

Experiential Learning through the Internet of Things in Primary STEM Education: Evidence from a Paired-Sample Pre-Test–Post-Test Study

Nor Hafeizah Hassan^{1*}, Azman Awang Teh², Mashanum Osman³, Muhammad Shahrilazlan Saleh⁴,
Muhammad Faheem Muhd Ezany⁵

^{1,3,5}Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Malaysia

²Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.1026EDU0617>

Received: 17 September 2026; Accepted: 22 September 2026; Published: 06 October 2026

ABSTRACT

Internet of Things (IoT) technologies are increasingly proposed as a vehicle for hands-on, data-rich STEM learning, yet empirical evidence from primary-school settings remains limited. Grounded in Kolb's experiential learning theory and Bandura's self-efficacy theory, this study examined whether self-reported computational thinking self-efficacy (CT-SE), learning attitude towards STEM (LAT-STEM), and programming self-efficacy (P-SE) changed following participation in an IoT-based STEM activity. A one-group pre-test–post-test design was used with 60 matched respondents: 45 primary-school pupils (mean age 11.09 years) and 15 adult observers (mean age 40.47 years). The 25-item instrument (CT-SE, 10 items; LAT-STEM, 8 items; P-SE, 7 items) used a five-point agreement scale and was internally consistent at both occasions (full-scale $\alpha = .92$ pre, $.95$ post). Paired-samples t-tests showed a significant increase in the overall composite from $M = 3.08$ to $M = 3.64$ (mean difference 0.56 , 95% CI $[0.38, 0.74]$, $t(59) = 6.38$, $p < .001$, $d_z = 0.82$). All three constructs increased, most strongly for P-SE ($d_z = 0.94$) and least for LAT-STEM ($d_z = 0.36$). At item level, 22 of 25 items increased at the unadjusted $.05$ level; after Holm correction for multiple comparisons, 16 items remained significant, all in the CT-SE and P-SE sections, whereas no LAT-STEM item survived correction. The LAT-STEM items had the highest baseline means, consistent with a ceiling constraint. Gains did not differ significantly between pupils and adult observers or between genders, although subgroup analyses were underpowered. The findings describe change over time rather than a causally established effect, but they show that a short IoT-based activity is associated with measurable shifts in pupils' self-efficacy beliefs and indicate that attitudinal outcomes shift least.

Keywords: Internet of Things; primary STEM education; experiential learning; computational thinking self-efficacy; pre-test–post-test design

INTRODUCTION

STEM education occupies a central position in contemporary schooling policy because the competencies it develops — reasoning with evidence, modelling, designing, and iterating — are treated as foundations for participation in technologically mediated economies. Reviews of integrated STEM approaches in K–12 settings have documented a broad and growing body of implementation work, while also noting that outcome evidence is uneven across educational levels, with primary schooling represented less often than secondary or tertiary contexts (Abichandani et al., 2022). The relative under-representation of primary education matters, because instructional decisions made in the primary years shape subsequent subject choice and self-concept in science and technology.

Digital technologies have been the principal instrument through which STEM curricula have been extended in the last decade. National policy has moved in the same direction: Malaysia's Digital Education Policy, launched by the Ministry of Education in 2023, sets out digital competence among teachers and pupils, infrastructure

development, and digital content as explicit system priorities (Ministry of Education Malaysia, 2023), building on the technology commitments of the Malaysia Education Blueprint 2013–2025 (Ministry of Education Malaysia, 2013). Bibliometric analysis confirms that research interest in IoT in education has grown steadily and accelerated from around 2020, though the literature remains concentrated in engineering and higher-education outlets rather than school-level educational research (Abdullah et al., 2024).

Within this broader digital turn, the Internet of Things occupies a distinctive pedagogical position. IoT configurations — low-cost microcontrollers, sensors, actuators, and network services — make physical phenomena measurable and visible to learners in real time. Abichandani et al. (2022), reviewing 60 studies of IoT curriculum, pedagogy, and assessment for STEM education, found that the most effective reported implementations combined low-cost hardware, open-source software, and active-learning instructional approaches. Meylani (2024) similarly surveys IoT-enabled learning environments and concludes that the promise of the technology is accompanied by unresolved questions of data security, privacy, and cost. At primary level specifically, Yeh, Cheng, and Lou (2020) developed and evaluated an IoT maker curriculum for primary-school pupils, and Spyropoulou et al. (2020) reported hands-on IoT activities aimed at STEM career awareness, in their case with secondary students.

