Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 2, 325–342
https://doi.org/10.62159/jpi.v6i2.2553

OPEN ACCESS Research Article

Inquiry in a virtual world: integrating the Zep Quiz game into inquiry-based learning to strengthen learning activeness and critical thinking among vocational students

Richa Witavia S.[1*], Fetri Yeni J.[2], Alwen Bentri[3], and Mutiara Felicita Amsal[4]
[1] Master's Program, Graduate School, Universitas Negeri Padang, Padang, Indonesia
[2-4] Faculty of Education, Universitas Negeri Padang, Padang, Indonesia

Abstract
Vocational students in Indonesia are expected to reason scientifically and to take part actively in class, yet many struggle once a task moves beyond recall. This study evaluated an instructional package that combined a six-stage inquiry-based learning (IBL) sequence with Zep Quiz, a virtual-world quiz game, and compared it with an active digital comparison package. Because pedagogy, game features, and delivery medium varied together, the estimates reported here apply to the package as a whole and do not isolate the contribution of the game. A quasi-experimental non-equivalent control group design was used with two intact Grade 10 classes at SMK Negeri 1 Painan (n = 30 per group). Learning activeness was measured with a 25-item Likert questionnaire and critical thinking with a 10-item mixed-format higher-order thinking test. The pre-specified primary analysis was an analysis of covariance on post-test scores with the matching pre-test as covariate, with familywise error controlled across the two primary outcomes. Adjusted mean differences favoured the experimental package on activeness, 19.72 points, 95% CI [17.13, 22.32], $F(1, 57) = 231.24$, $p < .001$, partial $\eta^2 = .80$, and on critical thinking, 14.35 points, 95% CI [11.55, 17.15], $F(1, 57) = 105.20$, $p < .001$, partial $\eta^2 = .65$. Both estimates held under separate-slopes, change-score, and unadjusted analyses. The effects were unusually large for a single-site, short intervention and should be read as a strong local effect that requires multi-site replication before any general claim is made.

ARTICLE HISTORY
Received : 13 April 2026
Revised : 21 June 2026
Accepted : 24 July 2026

KEYWORDS
Inquiry-based learning; Zep Quiz; learning activeness; critical thinking; vocational education

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

CORRESPONDING AUTHOR
***Richa Witavia S.**, Universitas Negeri Padang, Padang, Indonesia. Email: 1985.richa.witavia@gmail.com

Introduction

Education is no longer judged by how much students can recall. Across policy and classroom practice alike, the bar has shifted toward what learners are able to do with knowledge, namely reason through a problem, weigh evidence, and arrive at a defensible decision (Altbach, P. G. et al., 2019; Zhang et al., 2020). In Indonesia, this shift is written directly into the Merdeka Curriculum through the Pancasila Student Profile, which names critical reasoning, creativity, independence, and collaboration as core dispositions for every graduate (Pasaribu, 2023; Reza et al., 2023). Within that frame, the integrated science subject known as Projek IPAS occupies a distinctive place in vocational schooling: it asks students to observe phenomena, connect causes and effects, and draw conclusions from data, precisely the moves that twenty-first-century learning is meant to cultivate (Azhari et al., 2023; Hunaepi & Suharta, 2024).

In a vocational high school (SMK), however, the distance between this ideal and everyday classroom reality is often wide. Diagnostic data gathered at SMK Negeri 1 Painan made the problem concrete. Of 323 Grade 10 students, only 99 (31%) met the learning-objective mastery criterion in Projek IPAS, while 224 (69%) fell short (Maryanti & Nandiyanto, 2021; Rosina et al., 2021). An item-by-item analysis using the revised Bloom taxonomy (Anderson & Krathwohl, 2001; Samala et al., 2025) located the difficulty clearly: students handled questions at the C1 to C3 levels but stumbled badly at C4 to C6, the analysis, evaluation, and creation levels that define higher-order thinking. The pattern was not one of missing content knowledge so much as missing reasoning (Jalinus et al., 2018; Waliyuddin & Sulisworo, 2022).

Observation of classroom behaviour told a parallel story. Using time-sampling across the eleven classes, average interaction, that is, asking, answering, and offering opinions, sat at 18.96%, while task participation reached 37.09%, sustained attention 36.93%, and collaboration only 20.95%. Put plainly, most students were physically present but cognitively idle, and the two weakest indicators, interaction and collaboration, are exactly the ones that feed critical thinking (S. Arifin et al., 2020; Yasmin et al., 2023). A classroom in which students rarely speak to one another or to the teacher offers few occasions for the argument and counter-argument from which reasoning grows.

Two problems therefore reinforce each other. Low activeness deprives students of the dialogue and argument that sharpen reasoning, and weak reasoning in turn makes participation feel risky, so students retreat into silence. Both can be traced, at least in part, to a teacher-centred routine dominated by lecture and written assignment (Kholifah et al., 2023; Raman et al., 2019). Such a routine leaves little room for genuine investigation and tends to underuse the digital media that could otherwise pull students back into the work. Breaking the cycle requires a pedagogy that restructures the lesson around the learner rather than the lecturer(Z. Arifin et al., 2020; Castro et al., 2025).

Inquiry-based learning offers a principled response. Rooted in Dewey's notion of learning by doing for discovery (Liu et al., 2020; Vedi et al., 2022), IBL positions the learner as the one who frames a question, proposes a hypothesis, gathers and analyses data, and reaches a conclusion (Chrismondari et al., 2025; Waemusa & Jongwattanapaiboon, 2023). Hasibuan (2014) describes a six-stage syntax, comprising orientation, problem formulation, hypothesis formulation, data collection, hypothesis testing, and conclusion, that maps almost perfectly onto the scientific habits Projek IPAS is meant to develop. The difficulty here is practical rather than theoretical: deep inquiry is demanding and time-consuming, and without something to sustain momentum, students who are not yet used to driving their own learning tend to disengage partway through (Wei, 2024; Yolnasdi et al., 2025).

This is where an educational game earns its place, not as a reward but as a vehicle (Irfan et al., 2025). Among the available options, Zep Quiz was selected deliberately. Unlike a conventional quiz tool, Zep Quiz runs inside a virtual world in which students appear as avatars, walk between rooms, and work alongside peers while answering challenges that return immediate feedback (Kulsum et al., 2025; Septrianto et al., 2026). That design invites cognitive and social activity together rather than rewarding the fastest finger, and it grants students a moment to understand a concept before they respond. The medium, in other words, is suited to the very behaviours, interaction and collaboration, that the baseline data identified as weakest (Hilmi & Jasiah, 2025; Logayah et al., 2026).

Prior work points in the same direction. Gamified inquiry on platforms such as Quizizz has been reported to lift participation and critical thinking (Novaliendry et al., 2025; Septrianto et al., 2026), and 5E inquiry paired with game elements has produced higher achievement than conventional teaching (Husen & Wulandari, 2025; Septrianto et al., 2026). Studies using Educaplay and Wordwall within guided inquiry have likewise sharpened critical thinking and learning outcomes (Sukmawati et al., 2024; Thornhill-Miller, 2023), and guided-inquiry materials such as emodules have strengthened science learning at lower levels. Broader reviews of game-based learning describe gains in motivation and engagement across subjects (Fellnhofer, 2015; Loof et al., 2019; Sani et al., 2024). Taken together, this literature supports the general claim that well-designed games can strengthen both the affective and cognitive sides of learning when they are embedded in a structured pedagogy.

