Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 2, 264–288
https://doi.org/10.62159/jpi.v6i2.2745

OPEN ACCESS
Research Article

A Blended Instructional Design for Self-Reported Historical Literacy in South Asian History: Development and Initial Evaluation

Septina Alrianingrum[1*], Fadiyah Salsabila[2], Sumarno[3]
[1,3] History Education, Universitas Negeri Surabaya, Surabaya, Indonesia
[2] Social Science Education, Universitas Negeri Surabaya, Surabaya, Indonesia

Abstract
South Asian History courses often rely on dispersed, predominantly foreign-language resources that provide limited scaffolding for interpreting contested narratives and constructing evidence based historical arguments. This study developed and initially evaluated a discipline specific blended instructional package aligning synchronous guidance and asynchronous inquiry with six historical literacy indicators. Using ADDIE and an uncontrolled one-group pretest–posttest field evaluation, the study involved a convenience sample of 76 undergraduate students in three intact classes in the History Education Study Program at Universitas Negeri Surabaya. The package was designed for 16 planned sessions; however, actual instructional dosage and implementation-fidelity indicators were not systematically documented. One subject-matter validator, one media validator, and one learning validator appraised different domains using non-equivalent instruments; their domain-specific ratings ranged from 85.20% to 88.00%, with a descriptive cross-domain mean of 86.93%. Student acceptability averaged 84.82%. Aggregate self-reported historical-literacy scores increased from 51.03 (SD=14.55) to 88.93 (SD=4.91), with a mean change of 37.91, 95% CI [34.53, 41.29], t(75)=22.34, *sig* < .005, and Cohen's d_z = 2.56. The group-level normalized gain was 0.77 (77.39%), which was descriptively classified as high. These findings indicate high acceptability responses and substantial short term within-group gains in self-reported historical literacy. However, these findings do not establish causal effectiveness or demonstrated historical literacy performance.

ARTICLE HISTORY
Received : 7 April 2026
Revised : 19 June 2026
Accepted : 20 July 2026

KEYWORDS
Blended history learning; Historical literacy; Instructional design; South Asian history; History education

PUBLISHER'S NOTE
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

CORRESPONDING AUTHOR
***Septina Alrianingrum**, History Education, Universitas Negeri Surabaya, Surabaya, Indonesia. Email: septi@unesa.ac.id

Introduction

Teaching South Asian History to Indonesian history education students presents a discipline-specific challenge. Students must interpret long-term and interregional processes involving colonialism, nationalism, religion, political conflict, social transformation, and identity formation. At the same time, periodizations such as medieval, early modern, and modern, as well as explanations of modernity and colonial change, remain contested. Interpretations of nationalism, identity, and political agency also depend on the evidence and perspectives historians privilege. Learning the field therefore requires students to connect events across regions and periods, compare interpretations, and examine whose experiences are represented or marginalized, rather than treating South Asian history as a sequence of isolated countries, dynasties, and events (Bhargava & Nath, 2022; Jayaneththi & Dharmarathna, 2025).

A review of the existing course plan and learning materials identified a resource-related instructional problem in the target course. The available materials consisted mainly of lecturer-prepared slides, videos, and selected e-books, while structured guidance for engaging foreign-

language academic texts and connecting dispersed sources was limited. This evidence documents a gap in instructional provision, not the extent of individual students' language difficulty, because no separate student survey or interview measured that difficulty. The 76 participants' mean score of 51.03 out of 100 on the 25 item self-report pretest provided an additional but non-independent indication of need, comparable to the scores of 49.91–55.03 reported among Indonesian history education students (Kumalasari et al., 2022). Because the same pretest served as needs evidence and the outcome baseline, it introduced potential circularity; completing it may also have sensitized students to the targeted practices. The score was therefore treated only as preliminary contextual evidence, and possible testing reactivity was considered when interpreting change.

Historical literacy is related to but not synonymous with historical thinking, disciplinary literacy, or digital historical literacy. Disciplinary literacy denotes the epistemic conventions for reading, evaluating, and communicating knowledge within history (Moje, 2008; Shanahan & Shanahan, 2012); historical thinking denotes the reasoning practices used to interpret evidence (Seixas & Morton, 2013; Wineburg, 2001); and historical literacy is the broader capacity to mobilize historical knowledge and these practices to understand and communicate interpretations of the past (Maposa & Wassermann, 2009; Taylor & Young, 2003). Digital historical literacy extends this capacity to digitally mediated sources and media. Thus, disciplinary literacy provides the epistemic foundation, historical thinking supplies the reasoning processes, historical literacy is the intended learning outcome, and digital historical literacy describes its application in digital environments (Lesley et al., 2023; Nokes, 2022; Wissinger & Ciullo, 2018).

For this study, students' self-reported historical literacy was operationalized through six interrelated indicators: reading historical sources, understanding chronology, understanding historical context, connecting events, comparing perspectives, and constructing historical arguments. Together, these indicators represent students' perceived capacity to engage with historical sources, organize temporal relationships, situate events in context, identify connections among events, examine different perspectives, and develop historical arguments (Nokes, 2022; Taylor & Young, 2003; Wilson et al., 2023). In relation to Seixas & Morton's (2013) framework, the use of historical evidence and historical perspectives was represented directly, while continuity and change and cause and consequence were incorporated into understanding chronology and connecting events. Historical significance and the ethical dimension were excluded because the study focused on interpreting historical evidence and constructing historical arguments rather than evaluating the relative importance or moral implications of past events. The six indicators therefore define the specific scope of historical literacy examined in this study rather than representing all dimensions of historical thinking.

Prior blended-history interventions demonstrate pedagogical promise but leave this disciplinary alignment insufficiently specified. Prior blended history interventions demonstrate pedagogical promise but leave this disciplinary alignment insufficiently specified. Othman & Al-Hileh (2022) reported gains in general critical-thinking skills in a school history course, but did not assess a defined set of historical literacy processes. A connectivist blended design in African History supported conceptual understanding and digital literacy, yet did not explicitly align activities and assessment with sourcing, contextualization, corroboration, multiperspectivity, and evidence based historical argumentation (Alrianingrum et al., 2023). An ICARE based rotation model subsequently targeted historical literacy in Indonesian History course, but reported feasibility and aggregate

learning gains rather than a component-to-process alignment or direct performance evidence for each practice (Tanjung et al., 2023). Taken together, these studies have not sufficiently examined how blended learning components can be explicitly aligned with defined historical literacy processes and corresponding self-report outcome indicators in a higher-education regional history course. The present study addresses this gap through discipline specific design alignment and an initial evaluation based on expert appraisals, students' acceptability responses, and short term changes in self-reported historical literacy scores.

This study contributes a discipline-specific blended instructional package that explicitly aligns synchronous guidance and asynchronous inquiry with six historical literacy indicators and provides an initial evaluation of the package in a specific higher education context. Its novelty lies in the explicit alignment of synchronous guidance and asynchronous inquiry with six defined historical literacy indicators, supported by an initial evaluation in a higher-education context. Accordingly, this study aimed to develop and initially evaluate a blended instructional package for historical literacy in South Asian History. Specifically, it sought to identify instructional needs and translate them into design requirements, obtain domain specific expert appraisals, document implementation across three classes, and examine students' acceptability responses, and short-term within-group changes changes in students' self-reported historical literacy scores following implementation.

Method

This study employed an instructional product-development approach with an initial field evaluation. It focused on developing a context-specific blended instructional package for South Asian History and examining domain-specific expert appraisals, students' acceptability responses, and short-term within-group changes in self-reported historical literacy scores following implementation in three intact classes. Because the study was conducted in a single institutional setting, involved one development sequence, and did not undertake iterative research cycles intended to formulate or test transferable design principles, it was positioned as product development rather than educational design research.

At the time of the study, the institutional procedures applicable to this type of classroom-based instructional development study did not provide for formal ethics committee review; consequently, no ethical approval number was issued. Before the questionnaires were administered, students were verbally informed that participation was voluntary and that choosing to participate or decline would not affect their course grades or access to instruction. No written participant information sheet was provided, no signed consent form was obtained, and no contemporaneous record of the verbal notification was retained. Accordingly, the study did not involve documented written informed consent. The verbal notification was intended to minimize potential pressure arising from the lecturer–student power relationship. Questionnaire responses were handled confidentially, student identities were not disclosed in the analysis or reporting, and the findings were presented only in aggregate form.