The pedagogical rationale usually advanced for IoT in schools draws on experiential and constructivist traditions: learners act on a physical system, observe consequences immediately, and revise their approach (Kolb, 1984). Such repeated, successful action is also the principal source of self-efficacy — learners' beliefs in their capability to perform a task (Bandura, 1997). Adjacent evidence is consistent with this rationale without being specific to IoT. Meta-analytic work on programming instruction reports moderate-to-large gains in computational thinking among primary learners, with intra-group pre- to post-test effect sizes around $g = 1.04$ in one synthesis (Merino-Armero et al., 2022) and moderate effects on twenty-first-century skills more broadly ($g = 0.62$; Hu, 2024). Educational robotics, a close technological cousin of IoT, shows positive multilevel meta-analytic effects in STEM education (Ouyang & Xu, 2024) and on computational thinking and STEM attitudes (Zhang et al., 2021). These findings suggest, but do not establish, that IoT-based STEM activities at primary level will produce comparable shifts.

Three limitations recur in the existing evidence base. First, sample populations skew towards secondary and tertiary learners; the affordances documented at those levels may not transfer to eleven-year-olds, whose reading of technical language is a known constraint. Second, implementation studies in IoT education frequently report perceptions and satisfaction rather than learning-related constructs, and often without reliability evidence for the instruments used (Abichandani et al., 2022). Third, much of the school-level IoT literature is descriptive or single-occasion, so change over the course of an activity is rarely quantified.

The research gap addressed here is therefore narrow and concrete: there is little published evidence that quantifies, with paired data and reported effect sizes, how much primary pupils' computational thinking self-efficacy, programming self-efficacy, and attitude towards STEM change across an IoT-based STEM activity, and which of these domains change least. Identifying domains that do not move is as informative for instructional design as identifying those that do, and it is precisely the kind of finding that selective reporting tends to suppress.

The present study addresses that gap using a matched pre- and post-activity survey administered to 60 respondents in a primary-school context, conducted with the knowledge and approval of the State Education Department. Its purpose is to estimate and test the magnitude of change across three constructs and 25 individual items, to report reliability and effect sizes for each, to control for multiple item-level comparisons, and to examine whether change differs by respondent group or gender. The study makes no causal claim. Its contribution is a transparently bounded, fully reported baseline from an under-documented population, together with explicit identification of the items on which no robust change was detected.

LITERATURE REVIEW

IoT in education

In educational usage, the Internet of Things denotes networked physical objects — sensors, actuators, microcontrollers, and the services that connect them — deployed so that learners or institutions can capture, transmit, and act on real-world data. Abdullah et al. (2024), analysing 2,503 Scopus-indexed documents published between 2006 and 2021, report sustained growth in IoT-in-education publications with a marked increase from 2020, but also a dispersed literature spread across 160 outlets, much of it in engineering venues. Meylani (2024) reviews platforms and tools used in educational IoT — wearables, smart devices, immersive technologies, and gamified collaborative environments — and gives sustained attention to data security and the ethical and legal implications of collecting data in classrooms. Dake, Bada, and Dadzie (2023), surveying IoT applications in Ghanaian tertiary institutions, document benefits alongside implementation constraints including cost, infrastructure, and staff capability. The consistent picture across these reviews is of a technology whose educational affordances are plausible and increasingly documented, and whose implementation barriers are equally well documented.

IoT and STEM education

The alignment between IoT and STEM pedagogy is structural rather than incidental. Measurement, data representation, control logic, and iterative design are simultaneously the content of STEM subjects and the operational requirements of an IoT artefact. Abichandani et al. (2022) organise the reviewed literature by the sensing, networking, services, and interface layers of the IoT paradigm and identify active-learning instructional approaches as characteristic of the more effective implementations. Spyropoulou et al. (2020) report hands-on IoT activities designed to raise STEM career awareness among secondary students, with lessons learned that emphasise scaffolding and teacher support. What this literature establishes is that IoT-based STEM activity can be implemented in school settings and assessed; what it leaves uncertain is the size and durability of learning-related effects, particularly below secondary level.

