Journal of Mathematics Instruction, Social Research and Opinion
Vol. 5, No. 2, June 2026, pp. 1865 – 1878, https://doi.org/10.58421/misro.v5i2.1645
ISSN 2962-7842

Evaluating a Cloud-Based History Learning Model: A CIPP Study in Semarang Senior High Schools

Syaiful Amin[1], Imam Shofwan[2], Salman Rashid[3]
[1]Department of History Education, Faculty of Social Sciences and Political Science, Universitas Negeri Semarang, Indonesia
[2]Department of Nonformal Education, Faculty of Education and Psychology, Universitas Negeri Semarang, Indonesia
[3]Department of Education, University of Turbat, Turbat, Pakistan

Article Info

Article history:

Received 2026-05-11
Revised 2026-06-17
Accepted 2026-06-18

Keywords:

Cloud-Based Learning
CIPP Evaluation Model
Historical Thinking Skills
Secondary Analysis
Image Cloud Recognition
Public History

ABSTRACT

History education in the digital era cannot be confined to mere fact-reproduction and chronology; it must cultivate students' historical thinking skills through source analysis, contextualization, and evidence-based reasoning. Although prior research and development (R&D) established that Public History–Based Image Cloud Recognition (ICR) media improved students' historical thinking, that study did not appraise the model through a systematic program-evaluation framework, leaving its contextual fit, design quality, implementation process, and outcomes formally unevaluated. This study, therefore, conducts a secondary, evaluative reanalysis of that implementation using the Context–Input–Process–Product (CIPP) framework; no new primary data were collected. The reanalysis drew on the original survey, observation, interview, and pretest–post-test data gathered from two senior high schools in Semarang City, Indonesia (SMAN 6 and SMAN 12), each comprising one experimental and one control class. Quantitative differences in post-test historical thinking scores were examined with independent-sample t-tests and complemented by effect-size estimation (Cohen's d); qualitative records were thematically analysed. The findings show that teachers used varied media but in pedagogically passive ways, while students' baseline historical thinking was low. The cloud-based ICR media, anchored in local public-history content, produced statistically significant improvements in historical thinking, with $t = 8.778$ (SMAN 6) and $t = 10.239$ (SMAN 12), both exceeding the critical value (1.672), $p < .001$. The CIPP appraisal indicates strong contextual fit and systematic input design, while identifying onboarding, collaboration scaffolding, and data governance as conditions that must be met for adoption to be sustainable. As a secondary CIPP-based evaluation, the study is limited by its dependence on the original dataset and therefore reports conditional rather than conclusive evidence. The findings contribute an evaluative reframing that links contextual needs, programmatic inputs, implementation processes, and learning outcomes for cloud-based history learning.

Corresponding Author:

Syaiful Amin
Department of History, Faculty of Social Sciences and Political Science, Universitas Negeri Semarang
Email: syaifulamin@mail.unnes.ac.id

1. INTRODUCTION

History learning in senior high schools is no longer adequate when it is oriented merely toward the reproduction of facts and chronology. International scholarship on history education emphasizes that the core goal of school history is to develop *historical thinking skills*—the capacity to reason about the past through the use of evidence, contextual reading, cause–and–effect analysis, change-and-continuity, and consideration of ethical and perspectival dimensions [1], [2]. In classroom practice, however, recurrent challenges include the dominance of teacher-led lectures, the limited use of source-based learning, and the weakness of inquiry activities that require evidence-based argumentation. Cognitive studies of historical reasoning have long documented stark differences between novice and expert readers, particularly in the strategies of *sourcing, contextualizing, and corroborating* when evaluating historical evidence [3]. Facts, therefore, remain important, but as raw materials to be interpreted, tested, and debated rather than as the end-goal of learning.

The pedagogical challenge often originates in instructional design. The dominance of lectures, the reliance on textbooks as the sole authority, and the scarcity of source-based learning leave students with few opportunities to construct claims and justifications. Document-based interventions that systematically organize student work around source collections, guiding questions, and argumentative tasks have been shown to improve the quality of historical reasoning and document comprehension [4]. The issue, then, is not simply a lack of methodological variation but the absence of core activities that require students to engage in the disciplinary practices of selecting evidence, weighing perspectives, testing the consistency of sources, and constructing coherent explanations. If the curricular goal is historical thinking, the measure of success must shift from "being able to name an event" to "being able to explain why and how events are understood, based on verifiable evidence."

Within this context, the present study evaluates a specific cloud-based intervention: *Public History–Based Image Cloud Recognition* (ICR) media. ICR is a cloud-hosted learning medium in which curated historical images of local public-history sites and artefacts are stored in an online repository, tagged with descriptive metadata, and paired with guiding questions and structured discussion spaces. Operationally, students access the repository across devices, "read" each image for contextual cues (date, location, actors, and provenance), compare images across sources, and collaboratively construct evidence-based interpretations. ICR thus functions less as a content-delivery channel than as a scaffold for source-based historical reasoning. Defining ICR explicitly at the outset is necessary because its pedagogical logic—not its digital format—is what the evaluation in this study ultimately appraises.