ADDIE was used as the procedural framework, comprising analysis, design, development, implementation, and evaluation (Branch, 2009). Its phases corresponded to the study procedures: exploring instructional problems, specifying design requirements, developing and appraising the package, implementing it in three intact classes, and evaluating student responses and changes in self-reported historical literacy scores. To improve traceability across these phases, Table 1 links the

documented analysis findings with the corresponding design choices, development-stage revisions, and evaluation evidence.

The analysis stage was conducted as an exploratory preliminary study using evidence already available within the study program: previous institutional research on Vinesa and online history instruction, observations of existing learning practices, and a review of the South Asian History course plan and available teaching materials. The 2018 study conducted in the same study program and university reported that 80.69% of participating students perceived that Vinesa supported their active involvement in learning and the subsequent institutional review conducted in 2020 in ten courses, found that 70% of online history instruction still relied on WhatsApp Groups, whereas dedicated online-learning applications were used in only 15% of the reviewed courses. The course review also identified limited Indonesian-language resources, students' unfamiliarity with foreign-language literature, and the extensive scope of South Asian History as practical instructional problems.

These sources were used descriptively to formulate practical design requirements. The available records do not specify a formal protocol for selecting documentary evidence, a prespecified coding framework, independent coding, an intercoder-reliability check, or a formal audit of the resulting categories. This stage is therefore described as an exploratory preliminary analysis rather than a systematic documentary analysis or a formal student needs survey. The historical literacy pretest was not treated as needs-analysis evidence because it served as the baseline outcome measure for the field evaluation rather than as evidence used to develop the package.

Table 1. Documented ADDIE Decision Chain

Documented analysis finding	Design requirement or choice	Documented development-stage response	Evaluation criterion or evidence
A previous institutional study reported that 80.69% of students perceived active involvement when Vinesa was used	Retain Vinesa as the central digital learning environment	Technical guidance for accessing and completing Vinesa based activities was clarified	Media and learning expert appraisals; student responses concerning accessibility, usability, and instructional package design.
An institutional review found that 70% of online history instruction relied on WhatsApp Groups, whereas dedicated online learning applications were used in only 15% of ten reviewed courses	Organize resources, assignments, discussions, and complementary synchronous asynchronous activities through Vinesa	Asynchronous interaction was strengthened and technical instructions were clarified	Media and learning expert appraisals; student respons concerning blended learning implementation; available Vinesa participation and completion records
The review identified limited Indonesian language resources and students' unfamiliarity	Organize accessible course resources and provide guided source engagement activities	Access to linked and downloadable learning materials was improved. No separate language	Subject matter and media appraisals; student respons concerning the content and accessibility of learning

with foreign language literature	and worksheets	support revision was documented	materials
The extensive scope of South Asian History was identified as a practical instructional problem	Sequence materials and activities within a package designed for 16 planned sessions	No revision specifically addressing course scope or session allocation was documented	Subject matter and learning appraisals; student responses concerning package design. Actual session dosage was not documented

Note. The records do not provide a complete one-to-one audit trail linking every analysis finding to a specific revision. Only documented findings, design choices, and development-stage responses are shown. Alignment with the six historical literacy indicators was a theory-based design decision, not a result of the exploratory analysis.

The design stage translated the exploratory findings and the six historical literacy indicators into an alignment matrix linking intended learning outcomes, learning activities and modalities, instructional package components, historical literacy indicators, and learning evidence. Synchronous sessions focused on conceptual framing, discussion, clarification, and lecturer feedback, while asynchronous activities in Vinesa supported source reading, worksheet completion, chronological and contextual analysis, connections among events, comparison of interpretations, and evidence based writing. The instructional package included the semester learning plan, teaching materials, learning resources, student worksheets, assignments, discussion activities, historical literacy questionnaire used as the pretest and posttest, and a separate student response questionnaire. The resulting design alignment is presented in Table 2.

Table 2. Design Alignment Matrix for the Blended Instructional Package

Intended learning outcome	**Learning activities**	**Modality**	**Product components**	**Historical literacy indicator**	**Learning task or evidence and related questionnaire items**
Analyze historical sources as evidence	Guided source reading, source analysis, and discussion of findings	Asynchronous preparation followed by synchronous discussion	Selected learning resources, student worksheet, discussion activities	Reading historical sources	Source-analysis worksheet And related pretest–posttest items
Organize historical developments in a coherent sequence	Reading teaching materials constructing chronological sequences	Asynchronous task followed by lecturer feedback	Teaching materials and student worksheet	Understanding chronology	Chronology task and related pretest–posttest items
Interpret events in their historical settings	Context review, guided analysis, and class discussion	Asynchronous analysis and synchronous clarification	Teaching materials, worksheet, discussion activities	Understanding historical context	Context-analysis response and related pretest–posttest items
Explain relationships	Identifying explaining	Asynchronous assignment	Student	Connecting events	Written explanation of

among events, periods, and regions	connections among historical developments	followed by synchronous discussion	worksheet assignment		inter-event relationships and related test items
Compare different historical interpretations	Comparing accounts discussing differing perspectives	Asynchronous reading and synchronous group discussion	Learning resources, assignment, discussion activities	Comparing perspectives	Comparative response, group discussion output, and related test items
Construct an Evidence based historical argument	Selecting evidence, drafting an argument, receiving feedback, and revising explanation	Asynchronous writing supported by synchronous feedback	Student worksheet, written assignment, lecturer feedback	Constructing historical arguments	Evidence-based written argument and related pretest–posttest items

Table 2 shows that each learning outcome was aligned with complementary asynchronous and synchronous activities. Asynchronous tasks supported source analysis, chronology, contextualization, perspective comparison, and argument development, while synchronous sessions provided clarification, discussion, and feedback. The learning tasks demonstrated students' engagement with the six indicators, whereas the pretest and posttest questionnaire items measured their self-reported historical literacy.

During the development stage, the initial package was produced according to the design matrix. It comprised the semester learning plan, South Asian History teaching materials and resources, student worksheets, assignments, discussion activities, the historical literacy questionnaire, and the student-response questionnaire. One subject-matter validator, one media validator, and one learning validator appraised different domains using separate sheets containing 25, 20, and 40 items, respectively. Each item was rated on a five-point scale and the feasibility percentage was calculated by dividing the obtained score by the maximum possible score and multiplying it by 100. The available records identify the validators' domain roles but do not document their academic qualifications, selection criteria, independence from the research team, or years of relevant experience. Accordingly, the appraisals are treated as preliminary domain specific judgments.

The documented development-stage responses included improving access to downloadable materials, clarifying instructions for asynchronous activities, strengthening online interaction, and refining technical guidance for Vinesa. Because the validators assessed different domains with non-equivalent instruments, their scores were retained as separate judgments and no inter-rater agreement was calculated. Any arithmetic average across the three domain percentages is interpreted only as a descriptive cross-domain summary, not as an inter-rater or validity coefficient.

The field evaluation involved a convenience sample of 76 undergraduate students from the 2021 admission year who were enrolled in the South Asian History course in the History Education Study Program at Universitas Negeri Surabaya. Eligibility was defined by enrollment in the course. The participants were distributed across three intact classes: Class A = 27, Class B = 25, and Class C

= 24. No a priori power analysis or sample-size calculation was documented; the sample size reflected the number of enrolled students available for the field evaluation. Recruitment records, the number of students who declined participation, and any attrition before the analytic dataset was created were not documented. Consequently, the representativeness of the sample beyond this group of students and study program cannot be adequately assessed.

The implementation stage involved delivering the instructional package to three intact classes. The package was designed for 16 planned sessions combining synchronous conceptual guidance, discussion, and lecturer feedback with asynchronous source reading, worksheet completion, assignments, and evidence-based historical writing. Implementation was documented descriptively using the available Vinesa activity logs and records of student participation in discussions and completion of tasks and assessments.

However, the available records did not contain standardized class-level summaries of the number of sessions or instructional hours completed, attendance rates, activity-completion rates, technology-access problems, resource demands, or deviations from the planned sequence. No formal fidelity checklist or independent classroom observation was used, and the consistency of instructional delivery across the three classes could not be verified. Consequently, equivalence in instructional dosage and fidelity across classes could not be established, and the available evidence supports only a partial assessment of implementation feasibility. The outcome findings were therefore interpreted as preliminary short-term within-group changes rather than evidence of causal effectiveness.

