NUSRA: Jurnal Penelitian dan Ilmu Pendidikan
Volume 7, Issue 3, August 2026
DOI: https://doi.org/10.55681/10.55681/nusra.v7i3.6508
Homepage: ejournal.nusantaraglobal.ac.id/index.php/nusra

p-ISSN: 2715-114X
e-ISSN: 2723-4649
pp. 1998-2009

PEER ASSESSMENT OF STUDENT PERFORMANCE IN DESIGNING UNPLUGGED CODING WORKSHEETS BASED ON ENVIRONMENTAL ISSUES: A PLS-SEM ANALYSIS

Nenden Permas Hikmatunisa*, Fitri Nuraeni, Afridha Laily Alindra, Hafiziani Eka Putri, Tiara Yogiarni, Neneng Sri Wulan, Hisny Fajrussalam.
Program Studi Pendidikan Guru Sekolah Dasar, Universitas Pendidikan Indonesia, Indonesia
**Corresponding author email: nendenpermas17@upi.edu*

Article History

Received: 26 May 2026
Revised: 6 August 2026
Published: 22 August 2026

ABSTRACT

This study examines the structure and predictors of student performance in designing unplugged coding student worksheets integrated with environmental issues in a coding instruction course. This study employed a peer assessment approach to examine the interrelationships among Cognitive Engagement (CE), Behavioural Engagement (BE), Affective Engagement (AE), Academic Responsibility (AR), Collaboration (COL), and Communication (COM), drawing upon theories of student engagement, cooperative learning, computational thinking, peer assessment, and self-regulated learning. A total of 100 pre-service elementary school teachers enrolled in the course participated as respondents. Data were collected using a 12-item Likert-scale peer assessment rubric and analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM) via SmartPLS4. Results of the outer model confirmed convergent validity (outer loadings ≥ 0.70; AVE ≥ 0.50) and composite reliability (CR ≥ 0.843). The inner model revealed that Affective Engagement significantly predicted Academic Responsibility ($\beta = 0.606$, $T = 3.845$, $p < 0.001$) with a large effect size ($f^2 = 0.375$), and Behavioral Engagement significantly predicted Collaboration ($\beta = 0.571$, $T = 3.407$, $p = 0.001$) with the largest effect size ($f^2 = 0.598$). The model explained 55.2%, 72.7%, and 67.3% of the variance in Academic Responsibility, Collaboration, and Communication, respectively, with a GoF index of 0.78, indicating a strong overall model fit. These findings highlight the pivotal role of affective and behavioural dimensions of engagement in shaping collaborative competence and academic responsibility among pre-service teachers engaged in computational thinking design tasks.

Keywords: *PLS-SEM; Computational Thinking; Environmental Issues; Pre-Service Teacher Education.*

How to cite: Hikmatunisa, N.P., Nuraeni, F., Alindra, A. L., Putri, H. E., Yogiarni, T., Wulan, N. S., & Fajrussalam, H. (2026). PEER ASSESSMENT OF STUDENT PERFORMANCE IN DESIGNING UNPLUGGED CODING WORKSHEETS BASED ON ENVIRONMENTAL ISSUES: A PLS-SEM ANALYSIS. *NUSRA: Jurnal Penelitian Dan Ilmu Pendidikan*, *7* (3), 1998–2009. https://doi.org/10.55681/nusra.v7i3.6508

BACKGROUND

The integration of computational thinking (CT) into elementary school curricula has emerged as a global educational priority, with numerous countries embedding coding instruction across primary levels (Wing, 2006; CSTA, 2017). In Indonesia, the introduction of coding instruction in the elementary school curriculum has placed considerable demand on pre-service teacher education programs to equip future educators with both pedagogical and technical competencies in CT-based learning design. One significant challenge is preparing pre-service teachers to design high-quality learning materials; specifically, student worksheets that integrate unplugged coding activities with contextually relevant themes such as environmental issues.