In history education specifically, the use of digital platforms cannot be treated merely as a transfer of materials from face-to-face to online delivery [5]. History learning requires a process of historical thinking that involves source analysis, evidence-based argumentation, and interpretive dialogue. Indonesian studies of e-learning in non-formal and equivalency education programs likewise report that digital platforms do not

automatically improve learning interest or outcomes; their effectiveness depends on the activity design, the readiness of learners and tutors, and the quality of interaction afforded by the platform [6], [7], [8]. Studies of history learning that have utilized cloud-based platforms (e.g., Google Classroom) report practical benefits in accessibility, service integration, and task management, while also consistently noting constraints related to internet networks, device diversity, and the quality of interaction and content comprehension when communication is predominantly nonverbal or asynchronous [9], [10].

At the same time, the rapid development of digital ecosystems—particularly cloud computing services—has created new pedagogical opportunities. *Cloud learning* is the use of cloud services (storage, collaborative applications, learning management systems, and integrated digital resources) to provide cross-device access to learning resources, real-time collaboration, and adaptive multimedia presentation. State-of-the-art reviews indicate that cloud computing in education offers advantages in accessibility, scalability, and collaboration, but also raises issues of infrastructure readiness, data governance, and security [11]. The adoption of cloud-based tools for collaborative learning is shaped by activity design, the quality of participation, and users' readiness to exploit synchronous and asynchronous features [12], [13]. Indonesian research on community-based and equivalency learning has reported similar conditions: distance and virtual learning expand reach but require deliberate pedagogical mediation to translate access into meaningful learning [14], [15], [16]. Concerning risk, behavioural studies of cloud service use confirm that perceptions of security and privacy are important determinants of acceptance and sustained use [17].

Prior research from Amin [18] provides a strong empirical foothold for the Semarang context and underscores the need for evaluative inquiry. That study found that history learning media in Semarang City senior high schools were, in fact, varied (textbooks, videos, images, digital technology). However, their implementation remained suboptimal because they were often used as passive tools rather than as interactive means for developing critical and historical thinking. The same study tested the ICR medium and reported a highly significant impact on students' historical thinking skills compared to conventional media, with t-values exceeding the critical t-table value and significance below the .05 threshold in two schools.

Despite these promising results, a clear gap remains: the prior study demonstrated *that* the media worked, but it did not appraise *why, under what conditions, and how sustainably* it worked, because it was not analysed within a systematic program-evaluation framework. The Context–Input–Process–Product (CIPP) framework [19] has been widely applied to evaluate educational programs in Indonesia—including non-formal training, literacy, and skills programs—and provides a useful structure for connecting needs, design, implementation, and outcomes [20], [21]. Accordingly, the novelty of the present study does not lie in developing the media; the media already exist. Its novelty lies in the *CIPP-based evaluative reinterpretation* of the prior implementation, which repositions the earlier history-education findings within a structured evaluation that links contextual needs, programmatic inputs, implementation processes, and measurable outcomes.

This study therefore aims to: (i) identify the alignment of learning objectives with students' needs (*Context*); (ii) analysed the readiness of available resources and instructional design (*Input*); (iii) assess the quality of classroom implementation and learning dynamics (*Process*); and (iv) appraise the achievement of learning outcomes, particularly improvements in students' historical thinking skills (*Product*). The study contributes theoretically by demonstrating how a discipline-specific learning innovation can be appraised through a general program-evaluation lens and contributes practically by surfacing the implementation conditions—onboarding, facilitation, and governance—on which sustainable adoption depends.

2. METHOD

2.1 Research Design

This study employed an evaluative research design framed by the Context–Input–Process–Product (CIPP) model [19], operationalized as a secondary analysis (evaluative reanalysis) of an existing implementation dataset. The CIPP model has been used in the Indonesian educational context to evaluate diverse programs ranging from training and equivalency education to skills-based curricula [20], [21]. The design is explicitly non-experimental at the level of the present study: it does not generate new primary data but re-appraises documented evidence from a completed R&D study through a structured evaluation lens.

2.2 Source Study, Location, and Time

The dataset reanalysed here originates from Amin's [18] R&D study, which developed and tested the ICR media using ADDIE procedures in two public senior high schools in Semarang City, Central Java, Indonesia—SMAN 6 Semarang and SMAN 12 Semarang. The present secondary analysis was conducted after the completion of that study, drawing on its archived instruments, records, and statistical outputs. The present authors conducted no fieldwork.

2.3 Participants and Class Distribution

The original study involved history teachers and Grade X students across the two schools, with each school contributing one experimental class (using the ICR media) and one control class (using conventional media). The precise enrolment in each class, together with the number of participating teachers, is reported in Table 1 and should be transcribed from the original dataset prior to submission. Where the present article reports sample sizes (n) for each group, these values are taken directly from the source study and are not re-estimated.