The evaluation stage was primarily summative, with a limited formative function. The student-response questionnaire assessed perceptions of the learning materials, blended-learning implementation, and instructional-package design. The pretest and posttest assessed changes in aggregate self-reported historical literacy scores. Although student responses could identify aspects requiring improvement, the available records do not document specific revisions made after classroom implementation. The evaluation therefore provides evidence of responses acceptability and short-term within-group change, not a fully documented post-implementation revision cycle or causal effectiveness.

Data were collected using the three domain specific expert appraisal sheets, a 20 item student response questionnaire, and a 25 item historical literacy questionnaire administered as the pretest and posttest. The validation sheets were completed by a subject matter expert, a media expert, and a learning expert to assess the feasibility of the instructional package. The student response questionnaire examined students' perceptions of the learning materials, blended-learning implementation, and instructional-package design.

The historical literacy questionnaire was developed based on six indicators: reading historical sources, understanding chronology, understanding historical context, connecting events, comparing perspectives, and constructing historical arguments. Each item was rated on a four point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree). The same 25 item questionnaire and scoring procedure were used for the pretest and posttest. Item scores were summed, producing a possible total score ranging from 25 to 100. Because the available records do not contain factor-structure evidence supporting six dimensions or a higher order construct, the total is interpreted only as an aggregate self-reported historical literacy score.

The package also included source-analysis worksheets, historical argument assignments, and an analytic rubric designed to engage students in disciplinary historical practices. However, scores from these activities were not analyzed as outcome data, and the reliability of the scoring process was not examined. Consequently, these activities cannot be treated as direct performance evidence in the present evaluation.

Before the main field evaluation, both questionnaires were pilot-tested with 30 students who were not included in the main analytic cohort. Pearson product moment item total correlations were used as an item screening procedure. Items were retained when *sig* < .05 and the item total correlation exceeded the critical value of 0.361. Internal consistency was summarized using Cronbach's alpha. These statistics were not treated as evidence of content validity or dimensional structure. The available records do not include factor analysis, omega, inter-item correlations, or an item-redundancy analysis. Because the alpha coefficients were 0.973 and 0.979, their very high magnitude is interpreted cautiously as potentially indicating overlapping or redundant items. Table 3 summarizes the documented pilot results.

Table 3. Results of the Pilot Test for Item Total Correlations and Internal Consistency

Instrument	Number of items	Item Retained	Critical R	Item total R range	Significance	Cronbach's Alpha	Decision
Student response questionnaire	20	20	0.361	0.679-0.910	All *sig* < 0.05	0.973	All items retained; very high alpha warrants caution regarding redundancy
Historical literacy pretest–posttest questionnaire	25	25	0.361	0.712-0.895	All *sig* < 0.05	0.979	All items retained; very high alpha warrants caution regarding redundancy

Table 3 shows that all 20 items in the student response questionnaire and all 25 items in the historical literacy questionnaire met the predetermined item-screening criteria. Their item total correlations exceeded the critical value of 0.361 and were statistically significant at *sig* < 0.05. The item total correlations ranged from 0.679 to 0.910 for the student response questionnaire and from 0.712 to 0.895 for the historical literacy questionnaire; therefore, all items were retained for the field evaluation. The Cronbach's alpha coefficients were 0.973 and 0.979, respectively, indicating very high internal consistency in the pilot data. However, because coefficients of this magnitude may also indicate overlapping or redundant items, they were interpreted cautiously. These results support item retention and score consistency but do not, by themselves, establish content validity or the dimensional structure of either questionnaire.

All analyses were conducted using IBM SPSS Statistics version 21. Each expert-appraisal percentage was calculated and reported separately. Student-response percentages were calculated for each student and summarized by class and across the three classes. For both types of appraisal, the percentage P was calculated as follows:

$$\text{Percentage: } \frac{\text{total obtained score}}{\text{maximum score}} \times 100\%$$

The interpretation thresholds used in the original study are shown in Table 4. Although a descriptive cross domain average was calculated, the three validator percentages were not treated as interchangeable ratings or as an inter rater validity coefficient because each validator assessed a different domain using a separate instrument.

Table 4. Interpretation Criteria for Expert Appraisals and Student Responses

Range of percentage	Expert validation category	Student response category
81%-100%	Very Feasible	Very Good
61%-80%	Feasible	Good
41%-60%	Fair	Fair
21%-40%	Infeasible	Poor
0%-20%	Very Infeasible	Very Poor

The final analytic dataset contained 76 student-response records and 76 matched pretest–posttest total score pairs, as indicated by N = 76 at both measurements and df = 75 for the paired analysis. The available records do not document the identifier-matching procedure, the handling of item-level missing data, or whether partially completed questionnaires were excluded before the analytic dataset was created. Pretest and posttest total scores were summarized using means and standard deviations. The normality assumption for the paired-samples t test was examined using the Kolmogorov–Smirnov test on the posttest-minus-pretest difference scores. The test yielded a significance value of 0.200 ($sig > 0.05$), indicating no statistically significant departure from normality. Therefore, the normality assumption was considered satisfied, and the paired-samples t test was used to evaluate the mean change between the pretest and posttest scores.

Short term within group change in aggregate self-reported historical literacy scores was analyzed using a paired-samples t test. The analysis reported the posttest minus pretest mean difference, standard deviation and standard error of the difference, 95% confidence interval, t value, degrees of freedom, and two tailed *sig* value, with $sig < 0.05$ as the significance threshold. Class membership was not modeled: no intraclass correlation, multilevel analysis, cluster adjusted standard error, class specific paired outcome analysis, or sensitivity analysis was documented. Dependence within the three intact classes may therefore have affected the estimated uncertainty.

The standardized within group magnitude of change was estimated using Cohen's d_z:

$$d_z = \frac{D}{SD_D}$$

where D is the mean of the individual posttest-minus-pretest differences and SD_D is the standard deviation of those differences. Absolute dz values of approximately 0.20, .50, and .80 were interpreted as small, medium, and large, respectively.

Normalized gain was additionally calculated from the group mean scores as a supplementary descriptive measure of the proportion of attainable score improvement:

$$g = \frac{\text{posttest} - \text{pretest}}{100 - \text{pretest}}$$

Values below 0.30 were classified as low, values from 0.30 to 0.70 as medium, and values above 0.70 as high (Sundayana, 2014). Normalized gain was interpreted descriptively rather than as a standardized effect size estimate.

Result and Discussion

Needs Analysis Findings

The exploratory needs analysis drew on earlier institutional studies and a review of the South Asian History course plan and available teaching materials. A 2018 study conducted in the same study program and university reported that 80.69% of participating students perceived that Vinesa supported their active involvement in learning. A subsequent institutional review conducted in 2020 in the same study program and university covered ten courses. It found that 70% of online history instruction still relied on WhatsApp Groups and email, whereas dedicated online-learning applications were used in only 15% of the reviewed courses. These findings provided contextual evidence from different years and were not derived from the 76 participants in the present field evaluation.

The course review identified limited Indonesian-language resources, students' unfamiliarity with foreign-language literature, and the extensive scope of South Asian History as practical instructional problems. Available resources consisted mainly of lecturer-prepared slides, selected e-books, videos, and links distributed across course topics. Because the preliminary review did not employ a formal document-selection or coding protocol, these findings were treated descriptively. They informed requirements for an organized learning environment, accessible and guided engagement with course resources, complementary synchronous and asynchronous activities, and a sequenced package for the planned course sessions. The documented findings are summarized in Table 5.

Table 5. Summary of the Exploratory Needs Analysis

Evidence Source	Documented Finding	Design Requirement
2018 institutional study in the same study program and university	80.69% of participating students perceived active involvement through Vinesa	Retain Vinesa as the central digital learning environment
2020 institutional review in the same study program and university (10 courses)	70% relied on WhatsApp Groups and email; dedicated online-learning applications were used in 15% of the reviewed courses	Integrate resources, assignments, discussions, and feedback within an LMS
Review of the course plan and teaching materials	Limited Indonesian language resources, unfamiliarity with foreign-language literature, and extensive course scope	Organize accessible resources, provide guided source engagement activities, and sequence the package for 16 planned sessions

Note: The historical literacy pretest was the baseline outcome measure and was not treated as needs analysis evidence. Alignment with the six historical literacy indicators was a theory-based design decision.

