Cakrawala Pendidikan
Jurnal Ilmiah Pendidikan
Vol. 45 No. 3, October 2026, pp.575-589
https://journal.uny.ac.id/index.php/cp/issue/view/3181
DOI: https://doi.org/10.21831/cp.v45i3.88828

An augmented reality smart module for critical thinking in pre-service teachers learning megalithic history

Lukman Nadjamuddin[1]*, Aman[2], Idrus Rore[1], Windayanti[1], Muhammad Zaky[1], Wilman D. Lumangino[1], Nurgan Tadeko[1]
[1]Universitas Tadulako, Indonesia
[2]Universitas Negeri Yogyakarta, Indonesia
*Corresponding Author: lukman_n@untad.ac.id

ABSTRACT

Learning megalithic history requires students to connect artifact features with historical evidence. This study developed and appraised an augmented reality (AR) Smart Module using ADDIE and compared critical-thinking test performance in four intact classes at Universitas Tadulako. The pretest–posttest study included 122 undergraduate history education students, with 61 in each condition. Five experts rated ten assessment items; recalculated Aiken's V ranged from .600 to 1.000, with a mean of .853. AR posttest scores averaged 7.800 (SD 1.012), compared with 6.171 (SD 1.205) for lecture-based instruction. ANCOVA including pretest, cohort, and group × cohort gave an adjusted difference of 1.637 points (95% CI [1.245, 2.028]), $F(1,117) = 68.551$, $p = 2.30 \times 10^{-13}$, partial $\eta^2 = .369$. Mean gains were 2.801 and 1.170; the gain comparison gave $t(120) = 6.127$, $p = 1.17 \times 10^{-8}$, $d = 1.109$. The results show stronger critical-thinking test performance for students using the AR Smart Module in this classroom comparison. Globally, the study offers a design sequence for connecting digital heritage representations with contextual evidence and evidence-based historical reasoning.

Keywords: augmented reality, critical thinking, history education, cultural heritage, educational technology

Article history

Received:	*Revised:*	*Accepted:*	*Published:*
30 July 2025	*23 July 2026*	*07 September 2026*	*20 October 2026*

Citation (APA Style): Nadjamuddin, L., Aman, A., Rore, I., Windayanti, W., Zaky M., Lumangino, W. D., & Tadeko, N. (2026). An augmented reality smart module for critical thinking in pre-service teachers learning megalithic history. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan, 45*(3), pp.575-589. DOI: https://doi.org/10.21831/cp.v45i3.88828

INTRODUCTION

Historical inquiry asks learners to assess what a source shows, how it was produced, and which conclusions the evidence can support. In megalithic history, students must relate visible features of artifacts to contextual information about their use and past societies. This work calls for critical thinking: interpreting information, analyzing relationships, evaluating evidence, drawing warranted inferences, and explaining conclusions (Facione, 1990). Recent research on digital history literacy likewise connects the evaluation of historical information with critical thinking (Nurhasanah et al., 2025).

This disciplinary reasoning is important for pre-service history teachers because they need to practice evaluating historical claims as learners before guiding school students through similar inquiry. Their professional preparation makes this an especially relevant context for examining historical reasoning. Recent work on structured learning in higher education also underscores the role of instructional design when critical thinking is an intended outcome (Nadjamuddin et al., 2026).

The Kalamba stone jars and Tadulako statues of Central Sulawesi provide tangible material for such inquiry. Students can observe their form and features, then distinguish these observations from claims about function, chronology, and cultural meaning that require contextual evidence. Megalithic sites thus offer more than visual content: they give future history teachers a concrete

setting in which to practice moving from observation to a reasoned historical interpretation. Studies of local-history teaching materials and local cultural learning support the educational relevance of this contextual approach (Ma'unah et al., 2018; Pamungkas et al., 2023).

Augmented reality (AR) places digital information or three-dimensional objects within a view of the physical environment. In a history lesson, learners can rotate and inspect an artifact model, then relate its visible features to written historical information and other evidence. Recent research has connected AR information design with the presentation of historical knowledge (Ichsan et al., 2026), while an AR-supported living-museum study examined historical learning in a heritage setting (Shawarangkoon et al., 2024). These applications make AR relevant to the design of history learning, while the quality of historical reasoning still depends on the questions and evidence learners are asked to consider.

This question also matters for teacher education. Pre-service teachers increasingly encounter digital representations as part of historical learning and need to judge how such representations relate to sources and claims. Research on AR competencies among teacher-education students identifies preparation for AR integration as a relevant concern (Salmaini et al., 2025). In Indonesia, recent work describes the integration of digital technologies in history learning and its connection to cultural resources (Fahruddin et al., 2025). Together, these studies support examining AR as a learning resource for historical inquiry, rather than assuming that technology itself produces critical thinking.

Earlier reviews document AR applications in history education and cultural heritage (Challenor & Ma, 2019; Boboc et al., 2022). These reviews provide a context for the present module, but they do not evaluate its classroom outcomes. Karima (2024) examines cognitive bias among pre-service history teachers. This provides a close population comparison, although cognitive bias differs from the critical-thinking test performance assessed here. Gonnermann-Müller and Krüger (2025) review AR design and cognitive load. Their work highlights the importance of examining instructional design alongside the technology.

These studies address related but distinct contexts and outcomes. The present study focuses on Kalamba and Tadulako representations and pairs them with historical content and assessment tasks. It examines selected critical-thinking skills in one pre-service history program in Central Sulawesi. The classroom comparison evaluates the complete instructional package, including its historical content, learning activities, and AR representations.

The study develops and appraises an AR Smart Module based on Central Sulawesi megalithic heritage. It then compares students' critical-thinking outcomes with those in a lecture-based condition. Its contribution is a documented account linking local historical material, an AR-supported learning resource, expert appraisal, and a baseline-adjusted classroom comparison.

