

Content Validity and Inter-Rater Reliability of Computational Thinking and Green Computing Awareness Instruments in Software-Based Statistics Learning

Reny Refitaningsih Peby Ria¹, Kurnia Hidayati¹, Marsanda Eka Nurjayanti¹

¹Universitas Islam Negeri Kiai Ageng Muhammad Besari Ponorogo, Indonesia

ABSTRACT

Purpose – Computational Thinking (CT) and Green Computing Awareness (GCA) have thus far been studied separately and have rarely been developed within the same learning context; therefore, instruments that can assess both constructs within a shared technology-based Applied Statistics learning environment are needed. This study was an instrument development research project using a quantitative psychometric approach aimed at developing instruments for CT and GCA and evaluating their content validity and inter-rater agreement among experts.

Methods – The CT instrument comprises 12 case-study-based essay items covering abstraction, decomposition, pattern recognition, and algorithmic thinking. The GCA instrument comprises 10 questionnaire items covering environmental impact awareness, perceptions of efficiency and sustainability, and responsible digital behavior. Content validity was assessed through expert judgment involving two experts for each instrument using Aiken's V, while expert-rating consistency was assessed using the Intraclass Correlation Coefficient (ICC).

Findings – The CT instrument obtained an average Aiken's V of 0.958 and an ICC of 1.000, indicating excellent content validity and consistency among expert ratings. The GCA instrument obtained an average Aiken's V of 0.925 and an ICC of 0.561, indicating excellent content validity and moderate expert-rating consistency.

Research Implications – The findings provide initial evidence supporting further development of the instruments for technology-based Applied Statistics learning. Testing with student respondents is required to establish empirical validity and reliability before broader implementation.

 OPEN ACCESS

ARTICLE HISTORY

Received: 04-09-2026

Revised: 16-09-2026

Accepted: 05-10-2026

KEYWORDS

computational thinking, content validity, green computing awareness, inter-rater reliability, software-based statistics learning

Corresponding Author:

Reny Refitaningsih Peby Ria

Universitas Islam Negeri Kiai Ageng Muhammad Besari Ponorogo, Indonesia

Email: reny.refitaningsih@uinponorogo.ac.id

Introduction

Digital transformation in 21st-century education has expanded competency requirements beyond technical proficiency toward higher-order problem-solving and environmentally responsible technology use. Computational Thinking (CT) is increasingly recognized as a fundamental skill for addressing complex problems in digital learning, while growing computational demands have raised concerns about the environmental impact of technology use, highlighting the importance of Green Computing Awareness (GCA) as a relevant competency (Fidhyallah & Zakiah, 2025; Kang, 2024; Rina et al., 2025; Kee & Zhang, 2022; Obafemi et al., 2023; Tariq & Khalid, 2023). These competencies are particularly relevant to software-based Applied Statistics learning, where students routinely use computational tools to process, analyze, and interpret data. Such learning requires not only systematic problem-solving through CT but also awareness of the environmental implications of intensive digital technology use. However, CT and GCA have largely been examined as separate constructs, and instruments that assess both within the same statistics learning context remain limited. This gap highlights the need to develop and validate instruments specifically grounded in software-based Applied Statistics learning.

The urgency of this study stems from the gap between the importance of CT and GCA as 21st-century competencies and the limited availability of instruments capable of operationalizing both constructs within Applied Statistics education. Software-based statistical learning provides an authentic context for measuring CT, as students are required to abstract research problems, decompose analytical procedures, recognize patterns in data, and formulate analytical steps using computational tools (Aminah et al., 2023; Wu et al., 2024). At the same time, the use of software and digital technologies creates a relevant context for GCA, particularly awareness of environmental impacts, efficient and sustainable technology use, and responsible digital behavior (Atadoga et al., 2024; B & Ragini, 2025). In this context, JASP provides a concrete software environment in which students engage in statistical analysis while using digital resources. Therefore, integrating CT and GCA within software-based Applied Statistics learning provides a relevant context for examining computational problem-solving alongside environmentally responsible technology use.

Theoretically, CT is conceptualized as a multidimensional construct comprising four dimensions: decomposition, pattern recognition, abstraction, and algorithmic thinking (Pramesti et al., 2025; Selamat et al., 2024). Decomposition refers to breaking complex problems into smaller components, pattern recognition to identifying recurring patterns or relationships, abstraction to focusing on relevant information while excluding unnecessary details, and algorithmic thinking to formulating systematic steps for problem solving. These dimensions are relevant to software-based Applied Statistics learning, where students process statistical problems and data using computational tools.