2.4 Sampling Technique

Sample selection followed the procedure used in the source study [18], in which intact classes were assigned to experimental and control conditions (a quasi-experimental, purposive class-based assignment rather than individual random allocation). Readers

should interpret group comparisons in light of this intact-class design, which is common in school-based research but limits the possibility of strict causal inference.

2.5 Data Sources and Collection

The reanalysis used four categories of secondary data, mapped onto the CIPP components. Context: survey and interview data on teachers' use of history learning media, and diagnostic assessment of students' baseline historical thinking. Input: model design and development documents (R&D with ADDIE procedures), public-history content specifications, instructional tools, and device/access readiness records. Process: classroom observation notes documenting lesson flow, technical difficulties, and discussion dynamics across experimental and control classes. Product: pretest–post-test scores on historical thinking, assumption-testing outputs (normality and homogeneity), and independent-sample t-test results. All materials were obtained from the source study as archived documentation; no instruments were re-administered.

2.6 Instruments and Measurement

The measured construct is historical thinking, operationalized through indicators of chronological reasoning, historical comprehension, source interpretation, and inquiry capability, and assessed by the test instrument developed and validated in the source study [18]. Because the instrument and its validity/reliability evidence were established in the original R&D study, the present analysis treats them as given and reports them descriptively rather than re-validating them.

2.7 Trustworthiness and Coding Procedure

Qualitative records were re-coded using reflexive thematic analysis following six steps: (1) familiarization with the observation and interview records; (2) generation of initial codes; (3) searching for themes; (4) reviewing themes against the data extracts; (5) defining and naming themes; and (6) interpretation in relation to the CIPP components. To strengthen dependability, coding was conducted by two coders independently for an initial subset of the records, after which coding discrepancies were resolved through discussion, and a shared codebook was agreed upon; inter-coder agreement should be reported (e.g., percentage agreement or Cohen's κ) in the final manuscript once recomputed from the coded subset. Credibility was supported by triangulating survey, interview, and observation evidence; confirmability was supported by maintaining an audit trail linking each theme to its source extracts.

2.8 Data Analysis

Quantitative analysis proceeded in three steps. First, descriptive statistics (mean, standard deviation, and n) were compiled for the experimental and control groups in each school. Second, the assumptions of the independent-sample t-test—normality (tested separately for each school and each group) and homogeneity of variance—were verified using the outputs reported in the source study [18]. Third, an independent-sample t-test compared post-test historical thinking scores between the experimental and control groups in each school, and the practical magnitude of the difference was estimated using Cohen's d. Because only post-test comparisons between independent groups were available, the

analysis compares post-test scores rather than within-subject gain scores; this is stated explicitly to avoid overinterpretation. Qualitative analysis used the thematic procedure described in Section 2.7 to appraise process quality, identify implementation barriers, and surface cross-school success factors. The two strands were integrated within the CIPP framework to yield a coherent evaluative judgment. Figure 1 visualizes the operationalization of the four CIPP components and the three-layer interpretive framework that organizes the discussion.

2.9 Ethical Considerations

Because the study reused an existing dataset, no new participant contact was required. Permission to access and reanalyse the archived data was obtained from the original researcher/data custodian, and the data were used only in de-identified form, with no personally identifying student information reported. The authors affirm that the secondary use of these data is consistent with the consent and confidentiality conditions of the original study; the corresponding ethical clearance reference from the source study should be cited in the final manuscript.

