Innovative Educational Technologies and Pedagogies
Vol. 1 No. 1 (2026) pp. 21-36
Online ISSN: XXXX-XXXX

Original Research

From Virtual Heritage Visits to Civic Problem-Solving: Authentic Assessment of Historical HOTS and Civic Reasoning among Indonesian Vocational Students

Affan Budi Santosa[1]✉, Wildan Nurul Fajar[1]
[1] Master Program in Social Studies Education, Universitas Muhammadiyah Purwokerto, Indonesia
✉ affanbudi21@gmail.com
10.65840/ietp.vxix.xx

Article Info
Submitted:
12/22/2025
Revised:
02/20/2026
Accepted:
04/20/2026
Available Online:
04/24/2026

Abstract. *Digital museum environments are increasingly used in history education, yet their assessment value remains under-theorised: many classroom studies treat virtual tours as motivational media rather than as evidence-rich stimuli for assessing historical reasoning. This pilot case-based study redesigns a virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja, Purbalingga, as an authentic assessment ecology for measuring higher-order thinking skills (HOTS) and civic reasoning among 16 Indonesian vocational students undertaking internship placement. The assessment combined virtual artefact exploration, guided observation, a local heritage problem case, an analytic rubric, student written responses, field notes, and a 12-item post-activity perception questionnaire. Secondary descriptive analysis of the available questionnaire data showed a high overall perception index (M = 43.38/48, 90.36% of the maximum score; SD = 2.42; 95% descriptive interval = 42.09-44.66). The strongest dimensions were media use and independent engagement (both M = 3.76/4), while conceptual understanding was lower (M = 3.42/4), producing a media-activity/understanding gap of 0.34 points. Qualitative evidence mapping indicated that students could move from noticing artefacts to framing local cultural heritage as a public problem, evaluating preservation responsibilities, and designing vocationally relevant civic actions such as digital storytelling and community-based heritage campaigns. The study contributes a validity-aware model of virtual museum-based authentic assessment: virtual presence alone does not guarantee deep learning, but it becomes assessable civic-historical performance when mediated by explicit problem framing, evidence prompts, and analytic rubrics.*

Keywords: *authentic assessment; civic reasoning; historical thinking; higher-order thinking skills; local heritage; virtual museum; vocational education*

1. Introduction

History education in vocational schools is often caught between two institutional pressures. On the one hand, the subject is expected to cultivate historical consciousness, national identity, and civic responsibility. On the other hand, vocational students tend to prioritise competencies that they perceive as immediately connected to work placement, certification, and employability. This tension makes conventional history assessment vulnerable to two weaknesses: it may reward recall without interpretation, and it may fail to connect local historical knowledge to public problem-solving. The challenge is therefore not simply to make history lessons more attractive, but to make historical evidence operational within tasks that require analysis, evaluation, and action-oriented judgement.

Virtual museum tours offer a potentially productive response to this challenge because they provide accessible encounters with artefacts, spatial

narratives, and local heritage collections. The International Council of Museums defines museums as institutions that research, conserve, interpret, and exhibit tangible and intangible heritage while offering educational, reflective, and knowledge-sharing experiences (ICOM, 2022). UNESCO similarly frames museums and collections as public resources for cultural access, education, and the safeguarding of diverse heritage (UNESCO, 2015). In Purbalingga, the virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja provides digital access to exhibition spaces such as the lobby and display rooms, enabling students to encounter local cultural materials without the logistical constraints of a physical visit (Museum Soegarda, 2024a, 2024b).

Nevertheless, the mere presence of digital heritage does not automatically produce higher-order cognition. Research on immersive and multimedia learning repeatedly shows that visualisation, interactivity, and presence may support engagement, but learning depends on instructional design, cognitive scaffolding, task alignment, and the management of cognitive load (Mayer, 2020; Makransky & Petersen, 2021; Petersen et al., 2022; Radianti et al., 2020). For history education, this point is crucial: looking at artefacts is not equivalent to thinking historically. Students must be asked to use artefacts as evidence, interpret their significance, compare perspectives, identify cause-and-consequence relations, and deliberate about contemporary civic issues grounded in historical understanding (Seixas & Morton, 2013; Wineburg, 2001).

This article therefore reconceptualises a virtual museum tour as an authentic assessment environment rather than only as an instructional medium. Authentic assessment requires learners to demonstrate knowledge, judgement, and skill in tasks that approximate meaningful real-world practice (Gulikers et al., 2004; Wiggins, 1998). In the present case, the meaningful practice is not professional museum curation alone; it is the civic-historical practice of using local heritage evidence to reason about preservation, youth participation, and community responsibility. This design aligns with the Indonesian curricular emphasis on assessment that supports competency development and meaningful learning processes (Badan Standar, Kurikulum, dan Asesmen Pendidikan, 2025).

