Robotics Education and Computational Thinking in Middle Schools: A Gulf–East Asia Comparison

Aisha Al-Mansoori[1, *], Kenji Watanabe[2] and Yuki Nakamura[3]

[1]Zayed University, United Arab Emirates
[2]Kyoto University, Japan
[3]Osaka University, Japan
E-mail: aisha.al.mansoori@mock.iccmets.test; kenji.watanabe@mock.iccmets.test; yuki.nakamura@mock.iccmets.test

Abstract.

Educational robotics is widely used to develop computational thinking, but the instructional practices that account for its effects remain poorly specified. This comparative study of robotics programmes in the United Arab Emirates and Japan assesses computational thinking gains among 620 middle-school students enrolled in one-semester robotics courses in six schools, three in each country. Computational thinking was measured with a validated test administered at the beginning and end of the semester, and instructional practices were recorded through classroom observation and teacher logs, yielding measures of the intensity of structured pair programming, reflective journal use, teacher-led direct instruction and open-ended project time for each class. Students in both countries showed substantial and statistically significant gains, with an average effect size of 0.68 standard deviations, and the gains did not differ significantly between countries after controlling for prior achievement. Hierarchical regression revealed that structured pair programming and reflective journals explain most of the variance in outcomes across both contexts, jointly accounting for 41 per cent of the variance in gain scores at the class level, whereas direct instruction and unstructured project time contributed little. Both practices were effective in each setting, with no significant country interaction. The paper argues that the benefits of robotics education depend on specific collaborative and metacognitive practices rather than on the technology itself and discusses implications for curriculum design and teacher professional development in both regions.

Keywords: robotics education, computational thinking, pair programming, reflective journals, middle school

**Corresponding author: aisha.al.mansoori@mock.iccmets.test*

1. Introduction

Computational thinking, the set of thought processes involved in formulating problems and solutions so that they can be carried out by an information-processing agent, has become a prominent goal of school education, introduced into national curricula as foundational to a digital economy and to general problem-solving. Educational robotics, in which students design, build and program physical robots, has been one of the most popular means of pursuing it, offering tangible and motivating tasks that require students to translate ideas into precise instructions and test them against physical outcomes.

The United Arab Emirates and Japan have both invested heavily in robotics education at the middle-school level, though from different starting points and with different institutional arrangements. The UAE has introduced robotics through national initiatives linked to its innovation and artificial intelligence strategies, often delivered by specialist teachers or external providers using imported curricula, and has emphasised competitions and showcase events. Japan made programming education compulsory in primary schools from 2020 and expanded technology education in lower secondary schools, with robotics integrated into the technology and home economics curriculum and delivered by subject teachers within a strongly structured lesson culture. The two contexts therefore differ in teacher background, in pedagogical tradition and in the framing of robotics, while sharing a commitment to computational thinking as an outcome.

Despite the popularity of educational robotics, meta-analyses have found its effects on computational thinking to be positive but highly variable, and have called for research that identifies which instructional practices explain this variation. This study responds to that call. It assesses computational thinking gains among 620 middle-school students in robotics programmes in the UAE and Japan, measures the intensity of four instructional practices in each class, and examines which practices account for variance in gains and whether their effects differ between the two contexts. In doing so, it seeks to move the discussion from whether robotics works to how it works and under what conditions.

2. Literature Review

2.1 Computational thinking and its assessment

The contemporary concept is attributed to Wing (2006), who argued that computational thinking is a universally applicable skill set rather than a competence confined to computer scientists. Grover and Pea (2013) reviewed the field and called for clearer definitions and validated assessments, and Brennan and Resnick (2012) distinguished computational concepts, practices and perspectives. Assessment has followed two approaches, analysis of programming artefacts and standalone tests of computational reasoning; the Computational Thinking Test of Román-González et al. (2017) belongs to the second and has been validated for students aged 10 to 16 across several countries.

2.2 Educational robotics and computational thinking

Educational robotics draws on constructionist theory, which holds that learning is most effective when learners construct meaningful, shareable artefacts (Papert, 1980), and robots provide immediate physical feedback that supports testing and debugging. Reviews find generally positive but heterogeneous effects: Benitti (2012) noted that many studies lacked rigorous designs, and meta-analytic estimates cluster around a moderate effect size, larger for longer interventions with explicit instruction in computational concepts (Zhang et al., 2021). The variability suggests that the technology is necessary but not sufficient and that instructional design matters.