Table 5 shows that the documented findings from the three evidence sources were used to identify practical requirements for developing the instructional package. The 2018 finding concerning students' perceived active involvement through Vinesa supported the decision to retain Vinesa as the central digital learning environment. The 2020 institutional review indicated the need to integrate learning resources, assignments, discussions, and feedback within an LMS because online history instruction still relied largely on WhatsApp Groups and email, while dedicated online learning applications were used to a limited extent. The review of the course plan and teaching materials further indicated the need to organize accessible resources, provide guided engagement with historical sources, and sequence the package across 16 planned sessions.

Design of the Blended Instructional Package

The design of the instructional package is based on the results of a needs analysis that identifies three main issues. The learning device is not only designed to deliver South Asian History material but also to guide students in gradually engaging in historical literacy activities. The developed device includes lesson plans, teaching materials, learning resources, student worksheets, assignments, discussion activities, and evaluations. Each component is structured with reference to six indicators of historical literacy, namely reading historical sources, understanding chronology, understanding historical context, connecting events, comparing perspectives, and constructing historical arguments.

The semester learning plan organized the course into a chronological and thematic sequence. Weeks 1–2 introduced the geographical and cultural foundations of South Asia, Weeks 3–5 addressed Hindu–Buddhist kingdoms and influences, Weeks 6–9 examined the spread of Islam and the development of Mughal rule, Weeks 10–13 focused on European expansion, colonialism, and local resistance, and Weeks 14–15 addressed independence and postcolonial national identity. Each unit combined synchronous conceptual guidance with asynchronous access to materials and task completion through Vinesa.

A representative example was the Week 12 unit on conflict between South Asian societies and British colonial authorities. During the synchronous online session, the lecturer introduced British colonial policies and local conflicts using presentation slides and audiovisual materials. Students subsequently accessed the materials asynchronously through Vinesa and completed a task requiring them to analyze British colonial policies and explain the emergence of local resistance. The unit therefore moved from conceptual framing and contextual explanation to independent examination and written historical analysis across the two learning modalities.

Table 6. Example of the Blended Sequence in the Unit on British Colonialism and Local Resistance

Phase	Modality	Learning Activity	Historical Literacy Focus
Conceptual framing	Synchronous online	Introduction to British colonial policies and internal conflicts using slides and audiovisual materials	Historical context and Chronology
Material examination	Asynchronous through Vinesa	Students accessed and reviewed the provided learning materials	Reading historical information and understanding context

Analytical task	Asynchronous through Vinesa	Students analyzed colonial policies and explained the emergence of local resistance	Connecting events and constructing historical arguments

Table 6 illustrates that the two modalities served complementary purposes. The synchronous phase established the conceptual and historical setting, whereas the asynchronous phase provided time for students to examine the materials and develop a written analysis of the relationship between colonial policy and local resistance.

The teaching materials were developed with consideration of the breadth and interconnectedness of South Asian history. Rather than presenting the content solely as a sequence of events, the materials combined chronological and thematic organization. The major themes included the development of civilizations, religion and culture, colonialism, nationalism, political conflict, social change, and the formation of regional identities. Each theme presented its historical background, development, causal relationships, regional consequences, and connections with other periods. This organization was intended to support chronological, contextual, and multiperspective understanding and to make continuity and change across historical periods more explicit.

Learning resources were selected according to their relevance to the course topics and learning outcomes, identifiable authorship or institutional provenance, historical and contextual relevance, accessibility through Vinesa, and suitability for source comparison. The package included books, scholarly articles, maps, historical documents, images, digital archives, and selected online materials. Because Indonesian-language resources on South Asian history were limited, both Indonesian-language and selected foreign-language materials were included and integrated with lecturer-prepared explanations and guided worksheets.

Online materials were not included solely because they were easily accessible. During package development, their authorship or institutional origin, purpose, historical setting, evidence, and consistency with other sources were examined. Sources addressing the same issue were placed side by side to facilitate corroboration and perspective comparison. The selected materials were subsequently reviewed by the subject-matter expert for content relevance and conceptual accuracy and by the media expert for readability, accessibility, and usability. This review process reduced the likelihood that unsupported or low-quality online materials entered the instructional package.

The student worksheet was designed to scaffold historical literacy practices progressively. Students were guided to examine the authorship, origin, date, and purpose of sources; identify relevant information; organize events chronologically; explain their historical setting; compare information across sources; and construct evidence-based historical arguments. The completed worksheets were subsequently used during synchronous discussions and lecturer feedback. A representative worksheet prompt and its scoring criteria are presented in Table 7.

Table 7. Representative Worksheet Prompt and Scoring Criteria

Component	Representative prompt	Scoring criteria
Sourcing	Identify the author or creator, origin, date, intended audience, and	Accurate identification of authorship, origin, date, audience, and purpose

	purpose of each source concerning British colonial policy and local resistance	
Contextualization	Explain the political and social conditions in which each source was produced	Clear connection between the source and its relevant historical setting
Corroboration	Compare the two sources. What claims or evidence are consistent, different, or contradictory	Appropriate comparison of agreements, differences, and supporting evidence
Argumentation	Using evidence from both sources, explain how British colonial policies contributed to the emergence of local resistance	Clear claim, relevant evidence, logical reasoning, and acknowledgement of differing perspectives

Table 7 illustrates how the worksheet guided students through four interconnected historical literacy practices. Students first examined the authorship, origin, date, audience, and purpose of the sources before relating them to their historical setting. They then compared agreements, differences, and contradictions across the sources and used the resulting evidence to construct a historical argument.

Table 8. Analytic Scoring Rubric for the Historical Literacy Worksheet

Score	Criterion
0	No relevant response or unsupported response
1	Limited, unclear, or weakly connected to the sources
2	Generally accurate but incomplete or insufficiently supported
3	Complete, accurate, and supported by relevant historical evidence

Table 8 presents the analytic rubric used to assess students' responses to the historical literacy worksheet. Each dimension was scored independently on a scale from 0 to 3. A score of 0 indicated the absence of a relevant response, 1 indicated a limited or unclear response, 2 indicated a generally accurate but incomplete response, and 3 indicated a complete, accurate, and evidence supported response. The four dimensions produced a maximum total score of 12.

Individual and group assignments were aligned with the learning themes and the six historical literacy indicators. Individual assignments were designed to require students to examine historical sources and prepare initial interpretations. These interpretations were subsequently brought into group discussions, in which students compared perspectives and formulated historical arguments. During synchronous sessions, students were expected to present their analyses, defend

their claims using historical evidence, and refine their reasoning in response to peer and lecturer feedback. This sequence was intended to support the construction of more focused, logical, and evidence-based historical arguments. The worksheets, assignments, and discussion products served as components of the planned learning activities but were not analyzed as direct performance-outcome data in the present evaluation.

The evaluation used a 25 item historical literacy questionnaire administered before and after implementation of the instructional package and a separate 20 item student-response questionnaire administered after the learning activities. The historical literacy questionnaire covered six indicators: reading historical sources, understanding chronology, understanding historical context, connecting events, comparing perspectives, and constructing historical arguments. The same questionnaire and scoring procedure were used at the pretest and posttest stages to examine changes in students' total self-reported historical literacy scores. The student response questionnaire was used separately to assess students' perceptions of the instructional package after its implementation. The alignment among the instructional package components, planned learning activities, intended learning evidence, and historical literacy indicators is presented in Table 9.