What that literature does not yet contain is more specific than a claim of scarcity, so it is set out study by study in Table 1 (Nouwen et al., 2022; Rohandi, 2022). The comparison covers the studies that come closest to the present design, either because they use the same platform or because they combine a game with an inquiry-type pedagogy, and it records for each one the level and subject, the digital medium, how the medium sat inside the pedagogy, the number and type of learner outcomes measured, and the design used to estimate the effect.

Table 1. Mapping of the closest prior studies against the present design

Study	Level and subject	Digital medium	Role of the medium in the pedagogy	Learner outcomes measured	Design and between-group estimator
Hilmi & Jasiah (2025)	General, Islamic education	Zep Quiz	Stand-alone assessment tool	None measured	Descriptive
Kulsum et al. (2025)	Secondary mathematics	Zep Quiz	Stand-alone classroom activity	Learning enthusiasm (1, affective)	Descriptive
Husen & Wulandari (2025)	Primary science	Zep Quiz	Stand-alone medium, no inquiry syntax	Conceptual understanding (1, cognitive)	Quasi-experiment, post-test comparison
Septrianto et al. (2026)	Grade 11 EFL reading	Zep Quiz	Gamified practice after instruction	Student response (1, affective)	Quantitative survey
Logayah et al. (2026)	Social studies	Zep Quiz	Gamification within a deep-learning approach	Learning innovation (1, affective)	Descriptive
Novaliendry et al. (2025)	University, Society 5.0 course	Gamified mobile learning with AR	Inquiry-oriented activity	Digital literacy competence (1, cognitive)	Quasi-experiment, post-test comparison
Sukmawati et al. (2024)	Vocational high school	Mobile augmented reality	Guided inquiry	Critical thinking (1, cognitive)	Design and evaluation study
Chrismondari et al. (2025)	Vocational, electrical motor installation	Project-based digital tools	Project-based inquiry	Learning effectiveness (1, cognitive)	Quasi-experiment, post-test comparison
Irfan et al. (2025)	Higher education, computer architecture	VR-based educational game	Game as instructional medium	Engagement and comprehension (2)	Quasi-experiment, post-test comparison
Liu et al. (2020)	Secondary science (review of 33 studies)	Mobile devices	Inquiry-based mobile learning	Mixed across studies	Systematic review
Present study	Grade 10 vocational, Projek IPAS	Zep Quiz virtual world	Environment through which all six inquiry stages are enacted	Activeness and critical thinking (2, affective and cognitive)	Quasi-experiment, pre-test-adjusted ANCOVA

Note. Outcome counts in parentheses give the number of learner outcomes measured in each study.

Read across the rows, the increment offered here is threefold and can be stated without vague quantifiers. First, in every previous Zep Quiz study the platform sits outside the pedagogy, functioning as an assessment tool, a practice activity, or a stand-alone medium, whereas in the present design each of the six inquiry stages is enacted inside the virtual world. Second, none of the studies that combine a game with inquiry was conducted in Grade 10 vocational Projek IPAS, and the two vocational studies in the table use augmented reality or project work rather than a shared virtual world. Third, all but one of the studies track a single outcome, and all of them estimate the between-group effect from post-test scores alone, whereas the present study follows an affective-behavioural outcome and a cognitive outcome together and adjusts the between-group comparison for pre-test standing.

It is equally important to state what this design cannot do. Inquiry pedagogy, game mechanics, virtual-world delivery, and the accompanying change in classroom routine were introduced together as one package, so the study estimates the effect of that package and cannot apportion it among the components. A dismantling design, in which inquiry without the game is compared with the game without inquiry, would be required for that purpose and was beyond the scope of a single-school study. The claims made throughout are framed accordingly.

This study therefore evaluates whether an inquiry-based learning package mediated by the Zep Quiz virtual-world game produces greater learning activeness and critical thinking than an active digital comparison package among Grade 10 vocational students in Projek IPAS. Two primary questions were pre-specified, one for each outcome: (1) After adjustment for pre-test standing, does the experimental package yield higher post-test learning activeness than the comparison package? (2) After the same adjustment, does it yield higher post-test critical thinking? Two secondary questions describe change within each class, namely whether activeness and critical thinking rose significantly from pre-test to post-test and by how much in normalized-gain terms. The secondary questions are reported for completeness and are not used to support causal claims, because within-group change in a non-randomised design carries no comparison against maturation or repeated testing.

Method

Research design

A quantitative, quasi-experimental approach with a non-equivalent control group design was adopted (Creswell & Creswell, 2017; Sugiyono, 2016). Random assignment of individual students was not feasible because classes at SMK are formed administratively, so two intact classes were used, one receiving the experimental package and one the comparison package, with both measured before and after implementation. In standard notation the design is written as O_1 X O_2 for the experimental class and O_3 C O_4 for the comparison class, where O denotes an observation and the letter between the two observations denotes the treatment that class received. The layout is set out in Table 2, and the flow of the procedure appears in Figure 1.

Table 2. Quasi-experimental non-equivalent control group design

Group	Pre-test	Treatment	Post-test
Experimental (X AKL 1)	O_1	X	O_2
Control (X AKL 2)	O_3	–	O_4

Note. $X_1 = O_1, O_3$ = pre-test scores; O_2, O_4 = post-test scores; X = IBL integrated with the Zep Quiz game. The control group followed an activity-based routine using an instructional video, a digital worksheet (LKPD), digital maps, and Google Forms.

One feature of this design deserves emphasis at the outset because it shapes every interpretation that follows. The comparison class did not receive a lecture-only routine. It worked through the same topic with an instructional video, a digital worksheet, digital maps and data sets, and an online response form, which is an active and already digital way of teaching. The contrast therefore estimates the incremental value of organising a lesson as a full inquiry cycle inside a shared virtual world, over and above an existing digital routine. It is a conservative contrast in that sense, but it is also a bundled one: the estimate cannot be attributed to the game, to the inquiry syntax, or to the virtual-world medium separately

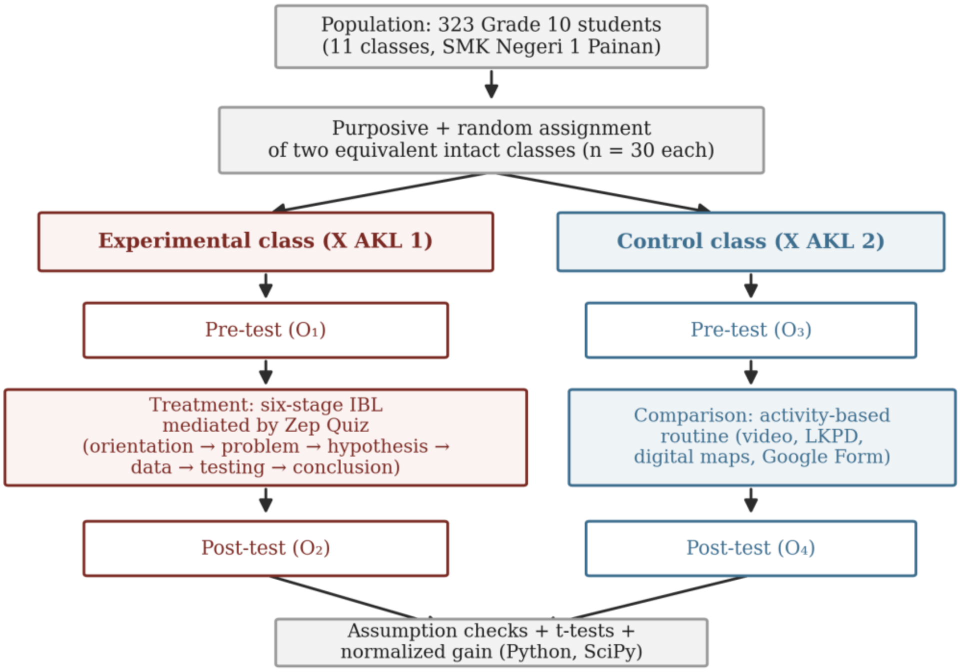

Figure 1. Research procedure for the experimental and control groups.