Unplugged coding refers to activities that develop computational thinking skills without digital devices, relying instead on physical, embodied, and collaborative tasks (Bell et al., 2009; Wang, Fan & Yu, 2023). Meta-analytic evidence indicates that unplugged coding activities yield large effects on CT skill development (Park et al., 2026), making them particularly appropriate for elementary school contexts where device access may be limited. Integrating environmental issues into unplugged coding further aligns with global sustainability education goals and the environmental literacy framework proposed by Hollweg et al. (2011), which encompasses ecological knowledge, cognitive skills, affective dispositions, and responsible environmental behavior.

A critical component of authentic learning in higher education is peer assessment—a formative evaluation approach in which learners assess each other's performance using structured criteria (Falchikov, 2005). Peer assessment has been shown to enhance metacognitive awareness, collaborative responsibility, and quality of academic output, particularly when operationalized through rubric-based instruments anchored in validated theoretical frameworks (Topping, 2009; Li, 2016). In the context of group-based worksheet design tasks, peer assessment provides a window into multidimensional student performance that instructor assessment may miss.

The theoretical framework of the peer assessment rubric employed in this study draws on five foundational frameworks: Fredricks, Blumenfeld, and Paris's (2004) tripartite model of school engagement (behavioral, cognitive, and affective dimensions); Zimmerman's (2000) self-regulated learning model emphasizing academic responsibility; Johnson and Johnson's (2009) social interdependence theory underpinning collaborative learning; Falchikov's (2005) peer assessment in higher education; and Wing's (2006) computational thinking dimensions. Together, these frameworks yield a 12-item instrument measuring six latent constructs: Cognitive Engagement, Behavioural Engagement, Affective Engagement, Academic Responsibility, Collaboration, and Communication.

Despite the growing body of literature on peer assessment and CT education independently, few studies have empirically examined the structural relationships among engagement dimensions, collaboration, and communication competencies within a CT-based design task context. Furthermore, the application of PLS-SEM to analyze peer assessment data in pre-service teacher education for coding instruction remains understudied. Addressing these gaps, this study seeks to map the predictive pathways among engagement constructs and performance outcomes using PLS-SEM, a method well-suited for exploratory structural analysis with latent variables (Hair et al., 2019).

This article contributes to the literature in three ways: (1) it introduces and validates a theory-driven, rubric-based peer assessment instrument for CT-integrated worksheet design tasks; (2) it examines how affective engagement, behavioral engagement, and cognitive engagement influence academic responsibility, collaboration, and communication among pre-service teachers; and (3) it provides empirical evidence to guide instructional design in coding education courses for elementary school teacher preparation programs.

Research Questions and Hypotheses

Based on the theoretical frameworks reviewed above, this study addresses the following research questions:

RQ1. Do the measurement instruments (outer model) for peer assessment of worksheet design performance demonstrate adequate convergent validity, discriminant validity, and reliability?

RQ2. What are the structural relationships and predictive pathways among Cognitive Engagement, Behavioral Engagement, Affective Engagement, Academic Responsibility, Collaboration, and Communication in the inner model?

RQ3. Which engagement dimensions serve as the strongest predictors of collaborative and academic performance outcomes in the Worksheet design task?

Table 1. Research Hypotheses and Theoretical Bases

H#	Hypothesized Path	Theoretical Basis
H1	Affective Engagement → Academic Responsibility (+)	Fredricks et al. (2004); Zimmerman (2000)
H2	Affective Engagement → Collaboration (+)	Fredricks et al. (2004); Johnson & Johnson (2009)
H3	Affective Engagement → Communication (+)	Fredricks et al. (2004); Falchikov (2005)
H4	Academic Responsibility → Collaboration (+)	Zimmerman (2000); Johnson & Johnson (2009)
H5	Academic Responsibility → Communication (+)	Zimmerman (2000); Falchikov (2005)
H6	Cognitive Engagement → Academic Responsibility (+)	Fredricks et al. (2004); Wing (2006)
H7	Cognitive Engagement → Collaboration (+)	Fredricks et al. (2004); Johnson & Johnson (2009)
H8	Cognitive Engagement → Communication (+)	Fredricks et al. (2004)
H9	Behavioral Engagement → Academic Responsibility (+)	Fredricks et al. (2004); Zimmerman (2000)
H10	Behavioral Engagement → Collaboration (+)	Fredricks et al. (2004); Johnson & Johnson (2009)
H11	Behavioral Engagement → Communication (+)	Fredricks et al. (2004)

