

Designing a Contextual AI-Based Deep Learning Model Grounded in Melayu Deli Tradition to Improve Fifth-Grade Elementary School Students' Creative Literary Literacy

Dewi Sartika Panggabean^{1,*}, Yusi Tri Utari Panggabean², Zulvia Misykah¹, Perti Wahyuni Halawa¹, Anisa Ramadhani¹, and Yuli Sutoto Nugroho³

¹ Elementary School Teacher Education, Universitas Battuta, Medan, Indonesia

² Corporate Management, Sekolah Tinggi Ilmu Ekonomi Al-Washliyah, Medan, Indonesia

³ School of Electronic Engineering and Computer Science at Queen Mary, University of London, London
Email: tikapanggabean.n@gmail.com (D.S.P.); yusitriutari@gmail.com (Y.T.U.P.); via.japanese@gmail.com (Z.M); halawaperty@gmail.com (P.W.H); anisaaramadhani30@gmail.com (A.R); y.nugroho@qmul.ac.uk (Y.S.N)

*Corresponding author

Abstract—Creative literary literacy among Indonesian elementary school students remains low, particularly in the appreciation of local cultural texts, because literary instruction is rarely connected to the cultural heritage that children experience in daily life. This study therefore designed and tested a Contextual AI-Based Deep Learning (CAIDL) model grounded in Melayu Deli tradition to improve the creative literary literacy of fifth-grade students. Following the five-stage Plomp research and development model, the study involved seven expert validators, 12 students in a small-group trial, three teachers, and 60 fifth-grade students at SDN Medan Deli assigned to an experimental group ($n = 30$) and a control group ($n = 30$). Data were collected through expert validation sheets, practicality questionnaires, and a 24-item creative literary literacy test (Cronbach's $\alpha = .91$). The CAIDL model was rated very valid (88.80%), very feasible (87.08%), and very practical (89.37%). On the posttest, the experimental group significantly outperformed the control group, $t(58) = 8.31$, $p < .001$, Cohen's $d = 2.14$, with a mean N -Gain of 0.582 (medium category); the largest indicator-level effects were found for creative fluency ($d = 1.71$) and symbolic interpretation ($d = 1.68$). These findings indicate that integrating Melayu Deli cultural context, Socratic AI scaffolding, and deep learning pedagogy strengthens both the appreciative and the creative dimensions of elementary students' literary literacy and offers a replicable model for culturally grounded literacy instruction.

Keywords—contextual AI, creative literary literacy, deep learning, elementary education, Melayu Deli tradition

Manuscript received August 13, 2026; revised September 25, 2026; accepted October 2, 2026

I. Introduction

Creative literary literacy is a foundational 21st-century competency that underpins elementary school students' reflective, imaginative, and adaptive thinking, yet it has not received adequate attention in Indonesian instructional practice (Beghetto, 2014; Cremin & Chappell, 2021; Kemendikbudristek, 2022). Elementary students who master creative literary literacy are not only able to comprehend texts literally, but can also interpret symbolic meaning, appreciate aesthetic value, and produce authentic literary expressions grounded in their cultural experience (Kaufman & Beghetto, 2009; Sawyer, 2012). Unfortunately, national and international literacy assessments indicate that Indonesian students' literacy skills, particularly in the appreciative and productive dimensions of literary texts, remain below the functional threshold (OECD, 2023; Mullis et al., 2017; Pratiwi & Mahdiansyah, 2017).

This problem is structural in nature. Indonesian language instruction at the elementary level remains dominated by a surface-learning pattern, namely memorizing character names, repeating story summaries, and answering closed-ended text questions with no room for reflective interpretation (Marton & Säljö, 1976; Biggs, 1987; Fullan & Langworthy, 2014). The literary materials presented in textbooks also tend to be generic and detached from children's cultural context, thereby failing to cultivate emotional and aesthetic engagement with

literary works (Purnamasari & Setiawan, 2017; Suryani et al., 2018). Yet, as emphasized by Vygotsky's (1978) sociocultural theory and contemporary creative-pedagogy frameworks, children's literary learning becomes more meaningful when rooted in the cultural heritage they know and live daily (Cremin et al., 2014; Craft, 2003).

North Sumatra, particularly the Medan area and its surroundings, holds an abundant wealth of Melayu Deli literary tradition in the form of *pantun* (traditional quatrains), *hikayat* (epic tales), *syair* (long lyric poems), *mantra* (incantations), and folktales rich in philosophical, ecological, and religious values (Sinar, 2011). This repository has the potential to become an authentic context for creative literary literacy learning among elementary students in the region, since children have absorbed the diction, rhythm, and aesthetics of Melayu Deli tradition from the pre-school years through family communication, customary ceremonies, and local media. However, the use of Melayu Deli tradition as a learning resource for creative literary literacy in elementary classrooms remains very limited; most teachers still lack adequate teaching materials and effective pedagogical strategies for integrating this tradition into Indonesian language instruction (Rahmawati & Ridwan, 2017; Parmin et al., 2016).

On the other hand, the emergence of generative artificial intelligence in the form of Large Language Models such as ChatGPT and Claude offers new opportunities for elementary education to provide Socratic tutors that accompany students in dialogue about literary texts through reflective questioning that was previously only possible in one-on-one tutoring (Kasneci et al., 2023; Holmes & Tuomi, 2022; Chiu et al., 2024). Recent studies show that AI configured to pose diagnostic questions, rather than to provide direct answers, can support children's reflective learning when guided by a teacher (Lo, 2023; Tlili et al., 2023). Unfortunately, most studies on AI integration in education still focus on mathematics and science contexts, while AI integration in children's literary learning grounded in local cultural contexts remains rarely explored systematically (Yan et al., 2024).

Prior studies have examined various aspects of children's literary learning and creativity separately. Amabile (1996) and Csikszentmihalyi (1996) laid the theoretical foundations of creativity that position socio-cultural context as the primary matrix. Beghetto and Kaufman (2007) subsequently developed the concept of mini-creativity, which is relevant to everyday elementary classroom learning contexts. In Indonesia, Suryani et al. (2018) and Purnamasari and Setiawan (2017) showed that local-wisdom-based learning improves elementary students' critical thinking skills. However, no study has comprehensively integrated the four pillars of contextual learning grounded in Melayu Deli tradition, Socratic AI tutor scaffolding, deep learning pedagogy, and a creative literary literacy assessment framework with valid and reliable instruments within the elementary school context of North Sumatra. This is precisely the gap that the present study addresses.

Based on the foregoing, this study was designed to answer four research questions: (1) what is the validity of expert judgment results for the Contextual AI-Based Deep Learning (CAIDL) model grounded in Melayu Deli tradition; (2) what is the operational feasibility of the model based on a small-group trial at SDN Medan Deli; (3) what is the practicality of the model based on teacher and student responses during field implementation; and (4) what is the effectiveness of the CAIDL model in improving fifth-grade elementary students' creative literary literacy, as measured through pretest, posttest, and N-Gain analysis. This study aims to produce a valid, feasible, practical, and effective learning model for improving elementary students' creative literary literacy grounded in Melayu Deli tradition, one that can be replicated in other local cultural contexts across Indonesia.

