence between claim and evidence, students gradually internalized the disciplinary norms of historical argumentation. Future interventions might strengthen this dimension by extending guided argumentation practice or incorporating sentence starters and argument frames that reduce extraneous cognitive load during the synthesis phase (De La Paz, 2005).

Relationship with Previous Research

The present findings broadly support earlier evidence on the value of inquiry-oriented history instruction. Tirado-Olivares et al. (2021), for example, reported that inquiry-based learning produced better academic outcomes than traditional lecturing among 240 prospective

history teachers, with a Cohen's d of 0.65. Their results provide an important point of comparison because both studies placed prospective teachers in learning environments that required more active engagement with historical problems rather than reliance on teacher explanation alone. The effect observed in the present study was even larger (Cohen's $d = 1.78$ for the posttest comparison; partial eta squared = 0.668 for the adjusted analysis). Although these effect-size indices come from different statistical procedures and should not be treated as directly interchangeable, the magnitude observed here suggests that the instructional condition accounted for a considerable proportion of variation in students' posttest historical literacy after controlling for baseline scores.

Several features of the intervention may help explain this pattern. One is the duration and repetition of the inquiry activities. Over the 12-week intervention, students repeatedly engaged in sourcing, contextualization, corroboration, and historical argumentation, including through structured assignments between the six face-to-face meetings. Repetition may have given them time to become more familiar with the inquiry process and to apply the same disciplinary routines across different historical materials. The use of varied sources may also have mattered. Political speeches, newspapers, memoirs, legal documents, and historical maps required students to approach evidence from different angles rather than apply the same reading strategy to a single document type. These explanations remain tentative, but they are compatible with the pattern of improvement observed across the four historical literacy sub-skills.

The findings are also consistent with Kumalasari et al. (2025), who found that inquiry involving primary sources can contribute to historical literacy and critical thinking. A common feature of the two studies is the role students play in the learning process. They were not simply

presented with a completed account of the past. Instead, they had to question evidence, examine the basis of historical claims, and justify the interpretations they developed.

A similar connection can be made with T. Wijaya et al. (2025), who emphasized primary sources' contribution to historical literacy. Lionar et al. (2025) likewise showed that active and project-oriented investigation can support students' historical understanding. Taken together, these studies suggest that historical literacy develops more readily when students work with evidence and construct meaning from it rather than only reproduce information from lectures or textbooks.

The present study adds a more specific contribution to this body of work. Inquiry was not treated as a general active-learning strategy, nor were historical sources used merely as supplementary materials. The two were deliberately combined in a sequence in which source analysis formed the core of the inquiry process. This integration was implemented in a South Asian History course, where competing political narratives, colonial records, nationalist perspectives, and memories of partition provide particularly rich opportunities for disciplinary historical reasoning. The findings therefore extend previous research by showing how a systematically organized, source-based inquiry model can develop historical literacy in a university-level regional history course.

Limitations of the Study

The findings need to be considered alongside several methodological limitations. The first concerns the quasi-experimental design. Because the study used two existing classes, students could not be randomly assigned individually to the experimental and control conditions. The pretest means were statistically equivalent, and the two classes were considered broadly comparable in their initial characteristics;

however, these similarities do not eliminate the possibility of selection effects. Unmeasured differences may still have influenced students' responses to the intervention.

A second limitation concerns events and learning experiences that may have occurred during the 12-week instructional period. Classroom-based quasi-experiments cannot completely isolate participants from experiences outside the intervention. For example, an external academic or social event could influence students in one class differently from those in another. No major event was identified during the implementation period, but this type of history threat to internal validity cannot be ruled out entirely. In addition, because the intervention was conducted over 12 weeks with six face-to-face meetings, learning activities between meetings were supported through structured assignments related to each instructional condition. However, these out-of-class activities could not be monitored with the same degree of control as classroom activities, which may have introduced variation in students' learning experiences during the intervention period.

A third limitation relates to the ANCOVA assumption of homogeneity of regression slopes. The pretest-by-group interaction was marginally significant ($p = 0.023$), indicating that the strength of the relationship between pretest and posttest scores differed somewhat between the two classes. Although the adjusted and unadjusted analyses converged on the same conclusion, and the normalized gain analysis, which does not depend on this assumption, showed the same pattern, interpret the ANCOVA estimate with this caveat in mind.

The researcher's dual role as instructor and observer presents another consideration. Direct involvement can help maintain consistency between the intended design and classroom implementation, but it may also introduce expectancy or interaction effects. Several steps were taken to minimize potential contamination

between the classes, including separate instructional schedules, restricted distribution of source packages and inquiry materials, and clear instructions regarding the learning activities assigned to each class. Nevertheless, because both classes belonged to the same study program, informal social interaction between students of the two groups could not be fully prevented. Therefore, some information exchange between conditions cannot be completely excluded and should be considered when interpreting the findings.