The objectives correspond to three research questions:

1. How were megalithic heritage resources incorporated into the Smart Module through the ADDIE development process?
2. What do expert ratings indicate about the content relevance of the assessment items and the suitability of the module design?
3. How do critical-thinking outcomes differ between the AR Smart Module and lecture-based comparison conditions after accounting for baseline performance and cohort?

METHOD

Research design

The module-development process followed the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) framework (Branch, 2009). The classroom evaluation used a nonrandomized pretest–posttest comparison between the AR Smart Module group and a lecture-based comparison group. The two conditions differed as instructional packages, combining their respective media and learning activities.

The AR module was developed to deliver historical content through interactive 3D representations of regional artifacts. By incorporating megalithic cultural heritage into the learning experience, the module aimed to enhance student engagement and deepen historical

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

understanding. Content and interface development followed an iterative design process, involving consultation with historians, pedagogy, and multimedia experts.

The outcome comprised ten historical assessment items organized under five intended indicators: interpretation, analysis, evaluation, inference, and explanation. It includes historical passages, migration maps, photographs, and artifact descriptions. Item 1 contains four true/false statements, item 6 permits multiple selections, and the remaining items use selected responses. Student performance is analyzed using the recorded pretest and posttest total scores.

The analysis reports sample counts, means, standard deviations, observed ranges, and gain scores (posttest minus pretest) by group. An independent-samples pretest comparison describes observed baseline differences. The main ANCOVA uses posttest as the dependent variable, pretest as a covariate, and group, cohort, and group × cohort as factors, following the supplied SPSS specification. Adjusted means are evaluated at the overall pretest mean and averaged equally over cohorts. Baseline-adjusted reporting follows Vickers and Altman (2001). Independent-samples gain and posttest tests are secondary contrasts; paired tests describe within-group change. Two-sided p values, degrees of freedom, 95% confidence intervals, and effect sizes are reported for each contrast.

Cohen's d for independent contrasts uses the pooled standard deviation of the corresponding outcome; paired d_z uses the standard deviation of individual change. The ANCOVA group term is summarized with partial η^2. Residual and model diagnostics were reviewed, including the group-by-pretest term. The numerical output was cross-checked against the final score file. The allocation structure is described with the results and informs interpretation of the classroom comparison.

The main model expresses expected posttest total score as a function of pretest, group, cohort, and the group-by-cohort term. Group distinguishes the AR and comparison conditions; cohort distinguishes first- and third-year students. The overall condition contrast averages the two cohort-specific differences equally, consistent with the estimated marginal means. Cohort-specific contrasts and the interaction are reported as features of this classroom comparison.

Five experts rated the relevance of ten assessment items on a four-point scale. Aiken's V was recalculated as $V = \Sigma(r - 1) / [n(c - 1)]$, with $n = 5$ raters, $c = 4$ response categories, and denominator 15 (Aiken, 1985). This coefficient summarizes content relevance, not internal consistency or the reliability of students' total scores. The module was appraised separately across six dimensions: perceived instructional effectiveness, engagement, support for learners, interactive learning, collaborative learning, and compatibility. Module averages are reported as continuous appraisal values without an undocumented validity cut off.

Assessment format and scoring documentation

Appendix A reproduces the assessment and its intended item-to-indicator mapping. Item 1 asks students to classify four statements as true or false after reading a migration passage; item 2 asks them to identify a map associated with an account of population origins. Student pretest and posttest totals are analyzed as recorded, and their observed ranges are reported in Table 9.

The outcome analysis uses the pretest and posttest total scores recorded in the study dataset. Aiken's V values describe experts' ratings of item relevance, while the item-to-indicator mapping documents the intended assessment structure.

Participants

The analyzed dataset comprises 122 undergraduate students in the History Education Department at Universitas Tadulako: 61 in the AR condition and 61 in the comparison condition. Four intact classes participated: classes 1 and 2 received the AR condition, and classes 3 and 4 received the comparison condition. Classes 1 and 3 comprise first-year students; classes 2 and 4 comprise third-year students. The sample included 58 male and 64 female students. Table 4 summarizes participants by condition and cohort.

All 122 records contain pretest and posttest totals. Pretest is retained as a continuous covariate in the main analysis, and the primary results are reported for the full sample.

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

Class allocation and implementation record

The four intact classes were allocated across condition and cohort as summarized in Table 1. The comparison therefore reflects the observed classroom allocation, and condition differences are interpreted in that context.

Table 1. Participants by condition and cohort

Category	Condition	Subcategory	n
Total	Both	Complete pretest and posttest	122
Class 1	AR	First year	31
Class 2	AR	Third year	30
Class 3	Control	First year	31
Class 4	Control	Third year	30
Gender	AR	Male / female	30 / 31
Gender	Control	Male / female	28 / 33

FINDINGS AND DISCUSSION

Findings

Analysis

The Analysis phase focused on the instructional problem of connecting megalithic artifacts with the historical information used to interpret them. Kalamba and Tadulako artifacts supplied a contextual foundation for the module. Learners need to identify relevant features, distinguish competing interpretations, and recognize what additional evidence a claim requires. This is the rationale for the design.

The identified design needs were therefore access to artifact representations and an explicit relationship between those representations and historical questions. These needs do not imply that every student had deficient reasoning or lacked prior exposure to technology. The study evaluates the developed package through expert appraisal and test performance, rather than estimating the prevalence of these needs in a wider population.

Design

The Design phase comprised conceptual and technical components. The conceptual design organized historical content around megalithic artifacts and connected descriptions with questions about evidence and interpretation. The technical design embedded AR markers in the module so that learners could access 3D content using mobile devices. Digital modeling and refinement belong to Development rather than to the initial specification of learning objectives. The module's educational rationale is that repeated artifact inspection can supply information for a historical task; this relationship is a design proposition and was not tested as a mediating mechanism.

Development

The Development phase converted historical content and artifact documentation into a module and Android-based AR resources. Photographs and video supplied visual material for digital representations, which were refined in Blender and integrated with the application. The development account distinguishes reconstruction of visible form from the historical interpretation attached to that form. No geometric accuracy study was conducted, so visual detail is not presented as proof of archaeological accuracy.