Meanwhile, GCA encompasses cognitive, affective, and behavioral aspects related to awareness of the environmental impacts of digital technology, concern for sustainability, and responsible technology use (Eid et al., 2022; Farliana et al., 2024; Liu & Zhang, 2024). The use of digital and software technologies provides a relevant context for examining these aspects because green computing emphasizes environmentally sustainable practices in technology use (Atadoga et al., 2024). Thus, CT and GCA provide complementary dimensions for examining computational problem-solving and environmentally responsible technology use within the same software-based Applied Statistics learning context.

Previous research has addressed CT and GCA from different perspectives, but important gaps remain in their integration and measurement. Lehtonen et al. (2025) developed digital learning materials and tasks to enhance students' CT in Grades 1 through 9; however, the development was not specifically designed for Applied Statistics learning or for measuring CT through a context-specific psychometric instrument. Christensen (2023) developed a learning progression that connected CT with environmental sustainability education, demonstrating the conceptual relevance of linking computational thinking with sustainability; however, the work did not develop or validate an instrument that simultaneously measures CT and GCA. More recently, Ria et al. (2025) mapped the relationship between CT, GCA, and statistical analysis software, indicating an emerging connection among these constructs. However, this bibliometric evidence does not provide an empirically validated instrument for measuring CT and GCA simultaneously in software-based Applied Statistics learning. Therefore, a specific instrument is needed to operationalize both CT and GCA within this learning context and to establish evidence of content validity and inter-rater reliability before further empirical testing with student respondents.

The identified research gaps encompass three dimensions: (1) an empirical gap, as limited empirical evidence is available on integrating CT instruments within software-based statistics learning; (2) a contextual gap, as research on the development of CT and GCA instruments among higher education students in Indonesia remains limited; and (3) a methodological gap, as the reviewed literature provides limited evidence of a psychometrically validated instrument that simultaneously operationalizes CT and GCA within the same software-based Applied Statistics learning context. This limitation constrains the empirical examination of the relationship between these constructs in subsequent student-based analyses.

Based on these three research gaps, the contribution of this study lies in three aspects. First, conceptually, this study offers a context-specific integration of CT and GCA instruments within a single software-based Applied Statistics learning environment, rather than examining these constructs independently in unrelated contexts (Christensen, 2023; Lehtonen et al., 2025). The two instruments are designed to capture

distinct constructs computational competence and environmental awareness while being grounded in the same instructional context, providing a basis for examining their relationship in subsequent student-based analyses. Notably, although the two instruments were evaluated by separate panels of domain experts reflecting the distinct expertise required for each construct, both were anchored within the same JASP-based statistics learning environment, ensuring contextual consistency across the two constructs. Second, methodologically, this study employs a dual-panel expert judgment approach in which instrument-specific validators evaluate each construct within a shared pedagogical context, addressing the methodological gap identified in previous research. Third, contextually, the study provides initial evidence of content validity and inter-rater consistency for CT and GCA instruments specifically developed for higher education students in Indonesia, a population that remains underrepresented in the existing literature.

Therefore, this study aims to address three research gaps by: (1) developing a case-based CT instrument with 12 essay questions; (2) developing a GCA questionnaire with 10 statements; and (3) evaluating content validity using Aiken's V coefficient and inter-rater consistency using ICC. The implications of this study are theoretical, methodological, and practical. Theoretically, the study provides a context-specific framework for examining CT and GCA as complementary constructs within technology-based learning, rather than as entirely independent domains. Methodologically, the resulting instruments provide initial evidence of content validity and consistency among expert ratings, while further empirical validation with student respondents is required to establish construct validity and internal consistency before broader application. Practically, the instruments may serve as a basis for future studies, including student profiling and Latent Profile Analysis (LPA), after sufficient empirical validation. They may also support the development of differentiated statistics instruction that considers students' computational and environmental competencies, with potential alignment with Outcome-Based Education and the Sustainable Development Goals (SDGs).