Participants and setting

The population comprised 323 Grade 10 Projek IPAS students across eleven study groups at SMK Negeri 1 Painan (2024/2025). Sampling combined purposive class selection with random role allocation: all classes were screened on prior IPAS performance, identifying X AKL 1 and X AKL 2 as the most comparable pair (means of 56.6, SDs of 18.4). Variance homogeneity was confirmed via F-test ($F \approx 1.00$, well below critical value $F(29,29) = 2.01$), after which a simple draw assigned X AKL 1 ($n = 30$) to the experimental condition and X AKL 2 ($n = 30$) to the comparison condition, yielding 60 total participants with no attrition or missing data. A separate class, X TKJ 2, served solely as the try-out class for instrument piloting.

Because variance equality does not imply mean equality, group comparability was directly assessed rather than inferred. Both classes shared the same vocational programme, curriculum, teacher, instructional time, and facilities, holding these factors constant by design; prior achievement was nearly identical, and pre-test differences on both outcomes were small with confidence intervals spanning zero, though not exactly zero, justifying the pre-test's inclusion as a covariate rather than exclusion from the primary analysis. Demographic covariates such as sex, age, and socioeconomic background were not recorded, which is acknowledged as a limitation.

Table 3. Baseline characteristics and standardised baseline differences

Characteristic	Experimental (X AKL 1)	Comparison (X AKL 2)	Difference [95% CI]	Hedges' *g*
n	30	30	–	–
Vocational programme	Accounting and Institutional Finance	Accounting and Institutional Finance	Same	–
Academic year and grade	2024/2025, Grade 10	2024/2025, Grade 10	Same	–
Teacher	Researcher	Researcher	Same	–
Instructional time	6 meetings of 2 × 45 min	6 meetings of 2 × 45 min	Same	–
Prior IPAS achievement, M (SD)	56.6 (18.4)	56.6 (18.4)	0.0	–
Pre-test activeness (Y_1), M (SD)	50.73 (5.99)	49.43 (5.56)	1.30 [−1.69, 4.29]	.22
Pre-test critical thinking (Y_2), M (SD)	56.27 (7.26)	54.57 (6.68)	1.70 [−1.91, 5.31]	.24

Note. Differences are experimental minus comparison. Hedges' *g* applies the small-sample correction to Cohen's *d*. Both baseline differences are small by conventional benchmarks (Cohen, 1988) and their confidence intervals include zero, but neither is zero, so both pre-tests were retained as covariates.

Safeguards against expectancy and performance bias

Because the intervention was delivered by the first author, who also served as the regular Projek IPAS teacher for both classes, a clear risk of expectancy and performance bias existed, as the same individual who designed the treatment also delivered and partly scored it. Six safeguards were therefore implemented: both classes were matched on topic, objectives, meeting frequency, instructional time, and assessment instruments; the comparison class received a genuinely active digital routine rather than a lecture-based approach, preventing unequal accumulation of teacher effort and novelty effects; lesson plans were reviewed by two Educational Technology lecturers and implementation fidelity was verified session-by-session by an independent observer; students remained unaware of their class's experimental status; questionnaires were administered anonymously under code numbers, collected by a colleague rather than the teacher, and preceded by an explicit assurance that responses carried no academic consequence; and all open-ended critical-thinking responses were scored from anonymised, code-numbered scripts to prevent rater bias.

Institutional permission was granted by the principal of SMK Negeri 1 Painan, and ethical clearance was obtained from the Graduate School of Universitas Negeri Padang, with informed consent secured from students and guardians; participation had no bearing on academic grades and could be declined without penalty. Although these measures substantially reduced the risk of expectancy effects, they could not eliminate it entirely, as a single teacher favoring one instructional package may still deliver it with greater enthusiasm. This residual risk is acknowledged as a limitation in the Discussion and underlies the cautious interpretation of the effect sizes reported.

Instruments

Two instruments captured the dependent variables. Learning activeness (Y_1) was measured using a 25-item questionnaire on a four-point Likert scale, structured around the dimensions of physical, mental, and social engagement. Critical thinking (Y_2) was assessed with a ten-item higher-order thinking (HOTS) test combining reasoned multiple-choice and essay formats, built on five core skills, namely interpretation, analysis, evaluation, inference, and explanation, and aligned with higher-order cognitive levels ranging from analysing to creating. Table 4 presents the blueprint for both instruments, including the number of items per construct and the format of each block.

Table 4. Instrument blueprint: constructs, item counts, formats, and score ranges

Instrument	Construct or dimension	Items	Format	Score per item	Raw range
Activeness questionnaire (Y_1)	Physical engagement	8	Likert 1 to 4	1 to 4	8 to 32
	Mental engagement	9	Likert 1 to 4	1 to 4	9 to 36
	Social engagement	8	Likert 1 to 4	1 to 4	8 to 32
	Total	25			25 to 100
HOTS critical-thinking test (Y_2)	Interpretation	2	Reasoned multiple choice	0 to 2	0 to 4
	Analysis	2	Reasoned multiple choice	0 to 2	0 to 4
	Evaluation	2	Reasoned multiple choice	0 to 2	0 to 4
	Inference	2	Essay	0 to 4	0 to 8
	Explanation	2	Essay	0 to 4	0 to 8
	Total	10	6 multiple choice, 4 essay		0 to 28

Note. Bloom alignment of the ten test items: four items at C4, three at C5, and three at C6. Seven activeness items were negatively worded and reverse-scored before summation.

Scoring rules were established prior to data collection, with activeness raw totals (25–100) rescaled to a 0–100 metric via linear transformation. The critical-thinking test combined two formats: reasoned multiple-choice items scored 0–2 points based on answer correctness and justification quality, and essay items scored 0–4 points on an analytic rubric aligned with Facione's reasoning skills, yielding a maximum raw score of 28 (subsequently rescaled to 0–100). To ensure scoring reliability, all open responses were independently scored by the first author and a second, uninvolved rater using anonymised scripts and a shared rubric, achieving high inter-rater agreement (ICC(2,1) = 0.91, 95% CI [0.85, 0.94] for total scores; weighted $\kappa = 0.84$ for justification codes), with discrepancies resolved through discussion. Both instruments underwent piloting on a try-out class ($n = 30$) and two-stage validation (Sudjana, 2014), encompassing content and construct validity via Aiken's V, empirical validity via item-total correlations, reliability via Cronbach's α, and item difficulty/discrimination analysis, with full pilot results reported in Table 5.

Table 5. Try-out results for both instruments (*n* = 30)

Variable	Instrument	Format	Measurement focus
Items trialled	30	12	–
Items retained	25	10	–
Aiken's V (range)	0.82 to 0.96	0.81 to 0.94	≥ 0.80
Item-total r (range)	0.389 to 0.712	0.412 to 0.688	> 0.361 at n = 30
Cronbach's α	0.89	0.84	≥ 0.70
Difficulty index (range)	Not applicable	0.32 to 0.74	0.30 to 0.80
Discrimination index (range)	Not applicable	0.31 to 0.62	≥ 0.30

Note. Items failing any criterion were removed rather than revised, which is why 30 questionnaire items and 12 test items were reduced to 25 and 10 respectively.