The study addresses three research questions. First, how can a virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja be structured as an authentic assessment task for vocational students? Second, what do the available perception data reveal about students' acceptance, engagement, and conceptual understanding after the assessment? Third, what evidence map can be constructed to explain how students' responses move from artefact observation toward HOTS and civic reasoning?

2. Conceptual Framework

2.1 Authentic assessment as validity-oriented task design

Authentic assessment is most useful when it is treated not as a decorative classroom activity but as an argument about the interpretation and use of evidence. Gulikers et al. (2004) identify task, physical context, social context, assessment form, and criteria as dimensions of authenticity. In this study, those dimensions are

translated into a museum-based task: students investigate a digital heritage space, select evidence from artefacts, respond to a contemporary local case, and justify a civic action proposal. The logic is consistent with Kane's (2013) argument-based view of validity: the issue is not whether the task is attractive, but whether the observed student performance can support the intended interpretation about HOTS and civic reasoning.

For this reason, the assessment was designed around performance evidence rather than recall. A student response was considered stronger when it demonstrated four linked moves: accurate selection of museum evidence, causal analysis of a heritage problem, evaluative judgement about preservation responsibilities, and creation of a feasible action proposal. These moves operationalise HOTS in the revised Bloom taxonomy, particularly analysing, evaluating, and creating (Anderson & Krathwohl, 2001; Brookhart, 2010).

2.2 Historical HOTS and civic reasoning

Historical thinking is a disciplined form of reasoning with evidence rather than a chronological accumulation of names and dates. Seixas and Morton's (2013) historical thinking concepts - including historical significance, evidence, cause and consequence, perspective taking, continuity and change, and ethical dimension - help clarify what can be assessed in a local museum task. Civic reasoning extends this logic into public judgement. The National Academy of Education frames civic reasoning and discourse as students' capacity to examine complex civic, political, and social issues using evidence and deliberation (Lee et al., 2021).

In the present study, civic reasoning was not reduced to civic knowledge. It was interpreted as the capacity to ask: What public problem is visible in this heritage case, whose responsibilities are implicated, what evidence supports the judgement, and what action is defensible? This definition is appropriate for vocational students because it invites them to connect civic responsibility with vocational expertise, for example by designing social media campaigns, digital storytelling, local tourism promotion, or community documentation of heritage assets.

2.3 Virtual museums as assessment ecologies

Virtual reality and virtual tours can create presence, interactivity, and learner control, but their pedagogical value depends on how these affordances are integrated into learning tasks. Radianti et al. (2020) show that VR applications need clearer design elements and learning theories, while Hamilton et al. (2021) find that VR can support learning outcomes when aligned with appropriate pedagogical conditions. The Cognitive Affective Model of Immersive Learning further explains that immersion and interactivity influence affective and cognitive processes, yet those processes can either facilitate or hinder learning depending on instructional support (Makransky & Petersen, 2021; Petersen et al., 2022).

This conceptual position leads to a core proposition of the article: a virtual museum tour becomes a high-quality assessment environment only when three layers are deliberately connected - the digital stimulus, the historical-civic problem, and the analytic evidence criteria. Figure 1 summarises this logic.

