International Journal of Educational Qualitative Quantitative Research (IJE-QQR)
ISSN: 2962-9918 (Online)
Vol. 5, No. 1, pp. 194-207

AI-Scaffolded Participatory Reflection: Nurturing Elementary Students' Civic Literacy and Responsible Character

Kusman Rukmana[1], Regina Lichteria Panjaitan[2], Yulianti Agustin[3]
[1,2,3]Universitas Pendidikan Indonesia, Indonesia

Article Info

Article history:
Received: April 10, 2026
Revised: Juni 27, 2026
Accepted: July 15, 2026

DOI: 10.58418/ijeqqr.v5i1.219

How to cite this article:
Rukmana, K., Panjaitan, R. L., & Agustin, Y. (2026). AI-Scaffolded Participatory Reflection: Nurturing Elementary Students' Civic Literacy and Responsible Character. *International Journal of Educational Qualitative Quantitative Research*, *5*(1), 194–207. https://doi.org/10.58418/ijeqqr.v5i1.219

Read online:

Scan this QR code with your smart phone or mobile device to read online.

ABSTRACT

While traditional civic education often fails to bridge conceptual knowledge with active democratic behaviors, integrating Artificial Intelligence (AI) offers novel pathways for metacognitive scaffolding. This study evaluates the effectiveness of AI-scaffolded participatory reflection in nurturing primary school students' civic literacy and responsible character. Utilizing a quasi-experimental pretest-posttest control group design, the research involved fifth-grade students (N = 64) divided into an experimental group (n = 32) employing a generative AI chatbot to scaffold reflection and a control group (n = 32) undergoing conventional learning. Data from validated tests, questionnaires, observations, and interviews were analyzed via ANCOVA to control initial variances. Inferential results demonstrated that the AI-scaffolded approach significantly outperformed conventional instruction in enhancing civic literacy ($F(1, 61) = 39.56$, $p < 0.001$, $d = 1.21$) and developing responsible character ($F(1, 61) = 55.06$, $p < 0.001$, $d = 1.34$). The experimental group achieved higher normalized gains, accelerating task completion and behavioral self-regulation. Qualitative insights confirmed that real-time AI feedback facilitated autonomous error correction while mitigating automation bias. Theoretically, this study contributes an empirical model merging AI-mediated metacognition with dialogic pedagogy, transitioning citizenship education from passive memorization to active character internalization. Pragmatically, it provides a scalable framework supporting United Nations' Sustainable Development Goal (SDG) 4.

Keywords: AI-Scaffolded Reflection, Generative AI Chatbots, Civic Literacy, Responsible Character, Primary Education, Sustainable Development Goals

Corresponding Author:
Kusman Rukmana
Universitas Pendidikan Indonesia, Indonesia
Email: kusmanrukmana@upi.edu

1. INTRODUCTION

Primary education carries a foundational responsibility to facilitate not only the acquisition of civic knowledge but also the internalization of democratic values and behavioral self-regulation. Within the volatile socio-political landscape of the 21st century, fostering civic literacy, which is defined as the capacity to interpret socio-political realities, evaluate structural rules, and engage in constructive dialogue, has emerged as an urgent global educational imperative. Concurrently, developing a robust sense of personal and social responsibility in young learners, manifested through task commitment, adherence to communal norms, and responsiveness to constructive feedback, serves as the bedrock of sustainable democratic citizenship. Recent investigations by Álamo-Bolaños et al. (2024), Hee-Young & Ji-Young (2023) and Sousa & Moss (2023) indicate that civic competence developed during early childhood contributes significantly to later civic participation, social cohesion, and the sustainability of democratic societies. However, traditional civic education frameworks globally, and particularly within developing

nations undergoing rapid democratic transitions such as Indonesia, remain heavily constrained by teacher-centered, didactic, and textual pedagogies. Andreouli et al. (2026), Donbavand & Hoskins (2021), Jerome et al. (2024), Littenberg-Tobias (2021) and Tian & Tang (2025) observe that instructional approaches emphasizing passive knowledge transmission enable students to acquire conceptual understanding but are less effective in fostering the transfer of civic knowledge into authentic civic participation and everyday responsible behavior. In contrast, participatory, active, and real-world learning experiences more effectively connect civic concepts with practical civic competence.

To address this disconnect, contemporary pedagogical theories advocate participatory reflective learning, a dual-engine framework grounded in Deweyan systematic reflection (Greenberger et al., 2025; Greenberger & Chavez, 2026) and Freirean dialogic participation (Chen et al., 2025; Pietersen, 2022). This approach encourages learners to reconstruct authentic experiences through critical self-reflection, collaborative dialogue, and collective action, thereby fostering meaningful connections between conceptual understanding and real-world practice. Within this framework, feedback is conceptualized as a dialogic and reflective process rather than a punitive evaluation, enabling learners to interpret, internalize, and apply feedback to regulate their learning, refine their performance, and support continuous improvement (Derrington et al., 2026; Hill & West, 2022; McGuire, 2026). Nevertheless, implementing such high-level participatory reflection in primary education remains a significant pedagogical challenge due to young learners' developmental and cognitive characteristics. Consistent with constructivist perspectives informed by Piaget and Vygotsky (Oogarah-Pratap et al., 2025; Yıldız, 2025), fifth-grade students typically operate within the concrete operational stage, where learning is more effectively supported through tangible experiences than abstract reasoning. As a result, many learners have not yet developed the metacognitive capacity to engage in sustained self-reflection or to independently interpret and act upon complex qualitative feedback. From the perspective of Vygotsky's Zone of Proximal Development (Raslan, 2024; Utami & Yus, 2025), meaningful learning is achieved when appropriate instructional scaffolding enables students to accomplish tasks beyond their current level of independent performance. However, translating this principle into everyday classroom practice remains challenging, as teachers often face limited instructional time and classroom management constraints that restrict their ability to provide individualized, real-time scaffolding for every student. Consequently, the reflective learning process may become uneven, reducing opportunities for learners to fully benefit from personalized feedback and guided reflection.

The rapid advancement of Generative Artificial Intelligence (GenAI) and AI-powered educational chatbots offers unprecedented opportunities to address this instructional bottleneck by providing personalized, immediate, and adaptive feedback. Recent evidence suggests that AI-supported learning environments enable students to receive timely formative feedback, engage in self-directed learning, and develop higher-order thinking through personalized guidance and adaptive scaffolding (Badger et al., 2026; Okonkwo & Ade-Ibijola, 2021; Suo et al., 2026). When embedded within well-designed pedagogical practices, AI-driven chatbots can further promote critical thinking, problem-solving, and learning performance by facilitating reflective inquiry rather than merely supplying answers (Li et al., 2026; Q. Zhang et al., 2026).

Despite these promising developments, empirical evidence regarding the pedagogical role of GenAI as a metacognitive scaffold in primary citizenship education remains limited. Existing research has predominantly examined the use of generative AI in domains such as computer programming, language learning, writing support, and higher education, while comparatively few studies have investigated its application in civic literacy among primary school students. Moreover, technology-based interventions in character and citizenship education have largely emphasized content delivery and instructional support, with relatively limited attention to AI-mediated reflective dialogue as a mechanism for fostering metacognitive development. Consequently, important theoretical and empirical questions remain regarding how young learners engage with automated, real-time feedback and whether such interactions promote authentic self-reflection or instead increase the risk of automation bias, whereby learners accept AI-generated feedback uncritically without sufficient cognitive evaluation.