IoT in primary education

Primary-level work is comparatively sparse. Yeh et al. (2020) developed and evaluated an IoT maker curriculum specifically for primary-school pupils, one of the few studies to treat this age group as its principal population. Three considerations distinguish the primary context. Developmentally, abstraction demands must be kept low, which is why block-based environments dominate at this level. Linguistically, technical vocabulary and English-language interfaces present a real access barrier where the language of instruction differs — a barrier that appeared spontaneously in the present study's open-ended responses. Infrastructurally, primary schools typically have less specialised equipment and fewer technically trained staff than secondary schools. Teacher capability is the binding constraint in most accounts, which is why frameworks such as TPACK (Mishra & Koehler, 2006) remain relevant to implementation planning even where, as here, teacher knowledge is not itself measured.

Evidence from adjacent interventions

Because IoT-specific outcome studies at primary level are few, the closest evidential neighbours are programming and robotics interventions. Merino-Armero et al. (2022) report a large intra-group pre- to post-test effect ($g = 1.044$) for computational-thinking interventions in K–12, with programming the most efficient learning tool among those examined. Hu (2024), synthesising 74 studies and 136 effect sizes, reports a moderate overall effect of programming instruction on twenty-first-century skills ($g = 0.621$). Li et al. (2022) compare unplugged activities and programming exercises and find both useful, with programming exercises more effective. Ouyang and Xu (2024) report positive multilevel meta-analytic effects of educational robotics in STEM education, and Zhang et al. (2021) find that educational robots improve computational thinking and STEM attitudes among K–12 students. Sun et al. (2021) report that STEM learning attitude predicts computational thinking skills among primary-school students, which is one reason attitudinal and self-efficacy measures of the kind used here retain educational relevance even though they are not achievement measures. These findings frame expectations for the present study; they do not substitute for IoT-specific evidence, and

the intra-group effects reported in meta-analyses of pre–post designs carry the same causal limitations as the present study.

Theoretical framing

Two complementary theories organise the study. *Experiential learning theory* (Kolb, 1984) describes learning as a cycle of concrete experience, reflective observation, abstract conceptualisation, and active experimentation. An IoT activity maps directly onto this cycle: pupils assemble or program a physical device (concrete experience), observe how sensors and actuators respond (reflective observation), relate that behaviour to programming constructs such as conditions and loops (abstract conceptualisation), and modify their code to test new behaviour (active experimentation). Kolb's theory predicts change across a single activity, which is what the present design measures.

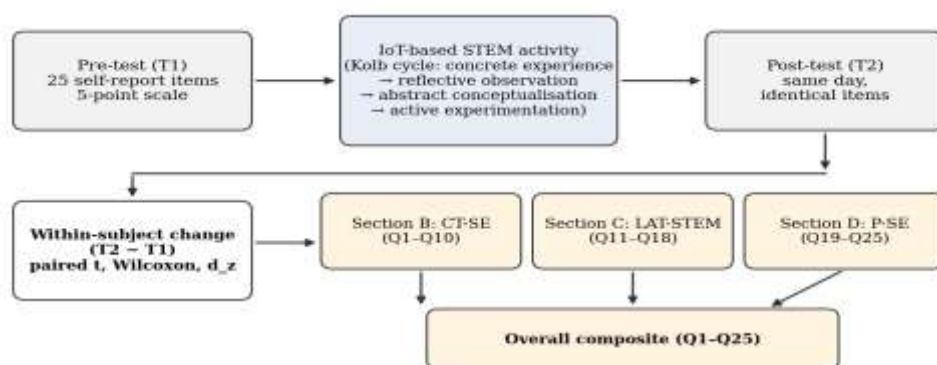
Self-efficacy theory (Bandura, 1997) specifies what that change should look like in the constructs measured. Self-efficacy is a task-specific belief in one's capability to organise and execute the actions needed to reach a goal, and its strongest source is mastery experience — successfully performing the task oneself. Because the IoT activity gives pupils direct, repeated experience of writing, running, and debugging code, self-efficacy theory predicts the largest shifts in programming self-efficacy, somewhat smaller shifts in the more general computational thinking self-efficacy, and the smallest shifts in broad attitudinal dispositions (enjoyment, interest, career aspiration) that are less tied to a single mastery episode. Technology-acceptance perspectives (Davis, 1989; Venkatesh et al., 2003) were considered but not adopted, because the instrument does not measure perceived usefulness, ease of use, or behavioural intention; TPACK (Mishra & Koehler, 2006) and Diffusion of Innovations (Rogers, 2003) concern teacher knowledge and institutional adoption, which are not measured here.