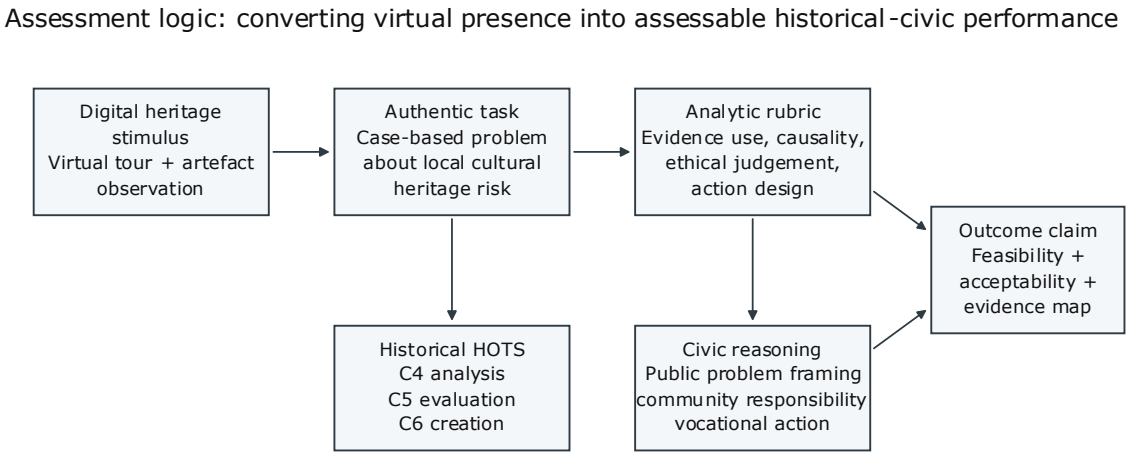

Figure 1. Assessment logic model for converting virtual museum presence into historical-civic performance evidence.

Accordingly, the model does not treat the virtual tour as an instructional innovation in itself, but as an evidentiary architecture for eliciting situated historical reasoning. The student's movement through the museum environment generates perceptual engagement; the civic problem transforms that engagement into a decision-making task; and the rubric converts the resulting interpretation into assessable evidence of HOTS and civic reasoning. This linkage is important because immersive media can easily remain at the level of visual attraction when it is not anchored in a problem, a disciplinary lens, and explicit performance criteria. In this study, therefore, the virtual museum functions as a mediated heritage space where local artefacts are not merely viewed, but mobilised as evidence for analysing public issues, evaluating preservation choices, and designing feasible civic responses.

3. Method

3.1 Research design and setting

This research used an embedded pilot case-study design. The design was appropriate because the study examined a bounded classroom assessment event involving a specific group of vocational students, one digital museum environment, and one case-based authentic assessment task. Case-study logic is suitable when the research question asks how an intervention works within a real educational context rather than whether a treatment causes a population-level effect (Yin, 2018).

The setting was a history learning and assessment activity using the virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja, Purbalingga, Central Java, Indonesia. Official museum information describes the museum as a local cultural and historical institution in Purbalingga and provides access to a virtual tour containing spaces such as the lobby and exhibition rooms (Museum Soegarda, 2024a, 2024b). The museum was selected because it offers locally meaningful artefacts that can be connected to students' civic identity and community heritage.

3.2 Participants

The participants were 16 vocational high school students from five vocational schools in Purbalingga who were undertaking internship placement. The inclusion criteria were active enrolment in vocational school, participation in internship placement, availability to complete the 90-minute assessment activity, and willingness to complete the observation and perception instruments. Because the sample was small and purposive, the study is interpreted as a pilot assessment validation case rather than as a generalisable experimental evaluation.

Table 1. Participants and data sources

Component	Description	Analytic function
Participants	16 vocational students undertaking internship placement in Purbalingga	Bounded pilot case; not intended for population-level inference
Digital stimulus	Virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja	Provides artefact-based evidence and spatial heritage context
Performance task	Case-based written response on local heritage preservation and youth participation	Generates evidence of historical HOTS and civic reasoning
Rubric	Four analytic dimensions: information identification, causal analysis, evidence-based civic argument, innovative solution	Converts written responses into interpretable performance indicators
Perception questionnaire	12 Likert-scale items, four response points, maximum total score = 48	Provides descriptive evidence of feasibility, acceptance, and perceived learning support
Field notes	Researcher notes during online/digital activity	Supports triangulation of implementation fidelity and student engagement

3.3 Assessment procedure

The assessment was conducted over approximately 90 minutes and consisted of five sequential phases: orientation, virtual exploration, guided observation, case-based written assessment, and reflection. The sequence was deliberately designed to avoid treating the virtual tour as a stand-alone entertainment medium. Students first received an orientation to the civic relevance of local heritage. They then explored the virtual museum and recorded relevant artefact observations. After this, they completed a case-based problem concerning threats to local cultural heritage, low youth participation, and the need for feasible preservation actions. Finally, they completed a reflection and perception questionnaire.