Methods

This study employed an instrument development design using an expert judgment-based validation approach. The development process consisted of four stages: (1) conceptual analysis and construct operationalization; (2) item development; (3) internal review and refinement; and (4) external validation by expert panels. The CT instrument consisted of 12 case-based essay items presenting statistical problem scenarios and was designed to assess four dimensions: abstraction (Items 1, 2, 3, 9, 10, and 11), decomposition (Items 2 and 5), pattern recognition (Items 8, 9, and 12), and algorithmic thinking (Items 3, 4, 6, 7, 10, and 12). The GCA instrument consisted of 10 statements rated on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) and assessed three

dimensions: Environmental Impact Awareness of JASP Usage (Items 1–4), Perceptions of Efficiency and Sustainability of JASP Use (Items 5–7), and Responsible Digital Behavior (Items 8–10).

Content validity was evaluated through expert judgment using separate expert panels for the two instruments. The CT instrument was evaluated by two experts with expertise relevant to Mathematics Education and Educational Research and Evaluation, whereas the GCA instrument was evaluated by two experts with expertise in Educational Technology. The experts were selected based on their academic expertise and experience relevant to the respective construct and educational context. Each expert independently evaluated the items of the assigned instrument. Content validity was quantified using Aiken's V coefficient (Aiken, 1985). For each item, Aiken's V was calculated using the following formula:

$$V = \frac{\sum s}{n(c - 1)}$$

Where $s = r - l_0$, r is the score assigned by an expert, l_0 is the lowest score on the rating scale, n is the number of experts, and c is the number of rating categories. Aiken's V ranges from 0 to 1, with higher values indicating stronger content validity. In this study, $V < 0.40$ was interpreted as low validity, $0.40 \leq V \leq 0.80$ as moderate validity, and $V > 0.80$ as high validity, following the criteria adopted in previous studies (Ria & Susilowati, 2023; Sridadi et al., 2023).

Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC) based on the ratings provided by the two experts for each instrument. A two-way mixed-effects model with absolute agreement for single measures was used because the same expert panel evaluated all items within each instrument and the analysis focused on the absolute agreement between expert ratings. The ICC was calculated separately for the CT and GCA instruments. ICC values were interpreted as <0.50 = poor, 0.50 – 0.75 = moderate, 0.75 – 0.90 = good, and >0.90 = excellent reliability (Koo & Li, 2016; Odabaşioğlu et al., 2026).

Result

1. Content Validity of the CT Instrument

The results of the Aiken's V coefficient calculations for the CT instrument are presented in Table 1. Ten of the 12 items obtained an Aiken's V value of 1.000, indicating high content validity, whereas Items 11 and 12 obtained a value of 0.750, indicating moderate content validity based on the criteria adopted in this study. The average Aiken's V was 0.958, indicating high overall content validity. Although Items 11 and 12 remained within the acceptable validity range, the experts' qualitative feedback indicated that these items placed greater emphasis on implications, recommendations, and broader

responses than on the targeted CT processes. Accordingly, Items 11 and 12 were revised to improve the specificity of the expected responses and strengthen their alignment with the corresponding CT indicators.

Table 1. Results of Aiken's V for the CT Instrument

Items	CT Dimensions	Operational Indicators	V1	V2	Aiken's V	Category
1	Abstraction	Identify and filter relevant information from statistical problems	5	5	1.000	High Validity
2	Abstraction + Decomposition	Identify and filter relevant information from statistical problems; break complex statistical problems into simpler parts	5	5	1.000	High Validity
3	Abstraction + Algorithmic Thinking	Identify and filter relevant information from statistical problems; formulate statistical problem-solving steps logically and systematically	5	5	1.000	High Validity
4	Algorithmic Thinking	Formulate statistical problem-solving steps logically and systematically	5	5	1.000	High Validity
5	Decomposition	Break complex statistical problems into simpler parts	5	5	1.000	High Validity
6	Algorithmic Thinking	Formulate statistical problem-solving steps logically and systematically	5	5	1.000	High Validity
7	Algorithmic Thinking	Formulate statistical problem-solving steps logically and systematically	5	5	1.000	High Validity
8	Pattern Recognition	Recognize patterns, trends, and relationships in statistical data	5	5	1.000	High Validity
9	Pattern Recognition + Abstraction	Recognize patterns, trends, and relationships in statistical data; identify and filter relevant information from statistical problems	5	5	1.000	High Validity