Table 9. Alignment of Instructional Package Components, Learning Activities, Evidence, and Historical Literacy Indicators

Instructional Package Component	Function in Instruction	Related Learning Activities and Modality	Learning Evidence or Measurement	Historical Literacy Indicators
Semester Learning Plan	Organizes learning outcomes, the sequence of topics, synchronous and asynchronous activities, assignments, and evaluation procedures	Synchronous conceptual guidance and discussion, asynchronous reading, source examination, worksheet completion, and written tasks	Planned learning activities, assignments, and evaluation procedures documented in the course sequence	All historical literacy indicators
Teaching Materials	Provides the chronological, thematic, and conceptual foundation of South Asian History	Asynchronous reading followed by synchronous explanation, clarification, and discussion	Chronological notes, contextual explanations, and responses to guiding questions	Understanding chronology, understanding historical context, connecting events
Learning Resources	Provides historical information and evidence from books, scholarly articles, maps, historical documents, images,	Asynchronous source examination and synchronous comparison of interpretations	Source analysis notes and comparisons of information across sources	Reading historical sources, understanding historical context, comparing perspectives

	digital archives, and selected online materials			
Student Worksheets	Scaffolds students' movement from source examination to contextualization, corroboration, and historical argumentation	Asynchronous worksheet completion followed by synchronous discussion and lecturer feedback	Completed worksheets assessed the analytic scoring rubric	All historical literacy indicators
Individual and Group Assignments	Supports independent source analysis and collaborative interpretation	Individual source examination and written analysis, group comparison, discussion, and presentation	Individual historical notes, group analyses presentations, and evidence based written arguments	Reading historical sources, connecting events, comparing perspectives, constructing historical arguments
Discussion Activities	Provides opportunities to compare interpretations, defend claims, and revise arguments through feedback	Synchronous presentation, peer response, evidence based discussion, and lecturer feedback	Oral presentations, responses to alternative interpretations, and revised arguments	Comparing perspectives, constructing historical arguments
Historical Literacy Questionnaire	Measured aggregate Self-reported historical literacy	Same 25 item questionnaire at pretest and posttest	Total questionnaire score ranging from 25 to 100	All historical literacy indicators
Student Response Questionnaire	Measured perceived acceptability of the package	20 item questionnaire after implementation	Percentage scores for perceived clarity, usability, and usefulness	Not a historical literacy performance me

Note. The worksheets, assignments, and discussion products were components of the planned learning activities. They were not analyzed as direct performance outcome data, and no documented performance-scoring procedure was used. The historical literacy questionnaire measured students' self-reported historical literacy rather than demonstrated disciplinary performance.

Table 9 shows that each component of the instructional package served a distinct but interconnected function. Teaching materials and learning resources provided the conceptual content and historical evidence, while worksheets, assignments, and discussions transformed these resources into structured historical literacy activities. However, because worksheet, assignment, and discussion scores were not analyzed using a documented performance-scoring procedure, the table demonstrates design alignment rather than improved performance. The pretest and posttest questionnaire measured students' self-reported historical literacy, while the student-response questionnaire measured their perceptions of the instructional package rather than their historical literacy performance.

Preliminary Expert Appraisal of the Instructional Package

The package was appraised by one subject-matter validator, one media validator, and one learning validator. Each validator assessed a different domain using a separate, non-equivalent appraisal instrument. Their roles, appraisal domains, and numbers of items are shown in Table 10.

Table 10. Validator Roles, Appraisal Domains, and Number of Items

Validator Role	Appraisal Aspects	Number of Items
Subject matter expert	Content, learning, language, and material graphics	25
Media expert	Learning-design, language, graphic, and learning-activities	20
Learning expert	Content; opening, core, closing activities; implementation; language; and instructional media graphics	40

Table 10 shows that the preliminary appraisal involved three validators with distinct areas of responsibility. The subject matter expert assessed 25 items covering content, learning, language, and material graphics. The media expert assessed 20 items related to learning design, language, graphics, and learning activities. The learning expert assessed 40 items covering content, opening, core, and closing activities, learning implementation, language, and instructional-media graphics. Because each validator examined a different domain of the instructional package, the appraisal results were interpreted separately before being summarized descriptively.

Each appraisal used a five-point response scale, and the obtained score was converted to a percentage of the maximum possible score. Table 11 reports the three domain specific results separately.

Table 11. Domain Specific Expert Appraisal Results

Validator	Percentage Score	Category
Subject matter expert	88%	Very feasible
Media expert	85.20%	Very feasible
Learning expert	87.60%	Very feasible
Overall average	86.93%	Very feasible

Note. The simple arithmetic mean across the three domain percentages was 86.93%. It is reported only as a descriptive cross-domain summary, not as an inter-rater agreement, validity coefficient, or pooled rating of a common construct.

Table 11 reports the three domain-specific appraisal results separately. The subject-matter expert assigned a score of 88.00%, the media expert 85.20%, and the learning expert 87.60%. Each result independently fell within the very feasible category according to the predetermined interpretation thresholds. An equal-weight arithmetic average of 86.93% is presented only as a descriptive cross-domain summary after the three scores were expressed on a common percentage scale. Because the validators assessed different domains using separate, non-equivalent instruments, this average was not interpreted as a pooled validity score or an inter-rater validity coefficient.

Although all three appraisal scores fell within the very feasible category, but they represented different domains and were not treated as interchangeable judgments. The available

records also documented recommendations concerning access to materials, asynchronous interaction, and technical guidance for Vinesa activities, as summarized in Table 12.

Table 12. Documented Expert Feedback on the Instructional Package

Validator	Feedback Focus	Documented Feedback
Media expert	Accessibility of learning materials	Learning materials linked through Vinesa should be easier for students to download
Media expert	Asynchronous interaction	Chat based activities should be strengthened to make student interaction more engaging and traceable
Learning expert	Technical guidance for Vinesa activities	Technical instructions for accessing and completing learning activities through Vinesa should be clearer and easier to understand

Table 12 shows that recommendations documented during the expert-appraisal process. The media expert emphasized the accessibility of downloadable materials and the need for stronger asynchronous interaction, while the learning expert emphasized clearer technical guidance for activities conducted through Vinesa. These recommendations provided practical directions for refining the accessibility and implementation of the instructional package before classroom use.

Implementation of the Instructional Package and Student Responses

The instructional package was implemented with 76 students enrolled in the South Asian History course. The participants were divided into three intact classes: Class A consisted of 27 students, Class B of 25 students, and Class C of 24 students. During synchronous activities, the lecturer provided conceptual framing, guided discussions of historical issues, and gave feedback on students' analyses. During asynchronous activities through Vinesa, students accessed teaching materials, examined additional sources, completed worksheets and assignments, and constructed source based historical arguments.

Students' responses to the instructional package were collected after implementation using a 20 item student-response questionnaire covering the content of the learning materials, blended learning activities, and instructional package design. Before the main implementation, the questionnaire was pilot tested with 30 students who were not included in the main study sample. All items met the validity criteria because their calculated item total correlation coefficients exceeded the critical value of 0.361 and their significance values were below 0.05. The questionnaire also demonstrated acceptable internal consistency, with a Cronbach's alpha coefficient above 0.60.

The questionnaire was completed by all 76 participating students, corresponding to a response rate of 100%. The resulting scores were interpreted as students' perceptions of the clarity, usability, and usefulness of the instructional package rather than as direct evidence of implementation fidelity, instructional quality, or learning effectiveness. Because the data were self-reported within a course setting, the favorable responses may have been influenced by social-desirability tendencies. Therefore, the findings were interpreted cautiously and considered complementary to, rather than a substitute for, the learning outcome analysis

Because the three subscales contained different numbers of items and consequently had different maximum score ranges, the subscale scores were converted into normalized percentages before interpretation. The normalized student response results across the three classes are presented in Table 12.

Table 13. Normalized Student Response Percentages Across the Three Classes

Class	Learning Material Content	Blended History Learning Implementation	Teaching Device Design	Overall Response	Category
A	86.07%	89.44%	82.83%	86.62%	Very good
B	77.87%	85.74%	85.74%	84.56%	Very good
C	77.67%	85.56%	82.40%	83.27%	Very good
Average	80.53%	86.92%	83.66%	84.82%	Very good

Note. The 20 item questionnaire used a four point scale. The learning-material content, blended learning activities, and instructional-package design subscales consisted of 3, 10, and 7 items, with maximum raw scores of 12, 40, and 28, respectively. Each subscale percentage was calculated by dividing the mean obtained subscale score by its maximum possible score and multiplying by 100. The overall response percentage was calculated using a maximum total score of 80.

Table 13 shows that the overall student-response scores were 86.62% for Class A, 84.56% for Class B, and 83.27% for Class C. The descriptive average across the three classes was 84.82%, which fell within the very good category. The average normalized percentages were 80.53% for learning-material content, 86.92% for blended-learning activities, and 83.66% for instructional-package design. These findings indicate generally favorable student responses to the three components of the instructional package. The class level scores were interpreted descriptively and were not used to infer meaningful differences among the classes or direct evidence of instructional effectiveness.