Table 2. Authentic assessment blueprint linking task phases, HOTS, and civic reasoning

Phase	Student activity	Targeted evidence	HOTS / civic reasoning indicator
Orientation	Receives task framing on local heritage, museum evidence, and public responsibility	Readiness to interpret heritage as a civic issue	Problem recognition; relevance framing
Virtual exploration	Navigates lobby, exhibition spaces, and collection information in the virtual tour	Artefact noticing and source selection	Evidence identification; historical significance
Guided observation	Records artefacts and related information on an observation sheet	Evidence accuracy and selection rationale	C2-C4 transition from noticing to categorising
Case-based response	Analyses local heritage problems, evaluates responsibility, and proposes a solution	Written argument and action proposal	C4 causality; C5 judgement; C6 solution design; civic reasoning
Reflection	Completes questionnaire and open feedback	Perceived feasibility, engagement, and conceptual clarity	Metacognitive evaluation of the assessment experience

3.4 Instruments and data analysis

The main performance instrument was an analytic rubric with four dimensions scored on a 1-4 scale: information identification, causal analysis, evidence-based civic argument, and innovative solution. The perception questionnaire consisted of 12 items organised into four dimensions: enjoyment of history learning, understanding of learning material, use of digital learning media, and independent student engagement. The available quantitative data were analysed descriptively using mean, median, standard deviation, range, percentage of maximum score, coefficient of variation, and a descriptive confidence interval for the total score mean. These statistics are interpreted as pilot evidence of acceptability and dispersion; they are not used to claim causal effectiveness.

The qualitative data were analysed through evidence mapping. The written responses and field notes were interpreted according to the rubric dimensions and then synthesised into cognitive-civic patterns. This strategy is consistent with qualitative display logic, where matrices help connect raw evidence, codes, and interpretive claims (Miles et al., 2020). Trustworthiness was supported through triangulation across task outputs, observation notes, and student perception data.

4. Results

4.1 Implementation fidelity: the assessment functioned as a structured evidence chain

The first result concerns implementation fidelity. The activity did not simply ask students to visit a website and report whether they liked it. It operated as a structured evidence chain: virtual exploration generated artefact observations; the

observation sheet directed attention toward relevant evidence; the case problem required interpretation and judgement; the rubric specified the performance criteria; and the reflection instrument captured students' perceived experience. This chain is important because it links digital access to assessable reasoning rather than assuming that visual exposure automatically creates learning.

The implementation design also responded to the vocational-school context. Students who were undertaking internship placement had limited time and uneven access to physical museum visits. The virtual tour therefore functioned as an access equaliser, while the case-based task preserved cognitive demand. In this sense, the central innovation is not the use of technology alone; it is the alignment among virtual access, local heritage evidence, and civic problem-solving.

4.2 Student acceptance was high but not psychometrically overclaimed

The reconstructed total perception scores ranged from 38 to 48 out of 48, with a mean of 43.38, median of 44, and standard deviation of 2.42. The mean represented 90.36% of the maximum possible score, indicating a very positive post-activity perception. The 95% descriptive interval for the mean was 42.09-44.66, and the coefficient of variation was only 5.58%, indicating low dispersion across participants. Nine students (9/16; 56.25%) scored 44 or higher, and six students (6/16; 37.50%) scored 45 or higher. Only four students scored 41 or below, which suggests that the positive reception was not driven by a single high-performing subgroup.

This pattern supports an acceptability claim: students generally perceived the virtual museum-based assessment as engaging and feasible. However, the data do not support a causal claim that the virtual tour was more effective than conventional instruction because the design did not include a pretest, control group, or delayed retention measurement. Reframing the claim in this way strengthens the article's methodological credibility and avoids a common weakness in technology-enhanced learning studies.

Table 3. Descriptive statistics of total perception scores

Indicator	Value	Interpretation
Number of students	16	Pilot classroom case
Score range	38-48	All students were above 79% of the maximum score
Mean	43.38/48	90.36% of the maximum score
Median	44	Central tendency close to the upper quartile cluster
Standard deviation	2.42	Low absolute dispersion
95% descriptive interval	42.09-44.66	Estimated mean range for this pilot group only
Coefficient of variation	5.58%	Relatively homogeneous positive perception
Students scoring >=44	9/16 (56.25%)	Majority positioned in the high-acceptance band

Indicator	Value	Interpretation
Students scoring <=41	4/16 (25.00%)	Small lower-response cluster requiring diagnostic attention