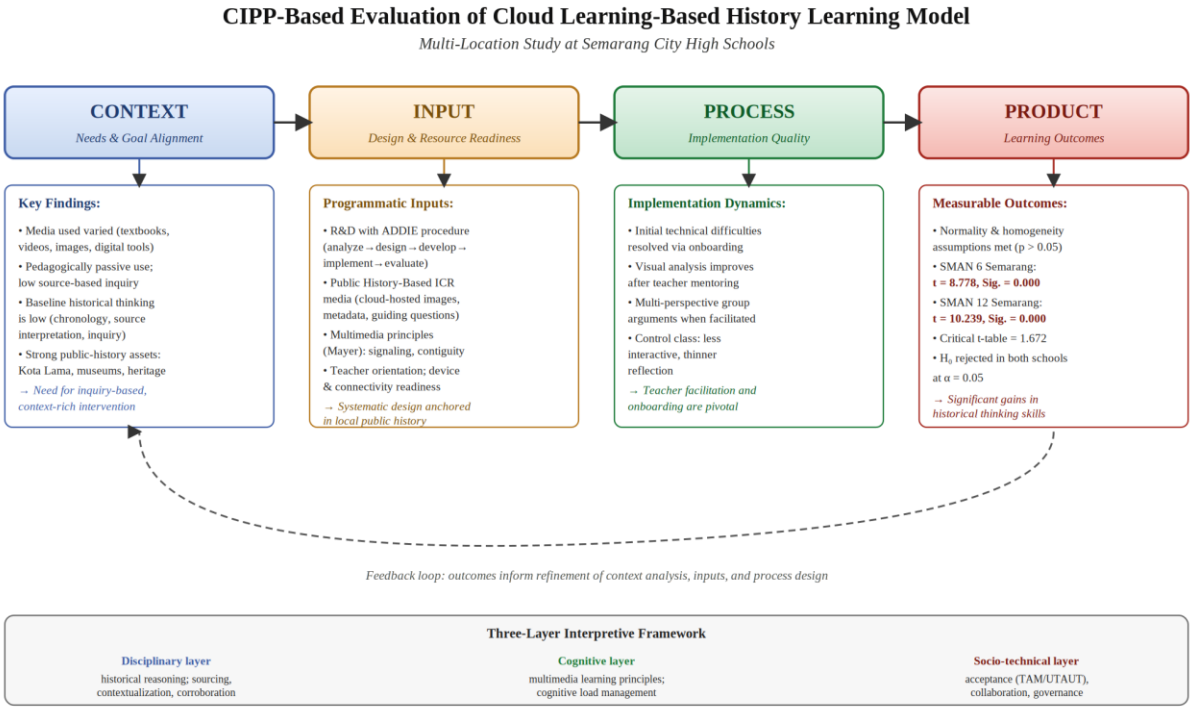

Figure 1. A CIPP-based evaluation framework applied to the cloud learning-based history learning model in Semarang City senior high schools, with the three-layer interpretive framework used in the discussion. [Insert figure file here.]

3. RESULTS AND DISCUSSION

3.1 Results

This section presents the findings of the CIPP-based reanalysis, organized according to the four components illustrated in Figure 1. It first summarizes the sample characteristics and baseline conditions (Context), then the design and readiness of the

intervention (Input), the implementation dynamics (Process), and finally the inferential and effect-size results on gains in historical thinking (Product).

3.1.1 Context: Gap between media variety and pedagogical use

Survey and interview data indicated that history teachers in senior high schools in Semarang City used a wide range of learning media, including printed textbooks, videos, still images, presentation slides, and several digital applications. However, observation notes and teacher self-reports converged on the same picture: in the majority of lessons, these media were deployed as one-way presentational tools. Source-based tasks, structured argumentation, and inquiry-oriented work were rare. The dominant activity pattern was teacher-narrated chronology supported by visual illustration rather than student-led interpretation of evidence.

Diagnostic mapping of students' historical thinking skills revealed relatively low performance on the core indicators: chronological reasoning, historical comprehension, source interpretation, and inquiry capability. Students were generally able to recall events and dates, but they struggled when asked to identify a source's perspective, weigh competing accounts, or formulate evidence-based explanations. This profile established a clear local need for an intervention that re-centres history learning on disciplinary practice rather than recall.

3.1.2 Input: Model design and readiness

The cloud learning model under evaluation, the ICR media [18], was developed through an R&D approach guided by the ADDIE procedure. The analysis phase produced the contextual gap and the historical-thinking diagnosis summarized above. The design phase generated learning objectives aligned with historical-thinking indicators, a content map drawn from Semarang's public-history sources (Kota Lama, museums, heritage buildings, oral accounts), and a sequence of activities anchored in source interpretation. The development phase produced the digital media itself: a cloud-hosted image repository with metadata, guiding questions, and structured discussion spaces. Similar ADDIE-based development trajectories have been documented in Indonesian studies of multimedia learning resources, where systematic design steps improve the alignment between content, activities, and assessment [22], [23].

The readiness assessment showed that participating schools had adequate connectivity and that most students could access the platform on their personal devices. Teachers received a brief orientation to the platform and the inquiry-based task structure prior to implementation. No major institutional barriers to access were reported, although individual variation in device quality and home internet was noted—echoing the access–quality tension repeatedly observed in Indonesian digital learning research [10], [14], [24].

3.1.3 Process: Implementation dynamics

Observation notes from the experimental classes documented two distinct phases. In the early meetings, students experienced technical difficulties with logging in, navigating the image repository, and uploading written responses. These difficulties produced a brief disruption but did not derail the lessons. By the third meeting, after

teacher mentoring and short demonstrations, students' visual analysis activity became more sustained: they began to read images for context (date, location, actors), to compare images across sources, and to connect visual elements with the textual material provided.

Group discussions, used to consolidate interpretations, produced mixed but informative patterns. Many groups developed multi-perspective arguments and explicitly cited image features as evidence. However, some students remained focused on surface visual description without contextual depth, and a smaller subset participated minimally in discussion threads. Teacher facilitation—reframing questions, prompting source comparison, and assigning roles—was associated with the higher-quality group outputs.

In the control classes, which used conventional history learning media without the ICR platform, observations indicated that discussions were less interactive, student participation was more limited, and reflection on historical context was less in-depth than in the experimental classes. Some critical thinking still occurred, but it was rarely organized around explicit source-based argumentation.