2.3 Collaborative and metacognitive practices

Pair programming, in which two students share a computer as driver and navigator and exchange roles, has been associated in university courses with higher completion, better program quality and increased confidence, especially for less experienced students (McDowell et al., 2006). School evidence is smaller but consistent, with structured role rotation important to ensure both partners engage (Denner et al., 2014). The mechanism is thought to involve articulating reasoning to a partner, immediate error detection and distributed cognitive load.

Reflective journals, in which students write regularly about what they attempted and learned, are grounded in metacognition and self-regulated learning; reflection prompts improve programming outcomes by encouraging students to name and generalise their strategies (Schraw et al., 2006). In robotics, where students may succeed through trial and error without understanding why, reflection is hypothesised to convert procedural success into conceptual understanding. Direct instruction and open-ended project time are also common, and some evidence indicates that unguided project time yields weaker gains than structured activity (Kirschner et al., 2006).

2.4 Cultural context

Comparative research, notably the video studies of Stigler and Hiebert (1999), has documented differences between East Asian and other classrooms in lesson structure, collaborative problem-solving and student talk. Japanese lesson culture, with its structured problem-solving and written reflection at the close of lessons, may favour reflective journals, while Gulf classrooms have been moving from teacher authority towards collaborative approaches, with robotics often positioned as a vehicle for this shift. Whether collaborative and reflective practices are equally effective across these contexts is an open question that the present study addresses by testing the same practices in both settings.

2.5 Hypotheses

Four hypotheses were formulated. H1 states that students in robotics programmes in both countries show significant gains in computational thinking over one semester. H2 states that the intensity of structured pair programming is positively associated with gains. H3 states that the intensity of reflective journal use is positively associated with gains. H4 states that the associations for pair programming and reflective journals do not differ significantly between the two countries.

3. Methodology

3.1 Participants and settings

The study was conducted in six schools, three in the UAE, in Abu Dhabi and Dubai, and three in Japan, in Kyoto and Osaka Prefectures, during the 2025 academic year. All schools offered a one-semester robotics course of approximately 30 lesson hours to students in grades 7 and 8 using block-based programming environments with commercially available robotics kits. Twenty-four classes participated, twelve in each country, with a total of 620 students who completed both the pre-test and the post-test. Table 1 presents the sample profile. The samples were balanced by gender in the UAE and slightly male-dominated in Japan, and few students in either country had substantial prior programming experience beyond introductory primary school activities. Informed consent was obtained from parents and assent from students, and the study was approved by the ethics committees of the participating universities.

Table 1: Sample profile (N = 620)

Characteristic	**UAE (n = 308)**	**Japan (n = 312)**	**Total (N = 620)**
Schools / classes	3 / 12	3 / 12	6 / 24
Grade 7 (%)	52.3	49.4	50.8
Grade 8 (%)	47.7	50.6	49.2
Female (%)	49.4	44.2	46.8
Mean age (years)	12.9	13.1	13.0
Prior programming experience beyond primary school (%)	14.6	11.2	12.9
Pre-test CT score, mean (SD)	13.8 (4.6)	14.4 (4.3)	14.1 (4.5)

Source: Authors' data (2025)

3.2 Measures

Computational thinking was measured with a 28-item adaptation of the Computational Thinking Test (Román-González et al., 2017), translated into Arabic and Japanese and back-translated, and administered in the first and last weeks of the semester. Internal consistency was acceptable in both languages, with Cronbach's alpha of 0.82 for the Arabic version and 0.84 for the Japanese version at pre-test. Gain scores were calculated as post-test minus pre-test. Instructional practices were measured at the class level. Trained observers attended six lessons per class distributed across the semester and recorded the proportion of lesson time devoted to structured pair programming with role rotation, teacher-led direct instruction and open-ended project time, using a time-sampling protocol; inter-observer agreement on a subset of lessons was 0.87. Reflective journal use was measured from teacher logs as the number of lessons in which students completed a structured written reflection, expressed as a proportion of lessons. Control variables comprised pre-test score, grade, gender, prior programming experience and country.

3.3 Analysis

Gains were tested with paired t-tests and effect sizes calculated as Cohen's d using the pooled pre-test standard deviation. Because students were nested within classes and practices were measured at the class level, hierarchical regression was used with gain score as the dependent variable, student-level controls entered at the first step, the four practice variables at the second step and country interaction terms at the third step. Standard errors were clustered by class, and the intraclass correlation for gain scores was 0.14, confirming that class-level variation was substantial. Class-level variance explained by the practice variables was estimated from a two-level model. Analyses were conducted in R. Assumptions of normality and homoscedasticity of residuals were checked and found acceptable, and variance inflation factors were below 2.0.