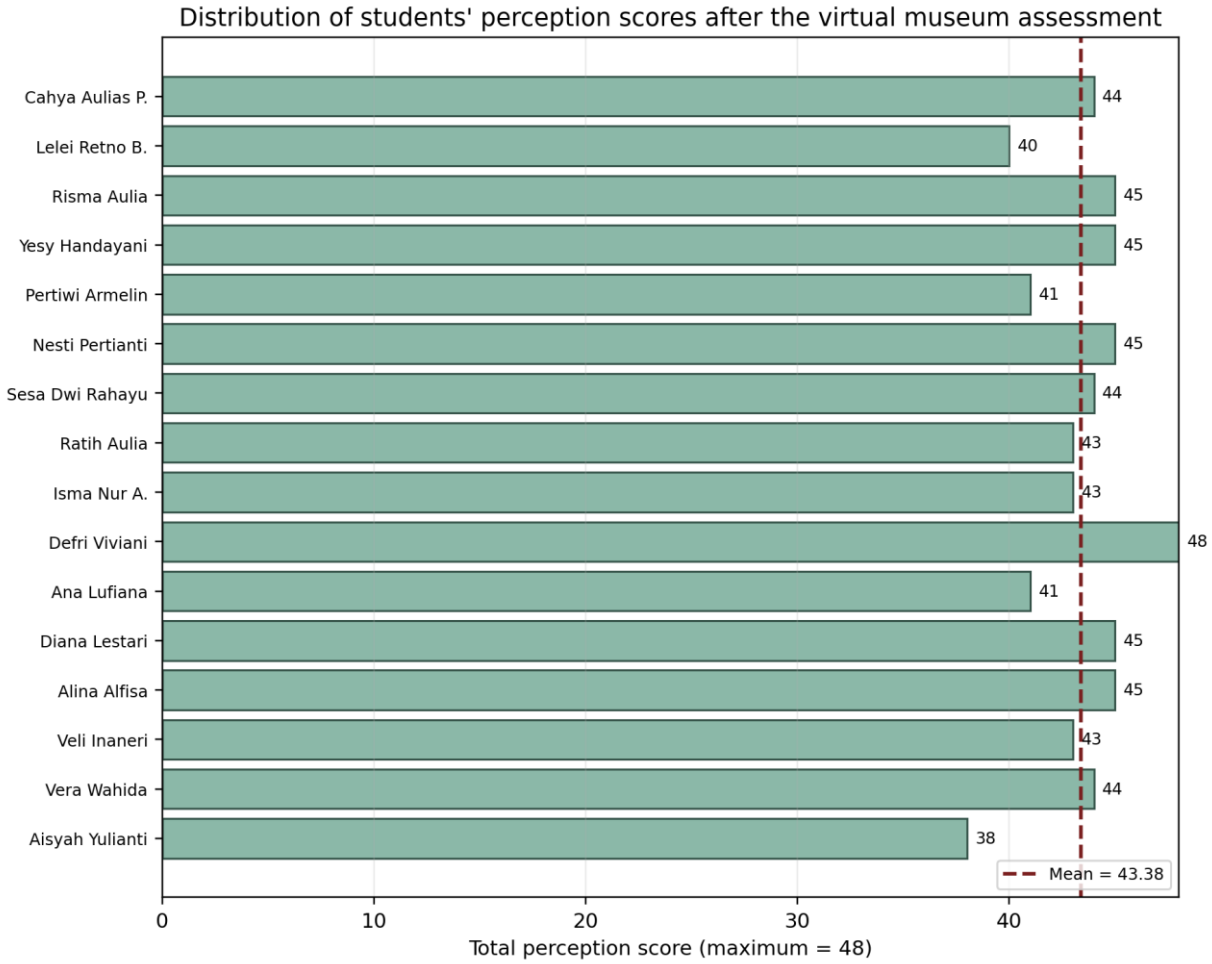

Figure 2. Distribution of students' total perception scores after completing the virtual museum-based authentic assessment

4.3 The strongest signal was media-supported engagement; the weakest signal was immediate conceptual clarity

Dimension-level analysis revealed a more diagnostic pattern than the total scores alone. Media use and independent engagement were the strongest dimensions (both M = 3.76/4; 94.00% index), followed closely by enjoyment of history learning (M = 3.73/4; 93.25% index). The lowest dimension was understanding of learning material (M = 3.42/4; 85.50% index). The difference between the strongest engagement/media dimensions and understanding was 0.34 points, equivalent to an 8.50 percentage-point gap on the four-point scale.

This media-activity/understanding gap is the most important quantitative finding in the article. It shows that the virtual tour successfully supported interest, activity, and perceived usability, but conceptual clarity remained comparatively weaker. Item-level signals sharpen this interpretation: students gave maximum

mean responses to the attractiveness of learning history through virtual media, the support for remembering material, and the interactivity of the media, while immediate clarity of understanding was lower and discussion/asking behaviour was only moderate. In practical terms, students could enjoy and remember the digital visit while still needing stronger prompts to transform observation into historical explanation.