3.1.4 Product: Effectiveness on historical thinking

Assumption testing was conducted separately for each school and each group. Normality testing yielded p-values above .05, and homogeneity-of-variance testing also yielded p-values above .05, indicating that the pretest scores were normally distributed and that the group variances were equal. The independent-sample t-test could therefore be applied to compare post-test historical thinking scores between the experimental and control classes.

The results are summarized in Table 1. In both schools, the observed t-value substantially exceeded the critical t-table value of 1.672, and significance was below .001. The decision in both cases is to reject the null hypothesis: the difference in post-test historical thinking scores between the experimental and control groups is statistically significant in favour of the experimental group. To convey practical magnitude—rather than statistical significance alone—Cohen's d is reported alongside each t-value; for two independent groups, it can be obtained from the reported statistics as $d = t \times \sqrt{(1/n_1 + 1/n_2)}$.

Table 1. Independent-sample t-test results on post-test historical thinking scores.

School	Group	n	Mean	SD	t	p	Cohen's d
SMAN 6 Semarang	Experimental	—	—	—	8.778	< .001	—
	Control	—	—	—	—	—	
SMAN 12 Semarang	Experimental	—	—	—	10.239	< .001	—
	Control	—	—	—	—	—	

Critical t-table value = 1.672 (α = .05, one-tailed). t-values are reported as in [18]; significance originally reported as .000 is presented as p < .001 because exact significance is never literally zero.

Author action required: the cells marked "—" for n, Mean, SD, and Cohen's d must be completed with the descriptive values from the original dataset [18] before submission. Cohen's d can be computed from the reported t and group sizes using $d = t \times \sqrt{(1/n_1 + 1/n_2)}$.

Read together, the four CIPP components produce a coherent empirical pattern: a clear need (Context), a systematically designed intervention (Input), an implementation that improved with onboarding and teacher facilitation (Process), and a statistically significant post-test advantage for the experimental group in both schools (Product). The strength of this evidence, however, is conditioned by the completeness of the descriptive statistics noted above; until effect sizes are reported, the magnitude of the advantage should be described as statistically significant rather than necessarily large.

3.2 Discussion

The findings can be interpreted at the three integrated layers shown at the bottom of Figure 1: a disciplinary layer rooted in historical thinking, a cognitive layer informed by multimedia learning principles, and a socio-technical layer that addresses acceptance, collaboration, and governance. Each layer offers a partial but complementary explanation of why the cloud-based model produced significant gains, and each is tied explicitly below to the corresponding CIPP component—its strength, its weakness, and the recommendation that follows.

3.2.1 Disciplinary layer (Context → Product): From media to historical reasoning

The Context findings echo a paradox well documented in the technology-adoption literature: media variety alone does not transform pedagogy [1], [25]. The fact that students in the diagnostic phase could recall events but struggled to weigh competing accounts is consistent with the long-standing distinction between novice and expert historical reading [3]. The intervention addressed this gap because it did not merely deliver content through a different channel; it restructured tasks around the disciplinary practices of sourcing, contextualization, and corroboration [1], [3]. By embedding image sources within a public-history context that is geographically and culturally proximate to learners, the model strengthens the bridge between "history in textbooks" and "history in students' living spaces" [18], a move consistent with progression models that treat historical understanding as a developmental achievement [26] and with broader pedagogies aimed at causal and interpretive reasoning [27]. In CIPP terms, the Context strength is a precisely diagnosed need; the corresponding recommendation is to keep task design, not platform features, at the centre of any scale-up.

Within the Indonesian context, similar trajectories are evident in studies of historical inquiry that scaffold students from source analysis to argumentation and reflective thinking [28], and in research on production-based and inquiry-oriented learning models that produce measurable gains in higher-order outcomes [29], [30]. The convergence of local and international evidence suggests that the Product gains are not artifacts of platform novelty; they are the predictable consequence of redesigning the learning task itself.

3.2.2 Cognitive layer (Process → Product): Multimedia integration and load management

The Process observation that students initially "got stuck" on surface visual features but later moved to contextual interpretation is best understood through multimedia

learning theory. Effective multimedia is not the mere addition of pictures to text; it is the design of materials in accordance with the cognitive principles of selection, organization, and integration [31]. The ICR model embodies several of these principles: words and images appear in complementary roles, guiding questions act as signalling cues that direct attention to historically relevant features, and the spatial and temporal contiguity of image-and-prompt pairs reduces extraneous load. Recent Indonesian work on digital learning media for community-based education similarly emphasizes that effectiveness depends on the deliberate alignment of media design with collaborative, literacy-oriented activities, not on the digital format alone [22], [32].

This cognitive interpretation also explains the documented learning curve. Initial technical difficulties produce an extraneous load that competes with the resources needed for source interpretation; once onboarding routines are in place, that load is reduced, and learners can devote attention to the historically meaningful task [33]. In CIPP terms, the Process weakness is the early-meeting friction, and the recommendation is to treat onboarding not as peripheral support but as a core Input that determines whether Process activities yield Product gains.