II. Literature Review

A. Contextual AI-Based Deep Learning (CAIDL)

Contextual AI-Based Deep Learning is a learning framework that integrates three pedagogical components. The first component is contextual learning, which situates learning material within authentic situations familiar to students, as outlined by Contextual Teaching and Learning (Crawford, 2001). The second component is the deep learning approach, which pushes students to move from surface memorization toward integrative meaning making through reflection and connections between prior and new knowledge (Marton & Säljö, 1976; Biggs, 1987; Fullan & Langworthy, 2014). The third component is AI scaffolding functioning as a Socratic tutor that poses diagnostic questions to guide students' reasoning, rather than simply providing answers (Kasneci et al., 2023; Holmes & Tuomi, 2022; Chiu et al., 2024). The combination of these three components produces a six-phase instructional syntax mapped onto the stages of concept formation from APOS theory (Dubinsky, 1991), namely context exposure, procedural exploration, object conceptualization, and reflective schema integration.

B. Melayu Deli Tradition as a Pedagogical Context

Melayu Deli tradition encompasses a range of oral and written literary works that developed in the region of the Deli sultanate, North Sumatra, since the 17th century and remain alive today (Sinar, 2011). This repository

includes *pantun*, with their *sampiran–isi* (couplet–content) structure that trains sensitivity to rhyme; *hikayat*, with heroic plots laden with moral values; *syair*, a long lyric form with spiritual themes; *mantra* (incantations), rich in symbolic diction; and folktales that record the ecological wisdom of coastal communities. Integrating this tradition as a pedagogical context is consistent with the principle of local-wisdom-based learning, which has proven effective across various Indonesian curriculum contexts (Purnamasari & Setiawan, 2017; Suryani et al., 2018; Rahmawati & Ridwan, 2017; Parmin et al., 2016). This approach simultaneously responds to the mandate of the Merdeka Curriculum, which encourages the integration of local content into the Pancasila Student Profile (Kemendikbudristek, 2022; Wahyudin et al., 2023).

C. Creative Literary Literacy

Creative literary literacy is a competency that integrates appreciative and productive capacities toward literary texts (Kaufman & Beghetto, 2009; Cremin et al., 2014). The appreciative dimension encompasses literal text comprehension, appreciation of aesthetic elements, identification of moral and cultural values, and symbolic interpretation. The creative dimension encompasses fluency of ideas, flexibility of storyline, originality of expression, and elaboration of narrative (Beghetto & Kaufman, 2007; Cropley, 2006; Sawyer, 2012). Kaufman and Beghetto's (2009) Four-C model of creativity serves as the primary reference, as it distinguishes mini-c (creativity within learning), little-c (everyday creativity), Pro-c (professional creativity), and Big-C (eminent creativity), with elementary student contexts focusing on mini-c and little-c. A number of recent studies show that structured creative pedagogy is capable of simultaneously improving both dimensions (Cremin & Chappell, 2021; Toivanen et al., 2013; Parsons et al., 2010).

D. Deep Learning Pedagogy for Children's Literary Instruction

The deep learning pedagogy referred to here is not deep learning in the sense of artificial neural network algorithms, but rather the deep approach to learning formulated by Marton and Säljö (1976) and extended by Biggs (1987) and Fullan and Langworthy (2014) as an approach that emphasizes meaning-making and connections between pieces of knowledge. Within the context of children's literary learning, this pedagogy is realized through structured reflection activities, reflective dialogue among students, reader journals, and creative-expression portfolios. Recent research shows that combining deep learning pedagogy with Socratic AI scaffolding fosters higher-order questioning and richer creative expression (Chiu et al., 2024; Lo, 2023; Webb et al., 2019). Cognitive load theory (Sweller et al., 2019) is also relevant for ensuring that the complexity of literary literacy tasks remains within the growth zone of fifth-grade elementary students.

III. Method

A. Research Design and Model

This study employed a Research and Development (R&D) design based on the Plomp model developed at the University of Twente (Plomp, 2013; Plomp & Nieveen, 2013). The Plomp model was selected because its iterative formative-evaluation cycle matches the developmental logic of a multi-pillar model such as CAIDL, which requires repeated refinement based on expert and user feedback. This model proceeds through five stages: (1) preliminary investigation, (2) design, (3) realization or construction, (4) test, evaluation, and revision, and (5) limited implementation. The complete sequence of stages and their respective activities is visualized in Figure 1. The study was conducted over seven months, from August 2024 to February 2025, with SDN Medan Deli's fifth-grade class serving as the primary implementation partner.



Figure 1. Stages of the Plomp Research and Development Model for CAIDL Melayu Deli Development

B. Population and Sample

The population and sample of this study were divided into seven participant groups corresponding to the Plomp stages, as detailed in Table 1. Participant selection used purposive sampling with inclusion criteria specific to each stage (Creswell & Creswell, 2018; Shadish et al., 2002; Sugiyono, 2019). The seven-member

expert panel was selected based on academic qualifications and expertise across the four CAIDL pillars. Students for the small-group trial were selected through stratified sampling across the range of baseline literacy ability. The experimental and control groups were drawn from two parallel fifth-grade classes at SDN Medan Deli with equivalent demographic backgrounds. Research ethics were satisfied through approval from the Universitas Battuta Research Ethics Committee (Number: [insert approval number]), permits from the local education office, and informed consent from students' parents/guardians.

Table 1. Research Participants by Plomp Model Stage

No	Research Stage	Participant	N	Participant Profile
1	Test, Evaluation, Revision	Literary content & language experts	3	Indonesian language & literature education lecturers, doctoral degree, >10 years of research experience, Sinta/Scopus-indexed publications
2	Test, Evaluation, Revision	Educational technology and AI experts	2	Lecturers specializing in instructional design and generative AI integration in elementary education
3	Test, Evaluation, Revision	Melayu Deli cultural experts	2	Academics or cultural practitioners with mastery of the Melayu Deli literary repository and the elementary curriculum
4	Small-group trial	Fifth-grade students at SDN Medan Deli	12	Stratified sampling: 4 high literacy ability, 4 medium, 4 low, across genders
5	Field implementation	Elementary Indonesian-language teachers	3	Fifth-grade Indonesian-language teachers with a minimum of 5 years teaching experience and adequate digital literacy
6	Field implementation	Experimental group	30	Fifth-grade students at SDN Medan Deli who received the CAIDL intervention for 8 weeks
7	Field implementation	Control group	30	Fifth-grade students at SDN Medan Deli who received conventional literary instruction with equivalent coverage

C. Data Collection Instrument

The data collection instruments consisted of five types, each validated prior to field use. First, an expert validation sheet developed into four sub-instruments for the literary-content validators (28 items), pedagogy validators (24 items), AI-integration validators (20 items), and language validators (20 items), each using a four-point Likert scale, with Aiken's V coefficients ranging from .82 to .94 (Aiken, 1985; Fauzi & Pradipta, 2018). Second, a 20-item practicality questionnaire for teachers and a 20-item practicality questionnaire for students, with reliability coefficients of Cronbach's $\alpha = .89$ and $\alpha = .87$, respectively. Third, a 24-item essay-based creative literary literacy test as the primary instrument measuring both dependent variables, with the blueprint presented in Table 2 and a reliability coefficient of Cronbach's $\alpha = .91$. Fourth, an implementation-fidelity observation sheet with inter-observer agreement of Cohen's $\kappa = .88$. Fifth, a semi-structured interview guide to explore teacher and student experiences with the AI tutor and reflections on Melayu Deli tradition (Creswell & Creswell, 2018).