Figure 1 documents the artifacts used as visual resources. The resulting models made selected features available for repeated inspection. Their instructional role is to support comparison with historical information, rather than to establish the original function, date, or meaning of an artifact from its appearance alone. Reconstruction decisions and missing contextual information remain relevant to interpretation.

As shown in Figure 2, the models were embedded into an Android-based AR application, allowing students to explore the artifacts through their mobile devices virtually. Figure 3 presents pages of the Smart Learning Module. The module combines historical explanation with access to the AR resources shown in Figure 2. These pages document the product; they do not constitute a

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

usability test or evidence of improved engagement. Expert appraisal and classroom score comparisons supply separate forms of evaluation.



Figure 1. Documentation of the Poso Megalithic Artifacts for Digital Modeling



Figure 2. Screenshot of the Android-based Augmented Reality Application Interface
(These images depict several pages from the AR Application that have been translated into English)

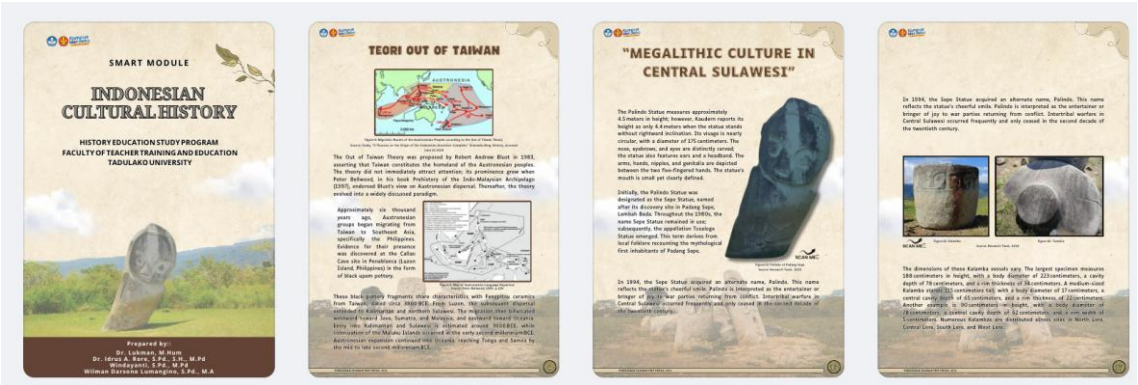


Figure 3. Digital Version of the Smart Learning Module
(These images depict several pages from the Smart Module that have been translated into English)

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

Prototype appraisal addressed content suitability, presentation, and alignment with the intended learning objectives. Five assessment-item raters comprised three specialists in metacognition and two in history education. The manuscript also reports four module validators working across media, pedagogy, and subject content. The three columns in Table 2 describe expertise-area summaries rather than four individual raters. The two panels and their purposes are distinct. Expert judgment evaluates suitability and relevance, whereas classroom effectiveness requires student outcome evidence.

The module summaries indicate which design dimensions experts appraised. They do not show agreement between individual validators or provide a student usability measure. Without the item-by-validator matrix, the three expertise-area columns cannot be reconstructed as four separate raters. This limits interpretation of differences among domain averages, even though their arithmetic means can be checked. The stated panel size and the reporting structure must therefore be reconciled before using the ratings as evidence of a broader measurement property.

Table 2. Module appraisal summaries by expertise area

Indicator	Media Expert	Pedagogy Expert	Content Expert	Average Score	Interpretation
Effectiveness	3.8	3.9	3.7	3.80	Expert appraisal only
Engagement	3.6	3.8	3.7	3.70	Expert appraisal only
Support for Learners	3.5	3.6	3.4	3.50	Expert appraisal only
Interactive Learning	3.7	3.9	3.8	3.80	Expert appraisal only
Collaborative Learning	3.4	3.5	3.3	3.40	Expert appraisal only
Compatibility	3.6	3.7	3.6	3.63	Expert appraisal only

Implementation

Implementation is reported as one classroom comparison involving 61 AR and 61 control students across four classes at Universitas Tadulako. Pretests preceded the instructional period, and posttests followed it. Expert appraisal is part of product development and is not counted as a second independent effectiveness trial.

The classroom evaluation compares the module-supported learning condition with the lecture-based comparison condition. Student test totals provide the outcome data. Observational impressions about interaction and usability are not used to claim an independently measured increase in engagement, collaboration, or cultural awareness. Instructor equivalence, instructional duration, and the implementation of specific learning tasks are important conditions for interpreting this contrast and affect the strength of the resulting inference.

Evaluation

Aiken's V was recalculated from the five expert ratings for each assessment item (Table 3). The coefficients range from 0.600 to 1.000; their arithmetic mean is 0.853. For items 1 and 2, the rating total is 19, so $\Sigma(r - 1) = 14$ and $V = 14/15 = 0.933$. Item 7 has a total of 14, giving $V = 9/15 = 0.600$. No undocumented categorical threshold is applied. Item 7 remains the lowest-rated item and a priority for revision; any revised version requires renewed appraisal.

For the remaining items, the corrected coefficients are .733 for items 3 and 4, 1.000 for item 5, .867 for items 6 and 8, and .933 for items 9 and 10. These corrections follow directly from the displayed ratings. Because the panel has five experts and only four relevance categories, a categorical decision requires a declared rule and appropriate justification. Reporting the coefficients continuously avoids overstating certainty from the earlier unexplained threshold.

Expert appraisal indicates that the module was considered suitable across the reported design dimensions. These ratings describe expert judgments of the module; student outcome differences are presented separately in the classroom analysis.