The two schools produced different effect magnitudes, with a higher t-value at SMAN 12 (10.239) than at SMAN 6 (8.778). Because both used the same media and task structure, the most plausible explanations lie in Process and Input conditions rather than in the medium itself: differences in teacher facilitation intensity, in the smoothness of onboarding, in baseline score dispersion, or in class size, can all raise or lower the t-statistic. The available documentation does not allow these explanations to be adjudicated definitively; the comparison is therefore offered as a hypothesis to be tested once the full descriptive statistics for each class are recovered from the original dataset.

3.2.3 Socio-technical layer (Input/Process → sustainability): Acceptance, collaboration, and governance

The third layer concerns the socio-technical conditions of cloud-based learning. Acceptance models such as TAM [34] and UTAUT [35] highlight that perceptions of usefulness, ease of use, social influence, and facilitating conditions shape sustained use. The orientation given to teachers and the brief demonstrations provided to students plausibly shifted these perceptions in a favourable direction, consistent with the observed transition from initial difficulty to active visual analysis. Indonesian studies of e-learning in equivalency and virtual programs similarly report that perceived usefulness and tutor mediation are decisive in translating access into engagement [6], [16].

Collaboration quality, however, was uneven: some groups produced multi-perspective arguments while others remained descriptive. This matches the cloud-collaboration literature—collaborative platforms do not automatically yield high-quality collaboration; clear role structures, contribution rubrics, and teacher facilitation are required [12], [13]. Indonesian research on participatory and adaptive learning in non-formal settings likewise points to facilitation, formative assessment, and role design as central conditions for the quality of participation [29], [36]. The contrast with the control classes—where discussions were less interactive and reflection thinner—reinforces the

argument that cloud learning works when it transforms tasks into inquiry, not when it merely migrates materials into a digital format.

Governance is a distinct sustainability condition. Cloud-education research consistently identifies infrastructure readiness, data governance, and security as prerequisites for durable adoption [11], [17]. For the ICR platform specifically, this is not abstract: the system collects and stores student accounts and login credentials, uploaded written responses and discussion contributions, and behavioural traces (access logs, time-on-task). These data would typically be managed by the school together with the platform provider, which raises concrete risks—unauthorized access to minors' records, unclear data-retention periods, and ambiguous ownership of student-generated interpretations. Sustainable institutionalization, therefore, requires explicit policies on account management, retention and deletion schedules, consent for behavioural logging, and a named data custodian. Comparable governance considerations recur in Indonesian evaluations of community learning programs, where institutional management determines whether well-designed programs persist [37], [38].

3.2.4 CIPP synthesis, implications, and limitations

Synthesizing across components, the evaluative judgment is positive but conditional. The model is contextually justified (it addresses a documented need), input-strong (ADDIE-based design grounded in local public history), process-sensitive (outcomes depend on facilitation and onboarding), and product-effective (significant post-test gains in both schools). For practice, professional development for history teachers should foreground multimedia design and inquiry facilitation rather than platform skills alone, and school leaders should treat onboarding and governance as first-class implementation tasks.

Several limitations qualify these conclusions. First, the evaluation is a secondary analysis of an existing dataset [18]; it introduces no new observations and inherits the original sampling decisions. Second, and consequently, the analysis depends heavily on the quality and completeness of the original study's documentation—where descriptive statistics or procedural details are missing, the present appraisal can only flag rather than resolve them. Third, effect sizes and retention of gains over time were not estimated here and should be examined in future work. Fourth, the equity of impact across student subgroups (gender, prior achievement, device quality) warrants targeted analysis. Fifth, transferability to other regional contexts, each with its own public-history repertoire, requires comparative case research. These limitations do not overturn the central judgment; they delineate the agenda for the next phase of design and inquiry.

4. CONCLUSION

This CIPP-based evaluation set out to appraise a cloud-based history learning model not by re-testing it, but by judging its contextual fit, design quality, implementation, and outcomes within a single evaluative frame. The appraisal indicates that the model meets a genuine local need, is systematically designed and grounded in local public history, depends on teacher facilitation and onboarding during implementation, and is associated with statistically significant gains in students' historical thinking in both

participating schools. The central conclusion is that the model is promising but conditional: its effectiveness depends on teacher facilitation, technical readiness, student onboarding, and governance support, rather than on the cloud platform as such. As the reviewers' framing makes explicit, the cloud-based ICR model can support historical thinking when implemented not merely as digital media delivery but as a scaffold for inquiry, source interpretation, and collaborative, evidence-based reasoning. Because this is a secondary evaluation, its claims are bounded by the original dataset. Future research should be prioritized accordingly: (1) report effect sizes to quantify practical magnitude; (2) conduct longitudinal studies of retention; (3) examine subgroup equity; and (4) compare implementation across other regions and across formal and non-formal education settings.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the participating institutions and the teachers and students at SMAN 6 and SMAN 12 Semarang and thank the original researcher for granting access to the dataset reanalysed in this study. Any errors that remain are the responsibility of the authors.