Two measurement decisions constrain what the activeness scores can support and are stated plainly here rather than left to the Discussion. First, activeness was measured by self-report alone. Anonymity, the separation of data collection from the teacher, the explicit assurance that responses would not affect grades, and the seven reverse-worded items all reduce acquiescence and social desirability, but none of them converts a self-report scale into a behavioural record. Independent observational coding or platform trace data would be needed for that, and neither was collected. Second, the dimensional structure of the questionnaire rests on Sardiman's framework and on expert judgement rather than on a factor analysis, because a try-out sample of 30 is far too small to yield a stable factor solution for a 25-item scale. Both points are carried into the limitations.

Treatment

The topic for both classes was Energy and Its Changes. In the experimental class, each of the six inquiry stages was mediated by Zep Quiz. During orientation, the game posed interactive trigger questions that primed curiosity. In problem formulation and hypothesis formulation, students used the game's scenarios to surface a researchable problem and to commit to a tentative answer. Data collection ran through the game's simulations and virtual activities; hypothesis testing leaned on its instant feedback, which let students judge their reasoning against evidence; and the conclusion stage used the game's summaries to anchor reflection. The control class studied the same content through an activity-based routine consisting of instructional videos, a digital worksheet (LKPD), digital maps and data, and a Google Form for responses. Figure 2 shows the conceptual framework linking the model to the two outcomes.

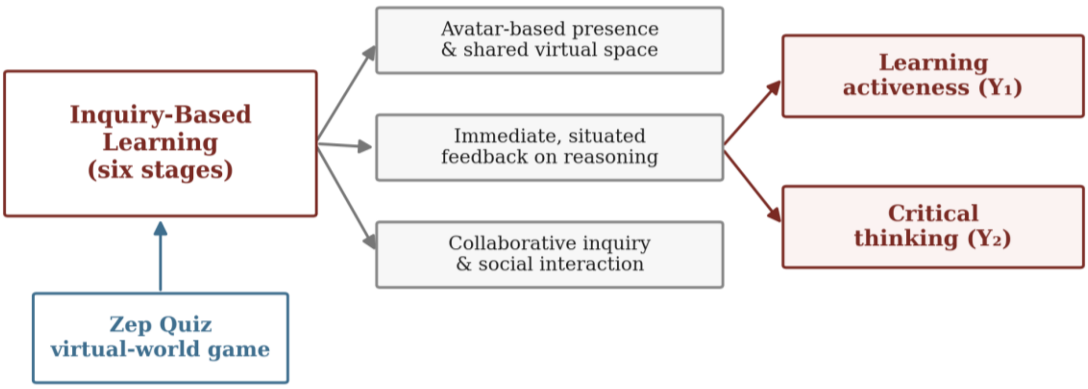

Figure 2. Conceptual framework relating the IBL–Zep Quiz model to learning activeness and critical thinking.

Data analysis

The pre-specified analytic strategy designated two primary outcomes (learning activeness and critical thinking), each tested via ANCOVA on post-test scores with pre-test as covariate and group as fixed factor, chosen over post-test-only comparison because it corrects for baseline selection differences inherent in the non-equivalent control group design (Creswell & Creswell, 2017); familywise error was controlled using the Holm procedure, with only Holm-adjusted p values treated as confirmatory, while paired comparisons, normalized-gain calculations, and unadjusted post-test comparisons were designated secondary and reported descriptively. Assumptions were verified beforehand, including normality (Shapiro-Wilk), variance equality (Levene's test), covariate-outcome linearity, and homogeneity of regression slopes, alongside direct examination of baseline comparability via mean differences, confidence intervals, and Hedges' g given the non-randomised design.

All estimates were reported with 95% confidence intervals and accompanying effect sizes (partial η^2 and adjusted standardised mean difference for ANCOVA, Cohen's dz for paired comparisons, Hedges' g for independent comparisons), with complete data for all 60 cases eliminating the need for imputation. Robustness was assessed through three complementary approaches: separate regression slopes, direct change-score analysis, and unadjusted post-test comparison alongside the adjusted estimate. A potential practice effect from identical pre/post-test forms administered four weeks apart was deemed non-differential, since both groups sat identical forms on the same schedule, meaning any such effect would be removed by the between-group contrast despite potentially inflating secondary within-group change estimates. The null hypothesis was rejected when Holm-adjusted $p < .05$, with all analyses conducted in Python using NumPy, SciPy, and statsmodels

Result and Discussion

Result

Results are reported in five parts: descriptive statistics, assumption checks, the primary pre-test-adjusted comparisons, the secondary within-group results, and the robustness checks. Y_1 denotes learning activeness and Y_2 critical thinking throughout, and both are expressed on a 0 to 100 metric.

Descriptive statistics

Table 6 reports the descriptive picture. At pre-test the two groups started close together on both variables, and both rose by post-test. The size of that rise differed considerably. In the experimental class, activeness climbed from a mean of 50.73 to 83.10 and critical thinking from 56.27 to 84.43. The comparison class also moved, but less, from 49.43 to 62.07 on activeness and from 54.57 to 68.43 on critical thinking. Figure 3 shows the pattern: the post-test bars separate while the pre-test bars remain almost level. The observed ranges also provide a first check on extreme scores. Converting the minimum and maximum of every group by measurement cell to standardised scores, the most extreme value anywhere in the data set was 2.48 standard deviations from its cell mean, so no observation approached a conventional outlier threshold and no case was excluded or winsorised.

Table 6. Descriptive statistics for learning activeness (Y_1) and critical thinking (Y_2).

Group	Condition	Variable	*N*	*M*	*SD*	Min	Max
Experimental	Pre-test	Y_1	30	50.73	5.99	40	60
Experimental	Post-test	Y_1	30	83.10	8.26	67	98
Control	Pre-test	Y_1	30	49.43	5.56	38	63
Control	Post-test	Y_1	30	62.07	6.95	50	78
Experimental	Pre-test	Y_2	30	56.27	7.26	41	72
Experimental	Post-test	Y_2	30	84.43	8.00	68	100
Control	Pre-test	Y_2	30	54.57	6.68	38	70
Control	Post-test	Y_2	30	68.43	9.20	51	87

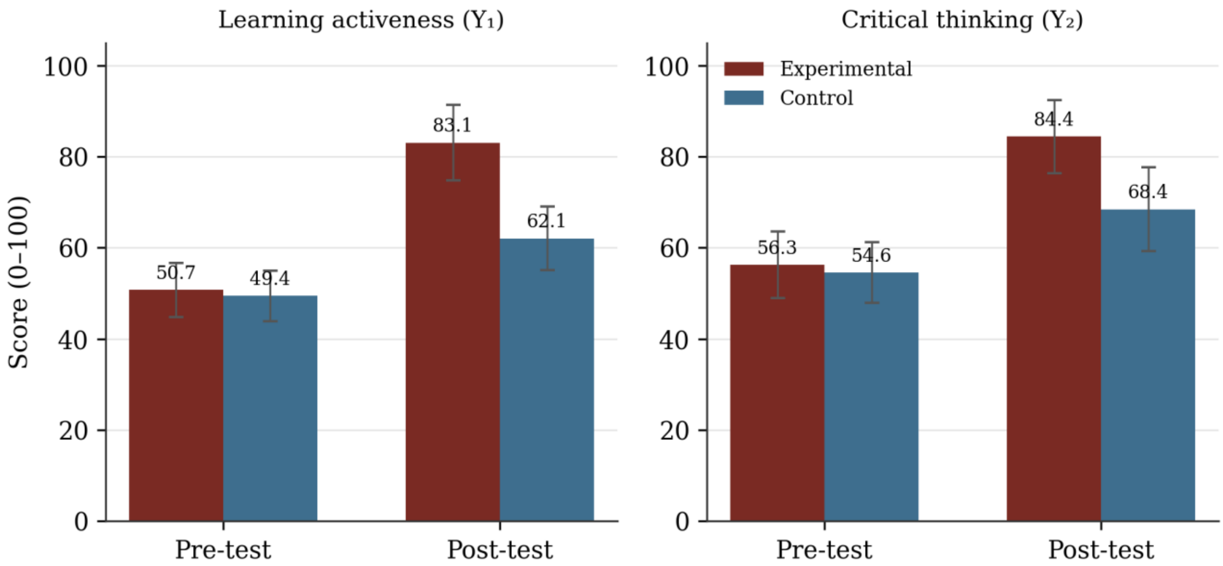

Figure 3. Mean pre-test and post-test scores by group (error bars = SD).