Table 2. Blueprint of the Fifth-Grade Creative Literary Literacy Test (24 Essay Items)

No	Variable	Indicator	Sub-Indicator	Items
1	Literary Appreciation	Literal comprehension of literary text	Identification of characters, setting, plot; accurate summary	3
2	Literary Appreciation	Appreciation of literary aesthetic elements	Identification of diction, rhyme, figurative language; aesthetic response	3
3	Literary Appreciation	Identification of moral & cultural values	Inference of Melayu Deli philosophical values; relevance of values	3
4	Literary Appreciation	Symbolic interpretation of text	Interpretation of Malay cultural symbols; interpretation of metaphor	3
5	Literary Creativity	Creative fluency	Number of varied literary ideas generated	3
6	Literary Creativity	Storyline flexibility	Variation in narrative structure; shifts in point of view	3
7	Literary Creativity	Originality of literary expression	Novelty of diction & metaphor; uniqueness of narrative ideas	3

8	Literary Creativity	Elaboration and narrative depth	Supporting detail of ideas; extension to related cases; reflection	3
Total Items				24

D. Research Procedure and Data Analysis

The research procedure followed the five Plomp stages sequentially, with formative evaluation cycles between stages. The preliminary investigation stage, in August 2024, carried out a needs analysis through initial observation at SDN Medan Deli, profiling of 84 fifth-grade students, a review of Melayu Deli oral literature, and a review of the deep learning and creative pedagogy literature (Sinar, 2011; Marton & Säljö, 1976; Fullan & Langworthy, 2014). The design stage, in September 2024, developed the six-phase CAIDL syntax, the Socratic AI prompt library, a curated corpus of Melayu Deli *pantun* and *hikayat*, and the creative literary literacy rubric. The realization stage, in October 2024, produced six instructional modules (advisory *pantun*, humorous *pantun*, heroic *hikayat*, ecological folktales, symbolic *mantra*, and a creative writing project), an AI tutor configured with the Socratic prompt library, an A4-format student activity book, and a teacher's guide. The test-evaluate-revise stage, in October–November 2024, ran two cycles: validation by the seven-member expert panel produced 23 revision items, and the small-group trial with 12 students produced five pragmatic revisions. The limited implementation stage, from December 2024 to February 2025, ran for eight effective weeks with 60 students, beginning with a pretest, followed by eight CAIDL sessions (twice weekly, 70 minutes per session), and concluding with a posttest and response questionnaires (Hake, 1998; Cohen, 1988).

Data analysis followed the R&D methodological standards applicable to elementary education research (Cheung & Slavin, 2013; Castles et al., 2018). First, expert validation data were analyzed using Aiken's V coefficient, with thresholds of very valid ($\geq .85$), valid (.70–.84), adequate (.55–.69), and weak ($< .55$) (Aiken, 1985). Second, feasibility and practicality were analyzed through mean response percentages, with categories of very feasible/very practical ($\geq 85\%$), feasible/practical (70–84%), adequate (55–69%), and weak ($< 55\%$). Third, effectiveness was analyzed through (a) Shapiro–Wilk tests for normality, reported with their degrees of freedom (df), and Levene's test for homogeneity of variance, (b) paired-sample t -tests for within-group pretest-posttest comparisons, (c) independent-sample t -tests for between-group posttest comparisons, (d) Cohen's d effect size, interpreted as small ($d = 0.20$), medium ($d = 0.50$), and large ($d = 0.80$) following Cohen (1988), with the extended benchmarks of very large ($d = 1.20$) and huge ($d = 2.00$) proposed by Sawilowsky (2009) applied to effects beyond the conventional range, and (e) normalized N-Gain analysis (Hake, 1998) with categories of high ($g \geq 0.70$), medium ($0.30 \leq g < 0.70$), and low ($g < 0.30$). Per-indicator effect sizes were calculated across all eight creative literary literacy sub-indicators to enable construct-level diagnosis. All quantitative analyses were conducted in SPSS v28 and R 4.3.2 at a significance level of $\alpha = .05$.

IV. Results and Discussion

A. Research Results

The realization or construction stage of the Plomp model produced a complete prototype of the Contextual AI-Based Deep Learning (CAIDL) model grounded in Melayu Deli tradition, ready for empirical testing at SDN Medan Deli. The findings are reported sequentially to answer the four research questions, covering expert validity, operational feasibility in the small-group trial, practicality based on teacher and student responses, and effectiveness in improving creative literary literacy. Table 3 presents the categorization criteria that convert numerical scores into qualitative categories consistent with the R&D literature in elementary education. The validity, feasibility, and practicality criteria were adopted from Indonesian R&D guidelines, while the N-Gain criteria follow the Hake (1998) framework as the international standard for measuring learning effectiveness (Cheung & Slavin, 2013).

Table 3. Categorization Criteria for Validity, Feasibility, Practicality, and N-Gain

Percentage / Score	Validity Category	Feasibility / Practicality Category	N-Gain Category
$\geq 85\%$	Very Valid	Very Feasible / Very Practical	High ($g \geq 0.70$)
70–84%	Valid	Feasible / Practical	Medium ($0.30 \leq g < 0.70$)
55–69%	Adequate	Adequate	Low ($g < 0.30$)
$< 55\%$	Weak	Weak	(not applicable)

Source: adapted from Indonesian R&D guidelines and Hake (1998)

Expert validation by a panel of seven, comprising three literary-content and language experts, two educational technology and AI experts, and two Melayu Deli cultural experts, produced a very valid category across all aspects, with an overall mean of 88.80% as shown in Table 4. The coherence of contextual and deep

learning pedagogy aspect obtained the highest score (91.00%), confirming that the four pillars of the CAIDL model were operationalized as a coherent instructional structure rather than loosely attached components. The AI-scaffolding integration as Socratic tutor aspect obtained a score of 89.00%, indicating that the AI prompt library functioned according to the pedagogical intent of stimulating reasoning rather than replacing it (Kasneci et al., 2023; Chiu et al., 2024). The accuracy of Melayu Deli literary substance aspect obtained 88.67%, implying that the selected corpus of *pantun* and *hikayat* authentically represented the traditional repository (Sinar, 2011). The completeness of creative literary literacy indicators aspect obtained 88.00%, and the clarity of language aspect obtained 87.33%, both falling within the very valid category, with minor revisions already accommodated prior to the feasibility stage.