4. Results

4.1 Computational thinking gains

Students in both countries made substantial gains. The mean score rose from 14.1 to 17.2 in the pooled sample, a gain of 3.1 points, and the effect size was $d = 0.68$. Gains were 3.2 points in the UAE ($d = 0.70$) and 3.0 points in Japan ($d = 0.67$), and the difference between countries was not statistically significant after controlling for pre-test score ($t = 0.81$, $p = 0.42$). H1 was therefore supported. Gains were slightly larger for students with lower pre-test scores, indicating some narrowing of the initial distribution, and did not differ significantly by gender or grade. Students with prior programming experience made gains similar to those without.

4.2 Instructional practice and gains

Table 2 presents the descriptive statistics for instructional practices by country and their bivariate correlations with class mean gain. Structured pair programming occupied on average 34 per cent of observed lesson time in the UAE classes and 29 per cent in the Japanese classes, while reflective journals were used in 41 per cent of lessons in the UAE and 63 per cent in Japan. Direct instruction was more prevalent in Japan and open-ended project time more prevalent in the UAE. Variation between classes within each country was considerable for all four practices, which provided the basis for the regression analysis. Both pair programming and reflective journal use were strongly correlated with class mean gain in each country, while direct instruction showed a weak positive correlation and project time a weak negative correlation.

Table 2: Instructional practice intensity by country and correlation with class mean gain (24 classes)

Practice	UAE mean % (SD)	Japan mean % (SD)	r with class gain, UAE	r with class gain, Japan
Structured pair programming (% of lesson time)	34 (11)	29 (9)	0.71	0.62
Reflective journals (% of lessons)	41 (18)	63 (15)	0.58	0.69
Teacher-led direct instruction (% of lesson time)	22 (8)	31 (7)	0.19	0.24
Open-ended project time (% of lesson time)	31 (12)	22 (9)	−0.21	−0.15

Source: Classroom observations and teacher logs (2025)

Table 3 presents the hierarchical regression results. At the first step, student-level controls explained 9 per cent of the variance in gain scores, with pre-test score negatively associated with gain. Adding the four practice variables at the second step increased the explained variance substantially. Structured pair programming ($\beta = 0.29$, $p < 0.001$) and reflective journal use ($\beta = 0.27$, $p < 0.001$) were both significantly and positively associated with gains, supporting H2 and H3. Direct instruction had a small positive coefficient that did not reach significance, and open-ended project time had a small negative coefficient that was also non-significant. At the class level, the two-level model indicated that pair programming and reflective journals together explained 41 per cent of the between-class variance in gains, with direct instruction and project time adding only 3 per cent.

Table 3: Hierarchical regression of computational thinking gain scores (N = 620, standardised coefficients)

Predictor	Step 1: controls	Step 2: practices	Step 3: country interactions	Hypothesis and outcome
Pre-test score	−0.24***	−0.22***	−0.22***	Control
Country (Japan = 1)	−0.03	−0.05	−0.04	Control
Structured pair programming	—	0.29***	0.32***	H2 supported
Reflective journals	—	0.27***	0.24***	H3 supported
Direct instruction	—	0.06	0.06	Not significant
Open-ended project time	—	−0.05	−0.05	Not significant
Pair programming × Japan	—	—	−0.06 (p = 0.21)	H4 supported
Reflective journals × Japan	—	—	0.07 (p = 0.14)	H4 supported
R^2 (student level)	0.09	0.27	0.28	

*Source: Authors' analysis (2025). *** p < 0.001; gender, grade and prior experience included as controls; standard errors clustered by class*

4.3 Country differences in practice effects

The third step added interaction terms between each practice and country. Neither interaction was statistically significant, supporting H4. The direction of the coefficients indicated that the effect of pair programming was slightly stronger in the UAE classes and that of reflective journals slightly stronger in the Japanese classes, but the confidence intervals overlapped substantially and the addition of the interaction terms increased the explained variance by only one percentage point. Both practices were therefore effective in both contexts. Supplementary analysis showed that classes combining high levels of both practices, defined as above the median on each, achieved a mean gain of 4.2 points, compared with 2.1 points in classes below the median on both, a difference equivalent to roughly half a standard deviation of the pre-test distribution.