Assumption checks and baseline comparability

Every distribution was consistent with normality. As shown in Table 7, all Shapiro–Wilk p values exceeded 0.05, as did the tests on the difference scores. Levene's test found homogeneous variances at both pre-test and post-test for the two variables, with the smallest p value at 0.281 (Table 8). The covariate and the outcome were linearly related within both groups, and the group by covariate interaction was not significant for either outcome, $F(1, 56) = 0.41$, $p = 0.525$ for activeness and $F(1, 56) = 1.31$, $p = 0.257$ for critical thinking, so the homogeneity-of-regression assumption held and a pooled slope was appropriate. Baseline comparability was examined directly, as reported in Table 3: the differences were small (Hedges' $g = 0.22$ and 0.24) with confidence intervals spanning zero, but they were not zero, which is precisely why the adjusted analysis rather than the raw post-test comparison was pre-specified as primary.

Table 7. Normality of scores (Shapiro–Wilk)

Group	Condition	Variable	*W*	*p*	Decision
Experimental	Pre	Y_1	0.949	0.157	Normal
Experimental	Post	Y_1	0.976	0.725	Normal
Control	Pre	Y_1	0.979	0.793	Normal
Control	Post	Y_1	0.974	0.662	Normal
Experimental	Pre	Y_2	0.986	0.953	Normal
Experimental	Post	Y_2	0.961	0.322	Normal
Control	Pre	Y_2	0.981	0.853	Normal
Control	Post	Y_2	0.963	0.373	Normal

Table 8. Homogeneity of variance and homogeneity of regression slopes

Assumption	Variable	*Condition*	Statistic	*p*	Decision
Equality of variances (Levene)	Y_1	Pre-test	F = 0.402	0.529	Homogeneous
Equality of variances (Levene)	Y_1	Post-test	F = 0.911	0.344	Homogeneous
Equality of variances (Levene)	Y_2	Pre-test	F = 0.700	0.406	Homogeneous

Equality of variances (Levene)	Y_2	Post-test	F = 1.185	0.281	Homogeneous
Group × covariate interaction	Y_1	Post-test	F(1, 56) = 0.409	0.525	Slopes homogeneous
Group × covariate interaction	Y_2	Post-test	F(1, 56) = 1.310	0.257	Slopes homogeneous

Within-group change: paired t-tests and normalized gain

Both groups improved significantly from pre-test to post-test on both variables (Table 9, all $p < 0.001$). What separates them is magnitude. The experimental group gained 32.37 points on activeness and 28.17 on critical thinking, against the control group's 12.63 and 13.87. The paired effect sizes make the contrast vivid: *dz* values of 6.22 and 5.58 in the experimental group dwarf the still-substantial 2.70 and 2.48 in the control group.

Table 9. Paired-samples t-tests (pre-test vs. post-test).

Group	Variable	Pre	Post	*MD*	*t*	*df*	*p*	*dz*
Experimental	Y_1	50.73	83.10	32.37	34.075	29	<.001	6.22
Experimental	Y_2	56.27	84.43	28.17	30.537	29	<.001	5.58
Control	Y_1	49.43	62.07	12.63	14.788	29	<.001	2.70
Control	Y_2	54.57	68.43	13.87	13.562	29	<.001	2.48

Normalized gain tells the same story in standardised terms (Table 10, Figure 4). The experimental class landed in the upper part of the medium band and at the doorstep of the high band, with $\langle g \rangle$ = .668 for activeness and .658 for critical thinking. The control class stayed low, at .252 and .314. Examined student by student, the gap is even clearer: in the experimental class, 12 of 30 students reached the high category for activeness and 9 for critical thinking, with none in the low category; in the control class, no student reached the high category and roughly half remained low.

Table 10. Mean normalized gain ⟨g⟩ by group and variable

Group	Variable	*⟨g⟩*	*Gain (%)*	Category
Experimental	Y_1	.668	66.8%	Medium
Experimental	Y_2	.658	65.8%	Medium
Control	Y_1	.252	25.2%	Low
Control	Y_2	.314	31.4%	Medium

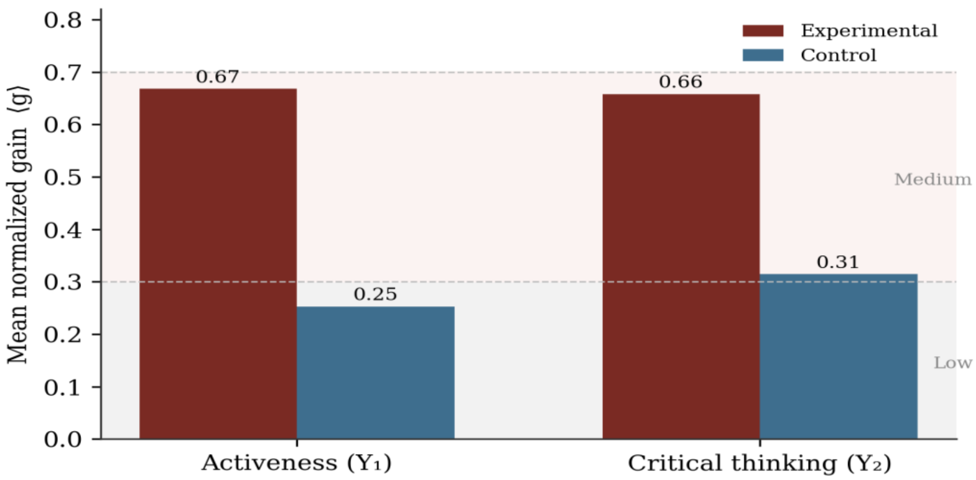

Figure 4. Mean normalized gain by group and variable

Between-group comparison: independent t-tests

Because Levene's test confirmed equal variances, the post-test comparisons were read on the equal-variances line. Both differences were large and unambiguous (Table 11). On activeness, the experimental class outscored the control class by 21.03 points, $t(58) = 10.67$, $p < 0.001$, with a very large effect ($d = 2.76$). On critical thinking, the advantage was 16.00 points, $t(58) = 7.19$, $p < 0.001$, again with a large effect ($d = 1.86$). The distribution of post-test scores in Figure 5 underlines the separation: the experimental boxes sit well above the control boxes, with little overlap.