REFERENCES

[1] J. van Drie and C. van Boxtel, "Historical reasoning: Towards a framework for analyzing students' reasoning about the past," *Educational Psychology Review*, vol. 20, no. 2, pp. 87–110, 2008, doi: 10.1007/s10648-007-9056-1.

[2] N. Popa, "How meaning making cultivates historical consciousness: Identifying a learning trajectory and pedagogical guidelines to promote it," *The Social Studies*, vol. 114, no. 4, pp. 139–159, 2023, doi: 10.1080/00377996.2022.2140641.

[3] S. Wineburg, "Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence," *Journal of Educational Psychology*, vol. 83, no. 1, pp. 73–87, 1991, doi: 10.1037/0022-0663.83.1.73.

[4] A. Reisman, "Reading like a historian: A document-based history curriculum intervention in urban high schools," *Cognition and Instruction*, vol. 30, no. 1, pp. 86–112, 2012, doi: 10.1080/07370008.2011.634081.

[5] N. Putri, E. T. Lestari, and T. Agustian, "Analysis of the use of the Google Classroom application in class X history learning at SMK Mandiri Pontianak," *Historica Didactica: Journal of Historical, Cultural and Social Education*, vol. 2, no. 3, pp. 58–70, 2022.

[6] I. Shofwan, S. Aminatun, E. Handoyo, and M. T. Kariadi, "The effect of e-learning on students' learning interest in the equivalence education program," *Journal of Non-formal Education*, vol. 7, no. 1, pp. 103–111, 2021, doi: 10.15294/jne.v7i1.26790.

[7] Y. Siswanto, J. Sutarto, and S. E. Mulyono, "E-learning in non-formal education programs," *Journal of Non-formal Education*, vol. 4, no. 2, pp. 168–179, 2018, doi: 10.15294/jne.v4i2.16005.

[8] T. Suminar, M. Arbarini, I. Shofwan, and N. Setyawan, "The effectiveness of production-based learning models in the ICARE approach to entrepreneurial literacy ability," *Journal of Non-formal Education*, vol. 7, no. 2, pp. 142–149, 2021, doi: 10.15294/jne.v7i2.31700.

[9] R. Rasmi, S. Bahri, and P. Sukino, "Analysis of history learning through learning media (Google Classroom) in Grade X students of State High School 1 Tebas, Sambas Regency," *Didactic Historica: Journal of Historical, Cultural and Social Education*, vol. 4, no. 1, pp. 140–159, 2024.

[10] S. Suwarni and S. Superman, "History learning using blended learning," *SOCIA: Journal of Social Sciences*, vol. 18, no. 2, pp. 153–162, 2021.

[11] J. A. González-Martínez, M. L. Bote-Lorenzo, E. Gómez-Sánchez, and R. Cano-Parra, "Cloud computing and education: A state-of-the-art survey," *Computers & Education*, vol. 80, pp. 132–151, 2015, doi: 10.1016/j.compedu.2014.08.017.

[12] H. Al-Samarraie and N. Saeed, "A systematic review of cloud computing tools for collaborative learning: Opportunities and challenges to the blended-learning environment," *Computers & Education*, vol. 124, pp. 77–91, 2018, doi: 10.1016/j.compedu.2018.05.016.

[13] N. S. Baanqud, H. Al-Samarraie, A. I. Alzahrani, and O. Alfarraj, "Engagement in cloud-supported collaborative learning and student knowledge construction: A modeling study," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, art. no. 56, 2020, doi: 10.1186/s41239-020-00232-z.

[14] M. Arbarini, A. Rahmat, I. Ismaniar, A. Isa, and Y. Siswanto, "Equivalency education: Distance learning and its impact in Indonesia," *Journal of Non-formal Education*, vol. 8, no. 1, pp. 12–22, 2022, doi: 10.15294/jne.v8i1.33932.

[15] I. Prasetyo, Y. Suryono, and S. Gupta, "The 21st century life skills-based education implementation at the non-formal education institution," *Journal of Non-formal Education*, vol. 7, no. 1, pp. 1–7, 2021, doi: 10.15294/jne.v7i1.26385.

[16] W. Yulianingsih, M. Roesminingsih, S. Mardliyah, M. Lutviatiani, and K. Rofiah, "Management of English courses in virtual learning during the COVID-19 pandemic," *Journal of Non-formal Education*, vol. 9, no. 1, pp. 23–30, 2023, doi: 10.15294/jne.v9i1.41737.

[17] I. Arpaci, K. Kilicer, and S. Bardakci, "Effects of security and privacy concerns on educational use of cloud services," *Computers in Human Behavior*, vol. 45, pp. 93–98, 2015, doi: 10.1016/j.chb.2014.11.075.

[18] S. Amin, "Development of public history-based Image Cloud Recognition (ICR) media to improve the historical thinking skills of high school students in Semarang City," Ph.D. dissertation, Universitas Sebelas Maret, Surakarta, Indonesia, 2024.