Research gap

Synthesising the above: IoT-in-education research is growing but concentrated above primary level and in non-education outlets; primary-level IoT studies exist but rarely quantify change with paired data, effect sizes, and reliability evidence; adjacent programming and robotics meta-analyses suggest moderate-to-large change but cannot be assumed to transfer. What remains insufficiently examined is a transparent, fully reported estimate of pre-to-post change in computational thinking self-efficacy, programming self-efficacy, and STEM learning attitude in a primary-level IoT STEM activity, including the items on which change is absent. That is the gap this study addresses.

Conceptual Framework

The framework (Figure 1) is deliberately minimal, because the dataset supports only a within-subject change model. The IoT-based activity, structured around Kolb's experiential cycle, is bracketed by identical pre- and post-test administrations. Change is examined on three constructs and their composite.



Grouping variables examined on gain scores only (not as moderators): respondent group (primary pupils, $n = 45$; adult observers, $n = 15$) and gender (pre-test coding).

Figure 1. Conceptual framework of the study: within-subject change across three self-reported constructs, with two observed grouping variables examined on gain scores.

Constructs. *Computational Thinking Self-Efficacy (CT-SE; Section B)* is defined as the learner's belief in their capability to apply core computational thinking practices — decomposition, pattern recognition, generalisation, abstraction, algorithm design, evaluation of alternatives, and explanation of solution logic (Wing, 2006; Bandura, 1997). *Learning Attitude towards STEM (LAT-STEM; Section C)* is defined as the learner's affective and motivational orientation towards STEM learning, including enjoyment, perceived relevance, reduced anxiety, interest, confidence in discussion, perseverance, perceived value of the tools, and career interest (Sun et al., 2021). *Programming Self-Efficacy (P-SE; Section D)* is defined as the learner's belief in their capability to perform specific programming tasks — writing conditionals, debugging, simulating and tracing code, testing, modifying example code, and persisting through errors (Ramalingam & Wiedenbeck, 1998). The sections are further justified empirically by their separation in the instrument and their differing internal reliabilities and change magnitudes.

Proposed relationships. Occasion of measurement is expected to be associated with higher construct scores, with the ordering $P-SE > CT-SE > LAT-STEM$ predicted by self-efficacy theory (Section 2.5). The relationships tested are differences in means, not paths in a structural model; no mediator or moderator is claimed.

Connection to the research questions. RQ1 tests the composite, RQ2 the three constructs, RQ3 the constituent items, RQ4 the grouping variables, and RQ5 the measurement quality that licenses the composites.

Research Questions And Hypotheses

- **RQ1.** Does the overall composite score differ between the pre-activity and post-activity occasions?
- **RQ2.** Do CT-SE, LAT-STEM, and P-SE differ between occasions, and do they differ in the magnitude of change?
- **RQ3.** Which individual items show a statistically significant change, and does this survive correction for multiple comparisons?
- **RQ4.** Do gain scores differ by respondent group (primary pupils vs adult observers) or by gender?
- **RQ5.** Is the instrument sufficiently internally consistent at both occasions to justify the composite scores used in RQ1 and RQ2?

The corresponding hypotheses are: **H1** — the overall composite differs between occasions (paired-samples *t*-test); **H2** — each construct composite differs between occasions (three paired-samples *t*-tests); **H3** — individual items differ between occasions (25 paired-samples *t*-tests with Holm and Benjamini–Hochberg adjustment, reported in full including non-significant results); **H4a** — gain scores differ between pupils and adult observers; **H4b** — gain scores differ between gender groups (independent-samples *t*-tests on gain).

METHODOLOGY

Research design

The study employed a quantitative, one-group pre-test–post-test design — a pre-experimental design — in which the same respondents completed an identical 25-item survey before and after participating in an IoT-based STEM activity. The design permits estimation of within-subject change but provides no counterfactual; it is therefore neither experimental nor quasi-experimental, and no causal inference is drawn (Shadish et al., 2002).

Population and sample

The analytic sample comprises 60 respondents with complete matched pre- and post-activity data, drawn by random sampling from selected schools in one district. Forty-five respondents were pupils at Sekolah Kebangsaan (national primary school) level, aged 11–12 years, and 15 were adults holding diploma, degree, or higher qualifications, aged 28–55 years, who attended the activity as observers. The two groups are structurally

different, so pupil-only results are reported alongside pooled results (Section 6.7), and the adult responses are interpreted cautiously.