This research gap is particularly relevant in the context of contemporary educational reforms, including Indonesia's recent implementation of the Independent Curriculum, which emphasizes formative assessment, differentiated instruction, and the development of civic values as integral components of primary education. Drawing on Kolb's experiential learning cycle (Kleinheksel et al., 2023; Morris, 2020) as the overarching theoretical framework, the proposed integration of an AI chatbot into participatory reflective learning follows a four-stage iterative process. Students first engage in authentic civic experiences, then interact with the AI chatbot to facilitate scaffolded reflection, construct conceptual understanding through immediate formative feedback, and subsequently apply their learning by experimenting with responsible civic behaviors in classroom practice. Examining this instructional model in primary education offers empirical evidence on how AI-supported reflective learning can facilitate competency-based instruction and formative assessment while supporting the implementation of contemporary curriculum reforms. Beyond its relevance to the Indonesian context, the findings may also

contribute to the broader discussion on the pedagogical integration of generative AI in citizenship education and other competency-oriented learning environments.

Driven by these theoretical and empirical exigencies, this study evaluates the effectiveness of AI-scaffolded participatory reflection in nurturing primary school students' civic literacy and responsible character. To ensure a rigorous investigation, a quasi-experimental approach is adopted to evaluate the operational impact of the AI scaffolding framework while controlling for potential confounding baseline differences among young learners. This study provides two crucial scholarly contributions. Theoretically, it establishes a novel empirical model that synthesizes AI-mediated metacognitive scaffolding with dialogic citizenship pedagogy, demonstrating how adaptive educational technology can transition citizenship education from passive textual memorization to active character internalization. Practically, it delivers a validated instructional blueprint for global educators to design adaptive learning environments in elementary settings. Ultimately, this research offers a concrete mechanism to advance the United Nations' Sustainable Development Goal 4 by transitioning citizenship education from passive textual absorption to verifiable, technology-enhanced character internalization.

2. METHOD

This study deployed a quantitative approach with a quasi-experimental pretest-posttest control group design to evaluate the structural efficacy of the intervention without disrupting the natural ecology of the classroom environment. The independent variable was the instructional framework, contrasted between AI-scaffolded participatory reflection and conventional teacher-centered learning. The dependent variables comprised two primary constructs, which were civic literacy and responsible character, while students' initial pretest scores for each respective construct were designated as continuous covariates to suppress confounding baseline statistical noise. The research population targeted fifth-grade elementary school students in West Java, Indonesia. Utilizing a cluster random sampling technique to adhere to institutional constraints against individual student disruption, two intact classroom units from the same academic track were selected. The final sample consisted of N = 64 students, who were equally distributed into an experimental group (n = 32) and a control group (n = 32). In strict compliance with ethical protocols involving minor participants, formal institutional approval was secured from the school board, written parental informed consent was acquired prior to the study, and absolute data anonymity was maintained throughout the computational pipeline.

To preserve strict treatment fidelity and mitigate internal validity threats, the instructional experiment was executed over a structured timeline of eight distinct learning sessions for both groups, excluding the baseline pretest and final posttest administrations, with both cohorts taught by the same certified primary educator. The experimental group underwent a participatory reflective learning framework integrated with an AI-driven educational chatbot built on a generative large language model framework. The chatbot was locked via pre-programmed system prompts to act exclusively as an interactive metacognitive peer that guided students through Kolb's experiential learning cycle (Kleinheksel et al., 2023; Morris, 2020) by offering real-time formative feedback, prompting critical self-evaluation of civic issues, and forcing students to articulate their own moral arguments to explicitly prevent automation bias. Conversely, the control group received conventional instructional delivery covering identical civic topics through teacher-centered lectures, passive textbook reading, and traditional group worksheets, completely devoid of interactive digital scaffolding or structured reflective feedback loops.

Data collection relied on a dual-construct instrumentation approach administered before and after the intervention matrix. The specific psychometric properties, structural blueprints, and statistical validation parameters for both the Civic Literacy Test and the Responsible Character Questionnaire are systematically detailed in Table 1.

Table 1. Psychometric and Structural Specifications of Research Instruments

Instrument Name	Target Construct	Measured Indicators	Item Total & Format	Content Validity Index (CVI)	Reliability (Cronbach's α)
Civic Literacy Test	Objective Civic Knowledge & Analytical Skills	Rights and obligations identification, social problem analysis, and democratic value interpretation.	22 Items; Multiple Choice	0.88	α = 0.82
Responsible Character Questionnaire	Affective Internalization & Behavioral Self-Regulation	Task commitment, adherence to rules, and positive acceptance of corrective feedback.	18 Items; 4-Point Likert Scale	0.91	α = 0.86

Statistical analysis and data processing were conducted through an integrated computational workflow utilizing IBM SPSS Statistics 29.0 and Jamovi 2.4. The analytical pipeline was operationalized through three sequential phases, beginning with descriptive statistics to calculate the mean (±SD) alongside the Normalized Gain (N-Gain) index to chart relative learning growth. The second phase involved rigorous parametric assumption testing, where normal distribution was verified using the Shapiro-Wilk test, homoscedasticity of variances was evaluated via Levene's test, and the homogeneity of regression slopes was tested to validate the linear relationship between the covariates and the dependent variables. The final phase executed an Analysis of Covariance (ANCOVA), where the instructional treatment served as the fixed between-subject factor, the posttest scores served as the dependent variables, and the pretest scores were controlled as the continuous covariate. The absolute magnitude of the AI scaffolding treatment effect was evaluated through Partial Eta Squared and Cohen's d calculations to ensure both statistical significance and practical educational meaningfulness.

3. RESULTS

3.1. Data Screening and Parametric Assumptions Testing

Before executing the primary inferential statistical procedures to evaluate the efficacy of the instructional intervention, a rigorous data screening pipeline was conducted using IBM SPSS Statistics 29.0 and Jamovi 2.4 to verify that the empirical data strictly adhered to the mathematical assumptions required for parametric Analysis of Covariance (ANCOVA). The first diagnostic phase evaluated the assumption of univariate normality across both dependent constructs using the Shapiro-Wilk test, which is highly recommended for sample sizes under N = 100. The distributional analysis confirmed that both the pretest and posttest scores for civic literacy and responsible character across the experimental and control cohorts did not deviate significantly from a normal distribution, as all computed significance values comfortably exceeded the conventional alpha threshold of 0.05. To complement the normality metrics, Levene's test was executed on the posttest scores to evaluate the homoscedasticity of variances across the treatment groups. The results indicated that the variance of the error terms was perfectly homogeneous across the cohorts, thereby satisfying the second fundamental prerequisitefor covariance analysis. The comprehensive psychometric outputs for both the Shapiro-Wilk and Levene procedures are systematically arranged in Table 2.

Table 2. Normality and Homogeneity of Variance Metrics for Civic Literacy and Responsible Character

Construct and Group Matrix	Measurement Phase	Sample Size (n)	Shapiro-Wilk statistic (W)	Normality (p-value)	Levene's statistic (F)	Variance Homogeneity (p-value)
Civic Literacy						
Experimental	Pretest	32	0.951	0.156		
	Posttest	32	0.962	0.312	0.142	0.708
Control	Pretest	32	0.948	0.124		
	Posttest	32	0.957	0.231		
Responsible Character						
Experimental	Pretest	32	0.968	0.441		
	Posttest	32	0.955	0.201	0.284	0.596
Control	Pretest	32	0.944	0.095		
	Posttest	32	0.960	0.284		

Beyond the baseline distribution tests, the most critical and mathematically sensitive prerequisite for an entry-level ANCOVA design is the evaluation of the homogeneity of regression slopes. This procedure evaluates whether a significant statistical interaction exists between the continuous covariate (the pretest scores) and the categorical independent variable (the instructional group) in predicting the posttest outcomes. A significant interaction effect would indicate that the relationship between the covariate and the dependent variable varies as a function of the treatment group, which would fundamentally violate the linearity assumptions of ANCOVA and render the model invalid. To test this, custom general linear models were constructed in IBM SPSS 29.0 to extract the explicit interaction terms (Group x Pretest) for each construct.