5. Discussion

5.1 Practices rather than technology

The central finding is that the gains observed in both countries were explained largely by structured pair programming and reflective journals rather than by direct instruction or free project time. This adds specificity to the meta-analytic observation that robotics interventions vary widely (Benitti, 2012; Zhang et al., 2021) and supports the view that the technology is a context in which effective practices can be deployed rather than a cause of learning in itself. The pair programming result is consistent with McDowell et al. (2006) and Denner et al. (2014) and extends them to robotics, where the physical artefact makes the navigator role concrete; the journal result supports the metacognitive account (Schraw et al., 2006) and suggests that the risk of trial-and-error success without understanding is mitigated when students must articulate what they learned.

The weak contribution of project time deserves comment. Constructionist theory (Papert, 1980) values student-directed design, and projects are often the most engaging element of robotics courses. The results do not show that project time is harmful, but on its own it contributes little to measured computational thinking, consistent with Kirschner et al. (2006) on minimally guided activity; observers noted that in classes with high project time and low use of the structured practices, students succeeded through repeated adjustment without reasoning about program behaviour. Projects should therefore be embedded in collaborative and reflective structures rather than reduced.

5.2 Cross-cultural consistency

The absence of country differences is notable given the different pedagogical traditions. Journals were used more in Japan and pair programming more in the UAE, yet each was associated with gains in the setting where it was less familiar as well as where it was more so. This suggests that the mechanisms, articulation of reasoning and metacognitive consolidation through writing, are not culturally specific and that pedagogical models can be transferred between the regions with reasonable confidence; the slightly larger coefficients for pair programming in the UAE and journals in Japan were small.

5.3 Implications for curriculum and teacher development

Robotics courses should specify structured pair programming with role rotation and regular written reflection as core components rather than optional enrichments, with the time devoted to them treated as a design parameter. Teacher development should focus on managing pair work, including role assignment and equitable participation, and on designing prompts that elicit strategy articulation rather than description of activity. In the UAE, where external providers often deliver robotics with a competition orientation, the priority is to embed reflection; in Japan, to expand structured collaboration. In terms of the framework of Brennan and Resnick (2012), these practices appear to support computational practices and perspectives, not merely concepts, which may explain their effect on a general test.

6. Conclusion

This comparative study assessed computational thinking gains among 620 middle-school students in robotics programmes in the United Arab Emirates and Japan and examined which instructional practices explained variation in those gains. Students in both countries made substantial gains, with an average effect size of 0.68, and the gains did not differ between countries. Structured pair programming and reflective journals explained most of the variance in outcomes, jointly accounting for 41 per cent of between-class variance, while direct instruction and open-ended project time contributed little. The effects of the two practices did not differ significantly between the two contexts. The study contributes to the literature on educational robotics by identifying specific practices that account for its effects and by demonstrating their consistency across two culturally distinct settings, and it offers practical guidance for curriculum design and teacher development.

Several limitations should be noted. The study was observational rather than experimental, and although the practices were measured independently of outcomes and controls were included, unmeasured teacher characteristics may be associated with both practice intensity and student gains. The number of classes, though adequate for the analysis, limits the precision of class-level estimates, and the schools were not randomly selected. Gains were measured over a single semester with a general test, and longer-term retention and transfer to other domains were not assessed. Future research should employ experimental designs that assign classes to practice conditions, extend the comparison to additional countries and to primary and upper-secondary levels, examine the quality as well as the quantity of pair interaction and reflective writing, and investigate whether the practices identified here have similar effects in text-based programming and in non-robotics contexts.

Acknowledgment

The authors thank Zayed University, Kyoto University and Osaka University for institutional support and the participating schools, teachers and students.

References

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, *58*(3), 978–988. https://doi.org/10.1016/j.compedu.2011.10.006

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 Annual Meeting of the American Educational Research Association* (pp. 1–25). AERA.

Denner, J., Werner, L., Campe, S., & Ortiz, E. (2014). Pair programming: Under what conditions is it advantageous for middle school students? *Journal of Research on Technology in Education*, *46*(3), 277–296. https://doi.org/10.1080/15391523.2014.888272

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

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

McDowell, C., Werner, L., Bullock, H. E., & Fernald, J. (2006). Pair programming improves student retention, confidence, and program quality. *Communications of the ACM, 49*(8), 90–95. https://doi.org/10.1145/1145287.1145293

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* Basic Books.

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1–2), 111–139. https://doi.org/10.1007/s11165-005-3917-8

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom.* Free Press.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. https://doi.org/10.1145/1118178.1118215

Zhang, Y., Luo, R., Zhu, Y., & Yin, Y. (2021). Educational robots improve K-12 students' computational thinking and STEM attitudes: Systematic review. *Journal of Educational Computing Research, 59*(7), 1450–1481. https://doi.org/10.1177/0735633121994070