Generalizability is also limited by the setting of the study. All 58 participants came from a single History Education Study Program at one university and were enrolled in the same South Asian History course. The findings should therefore not be assumed to apply automatically to students in other institutions, disciplines, or historical contexts. Source-based inquiry may operate differently depending on students' prior historical knowledge, the availability and complexity of sources, instructor experience, course objectives, and institutional learning conditions.

There are limitations related to measurement as well. Historical literacy was assessed primarily through an essay-based test. The instrument demonstrated good content validity and high internal consistency, but an essay test captures only part of historical inquiry. It does not fully represent students' participation in classroom discussion, their engagement with uncertainty, the way they negotiate interpretations with peers, or whether the skills demonstrated in the posttest persist over time. The findings therefore describe students' measured historical literacy performance at the end of the intervention rather than every dimension of their historical learning.

Directions for Future Research

Several directions follow from these limitations. Where institutional conditions allow,

future studies could use randomized designs to reduce selection effects and strengthen causal inference. Replication across different universities would be equally important. Studies involving multiple history education programs, instructors, and student cohorts could show whether the pattern observed here is specific to this setting or can be reproduced under different academic conditions. In addition, future studies could examine instructional implementations in settings where experimental and control groups are more clearly separated, or where procedures for monitoring interaction between groups can be strengthened, to reduce the possibility of treatment contamination. Involving independent instructors or multiple researchers could also help determine whether the effects of source-based inquiry remain consistent when instructor-related influences are reduced.

The intervention should also be examined across different historical domains. South Asian History offers a particularly rich setting because of its contested narratives and diverse source traditions. However, it remains unclear whether a similar pattern would emerge in courses on Indonesian history, European history, local history, or other regional histories. Comparative research across these contexts could reveal whether the effectiveness of source-based inquiry depends partly on the nature of historical content, students' prior knowledge, and the types of evidence available for analysis.

Methodologically, mixed-methods research could provide a more detailed account of how historical literacy develops during inquiry. Test scores can show whether improvement occurs, but they reveal less about the reasoning processes that drive it. Classroom observations, interviews, think-aloud protocols, students' written work, and analysis of inquiry discussions could help trace how students move from identifying source characteristics to contextualizing evidence, comparing historical accounts, and eventually constructing evidence-based historical arguments.

This is particularly relevant to the pattern found in the present study. Sourcing and contextualization showed the highest gains, whereas historical argumentation improved more modestly. Future research could examine whether this difference reflects increasing cognitive complexity, the amount of instructional time devoted to each skill, students' prior experience with historical writing and argumentation, or some combination of these factors. Process-oriented evidence would make it possible to move beyond identifying the pattern toward explaining how and why these competencies develop through source-based inquiry.

Longitudinal research is another important next step. The present study measured historical literacy immediately after the intervention, so it cannot determine whether students retained the observed gains over time. Follow-up assessments could establish whether students continue to apply sourcing, contextualization, corroboration, and evidence-based argumentation in later history courses. For prospective history teachers, an even more important question is whether these competencies eventually influence their own instructional practices and approaches to teaching history.

Future studies may also broaden the range of outcomes examined. Historical empathy, epistemic beliefs, learning engagement, and the quality of evidence-based argumentation are particularly relevant because they may help explain how students respond to conflicting historical accounts and construct interpretations of the past. Examining these variables alongside historical literacy could provide a more comprehensive understanding of what students learn through source-based inquiry and how such learning develops across different educational contexts.

■ CONCLUSION

The findings indicate that source-based inquiry learning was associated with greater

improvements in university students' historical literacy than conventional instruction in the South Asian History course. Students in the experimental group achieved a higher posttest mean, demonstrated a larger increase from their initial scores, and reached a moderate level of normalized gain. By contrast, improvement in the control group remained lower. Both the independent-samples t-test and ANCOVA supported this pattern, with the group difference remaining statistically significant after accounting for students' initial performance. The large effect size further suggests that the instructional condition was meaningfully associated with students' historical literacy outcomes.

The improvement was not confined to overall performance. A consistent pattern also appeared across the four dimensions examined in this study: sourcing, contextualization, corroboration, and evidence-based historical argumentation. Sourcing showed the strongest gain. This pattern suggests that repeated engagement with different forms of historical evidence, including speeches, newspapers, memoirs, legal documents, and historical maps, provided students with sustained opportunities to examine who produced a source, why it was created, how credible it was, and what kinds of bias or limitations it might contain. Contextualization and corroboration also improved, indicating that students became better able to situate evidence within its historical setting and compare different accounts of the same or related events.