RESEARCH METHODOLOGY

Research Design and Participants

This study employed a quantitative, survey-based cross-sectional design using PLS-SEM for data analysis. Participants were 100 pre-service elementary school teachers enrolled in the Coding Instruction for Elementary Schools course at a state university in Indonesia. All participants engaged in a structured group-based task requiring the collaborative design of two worksheets for unplugged coding activities contextualized within environmental issues (e.g., water pollution, deforestation, waste management). Groups consisted of 6–8 members, and each member assessed the performance of all peers within their group using the structured rubric.

A sample of 100 is considered adequate for PLS-SEM analysis, exceeding the minimum recommended by Hair et al. (2019) based on the 10-times rule (maximum formative indicators or paths to any construct × 10), which in this model yields a minimum n = 40. Ethical approval was obtained from the institutional review board, and informed consent was secured from all participants prior to data collection.

Instrumentation: Peer Assessment Rubric

The peer assessment instrument consisted of 12 items across six theoretically derived dimensions: Cognitive Engagement (KK; items KK1, KK2, KK3), Behavioral Engagement (KP; items KP4, KP5, KP6), Affective Engagement (KA; items KAF7, KAF8), Academic Responsibility (KAK; item KAK12), Collaboration (KOL; items KOL9, KOL10), and Communication (KOM; item KOM11). Items were rated on a 4-point Likert scale (1 = Poor, 2 = Adequate, 3 = Good, 4 = Excellent). Each item was accompanied by detailed behavioral anchor descriptors to ensure inter-rater consistency. The instrument was developed from and validated against the five theoretical frameworks described in Section 2, following the expert judgment and construct validation protocols recommended by Falchikov (2005) and Ung et al. (2021).

Analytical Approach: PLS-SEM

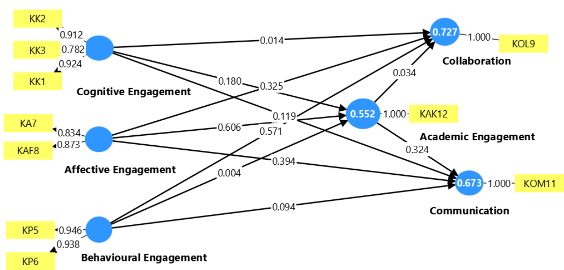
Data were analyzed using SmartPLS 4.0 (Ringle, Wende & Becker, 2022) following the two-stage assessment procedure recommended by Hair et al. (2019): (1) outer model evaluation for reliability and validity, and (2) inner model evaluation for structural path significance and explanatory power. The bootstrapping procedure (5,000 subsamples, two-tailed, bias-corrected confidence intervals) was used for hypothesis testing. Model fit was assessed using the SRMR (< 0.10 threshold), NFI, and the GoF index. Effect sizes were interpreted using Cohen's (1988) benchmarks: $f^2 \geq 0.02$ (small), ≥ 0.15 (medium), ≥ 0.35 (large).

RESULT

Descriptive Statistics

In the present study, inferential analysis was conducted using SmartPLS version 4, which implements the Partial Least Squares Structural Equation Modeling (PLS-SEM) approach. This method enables simultaneous examination of the measurement model (outer model) and the structural model (inner model), making it particularly suitable for complex relational structures among latent variables.

Figure 1. Graphical Output



The analysis proceeded through two principal stages: the Outer Model Assessment and the Inner Model Assessment. The Outer Model evaluates the validity and reliability of the measurement indicators used to operationalize each latent construct, encompassing Convergent Validity, Discriminant Validity, and Construct Reliability tests. The Inner Model, conversely, assesses the structural relationships among latent variables, including explanatory power (R^2), path significance, and effect size (f^2).