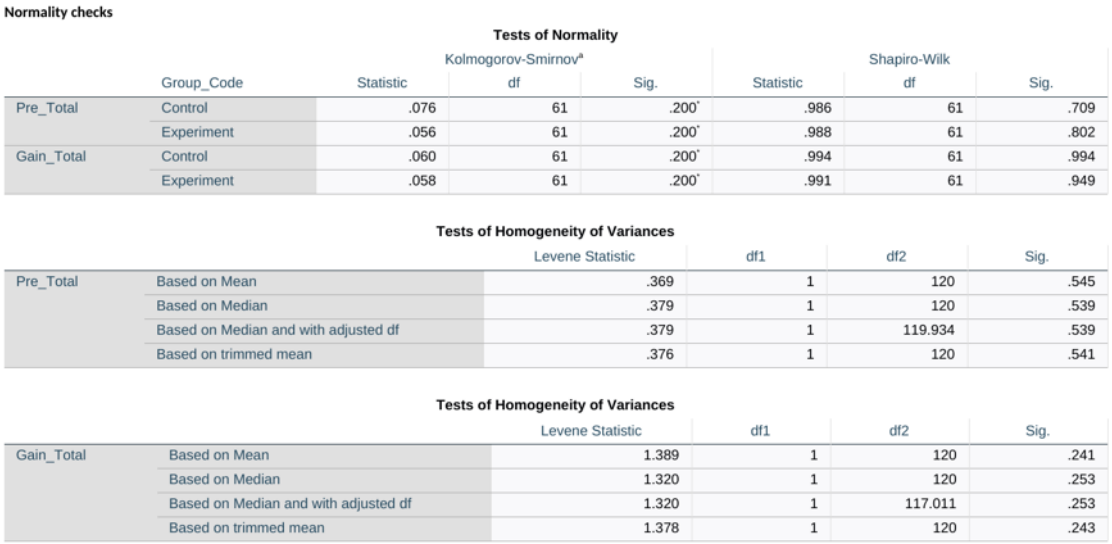
Shapiro–Wilk tests did not reject normality for either condition. Pretest p values were .802 for AR and .709 for control; posttest p values were .626 and .287; gain-score p values were .949 and .994. Levene's test for gain scores gave $F(1,120) = 1.389$, $p = .241$. The supplied four-cell posttest Levene test gave $F(3,118) = 0.311$, $p = .818$. Additional checks gave ANCOVA residual

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

W = .986, p = .263, and group × pretest F(1,116) = 0.758, p = .386. These checks did not identify the tested distributional departures, but they do not establish independence of students within classes. Figure 4 reproduces the supplied SPSS assumption-test output.

Table 3. Recalculated content-relevance coefficients for the critical-thinking instrument

Item	Indicator	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Aiken's V	Appraisal note
1	Interpretation	4	4	3	4	4	0.933	Content-relevance estimate
2	Interpretation	4	4	4	3	4	0.933	Content-relevance estimate
3	Analysis	3	3	3	4	3	0.733	Content-relevance estimate
4	Analysis	3	4	3	3	3	0.733	Content-relevance estimate
5	Evaluation	4	4	4	4	4	1.000	Content-relevance estimate
6	Evaluation	3	4	4	4	3	0.867	Content-relevance estimate
7	Inference	2	3	3	3	3	0.600	Lowest relevance; revise
8	Inference	3	4	4	3	4	0.867	Content-relevance estimate
9	Explanation	4	4	3	4	4	0.933	Content-relevance estimate
10	Explanation	4	4	4	4	3	0.933	Content-relevance estimate



Normality checks

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group_Code	Statistic	df	Sig.	Statistic	df	Sig.
Pre_Total	Control	.076	61	.200*	.986	61	.709
	Experiment	.056	61	.200*	.988	61	.802
Gain_Total	Control	.060	61	.200*	.994	61	.994
	Experiment	.058	61	.200*	.991	61	.949

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Pre_Total	Based on Mean	.369	1	120	.545
	Based on Median	.379	1	120	.539
	Based on Median and with adjusted df	.379	1	119.934	.539
	Based on trimmed mean	.376	1	120	.541

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Gain_Total	Based on Mean	1.389	1	120	.241
	Based on Median	1.320	1	120	.253
	Based on Median and with adjusted df	1.320	1	117.011	.253
	Based on trimmed mean	1.378	1	120	.243

Figure 4. SPSS Normality Output for the Classroom Score Data

The secondary independent-samples gain comparison gave an AR-minus-control difference of 1.630 points, $t(120) = 6.127$, $p = 1.17 \times 10^{-8}$, 95% CI [1.103, 2.157]. Cohen's d was 1.109, using the pooled gain SD of 1.469. This is the test corresponding to the previously rounded t = 6.14; it compares gain scores rather than posttest totals. An unadjusted posttest contrast gave a difference of 1.629 points, $t(120) = 8.088$, $p = 5.54 \times 10^{-13}$, 95% CI [1.230, 2.028], d = 1.464. Baseline-adjusted results are the main between-condition comparison.

Within-group paired tests showed increases in both conditions. AR increased by 2.801 points, $t(60) = 16.206$, $p = 1.37 \times 10^{-23}$, 95% CI [2.455, 3.146], $d_z = 2.075$. Control increased by 1.170 points, $t(60) = 5.786$, $p = 2.79 \times 10^{-7}$, 95% CI [0.766, 1.575], $d_z = 0.741$. These tests

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

describe change within a condition and should not be interpreted as separate tests of the intervention effect. Figure 5 reproduces the SPSS gain-comparison output.

This shift indicates that administrative complexity under Merdeka Belajar has direct implications for teaching practice. The paradox of administrative simplification thus extends beyond governance structures and enters the pedagogical domain, affecting how teachers prioritize their professional tasks.