Table 4. Expert Validation Results for the CAIDL Model Grounded in Melayu Deli Tradition

No	Validation Aspect	V1	V2	V3	Mean (%)
1	Accuracy of Melayu Deli literary substance	89	88	89	88.67
2	Completeness of creative literary literacy indicators	88	89	87	88.00
3	Coherence of contextual and deep learning pedagogy	92	90	91	91.00
4	AI scaffolding integration as Socratic tutor	90	89	88	89.00
5	Language clarity for fifth-grade students	87	88	87	87.33
Expert Validity Mean		89.20	88.80	88.40	88.80

Source: primary analysis using Aiken's *V* coefficient, 2024

A visual summary across the five aspects and the overall mean is presented in Figure 2. The pedagogical coherence aspect received the highest score, followed by AI integration, then substance accuracy, indicator completeness, and language clarity. All five aspects exceeded the 85% threshold for the very valid category (Aiken, 1985). The 88.80% mean across seven experts qualified the model design to proceed to the operational feasibility testing stage in the elementary student small-group trial.

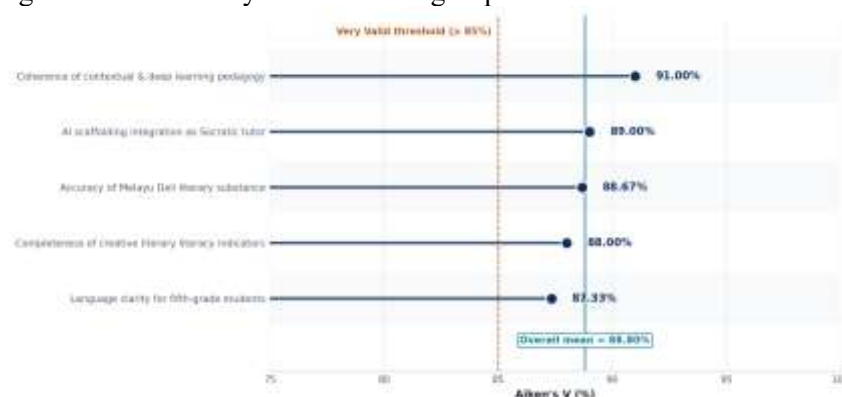


Figure 2. Lollipop Chart of Expert Validity Scores for the CAIDL Model

The small-group trial with twelve fifth-grade students at SDN Medan Deli, selected through stratified sampling across three ranges of baseline literacy ability (high, medium, low), tested operational feasibility prior to field implementation. As shown in Table 5, this trial produced a feasibility mean of 87.08%, falling within the very feasible category. The relevance of Melayu Deli tradition to student experience aspect obtained the highest indicator score (89.58%), confirming that the framing of *pantun*, *hikayat*, and folktale texts resonated with the everyday world of Medan Deli children. The student engagement in creative tasks aspect (88.54%) and the clarity of learning stages aspect (87.50%) also fell within the very feasible category, indicating that the CAIDL syntax could be carried out by students with minimal teacher guidance. The ease of AI tutor use aspect (86.46%) fell within the very feasible category, while the appropriateness of cognitive challenge aspect (83.33%) fell within the feasible category, indicating that some students initially found the symbolic-interpretation task somewhat challenging.

Table 5. Operational Feasibility Results in the Small-Group Trial ($n = 12$)

No	Feasibility Indicator	Max Score	Score Obtained	Percentage	Category
1	Relevance of Melayu Deli tradition to student experience	48	43.00	89.58%	VF
2	Ease of Socratic AI tutor use	48	41.50	86.46%	VF

3	Clarity of learning stages within the CAIDL syntax	48	42.00	87.50%	VF
4	Student engagement in creative literary literacy tasks	48	42.50	88.54%	VF
5	Appropriateness of cognitive challenge for fifth grade	48	40.00	83.33%	F
Operational Feasibility Mean ($n = 12$)		240	209.00	87.08%	VF

Source: primary analysis, 2024. Note: VF = Very Feasible; F = Feasible.

The bullet chart in Figure 3 visualizes the distribution of feasibility scores per indicator, showing that four of the five indicators exceeded the 85% threshold for the very feasible category. Appropriateness of cognitive challenge was the sole lagging indicator, with a qualitative note that the *mantra* symbolic-interpretation task required gentler initial scaffolding before students were ready to enter the narrative elaboration phase. Five pragmatic revisions implemented prior to the field trial included: (1) adding examples of symbolic interpretation to the activity sheets, (2) splitting the symbolic-interpretation task into two sub-tasks, (3) refining the Socratic AI prompts with more child-friendly language, (4) adding a routine of paired discussion prior to creative writing, and (5) adding a moment of peer-work appreciation at the end of each session.

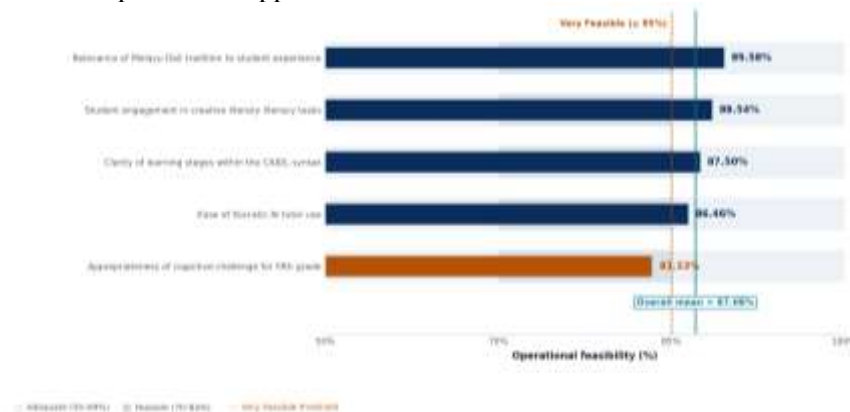


Figure 3. Bullet Chart of Operational Feasibility for the CAIDL Model in the Small-Group Trial

The practicality trial at the limited-implementation stage involved three fifth-grade Indonesian-language teachers and thirty experimental-group students who completed response questionnaires after eight weeks of CAIDL intervention. Table 6 presents teacher responses across five dimensions, with a mean of 89.33% in the very practical category. The relevance of Melayu Deli tradition to curriculum objectives dimension obtained the highest score (92.50%), confirming teachers' perception that the literary corpus used fully supported Merdeka Curriculum Indonesian-language learning outcomes (Kemendikbudristek, 2022). The coherence of learning stages dimension (91.67%) and ease of AI tutor integration dimension (89.17%) also fell within the very practical category, indicating that teachers were able to implement the design without extensive prior AI training. The lowest score from the teacher perspective emerged on the appropriateness of time allocation dimension (85.00%), reflecting the pragmatic challenge of accommodating deep reflection cycles within an already packed elementary school schedule.