Among the 100 respondents, the majority of peer assessment scores clustered in the 'Good' (3) and 'Excellent' (4) categories across all 12 items, indicating generally high perceived performance by peers during the worksheet design process. Items belonging to Behavioral Engagement (KP4–KP6) showed the highest mean ratings, while a subset of respondents (approximately 8%) received scores of (2) 'adequate' on items related to Cognitive Engagement and Affective Engagement, suggesting moderate variability in self-regulated and cognitively demanding dimensions of the task.

Outer Model Evaluation

1. Convergent Validity.

Convergent validity was assessed through outer loadings and Average Variance Extracted (AVE). All retained indicators demonstrated outer loadings ≥ 0.70, meeting the threshold recommended by Hamid and Anwar (2019) and Hair et al. (2019). The full matrix of outer loadings is presented in Table 2.

Table 2. Outer Loadings Matrix (Convergent Validity — Stage 2 Final Model). All loadings ≥ 0.70.

Item	Affective Eng.	Academic Resp.	Cognitive Eng.	Behavioral Eng.	Coll	Comm.
KA7	0.847*	—	—	—	—	—
KAF8	0.862*	—	—	—	—	—
KAK12	—	1.000*	—	—	—	—
KK1	—	—	0.928*	—	—	—
KK2	—	—	0.915*	—	—	—
KK3	—	—	0.774*	—	—	—
KOL9	—	—	—	—	1.000*	—
KOM11	—	—	—	—	—	1.000*
KP5	—	—	—	0.946*	—	—
KP6	—	—	—	0.938*	—	—

Note: Two items were removed during the iterative CL evaluation: KOL10 (cross-loading > own loading; Stage 1 discriminant validity failure) and KP4 (cross-loading to Cognitive Engagement exceeded its loading on Behavioral Engagement). The final model retained 10 indicators across 6 constructs.

AVE values for all constructs exceeded the 0.50 threshold (Fornell & Larcker, 1981): Affective Engagement = 0.730; Cognitive Engagement = 0.766; Behavioral Engagement = 0.887 (updated after item removal); Collaboration = 1.000 (single item); Communication = 1.000 (single item). These values confirm that each construct explains more than half of its indicator variance, satisfying convergent validity.

2. Discriminant Validity.

Discriminant validity was assessed using the cross-loadings criterion (Haryono, 2016) and the Fornell-Larcker criterion (1981). After removing KOL10 and KP4 in Stage 1, all remaining indicators loaded higher on their own construct than on any other construct in Stage 2 (Table 3). The Fornell-Larcker matrix (Table 4) confirms that the square root of each AVE (diagonal) exceeds all inter-construct correlations, indicating adequate discriminant validity.

Table 3. Fornell-Larcker Criterion Matrix. Diagonal values (green) = √AVE; off-diagonal = inter-construct correlations.

Construct	AE (KA)	AR (KAK)	CE (KK)	BE (KP)	COL (KOL)	COM (KOM)
Affective Eng. (AE)	0.854	0.731	0.682	0.651	0.731	0.773
Academic Resp. (AR)	0.731	1.000	0.595	0.515	0.573	0.731
Cognitive Eng. (CE)	0.682	0.595	0.875	0.646	0.624	0.641
Behavioral Eng. (BE)	0.651	0.515	0.646	0.942	0.809	0.594
Collaboration (COL)	0.731	0.573	0.624	0.809	1.000	0.661
Communication (COM)	0.773	0.731	0.641	0.594	0.661	1.000

3. Reliability.

Internal consistency was evaluated using Cronbach's alpha (α) and composite reliability (CR; ρc). For multi-item constructs, Cognitive Engagement showed α = 0.844 and CR = 0.907; Behavioral Engagement showed α = 0.872 and CR = 0.940; Affective Engagement showed α = 0.630 and CR = 0.843. While the alpha for Affective Engagement falls slightly below the conventional 0.70 threshold, the composite reliability of 0.843 satisfies the minimum of 0.70 (Hair et al., 2019), and the two-item limitation of this construct is acknowledged as a constraint. All three multi-item constructs demonstrate acceptable to strong reliability.