The inferential summary presented in Table 3 reveals that the interaction effects for both civic literacy and responsible character were completely non-significant, with p-values falling well above the 0.05 threshold. This lack of statistical significance confirms that the regression lines for the experimental and control groups are parallel, thereby providing a robust mathematical justification to proceed with the main ANCOVA model to isolate the true operational impact of the AI scaffolding framework.

Table 3. Inferential Assessment of the Homogeneity of Regression Slopes

Dependent Construct Source	Evaluated Source of Variance	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Square (MS)	Computed F-value	Interaction (p-value)
Civic Literacy Posttest	Group x Civic Literacy Pretest	53.24	1	53.24	1.15	0.288
	Error Term	2747.30	60	45.79		
Responsible Character Posttest	Group x Responsibility Pretest	29.98	1	29.98	0.74	0.393
	Error Term	2413.34	60	40.22		

3.2. Baseline Equivalence Analysis

Following the verification of the underlying parametric assumptions, a comprehensive baseline equivalence analysis was executed to mitigate potential selection bias and protect the internal validity of the quasi-experimental design. The primary objective of this diagnostic phase was to empirically establish that the experimental and control cohorts commenced the study from statistically identical baselines regarding both targeted constructs, thereby neutralizing any suspicion that one group possessed an initial cognitive or behavioral advantage over the other.

Descriptive statistics were computed to map the baseline profile, revealing that the initial civic literacy and responsibility skills of students across both groups were at relatively equal levels prior to the implementation of the AI-scaffolded intervention. Specifically, on the pretest for civic literacy, the experimental cohort achieved a mean score of 62.50 ± 8.12, while the control cohort recorded a highly comparable mean score of 61.88 ± 7.95. For the responsible character construct, the baseline mean scores were logged at 65.16 ± 9.04 for the experimental group and 64.69 ± 8.78 for the control group.

To statistically confirm whether these minor descriptive variations carried any mathematical weight, an Independent Samples t-test was conducted on the raw pretest scores of both groups. The inferential metrics, which are systematically structured in Table 4, demonstrated that there was no statistically significant difference between the two groups on either measurement scale.

Table 4. Inferential Baseline Equivalence of Pretest Scores Between Cohorts

Target Construct Variable	Group Assignment	n	Mean Pretest Score (±SD)	df	Computed t-value	Significance (p-value)	Statistical Decision
Civic Literacy	Experimental	32	62.50±8.12	62	0.31	0.758	Non-Significant (p>0.05)
	Control	32	61.88±7.95				
Responsible Character	Experimental	32	65.16±9.04	62	0.21	0.834	Non-Significant (p>0.05)
	Control	32	64.69±8.78				

The inferential calculation for the civic literacy pretest yielded $t(62) = 0.31$, $p = 0.758$, while the evaluation of the responsible character pretest produced $t(62) = 0.21$, $p = 0.834$. Because both calculated significance values fell substantially above the conventional alpha threshold ($p > 0.05$), the null hypothesis of group divergence was confidently retained. These statistical outputs provide robust empirical verification that both student groups started from the exact same baseline line of aptitude and character disposition. Consequently, any statistically significant upward shifts observed in the subsequent posttest phase can be legitimately attributed to the operational impact of the AI-scaffolded participatory reflection treatment, rather than confounding fluctuations in the participants' initial competencies.

3.3. Hypothesis Testing 1: Efficacy of AI Scaffolding on Civic Literacy

To evaluate the first research hypothesis regarding the structural efficacy of the interventional model on students' civic literacy, an Analysis of Covariance (ANCOVA) was executed. This inferential procedure was critical to isolate the true operational impact of the generative AI-based educational chatbot by statistically filtering out the potential confounding variance of students' initial baseline competencies captured during the pretest phase. In this statistical arrangement, the posttest civic literacy scores served as the dependent variable, the instructional treatment functioned as the fixed between-subject factor, and the pretest civic literacy scores were designated as the continuous covariate. The comprehensive inferential outputs generated through the general linear model workflow are systematically structured in Table 5.

Table 5. ANCOVA Summary for Civic Literacy Achievement Between Cohorts

Source of Variance	Sum of Squares (SS)	df	Mean Square (MS)	Computed F-value	Significance (p-value)	Partial Eta Squared
Civic Literacy Pretest (Covariate)	412.18	1	412.18	8.98	0.004	0.128
Instructional Group (Treatment)	1816.42	1	1816.42	39.56	< 0.001	0.393
Error Term	2800.54	61	45.91			
Total Variance	412318.00	64				

The inferential metrics displayed in Table 5 reveal that after controlling for the initial baseline differences, the main effect of the instructional intervention was highly significant, $F(1, 61) = 39.56$, $p < 0.001$. Consequently, the null hypothesis stating that there is no significant difference in civic literacy between the groups was confidently rejected, confirming that the AI-scaffolded participatory reflection model yielded a statistically superior cognitive advancement compared to conventional instruction. The effect size magnitude, captured via Partial Eta Squared, reached 0.393, implying that the treatment alone accounted for 39.3% of the total variance in students' posttest civic literacy performance after extracting the covariate noise.

To further chart the efficiency of this cognitive growth, a Normalized Gain (N-Gain) analysis was calculated. The experimental cohort achieved an average N-Gain index of 0.58, placing their intellectual progress firmly within the moderate growth category. In stark contrast, the control cohort recorded an average N-Gain index of 0.31, positioning their academic development within the lower tier of the spectrum. To determine the practical educational meaningfulness of these statistical disparities for field implementation, Cohen's d was computed. The operational effect size yielded a massive value of $d = 1.21$, which sits substantially above the standard international high-impact threshold of 0.80. This robust psychometric profile demonstrates that the implementation of AI-scaffolded participatory reflection over the eight-meeting intervention timeline does not merely yield a technical statistical significance, but triggers a powerful, scalable, and practically transformative impact on primary students' capacity to interpret regulations and analyze civic issues effectively.

3.4. Hypothesis Testing 2: Efficacy of AI Scaffolding on Responsible Character

To address the second research hypothesis concerning the developmental shifts in students' responsible character, a secondary Analysis of Covariance (ANCOVA) was executed. This inferential stage was operationalized to evaluate the direct statistical impact of the instructional treatment while systematically filtering out the potential confounding influence of students' baseline behavioral tendencies captured during the pretest administration. Within this general linear model matrix, the posttest scores for responsible character functioned as the dependent variable, the instructional grouping served as the fixed between-subject factor, and the pretest responsibility scores were controlled as the continuous covariate. The comprehensive psychometric outputs and variance indices calculated for this behavioral construct are systematically arranged in Table 6.

