

Validating a Computational Thinking Assessment in Malaysia: Evidence of Concurrent Validity from a Multiple-Choice Instrument

Ahmad Sarji Abdul HAMED^a, Su Luan WONG^{a*}, Mohd Zariat Abdul RANI^b, Mas Nida Md KHAMBARI^a, Nur Aira Abd RAHIM^a, Priscilla MOSES^c, & Nur Izzati Mat ZIN^a

^a*Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia*

^b*Faculty of Modern Language and Communication, Universiti Putra Malaysia, Malaysia*

^c*Faculty of Education, Universiti Tunku Abdul Rahman, Malaysia*

[*suluan@upm.edu.my](mailto:suluan@upm.edu.my)

Abstract: This study investigates the concurrent validity of a newly developed multiple-choice computational thinking (CT) instrument tailored to the Malaysian educational framework, which uniquely incorporates evaluation as a sixth component alongside decomposition, abstraction, pattern recognition, algorithmic thinking, and logical reasoning. Administered to 52 undergraduates together with the established Computational Thinking Questionnaire (CTq), the instrument showed a strong positive correlation with the CTq ($r = .741, p < .001$), providing robust evidence of validity. Beyond statistical alignment, the instrument reflects Malaysia's exam-oriented and collectivist educational culture by embedding locally meaningful scenarios within a familiar multiple-choice format, thereby ensuring both cultural relevance and policy alignment. Its potential applications include diagnosing student proficiency, informing curriculum integration, and guiding national education reforms, making it a valuable tool for teachers, curriculum developers, and policymakers. While the findings highlight the instrument's promise, the modest sample size limits generalizability, warranting replication with larger and more diverse groups. Overall, this research advances the global discourse on CT assessment by demonstrating how culturally responsive, curriculum-aligned instruments can bridge international frameworks with local educational realities in Southeast Asia.

Keywords: Computational thinking, assessment tools, multiple choice question, computational thinking questionnaire.

1. Introduction

In an increasingly digital and interconnected world, the ability to solve complex problems systematically has emerged as a defining competency for the twenty-first century. Computational thinking (CT) lies at the heart of this skill set, offering a structured approach to problem-solving that transcends disciplinary boundaries. Since Wing's (2006) seminal work, CT has been conceptualized as a collection of cognitive processes for framing problems and designing solutions that can be executed either by computers or humans following a logical sequence of steps. Five core components (decomposition, abstraction, pattern recognition, algorithmic design, and logical reasoning) are widely recognized as essential to CT (Grover & Pea, 2013; Shute et al., 2017). In the Malaysian context, the Ministry of Education (MOE, 2017) expanded this framework by adding evaluation as a sixth component, reflecting national priorities that emphasize reflective problem-solving, self-assessment, and solution refinement. A unique feature of the Malaysian and broader Southeast Asian educational context is the cultural orientation towards collectivism, respect for authority, and exam-centered learning traditions (Hofstede, 2011; Yusof & Halim, 2019). These cultural values shape not only classroom interactions but also the way assessment is perceived and used. In Malaysia, where high-stakes examinations have historically played a central role in determining

academic progression, multiple-choice questions (MCQs) represent a familiar and culturally congruent format. Their scalability, objectivity, and alignment with exam-driven traditions make them especially suitable for measuring CT in a way that resonates with students and educators. At the same time, embedding MCQ items within locally relevant problem scenarios such as community issues, everyday problem-solving, and curriculum-linked contexts, ensures that the assessment reflects both the cognitive and cultural realities of Malaysian learners.