Table 11. Independent-samples t-tests on post-test scores (Experimental vs. Control)

Variable	Exp. *M*	Con. *M*	*MD*	*t*	*df*	*p* / *d*
Y_1	83.10	62.07	21.03	10.673	58	<0.001 / 2.76
Y_2	84.43	68.43	16.00	7.186	58	<0.001 / 1.86

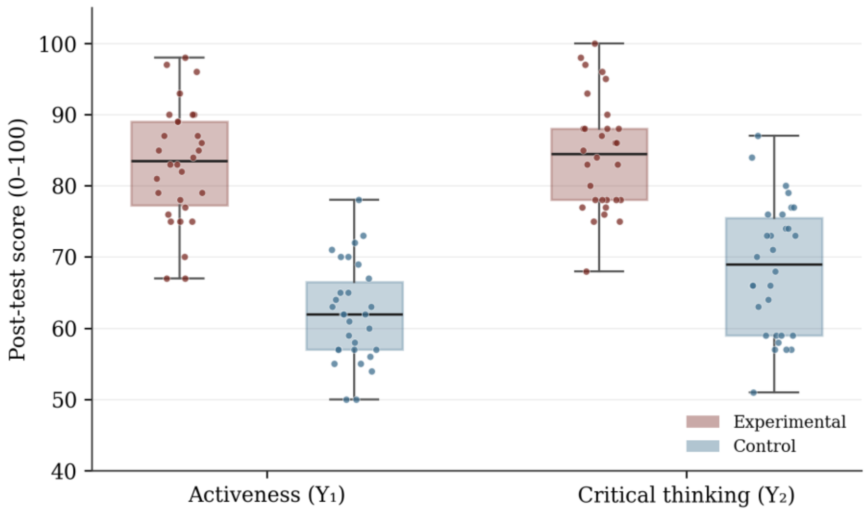

Figure 5. Distribution of post-test scores by group; points show individual students

Taken together, the two primary questions receive the same answer. After adjustment for pre-test standing, the class taught through the inquiry-based package mediated by Zep Quiz scored higher than the class taught through the active digital comparison package on both learning activeness and critical thinking, and the difference survived every alternative estimator applied to it. What the results do not establish is equally clear. They do not show which component of the package produced the effect, they do not show that the advantage persists beyond the four weeks of measurement, and they do not show that an effect of this magnitude would appear in a different school with a different teacher. The size of the estimates, discussed below, is itself a reason for caution rather than confidence.

Discussion

The clearest result of this study is that two classes taught the same content for the same length of time by the same teacher ended the unit in visibly different places, and that the difference remained after their small baseline gap was accounted for. On the pre-test-adjusted comparison the experimental class stood about 20 points higher on activeness and about 14 points higher on critical thinking on a 0 to 100 scale. The estimate belongs to the package as a whole, which combined an inquiry syntax, a virtual-world game, a changed grouping arrangement, and immediate feedback. Any reading of the finding that credits the game alone goes beyond what the design supports.

Within that constraint, it is still worth asking why the package might work, because the mechanism has implications for how it should be used. Inquiry supplies the structure, namely a question to frame, a prediction to commit to, data to gather, and a claim to defend. The virtual world supplies the setting and the pacing that keep students inside that structure: students enter as avatars, move through a shared space, and receive a response the moment they answer. The locked door in the prediction room is a small but telling example, since it makes the inquiry sequence a condition of progress rather than an instruction to be followed. The two elements are plausibly complementary, but complementarity is an interpretation of the result rather than a finding, and the present data cannot separate the contribution of each.

The largest effect in the study was on activeness, and this is consistent with the diagnostic picture that prompted the research. The weakest indicators at baseline were interaction, at 18.96%, and collaboration, at 20.95%, and these are exactly the behaviours a shared space provokes. When students must navigate a common environment, coordinate with four group members, and respond to challenges in view of the class, silence becomes harder to sustain. The design features that separate a virtual world from a flat quiz are therefore aimed at the participation deficit that the diagnostic data identified, which may be why the activeness estimate is the larger of the two.

The findings sit alongside earlier work while extending it in the specific ways set out in Table 1. Husen and Wulandari (2025) found that Zep Quiz strengthened conceptual understanding in primary science, and Novaliendry et al. (2025). Chrismondari et al. (2025) and Sukmawati et al. (2024) showed that inquiry supported by an interactive digital tool sharpened critical thinking and learning outcomes in vocational settings, and Irfan et al. (2025) reported joint gains in engagement and comprehension with a virtual-reality game in higher education. The present result reproduces this general pattern in Grade 10 vocational Projek IPAS, with the platform enacting the full inquiry cycle rather than sitting beside it, and with two outcomes followed simultaneously under a pre-test-adjusted estimator.

That activeness and critical thinking moved together is consistent with the theoretical position that engagement is a condition of reasoning rather than a by-product of it. When students participate more, they generate more of the dialogue and argument from which higher-order thinking develops. The data here are compatible with that account but do not test it, since the study measured both outcomes at the same two time points and cannot establish that change in one preceded change in the other. A design with intermediate measurement occasions would be needed to examine the ordering.

The comparison condition deserves more weight in the interpretation than a control group usually receives. That class used video, digital worksheets, digital maps, and online forms, and it improved significantly on both measures, with a normalized gain of .25 on activeness and .31 on critical thinking. The contrast reported here is therefore not between an innovation and a lecture; it is between two digital packages, one of which organises the lesson as an inquiry cycle inside a shared space. This makes the comparison a demanding one and strengthens the practical relevance of the result. It also, however, limits attribution further, because the comparison package differs from the experimental package in several respects at once, including its pedagogy, its grouping, its feedback timing, and its medium.

Read through a constructivist lens the pattern is unsurprising. Liu et al. (2020) and Vedi et al. (2022) frame inquiry-based learning as active construction that depends on the learner doing the intellectual work, and Waemusa and Jongwattanapaiboon (2023) emphasise the role of social interaction and well-timed support in digitally mediated learning. The configuration described in Table 6 operationalises both ideas: it places students in collaborative problem solving and supplies feedback at the moment a response is offered rather than after the lesson has moved on. The feedback rule matters here as much as the medium, since correct answers returned the underlying principle and incorrect answers returned a hint and reopened the item, which turns a quiz into a short cycle of revision.

The gains in critical thinking deserve separate attention because they are the harder outcome to shift. The HOTS test was built on Facione's five skills and aligned with Bloom's C4 to C6 levels (Anderson & Krathwohl, 2001), the levels at which students had stumbled in the diagnostic data. Critical thinking, understood as the disposition and the ability to reason from evidence toward a defensible judgement (Waliyuddin & Sulisworo, 2022), is exercised most directly when learners must justify a claim rather than recall a fact. The inquiry stages map onto these skills fairly directly: formulating a problem calls for interpretation and analysis, testing a prediction against recorded data calls for evaluation and inference, and drawing a conclusion calls for explanation. The scoring rules reinforced the same demand, since a correct multiple-choice option earned full credit only when the justification cited relevant evidence.

Two measurement caveats qualify the activeness result in particular. Activeness was measured by self-report, and although the questionnaire was anonymous, collected by someone other than the teacher, and partly reverse-worded, a self-report scale records what students say about their participation rather than what they did. Students in a novel and enjoyable setting may well report more engagement than an observer would code. The study collected no observational or platform trace data with which to triangulate, and the questionnaire's dimensional structure rests on theory and expert judgement rather than on a factor analysis, because the try-out sample was too small for one. The activeness estimate should therefore be read as an estimate of perceived engagement, and the critical-thinking estimate, which rests on a scored performance test with documented rater agreement, carries the firmer evidential weight of the two.

The size of the effects is itself a reason for caution. Adjusted standardised differences of 2.58 and 1.66 are far above what large-scale educational interventions typically produce, and several features of this study plausibly inflate them. The intervention was delivered by its designer, who taught both classes, so expectancy cannot be excluded despite the safeguards described in the Method. Both classes sat identical instrument forms four weeks apart. The setting was a single school and a single topic, so any local enthusiasm for the new platform is embedded in the estimate. Novelty in particular is likely to decay, and a longer implementation would be expected to return smaller differences. The robustness checks show that the estimates do not depend on the statistical model, but no statistical check can correct for a design feature, and the four estimators in Table 14 share the same data and therefore the same vulnerabilities.