Table 6. ANCOVA Summary for Responsible Character Development Between Cohorts

Source of Variance	Sum of Squares (SS)	df	Mean Square (MS)	Computed F-value	Significance (p-value)	Partial Eta Squared
Responsibility Pretest (Covariate)	502.64	1	502.64	12.55	<0.001	0.171
Instructional Group (Treatment)	2205.18	1	2205.18	55.06	< 0.001	0.474
Error Term	2443.32	61	40.05			
Total Variance	441208.00	64				

The inferential parameters displayed in Table 6 demonstrate that after eliminating the statistical noise of initial baseline variances, the main effect of the instructional intervention on students' responsible character was exceptionally profound, $F(1, 61) = 55.06$, $p < 0.001$. Consequently, the second null hypothesis was confidently rejected, providing robust statistical evidence that the AI-scaffolded participatory reflection framework triggers a highly superior advancement in student behavioral internalization compared to the conventional instructional setup. The calculated Partial Eta Squared value reached 0.474, which implies that the instructional treatment alone accounts for 47.4% of the total variance in posttest responsibility scores, demonstrating a very strong explanatory power within the model. This statistical shift substantiates the finding that students engaging with the interactive framework experienced substantial behavioral

improvements, particularly in terms of completing tasks on time, adhering to communal classroom regulations, and processing teacher feedback constructively.

To examine the specific growth trajectories and efficiency of this character development, a comparative Normalized Gain (N-Gain) analysis was operationalized. The experimental cohort achieved an average N-Gain index of 0.62, which positions their behavioral self-regulation firmly within the moderate growth category. Conversely, the control group exhibited a vastly lower average N-Gain index of 0.29, placing their development within the low-growth boundaries of the psychometric spectrum. To determine the practical educational meaningfulness of this divergence, Cohen's d was computed, yielding a massive effect size value of $d = 1.34$. This large operational effect size indicates that the integration of the generative AI chatbot within the participatory reflection cycle does not merely produce an artificial statistical significance, but exerts a massive, highly transferable impact on primary school learners. The real-time automated feedback loop provided by the chatbot effectively served as an immediate metacognitive scaffold, allowing young learners to autonomously evaluate their errors, fulfill group commitments, and internalize personal accountability more rapidly than those exposed to passive, un-scaffolded instruction.

3.5. Pedagogical Observations and Qualitative Triangulation

To complement and triangulate the robust inferential metrics obtained from the covariance analyses, qualitative data from structured classroom observations and semi-structured interviews were analyzed to capture the behavioral dynamics within the instructional ecosystem. The longitudinal observation of the experimental cohort revealed a profound evolution in student engagement and behavioral adaptation across the eight-session interventional matrix. In the initial phase, spanning the first two sessions, a notable portion of the fifth-grade students exhibited passive tendencies and were highly hesitant to express their opinions or interact deeply with the generative AI chatbot. This baseline hesitation was primarily driven by unfamiliarity with dialogic technology and a conventional fear of providing incorrect answers. However, as the students became accustomed to the structured routine of discussion, automated reflection, and follow-up activities from the third session onward, their learning engagement increased significantly from meeting to meeting. By the final sessions, the young learners became highly active, confident, and extraordinarily open to receiving and processing feedback. The generative AI educational chatbot successfully transitioned into a trusted conversational peer, motivating the students to explore complex civic issues critically and conduct autonomous self-evaluations of their learning progress.

This observed behavioral maturation was heavily supported by the real-time formative feedback delivered by the AI platform, which shifted the students' perception of error correction. Interviews with the participating primary educators indicated that this technological scaffolding helped students understand mistakes not as a form of punishment, but as an essential, constructive part of the learning process, thereby encouraging them to improve their work independently. Because the AI chatbot enabled students to obtain instantaneous responses regarding their inputs, it facilitated rapid, independent identification of behavioral and conceptual errors. To provide concrete empirical evidence of this metacognitive breakthrough, a brief transcript excerpt from a semi-structured interview with a fifth-grade student from the experimental group demonstrates this autonomous self-regulation: "*When I told the chatbot that I did not want to finish my class duty because I was tired, it did not give me a grade or get angry. Instead, it asked me what would happen to our school community if every citizen ignored their promises just because they were tired, which suddenly made me see that my action was wrong, so I immediately closed the chat and finished my team assignment.*"

This explicit student testimony provides a clear qualitative triangulation for the massive statistical effect size ($d = 1.34$) logged within the responsible character variable. It vividly demonstrates that the pre-programmed Socratic dialogue of the chatbot effectively circumvented automation bias by refusing to supply passive answers, forcing the student instead into an active cycle of moral reasoning and independent behavioral modification.

4. DISCUSSION

4.1. The Role of AI Metacognitive Scaffolding in Enhancing Civic Literacy

The remarkable statistical superiority observed in the experimental cohort concerning civic literacy, $F(1, 61) = 39.56$, $p < 0.001$, Partial Eta Squared = 0.393, can be theoretically illuminated through the lens of Vygotsky's Zone of Proximal Development (Raslan, 2024; Utami & Yus, 2025) and the mechanism of adaptive scaffolding (Puntambekar, 2022; van Nooijen et al., 2024). In this study, the generative AI-based educational chatbot functioned not merely as a static repository of digital information, but as an active digital More Knowledgeable Other (Davar et al., 2025; Tan et al., 2026; Tran et al., 2025). Primary school students, particularly fifth-graders, frequently encounter severe cognitive bottlenecks when required to translate abstract socio-political or democratic concepts into real-world applications. Traditional instructional settings usually fail to provide immediate, individualized interventions when these mental impasses occur due to the logistical and temporal constraints faced by classroom teachers. The AI chatbot successfully bypassed this operational barrier by providing real-time, personalized formative feedback

tailored to the specific conceptual trajectory of each student. When a student experienced an analytical standstill regarding rights, obligations, or social problem-solving, the AI stepped in as an interactive peer to offer intellectual scaffolding. Instead of oversimplifying the problem or presenting ready-made answers that might induce automation bias, the digital More Knowledgeable Other utilized Socratic questioning (Overholser & Beale, 2023; Tran et al., 2025; Vittorio et al., 2022) to prompt deeper critical inquiry, thereby guiding the learner from their current level of independent performance to a higher level of potential development. This systematic interaction explains the moderate growth manifested in the experimental group's Normalized Gain index of 0.58 compared to the low growth of the control group.

Furthermore, situating these findings within the broader international literature highlights the theoretical contribution of the proposed model. Contemporary research has consistently questioned instructional approaches to citizenship education that rely primarily on knowledge transmission and rote learning, as such approaches are less effective in fostering meaningful civic competence and active citizenship (Donbavand & Hoskins, 2021; Jerome et al., 2024). When civic learning is limited to unidirectional instruction, students may acquire conceptual knowledge but often struggle to transfer that knowledge into authentic civic reasoning, participation, and responsible action in everyday contexts (Álamo-Bolaños et al., 2024; Tian & Tang, 2025; White et al., 2023). Prior research on civic competence has predominantly focused on adolescents and older learners, with comparatively limited empirical attention devoted to primary school students. Consequently, evidence regarding the development of civic competence during the elementary years remains relatively scarce, despite growing recognition that citizenship learning begins in early childhood and is shaped through everyday social interactions and classroom experiences (Fitzgerald et al., 2021; Krijnen et al., 2024). This study successfully challenges that empirical paradigm. By introducing dialogic AI as a scaffold, this research proves that primary school students can actively engage with, interpret, and evaluate democratic values effectively if given a highly responsive, non-punitive digital environment. The substantial effect size observed for this variable ($d = 1.21$) reinforces contemporary evidence that effective citizenship education extends beyond the transmission of civic knowledge to integrate reflective inquiry, authentic participation, and opportunities for active civic engagement. Rather than fostering compliance with institutional rules alone, meaningful civic learning requires students to critically interpret, apply, and enact democratic values through participatory experiences.