[19] D. L. Stufflebeam, "The CIPP model for evaluation," in *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, D. L. Stufflebeam, G. F. Madaus, and T. Kellaghan, Eds., Dordrecht: Springer Netherlands, 2000, pp. 279–317, doi: 10.1007/0-306-47559-6_16.

[20] D. Y. Akhadi and I. Shofwan, "Evaluation of the CIPP model of the sewing training program," *Journal of Non-formal Education*, vol. 10, no. 1, pp. 33–48, 2024, doi: 10.15294/jone.v10i1.1723.

[21] A. Heryahya, B. Sujanto, and R. Rugayah, "Implementation of CIPP evaluation on the Keaksaraan Usaha Mandiri program," *Journal of Non-formal Education*, vol. 6, no. 1, pp. 9–18, 2020, doi: 10.15294/jne.v6i1.22036.

[22] K. D. Dhiu and V. O. Mere, "Development of pop-up book media social aspects in developing early children's cooperation capability," *Journal of Non-formal Education*, vol. 11, no. 2, pp. 204–215, 2025, doi: 10.15294/jone.v11i2.29584.

[23] R. A. H. Cahyadi, "Development of teaching materials based on the ADDIE model," *Halaqa: Islamic Education Journal*, vol. 3, no. 1, pp. 35–42, 2019, doi: 10.21070/halaqa.v3i1.2124.

[24] N. Hazizah and Yaswinda, "Teachers' strategies in preparing online learning digital media for developing children's literacy skills," *Journal of Non-formal Education*, vol. 7, no. 2, pp. 187–194, 2021, doi: 10.15294/jne.v7i2.26203.

[25] P. Seixas, "A model of historical thinking," *Educational Philosophy and Theory*, vol. 49, no. 6, pp. 593–605, 2017, doi: 10.1080/00131857.2015.1101363.

[26] P. Lee and D. Shemilt, "A scaffold, not a cage: Progression and progression models in history," *Teaching history*, no. 113, pp. 13–23, 2003.

[27] G. Stoel, J. van Drie, and C. van Boxtel, "Teaching towards historical expertise: Developing a pedagogy for fostering causal reasoning in history," *Journal of Curriculum Studies*, vol. 47, no. 1, pp. 49–76, 2015, doi: 10.1080/00220272.2014.968212.

[28] L. Armiyati and H. Purwanta, "Strengthening historical literacy through historical inquiry: Learning from Singapore," *Essays: Journal of Social Science and Humanities*, vol. 5, no. 1, pp. 664–682, 2024.

[29] M. Arbarini, S. Jutmini, S. Joyoatmojo, and Sutarno, "Effect of participatory learning model on functional literacy education," *Journal of Non-formal Education*, vol. 4, no. 1, pp. 87–96, 2018, doi: 10.15294/jne.v4i1.13570.

[30] B. A. VanSledright, *Assessing Historical Thinking and Understanding: Innovative Designs for New Standards*. New York, NY, USA: Routledge, 2014.

[31] R. E. Mayer, "Using multimedia for e-learning," *Journal of Computer Assisted Learning*, vol. 33, no. 5, pp. 403–423, 2017, doi: 10.1111/jcal.12197.

[32] D. Darmawan, P. Hadiyanti, S. Wibowo, R. Syah, and M. F. Abdillah, "Development of digital learning media for community-based education: A collaborative innovation and literacy framework," *Journal of Non-formal Education*, vol. 11, no. 2, pp. 289–300, 2025, doi: 10.15294/jone.v11i2.32060.

[33] S. Liu, "Developing digital literacy in history classrooms: A blended approach to integrating technology and historical thinking," *Journal of Educational Technology & Society*, vol. 24, no. 2, pp. 101–114, 2021.

[34] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989, doi: 10.2307/249008.

[35] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, 2003, doi: 10.2307/30036540.

[36] M. A. Adeoye, E. F. Prastikawati, and R. Riwayatiningsih, "Beyond traditional testing: Exploring the efficacy of adaptive assessments in shaping future learners," *Journal of Non-formal Education*, vol. 11, no. 1, pp. 153–169, 2025, doi: 10.15294/jone.v11i1.19934.

[37] A. Ridwan, E. Sustiyatik, and A. R. Pamungkas, "Analysis of methods and approaches in project management in non-formal education in Indonesia," *Journal of Non-formal Education*, vol. 10, no. 2, pp. 297–308, 2024, doi: 10.15294/jone.v10i2.13427.

[38] M. Melania, A. Kadir, A. R. Pamungkas, and S. Gupta, "Contribution of non-formal education to improve the quality of human resources," *Journal of Non-formal Education*, vol. 10, no. 2, pp. 169–179, 2024, doi: 10.15294/jone.v10i2.2015.

[39] T. C. Hammond, "Doing history with digital tools: A new model," *The History Teacher*, vol. 47, no. 2, pp. 209–236, 2014.

[40] B. Kisworo, Implementation of learning media based on adult education principles at PKBM Indonesia Pusaka Ngaliyan, Semaran