Items	CT Dimensions	Operational Indicators	V1	V2	Aiken's V	Category
10	Algorithmic Thinking + Abstraction	Formulate statistical problem-solving steps logically and systematically; identify and filter relevant information from statistical problems	5	5	1.000	High Validity
11	Abstraction	Identify and filter relevant information from statistical problems	4	4	0.750	Moderate Validity
12	Algorithmic Thinking + Pattern Recognition	Formulate statistical problem-solving steps logically and systematically; recognize patterns, trends, and relationships in statistical data	4	4	0.750	Moderate Validity
Mean	—		—	—	0.958	High Validity

2. Content Validity of the GCA Instrument

The results of the Aiken's V coefficient calculations for the GCA instrument are presented in Table 2. Six of the ten items obtained an Aiken's V value of 1.000, indicating high content validity. Items 1 and 6 obtained a value of 0.875, also indicating high content validity, whereas Items 4 and 5 obtained a value of 0.750, indicating moderate content validity based on the criteria adopted in this study. The average Aiken's V was 0.925, indicating high overall content validity. However, the experts identified that Items 4 and 5 required further clarification to make the representation of the GCA construct more specific and reduce potential variation in interpretation. Specifically, Item 4 was revised to clarify the meaning of energy efficiency in relation to ICT energy use, while Item 5 was refined to specify the types of ICT resources considered.

Table 2. Results of Aiken's V for the GCA Instrument

Items	GCA Dimensions	Operational Indicators	V1	V2	Aiken's V	Category
1	Environmental Impact	Understanding of the ecological impacts of JASP compared with	5	4	0.875	High Validity
2	Awareness of JASP Usage	conventional methods, paperless practices, and	5	5	1.000	High Validity
3		green computing features	5	5	1.000	High Validity

Items	GCA Dimensions	Operational Indicators	V1	V2	Aiken's V	Category
4			4	4	0.750	Moderate Validity
5	Perceived Efficiency and Sustainability of JASP	Beliefs about JASP's energy efficiency and its contribution to reducing ICT resource waste	4	4	0.750	Moderate Validity
6			4	5	0.875	High Validity
7			5	5	1.000	High Validity
8			5	5	1.000	High Validity
9	Responsible Digital Behavior Intentions	Preferences and intended actions in choosing and promoting JASP as an environmentally friendly software	5	5	1.000	High Validity
10			5	5	1.000	High Validity
Mean	—		—	—	0.925	High Validity

The GCA instrument was intentionally contextualized to JASP-based Applied Statistics learning. Therefore, the instrument assesses green computing awareness as manifested in students' perceptions and intended behaviors regarding JASP use, rather than general awareness of all statistical software. This contextualization was adopted to align the instrument with the actual technology-mediated learning environment in which the construct was examined.

3. Inter-Rater Reliability of the CT Instrument: ICC Analysis.

The results of the ICC analysis for the CT instrument are presented in Table 3. Using a two-way mixed-effects model with absolute agreement for single measures (ICC3,1), the ICC estimate was 1.000, with a 95% confidence interval of 1.000–1.000. This value indicates excellent inter-rater reliability between the two experts, demonstrating complete consistency in their ratings across the 12 CT instrument items.

Table 3. ICC Results for CT Instrument Items

Type	Point Estimate	Lower 95% CI	Upper 95% CI
ICC3,1	1.000	1.000	1.000

Note. 12 items and 2 raters/measurements.

4. Inter-Rater Reliability of the GCA Instrument: ICC Analysis.

The results of the ICC analysis for the GCA instrument are presented in Table 4. Using a two-way mixed-effects model with absolute agreement for single measures (ICC3,1), the ICC estimate was 0.5610, with a 95% confidence interval ranging from -0.06202 to 0.8694. This value indicates moderate inter-rater reliability between the two experts' ratings of the 10 GCA instrument items.

Table 4. ICC Results for GCA Instrument Items

Type	Point Estimate	Lower 95% CI	Upper 95% CI
ICC3,1	0.5610	-0.06202	0.8694

Note. 10 items and 2 raters/measurements.