Table 4. Reliability and AVE Summary.

Construct	Cronbach's α	CR (ρa)	CR (ρc)	AVE
Affective Engagement	0.630	0.636	0.843	0.729
Cognitive Engagement	0.844	0.852	0.907	0.766
Behavioral Engagement	0.872	0.875	0.940	0.887
Academic Resp.	— (single)	—	1.000	1.000
Collaboration	— (single)	—	1.000	1.000
Communication	— (single)	—	1.000	1.000

Note: Green = meets threshold (CR ≥ 0.70; AVE ≥ 0.50).

Model Fit

The overall model fit was evaluated using the SRMR, NFI, and GoF indices. The SRMR value of 0.104 marginally exceeds the strict threshold of 0.10 but falls within the acceptable range of 0.10–0.12 commonly reported for complex PLS-SEM models (Henseler et al., 2015). The NFI of 0.697 falls below the 0.90 ideal threshold, which is expected given that NFI is sensitive to sample size and model complexity (Hair et al., 2019). Importantly, the GoF index of 0.78 substantially exceeds the large-fit threshold of 0.36 (Wetzels, Odekerken-Schröder & Van Oppen, 2009), indicating strong overall predictive relevance of the model.

Table 5. Model Fit Indices.

Index	Saturated Model	Estimated Model	Interpretation
SRMR	0.104	0.104	Marginally acceptable (< 0.12)
NFI	0.697	0.696	Below ideal; expected for complex models
GoF	—	0.780	Strong fit (threshold > 0.36)

Inner Model: Predictive Relevance (R^2 and f^2)

The inner model was evaluated through R^2 and adjusted R^2 values for endogenous constructs, and f^2 effect sizes for each path. The model explained substantial variance in all three endogenous constructs:

Table 6. R^2 Values for Endogenous Constructs.

Endogenous Construct	R^2	R^2 (Adj.)	Interpretation (Cohen, 1988)
Academic Responsibility (AR) / (KAK)	0.552	0.538	Moderate-to-substantial
Collaboration (COL) / (KOL)	0.727	0.716	Substantial
Communication (COM) / (KOM)	0.673	0.660	Substantial

Effect size analysis (f^2) revealed that Behavioral Engagement → Collaboration had by far the largest effect ($f^2 = 0.598$), classifying as a large effect. Affective Engagement → Academic Responsibility also yielded a large effect ($f^2 = 0.375$). All remaining paths produced small-to-negligible effects (f^2 = 0.000–0.158).

Hypothesis Testing Results

Hypothesis testing was conducted using bootstrapping (5,000 subsamples, two-tailed). Results are summarized in Table 7. There are 11 hypothesized paths; two were statistically supported: H1 (Affective Engagement → Academic Responsibility: $\beta = 0.606$, $T = 3.845$, $p < 0.001$) and H10 (Behavioral Engagement → Collaboration: $\beta = 0.571$, $T = 3.407$, $p = 0.001$). All remaining nine paths did not reach statistical significance at the $\alpha = 0.05$ level.

Table 7. Hypothesis Testing Results

H	Path	β (O)	M	STDEV	T stat.	p-val.	Decision
H1	AE → AR	0.606	0.578	0.158	3.845	0.000	Supported***
H2	AE → COL	0.325	0.317	0.174	1.870	0.062	Not Supported
H3	AE → COM	0.394	0.392	0.257	1.530	0.126	Not Supported
H4	AR → COL	0.034	0.043	0.130	0.258	0.796	Not Supported
H5	AR → COM	0.324	0.305	0.179	1.810	0.070	Not Supported
H6	CE → AR	0.180	0.170	0.154	1.163	0.245	Not Supported
H7	CE → COL	0.014	0.020	0.159	0.085	0.932	Not Supported
H8	CE → COM	0.119	0.116	0.143	0.834	0.404	Not Supported
H9	BE → AR	0.004	0.017	0.164	0.026	0.980	Not Supported
H10	BE → COL	0.571	0.549	0.168	3.407	0.001	Supported***
H11	BE → COM	0.094	0.103	0.175	0.535	0.593	Not Supported