Instrument

The instrument was a Malay-language questionnaire prepared by the research team as part of a wider project on interactive programming tools and STEM learning. It comprises five sections. **Section A (demographics)** records gender, age, education level (primary with year, secondary with form, or post-secondary diploma/degree), school type (SK, SJKC, or SJKT), prior programming experience (none; little, under six months; moderate, six months to one year; experienced, over one year), programming tools previously used (Scratch, Python, block-based coding such as Code.org, other, or none), and home access to devices and the internet. **Sections B, C, and D** contain the 25 closed items (Table 1), each rated on a five-point agreement scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). **Section E** is one optional open-ended item inviting comments on the study or questionnaire. The respondent instructions state that all responses are confidential and that a summary of group results will be made available to interested participants.

The instrument was administered in Malay. For reporting purposes the items were translated into English by the authors; the English wording in Table 1 is a faithful rendering of the Malay items. In the items, "interactive tools" refers to the IoT kit and block-based programming environment used in the activity. Internal consistency was computed from the raw data for this study and is reported in Section 6.3; no reliability coefficient was imported from an external source.

Table 1. Constructs and Items of the Instrument (Sections B–D)

Item	Statement (English translation of the Malay original)
Section B — Computational Thinking Self-Efficacy (CT-SE)	
Q1	I can break down a complex problem into smaller, more manageable parts.
Q2	I can identify patterns and similarities across different problems.
Q3	I can generalise a solution so that it works for new, related problems.
Q4	I can select and define suitable variables to represent a problem.
Q5	I can design a step-by-step procedure (algorithm) to solve a problem.
Q6	I can decide which details are important and ignore irrelevant information (abstraction).
Q7	I can trace a solution and explain why each step is necessary.
Q8	I can restructure a problem when my first approach does not work.
Q9	I can evaluate alternative solution strategies and choose the best one.
Q10	I can explain the logic of my problem solving clearly to others.
Section C — Learning Attitude towards STEM (LAT-STEM)	
Q11	I enjoy learning with interactive programming tools in my STEM lessons.
Q12	STEM topics feel relevant to my future goals.
Q13	I feel less anxious when experimenting with ideas using interactive tools.

Q14	Working with these tools makes STEM more interesting for me.
Q15	I feel confident discussing STEM ideas with my classmates.
Q16	I persevere when STEM tasks are challenging.
Q17	I believe these tools help me learn STEM concepts more deeply.
Q18	I am interested in exploring a STEM career in the future.
Section D — Programming Self-Efficacy (P-SE)	
Q19	I can write and evaluate conditional statements (e.g., if/else) correctly.
Q20	I can debug a program by finding and fixing errors efficiently.
Q21	I can use interactive tools to simulate the behaviour of my code.
Q22	I can trace the execution of a program with nested loops or conditionals.
Q23	I can design tests to check whether my program works as expected.
Q24	I can modify example code to solve a new problem.
Q25	Even after encountering several errors, I can keep trying to produce a working program.

Note. All items rated 1 = Strongly Disagree to 5 = Strongly Agree. Items are presented in the order administered.

Intervention

The IoT-based STEM activity was a structured, hands-on session in which participants worked with an IoT kit through a block-based programming environment. Consistent with Kolb's cycle, participants built and ran code blocks to control the device, observed the device's response, and revised their code to complete a small project. The open-ended responses refer to code, code blocks, and a project, consistent with this format. The pre-test and post-test were administered on the same day, immediately before and immediately after the activity.

Data collection

The pre-activity survey was administered before, and the post-activity survey after, participation; both were linked by respondent identifier, yielding a fully matched dataset with no attrition and no missing item responses. Administration was monitored by the research team, who were available to clarify wording for younger respondents.

Ethical considerations

The study was conducted as one of the STEM activities approved by the State Education Department. Participation was voluntary, respondents were informed in the questionnaire instructions that their responses would be kept confidential, and no identifying information is reported.

Data analysis

All analyses were conducted on the matched raw data, with $\alpha = .05$, two-tailed. Descriptive statistics were computed for each item and construct at both occasions. Cronbach's alpha was computed for each section and for the full scale at each occasion (RQ5). Shapiro–Wilk tests, skewness, and kurtosis were computed on the difference scores, because the paired *t*-test's normality assumption concerns the differences. Paired-samples *t*-tests were used to test change at composite and item level (H1–H3), with Wilcoxon signed-rank tests as a non-

parametric check given the ordinal nature of single Likert items. Cohen's d_z (mean difference $\div$ SD of differences) and 95% confidence intervals were computed for every comparison and interpreted against benchmarks of 0.20, 0.50, and 0.80 (Cohen, 1988).