Discussion

The CT and GCA instruments demonstrated high overall content validity, with average Aiken's V values of 0.958 and 0.925, respectively, both exceeding the criterion for high validity ($V > 0.80$) adopted in this study (Ria & Susilowati, 2023; Sridadi et al., 2023). For the CT instrument, ten items obtained an Aiken's V of 1.000, while Items 11 and 12 obtained 0.750, indicating moderate content validity; these items were subsequently refined based on expert feedback to improve the specificity of the expected responses and their alignment with the targeted CT dimensions. For the GCA instrument, six items obtained an Aiken's V of 1.000, two items obtained 0.875, and two items obtained 0.750, with the latter indicating moderate content validity; expert feedback on Items 4 and 5 concerned the clarity of the terms "energy efficiency" and "ICT resource waste," which were subsequently refined to reduce potential variation in interpretation.

The ICC results provide additional evidence regarding the consistency of expert ratings for the two instruments. The CT instrument demonstrated excellent inter-rater reliability, with an ICC3,1 of 1.000 (95% CI = 1.000–1.000), whereas the GCA instrument demonstrated moderate inter-rater reliability, with an ICC3,1 of 0.561 (95% CI = -0.06202–0.8694), based on the interpretation criteria of Koo & Li (2016). The high ICC value for the CT instrument indicates strong consistency between the two experts' ratings, while the moderate ICC value for the GCA instrument indicates greater variation in expert ratings (Odabaşioğlu et al., 2026). For the CT instrument, the high level of agreement is consistent with the clear operationalization of the CT dimensions and the case-based format of the items. For the GCA instrument, the moderate inter-rater reliability is consistent with the experts' feedback on Items 4 and 5, particularly regarding the need to clarify the terms "energy efficiency" and "ICT resource waste."

Methodologically, the ICC analysis provides additional evidence regarding the consistency of expert ratings across the two instruments (Koo & Li, 2016). In the present study, the ICC results were based on ratings from two fixed experts, with ICC3,1 used to

assess single-measure inter-rater reliability. The findings should therefore be interpreted specifically as evidence of inter-rater reliability within the expert-validation process, rather than as evidence of overall instrument reliability. Further empirical testing with student respondents is required to establish the psychometric properties of the instruments.

From the perspective of sustainable digital education, the findings provide initial evidence that CT and GCA can be operationalized within a shared software-based Applied Statistics learning context. The integration of cognitive problem-solving dimensions with environmental awareness and responsible technology use is consistent with the growing emphasis on sustainability competencies in technology-based education (Peters et al., 2024; Mahrishi et al., 2025). However, the present findings are limited to evidence of content validity and inter-rater reliability; further empirical testing with student respondents is required to establish the construct validity, internal consistency, and other psychometric properties of the instruments.

Furthermore, the integration of CT and GCA establishes a conceptual link between computational problem-solving and sustainable technology use within software-based Applied Statistics learning. In this context, CT is operationalized through abstraction, decomposition, pattern recognition, and algorithmic thinking, whereas GCA is operationalized through awareness of environmental impacts, perceptions of efficiency and sustainability, and responsible digital behavior. Bringing these dimensions together connects systematic statistical problem-solving with awareness of the environmental and responsible-use aspects of digital technology, consistent with the emphasis on sustainability-oriented competencies in education (Hammer & Lewis, 2023; Idoiaga Mondragon et al., 2023; Peters et al., 2024; Mahrishi et al., 2025). Accordingly, the instruments developed in this study provide a context-specific basis for assessing these complementary constructs, while their broader pedagogical effectiveness requires further empirical investigation.

Conclusion

This study developed two instruments for assessing Computational Thinking (CT) and Green Computing Awareness (GCA) within software-based Applied Statistics learning and provided initial evidence of their content validity and inter-rater reliability based on expert judgment. The CT instrument consists of 12 case-based essay items and obtained an average Aiken's V of 0.958 and an ICC_{3,1} of 1.000, indicating high overall content validity and excellent inter-rater reliability. The GCA instrument consists of 10 questionnaire items and obtained an average Aiken's V of 0.925 and an ICC_{3,1} of 0.561, indicating high overall content validity and moderate inter-rater reliability. These findings support the contextual relevance of the instruments for further development and empirical testing; however, testing with student respondents is required to establish

construct validity, internal consistency, and other psychometric properties before broader application.