Note: (Two-tailed Bootstrapping, n = 5,000). *** $p < 0.001$. AE = Affective Engagement, AR = Academic Responsibility, CE = Cognitive Engagement, BE = Behavioral Engagement, COL = Collaboration, COM = Communication.

DISCUSSION

Affective Engagement as a Strong Driver of Academic Responsibility

The most striking finding of this study is the large and significant effect of Affective Engagement on Academic Responsibility ($\beta = 0.606$, $f^2 = 0.375$, $p < 0.001$). This result is consistent with Fredricks et al.'s (2004) theorization that affective engagement—characterized by enthusiasm, sense of belonging, and valuing of the task—creates motivational momentum that translates into responsible academic behavior. In the context of worksheet design, students who exhibited high enthusiasm for developing unplugged coding activities contextualized in environmental issues were also those rated by peers as maintaining higher content accuracy and quality vigilance.

This finding resonates with Zimmerman's (2000) SRL framework: intrinsic motivation (a hallmark of affective engagement) facilitates the self-monitoring and self-evaluation processes that underpin academic responsibility. The environmental issue context appears to have amplified this relationship—when students personally care about environmental problems, their affective investment in accurately representing those issues in pedagogical materials naturally intensifies. This is consistent with Hollweg et al.'s (2011) observation that environmental affective dispositions are a prerequisite to responsible environmental behavior, here manifested as responsible academic behavior in worksheet design.

The explained variance in Academic Responsibility ($R^2 = 0.552$) indicates that the combination of engagement constructs accounts for over half of the variability in this outcome, a moderate-to-substantial level of predictive power. This suggests that the remaining variance is attributable to individual differences, group-level dynamics, or instructor-mediated factors not captured in the present model.

Behavioral Engagement as the Dominant Predictor of Collaboration

The second supported hypothesis—Behavioral Engagement → Collaboration ($\beta = 0.571$, $f^2 = 0.598$, $p = 0.001$)—yielded the largest effect size in the model, explaining a dominant portion of the 72.7% variance in Collaboration. This finding powerfully corroborates Johnson and Johnson's (2009) Social Interdependence Theory: students who consistently attended discussions, completed assigned tasks on time, and actively participated in worksheet simulation and revision processes were rated by peers as significantly more collaborative. Behavioral engagement, as measured here through attendance, task completion, and simulation involvement, represents the visible enactment of positive interdependence—the mechanism through which SIT predicts collaborative success.

The magnitude of this effect ($f^2 = 0.598$, classified as large by Cohen's (1988) benchmarks) is noteworthy and suggests that behavioral engagement is not merely one of many contributors to collaboration but is its primary behavioral foundation. This is particularly meaningful in the unplugged coding context, where physical, hands-on task execution (simulating algorithms, testing worksheet steps with peers) is the primary medium of collaborative learning. The findings align with (Zhang et al., 2024) meta-analytic observation that collaborative implementations of unplugged activities produce stronger CT outcomes, likely because the behavioral co-engagement required in physical task execution is itself a driver of collaborative learning.

Interestingly, neither Cognitive Engagement nor Academic Responsibility significantly predicted Collaboration. This suggests that the collaborative performance evaluated by peers in the present context is primarily driven by observable, present-in-the-room behavior rather than intellectual ideation or conscientiousness about quality. Peer raters may find behavioral indicators more salient and observable, lending them greater discriminating power in peer ratings.