Globally, education systems have recognized CT as a foundational skill for participation in the knowledge economy, with integration efforts seen in the UK's Computing at School initiative, the US K–12 Computer Science Framework, and curricular reforms across Asia and Europe (Voogt et al., 2015; Yadav et al., 2021). Yet, the success of such reforms depends not only on curriculum design but also on the availability of valid and contextually relevant assessment instruments. While CT assessments such as the Computational Thinking Scale (CTS) (Korkmaz et al., 2017), the Computational Thinking Performance Test (Román-González et al., 2017), and the competent Computational Thinking test (cCTt) (El-Hamamsy et al., 2022) have contributed significantly to the field, most were developed in Western, high-resource contexts. Their direct applicability to Southeast Asian classrooms, where cultural and curricular frameworks differ substantially, remains uncertain. This study addresses these gaps by providing empirical evidence of concurrent validity for a newly developed multiple-choice CT instrument that incorporates all six Malaysian-defined components. Unlike prior studies that emphasize construct validity or internal consistency, this research explicitly investigates the degree of convergence between the new instrument and an established benchmark (CTq) administered within the same session. By doing so, it contributes two key advances: (1) the validation of an assessment tool that reflects both national policy priorities and cultural characteristics of Malaysian education, and (2) a demonstration of how MCQs, when carefully designed, can capture CT proficiency in ways that are both psychometrically sound and culturally appropriate. Accordingly, this study is guided by the following research questions:

1. To what extent do scores from the newly developed multiple-choice CT instrument correlate with scores from an established and validated CT questionnaire?
2. What are the implications of the concurrent validity findings for the adoption and application of the new instrument in educational contexts??

2. Literature Review

Over the past decade, CT assessment has become a central focus of educational research as systems worldwide search for reliable and valid ways to measure this critical twenty-first-century skill. A wide variety of approaches have been explored, from self-report surveys to performance-based tasks, each with its own strengths and limitations. Among these, multiple-choice assessments have gained particular prominence because of their scalability, efficiency, and objectivity, making them well suited for large-scale educational settings (Weese & Feldhausen, 2017). When carefully designed and validated, such assessments can effectively capture the cognitive dimensions of CT (Shute et al., 2017). Several notable instruments illustrate this promise: the Computational Thinking Scale (CTS) by Korkmaz et al. (2017), which evaluates creativity, algorithmic thinking, cooperation, and problem-solving using a Likert-type format; the Computational Thinking Performance Test (CTPT) by Román-González et al. (2017), a 28-item multiple-choice measure validated through Classical Test Theory (CTT); and the competent Computational Thinking Test (cCTt) by El-Hamamsy et al. (2022), an unplugged multiple-choice test validated with both CTT and Item Response Theory (IRT). Collectively, these instruments demonstrate that multiple-choice formats, when psychometrically sound, can serve as robust tools for assessing CT across diverse contexts. Despite these contributions, the literature reveals a critical gap in the validation of CT assessments concurrent validity.

Concurrent validity refers to the degree to which results from a new instrument converge with those of an established benchmark when both are administered simultaneously (Anastasi & Urbina, 1997). While studies frequently report internal consistency and construct validity, evidence of convergence with validated measures remains comparatively rare. One

important exception is Rowe et al. (2021), who demonstrated strong correlations between the Inclusive Assessment of Computational Thinking (IACT) and the Bebras CT Challenge, confirming the applicability of IACT to diverse learners, including those with special educational needs. Yet, examples of this kind are limited, especially outside Western educational systems. In the Malaysian context, the Computational Thinking Questionnaire (CTq) developed by Hamed et al. (2024) has emerged as a key benchmark aligned with national curricular priorities. Initially comprising 60 items across six CT constructs (abstraction, decomposition, pattern recognition, algorithmic thinking, logical reasoning, and evaluation), the CTq showed excellent internal consistency (Cronbach's $\alpha = .93$). Its updated version further incorporates readiness and motivation, expanding the scope to include not only cognitive skills but also affective and motivational dimensions of CT. This makes it uniquely relevant in Malaysia, where educational policy emphasizes not only skill acquisition but also student preparedness for integrating CT into teaching and learning practices.