Table 4. Dimension-level perception results and diagnostic interpretation

Dimension	Mean / 4	Index	Diagnostic meaning
Enjoyment of history learning	3.73	93.25%	The virtual tour reduced affective distance from history and increased initial interest
Understanding of learning material	3.42	85.50%	The weakest dimension; conceptual explanation requires stronger scaffolding
Use of digital learning media	3.76	94.00%	The media was perceived as accessible, interactive, and suitable for the task
Independent student engagement	3.76	94.00%	Students were willing to explore and complete the task independently
Overall mean	3.67	91.75%	Strong acceptability with a visible cognition gap
Media-activity/understanding gap	0.34	8.50 percentage points	A signal that technology must be coupled with guided reasoning prompts

Table 5. Item-level signals used to interpret the media-activity/understanding gap

Item signal	Mean	Interpretive function
P2: History learning through virtual media is interesting	4.00	Affective entry point; strong novelty and attractiveness
P5: The activity helps students remember material after learning	4.00	Retention support; visual memory is strong
P9: The virtual tour is more interactive than conventional media	4.00	Media affordance; interactivity is recognised
P12: Students explore learning material independently	3.82	Independent engagement is high but not complete
P11: Students ask questions or discuss during the activity	3.55	Dialogic engagement requires stronger facilitation
P4: Students understand the material clearly during the activity	3.00	Core diagnostic weakness; observation has not fully become conceptual explanation

4.4 HOTS emerged when artefact observation was converted into causal, evaluative, and creative reasoning

The written case responses indicated that the assessment task enabled a movement from artefact noticing to higher-order reasoning. The key evidence was not the number of artefacts mentioned but how students used those artefacts to make claims. Stronger responses showed three features. First, students identified heritage objects or museum narratives as evidence rather than as decorative details. Second, they linked heritage problems to contemporary social causes, such as low youth participation, weak promotion, modern entertainment preferences, and limited public awareness. Third, they proposed preservation actions that connected civic responsibility with vocationally relevant skills.

This pattern shows that HOTS was not located in the virtual tour itself. It emerged in the conversion zone between museum evidence and case-based problem-solving. Students demonstrated C4 analysis when they explained why cultural heritage becomes vulnerable. They demonstrated C5 evaluation when they judged the adequacy of community or government responses. They demonstrated C6 creation when they proposed digital storytelling, social media promotion, youth campaigns, or school-museum collaboration. These proposals are important because they connect the vocational-school context to local civic action rather than treating history as a detached academic subject.

Table 6. Evidence map of HOTS and civic reasoning in the authentic assessment task

Reasoning layer	Observable response feature	HOTS level	Civic reasoning marker	Quality criterion
Artefact-based noticing	Mentions museum collections, figures, places, or local heritage narratives	C2-C3	Recognises heritage as public memory	Accuracy of evidence selection
Causal analysis	Explains why young people ignore museums or why local heritage is at risk	C4	Frames heritage loss as a public problem	Specific cause-effect link rather than generic complaint
Evidence-based judgement	Evaluates the role of schools, local government, museum managers, community, or youth	C5	Distributes responsibility across civic actors	Argument supported by museum evidence and local context
Vocational civic solution	Designs digital storytelling, promotion media, community campaign, or documentation project	C6	Turns historical understanding into feasible action	Feasibility, originality, and alignment with vocational skill
Reflective stance	Explains why local heritage matters for identity and future community life	C5-C6	Connects self, community, and public good	Ethical and identity-based justification

4.5 A revised assessment model: guided virtual evidence inquiry

The results support a revised model that can be called guided virtual evidence inquiry. The model contains four stages: encounter, selection, transformation, and action. In the encounter stage, the virtual tour provides access and affective engagement. In the selection stage, students choose artefacts or narratives that can function as evidence. In the transformation stage, they use that evidence to analyse causes and evaluate responsibilities. In the action stage, they produce a civic solution relevant to their community and vocational capacity.

The model explains why the understanding score was lower than the engagement score. Without strong selection and transformation prompts, virtual tours may remain at the encounter stage. The instructional implication is clear: teachers need to design observation questions that ask not only 'What do you see?' but also 'What does this evidence prove?', 'What public problem does it illuminate?', 'Whose responsibility is implicated?', and 'What action would be feasible in your vocational context?

Table 7. Guided virtual evidence inquiry model for future classroom implementation.