Non-Significant Paths: Theoretical and Methodological Considerations

Nine of eleven hypothesized paths did not achieve statistical significance. Several theoretical and methodological explanations merit consideration. First, the non-significance of Affective Engagement → Collaboration ($p = 0.062$) and Affective Engagement → Communication ($p = 0.126$) may reflect the primacy of behavioral engagement in driving these outcomes: when behavioral engagement is in the model alongside affective engagement, it may absorb much of the explanatory variance in collaboration, rendering affective engagement non-significant by comparison—a suppression or redundancy effect not uncommon in engagement research (Skinner et al., 2009).

Second, the near-zero and non-significant path from Cognitive Engagement → Collaboration ($\beta = 0.014$, $p = 0.932$) challenges the assumption that intellectual contribution translates directly into peer-perceived collaboration. This may reflect a measurement issue: the single remaining indicator for Collaboration (KOL9, measuring cooperative task division and flexibility) captures cooperative behavior rather than the cognitive quality of contributions, thus diverging from what cognitive engagement most directly influences. Future research might employ a more granular multi-item collaboration scale to detect cognitive contributions to collaborative quality.

Third, the non-significance of Academic Responsibility → Communication ($p = 0.070$) is marginally outside the $\alpha = 0.05$ threshold and may achieve significance in larger samples. The positive and moderate path coefficient ($\beta = 0.324$) is theoretically sensible and suggests a trend worth investigating with increased statistical power.

Fourth, the non-significance of all Behavioral Engagement → Communication ($p = 0.593$) and Behavioral Engagement → Academic Responsibility ($p = 0.980$) paths reflects the specificity of behavioral engagement's role: its strong effect is channelled specifically through collaboration rather than diffused across all performance outcomes. This selective effect pattern suggests that behavioral engagement may be most salient precisely in the tasks requiring physical co-participation—like unplugged simulation—rather than in individual accountability or communicative performance.

Implications for Pre-Service Teacher Education

The empirical structure of student engagement in worksheet design tasks has several actionable implications for instructors in coding education courses. First, instructors should strategically cultivate affective engagement early in the course—through the selection of environmental issues that resonate personally with students, exposure to authentic local environmental contexts, and affective scaffolding through reflective activities—to build the motivational foundation for academic responsibility.

Second, the primacy of behavioral engagement in predicting collaboration suggests that attendance and active participation policies should be enforced, and that simulation/testing phases of worksheet design should be mandatory, structured, and graded rather than optional.

Third, the finding that cognitive engagement did not significantly predict collaboration or academic responsibility invites instructors to reconsider how cognitive contributions are made visible in group processes—through think-aloud protocols, contribution logs, or structured ideation sessions that make individual cognitive contributions attributable and assessable.

Fourth, the rubric and PLS-SEM structure validated here can serve as a replicable assessment framework for peer evaluation in CT-based design tasks across similar courses in Indonesia and internationally.

Limitations and Future Directions

This study has several limitations that should be acknowledged. First, the use of single-indicator constructs for Academic Responsibility, Collaboration, and Communication (after item removal) limits the depth of measurement and precludes composite reliability calculation for these constructs. Future work should develop multi-item operationalizations validated through confirmatory factor analysis. Second, the marginally elevated SRMR (0.104) and modest NFI (0.697) indicate that the structural model, while satisfactory by GoF criteria, may benefit from additional indicators or refined pathway specifications in replication studies. Third, the sample is drawn from a single institution and course, limiting generalizability; multi-institutional replication is recommended.

Fourth, the peer assessment ratings may be subject to social desirability bias and peer leniency effects documented in the peer assessment literature (Falchikov, 2005; Li, 2016), particularly in cohesive groups. Triangulation with instructor assessment, product quality ratings of the worksheet outputs, and objective engagement trace data (e.g., participation logs) would strengthen construct validity in future studies. Fifth, the cross-sectional design precludes causal inference; longitudinal studies tracking engagement trajectories across the worksheet design process would allow more dynamic modelling of engagement-performance relationships.