Table 6. Teacher Practicality Response Results for the CAIDL Model ($n = 3$)

No	Teacher Practicality Dimension	Max Score	Score Obtained	Percentage	Category
1	Coherence of learning stages within the CAIDL syntax	60	55.00	91.67%	VP
2	Ease of AI tutor integration in instruction	60	53.50	89.17%	VP
3	Relevance of Melayu Deli tradition to curriculum objectives	60	55.50	92.50%	VP
4	Support for students' creative literary literacy development	60	53.00	88.33%	VP
5	Appropriateness of instructional time allocation	60	51.00	85.00%	VP
Teacher Practicality Mean ($n = 3$)		300	268.00	89.33%	VP

Source: primary analysis, 2025. Note: VP = Very Practical.

Student practicality responses are presented in Table 7, with a mean of 89.40% in the very practical category. The usefulness of the AI tutor for understanding literary texts dimension obtained the highest score (91.00%), consistent with recent findings that generative AI tutors can support conceptual understanding when configured as Socratic questioners rather than direct-answer providers (Lo, 2023; Chiu et al., 2024). The engagement with Melayu Deli traditional context dimension (90.25%) ranked second, confirming that authentic cultural framing was able to sustain student attention throughout the eight-week intervention. Interestingly, teacher and student perspectives converged on the appropriateness of time allocation dimension as the lowest (teachers 85.00%; students 86.50%), indicating a structural rather than a design-specific issue. The combined mean of teacher practicality (89.33%) and student practicality (89.40%) yielded an overall practicality of 89.37% in the very practical category. Figure 4 visualizes a dumbbell comparison of practicality from both perspectives.

Table 7. Student Practicality Response Results for the CAIDL Model ($n = 30$)

No	Student Practicality Dimension	Max Score	Score Obtained	Percentage	Category
1	Usefulness of AI tutor for understanding literary texts	450	409.50	91.00%	VP
2	Clarity of progress stages within the CAIDL syntax	450	402.75	89.50%	VP
3	Engagement with Melayu Deli traditional context	450	406.13	90.25%	VP
4	Perceived strengthening of creative literary literacy	450	403.88	89.75%	VP
5	Appropriateness of time allocation per session	450	389.25	86.50%	VP
Student Practicality Mean ($n = 30$)		2,250	2,011.51	89.40%	VP

Source: primary analysis, 2025. Note: VP = Very Practical.

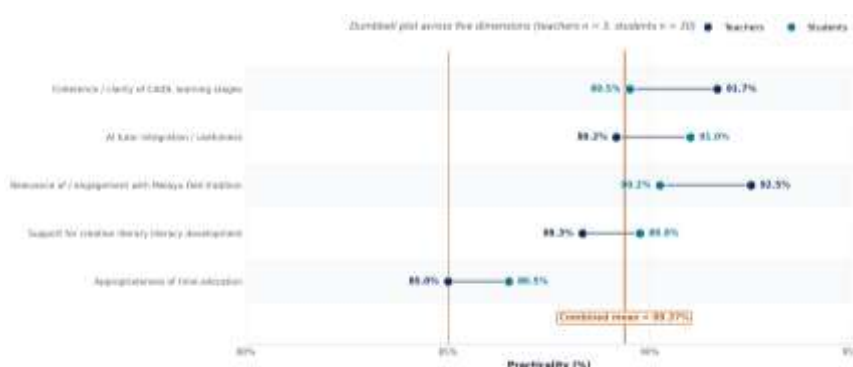


Figure 4. Dumbbell Plot Comparing Practicality between Teacher and Student Perspectives

The effectiveness of the CAIDL model on fifth-grade elementary students' creative literary literacy was analyzed through descriptive statistics for pretest and posttest, paired-sample t -tests, independent-sample t -tests, Cohen's d effect size, and normalized N-Gain analysis (Hake, 1998; Cohen, 1988). Table 8 presents the descriptive statistics, showing baseline equivalence between the two groups, with an experimental pretest mean of 54.20 ($SD = 7.85$) and a control pretest mean of 53.80 ($SD = 7.62$). The pretest t -test produced a non-significant difference ($t(58) = 0.20, p = .842$), satisfying the equivalence prerequisite for the quasi-experimental design (Shadish et al., 2002). After eight weeks of intervention, the experimental group reached a posttest mean of 80.10 ($SD = 7.20$), a gain of 25.90 points, well above the control posttest mean of 64.50 ($SD = 7.41$), a gain of 10.70 points. Prerequisite tests confirmed a normal distribution across all four score sets (Shapiro–Wilk, $df = 30$ for each score set, $p = .142$ to $.418$) and homogeneity of variance (Levene's $F(1, 58) = 1.12, p = .294$), satisfying the parametric assumptions for the subsequent inferential analysis.

Table 8. Descriptive Statistics of Pretest and Posttest Creative Literary Literacy Scores

Group	N	Pretest M (SD)	Posttest M (SD)	Mean Gain
Experimental	30	54.20 (7.85)	80.10 (7.20)	25.90
Control	30	53.80 (7.62)	64.50 (7.41)	10.70

Table 9 reports the paired-sample *t*-test for each group and the independent-sample *t*-test on posttest scores, confirming a significant improvement in the experimental group. In the experimental group, the pretest-posttest difference of 25.90 points yielded $t(29) = 14.10$, $p < .001$, with Cohen's $d = 1.62$, a very large effect (Sawilowsky, 2009). The control group also showed a significant increase of 10.70 points ($t(29) = 6.42$, $p < .001$, $d = 0.71$), but with a medium magnitude far below that of the experimental group. The independent *t*-test on posttest scores yielded a significant difference of 15.60 points ($t(58) = 8.31$, $p < .001$, $d = 2.14$), which falls within the huge effect category ($d \geq 2.00$) according to the extended benchmarks of Sawilowsky (2009). Because both groups completed identical pretests and posttests, this between-group difference cannot be attributed to repeated-testing effects alone and is consistent with an intervention effect of the CAIDL model. Figure 5 visualizes the trajectories and N-Gain distribution between groups.

Table 9. Paired Sample and Independent Sample *t*-Test Results

Comparison	Mean Diff	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
Pretest-Posttest, Experimental	25.90	14.10	29	< .001	1.62
Pretest-Posttest, Control	10.70	6.42	29	< .001	0.71
Posttest Between Groups	15.60	8.31	58	< .001	2.14

The normalized N-Gain analysis (Hake, 1998) in Table 10 offers a granular view of improvement at the individual student level. The experimental group's mean N-Gain of 0.582 falls within the medium category, while the control group's 0.215 falls within the low category. The individual N-Gain distribution in the experimental group shows 43.3% of students reaching the high category ($g \geq 0.70$), 53.3% the medium category, and only 3.3% the low category. In the control group, 60.0% of students fell into the low category, none reached the high category, and 40.0% reached the medium category. This distribution indicates not only group-level effectiveness but also consistency of improvement across the range of baseline literacy ability in the experimental group, implying that the CAIDL model supports creative literary literacy development across diverse student profiles.