These results are particularly relevant for teaching histories characterized by competing narratives and multiple perspectives. South Asian History includes colonial, nationalist, political, religious, and cultural interpretations that cannot always be understood through a single account. By requiring students to examine multiple sources and justify their interpretations with evidence, source-based inquiry created opportunities for

students to recognize the multivocal nature of the past rather than approach history as a fixed narrative.

The study contributes to historical literacy research by examining source-based inquiry within a university-level regional history course. This area has received comparatively limited attention in the Indonesian context. Its findings provide preliminary support for integrating structured source investigation into history teacher education, particularly in courses where students are expected to evaluate competing accounts and construct evidence-based interpretations. For prospective history teachers, this is important not only for their own academic development but also for the kinds of historical learning experiences they may later design for their students.

Further research is needed to determine how far these findings can be extended beyond the present setting. Studies with randomized designs, larger samples, and multiple institutions could provide stronger evidence of the intervention's generalizability. Longitudinal research would also help establish whether gains in sourcing, contextualization, corroboration, and argumentation are retained over time and transferred to subsequent coursework or teaching practice. In addition, mixed-methods studies could examine the reasoning processes behind these improvements more closely, particularly how students move from evaluating individual sources to constructing increasingly sophisticated historical arguments.

■ DECLARATION OF GENERATIVE AI USAGE IN THE WRITING PROCESS

During the drafting and revision of this manuscript, the authors used generative artificial intelligence (AI) tools, particularly ChatGPT, to organize the manuscript structure, improve the flow of ideas between sections, refine academic language, align terminology, and support the

systematic presentation of concepts. After using these tools, the authors critically reviewed, verified, and revised all AI-assisted outputs to ensure accuracy, relevance, and consistency with the research objectives and scientific evidence. The authors take full responsibility for the final manuscript content.

■ REFERENCES

- Åhsberg, C. S. (2024). Students' views of historical significance—a narrative literature review. *History Education Research Journal*, 21(1). <https://doi.org/10.14324/HERJ.21.1.02>
- Ana Isabel, M., & Barca, I. (2024). Historical narratives: how Portuguese students aged 10 and 15 make use of national history. *History Education Research Journal*, 21(1). <https://doi.org/10.14324/HERJ.21.1.10>
- Barreiro, A., & Pérez-Manjarrez, E. (2025). Collective memory, moral judgments, and history education. In *Current Opinion in Psychology* (Vol. 66). Elsevier Science. <https://doi.org/10.1016/j.copsyc.2025.102113>
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. (pp. 453–494). Lawrence Erlbaum Associates, Inc.
- De La Paz, S. (2005). Effects of Historical Reasoning Instruction and Writing Strategy Mastery in Culturally and Academically Diverse Middle School Classrooms. *Journal of Educational Psychology*, 97(2), 139–156. <https://doi.org/10.1037/0022-0663.97.2.139>
- Fadkhurosi, A. (2023). Analisis Pearson Product Moment dan Alpha Cronbach dalam Pengembangan Need Assesment Bimbingan dan Konseling. *Jurnal Muria Research Guidance and Counseling (MRGC)*, 2(1), 40–47. <https://doi.org/10.24176/mrgc.v2i1.9451>
- Gibson, L., Peck, C. L., Miles, J., & Duquette, C. (2025). Historical thinking: Trends, critiques, and future directions. *Current Opinion in Psychology*, 65, 102088. <https://doi.org/https://doi.org/10.1016/j.copsyc.2025.102088>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., Shanahan, C., & READI, P. (2016). Disciplinary Literacies and Learning to Read for Understanding: A Conceptual Framework for Disciplinary Literacy. *Educational Psychologist*, 51(2), 219–246. <https://doi.org/10.1080/00461520.2016.1168741>
- 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*, 1(68), 64–74. <https://files.eric.ed.gov/fulltext/ED441679.pdf>
- Harnes, H. B. (2025). What can empirical research tell us about how to develop students' historical empathy? A scoping review. *History Education Research Journal*, 22(1). <https://doi.org/10.14324/HERJ.22.1.18>
- Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (Vol. 3). The Guilford Press.
- Hidayati, R., Kusmanto, A. S., & Kiswanto, A. (2023). Development and Construct Validation of Indonesian Students Self-Confidance Scale Using Pearson Product Moment. *Pegem Journal of Education and Instruction*, 13(3 SE-Article), 94–103. <https://doi.org/10.47750/pegegog>