Because 25 item-level tests inflate the family-wise error rate, item-level p -values were adjusted using two procedures: the Holm (1979) sequential procedure, which controls the family-wise error rate and is uniformly more powerful than Bonferroni, and the Benjamini–Hochberg (1995) procedure, which controls the false discovery rate. The four composite tests (three constructs and overall) were likewise Holm-adjusted as a family. Pearson correlations among construct composites described the covariance structure at each occasion. Independent-samples t -tests on gain scores were used for H4a and H4b, with Levene's test screening the equal-variance assumption, and within-group paired tests were run separately for pupils and adult observers.

RESULTS

Respondent characteristics

Table 2 presents the demographic composition of the sample.

Table 2. Demographic Characteristics of Respondents (N = 60)

Characteristic	Category	n	%
Respondent group	Primary pupils (Sekolah Kebangsaan)	45	75.0
	Adult observers (diploma/degree and above)	15	25.0
Gender (pre-test coding)	Male	34	56.7
	Female	26	43.3
Age group	11–12 years	45	75.0
	28–55 years	15	25.0
Total		60	100.0

Note. Age: primary pupils $M = 11.09$ years, $SD = 0.29$; adult observers $M = 40.47$ years, $SD = 8.98$.

Descriptive statistics

Responses ranged from 1 to 5 on every item at both occasions. Across all 1,500 item responses per occasion, the proportion in the two lowest categories fell from 31.7% at pre-test to 11.7% at post-test, while the proportion in the two highest categories rose from 37.1% to 59.8%. The three constructs differed substantially at baseline: LAT-STEM recorded the highest pre-test mean ($M = 3.69$, $SD = 0.66$), whereas CT-SE ($M = 2.83$, $SD = 0.77$) and P-SE ($M = 2.75$, $SD = 0.72$) started below the scale midpoint. Pupils thus began the activity positive about STEM but unsure of their own computational and programming capability — a difference that is central to interpreting the change results.

Reliability

Table 3. Internal Consistency of the Instrument at Both Occasions (N = 60)

Scale	Items (k)	Cronbach's α (pre)	Cronbach's α (post)
CT-SE (Section B; Q1–Q10)	10	.905	.913

LAT-STEM (Section C; Q11–Q18)	8	.848	.903
P-SE (Section D; Q19–Q25)	7	.853	.866
Full scale (Q1–Q25)	25	.923	.945

All coefficients exceed the conventional .70 threshold, so the construct composites are defensible on reliability grounds (RQ5). Two cautions apply: alpha rises with the number of items, so the full-scale value should not be read as evidence of one-dimensionality; and alpha is a property of these responses in this sample, not of the instrument in general.

Correlation analysis

Table 4. Pearson Correlations Among Construct Composites (N = 60)

	1	2	3	4	5	6
1. Pre CT-SE	—					
2. Pre LAT-STEM	.434**	—				
3. Pre P-SE	.683**	.370**	—			
4. Post CT-SE	.335**	.374**	.237	—		
5. Post LAT-STEM	.157	.092	.146	.540**	—	
6. Post P-SE	.230	.271*	.235	.794**	.572**	—

Note. $p < .05$, $p^* < .01$, two-tailed.

Within each occasion, the construct composites are positively intercorrelated, and the particularly strong association between CT-SE and P-SE ($r = .683$ pre, $.794$ post) is consistent with both being capability beliefs about related computing tasks. Across occasions, associations are weaker: the pre–post correlation is .335 ($p = .009$) for CT-SE, .092 ($p = .483$) for LAT-STEM, and .235 ($p = .070$) for P-SE. Respondents therefore did not simply shift upward while preserving their rank order. These are associations only.

Change by construct and overall

Shapiro–Wilk tests on the difference scores were non-significant for all four composites (CT-SE $W = .991$, $p = .925$; LAT-STEM $W = .973$, $p = .199$; P-SE $W = .986$, $p = .744$; overall $W = .982$, $p = .540$), and skewness and kurtosis were within ± 1.3 , supporting the paired t -test. Wilcoxon signed-rank tests agreed with every t -test decision.