The favorable responses suggested that students generally perceived the complementary use of synchronous and asynchronous activities positively. Synchronous sessions provided conceptual explanations, historical discussions, and lecturer feedback, whereas asynchronous activities provided opportunities to access materials, examine additional sources, complete worksheets, and prepare historical arguments. This pattern was consistent with the intended design of blended learning as an integrated learning experience rather than as a collection of separate synchronous and asynchronous activities (Istenič, 2024).

Nevertheless, the favorable responses were not interpreted as direct evidence of instructional-design quality or effectiveness. Because the data were self-reported within a course setting, the responses may have been influenced by the novelty of the blended learning activities, students' expectations regarding the lecturer's preferred responses, or perceptions that their responses could be associated with course assessment. The student-response results were therefore treated as complementary evidence and interpreted alongside the expert appraisals and pretest–posttest findings.

Short-Term Changes in Students' Self-Reported Historical Literacy

Short-term changes in students' self-reported historical literacy were examined using the same 25-item historical literacy questionnaire administered before and after the implementation of the instructional package. Each item was rated on a four-point scale, producing a possible total score ranging from 25 to 100. The normality assumption for the paired samples t test was examined using the distribution of the posttest minus pretest difference scores. The documented significance value is shown in Table 14.

Table 14. Normality Test for the Paired Difference Scores

Variable	Statistic	df	*Sig*	Interpretation
Pretest	0.086	76	0.200	No statistically significant departure from normality

Table 14 shows that the test did not indicate a statistically significant departure from normality (*sig* > .200); therefore, the normality assumption for the paired samples t test was considered satisfied. Descriptive statistics for the two measurements are reported in Table 15.

Table 15. Descriptive Statistics for Self-Reported Historical Literacy Scores

Measurement	N	Minimum	Maximum	Mean	Std. Deviation
Pretest	76	26	89	51.03	14.55
Posttest	76	77	100	88.93	4.91

Table 15 shows that the mean self-reported historical literacy score increased from 51.03 with a standard deviation of 14.55 at pretest to 88.93 with a standard deviation of 4.91 at posttest. This represents an increase of approximately 37.90 points in the mean score. The posttest scores were more tightly clustered, and the observed maximum was 100. The smaller posttest standard deviation describes score compression near a high mean; it does not by itself demonstrate convergence in students' historical literacy competence.

The paired samples t test was conducted to determine whether the difference between the pretest and posttest scores was statistically significant. The analysis also included the 95% confidence interval and Cohen's d_z estimate the magnitude of the change. The results are presented in Table 15.

Table 16. Paired-Samples Test and Effect Size for Self-Reported Historical Literacy Scores

Comparison	Mean Difference	SD Difference	SE Difference	95% CI Lower	95% CI Upper	*T*	df	*Sig*	Cohen's d_z
Posttest – Pretest	37.91	14.79	1.70	34.53	41.29	22.34	75	< 0.05	2.56

Table 16 shows that the mean paired difference, calculated as posttest minus pretest, was 37.91 points, with a standard deviation of 14.79. The paired samples t-test indicated that the posttest scores were significantly higher than the pretest scores, with a t-value of 22.34, 75 degrees of freedom, and a significance value below 0.05. The 95% confidence interval for the mean difference ranged from 34.53 to 41.29. Cohen's d_z was 2.56, indicating an exceptionally large standardized within-group change. This estimate quantifies change within the observed group without a comparison group, it does not establish that the instructional package caused the increase.

As a supplementary descriptive analysis, normalized gain was calculated from the group mean pretest and posttest scores. The result is presented in Table 17.

Table 17. Group Level Normalized Gain

Pretest Mean	Posttest Mean	Maximum Score	N-gain	Percentage	Category
51.03	88.93	100	0.77	77.39%	High

Table 17 shows that the group level N-gain was 0.77, equivalent to 77.39%, and was classified as high. The N-gain was interpreted as supplementary descriptive evidence rather than as a

standardized effect size. Because it was based on the aggregate questionnaire score, it did not indicate separate changes in each of the six historical literacy indicators.

The instructional package was designed to align materials, worksheets, assignments, and discussions with the six historical literacy indicators. These components provided planned opportunities for students to examine sources, understand chronology and context, connect events, compare perspectives, and construct evidence-based arguments. This alignment may have contributed to the observed increase in aggregate self-reported historical literacy scores; however, the proposed mechanism was not tested in the present study. The study did not isolate the contribution of individual components, analyze changes in each indicator separately, or examine students' learning-task performance. Therefore, explaining how the package may have supported the observed change requires further process and performance data.

Discussion

The findings should be interpreted across several distinct forms of evidence. The domain specific expert appraisals were favorable, with the descriptive cross domain average of 86.93%, while the overall student-response score of 84.82% indicated generally positive acceptability of the instructional package. Separately, aggregate self-reported historical literacy scores increased from 51.03 to 88.93. These findings address different constructs and should not be treated as interchangeable evidence of instructional effectiveness (de Bruijn-Smolders & Prinsen, 2024; McCarthy & Palmer, 2023). In particular, the expert appraisals represent preliminary judgments of different aspects of the package, the student-response questionnaire reflects perceived acceptability, and the pretest-posttest analysis represents short-term within group change in self-reported historical literacy. Implementation feasibility was assessed only partially because the available records did not provide standardized information on completed instructional dosage, attendance, activity completion rates, technology access problems, or deviations from the planned sequence (Anthony et al., 2022; Rasheed et al., 2020).

The instructional design nevertheless illustrates how synchronous and asynchronous activities can be organized as complementary rather than separate components of blended learning. Synchronous activities provided conceptual framing, discussion, clarification, and lecturer feedback, whereas asynchronous activities provided additional opportunities to access materials, examine historical sources, complete analytical tasks, and develop written interpretations. This arrangement is consistent with Istenič (2024) emphasis on purposeful integration in blended learning rather than the simple addition of a digital platform to conventional instruction (Anthony et al., 2022; Heilporn et al., 2021; McCarthy & Palmer, 2023; Ridha et al., 2024). The present study adds a discipline-specific dimension to that principle by aligning the two modalities with defined historical literacy processes (Kaviza, 2020; Tanjung et al., 2024). However, because implementation fidelity and instructional dosage were not systematically documented, the extent to which this intended integration was enacted consistently across all three classes cannot be established.

The magnitude of the observed within-group change also requires cautious interpretation. Cohen's *dz * of 2.56 represents an exceptionally large standardized change, but it should not be interpreted straightforwardly as an equally large improvement in demonstrated historical literacy competence. The posttest mean of 88.93, SD=4.91, and maximum score of 100 indicate substantial score compression near the upper boundary of the scale and raise the possibility of a ceiling effect (Staus et al., 2021). In addition, the same self-report questionnaire was administered before and after

instruction. Repeated exposure may have sensitized students to the targeted practices, while instruction itself may have changed how students understood and judged the questionnaire items, producing a possible response-shift effect (Staus et al., 2021). Demand characteristics, lecturer influence, and the use of the same measurement method at both time points may also have contributed to common method inflation (Baumgartner et al., 2021). The group level normalized gain of 0.77 should therefore be treated only as a descriptive transformation of the same pretest and posttest means rather than as independent corroborating evidence of instructional effectiveness.

At the disciplinary level, the package was designed to move students beyond access to historical information by structuring engagement with source provenance, historical setting, evidence, and competing accounts (Cinnamon et al., 2021; Mulyana & Kurniawati, 2020; Santiago & Dozono, 2022; Widiadi et al., 2022; Wilson et al., 2023). This design is consistent with Eem et al. (2023), who emphasize the procedural knowledge involved in evaluating the trustworthiness of historical sources. The convergence, however, is primarily at the level of instructional principles. Whereas procedural knowledge in prior research can be examined through students' performance in source-evaluation tasks, the present evaluation relied on an aggregate self-report questionnaire. Consequently, the increase in the total score cannot establish that sourcing, chronological understanding, contextualization, connections among events, perspective comparison, and argumentation improved to different degrees, or that any particular indicator improved at all. Indikator level analyses and performance based measures would be required to determine changes in specific disciplinary practices (Fikri et al., 2023; Purwanta, 2023; ter Beek et al., 2022).