Two limitations in the current literature highlight the need for the present study. First, most validation efforts for CT instruments have emphasized internal consistency and construct validity, often neglecting the crucial dimension of concurrent validity (Korkmaz et al., 2017; Román-González et al., 2017). Second, concurrent validity research has largely been conducted in Western, high-resource contexts (Cutumisu et al., 2019), which differ significantly from Southeast Asian settings where exam-oriented traditions, multilingual environments, and collectivist practices shape teaching and assessment. Instruments developed elsewhere may therefore lack cultural and curricular relevance in Malaysia. To address these gaps, the present study validates a newly developed multiple-choice CT instrument tailored specifically to the Malaysian educational landscape. Its novelty lies in two aspects: the integration of "evaluation" as a sixth CT component, uniquely recognized by the Malaysian Ministry of Education (MOE, 2017), and the embedding of locally meaningful, culturally relevant problem scenarios in its design. Beyond establishing internal psychometric properties, this study provides empirical evidence of concurrent validity by directly comparing the new instrument with the CTq. In doing so, it not only contributes to the global body of CT assessment research but also advances the development of culturally grounded tools that can inform educators, curriculum designers, and policymakers in Malaysia and across Southeast Asia.

3. Methodology

This study adopted a quantitative, correlational design to examine the concurrent validity of a newly developed multiple-choice computational thinking (CT) instrument. The goal was to determine the extent to which its scores aligned with the established CTq. Both instruments were administered in a single session to the same participants, consistent with validation principles outlined by Anastasi and Urbina (1997) and the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014). The sample comprised 52 purposively selected undergraduate engineering students at Universiti Putra Malaysia who met criteria of active enrollment, basic digital literacy, and informed consent. While sufficient for correlational analysis, this sample size limits statistical power and generalizability, highlighting the need for future research with larger and more diverse groups, including pre-service teachers and learners from varied sociocultural backgrounds. The new multiple-choice instrument contained 30 items, each situated in real-world or curricular scenarios, with four options (one correct, three distractors) following best-practice guidelines (Haladyna & Rodriguez, 2013). Items assessed six CT components: decomposition, abstraction, pattern recognition, algorithmic thinking, logical reasoning, and evaluation, the latter reflecting the Malaysian Ministry of Education's (MOE, 2017) unique emphasis.

Content validity was established through expert review by three specialists in educational technology and CT. As the criterion measure, the study employed the CTq (Hamed et al., 2024), an 80-item Likert-type questionnaire measuring eight CT dimensions (including readiness and motivation), previously validated with strong reliability (Cronbach's $\alpha = .95$). Its alignment with Malaysia's curriculum and policy further justified its use as a benchmark. Data collection took place in classrooms under standardized conditions. To minimize order effects, participants alternated between completing the new instrument first or

the CTq first, with instructions read aloud and a short break in between. Ethical clearance was obtained from Universiti Putra Malaysia's Ethics Committee. Data were analyzed using IBM SPSS Statistics (Version 27). Descriptive statistics summarized performance, and Pearson's correlation coefficient tested concurrent validity between the two instruments. Effect sizes were interpreted using Cohen's (1988) benchmarks (.10 small, .30 medium, .50 large), with significance set at $p < .05$. Interpretation emphasized both statistical convergence and the practical implications of using the new instrument to support CT assessment in Malaysian classrooms.

4. Result and Discussion

4.1 Descriptive Statistics

The descriptive statistics provided an overview of participant performance on both the newly developed multiple-choice instrument (MCQ) and the established Computational Thinking Questionnaire (CTq). Scores on the MCQ ranged from 43 to 90, with a mean of 72.65 (SD = 10.21), suggesting that participants demonstrated relatively strong CT proficiency across the six measured components. On the CTq, mean scores ranged from 2.58 to 4.60 on a five-point scale, with an overall mean of 3.67 (SD = 0.54), indicating that participants generally self-reported moderate to high levels of CT skills. These descriptive results show consistency between students demonstrated proficiency on an objective assessment and their perceived competence on a self-report questionnaire.