Independent samples gain comparison

Group Statistics

	Group_Code	N	Mean	Std. Deviation	Std. Error Mean
Gain_Total	Control	61	1.1705	1.57995	.20229
	Experiment	61	2.8007	1.34976	.17282

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Significance One-Sided p
Gain_Total	Equal variances assumed	1.389	.241	-6.127	120	<.001
	Equal variances not assumed			-6.127	117.143	<.001

Independent Samples Test

		t-test for Equality of Means				
		Significance Two-Sided p	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Gain_Total	Equal variances assumed	<.001	-1.63016	.26606	-2.15695	-1.10338
	Equal variances not assumed	<.001	-1.63016	.26606	-2.15708	-1.10325

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
Gain_Total	Cohen's d	1.46937	-1.109	-1.489	-.726
	Hedges' correction	1.47864	-1.102	-1.480	-.721
	Glass's delta	1.34976	-1.208	-1.619	-.789

Figure 5. SPSS Independent-Samples Gain Comparison

The descriptive results distinguish pretest, posttest, and gain totals (Table 4). The AR mean posttest score was 7.800 (SD 1.012), compared with 6.171 (SD 1.205) for control. Pretest means were 5.000 in both conditions. Similar observed baseline means do not imply random assignment or eliminate unmeasured differences between classes.

Table 4. Total-score descriptive statistics by condition and measurement occasion

Measure	AR n	AR mean	AR SD	Control n	Control mean	Control SD
Pretest	61	5.000	0.993	61	5.000	0.947
Posttest	61	7.800	1.012	61	6.171	1.205
Gain	61	2.801	1.350	61	1.170	1.580

Table 5. Estimates on dependent variabel post_total

			95% Confidence Interval	
Group_Code	Mean	Std. Error	Lower Bound	Upper Bound
Control	6.165[a]	.140	5.888	6.442
Experiment	7.801[a]	.140	7.524	8.078

Note: Covariate evaluated at Pre_Total = 5.0000.

At an overall pretest mean of 5.000 and with equal averaging across cohorts, SPSS estimated marginal means were 7.801 for AR and 6.165 for control (Table 5). The adjusted difference was 1.637 points, 95% CI [1.245, 2.028], $F(1,117) = 68.551$, $p = 2.30 \times 10^{-13}$, partial $\eta^2 = .369$. This contrast accounts for the measured baseline score and cohort within the specified model. Its standard error and confidence interval are student-level estimates without adjustment for class clustering.

The AR-minus-control pretest difference was −0.001 points, 95% CI [−0.349, 0.347], $t(120) = -0.006$, $p = .9955$, $d = -0.001$. This check reports baseline similarity on the measured score without proving equivalence of the groups. The cohort term gave $F(1,117) = 2.533$, $p = .1142$, partial $\eta^2 = .021$. Figure 7 provides the model output for the group, cohort, and interaction

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

terms. Full original SPSS tables and additional numerical checks are retained in Appendices C and D.

ANCOVA output

Between-Subjects Factors

		Value Label	N
Group_Code	0	Control	61
	1	Experiment	61
Cohort_Code	1	Year 1	62
	3	Year 3	60

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Post_Total

F	df1	df2	Sig.
.311	3	118	.818

Tests of Between-Subjects Effects

Dependent Variable: Post_Total

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	90.083[a]	4	22.521	18.905	<.001	.393
Intercept	202.306	1	202.306	169.822	<.001	.592
Pre_Total	.056	1	.056	.047	.829	.000
Group_Code	81.664	1	81.664	68.551	<.001	.369
Cohort_Code	3.018	1	3.018	2.533	.114	.021
Group_Code * Cohort_Code	6.027	1	6.027	5.059	.026	.041
Error	139.380	117	1.191			
Total	6182.848	122				
Corrected Total	229.463	121				

Figure 7. SPSS ANCOVA Tests for Group, Cohort, and Their Interaction

Direction conventions explain the signs in the reproduced SPSS output. The independent-samples output lists control before experiment, giving a negative control-minus-AR difference. The narrative reverses the direction to AR minus control, so the corresponding t statistic and standardized difference are positive. Similarly, paired SPSS output uses pretest minus posttest, whereas the narrative reports gain as posttest minus pretest. Reversing a contrast changes its sign and reverses the interval endpoints, but leaves the two-sided p value unchanged. Experiment in the output denotes the AR condition.

Table 6. Intended construct coverage of the historical assessment

Dimension	Items	Coverage	Response format	Evidence boundary	Outcome status
Interpretation	1–2	Passage interpretation; migration-map identification	True/false; selected response	Intended mapping from validation matrix	No verified subscale effect
Analysis	3–4	Evidence and population-origin accounts	Selected response	Recognition differs from written argument	No verified subscale effect
Evaluation	5–6	Assessment of historical information	Selected and multiple response	Administered key and weights need confirmation	No verified subscale effect
Inference	7–8	Conclusions from historical stimulus material	Selected response	Item 7 has V = .600	No verified subscale effect
Explanation	9–10	Artifact-related descriptions and conclusions	Selected response	Does not test independent written explanation	No verified subscale effect

The assessment's intended construct coverage is documented in Table 6. Two items are assigned to each of five dimensions in the original validation matrix. The final score file contains pretest and posttest totals rather than item-level responses. It therefore does not support verified indicator-specific means or claims that one critical-thinking dimension improved more than another.

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

Interpretation, analysis, evaluation, inference, and explanation identify the intended content coverage of the assessment, not five independently estimated treatment effects. Item appraisal supports discussion of relevance; the learning comparison concerns the total score. Item-level responses and the administered scoring key would be needed to calculate indicator scores, evaluate internal consistency, or compare changes across dimensions (Table 7).

Table 7. Effect sizes for total score comparison

Metric	Value
Gain-score difference	1.630 points (95% CI 1.103–2.157)
Pooled gain-score SD	1.469
Gain-score Cohen's d	1.109 (SPSS 95% CI 0.726–1.489)
Adjusted group partial η^2	.369

The gain-score effect size of d = 1.109 expresses the mean difference in gain scores in pooled standard-deviation units. The adjusted partial η^2 summarizes the group term in the ANCOVA; these statistics describe different contrasts and are interpreted accordingly.