Table 5. Paired-Samples t -Tests for Construct and Overall Composites (N = 60)

Scale	M pre (SD)	M post (SD)	Diff.	95% CI	$t(59)$	p	p_{Holm}	d_z
CT-SE (Q1–Q10)	2.83 (0.77)	3.40 (0.66)	0.58	[0.36, 0.79]	5.37	$< .001$	$< .001$	0.69
LAT-STEM (Q11–Q18)	3.69 (0.66)	4.00 (0.66)	0.32	[0.09, 0.55]	2.75	.008	.008	0.36
P-SE (Q19–Q25)	2.75 (0.72)	3.57 (0.69)	0.82	[0.59, 1.04]	7.26	$< .001$	$< .001$	0.94
Overall (Q1–Q25)	3.08 (0.59)	3.64 (0.58)	0.56	[0.38, 0.74]	6.38	$< .001$	$< .001$	0.82

Note. Diff. = post – pre. d_z = Cohen's d for paired samples. p_{Holm} = Holm-adjusted across the four composite tests. All tests two-tailed.

H1 is supported. The overall composite increased by 0.56 scale points, $t(59) = 6.38, p < .001$, with a large paired effect ($d_z = 0.82$). **H2 is supported for all three constructs**, including after Holm adjustment, but the magnitudes differ markedly and follow the ordering predicted by self-efficacy theory: P-SE shows a large effect ($d_z = 0.94$), CT-SE a medium-to-large effect ($d_z = 0.69$), and LAT-STEM a small-to-medium effect ($d_z = 0.36$) whose confidence interval extends close to zero. LAT-STEM also had the highest pre-test mean, leaving less room for measured increase.

Item-level change and correction for multiple comparisons

Table 6. Item-Level Paired-Samples t-Tests with Multiple-Comparison Adjustment (N = 60)

Item	<i>M</i> pre (<i>SD</i>)	<i>M</i> post (<i>SD</i>)	Diff.	95% CI	<i>t</i> (59)	<i>p</i>	<i>p</i> _{Holm}	<i>q</i> _{BH}	<i>d_z</i>
CT-SE (Section B)									
Q1	2.60 (1.03)	3.23 (0.91)	0.63	[0.30, 0.97]	3.77	< .001	.005	< .001	0.49
Q2	2.70 (0.98)	3.32 (0.72)	0.62	[0.32, 0.91]	4.20	< .001	.002	< .001	0.54
Q3	2.50 (1.05)	3.18 (0.98)	0.68	[0.38, 0.99]	4.46	< .001	< .001	< .001	0.58
Q4	2.77 (1.16)	3.43 (0.81)	0.67	[0.35, 0.99]	4.15	< .001	.002	< .001	0.54
Q5	2.93 (0.99)	3.55 (0.98)	0.62	[0.32, 0.91]	4.15	< .001	.002	< .001	0.54
Q6	2.80 (1.16)	3.40 (0.87)	0.60	[0.23, 0.97]	3.28	.002	.019	.003	0.42
Q7	2.85 (0.92)	3.32 (0.83)	0.47	[0.21, 0.72]	3.68	< .001	.007	.001	0.48
Q8	3.08 (1.03)	3.65 (0.92)	0.57	[0.24, 0.90]	3.43	.001	.013	.002	0.44
Q9	2.97 (1.07)	3.48 (0.93)	0.52	[0.17, 0.86]	2.97	.004	.043	.007	0.38
Q10	3.05 (1.05)	3.43 (0.89)	0.38	[0.08, 0.68]	2.55	.014	.121†	.020	0.33
LAT-STEM (Section C)									
Q11	3.88 (0.99)	4.20 (0.75)	0.32	[0.01, 0.63]	2.04	.045	.214†	.052†	0.26
Q12	3.55 (0.85)	3.88 (0.88)	0.33	[0.07, 0.60]	2.53	.014	.121†	.020	0.33
Q13	3.32 (1.00)	3.80 (0.99)	0.48	[0.08, 0.89]	2.40	.019	.137†	.026	0.31
Q14	3.80 (0.95)	4.05 (0.77)	0.25	[−0.07, 0.57]	1.56	.125	.372†	.135†	0.20
Q15	3.92 (0.81)	4.10 (0.82)	0.18	[−0.10, 0.47]	1.29	.200	.372†	.202†	0.17
Q16	3.45 (1.08)	3.87 (0.93)	0.42	[0.05, 0.78]	2.30	.025	.150†	.031	0.30
Q17	3.75 (0.97)	4.07 (0.86)	0.32	[0.01, 0.62]	2.07	.043	.214†	.051†	0.27
Q18	3.82 (0.95)	4.05 (0.81)	0.23	[−0.11, 0.58]	1.36	.180	.372†	.187†	0.18
P-SE (Section D)									
Q19	2.75 (1.14)	3.58 (0.91)	0.83	[0.50, 1.17]	4.95	< .001	< .001	< .001	0.64