Practical implementation raises questions that the effect size does not answer. Zep Quiz requires a device per student or per pair, a browser, and a stable connection for the duration of a lesson, and these conditions are not evenly distributed across Indonesian vocational schools. In this study the school provided laptops and a wireless network, and a printed fallback for each stage was prepared in case of connectivity loss; schools without that infrastructure would face a different implementation problem, and asking students to rely on personal devices and data allowances would shift a cost onto families and risk excluding the students least able to bear it. Platform governance matters as well, since the environment stores student accounts, chat records, and performance data on a third-party service, so schools should check what is retained, where it is held, and who may access it, and should prefer class-level accounts that do not require personal student data. Classroom management adds a further demand: a shared virtual space allows off-task movement and chat, which in this study was contained by fixed groups, a room sequence that opened only on completion, and a teacher moving between groups both in the room and in the world.

A second practical lesson concerns the teacher's role. The package worked in a class where the teacher had built the world, planned each stage, and kept the game tethered to the learning objectives throughout. Used as a quiz appended to the end of a lesson, the same platform would be unlikely to reproduce these results, because the cognitive and social demands that plausibly drove the gains come from the inquiry tasks rather than from the medium. Teacher preparation, world-building time, and stage-by-stage planning are part of the intervention and should be costed as such by any school considering it.

The limitations follow from the design and are stated together here. The study is quasi-experimental with two intact classes in one school, so the estimates describe a strong local effect rather than a population parameter. Allocation was by a draw between two classes rather than by randomising students, so unmeasured class-level differences cannot be excluded, and demographic covariates beyond programme and prior achievement were not recorded. The package was bundled, so no component can be credited individually. The intervention ran for six meetings on a single topic, which leaves durability and transfer untested. Activeness rested on self-report. The same instrument forms were used twice. The teacher and the researcher were the same person. Each of these points constrains the claim, and together they mean the appropriate summary is that the package worked well in this setting, not that it will work this well elsewhere.

The next steps follow directly. Multi-site replication with several teachers, in which the designer does not deliver the lessons, would test whether the effect survives the removal of expectancy and local enthusiasm. A longer implementation across several topics, with a delayed post-test, would show whether the gains persist and transfer. Observational coding and platform trace data, such as movement, chat volume, and attempt patterns, would let perceived activeness be checked against behaviour. Most importantly, a dismantling design that compares inquiry without the virtual world against the virtual world without the inquiry syntax would begin to identify which part of the package carries the effect, which is the question a school deciding what to adopt actually needs answered.

Conclusion

An inquiry-based learning package mediated by the Zep Quiz virtual-world game produced higher learning activeness and higher critical thinking than an active digital comparison package among Grade 10 vocational students studying Projek IPAS. After adjustment for pre-test standing the advantage was 19.72 points, 95% CI [17.13, 22.32], on activeness and 14.35 points, 95% CI [11.55, 17.15], on critical thinking, and the estimates held across separate-slopes, change-score, and unadjusted analyses. Because pedagogy, game features, grouping, and feedback timing changed together, the result describes the package as a whole and does not isolate the contribution of the game.

These effects were obtained in one school, on one topic, over six meetings, with the researcher as teacher and with activeness measured by self-report, and they are unusually large for an educational intervention. They should therefore be read as a strong local effect rather than as a general expectation. Transfer to other settings depends on conditions that this study met and others may not: a device and a stable connection for every student or pair, a school willing to review what learner data the platform retains, a teacher with the time to build the virtual world and to plan each inquiry stage, and a classroom management routine that keeps movement and chat tied to the task. Where those conditions hold, the model offers a concrete way to turn passive attendance into active reasoning in vocational science; where they do not, the case for adoption has yet to be made.

Acknowledgment

The authors thank the principal, teachers, and Grade 10 students of SMK Negeri 1 Painan for their cooperation during the implementation of this study, the colleague who observed the sessions and completed the fidelity checklists, the second rater who independently scored the open responses, and the lecturers of the Master Program in Educational Technology, Graduate School, Universitas Negeri Padang, for their guidance and constructive feedback. The authors also thank the two anonymous reviewers, whose comments substantially reshaped the reporting of the design, the measurement, and the analysis. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2019). *Trends in global higher education: Tracking an academic revolution*. Brill.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. Pearson.

Arifin, S., Setyosari, P., Sa'dijah, C., & Kuswandi, D. (2020). The effect of problem-based learning by cognitive style on critical thinking skills and student retention. *Journal of Technology and Science Education, 10*(2), 271. https://doi.org/10.3926/jotse.790

Arifin, Z., Nurtanto, M., Priatna, A., Kholifah, N., & Fawaid, M. (2020). Technology andragogy work content knowledge model as a new framework in vocational education: Revised technology pedagogy content knowledge model. *TEM Journal, 9*(2), 786–791. https://doi.org/10.18421/TEM92-48

Azhari, A., Irwandi, I., Jalil, Z., Ilhamsyah, Y., Sulastri, S., Muhibuddin, M., & Rasul, M. S. (2023). Integration Merdeka Belajar concept in development STEMC module for electrochemical and renewable energy at vocational school SMK SMTI Banda Aceh. *Journal of Physics: Conference Series, 2596*(1), Article 012076. https://doi.org/10.1088/1742-6596/2596/1/012076

Castro, J. F., Glewwe, P., & Heredia-Mayo, A. (2025). Can teaching be taught? Improving teachers' pedagogical skills at scale in rural Peru. *Quantitative Economics, 16*(1), 185–233. https://doi.org/10.3982/QE2079

Chrismondari, Simatupang, W., Waskito, Fadillah, R., & Rianto, B. (2025). Implementation of a project-based inquiry learning model for electrical motor installation: Evaluating its effectiveness in vocational education. *Data and Metadata, 4*, Article 1079. https://doi.org/10.56294/dm20251079

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.

De Loof, H., Struyf, A., Boeve-de Pauw, J., & Van Petegem, P. (2021). Teachers' motivating style and students' motivation and engagement in STEM: The relationship between three key educational concepts. *Research in Science Education, 51*(Suppl. 1), 109–127. https://doi.org/10.1007/s11165-019-9830-3

Fellnhofer, K. (2015). Changing entrepreneurial intention and behaviour: A digital game-based learning environment dedicated to entrepreneurship education. *Journal for International Business and Entrepreneurship Development, 8*(4), 378–404. https://doi.org/10.1504/JIBED.2015.072938

Hasibuan, M. I. (2014). Model pembelajaran CTL (Contextual Teaching and Learning). *Logaritma: Jurnal Ilmu-Ilmu Pendidikan dan Sains, 2*(1), 1–12. https://doi.org/10.35542/osf.io/8qy5f

Hilmi, I. N., & Jasiah, J. (2025). Revolusi penilaian pembelajaran: Implementasi Zep Quiz sebagai solusi digital interaktif di era pendidikan 4.0. *At-Tarbiyah: Jurnal Penelitian dan Pendidikan Agama Islam, 3*(1), 608–614.