The group-by-cohort interaction was $F(1,117) = 5.059$, $p = .0264$, partial $\eta^2 = .041$ (Figure 7). The cohort-specific contrasts are reported in Table 8. Each cohort-by-condition cell comprises one class, so cohort and class variation are represented together in this comparison.

Adjusted AR-minus-comparison differences were positive in both cohorts: 1.189 points for first-year students and 2.084 points for third-year students. The corresponding intervals and p values are reported in Table 8.

Table 8. Adjusted condition contrasts by Cohort

Cohort	Adjusted AR−control	95% CI	t(117)	Exact p
First year	1.189	0.639–1.740	4.278	3.87×10^{-5}
Third year	2.084	1.524–2.643	7.372	2.60×10^{-11}

The outcome analysis concerns total scores on the historical assessment. The module appraisal and developmental observations are reported as separate evidence about the learning resource and its implementation.

Table 9. Observed score ranges by condition

Condition	Measure	Mean	SD	Minimum	Maximum
AR	Pretest	5.000	0.993	3.010	7.700
AR	Posttest	7.800	1.012	5.600	10.000
AR	Gain	2.801	1.350	−0.680	6.520
Control	Pretest	5.000	0.947	2.280	7.610
Control	Posttest	6.171	1.205	3.370	9.490
Control	Gain	1.170	1.580	−2.760	5.060

The observed ranges indicate that change was not uniform. AR gains ranged from −0.680 to 6.520, and control gains ranged from −2.760 to 5.060 (Table 9). Thus, higher average gain does not mean that every AR student improved or that improvement occurred only in the AR condition. These ranges describe individual variation without inferring its causes.

Variation in performance and limits of subgroup inference

The main model retains measured pretest performance continuously and includes cohort. The overall condition contrast is the primary result; gender-specific and prior-knowledge subgroup effects are not part of the analysis.

Discussion

The findings document an ADDIE-developed module combining historical explanation, local artifact representations, and Android-based AR access. Expert ratings describe item relevance and module design, and the classroom comparison shows higher average critical-

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

thinking test performance in the AR condition, including a positive baseline-adjusted contrast. The mean gains of 2.801 and 1.170, together with d = 1.109, summarize the observed difference. Recent AR history and heritage studies provide a relevant context for this work (Shawarangkoon et al., 2024; Fahruddin et al., 2025), while the present results add evidence from pre-service history students working with Central Sulawesi megalithic material.

The module places Kalamba and Tadulako representations alongside historical information used to interpret them. This arrangement connects local heritage to an explicit learning task: students examine artifact features and consider what the available context supports. Recent work on AR-supported museum learning and educational technology in Indonesian history provides related examples of digital heritage in learning environments (Shawarangkoon et al., 2024; Fahruddin et al., 2025). In the present study, the measured outcome is students' performance on historical critical-thinking tasks.

A plausible instructional explanation is that artifact models offer opportunities to inspect relevant features and connect them with contextual information. Historical AR applications show different ways of combining representations with learning activity (Blanco-Fernández et al., 2014; Keil et al., 2011; Shawarangkoon et al., 2024). The AR Smart Module comparison evaluates the combined content, representations, and learning activities. The measured score difference therefore describes this instructional package in the participating classes.

Augmented reality can connect a digital representation with the learner's physical setting, allowing historical objects to be examined alongside written explanations. Aji and Purbojo (2023) describe AR use in Indonesian higher education. Dunleavy and Dede (2014) provide a broader account of AR teaching and learning, while Bacca-Acosta et al. (2019) examine motivational application design. These works make instructional design central to the use of the medium. For the present module, the pertinent relationship is between artifact observation, contextual information, and the historical question students must answer. An interactive display offers opportunities for inspection, but engagement with the interface alone does not demonstrate critical thinking.

Instrument appraisal and outcome analysis provide complementary information. Aiken's V summarizes expert-rated content relevance, and the recorded total scores provide the classroom outcome measure. Item 7 had the lowest relevance coefficient at .600, while the mean coefficient was .853. Expert ratings and classroom score differences describe distinct aspects of the study: assessment content appraisal and student performance.

Karima (2024) offers a closer history-teacher-education comparison through a study of AR and cognitive bias. The present study differs in its megalithic content and total-score assessment. This makes the outcome definition important: the label critical thinking refers to performance on the administered historical tasks, not a demonstrated change in all components of the construct. The two studies share a population context while evaluating different outcomes.

The local heritage content gives the historical inquiry a concrete subject. Students can distinguish a feature visible in a digital model from an interpretation that depends on archaeological and historical context. Recent AR work in history and heritage learning similarly places digital representations within a broader explanatory setting (Shawarangkoon et al., 2024; Fahruddin et al., 2025). This relation between visual observation and contextual evidence is central to the module's critical-thinking tasks.

Local historical content can make a learning problem concrete without replacing the need to evaluate sources. Ma'unah et al. (2018) examine local-history interactive materials, while Dewi and Primayana (2019) study contextual learning modules and conceptual understanding. Rini (2020) considers contextual learning through lesson study. Within history education, Afriani et al. (2022) address character education through the Diponegoro War, and Asri et al. (2020) discuss historical empathy. These outcomes should be distinguished from the critical-thinking score used here. The common instructional consideration is how learners connect information to a meaningful context; the present evaluation specifically examines performance on historical assessment tasks.

Research on heritage technology also spans applications whose purposes differ from classroom assessment. Jung and Dieck (2017) discuss visitor experience through AR, VR, and

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

3D printing; Mendoza et al. (2020) report a heritage-education experience supported by AR. Boboc et al. (2019) describe mobile heritage interpretation across locations, and Abdul-Jabbar and Alwehab (2023) consider the presentation of Kirkuk Citadel. Neumüller et al. (2014) discuss artifact replicas and 3D printing. These applications show different ways to make cultural material available for inspection. Availability, however, must be distinguished from evidence that a learner can evaluate an interpretation. A classroom module therefore needs an assessment aligned with the historical reasoning it intends to support.