13.03.11

- Karn, S. (2024). Designing historical empathy learning experiences: a pedagogical tool for history teachers. *History Education Research Journal*, 21(1). <https://doi.org/10.14324/HERJ.21.1.06>
- Kim, W., Dean, D., Janet, D., Renee, P.-M., & Elizabeth, P. (2023). A systematic review of pedagogical interventions on the learning of historical literacy in schools. *History Education Research Journal*, 20(1). <https://doi.org/10.14324/HERJ.20.1.09>
- Kumalasari, D., Gunartati, G., Wilis, R., Dewi, I. M., & Yudi, S. (2025). Enhancing historical literacy and critical thinking through primary source-based inquiry: evidence from the jogja library center tour dyah. *Jurnal Pendidikan Progresif*, 15(03), 2028–2048. <https://doi.org/10.23960/jpp.v15i3.pp2028-2048>
- Lionar, U., Mulyana, A., Sapriya, Winarti, M., Yefterson, R. B., Syafrina, Y., Naldi, H., & Yuhardi. (2025). Investigating local historical figures: the implementation of the pjbl model to increase historical literacy. *Paramita*, 35(1), 116–128. <https://doi.org/10.15294/paramita.v35i1.23411>
- Miralles-Sánchez, P., Rodríguez-Medina, J., & Gómez-Carrasco, C. J. (2025). Historical thinking and teacher discourse in secondary education: an exploratory observational study. In *Education Sciences* (Vol. 15, Issue 3, p. 394). <https://doi.org/10.3390/educsci15030394>
- Nayeri, N. D., Noodeh, F. A., Nia, H. S., Yaghoobzadeh, A., & Goudarzian, A. H. (2024). Statistical procedures used in pretest-posttest control group design: a review of papers in five iranian journals. *Acta Medica Iranica*, 61(10 SE-Review Article(s)). <https://doi.org/10.18502/acta.v61i10.15657>
- Ong, E. T., Govindasamy, D., Singh, C. K. S., Ibrahim, M. N., Wahab, N. A., Borhan, M. T., & Tho, S. W. (2021). The 5E inquiry learning model: Its effect on the learning of electricity among Malaysian students. *Cakrawala Pendidikan*, 40(1), 170–182. <https://doi.org/10.21831/cp.v40i1.33415>
- Puspito, D. R. A., Supardi, K. I., & Sulhadi Sulhadi. (2021). The influence of guided inquiry models on science process skills dhimas. *Journal of Primary Education*, 10(1), 36–43. <https://doi.org/10.15294/jpe.v9i3.39442>
- Reisman, A. (2012). Reading like a Historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112. <https://doi.org/10.1080/07370008.2011.634081>
- Smets, W. (2024). Four design principles for student learning of substantive historical concepts – a realistic review study. *History Education Research Journal*, 21(1). <https://doi.org/10.14324/HERJ.21.1.04>
- Sweller, J. (2011). Cognitive load theory. In *The psychology of learning and motivation: Cognition in education*, Vol. 55 (pp. 37–76). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Syah, M. A., Mulyani, F. F., Wijayadi, W., & Arfaton, A. (2025). Research trends in history education on historical thinking in Indonesia, 2019-2024: A systematic literature review and bibliometric. *Agastya: Jurnal Sejarah Dan Pembelajarannya*, 15(2), 207–225. <https://doi.org/10.25273/ajsp.v15i2.21073>
- Tirado-Olivares, S., Czar-Gutiérrez, R., García-Olivares, R., & González-Calero, J. A. (2021). Active learning in history teaching in higher education: The effect of inquiry-based learning and a student response system-based formative assessment in

- teacher training. *Australasian Journal of Educational Technology*, 37(5), 61–76. <https://doi.org/10.14742/ajet.7087>
- Tribukait, M. (2024). History education as prevention: the topic of right-wing extremism in German educational media. *History Education Research Journal*, 21(1). <https://doi.org/10.14324/HERJ.21.1.12>
- VanSledright, B. (2015). What does it mean to think historically . . . and how do you teach it? *Social Studies Today: Research and Practice: Second Edition*, 144–150. <https://doi.org/10.4324/9781315726885-24>
- Wijaya, A. M., Satriya, R. N., & Arifin, I. F. (2026). Reconstructing Gen Z digital historical literacy through AI- based pedagogy in higher education. *Harmoni Sosial: Jurnal Pendidikan IPS*, 13(1), 60–70. <https://doi.org/https://doi.org/10.21831/hsjpi.v13i1.93395>
- Wijaya, T., Alauwiyah, F., & Pahlevi, M. R. (2025). Utilization of primary sources in history learning/ : teachers ' efforts to improve high school students ' historical literacy. *J-PIPS (Jurnal Pendidikan Ilmu Pengetahuan Sosial)*, 12(1), 1–15. <https://doi.org/http://dx.doi.org/10.18860/jpips.v12i1.37197>
- Wineburg, S. (2010). Historical thinking and other unnatural acts. *Phi Delta Kappan* 80, 488–499. kappanmagazine.org