4.2. Behavioral Self-Regulation and the Internalization of Responsible Character

Beyond cognitive gains, the profound behavioral shift observed in the experimental cohort's responsible character, marked by an impressive covariance metric of $F(1, 61) = 55.06$, $p < 0.001$, Partial Eta Squared = 0.474, can be comprehensively decoded through Lickona's classic Character Education Theory. According to this theoretical paradigm, authentic character internalization requires a harmonious alignment of three interconnected psychological dimensions, which are moral knowing, moral feeling, and moral action (Anisa et al., 2026; Arif et al., 2023; Prabowo, 2024). The interactive architecture of the AI-scaffolded participatory reflection model directly operationalized this tripartite cycle. First, the chatbot facilitated moral knowing by prompting students to conceptually articulate their duties and understand classroom rules. Second, through its automated, non-judgmental questioning, the AI stimulated moral feeling, forcing young learners to experience an internal affective awareness of how their personal negligence or procrastination would negatively impact their peer community. Third, this psychological friction was immediately translated into moral action, as observed in the students' accelerated task completion, disciplined adherence to communal guidelines, and consistent fulfillment of group commitments. This complete alignment successfully drove the experimental group to achieve a moderate growth index of 0.62, while the control group languished in the low-growth tier of 0.29.

The psychological catalyst for this rapid character development lies primarily in the unique, non-threatening nature of real-time automated feedback. A persistent challenge in elementary character education is ensuring that corrective feedback is interpreted as developmental guidance rather than punitive judgment. When students perceive teacher corrections as criticism or punishment, they are more likely to respond with negative emotions and superficial compliance than with reflective engagement and authentic behavioral change (Bai et al., 2026; Singh & Mueller, 2021; Yu & Yang, 2021). Conversely, the generative AI chatbot provided a hyper-personalized, private conversational space where feedback was strictly objective and formative. Because the machine is devoid of human emotional bias, impatience, or punitive authority, the students' affective filters were drastically lowered (DiBerardino & Stark, 2023; Jonas et al., 2025). Rather than experiencing corrective feedback as a potentially threatening social evaluation, students interacted with the AI chatbot in a more private and supportive learning environment. This individualized form of digital scaffolding encouraged learners to engage with feedback as a resource for self-improvement, strengthening self-regulated learning, intrinsic motivation, and reflective engagement rather than merely complying with external expectations (Guan et al., 2025; Ng et al., 2024). Empowered by timely, personalized, and metacognitive feedback, learners became more willing to recognize mistakes as opportunities for learning, supporting greater cognitive autonomy and more active self-correction (Ortega-

Ochoa et al., 2024; Yin et al., 2024). This profound transformation explains the massive operational effect size of d = 1.34, proving that AI-mediated scaffolding fosters an enduring internal locus of responsibility that far outperforms conventional, teacher-centered disciplinary measures.

4.3. Mitigating Automation Bias in Young Learners: A Dialogic Paradigm

A critical dimension of this study that represents its core theoretical novelty within the field of Artificial Intelligence in Education (AIEd) is the systematic mitigation of automation bias among young learners. Recent educational technology research has identified automation bias as an emerging challenge associated with the use of Large Language Models (LLMs). In educational settings, automation bias refers to learners' tendency to rely uncritically on AI-generated recommendations or responses, potentially reducing reflective thinking, cognitive engagement, and independent evaluation unless appropriate pedagogical scaffolding is provided (Lee et al., 2024; Yan et al., 2024). This cognitive vulnerability may be particularly salient among elementary school students, whose reasoning abilities are still developing and whose epistemic trust is becoming more sophisticated. As a result, they may be more susceptible to over-relying on AI-generated information when appropriate opportunities for critical evaluation and guided reflection are absent (Hoehl et al., 2024; Schmid et al., 2024). However, the qualitative observations and inferential evidence obtained in this study suggest that the anticipated cognitive dependency did not emerge. Instead, the instructional design appeared to mitigate the risk of automation bias by intentionally configuring the generative AI chatbot as a metacognitive scaffold rather than an answer-generation system. Consistent with recent recommendations for the educational use of large language models, the chatbot's system prompts were deliberately constrained to support Socratic dialogue, reflective inquiry, self-evaluation, and structured formative feedback rather than providing direct solutions (Ng et al., 2024; Yan et al., 2024). Consequently, when students presented civic dilemmas or behavioral questions, the chatbot responded with reflective prompts that encouraged learners to examine their own reasoning, consider alternative perspectives, and justify their decisions instead of passively accepting AI-generated answers. Such dialogic interactions are consistent with evidence showing that reflective questioning and metacognitive scaffolding promote deeper cognitive engagement and self-regulated learning (Guan et al., 2025; Wang et al., 2025). In this way, the chatbot functioned as a cognitive partner that supported students' reflective thinking while preserving learner agency, rather than replacing independent reasoning.

By neutralizing the risk of cognitive passivity through Socratic constraints, the AI chatbot effectively operationalized the experiential learning cycle developed by Kolb into a functional classroom mechanism. The four stages of Kolb’s learning cycle (Kleinheksel et al., 2023; Morris, 2020), which places concrete experience, reflection, conceptualization, and active experimentation as a continuous loop, were executed within a highly responsive instructional framework. First, students encountered or recalled a concrete civic conflict within their school environment, establishing the concrete experience stage of the cycle. Second, when students interacted with the educational chatbot, the machine's refusal to provide direct solutions forced them into deep, structured self-interrogation, which successfully fulfilled the reflective observation stage by helping them identify conceptual and behavioral errors independently. Third, the iterative dialogue enabled young learners to autonomously synthesize automated feedback and conceptualize generalized ethical concepts regarding mutual cooperation, rights, and communal rules, thereby driving the abstract conceptualization stage. Fourth, because the AI scaffolding was designed to conclude by prompting the user for an actionable commitment, students immediately transitioned into the active experimentation stage by modifying their actual behavioral choices, completing assignments on time, adhering to group regulations, and accepting feedback positively within the classroom. This empirical alignment proves that when generative AI is pedagogically restricted to deliver interactive, reflective feedback rather than direct content, it can serve as a powerful catalyst for human critical thinking and self-regulation. Consequently, this study shifts the global AIEd narrative from utilizing technology as a mere delivery vehicle to positioning AI as an active, metacognitive scaffold capable of transforming passive textual memorization into verifiable character internalization.

4.4. Pedagogical Implications for Competency-Based Curriculum and SDG 4

The remarkable statistical and qualitative improvements associated with the AI-scaffolded participatory reflection framework have important implications for contemporary educational reform. In Indonesia, the Independent Curriculum emphasizes competency-based learning through formative assessment, differentiated instruction, and the development of the Pancasila Student Profile (Sari et al., 2025). However, implementing these pedagogical principles in everyday classroom practice remains challenging because teachers frequently face time constraints and heavy instructional workloads that limit opportunities to provide individualized formative feedback (Derrington et al., 2026; Lei & Lei, 2026; Singh & Mueller, 2021). Within this context, AI-supported reflective scaffolding offers a practical mechanism for extending personalized feedback without replacing the pedagogical role of teachers. These structural constraints often compel teachers to rely on more generalized assessment practices, limiting opportunities to provide continuous, individualized feedback that supports students' character development and competency growth (Van Orman et al., 2025; H. Zhang et al., 2024). The pedagogical architecture validated

in this study directly resolves this policy-practice paradox. By deploying generative AI chatbots locked as Socratic dialogue partners, schools can effectively decentralize the feedback loop. The technology automates the preliminary scaffolding process, allowing each child to receive immediate, private, and tailored metacognitive guidance while simultaneously relieving teachers from crushing cognitive overloads. Consequently, this model offers an empirical blueprint for realizing the core philosophy of student-centered, differentiated instruction within resource-constrained educational settings.