The wider relevance of the study lies in its instructional arrangement: connect a represented heritage object with contextual information and ask learners to justify a historical interpretation. Educators can examine this sequence when adapting digital heritage resources to their own material and curricula. The classroom results provide a context-specific example for that design question.

Using this arrangement in other settings requires attention to device availability, connectivity, technical support, and the preparation of accurate digital models. These conditions shape how heritage materials can be represented and used in classroom inquiry. Teacher-education research on AR competencies also points to the need to prepare future educators for integrating the medium into learning (Salmaini et al., 2025).

Bacca et al. (2014) synthesize educational uses, benefits, and limitations of AR across early studies. More recent teacher-education evidence also makes implementation constraints visible. Sat et al. (2023) studied 55 pre-service teachers attending an online seminar about AR tools. Their findings concern perceptions of tools and design experiences rather than the historical reasoning assessed here. They report challenges involving application affordability, educational content, connectivity, and device access. These are reasons to examine feasibility separately from score differences. A successful classroom comparison at one university cannot establish that the same resources are available or workable elsewhere.

Several features of the study define priorities for subsequent research: 1) The study involved 122 students in four intact classes, with one class in each condition-by-cohort cell. Future studies involving more classes can examine classroom variation alongside student pretest performance and provide a broader basis for estimating condition differences. 2) The assessment measured immediate pretest and posttest performance using selected-response historical tasks. Future work can add delayed measures and written responses to examine retention and students' independently constructed historical arguments. 3) The module uses localized megalithic heritage content. Adaptation to other contexts can review the accuracy of representations, historical sources, instructional questions, and assessment items with local experts.

Future research can compare instructional conditions that systematically vary access to AR while keeping historical content, instructional time, questions, and teacher support comparable. Studies across more classes and institutions can also examine persistence of outcomes, implementation conditions, and transfer to written historical reasoning.

CONCLUSION

The study documents an ADDIE-developed AR Smart Module that connects megalithic heritage representations with historical information. Expert appraisal describes the relevance of assessment content and the design suitability of the module.

Students in the AR condition achieved higher average posttest and gain totals than students in the lecture-based comparison condition. The baseline- and cohort-adjusted difference was 1.637 points, 95% CI [1.245, 2.028], with a gain-score standardized difference of $d = 1.109$. These results concern performance on selected historical critical-thinking tasks in the participating classes. The educational contribution is an instructional arrangement in which pre-service history teachers examine representations of local artifacts alongside historical information and use evidence to justify interpretations. The global implication is that heritage educators can adapt this sequence to local artifacts and curricula, then assess students' evidence-based historical reasoning. The classroom comparison involved four intact classes, with one class in each

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

condition-by-cohort cell. This allocation structure is relevant when interpreting the observed contrasts and informs the design of subsequent evaluations across more classes.

Further studies can compare instructional conditions that vary AR access while retaining comparable historical tasks, document implementation across classes, and include delayed and written-response assessment. Such work can examine the module's contribution to historical reasoning in additional learning contexts.

ACKNOWLEDGMENT

The authors would like to sincerely thank Universitas Tadulako for supporting this research. Special thanks are extended to the faculty members of the History Education Department and all students who participated in the study. Appreciation is also given to the expert validators for their valuable contributions during the instrument and module validation processes. This study was made possible through collaborative efforts across academic, technical, and cultural domains. Universitas Tadulako supported this research through the Internal Research Fund (Grant Number: 2653/UN28/KU/2024).

REFERENCES

Abdul-Jabbar, S. N., & Alwehab, A. A. (2023). Augmented reality's role in highlighting historical heritage: Kirkuk Citadel. *Journal of Engineering*, *29*(12), 104–122. DOI: https://doi.org/10.31026/j.eng.2023.12.07

Afriani, Y., Agustiningsih, N., & Karmela, S. H. (2022). Character education in learning history of the Diponegoro War material. *Journal of Research in Instructional*, *2*(1), 19–32. DOI: https://doi.org/10.30862/jri.v2i1.39

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement, 45*(1), 131–142. DOI: https://doi.org/10.1177/0013164485451012

Aji, K. A., & Purbojo, R. (2023). Learning through augmented reality: The Indonesia Open University experience. *Jurnal Teknologi Pendidikan: Jurnal Penelitian dan Pengembangan Pembelajaran*, *8*(3), 678. DOI: https://doi.org/10.33394/jtp.v8i3.7800

Asri, Z., Aisiah, & Aman. (2020). Historical empathy learning model for strengthening character education 2013 curriculum. In *Proceedings of the 3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019)*. DOI: DOI: https://doi.org/10.2991/assehr.k.200803.070

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented reality trends in education: A systematic review of research and applications. *Educational Technology & Society, 17*(4), 133–149. DOI: https://dugi-doc.udg.edu/handle/10256/17763

Bacca-Acosta, J., Navarro, S. M. B., Gesa, R. F., & Kinshuk, K. (2019). Framework for designing motivational augmented reality applications in vocational education and training. *Australasian Journal of Educational Technology*, *35*(3). https://doi.org/10.14742/ajet.4182

Blanco-Fernández, Y., López-Nores, M., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrer, M., & García-Duque, J. (2014). REENACT: A step forward in immersive learning about human history by augmented reality, role playing and social networking. *Expert Systems with Applications*, *41*(10), 4811–4828. DOI: https://doi.org/10.1016/j.eswa.2014.02.018

Boboc, R. G., Băutu, E., Gîrbacia, F., Popovici, N., & Popovici, D.-M. (2022). Augmented reality in cultural heritage: An overview of the last decade of applications. *Applied Sciences*, *12*(19), 9859. DOI: https://doi.org/10.3390/app12199859

Boboc, R. G., Duguleană, M., Voinea, G.-D., Postelnicu, C.-C., Popovici, D.-M., & Carrozzino, M. (2019). Mobile augmented reality for cultural heritage: Following the footsteps of Ovid among different locations in Europe. *Sustainability*, *11*(4), 1167. DOI: https://doi.org/10.3390/su11041167

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

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

Challenor, J., & Ma, M. (2019). A review of augmented reality applications for history education and heritage visualisation. *Multimodal Technologies and Interaction*, *3*(2), 39. DOI: https://doi.org/10.3390/mti3020039

Dewi, P. Y. A., & Primayana, K. H. (2019). Effect of learning module with setting contextual teaching and learning to increase the understanding of concepts. *International Journal of Education and Learning*, *1*(1), 19–26. DOI: https://doi.org/10.31763/ijele.v1i1.26

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector et al. (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 735–745). Springer. DOI: https://doi.org/10.1007/978-1-4614-3185-5_59

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report)*. American Philosophical Association.