This distinction is particularly important for historical argumentation, one of the disciplinary practices explicitly incorporated into the instructional package. The worksheets and assignments guided students to identify the provenance and purpose of historical sources, relate information to its historical setting, compare agreements and contradictions across sources, and formulate claims supported by relevant evidence. These activities show that historical argument construction was deliberately scaffolded within the instructional design (Allagui, 2024; Cinnamon et al., 2021; Wilson et al., 2023). This design is consistent with research showing that source-based argumentative writing requires students to synthesize information across texts, select relevant evidence, and explain the relationship between evidence and historical claims (Steiss et al., 2024). However, the present study did not analyze students' worksheet or written-assignment performance as outcome data or examine the reliability of rubric based scoring. Consequently, the findings support the presence of explicit scaffolding for historical argumentation in the instructional design, but they do not establish improvement in students' demonstrated historical argumentation performance.

The findings also reinforce the distinction between access to technology and disciplinary use of technology (Breakstone et al., 2021; Hahnel et al., 2020; McGrew, 2021; Walsh-Moorman et al., 2020). Kumalasari et al. (2024) show that familiarity with digital tools does not necessarily produce digital historical literacy when guidance in evaluating and interpreting information is limited. In this study, the package's design contribution lay in organizing engagement with sources through curated materials, guided tasks, synchronous discussion, and asynchronous analysis rather than simply increasing access to digital resources (Hoelscher, 2024; Lesley et al., 2023; Widiadi et al., 2022; Yulifar & Aman, 2023). This interpretation also accords with Lionar et al. (2024), who emphasize inquiry, historical writing, and the construction of historical understanding as pathways for developing historical literacy. Yet the present results show only that these practices were

incorporated into the instructional design and that students subsequently reported higher aggregate historical literacy scores, they do not demonstrate performance gains in those practices.

The comparison with Alrianingrum et al. (2023) further clarifies the specific contribution of the present design. Both studies used digital resources and complementary synchronous and asynchronous activities in history education, demonstrating convergence in their broader blended and connectivist orientation (Firmansyah & Sastra Atmaja, 2025; Kaviza, 2020; Rahman et al., 2021; Tanjung et al., 2024). The present package differed by explicitly mapping learning objectives, instructional components, modalities, activities, learning evidence, and six historical literacy indicators, while the worksheets further specified sourcing, contextualization, corroboration, and evidence-based argumentation. The studies also examined different outcome constructs and used different measurement procedures: Alrianingrum et al. (2023) emphasized conceptual learning outcomes, digital literacy, and student responses, whereas the present study examined preliminary expert appraisals, student acceptability, and aggregate self-reported historical literacy. Consequently, the present study extends the earlier work primarily through more explicit discipline-specific alignment rather than through evidence of superior learning effects, and numerical effect-size comparisons between the two studies would not be warranted.

Variation in the class-level student responses provides an additional reason for caution. Overall response scores were 86.62% for Class A, 84.56% for Class B, and 83.27% for Class C. Although all were categorized as very good, the variation indicates that students' experiences were not completely uniform across the three intact classes. The available records did not permit examination of whether class-specific pretest–posttest distributions or standardized changes showed a similar pattern, nor did they provide standardized class-level information on attendance, task completion, instructional dosage, platform participation, or implementation deviations (de Bruijn-Smolders & Prinsen, 2024; Heilporn et al., 2021; Rasheed et al., 2020). Because students were nested within three intact classes and the paired analysis was conducted at the individual level without adjustment for clustering, the reported uncertainty may also have been affected by within-class dependence. These issues limit conclusions about consistency of implementation and outcomes across classes.

Accordingly, the short-term increase in aggregate self-reported historical literacy scores should be interpreted as an association observed after implementation rather than as evidence that the package caused the change (Reynolds, 2022). The study used a convenience sample from one study program, lacked a comparison group and delayed posttest, relied on self-reported scores measured using the same questionnaire before and after implementation, and did not analyze performance-based outcomes or systematic process data. Maturation, testing reactivity, response shift, demand characteristics, lecturer influence, common-method effects, and class-level clustering remain plausible alternative explanations. The present evaluation therefore provides preliminary evidence that the discipline-specific blended package was favorably appraised and accepted by students and was accompanied by a substantial short-term change in self-reported historical literacy. Whether the same design produces reliable changes in demonstrated historical literacy performance, across individual indicators, classes, institutions, and longer periods remains a question for subsequent controlled and process-oriented research (Wilson et al., 2023).

Conclusion

This study developed a discipline-specific blended instructional package for South Asian History that aligned synchronous guidance and asynchronous inquiry with six historical literacy indicators. Domain-specific expert appraisals and student responses indicated that the package was preliminarily feasible and acceptable for use in the study context. Its implementation was also associated with increased aggregate self-reported historical literacy scores in the short term. However, these findings should not be interpreted as evidence of improvement in demonstrated historical literacy performance or as a causal effect of the instructional package, because the evaluation relied on self-reported outcomes, involved a convenience sample from a single study program, and did not include a comparison group.

The explicit alignment among learning outcomes, instructional activities, learning modalities, and historical literacy indicators provides a design proposition for further evaluation rather than an established transferable contribution. Future research should test this approach across multiple institutions using controlled designs, delayed measurements, indicator-level analyses, systematic implementation-fidelity assessment, and authentic historical tasks scored by blinded raters. Such evaluation is necessary to determine whether the observed short-term changes in self-reported scores are accompanied by sustained and demonstrable changes in specific historical literacy practices.

References

Allagui, B. (2024). A scaffolding intervention to improve self-efficacy in source-based argumentative writing. *Frontiers in Psychology, 15*, Article 1454104. https://doi.org/10.3389/fpsyg.2024.1454104

Alrianingrum, S., Rianto, Y., & Bahri, B. S. (2023). Blended learning development based on connective theory to improve results of learning concepts and digital literacy. In *Proceedings of the International Joint Conference on Arts and Humanities 2022 (IJCAH 2022)* (pp. 1129–1138). Atlantis Press. https://doi.org/10.2991/978-2-38476-008-4_121

Anthony, B., Jr., Kamaludin, A., Romli, A., Mat Raffei, A. F., Eh Phon, D. N. A. L., Abdullah, A., & Gan, L. M. (2022). Blended learning adoption and implementation in higher education: A theoretical and systematic review. *Technology, Knowledge and Learning, 27*, 531–578. https://doi.org/10.1007/s10758-020-09477-z

Baumgartner, H., Weijters, B., & Pieters, R. (2021). The biasing effect of common method variance: Some clarifications. *Journal of the Academy of Marketing Science, 49*(2), 221–235. https://doi.org/10.1007/s11747-020-00766-8

Bhargava, M., & Nath, P. (Eds.). (2022). *The early modern in South Asia: Querying modernity, periodization, and history*. Cambridge University Press. https://doi.org/10.1017/9781009215381

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher, 50*(8), 505–515. https://doi.org/10.3102/0013189X211017495

Cinnamon, S. A., Rivera, M. O., & Sellers, H. K. D. (2021). Teaching disciplinary literacy through historical inquiry: Training teachers in disciplinary literacy and historical inquiry instructional practices. *The Journal of Social Studies Research, 45*(4), 241–252. https://doi.org/10.1016/j.jssr.2021.03.001

de Bruijn-Smolders, M., & Prinsen, F. R. (2024). Effective student engagement with blended learning: A systematic review. *Heliyon, 10*(23), Article e39439. https://doi.org/10.1016/j.heliyon.2024.e39439

Fikri, A., Aman, A., Suarman, S., Isjoni, I., & Sumarno, S. (2023). Development of a project assessment instrument to measure skills using historical sources for history study program students at the University of Riau. *Jurnal Ilmu Pendidikan, 29*(1), 48–57. https://doi.org/10.17977/um048v29i12023p48-57

Firmansyah, H., & Sastra Atmaja, T. (2025). Transformation of history learning methods in the digital era: Challenges and opportunities in schools. *Indonesian Journal of Educational Development (IJED), 6*(1), 264–277. https://doi.org/10.59672/ijed.v6i1.4791

Hahnel, C., Eichmann, B., & Goldhammer, F. (2020). Evaluation of online information in university students: Development and scaling of the screening instrument EVON. *Frontiers in Psychology, 11*, Article 562128. https://doi.org/10.3389/fpsyg.2020.562128