On a global scale, the systematic integration of this low-cost digital architecture has important implications for achieving the United Nations' Sustainable Development Goal (SDG) 4, which promotes inclusive, equitable, and quality education for all. Although immersive technologies such as virtual reality and other advanced digital learning environments have demonstrated considerable educational potential, their widespread implementation often requires substantial financial investment, specialized hardware, and reliable digital infrastructure. These requirements present significant barriers for many schools in developing regions and may exacerbate existing educational inequalities if equitable access is not addressed (Ali et al., 2026; Luo et al., 2021; Matjie et al., 2026). This research presents an alternative and more equitable instructional paradigm by demonstrating that advanced metacognitive scaffolding can be achieved through lightweight, software-based generative AI systems operating on existing school devices. Rather than serving merely as a content presentation tool, the chatbot functioned as a dialogic and metacognitive scaffold that encouraged reflective inquiry, critical thinking, and socially responsible decision-making. These characteristics closely align with the objectives of SDG Target 4.7, which emphasizes education for sustainable development and global citizenship through transformative learning processes (Holst et al., 2024; Leite, 2022; Santamaría-Cárdaba et al., 2024). The findings further support a complementary human–AI instructional ecosystem in which adaptive technology extends teachers' capacity to provide personalized formative support while preserving their central pedagogical role, thereby accommodating diverse learning needs within resource-constrained classrooms. Collectively, this study contributes to the growing body of AIEd research by suggesting that low-cost generative AI can offer a scalable approach to fostering reflective learning, character development, and digital literacy through pedagogically guided interaction, particularly in educational contexts where access to advanced technological infrastructure remains limited.

4.5. Limitations and Future Research Directions

Notwithstanding the compelling inferential and qualitative breakthroughs documented in this study, several inherent limitations must be transparently acknowledged to maintain scholarly rigor and contextualize the generalizability of the findings. First, the empirical evaluation was restricted to a sample size of N = 64 students drawn from a single geographic region in West Java, Indonesia, which inevitably constrains the immediate extrapolation of the statistical data to highly diverse socioeconomic or varied institutional landscapes. Second, the instructional intervention was operationalized across a compact timeline of eight learning sessions. While this operational duration proved highly sufficient to capture immediate cognitive adjustments in civic literacy and initial behavioral self-regulation, it remains methodologically limited in guaranteeing the long-term retention of responsible character traits and civic habits once the interactive digital scaffolding is permanently removed from the daily classroom ecosystem.

To address these empirical boundaries and expand the scholarly horizon of Artificial Intelligence in Education (AIEd), subsequent investigations must deliberately broaden both the methodological architecture and participant demographics. Future research should transition toward a longitudinal mixed-methods design, combining robust time-series quantitative tracking with deep qualitative observations to evaluate the endurance of moral and character development over multiple academic semesters. Furthermore, deploying this AI-scaffolded framework across a more massive, cross-cultural sample involving diverse regional and international student cohorts would provide invaluable empirical insights into how varying cultural backgrounds interact with automated Socratic feedback loops. Such expanded scholarly efforts will be pivotal in establishing a truly scalable, universally generalizable, and cross-culturally validated instructional blueprint for technology-enhanced citizenship education.

5. CONCLUSION

In conclusion, this study provides definitive empirical evidence that the implementation of AI-scaffolded participatory reflection is highly effective in nurturing both civic literacy and responsible character among primary school students. The inferential metrics demonstrate that integrating a generative AI educational chatbot as a Socratic dialogue partner significantly outperforms conventional, teacher-centered instructional methods. By filtering out initial baseline abilities, the framework achieved substantial cognitive and behavioral growth, moving students from low-growth tiers to moderate growth categories with exceptionally strong operational effect sizes. This success is driven by the synergistic alignment of Vygotsky’s Zone of Proximal Development and Kolb’s experiential learning cycle, wherein the AI chatbot functions as a private, non-punitive digital More Knowledgeable Other. The platform successfully guides

young learners through a cycle of concrete experience, structured reflection, and abstract conceptualization, which ultimately manifests as active classroom experimentation and responsible self-regulation.

The core scholarly value of this research is established through its dual-level contributions to educational theory and macro policy execution. Theoretically, this study delivers a novel, validated empirical model that synthesizes AI-mediated metacognitive scaffolding with dialogic citizenship pedagogy, effectively shifting the instructional paradigm from passive textual memorization to active character internalization. Practically, it offers a concrete, low-cost solution to the operational bottleneck embedded within competency-based curriculum transformations, such as Indonesia's Independent Curriculum. The interactive framework decentralizes the feedback loop, automating personalized formative assessments and resolving the severe limitations of teacher temporal bandwidth in crowded classrooms. Furthermore, by utilizing lightweight, software-based systems rather than cost-prohibitive hardware, this study provides a highly equitable blueprint that directly advances United Nations' Sustainable Development Goal (SDG) 4, proving that transformative quality education and advanced digital literacy can be democratized in resource-constrained developing nations.

However, the localized boundaries of this research must be transparently acknowledged to guide future scholarly inquiries. The empirical scope was constrained by a sample size of N = 64 students situated within a single geographic region in West Java, Indonesia and the intervention was bounded by a compact timeline of eight learning sessions, which limits the immediate assessment of long-term character retention once the digital scaffold is removed. To overcome these limitations, future research directions must transition toward longitudinal mixed-methods designs capable of tracking behavioral durability across multiple academic semesters. Additionally, subsequent studies should deploy this Socratic AI framework across massive, cross-cultural, and international student cohorts. Such expanded research efforts will be essential to evaluate how diverse cultural background variables interact with automated feedback loops, thereby establishing a universally generalizable and globally scalable instructional blueprint for 21st-century citizenship education.

ACKNOWLEDGMENTS

This research is funded by the Universitas Pendidikan Indonesia through the research scheme/program of *Riset Afirmasi Dosen* with Contract No. 393 /UN40.F7/PT.01.01/2026.

REFERENCES

Álamo-Bolaños, A., Mulero-Henríquez, I., & Morata Sampaio, L. (2024). Childhood, Education, and Citizen Participation: A Systematic Review. *Social Sciences*, *13*(8), 399. https://doi.org/10.3390/socsci13080399

Ali, A. S., Alemayehu, Y. A., Mersha, G. T., & Mengistie, B. M. (2026). Opportunities and barriers of artificial intelligence in transforming education in developing regions: a systematic scoping review. *Discover Education*, *5*(1), 584. https://doi.org/10.1007/s44217-026-01816-z

Andreouli, E., Obradović, S., Young, K., Burton, T., & Hecht, A. (2026). Active Citizens and Passive Learning: A Qualitative Study of Students' Perspectives on Citizenship Education Across England and Wales. *Youth & Society*, *58*(1), 152–175. https://doi.org/10.1177/0044118X251377333

Anisa, P., Rahmi, R., Agustina, M., Azhari, M., & Yacob, H. H. (2026). Character Education from Thomas Lickona's Perspective in Building Moral Intelligence in Primary Education. *Journal of Education Research*, *7*(2), 322–327. https://doi.org/10.37985/jer.v7i2.3467