Fahruddin, F., Salamah, S., Nurgiansah, T. H., Rosidi, M. I., Fitroh, I., Zulkarnain, Z., Darsono, D., Judijanto, L., & Saefudin, A. (2025). Integrating educational technology in history learning: Advancing cultural awareness and preservation in Indonesia. *International Journal of Information and Education Technology, 15*(9), 1952–1966. DOI: https://doi.org/10.18178/ijiet.2025.15.9.2395

Gonnermann-Müller, J., & Krüger, J. M. (2025). Unlocking augmented reality learning design based on evidence from empirical cognitive load studies: A systematic literature review. *Journal of Computer Assisted Learning, 41*(1), e13095. https://doi.org/10.1111/jcal.13095

Ichsan, M., Haris, F., & Qamaruzzaman, M. H. (2026). Augmented reality and information design: Visualizing historical knowledge for classroom learning. *InfoDesign, 23*(1), 1–16. DOI: https://doi.org/10.51358/id.v23i1.1278

Jung, T., & Dieck, M. C. T. (2017). Augmented reality, virtual reality and 3D printing for the co-creation of value for the visitor experience at cultural heritage places. *Journal of Place Management and Development*, *10*(2), 140–151. DOI: https://doi.org/10.1108/jpmd-07-2016-0045

Karima, E. M. (2024). Reducing cognitive bias of pre-service history teachers through augmented reality. *JOIV: International Journal on Informatics Visualization*, *8*(1), 520–526. DOI: https://doi.org/10.62527/joiv.8.1.2222

Keil, J., Zöllner, M., Becker, M., Wientapper, F., Engelke, T., & Wuest, H. (2011). The House of Olbrich: An augmented reality tour through architectural history. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities* (pp. 15–18). IEEE. DOI: https://doi.org/10.1109/ISMAR-AMH.2011.6093651

Ma'unah, S., Umamah, N., Sumardi, S., & Surya, R. A. (2018). The enhancement of attractiveness and effectiveness of history learning using local history interactive teaching material. *American Journal of Educational Research*, *6*(11), 1531–1538. DOI: https://doi.org/10.12691/education-6-11-11

Mendoza, R., Cabarcas, A., Puello-Beltran, J. J., Fabregat, R., & Navarro, S. M. B. (2020). Heritage education experience supported in augmented reality. *Revista Facultad de Ingeniería Universidad de Antioquia*, *99*, 52–62. DOI: https://doi.org/10.17533/udea.redin.20200582

Nadjamuddin, L., Rore, I., Mahmud, J., Zaky, M., Tadeko, N., & Nirmala, B. (2026). Effectiveness of hybrid discovery learning in enhancing students' comprehension and critical thinking skills. *International Journal of Interactive Mobile Technologies, 20*(1), 87–104. DOI: https://doi.org/10.3991/ijim.v20i01.57485

Neumüller, M., Reichinger, A., Rist, F., & Kern, C. (2014). 3D printing for cultural heritage: Preservation, accessibility, research and education. In M. Ioannides et al. (Eds.), *Cultural heritage and new technologies* (pp. 119–134). Springer. DOI: https://doi.org/10.1007/978-3-662-44630-0_9

Nurhasanah, A., & Maksum, A. (2025). Bridging cognition and ethics: Socio-emotional skills and digital history literacy in fostering critical thinking. *Social Sciences & Humanities Open, 12*, Article 101786.DOI: https://doi.org/10.1016/j.ssaho.2025.101786

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370

Pamungkas, J., Harun, H., & Manaf, A. (2023). A systematic review and meta-analysis group contrasts: Learning model based on local cultural wisdom and student learning outcomes. *International Journal of Instruction*, *16*(2), 53–70. https://doi.org/10.29333/iji.2023.1624a

Rini, T. A. (2020). Implementation of contextual learning based on lesson study model. *International Research-Based Education Journal*, *2*(1), 25–28. DOI: https://doi.org/10.17977/um043v2i1p25-28

Salmaini, S., Handrianto, C., Sartono, S., & Nor-Azhar, N. F. (2025). Augmented reality competencies and integration challenges among primary school teacher education students: A mixed-methods study. *Perspektivy Nauki i Obrazovania, 73*(1), 651–664. DOI: https://doi.org/10.32744/pse.2025.1.42

Sat, M., Ilhan, F., & Yukselturk, E. (2023). Comparison and evaluation of augmented reality technologies for designing interactive materials. *Education and Information Technologies, 28*(9), 11545–11567. DOI: https://doi.org/10.1007/s10639-023-11646-3

Shawarangkoon, S., Kucharoenthavorn, A., Wannapiroon, N., Chawarangkoon, C., Linthaluek, S., & Wannapiroon, P. (2024). Immersive living museum with augmented reality technology, integrated traditional art and digital art to enhance historical knowledge and learning attraction. *International Journal of Information and Education Technology, 14*(11), 1638–1651. DOI: https://doi.org/10.18178/ijiet.2024.14.11.2194

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ, 323*(7321), 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123

Copyright © 2026, *author*, e-ISSN 2442-8620, p-ISSN 0216-1370