CONCLUSION

This study examined the structural determinants of pre-service teacher performance in designing unplugged coding worksheets integrated with environmental issues, using peer assessment data from 100 participants analyzed through PLS-SEM. The validated measurement model confirmed adequate convergent validity, discriminant validity, and reliability for the six engagement constructs derived from the theoretical frameworks of Fredricks et al. (2004), Zimmerman (2000), Johnson and Johnson (2009), Falchikov (2005), and Wing (2006).

Two significant structural pathways were identified: Affective Engagement strongly predicted Academic Responsibility ($\beta = 0.606$, $f^2 = 0.375$), and Behavioral Engagement was the dominant predictor of Collaboration ($\beta = 0.571$, $f^2 = 0.598$). These findings emphasize that in CT-based collaborative design tasks, the emotional investment students bring to environmental problem-solving and the observable behavioral participation they exhibit in group work are the most consequential dimensions of student engagement for performance outcomes. The model explained 55.2%, 72.7%, and 67.3% of variance in Academic Responsibility, Collaboration, and Communication, respectively, with a GoF of 0.78 indicating strong predictive relevance.

The study contributes a theoretically grounded, empirically validated peer assessment framework for pre-service teacher education in coding instruction, with practical implications for instructional design and assessment practice in elementary school teacher preparation programs globally. Future research should address measurement depth, multi-institutional replication, and longitudinal designs to extend and consolidate these findings.

REFERENCES

Bell, T., Witten, I. H., & Fellows, M. (2009). CS Unplugged: An enrichment and extension programme for primary-aged students. University of Canterbury.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Computer Science Teachers Association (CSTA). (2017). CSTA K-12 computer science standards. CSTA.

Falchikov, N. (2005). Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education. Routledge.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European Business Review, 31(1), 2–24.

https://doi.org/10.1108/EBR-11-2018-0203

Hamid, R., & Anwar, S. M. (2019). Partial least square structural equation modeling (PLS-SEM) using SmartPLS 3. Journal of Physics: Conference Series, 1372(1), 012019.

Haryono, S. (2016). Metode SEM untuk penelitian manajemen dengan AMOS LISREL PLS. Intermedia Personalia Utama.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135.

Hollweg, K. S., Taylor, J., Bybee, R. W., Marcinkowski, T. J., McBeth, W. C., & Zoido, P. (2011). Developing a framework for assessing environmental literacy. NAAEE.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. Educational Researcher, 38(5), 365–379.

Li, H. (2016). Peer assessment. In C. Gipps & B. McCallum (Eds.), Enhancing student-centered learning. Routledge.

Park, H., Brannock, E., Im, T., Jin, W., Shin, S., Xu, X., & Kerven, D. (2026). CS Unplugged in Gateway Computing Courses: A Collaborative, Active Learning Approach in Introductory Computing. Proceedings of the 57th ACM Technical Symposium on Computer Science Education V.1, 811–816. https://doi.org/10.1145/3770762.3772672

Ringle, C. M., Wende, S., & Becker, J.-M. (2022). SmartPLS 4. SmartPLS GmbH.

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2009). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? Journal of Educational Psychology, 100(4), 765–781.

Topping, K. J. (2009). Peer assessment. Theory Into Practice, 48(1), 20–27.

Ung, L.-L., Labadin, J., & Nizam, S. (2021). Development of a rubric to assess computational thinking skills among primary school students in Malaysia. ESTEEM Academic Journal, 17, 11–22.

Wang, J., Fan, W., & Yu, J. (2023). Fostering computational thinking through unplugged activities: A systematic literature review and meta-analysis. International Journal of STEM Education, 10(1), 47.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. MIS Quarterly, 33(1), 177–195.

Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35.

Zhang, Y., Liang, Y., Tian, X., & Yu, X. (2024). The effects of unplugged programming activities on K-9 students' computational thinking: Meta-analysis. Educational Technology Research and Development, 72(3), 1331–1356.

https://doi.org/10.1007/s11423-023-10339-5

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13–39). Academic Press.