Arif, M., Abdurakhmonovich, Y. A., & Dorloh, S. (2023). Character Education in the 21st Century: The Relevance of Abdul Wahhab Ash Syarani's and Thomas Lickona's Concepts. *Islamic Review: Jurnal Riset Dan Kajian Keislaman*, *12*(1), 35–58. https://doi.org/10.35878/islamicreview.v12i1.690

Badger, M., López, N., & Nissen, C. F. R. (2026). The Impact of GenAI Chatbots on Student Learning in Higher Education: A Literature Review. *International Journal of Technology in Education*, *9*(1), 43–69. https://doi.org/10.46328/ijte.5111

Bai, J., Huang, K., & Bi, Y. (2026). The influence of teachers' criticism styles on students' intention to improve disruptive classroom behavior: A dual-mediation model of negative academic emotions and teacher-student relationship. *Acta Psychologica*, *264*, 106455. https://doi.org/10.1016/j.actpsy.2026.106455

Chen, M.-K., Hsueh, C.-M., Chiang, W.-W., Chiu, K.-I., & Shih, Y.-H. (2025). Paulo Freire's dialogic pedagogy and the democratization of pedagogy in Taiwan's higher education: Challenges and possibilities. *Edelweiss Applied Science and Technology*, *9*(5), 1264–1273. https://doi.org/10.55214/25768484.v9i5.7134

Davar, N. F., Dewan, M. A. A., & Zhang, X. (2025). AI Chatbots in Education: Challenges and Opportunities. *Information*, *16*(3), 235. https://doi.org/10.3390/info16030235

Derrington, M. L., Saleh, A., & Murphy, M. (2026). Supporting Teacher Reflective Practice Through Instructional Supervision: A Qualitative Study. *Education Sciences*, *16*(5), 726. https://doi.org/10.3390/educsci16050726

DiBerardino, N., & Stark, L. (2023). (Anti)-Intentional Harms: The Conceptual Pitfalls of Emotion AI in

Education. *2023 ACM Conference on Fairness Accountability and Transparency*, 1386–1395. https://doi.org/10.1145/3593013.3594088

Donbavand, S., & Hoskins, B. (2021). Citizenship Education for Political Engagement: A Systematic Review of Controlled Trials. *Social Sciences*, *10*(5), 151. https://doi.org/10.3390/socsci10050151

Fitzgerald, J. C., Cohen, A. K., Maker Castro, E., & Pope, A. (2021). A Systematic Review of the Last Decade of Civic Education Research in the United States. *Peabody Journal of Education*, *96*(3), 235–246. https://doi.org/10.1080/0161956X.2021.1942703

Greenberger, S. W., Anderson, P., Hogan, S., & Pasley, M. (2025). Ready for reflection: strategies to enhance open-mindedness in Deweyan reflective practice. *Reflective Practice*, *26*(5), 633–644. https://doi.org/10.1080/14623943.2025.2494318

Greenberger, S. W., & Chavez, T. E. (2026). Judgments of practice: evaluating and deciding in Deweyan reflection. *Reflective Practice*, 1–11. https://doi.org/10.1080/14623943.2026.2632665

Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies*, *30*(4), 4493–4518. https://doi.org/10.1007/s10639-024-12881-y

Hee-Young, K., & Ji-Young, L. (2023). The influence of pre-service early childhood teachers' civic awareness on their ability to teach democratic citizenship education: The mediating role of multicultural beliefs. *International Journal of Advanced Culture Technology*, *11*(4), 131–140. https://doi.org/10.17703/IJACT.2023.11.4.131

Hill, J., & West, H. (2022). Dialogic Feed-Forward in Assessment: Pivotal to Learning but not Unproblematic. *Teaching and Learning Inquiry*, *10*, 1–19. https://doi.org/10.20343/teachlearninqu.10.20

Hoehl, S., Krenn, B., & Vincze, M. (2024). Honest machines? A cross-disciplinary perspective on trustworthy technology for children. *Frontiers in Developmental Psychology*, *2*, 1308881. https://doi.org/10.3389/fdpys.2024.1308881

Holst, J., Singer-Brodowski, M., Brock, A., & de Haan, G. (2024). Monitoring SDG 4.7: Assessing Education for Sustainable Development in policies, curricula, training of educators and student assessment (input-indicator). *Sustainable Development*, *32*(4), 3908–3923. https://doi.org/10.1002/sd.2865

Jerome, L., Hyder, F., Hilal, Y., & Kisby, B. (2024). A systematic literature review of research examining the impact of citizenship education on active citizenship outcomes. *Review of Education*, *12*(2), e3472. https://doi.org/10.1002/rev3.3472

Jonas, A., De Jesús, A., & Hardy, J. (2025). Student Agency and Punishment Logics in Imagined Future Classroom Technologies. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3706598.3713169

Kleinheksel, A. J., Chen, W., Rudd, M. J., Drowos, J., Gupta, S., Minor, S., & Bailey, J. M. (2023). Putting reflection back into practice: Kolb's theory of experiential learning as a theoretical framework for just-in-time faculty development. *SN Social Sciences*, *3*(3), 59. https://doi.org/10.1007/s43545-023-00649-z

Krijnen, M. A., Wansink, B. G. J., van den Berg, Y. H. M., van Tartwijk, J., & Mainhard, T. (2024). Citizenship in the Elementary Classroom Through the Lens of Peer Relations. *Educational Psychology Review*, *36*(1), 34. https://doi.org/10.1007/s10648-024-09870-5

Lee, J., Hicke, Y., Yu, R., Brooks, C., & Kizilcec, R. F. (2024). The life cycle of large language models in education: A framework for understanding sources of bias. *British Journal of Educational Technology*, *55*(5), 1982–2002. https://doi.org/10.1111/bjet.13505

Lei, W., & Lei, Z. (2026). A systematic review of feedback literacy in higher education: concepts, measurement, and interventions. *Assessment & Evaluation in Higher Education*, 1–22. https://doi.org/10.1080/02602938.2026.2656285

Leite, S. (2022). Using the SDGs for global citizenship education: definitions, challenges, and opportunities. *Globalisation, Societies and Education*, *20*(3), 401–413. https://doi.org/10.1080/14767724.2021.1882957

Li, C., Cui, H., & Hagedorn, L. S. (2026). The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. *Computers and Education: Artificial Intelligence*, *10*, 100571. https://doi.org/10.1016/j.caeai.2026.100571

Littenberg-Tobias, J. (2021). Teaching Citizens: What Can NAEP Civics Tell Us about Active Learning in Civics? *Peabody Journal of Education*, *96*(3), 247–260. https://doi.org/10.1080/0161956X.2021.1942704

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning*, *37*(3), 887–901. https://doi.org/10.1111/jcal.12538

Matjie, M. A., Nethavhani, A., & Matlakala, M. (2026). AI and the digital divide in education. *Frontiers in Computer Science*, *8*, 1759027. https://doi.org/10.3389/fcomp.2026.1759027

McGuire, M. (2026). Feedback, reflection and psychological safety: rethinking assessment for student well-being in higher education. *Assessment & Evaluation in Higher Education*, *51*(3), 532–556. https://doi.org/10.1080/02602938.2025.2548590

Morris, T. H. (2020). Experiential learning – a systematic review and revision of Kolb's model. *Interactive Learning Environments*, *28*(8), 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Oogarah-Pratap, B., Bholoa, A., & Ramma, Y. (2025). Stage Theory of Cognitive Development—Jean Piaget. In B. Akpan & T. J. Kennedy (Eds.), *Science Education in Theory and Practice: An Introductory Guide to Learning Theory* (pp. 125–142). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-81351-1_8