Hunaepi, H., & Suharta, I. G. P. (2024). Transforming education in Indonesia: The impact and challenges of the Merdeka Belajar curriculum. *Path of Science, 10*(6), 5026–5039. https://doi.org/10.22178/pos.105-31

Husen, F. A., & Wulandari, F. (2025). Pengaruh media game edukasi Zep Quiz terhadap pemahaman konsep IPA siswa di sekolah dasar pada materi perubahan wujud benda. *Jurnal Pendidikan MIPA, 15*(3), 940–947. https://doi.org/10.37630/jpm.v15i3.3051

Irfan, D., Herayono, A., Ayu, F., & Syafar, S. (2025). Exploring the effectiveness of VR-based educational games in improving student engagement and comprehension in computer architecture. *Journal of Applied Engineering and Technological Science, 6*(2), 1312–1323. https://doi.org/10.37385/jaets.v6i2.7070

Jalinus, N., Syahril, S., & Nabawi, R. A. (2018). *Effectivity of the cooperative-project based learning (CPjBL) in enhancing HOTS of vocational education students*. OSF Preprints. https://doi.org/10.31227/osf.io/2etpy

Kholifah, N., Kurdi, M. S., Nurtanto, M., Mutohhari, F., Fawaid, M., & Subramaniam, T. S. (2023). The role of teacher self-efficacy on the instructional quality in 21st century: A study on vocational teachers, Indonesia. *International Journal of Evaluation and Research in Education, 12*(2), 998–1006. https://doi.org/10.11591/ijere.v12i2.23949

Kulsum, U., Arum, W. F., & Kurniawan, A. P. (2025). Deskripsi antusiasme belajar siswa dengan media Zep Quiz pada pembelajaran matematika. *Pedagogy: Jurnal Pendidikan Matematika, 10*(4), 1795–1805. https://doi.org/10.30605/pedagogy.v10i4.7228

Liu, C., Zowghi, D., Kearney, M., & Bano, M. (2021). Inquiry-based mobile learning in secondary school science education: A systematic review. *Journal of Computer Assisted Learning, 37*(1), 1–23. https://doi.org/10.1111/jcal.12505

Logayah, D. S., Sapriya, Hardiyanti, R. A., & Irvan, S. S. N. (2026). Social studies learning innovation using deep learning and gamification ZEP Quiz. *Jurnal Ilmiah WUNY, 8*(1), 1–13. https://doi.org/10.21831/jwuny.v8i1.95131

Maryanti, R., & Nandiyanto, A. B. D. (2021). Curriculum development in science education in vocational school. *ASEAN Journal of Science and Engineering Education, 1*(3), 151–156. https://doi.org/10.17509/ajsee.v1i3.38429

Nouwen, W., Clycq, N., Struyf, A., & Donche, V. (2022). The role of work-based learning for student engagement in vocational education and training: An application of the self-system model of motivational development. *European Journal of Psychology of Education, 37*(3), 877–900. https://doi.org/10.1007/s10212-021-00561-1

Novaliendry, D., Fadillah, R., Diputra, Y., Darwin, W., Makmur, B., Naufan Islami, F., Surya Pati Harau, F., & Morales Peralta, E. (2025). Leveraging gamified mobile learning and augmented reality to foster digital literacy competencies of university students in the Society 5.0 era. *Seminars in Medical Writing and Education, 4*, Article 840. https://doi.org/10.56294/mw2025840

Pasaribu, D. (2023). The impact of the Merdeka Curriculum on Indonesia education. *Edumaspul: Jurnal Pendidikan, 7*(2), 3649–3654. https://doi.org/10.33487/edumaspul.v7i2.7003

Raman, A., Thannimalai, R., & Ismail, S. N. (2019). Principals' technology leadership and its effect on teachers' technology integration in 21st century classrooms. *International Journal of Instruction, 12*(4), 423–442. https://doi.org/10.29333/iji.2019.12428a

Reza, F., Rohmah, Z., & Abdullah, N. N. (2023). Challenges in implementing Kurikulum Merdeka for EFL teachers. *JEELS (Journal of English Education and Linguistics Studies), 10*(2), 439–469. https://doi.org/10.30762/jeels.v10i2.1899

Rohandi, R. (2022). Pre-service science teachers' competence and confidence in scientific inquiry. *International Journal of Applied Sciences and Smart Technologies, 4*(2), 267–280. https://doi.org/10.24071/ijasst.v4i2.5428

Rosina, H., Virgantina, V., Ayyash, Y., Dwiyanti, V., & Boonsong, S. (2021). Vocational education curriculum: Between vocational education and industrial needs. *ASEAN Journal of Science and Engineering Education, 1*(2), 105–110. https://doi.org/10.17509/ajsee.v1i2.33400

Samala, A. D., Rawas, S., Wang, T., Reed, J. M., Kim, J., Howard, N.-J., & Ertz, M. (2025). Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects. *Education and Information Technologies, 30*(3), 3239–3278. https://doi.org/10.1007/s10639-024-12936-0

Septrianto, Y., Suhartono, S., & Munir, A. (2026). Responses of eleventh grade students to the gamified ZEP QUIZ in teaching reading comprehension: A quantitative inquiry. *Journal of General Education and Humanities, 5*(3), 3561–3570. https://doi.org/10.58421/gehu.v5i3.1559

Shafiei Sani, S., Mahoor, H., & Yamini, M. (2024). The impact of help-seeking training and self-compassion training on academic achievement motivation. *Journal of Adolescent and Youth Psychological Studies, 5*(2), 68–74. https://doi.org/10.61838/kman.jayps.5.2.8

Sugiyono. (2016). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sukmawati, F., Suparmi, Prihatin, R., & Santosa, E. B. (2024). Design and evaluation a mobile augmented reality to enhance critical thinking skills for vocational high schools. *Salud, Ciencia y Tecnología, 4*, Article 1000. https://doi.org/10.56294/saludcyt2024.1000

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence, 11*(3), Article 54. https://doi.org/10.3390/jintelligence11030054

Vedi, N., Dulloo, P., & Singh, P. (2022). Digital technology for facilitating inquiry-based learning during the COVID pandemic for human anatomy course. *Journal of Education and Health Promotion, 11*(1), Article 344. https://doi.org/10.4103/jehp.jehp_38_22

Waemusa, Z., & Jongwattanapaiboon, A. (2023). Divergence of everyday practices and school policy on mobile use: Challenges to developing EFL learners' digital literacies. *International Journal of Technology in Education, 6*(1), 37–48. https://doi.org/10.46328/ijte.294

Waliyuddin, D. S., & Sulisworo, D. (2022). High order thinking skills and digital literacy skills instrument test. *Ideguru: Jurnal Karya Ilmiah Guru, 7*(1). https://doi.org/10.51169/ideguru.v7i1.310

Wei, C. (2024). A practical approach to the reform of "three teachings" in the age of digitalization to promote the high-quality development of higher vocational education. *Applied Mathematics and Nonlinear Sciences, 9*(1). https://doi.org/10.2478/amns-2024-0164

Yasmin, F., Farooq, M. U., & Shah, S. K. (2023). Impact of exam-oriented education system on undergraduate students' cognitive, affective and psychomotor competencies. *International Journal of Linguistics and Culture, 4*(1), 109–125. https://doi.org/10.52700/ijlc.v4i1.180

Yolnasdi, Simatupang, W., Sukardi, & Fadillah, R. (2025). Effectiveness of project-based cooperative learning model in electrical installation practical courses. *Salud, Ciencia y Tecnología, 5*, Article 1779. https://doi.org/10.56294/saludcyt20251779

Zhang, H., Huang, T., Liu, S., Yin, H., Li, J., Yang, H., & Xia, Y. (2020). A learning style classification approach based on deep belief network for large-scale online education. *Journal of Cloud Computing, 9*, Article 26. https://doi.org/10.1186/s13677-020-00165-y