Stage	Teacher design move	Student performance evidence	Risk if omitted
Encounter	Provide access to virtual museum spaces and clarify task relevance	Student explores artefacts and recognises local heritage	Virtual visit becomes recreational browsing
Selection	Use guided observation sheet requiring evidence choice and justification	Student selects artefacts as claim-supporting evidence	Students list objects without interpretation
Transformation	Provide case prompts requiring causality, evaluation, and ethical judgement	Student constructs historical-civic argument	Understanding remains shallow despite high interest
Action	Require feasible civic solution linked to vocational skill	Student designs digital/community action proposal	Civic reasoning stops at moral statement

5. Discussion

5.1 The study shifts the value of virtual museums from access to assessable reasoning

The main contribution of the study is a shift in how virtual museums are positioned in history education. Previous technology-enhanced learning studies often emphasise motivation, presence, and interactivity. Those dimensions matter, and the present data confirm that media-supported engagement was high. However, the stronger contribution is methodological: the virtual museum is treated as an evidence environment for authentic assessment. This reframing matters because an educational technology becomes academically meaningful only when it generates interpretable learning evidence, not merely positive student reactions.

The findings support the broader literature on multimedia and immersive learning, which warns against assuming that richer media automatically produce deeper learning (Mayer, 2020; Makransky & Petersen, 2021). In this pilot, students' perceived interest and interactivity were higher than conceptual clarity. The article therefore adds a local-history example to a wider international problem: technology can increase engagement while leaving cognition under-scaffolded. For Q1-level contribution, this tension is more important than a simple success narrative.

5.2 The media-activity/understanding gap is a productive diagnostic finding

The most analytically useful result is the media-activity/understanding gap. Rather than treating the lower understanding score as a weakness of the intervention, the study interprets it as a design signal. The virtual tour successfully solved an access and engagement problem; it did not fully solve the problem of historical explanation. That distinction allows the study to generate a practical theory of improvement: the next iteration should strengthen guided inquiry prompts, artefact-to-claim scaffolds, peer discussion, and rubric-based formative feedback.

This interpretation also protects the manuscript from overclaiming. A small pilot with no comparison group cannot demonstrate that virtual tours are more effective than conventional teaching. What it can demonstrate is that a virtual museum-based assessment was feasible, positively received, and diagnostically informative. It can also show that students' written responses contained assessable patterns of C4-C6 reasoning when the virtual tour was tied to a local civic problem. These are more defensible claims and therefore stronger for international publication.

5.3 Authenticity was produced by the relationship between local heritage, vocational identity, and public action

The authenticity of the assessment did not come only from the museum context. It came from the relationship between local heritage, vocational identity, and public action. For vocational students, history can feel distant when it is assessed through recall of national chronology. It becomes more meaningful when students are asked to use historical evidence to design practical actions that resemble skills they value: digital communication, promotion, documentation, service design, tourism support, or community engagement. This is the point at which HOTS and civic reasoning intersect.

The study therefore contributes to civic education by showing that civic reasoning can be assessed through local heritage problems. Civic reasoning is often discussed in relation to political discourse or controversial public issues (Hess & McAvoy, 2015; Lee et al., 2021). This article extends that conversation by showing that cultural preservation is also a civic problem requiring evidence, judgement, responsibility, and collective action. The museum becomes not only a site of memory but also a civic laboratory for vocational learners.

5.4 Practical implications

For teachers, the practical implication is to integrate virtual tours with explicit evidence prompts. A strong prompt should require students to identify one artefact, state what historical claim it supports, connect it to a local public problem, evaluate

the relevant civic actors, and propose an action. For museum managers, the implication is to enrich virtual tour pages with inquiry prompts, thematic routes, short evidence cards, and downloadable worksheets. For curriculum designers, the implication is that local heritage can support project-based and authentic assessment within vocational education without requiring a full-day physical visit.

For future implementation, the model should include peer discussion before the written response. The moderate score for asking or discussion indicates that students need a structured dialogic space. This can be implemented through small-group evidence debates, digital Padlet-style boards, or a museum inquiry forum embedded in the virtual tour. The goal is not only to increase interaction but to make students justify claims before finalising their written civic action proposals.

5.5 Limitations and future research

The study has several limitations. The sample was small and purposive, the design did not include a control group, and the available quantitative data were perception-based rather than performance-score distributions. Consequently, the study cannot estimate causal learning gains or generalise to all vocational students. The article also does not claim psychometric validation of the rubric because inter-rater reliability and item-level rubric data were not available in the current dataset.

Future research should develop a quasi-experimental or design-based research sequence involving pretest-posttest HOTS measures, rubric-scored written responses, inter-rater reliability, delayed retention, and comparison with non-virtual museum tasks. A larger study should also test whether vocationally aligned civic action prompts produce stronger C6 performance than generic preservation prompts. These steps would allow the pilot model developed here to move from feasibility evidence toward stronger validity and effectiveness evidence.