Ortega-Ochoa, E., Quiroga Pérez, J., Arguedas, M., Daradoumis, T., & Marquès Puig, J. M. (2024). The effectiveness of empathic chatbot feedback for developing computer competencies, motivation, self-regulation, and metacognitive reasoning in online higher education. *Internet of Things*, *25*, 101101. https://doi.org/10.1016/j.iot.2024.101101

Overholser, J. C., & Beale, E. (2023). The art and science behind socratic questioning and guided discovery: a research review. *Psychotherapy Research*, *33*(7), 946–956. https://doi.org/10.1080/10503307.2023.2183154

Pietersen, D. (2022). Engaging Paulo Freire on deliberative democracy: Dialogical pedagogy, deliberation and inclusion in a transformative higher education online education space. *Transformation in Higher Education*, *7*(0), a211. https://doi.org/10.4102/the.v7i0.211

Prabowo, C. D. (2024). Moral Values in Digital Learning: Applying Thomas Lickona's Framework in Online Higher Education. *Journal of Islamic Education Management Research*, *2*(2), 165–180. https://doi.org/10.14421/jiemr.2024.22-07

Puntambekar, S. (2022). Distributed Scaffolding: Scaffolding Students in Classroom Environments. *Educational Psychology Review*, *34*(1), 451–472. https://doi.org/10.1007/s10648-021-09636-3

Raslan, G. (2024). The Impact of the Zone of Proximal Development Concept (Scaffolding) on the Students Problem Solving Skills and Learning Outcomes. In K. Al Marri, F. A. Mir, S. A. David, & M. Al-Emran (Eds.), *BUiD Doctoral Research Conference 2023* (pp. 59–66). Springer Nature Switzerland.

Santamaría-Cárdaba, N., Gajardo-Espinoza, K., & Cáceres-Iglesias, J. (2024). Global Citizenship Education: A Systematic Review of Educational Research. *Educational Process International Journal*, *13*(4), 115–131. https://doi.org/10.22521/edupij.2024.134.7

Sari, R. E., Ambiyar, Giatman, M., & Adri, M. (2025). Development of a Formative Assessment Model for the Pancasila Student Profile Strengthening Project for the Independent Curriculum. *Jurnal Penelitian Pendidikan IPA*, *11*(1), 664–672. https://doi.org/10.29303/jppipa.v11i1.9665

Schmid, B., Bleijlevens, N., Mani, N., & Behne, T. (2024). The cognitive underpinnings and early development of children's selective trust. *Child Development*, *95*(4), 1315–1332. https://doi.org/10.1111/cdev.14073

Singh, K., & Mueller, J. (2021). Taking a nuanced view of the role of teacher feedback in the elementary classroom. *Teachers and Teaching*, *27*(1–4), 95–115. https://doi.org/10.1080/13540602.2021.1933415

Sousa, D., & Moss, P. (2023). Towards democratic culture and political practice in early childhood education: The case for transformative change at a time of converging crisis. *ETD - Educação Temática Digital*, *25*, e023067. https://doi.org/10.20396/etd.v25i00.8672067

Suo, X., Yin, B., Wang, W., & Guo, L. (2026). Exploring the Impact of Generative AI-Powered Conversational Agents on Student Learning: A Systematic Review and Meta-Analysis Grounded in Activity Theory. *Journal of Educational Computing Research*, *64*(6), 1672–1702. https://doi.org/10.1177/07356331261433396

Tan, S. C., Tan, Y. Y., Teo, C. L., & Yuan, G. (2026). Teachers' professional agency in learning with AI: A case study of a generative AI-based knowledge building learning companion for teachers. *British Journal of Educational Technology*, *57*(4), 943–964. https://doi.org/10.1111/bjet.70013

Tian, X., & Tang, Y. (2025). From Awareness to Behavior: The Empirical Effects of Real Problem-Oriented Learning in Civic and Moral Education. *Sage Open*, *15*(2), 21582440251338948. https://doi.org/10.1177/21582440251338948

Tran, M., Balasooriya, C., Semmler, C., & Rhee, J. (2025). Generative artificial intelligence: the 'more knowledgeable other' in a social constructivist framework of medical education. *Npj Digital Medicine*, *8*(1), 430. https://doi.org/10.1038/s41746-025-01823-8

Utami, A. D., & Yus, A. (2025). Interpreting Vygotsky's zone of proximal development in the context of children's learning through play. *Journal of Early Childhood Teacher Education*, 1–14.

https://doi.org/10.1080/10901027.2025.2591098
van Nooijen, C. C. A., de Koning, B. B., Bramer, W. M., Isahakyan, A., Asoodar, M., Kok, E., van Merrienboer, J. J. G., & Paas, F. (2024). A Cognitive Load Theory Approach to Understanding Expert Scaffolding of Visual Problem-Solving Tasks: A Scoping Review. *Educational Psychology Review*, *36*(1), 12. https://doi.org/10.1007/s10648-024-09848-3
Van Orman, D. S. J., Gotch, C. M., & Carbonneau, K. J. (2025). Preparing Teacher Candidates to Assess for Learning: A Systematic Review. *Review of Educational Research*, *95*(3), 427–463. https://doi.org/10.3102/00346543241233015
Vittorio, L. N., Murphy, S. T., Braun, J. D., & Strunk, D. R. (2022). Using Socratic Questioning to promote cognitive change and achieve depressive symptom reduction: Evidence of cognitive change as a mediator. *Behaviour Research and Therapy*, *150*, 104035. https://doi.org/10.1016/j.brat.2022.104035
Wang, P., Chen, F., Wang, D., & Chen, G. (2025). Enhancing students' dialogic reflection through classroom discourse visualisation. *International Journal of Computer-Supported Collaborative Learning*, *20*(3), 293–315. https://doi.org/10.1007/s11412-024-09443-2
White, G., Dong, D., Campbell, D. E., & Lee, C. D. (2023). The Future of Civic Education: Implications for Research, Policy, and Practice. *The ANNALS of the American Academy of Political and Social Science*, *705*(1), 249–257. https://doi.org/10.1177/00027162231193538
Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370
Yin, J., Zhu, Y., Goh, T.-T., Wu, W., & Hu, Y. (2024). Using Educational Chatbots with Metacognitive Feedback to Improve Science Learning. *Applied Sciences*, *14*(20), 9345. https://doi.org/10.3390/app14209345
Yıldız, T. (2025). From Constructivism To Cultural Cognition: A Comparative Analysis Of Piaget, Vygotsky, And Tomasello's Theories Of Cognitive Development. *HUMANITAS - Uluslararası Sosyal Bilimler Dergisi*, *13*(25), 411–429. https://doi.org/10.20304/humanitas.1601228
Yu, R., & Yang, L. (2021). ESL/EFL Learners' Responses to Teacher Written Feedback: Reviewing a Recent Decade of Empirical Studies. *Frontiers in Psychology*, *12*, 735101. https://doi.org/10.3389/fpsyg.2021.735101
Zhang, H., Ge, S., & Mohd Saad, M. R. Bin. (2024). Formative assessment in K-12 English as a foreign language education: A systematic review. *Heliyon*, *10*(10), e31367. https://doi.org/10.1016/j.heliyon.2024.e31367
Zhang, Q., Zhang, N., & Lu, C. (2026). How do pedagogical approaches affect the impact of chatbots on learning performance? A meta-analysis and research synthesis. *Educational Research Review*, *51*, 100783. https://doi.org/10.1016/j.edurev.2026.100783