Despite these findings, several limitations should be acknowledged. First, the expert validation involved only two raters for each instrument, which limits the generalizability of the inter-rater reliability estimates. Second, the present study provides evidence of content validity and inter-rater reliability but does not yet establish empirical construct validity or score reliability based on student responses. Future research should therefore conduct empirical validation with student respondents. For the GCA questionnaire, exploratory factor analysis (EFA) may first be conducted to examine the empirical structure of the newly developed items, followed by confirmatory factor analysis (CFA) using an independent sample to confirm the hypothesized three-dimensional structure, together with Composite Reliability and internal consistency analysis using Cronbach's alpha and/or McDonald's omega.

For the CT essay test, the hypothesized four-dimensional structure should be examined using ordinal CFA with theoretically specified cross-loadings, followed by Composite Reliability. Because the CT instrument uses polytomous performance levels and essay responses, scoring reliability should also be examined using a standardized scoring rubric and inter-rater reliability, with Generalizability Theory considered to evaluate measurement error associated with persons, tasks, and raters. Once sufficient empirical evidence supports the measurement properties of both instruments, subsequent research can use the resulting CT and GCA dimension scores in LPA to identify distinct configurations of students based on their combined CT and GCA characteristics. This sequence will provide a psychometrically supported basis for examining student heterogeneity through a person-centered analysis. Finally, the contextual focus on JASP-based Applied Statistics learning may limit the generalizability of the instruments to other statistical software and learning contexts.

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>
- Aminah, N., Sukestiyarno, Y. L., Cahyono, A. N., & Maat, S. M. (2023). Student activities in solving mathematics problems with a computational thinking using Scratch. *International Journal of Evaluation and Research in Education (IJERE)*, 12(2), 613. <https://doi.org/10.11591/ijere.v12i2.23308>
- Atadoga, A., Umoga, U. J., Lottu, O. A., & Sodiya, E. O. (2024). Advancing green computing: Practices, strategies, and impact in modern software development for environmental sustainability. *World Journal of Advanced Engineering Technology and Sciences*, 11(1), 220–230. <https://doi.org/10.30574/wjaets.2024.11.1.0052>

- B, D., & Ragini. (2025). Greening the workplace: Exploring the dynamics of environmental knowledge, individual green values, supervisory support, and green behavior in IT organizations. *Sage Open*, 15(3), 21582440251357384. <https://doi.org/10.1177/21582440251357383>
- Christensen, D. (2023). Computational thinking to learn environmental sustainability: A learning progression. *Journal of Science Education and Technology*. Springer. <https://doi.org/10.1007/s10956-022-10004-1>
- Eid, A., Salah, M., Barakat, M., & Obrecht, M. (2022). Airport sustainability awareness: A theoretical framework. *Sustainability*, 14(19), 11921. <https://doi.org/10.3390/su141911921>
- Farliana, N., Hardianto, H., Rusdarti, R., & Sakitri, W. (2024). Sustainability consciousness in higher education: Construction of three-dimensional sustainability and role of locus of control. *Sustinere Journal of Environment and Sustainability*, 8(1), 80–90. <https://doi.org/10.22515/sustinere.jes.v8i1.374>
- Fidhyallah, N. F., & Zakiah, R. (2025). Rethinking assessment in the digital age: Pedagogical implications and challenges. *Journal of Educational Technology Innovation and Applications*, 1(1), 21–26. <https://doi.org/10.56741/jetia.v1i01.985>
- Hammer, T., & Lewis, A. L. (2023). Which competencies should be fostered in education for sustainable development at higher education institutions? Findings from the evaluation of the study programs at the University of Bern, Switzerland. *Discover Sustainability*, 4(1), 19. <https://doi.org/10.1007/s43621-023-00134-w>
- Idoiaga Mondragon, N., Yarritu, I., Saez de Cámara, E., Beloki, N., & Vozmediano, L. (2023). The challenge of education for sustainability in higher education: Key themes and competences within the University of the Basque Country. *Frontiers in Psychology*, 14, 1158636. <https://doi.org/10.3389/fpsyg.2023.1158636>
- Kang, Y. (2024). Enhancing computational and data science thinking skills for K-12 education. *Lecture Notes in Education Psychology and Public Media*, 43(1), 20–25. <https://doi.org/10.54254/2753-7048/43/20240609>
- Kee, T., & Zhang, H. (2022). Digital experiential learning for sustainable horticulture and landscape management education. *Sustainability*, 14(15), 9116. <https://doi.org/10.3390/su14159116>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lehtonen, D., Parviainen, M., Kaarto, H., Räsänen, P., & Dagienė, V. (2025). Development of digital learning materials to promote computational thinking skills in grades 1–9. *INTED2025 Proceedings*, 6643–6650. <https://doi.org/10.21125/inted.2025.1720>
- Liu, J., & Zhang, X. (2024). Enhancing environmental awareness through digital tools in environmental education in China. *Environment-Behaviour Proceedings Journal*, 9(28), 123–129. <https://doi.org/10.21834/e-bpj.v9i28.5820>