4.2 Concurrent Validity Analysis (RQ1)

The core analysis of concurrent validity was conducted using Pearson's correlation coefficient to determine the relationship between scores on the MCQ and the CTq. The results revealed a statistically significant, strong positive correlation ($r = .741$, $p < .001$). According to Cohen's (1988) guidelines, this represents a large effect size, providing robust evidence that the newly developed multiple-choice instrument measures CT proficiency in a manner highly consistent with the validated CTq. This finding aligns with Rowe et al. (2021), who similarly reported strong convergence between the Inclusive Assessment of Computational Thinking (IACT) and the Bebras CT Challenge, demonstrating that meaningful alignment can be achieved even when assessment formats differ. In the present case, the MCQ provides performance-based evidence, while the CTq captures self-reported skills, yet both converge strongly on the construct of computational thinking.

4.3 Interpretation and Implications (RQ2)

The core analysis of concurrent validity was conducted using Pearson's correlation coefficient to determine the relationship between scores on the MCQ and the CTq. The results revealed a statistically significant, strong positive correlation ($r = .741$, $p < .001$). According to Cohen's (1988) guidelines, this represents a large effect size, providing robust evidence that the newly developed multiple-choice instrument measures CT proficiency in a manner highly consistent with the validated CTq. This finding aligns with Rowe et al. (2021), who similarly reported strong convergence between the Inclusive Assessment of Computational Thinking (IACT) and the Bebras CT Challenge, demonstrating that meaningful alignment can be achieved even when assessment formats differ. In the present case, the MCQ provides performance-based evidence, while the CTq captures self-reported skills, yet both converge strongly on the construct of computational thinking. The significance of these findings lies in both methodological and practical domains. Methodologically, the study provides one of the few demonstrations of concurrent validity for a CT assessment in Southeast Asia. Validation research in this region is limited, with most CT assessment studies situated in Western, high-resource contexts (Cutumisu et al., 2019). By situating the validation process within Malaysia's national framework, this research provides contextually grounded evidence that strengthens

the case for culturally relevant CT assessments. The inclusion of “evaluation” as a sixth CT component, unique to the Malaysian Ministry of Education’s framework (MOE, 2017), further illustrates how the instrument aligns with local curricular priorities in ways that many global instruments do not.

The implications for teachers are particularly noteworthy. A validated multiple-choice instrument offers educators a practical tool for diagnosing students’ CT strengths and weaknesses. Because the instrument is aligned with the national curriculum, it can serve as both a diagnostic and formative assessment tool, guiding teachers in tailoring instruction to address specific skill gaps. For example, if a student demonstrates strong performance in pattern recognition but weaker ability in evaluation, teachers can design interventions to strengthen reflective and iterative problem-solving. Moreover, the objectivity and scalability of MCQs allow for efficient assessment in large classes, making the instrument feasible for widespread use in Malaysian schools and universities. From a curriculum design perspective, the results reinforce the importance of embedding CT assessment within national education systems. The strong alignment between the MCQ and the CTq suggests that the instrument not only captures the intended constructs but also reflects the broader educational emphasis on CT as outlined in Malaysian policy documents. Curriculum developers can use the instrument as a benchmark to evaluate the effectiveness of CT integration across subjects and grade levels. Its scalability further means that it can be adapted for both formative classroom uses and summative evaluation at institutional or national levels.

Cultural implications also emerge from the findings. Malaysia’s exam-oriented learning culture and collectivist educational practices mean that students are highly accustomed to MCQ-based testing. By designing items that are embedded in local contexts such as daily life situations or community-based problem-solving, the instrument reflects cultural characteristics while maintaining psychometric rigor. This cultural alignment not only enhances the face validity of the instrument but also increases its acceptance among educators and students, thereby facilitating smoother integration into teaching and learning practices. At the same time, the study acknowledges that the small sample size ($n = 52$) poses limitations on the generalizability of these findings. While the results provide strong preliminary evidence, further replication with larger and more diverse groups, including secondary school students, pre-service teachers, and learners from different sociocultural backgrounds, is necessary. Such replication will ensure that the instrument is robust across varied populations and reduce the risk of cultural or contextual bias.