6. Conclusion

This article optimises the original classroom implementation report into a theoretically grounded and data-conscious pilot study of virtual museum-based authentic assessment. The virtual tour of Museum Prof. Dr. R. Soegarda Poerbakawatja functioned as an accessible local heritage stimulus for vocational students undertaking internship placement. Descriptive analysis of the available perception data showed high acceptance, low score dispersion, and strong media-supported engagement. The most important finding, however, was not the high overall score; it was the gap between media/engagement and conceptual understanding. This gap demonstrates that virtual tours need guided evidence inquiry to produce historical explanation.

The study's contribution is the guided virtual evidence inquiry model. In this model, students move from encountering digital heritage to selecting evidence, transforming that evidence into causal and evaluative reasoning, and designing civic action. The model makes HOTS and civic reasoning assessable within a local-history task and provides a feasible pathway for integrating museums, technology, and vocational education. With stronger performance data and comparative design in future studies, this model has the potential to contribute to international debates on authentic assessment, digital heritage learning, and civic-historical reasoning.

Declarations

Ethics statement

The study involved a classroom-based educational assessment activity with vocational students. Future submission should include the institutional ethics approval number or school-level permission letter, student assent, and parental/guardian consent where required by the target journal. No personally sensitive data should be reported in the manuscript.

Funding

No external funding was reported for this study.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability

The aggregated questionnaire scores, rubric blueprint, and anonymised coding matrix can be made available by the corresponding author upon reasonable request. Individual student identifiers should be removed before data sharing.

Author contributions

Affan Budi Santosa contributed to conceptualisation, data collection, classroom implementation, and initial drafting. Wildan Nurul Fajar contributed to supervision, methodology refinement, interpretation, and manuscript review. Both authors approved the final manuscript.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Badan Standar, Kurikulum, dan Asesmen Pendidikan. (2025). Panduan pembelajaran dan asesmen. Kementerian Pendidikan Dasar dan Menengah Republik Indonesia.

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in reflexive thematic analysis? Qualitative Research in Psychology, 18(3), 328-352. https://doi.org/10.1080/14780887.2020.1769238

Brookhart, S. M. (2010). How to assess higher-order thinking skills in your classroom. ASCD.

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE.

Darling-Hammond, L., & Adamson, F. (2014). Beyond the bubble test: How performance assessments support 21st century learning. Jossey-Bass.

Falk, J. H., & Dierking, L. D. (2016). The museum experience revisited. Routledge.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52(3), 67-86. https://doi.org/10.1007/BF02504676

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. Journal of Computers in Education, 8, 1-32. https://doi.org/10.1007/s40692-020-00169-2

Hess, D. E., & McAvoy, P. (2015). The political classroom: Evidence and ethics in democratic education. Routledge.

International Council of Museums. (2022). Museum definition. ICOM.

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. American Educational Research Journal, 54(1), 3-34. https://doi.org/10.3102/0002831216679817

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1-73. https://doi.org/10.1111/jedm.12000

Lee, C. D., White, G., & Dong, D. (Eds.). (2021). Educating for civic reasoning and discourse. National Academy of Education.

Lévesque, S. (2008). Thinking historically: Educating students for the twenty-first century. University of Toronto Press.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. Educational Psychology Review, 33, 937-958. https://doi.org/10.1007/s10648-020-09586-2

Mayer, R. E. (2020). Multimedia learning (3rd ed.). Cambridge University Press.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). Qualitative data analysis: A methods sourcebook (4th ed.). SAGE.

Monte-Sano, C. (2010). Disciplinary literacy in history: An exploration of the historical nature of adolescents' writing. The Journal of the Learning Sciences, 19(4), 539-568. https://doi.org/10.1080/10508406.2010.481014

Museum Soegarda. (2024a). Home: Museum Prof. Dr. R. Soegarda Poerbakawatja. Museum Soegarda.

Museum Soegarda. (2024b). Virtual tour museum. Museum Soegarda.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing what students know: The science and design of educational assessment. National Academies Press.

Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. Computers & Education, 179, Article 104429. https://doi.org/10.1016/j.compedu.2021.104429

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education, 147, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

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

UNESCO. (2015). Recommendation concerning the protection and promotion of museums and collections, their diversity and their role in society. UNESCO.

VanSledright, B. A. (2014). Assessing historical thinking and understanding: Innovative designs for new standards. Routledge.

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. Jossey-Bass.

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

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.