Heilporn, G., Lakhal, S., & Bélisle, M. (2021). An examination of teachers' strategies to foster student engagement in blended learning in higher education. *International Journal of Educational Technology in Higher Education, 18*, Article 25. https://doi.org/10.1186/s41239-021-00260-3

Hoelscher, C. (2024). Digital primary source literacy: Developing skills for online archival research. *New Directions for Teaching and Learning, 2024*(180), 39–47. https://doi.org/10.1002/tl.20607

Istenič, A. (2024). Blended learning in higher education: The integrated and distributed model and a thematic analysis. *Discover Education, 3*, Article 165. https://doi.org/10.1007/s44217-024-00239-y

Jayaneththi, R. (2025). South Asian historiography from the British period to today. *História da Historiografia: International Journal of Theory and History of Historiography, 18*, 1–24. https://doi.org/10.15848/hh.v18.2587

Kaviza, M. (2020). Pelaksanaan kelas flipped dengan sumber digital dalam mata pelajaran sejarah: Kesannya terhadap kemahiran pemikiran sejarah. *Journal of ICT in Education, 7*(1), 30–42. https://doi.org/10.37134/jictie.vol7.1.4.2020

Kumalasari, D., Purwanta, H., & Aw, S. (2024). Comparative analysis of Generation Z's digital history literacy in history education majors on Java Island: A study of history digital literacy. *Journal of Education and E-Learning Research, 11*(1), 90–96. https://doi.org/10.20448/jeelr.v11i1.5342

Kumalasari, D., Purwanta, H., Aw, S., & Agustinova, D. E. (2022). Historical literacy competencies of history education students: Case studies at Surabaya and Yogyakarta State Universities. *Journal of Curriculum and Teaching, 11*(8), 339–348. https://doi.org/10.5430/jct.v11n8p339

Lesley, M., Stewart, E. A., & Keene, J. (2023). Exploring the connections between disciplinary and digital literacies in history. *Journal of Adolescent & Adult Literacy, 67*(2), 53–64. https://doi.org/10.1002/jaal.1308

Lionar, U., Supriatna, N., Mulyana, A., Winarti, M., & Sumaludin, M. M. (2024). Historical literacy in social studies learning at junior high school in Bandung City: Analysis of implementation and challenges. *Indonesian Journal of Social Science Education, 6*(2), 97–109. https://doi.org/10.29300/ijsse.v6i2.3953

Maposa, M., & Wassermann, J. (2009). Conceptualising historical literacy: A review of the literature. *Yesterday and Today, 4*, 41–66.

McCarthy, S., & Palmer, E. (2023). Defining an effective approach to blended learning in higher education: A systematic review. *Australasian Journal of Educational Technology, 39*(2), 98–114. https://doi.org/10.14742/ajet.8489

McGrew, S. (2021). Skipping the source and checking the contents: An in-depth look at students' approaches to web evaluation. *Computers in the Schools, 38*(2), 75–97. https://doi.org/10.1080/07380569.2021.1912541

Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*(2), 96–107. https://doi.org/10.1598/JAAL.52.2.1

Mulyana, A., & Kurniawati, Y. (2020). Phenomenography analysis of students' historical thinking ability in studying social history. *Jurnal Cakrawala Pendidikan, 39*(3), 666–678. https://doi.org/10.21831/cp.v39i3.28982

Nokes, J. D. (2022). *Building students' historical literacies: Learning to read and reason with historical texts and evidence*. Routledge. https://doi.org/10.4324/9781003183495

Othman, F. A., & Al-Hileh, M. M. (2022). The effect of blended learning in developing critical thinking skills among 8th grade students in history subject. *Journal of Positive School Psychology, 6*(2), 2409–2428.

Purwanta, H. (2023). Indonesian secondary history literacy skills level: Case study of public and private high schools in Sukoharjo Regency, Central Java. *Multidisciplinary Science Journal, 5*, Article e2023050. https://doi.org/10.31893/multiscience.2023050

Rahman, A., Kurniawati, K., & Winarsih, M. (2021). Penerapan literasi sejarah dalam pembelajaran sejarah pada masa pembelajaran jarak jauh di SMA. *Jurnal Pendidikan Sejarah, 10*(1), 57–76. https://doi.org/10.21009/JPS.101.04

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & Education, 144*, Article 103701. https://doi.org/10.1016/j.compedu.2019.103701

Reynolds, J. J. (2022). Improving the assessment of teaching effectiveness with the nonequivalent dependent variables approach. *Teaching of Psychology, 49*(4), 381–387. https://doi.org/10.1177/00986283211022390

Ridha, M., Rahman, F., & Islamy, M. I. (2024). Students learning engagement in the flipped classroom: Systematic literature review. *Jurnal Inovasi Teknologi Pendidikan, 11*(3), 267–284. https://doi.org/10.21831/jitp.v11i3.70562

Santiago, M., & Dozono, T. (2022). History is critical: Addressing the false dichotomy between historical inquiry and criticality. *Theory & Research in Social Education, 50*(2), 173–195. https://doi.org/10.1080/00933104.2022.2048426

Seixas, P., & Morton, T. (2013). *The big six historical thinking concepts*. Nelson Education.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32*(1), 7–18. https://doi.org/10.1097/TLD.0b013e318244557a

Staus, N. L., O'Connell, K., & Storksdieck, M. (2021). Addressing the ceiling effect when assessing STEM out-of-school time experiences. *Frontiers in Education, 6*, Article 690431. https://doi.org/10.3389/feduc.2021.690431

Steiss, J., Krishnan, J., & Wang, J. (2024). Designing writing prompts to elicit students' historical thinking. *The Social Studies, 115*(6), 323–339. https://doi.org/10.1080/00377996.2024.2324926

Sundayana, R. (2014). *Statistika penelitian pendidikan*. Alfabeta.

Tanjung, S., Ane, L., & Sumantri, P. (2024). ICARE-based blended learning model: Improving historical literacy in Indonesian history education courses. In *Proceedings of the 5th International Conference on Innovation in Education, Science, and Culture, ICIESC 2023*. EAI. https://doi.org/10.4108/eai.24-10-2023.2342076

Taylor, T., & Young, C. (2003). *Making history: A guide for the teaching and learning of history in Australian schools*. Curriculum Corporation.

ter Beek, M., Opdenakker, M.-C., Deunk, M. I., Strijbos, J.-W., & Huijgen, T. (2022). Relationships between adolescent students' reading skills, historical content knowledge and historical reasoning ability. *History Education Research Journal, 19*(1), 1–16. https://doi.org/10.14324/HERJ.19.1.02

van der Eem, M., van Drie, J., Brand-Gruwel, S., & van Boxtel, C. (2023). Students' evaluation of the trustworthiness of historical sources: Procedural knowledge and task value as predictors of student performance. *The Journal of Social Studies Research, 47*(1), 64–76. https://doi.org/10.1016/j.jssr.2022.05.003

Walsh-Moorman, E., Pytash, K. E., & Ausperk, M. (2020). Naming the moves: Using lateral reading to support students' evaluation of digital sources. *Middle School Journal, 51*(5), 29–34. https://doi.org/10.1080/00940771.2020.1814622

Widiadi, A. N., Sheehan, M., & Shep, S. (2022). The potential of web-based historical sources as learning resources to foster students' historical thinking skills. *Paramita: Historical Studies Journal, 32*(1), 138–148. https://doi.org/10.15294/paramita.v32i1.31048

Wilson, K., Dudley, D. A., Dutton, J., Preval-Mann, R., & Paulsen, E. (2023). A systematic review of pedagogical interventions on the learning of historical literacy in schools. *History Education Research Journal, 20*(1), 1–27. Article 9. https://doi.org/10.14324/HERJ.20.1.09

Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Temple University Press.

Wissinger, D. R., & Ciullo, S. (2018). Historical literacy research for students with and at risk for learning disabilities: A systematic review. *Learning Disabilities Research & Practice, 33*(4), 237–249. https://doi.org/10.1111/ldrp.12182

Yulifar, L., & Aman, A. (2023). Resources of history learning in conventional and modern continuum lines. *Jurnal Cakrawala Pendidikan, 42*(3), 586–600. https://doi.org/10.21831/cp.v42i3.63713