Table 10. Distribution of Normalized N-Gain across Both Groups

N-Gain Category	Range	Experimental <i>f</i>	Experimental %	Control <i>f</i>	Control %
High	$g \geq 0.70$	13	43.3%	0	0.0%
Medium	$0.30 \leq g < 0.70$	16	53.3%	12	40.0%
Low	$g < 0.30$	1	3.3%	18	60.0%
N-Gain Mean	(not applicable)	0.582	Medium	0.215	Low

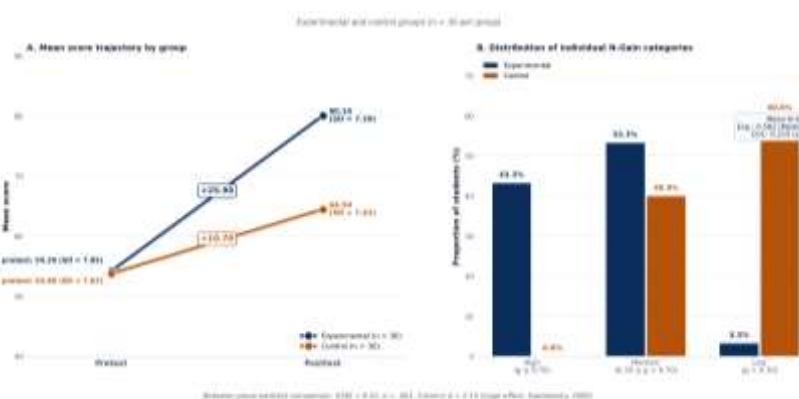


Figure 5. Slope Chart of Pretest-Posttest Trajectories and Individual N-Gain Distribution by Group

Effect size analysis was conducted for each of the eight indicators across the two target variables (four for appreciation, four for creativity) to provide construct-level diagnosis of effectiveness. Table 11 reports the results, showing that all eight indicators exceeded the conventional large-effect threshold ($d \geq 0.80$; Cohen, 1988) and fell within the very large range ($d \geq 1.20$; Sawilowsky, 2009). The largest effect appeared on creative fluency ($d = 1.71$), followed by symbolic text interpretation ($d = 1.68$), indicating that the CAIDL model strengthened students' capacity to both generate a range of narrative ideas and interpret the symbolic meaning of literary texts. Storyline flexibility ($d = 1.62$), appreciation of literary aesthetic elements ($d = 1.55$), and narrative elaboration ($d = 1.55$) also recorded very large effects. The remaining indicators, namely originality of literary expression ($d = 1.49$), identification of moral and cultural values ($d = 1.48$), and literal comprehension of literary texts ($d = 1.42$), complete the set of very large effects. Figure 6 visualizes the effect size per indicator

grouped by target variable, showing the consistent dominance of the experimental group across both the appreciative and creative dimensions.

Table 11. Effect Size per Creative Literary Literacy Indicator

No	Indicator	Variable	Exp. Gain	Ctrl. Gain	Cohen's <i>d</i>
1	Literal comprehension of literary text	Appreciation	23.50	9.80	1.42
2	Appreciation of literary aesthetic elements	Appreciation	27.10	11.20	1.55
3	Identification of moral & cultural values	Appreciation	25.40	10.50	1.48
4	Symbolic interpretation of text	Appreciation	28.30	10.90	1.68
5	Creative fluency	Creativity	28.70	11.10	1.71
6	Storyline flexibility	Creativity	27.50	10.80	1.62
7	Originality of literary expression	Creativity	25.60	10.40	1.49
8	Elaboration and narrative depth	Creativity	26.40	10.70	1.55

Source: primary analysis, 2025. Note: All values fall within the very large range ($d \geq 1.20$; Sawilowsky, 2009).

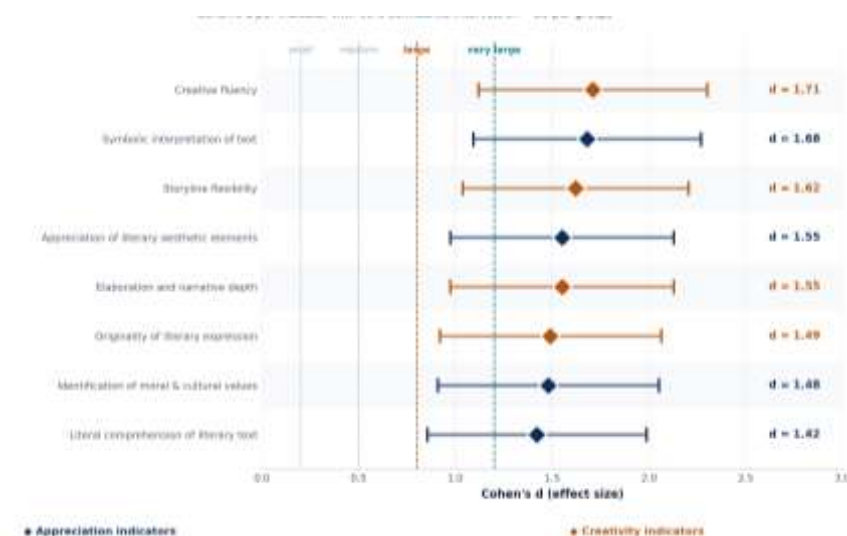


Figure 6. Forest Plot of Effect Size for the CAIDL Model across Creative Literary Literacy Indicators

B. Discussion

The 88.80% validity in the very valid category confirms that the four-pillar architecture of the CAIDL model stands as a coherent instructional theory. The framing of Melayu Deli traditional context (Sinar, 2011; Crawford, 2001), Socratic AI scaffolding (Kasneci et al., 2023; Holmes & Tuomi, 2022), deep learning pedagogy (Marton & Säljö, 1976; Biggs, 1987; Fullan & Langworthy, 2014), and the creative literary literacy formation framework (Kaufman & Beghetto, 2009; Cremin & Chappell, 2021) operate as mutually complementary rather than competing components. The highest score on pedagogical coherence (91.00%) is significant because integrative models frequently suffer internal contradictions when contextual and reflective elements are forced into a single syntax. The strong AI-integration score (89.00%) indicates that the Socratic configuration of the AI tutor met expert standards as a learning partner rather than a teacher replacement, addressing a primary concern in the generative-AI-in-education literature (Lo, 2023; Yan et al., 2024).

The 87.08% operational feasibility in the small-group trial shows that the CAIDL model can be carried out by fifth-grade elementary students with diverse literacy profiles. The highest score on the relevance of Melayu Deli tradition (89.58%) reinforces the position of Crawford (2001) and Purnamasari and Setiawan (2017) that children's literary learning improves when the texts used are connected to authentic cultural experience. The student-engagement score (88.54%) is meaningfully important because sustained engagement is a prerequisite for the iterative reflection cycle required by deep learning pedagogy (Biggs, 1987; Parsons et al., 2010). The lower score on appropriateness of cognitive challenge (83.33%) had already been anticipated and indicates that the symbolic-interpretation stage requires additional scaffolding when first introduced, a pattern documented in prior deep-learning-approach implementation studies (Sweller et al., 2019; Vansteenkiste et al., 2009). The five pragmatic revisions implemented after the small-group trial reflect the strength of the Plomp model in supporting evidence-based refinement prior to broader implementation (Plomp & Nieveen, 2013).