- Mahrishi, M., Ramakrishna, S., Hosseini, S., & Abbas, A. (2025). A systematic literature review of the global trends of outcome-based education (OBE) in higher education with an SDG perspective related to engineering education. *Discover Sustainability*, 6(1), 620. <https://doi.org/10.1007/s43621-025-01496-z>
- Obafemi, S., Oyetade, K., & Nyakudya, M. (2023). Awareness and practice of green computing in higher education institutions. In A. K. Nagar, D. Singh Jat, D. K. Mishra, & A. Joshi (Eds.), *Intelligent Sustainable Systems* (pp. 511–519). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-7660-5_44
- Odabaşioğlu, M. E., Kasırga, Z., & Kanter, A. G. (2026). Concurrent validity and reliability of a smartphone app for assessing weight-bearing ankle dorsiflexion using 2D video analysis. *Journal of Back and Musculoskeletal Rehabilitation*, 39(3), 835–843. <https://doi.org/10.1177/10538127251395837>
- Peters, A. K., Capilla, R., Coroamă, V. C., Heldal, R., Lago, P., Leifer, O., Moreira, A., Fernandes, J. P., Penzenstadler, B., Porras, J., & Venters, C. C. (2024). Sustainability in computing education: A systematic literature review. *Computing Education*. <https://doi.org/10.1145/3639060>
- Pramesti, S. L. D., Dewi, H. L., Mariani, S., & Amrina, D. (2025). Self-regulated learning and computational thinking: Analyzing the performance of junior high school students on PISA-like problems. *Matematika Dan Pembelajaran*, 13(2), 257–289. <https://doi.org/10.33477/mp.v13i2.11343>
- Ria, R. R. P., & Susilowati, D. (2023). Validity and reliability of a diagnostic test of computational thinking based on local wisdom. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, 9(4), 1142–1149. <https://doi.org/10.3334/jk.v9i4.8991>
- Ria, R. R. P., Yuda, L. S., & Patty, E. N. S. (2025). Bibliometric insights into computational thinking and green computing awareness: Emerging trends toward sustainable digital pedagogy (2015–2024). *Jurnal Kolaboratif Sains*, 8(11), 6849–6858. <https://doi.org/10.56338/jks.v8i11.9055>
- Rina, C., Rossalina, C. M., Zubainur, M., Subianto, D., & Fadhiliani. (2025). Unpacking research on computational thinking in mathematics education: A systematic literature review. *Jurnal Elemen*. <https://doi.org/10.29408/jel.v11i2.29183>
- Salamat, S., Nasir, M. K. M., & Adnan, N. H. (2024). Investigation of computational thinking skills through instructional techniques, games and programming tools. *International Journal of Learning Teaching and Educational Research*, 23(10), 435–452. <https://doi.org/10.26803/ijlter.23.10.21>
- Sridadi, S., Hermawan, H. A., & Utama, A. M. B. (2023). Development of basic swimming skills assessment instruments for PJKR FIKK UNY students. *Proceedings of the 6th Yogyakarta International Seminar on Health, Physical Education and Sport Science (YISHPESS 2023)*, 90–96. https://doi.org/10.2991/978-94-6463-356-6_11
- Tariq, M. S., & Khalid, A. (2023). Placing Green IT awareness and practices among universities' librarians: A NAT perspective. *The Journal of Academic Librarianship*, 49(5),

102770. <https://doi.org/10.1016/j.acalib.2023.102770>

Wu, T. T., Asmara, A., Huang, Y. M., & Hapsari, I. P. (2024). Identification of problem-solving techniques in computational thinking studies: Systematic literature review. *Sage Open*. <https://doi.org/10.1177/21582440241249897>