5. Conclusion

This study provides strong empirical evidence for the concurrent validity of a newly developed multiple-choice CT instrument aligned with the Malaysian educational framework. The instrument, which uniquely incorporates evaluation as a sixth CT component, reflects national curricular priorities and demonstrates a strong positive correlation with the established CTq, thereby confirming its credibility as a valid assessment tool. By embedding items in locally meaningful scenarios and employing a multiple-choice format familiar within Malaysia’s exam-oriented, collectivist educational culture, the instrument enhances both cultural relevance and policy alignment, addressing the scarcity of validation studies in Southeast Asia. Its practical utility lies in offering teachers a scalable means of diagnosing CT skills, supporting curriculum developers in benchmarking integration across subjects, and guiding policymakers in implementing national education reforms. At the same time, the modest sample size of 52 limits generalizability, underscoring the need for replication with larger and more diverse populations to ensure robustness and fairness. Overall, the study advances both theory and practice by demonstrating how culturally responsive, curriculum-aligned assessments can bridge global CT research traditions with local educational realities, contributing a Southeast Asian perspective to the global discourse on CT education.

Acknowledgements

This study is supported by the Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS) (FRGS/1/2023/SSI07/UPM/01/1). The assistance of the Research Management Center (RMC) of Universiti Putra Malaysia in coordinating and distributing the fund for this research is greatly appreciated.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. *American Educational Research Association*.
- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Prentice Hall.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.)*. Lawrence Erlbaum Associates.
- Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology*, 28(6), 651–676. <https://doi.org/10.1007/s10956-019-09799-3>
- El-Hamamsy, L., Zapata-Cáceres, M., Barroso, E. M., Mondada, F., Zufferey, J. D., & Bruno, B. (2022). The competent computational thinking test: Development and validation of an unplugged computational thinking test for Upper Primary School. *Journal of Educational Computing Research*, 60(7), 1818-1866. <https://doi.org/10.1177/07356331221081753>
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43. <https://doi.org/10.3102/0013189X12463051>
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Hamed, A. S. A., Wong, S. L., Rani, M. Z. A., Khambari, M. N. M., Rahim, N. A. A., Khalid, F., Moses, P. (2024). Computational thinking proficiency among pre-service teachers at Universiti Putra Malaysia. *Bulletin of the Technical Committee on Learning Technology*, 24 (1), 7-12. <https://tc.computer.org/tclt/10-1109-2024-0105002/>
- Korkmaz, O., Cakir, R., & Ozden, M.Y. (2017). A validity and reliability study of the computational thinking scales (CTS). *Computer in Humanity Behaviour*, 72, 558-569.
- Ministry of Education, Malaysia. (2017). *Kemahiran pemikiran komputasional (computational thinking) salah satu pembelajaran abad ke-21 yang perlu dikuasai oleh semua*. Putrajaya.
- Román-González, M., Moreno-León, J., Robles, G. (2019). *Combining assessment tools for a comprehensive evaluation of computational thinking interventions*. In: Kong, SC., Abelson, H. (eds) *Computational Thinking Education*. Springer, Singapore. https://doi.org/10.1007/978-981-13-6528-7_6
- Rowe, E., Asbell-Clarke, J., Almeda, M. V., Gasca, S., Edwards, T., Bardar, E., Shute, V., & Ventura, M. (2021). Interactive assessments of CT (IACT): Digital interactive logic puzzles to assess computational thinking in grades 3–8. *International Journal of Computer Science Education in Schools*, 5(2), 28–73. <https://doi.org/10.21585/ijcses.v5i1.149>
- Sands, P., Yadav, A., Good, J. (2018). *Computational thinking in K-12: In-service teacher perceptions of computational thinking*. In: Khine, M. (eds) *Computational thinking in the STEM Disciplines*. Springer, Cham. https://doi.org/10.1007/978-3-319-93566-9_8
- Shute, V.J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158. <https://doi.org/10.1016/J.EDUREV.2017.09.003>
- Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. *Education and Information Technologies*, 20(4), 715-728. <https://doi.org/10.1007/s10639-015-9412-6>
- Wing, J. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–36.
- Yadav, A., & Berthelsen, U. (Eds.). (2021). *Computational thinking in education: A pedagogical perspective (1st ed.)*. Routledge. <https://doi.org/10.4324/9781003102991>