The very practical results from teachers (89.33%) and students (89.40%) show that the CAIDL model can be realistically integrated into the fifth-grade elementary Indonesian-language syllabus. The convergence of teachers' high score on coherence of learning stages (91.67%) and students' high score on AI-tutor usefulness (91.00%) reflects two complementary perspectives aligned with Self-Determination Theory (Ryan & Deci, 2020), in which sustained learning engagement requires the simultaneous fulfillment of the needs for autonomy, competence, and relatedness (Vansteenkiste et al., 2009). Teachers prioritized the structural integrity of the design, while students prioritized the scaffolding that helped them traverse the meaning-making stages. Both groups converged on time allocation as the lowest-scoring dimension, reflecting the structural constraint of Indonesia's relatively packed elementary school class schedule (Wahyudin et al., 2023). The resulting practical recommendation is to split the symbolic-interpretation and narrative-elaboration stages into two separate sessions in subsequent modules.

The effectiveness of CAIDL, with Cohen's $d = 2.14$ on the independent t -test and a mean N-Gain of 0.582 in the experimental group, exceeds the average effect size from meta-analyses of children's literacy interventions, which typically range from 0.30 to 0.60 (Cheung & Slavin, 2013; Castles et al., 2018). Four theoretical factors help explain this unusually large effect. First, the structured creative literary literacy concept framework provides explicit cognitive scaffolding for traversing the stages of literal comprehension, aesthetic appreciation, and symbolic interpretation, a pathway rarely made explicit in conventional literary instruction (Kaufman & Beghetto, 2009; Cremin & Chappell, 2021). Second, the AI tutor configured as a Socratic questioner sustained the reflective dialogue that prior research identifies as central to the development of appreciative reasoning (Chiu et al., 2024; Webb et al., 2019). Third, the contextual pillar situated literary work within situations familiar to Medan Deli students through the *pantun-hikayat-folktale* corpus, reducing the cognitive load required to instill new literary aesthetic concepts (Sweller et al., 2019). Fourth, the deep learning prompts in each session pushed students from procedural reproduction toward integrative meaning making, the same shift documented across four decades of deep-learning-approach research (Marton & Säljö, 1976; Biggs, 1987; Fullan & Langworthy, 2014).

The pattern of effect sizes across the eight indicators in Table 11 and Figure 6 reveals informative differences between the appreciation and creativity domains. Creative fluency ($d = 1.71$) and symbolic interpretation ($d = 1.68$) emerged as the two largest effects, indicating that the CAIDL model expanded both the quantity of narrative ideas students could generate and their capacity to interpret layers of meaning in literary texts. This dual-gain pattern rarely appears in interventions that target only one of the two domains (Amabile, 1996; Beghetto & Kaufman, 2007). The underlying mechanism is an iterative cycle between AI-tutor-mediated divergent exploration (which generates many candidate ideas) and Melayu-Deli-tradition-grounded convergent reflection (which frames those ideas within an authentic cultural context). Sawyer (2012) and Csikszentmihalyi (1996) have long argued that creativity and appreciation should be developed in tandem rather than sequentially, and the CAIDL model offers empirical confirmation of that proposition.

A finding with significant policy implications is the equalizing effect across the range of baseline literacy ability. Unlike the Matthew effect commonly observed in literacy instruction, in which high-ability students extend their advantage while low-ability students fall further behind (Castles et al., 2018), the CAIDL experimental group showed an even distribution of gains. The mechanism underlying this finding is the adaptive scaffolding enabled by the AI tutor: higher-ability students engaged with the Socratic prompts at a deeper conceptual level, while lower-ability students received a sequence of more granular questions that bridged their initial knowledge gaps (Chiu et al., 2024; Holmes & Tuomi, 2022). This finding is relevant to elementary education quality-equity policy that pursues equitable learning outcomes across student profiles, and aligns with the universal design for learning principle advocated within the Merdeka Curriculum (Kemendikbudristek, 2022; Wahyudin et al., 2023).

Several limitations of these positive findings should be transparently acknowledged. The eight-week intervention duration may not capture long-term retention, leaving open the question of whether the gains persist beyond the immediate posttest (Lo, 2023; Tlili et al., 2023); follow-up research with delayed posttests at the third and sixth month is strongly recommended. The sample was drawn from a single elementary school in Medan with a specific student profile, so generalization to elementary schools in other provinces or with different socio-economic profiles should be made with caution; multi-school replication across Indonesia's diverse local cultural contexts is an important next step. The focus on Melayu Deli literary literacy limits generalization to other literary traditions across the archipelago, such as Minangkabau *pantun*, Acehese *hikayat*, or Bugis-Makassar folktales; parallel development for those cultural contexts would strengthen the empirical base (Rahmawati & Ridwan, 2017). Elementary teachers' capacity to facilitate AI-mediated dialogue is an important moderating factor that warrants deeper investigation through qualitative or multilevel studies (Webb et al., 2019; Zulaiha et al., 2020). Finally, the ethical dimensions of using generative AI in elementary

education, including questions of academic integrity and equitable access, require sustained scholarly and institutional attention (Kasneji et al., 2023; Yan et al., 2024).

V. Conclusion

This research and development study produced a Contextual AI-Based Deep Learning (CAIDL) model grounded in Melayu Deli tradition to improve the creative literary literacy of fifth-grade elementary students, satisfying all four R&D quality criteria. First, validity reached a mean of 88.80% in the very valid category based on the assessment of seven experts, with coherence of contextual and deep learning pedagogy at 91.00%, Socratic AI integration at 89.00%, accuracy of Melayu Deli literary substance at 88.67%, indicator completeness at 88.00%, and language clarity at 87.33%. Second, operational feasibility was confirmed in the small-group trial with a mean of 87.08% in the very feasible category, supported by the relevance of Melayu Deli tradition (89.58%) and student engagement (88.54%). Third, practicality in the limited implementation reached a combined mean of 89.37%, as a combination of teacher responses (89.33%) and student responses (89.40%), both in the very practical category. Fourth, effectiveness was demonstrated through the significant improvement of the experimental group (pretest $M = 54.20$ to posttest $M = 80.10$) compared with the control group (pretest $M = 53.80$ to posttest $M = 64.50$), with an independent t -test of $t(58) = 8.31, p < .001$, Cohen's $d = 2.14$, and a mean N-Gain of 0.582 in the medium category. Per-indicator effect sizes ranged from $d = 1.42$ to $d = 1.71$, with the largest effects on creative fluency ($d = 1.71$) and symbolic text interpretation ($d = 1.68$), indicating that the CAIDL model simultaneously strengthens both the appreciative and creative dimensions of students' literary literacy. These findings confirm that integrating contextual learning grounded in Melayu Deli tradition, Socratic AI scaffolding, deep learning pedagogy, and a creative literary literacy concept framework can address the dual-deficit pattern in elementary education, while also implying the need to integrate similar models into the development of the Merdeka Curriculum's local-content- and 21st-century-competency-based approach. Further research is recommended to test long-term retention through a six-month delayed posttest, to develop parallel models for other local literary traditions across the archipelago such as Minangkabau *pantun* and Acehnese *hikayat*, and to explore the moderating role of teacher capacity in facilitating AI-mediated dialogue within the elementary school context.

Conflict of Interest

The authors declare that there is no conflict of interest related to this study.

Author Contributions

D.S.P. designed the study, developed the CAIDL model, and drafted the manuscript. Y.T.U.P. contributed to data analysis and research administration. Z.M. contributed to instrument validation and the literature review. P.W.H. contributed to field data collection. A.R. contributed to manuscript review and editing. All authors have read and approved the final version of this manuscript.

Funding

This article was funded by the Directorate of Research, Technology, and Community Service (DRTPM), Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kemdikbudristek), under Master Contract Number 280/C3/DT.05.00/PL-BARU/2026, dated January 30, 2026.

Acknowledgment

The authors would like to thank Universitas Battuta and Sekolah Tinggi Ilmu Ekonomi Al-Washliyah for their institutional support, SDN Medan Deli for its willingness to serve as an implementation partner, and the seven expert validators, three Indonesian-language teachers, and all fifth-grade students who participated in this study.

References

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Westview Press.
- Beghetto, R. A. (2014). Creative mortification: An initial exploration. *Psychology of Aesthetics, Creativity, and the Arts*, 8(3), 266–276. <https://doi.org/10.1037/a0036618>
- Beghetto, R. A., & Kaufman, J. C. (2007). Toward a broader conception of creativity: A case for "mini-c" creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 1(2), 73–79. <https://doi.org/10.1037/1931-3896.1.2.73>

- Biggs, J. B. (1987). Student approaches to learning and studying. Australian Council for Educational Research.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing reading achievement in K–12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113. <https://doi.org/10.1016/j.edurev.2013.01.001>
- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2024). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 6, 100118. <https://doi.org/10.1016/j.caeai.2023.100118>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Craft, A. (2003). The limits to creativity in education: Dilemmas for the educator. *British Journal of Educational Studies*, 51(2), 113–127. <https://doi.org/10.1111/1467-8527.t01-1-00229>
- Crawford, M. L. (2001). *Teaching contextually: Research, rationale, and techniques for improving student motivation and achievement in mathematics and science*. CCI Publishing.
- Cremin, T., & Chappell, K. (2021). Creative pedagogies: A systematic review. *Research Papers in Education*, 36(3), 299–331. <https://doi.org/10.1080/02671522.2019.1677757>
- Cremin, T., Mottram, M., Collins, F., Powell, S., & Safford, K. (2014). *Building communities of engaged readers: Reading for pleasure*. Routledge. <https://doi.org/10.4324/9781315772585>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal*, 18(3), 391–404. https://doi.org/10.1207/s15326934crj1803_13
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. Harper Collins.
- Dubinsky, E. (1991). Reflective abstraction in advanced mathematical thinking. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 95–126). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47203-1_7
- Fauzi, A., & Pradipta, R. F. (2018). Research methods and data analysis techniques in education articles published by Indonesian biology educational journals. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 4(2), 123–134. <https://doi.org/10.22219/jpbi.v4i2.5889>
- Fullan, M., & Langworthy, M. (2014). *A rich seam: How new pedagogies find deep learning*. Pearson.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <https://doi.org/10.1111/ejed.12533>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four C model of creativity. *Review of General Psychology*, 13(1), 1–12. <https://doi.org/10.1037/a0013688>
- Kemendikbudristek. (2022). *Kurikulum Merdeka: Kerangka dasar dan struktur kurikulum pendidikan dasar dan menengah* [Merdeka Curriculum: Basic framework and structure of the primary and secondary education curriculum].
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. <https://doi.org/10.3390/educsci13040410>
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46(1), 4–11. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading. TIMSS & PIRLS International Study Center, Boston College*.
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>

- Parmin, P., Sajidan, S., Ashadi, A., Sutikno, S., & Maretta, Y. A. (2016). Preparing prospective teachers in integrating science and local wisdom through practicing open inquiry. *Journal of Turkish Science Education*, 13(2), 3–14. <https://doi.org/10.12973/tused.10163a>
- Parsons, S., Reid, A., & Belski, I. (2010). Engaging beginning readers in the formal study of literature through a play-based instructional approach. *Reading Research Quarterly*, 45(2), 156–179. <https://doi.org/10.1598/RRQ.45.2.2>
- Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *Educational design research, Part A: An introduction* (pp. 10–51). SLO.
- Plomp, T., & Nieveen, N. (Eds.). (2013). *Educational design research, Part B: Illustrative cases*. SLO.
- Pratiwi, I., & Mahdiansyah. (2017). The impact of school environment on Indonesian student literacy achievement: Analysis of PISA 2015 data. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(2), 175–187. <https://doi.org/10.21831/pep.v21i2.13851>
- Purnamasari, I., & Setiawan, A. (2017). Local wisdom-based learning material development in elementary schools. *Journal of Education and Learning (EduLearn)*, 11(2), 199–207. <https://doi.org/10.11591/edulearn.v11i2.5829>
- Rahmawati, Y., & Ridwan, A. (2017). Empowering students' chemistry learning: The integration of ethnochemistry in culturally responsive teaching. *Bulgarian Journal of Science and Education Policy*, 11(2), 246–262. <https://doi.org/10.36681/tused.2020.30>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599. <https://doi.org/10.22237/jmasm/1257035100>
- Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation* (2nd ed.). Oxford University Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Sinar, T. L. (2011). *Sejarah Medan tempo doeloe [The history of old Medan]* (10th ed.). Lembaga Penelitian dan Pengembangan Seni Budaya Melayu.
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D [Quantitative, qualitative, and R&D research methods]* (26th ed.). Alfabeta.
- Suryani, N., Sudiyanto, S., & Hartati, S. (2018). Local wisdom-based teaching materials and the effectiveness in improving students' critical thinking. *Cakrawala Pendidikan*, 37(3), 411–423. <https://doi.org/10.21831/cp.v38i3.21097>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Toivanen, T., Halkilahti, L., & Ruismäki, H. (2013). Creative pedagogy — Supporting children's creativity through drama. *The European Journal of Social & Behavioural Sciences*, 7(4), 1168–1179. <https://doi.org/10.15405/ejsbs.97>
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology*, 101(3), 671–688. <https://doi.org/10.1037/a0015083>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wahyudin, D., Subkhan, E., Malik, A., Hakam, K. A., Sutarsih, C., Rusman, R., ... Dewi, L. (2023). Curriculum reform in Indonesia: Moving from an era of competence to character education. *Cakrawala Pendidikan*, 42(3), 599–613. <https://doi.org/10.21831/cp.v42i3.55326>

- Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., & Zimmerman, J. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *International Journal of Educational Research*, 97, 176–186. <https://doi.org/10.1016/j.ijer.2017.07.009>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... 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>
- Zulaiha, S., Mulyono, H., & Ambarwati, E. K. (2020). Exploring junior high school EFL teachers' training needs of assessment literacy. *Cogent Education*, 7(1), 1772943. <https://doi.org/10.1080/2331186X.2020.1772943>



Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.