Q20	2.52 (0.91)	3.30 (0.94)	0.78	[0.44, 1.13]	4.57	< .001	< .001	< .001	0.59
Q21	2.67 (0.93)	3.55 (0.91)	0.88	[0.57, 1.20]	5.60	< .001	< .001	< .001	0.72
Q22	2.67 (0.91)	3.45 (0.93)	0.78	[0.44, 1.13]	4.57	< .001	< .001	< .001	0.59
Q23	2.80 (0.94)	3.65 (0.88)	0.85	[0.55, 1.15]	5.67	< .001	< .001	< .001	0.73
Q24	2.73 (0.92)	3.52 (1.03)	0.78	[0.46, 1.10]	4.91	< .001	< .001	< .001	0.63
Q25	3.13 (1.11)	3.95 (0.91)	0.82	[0.49, 1.14]	5.04	< .001	< .001	< .001	0.65

Note. p = unadjusted two-tailed p ; p_{Holm} = Holm (1979) family-wise adjusted p across 25 items; q_{BH} = Benjamini–Hochberg (1995) false-discovery-rate adjusted value. † = not significant at .05 after adjustment. Adjusted values were computed from the reported t statistics ($df = 59$). Wilcoxon signed-rank tests agreed with every unadjusted decision except Q17 (Wilcoxon $p = .075$).

H3 is partly supported. At the unadjusted .05 level, 22 of 25 items increased significantly; Q14 ($p = .125$), Q15 ($p = .200$), and Q18 ($p = .180$) did not. After Holm correction, **16 items remained significant — all ten CT-SE items and all seven P-SE items — whereas none of the eight LAT-STEM items survived.** Under the less conservative Benjamini–Hochberg procedure, 20 items remained significant; the five that did not (Q11, Q14, Q15, Q17, Q18) are all LAT-STEM items. The strongest item-level effects were concentrated in P-SE: using interactive tools to simulate code behaviour (Q21, $d_z = 0.72$), designing tests for a program (Q23, $d_z = 0.73$), and persisting through errors (Q25, $d_z = 0.65$). The items that did not change concern whether the tools make STEM more interesting (Q14), confidence in discussing STEM with classmates (Q15), and interest in a STEM career (Q18); all three had pre-test means at or above 3.80, the highest baselines in the instrument.

Gain scores by respondent group and gender

Table 7. Independent-Samples t-Tests on Gain Scores (N = 60)

Comparison	Group 1	Group 2	$t(58)$	p	Cohen's d
Overall gain by group	Pupils (n = 45): $M = 0.57$	Adults (n = 15): $M = 0.55$	0.09	.928	0.03
CT-SE gain by group	Pupils: $M = 0.60$	Adults: $M = 0.49$	0.44	.663	0.13
LAT-STEM gain by group	Pupils: $M = 0.32$	Adults: $M = 0.31$	0.04	.967	0.01
P-SE gain by group	Pupils: $M = 0.79$	Adults: $M = 0.90$	−0.39	.700	−0.12
Overall gain by gender	Male (n = 34): $M = 0.50$	Female (n = 26): $M = 0.64$	−0.74	.462	−0.19

Note. Levene's test was non-significant for every comparison (all $p > .09$). Gender uses the pre-test coding.

Neither H4a nor H4b is supported. Gains did not differ significantly between pupils and adult observers on any composite, nor between genders. Because it was rightly noted that pooling two structurally different populations complicates interpretation, the groups were also analysed separately. Among pupils alone, the overall composite rose significantly ($M_{\text{diff}} = 0.57$, $t(44) = 5.30$, $p < .001$, $d_z = 0.79$), so the main finding does not depend on the adult observers. Adult observers also changed ($M_{\text{diff}} = 0.55$, $t(14) = 3.63$, $p = .003$, $d_z = 0.94$), but with 15 respondents this estimate is imprecise. A non-significant group difference is absence of evidence for a difference, not evidence of equivalence.

Open-ended responses (Section E)

Five respondents commented at pre-test and nine at post-test (8.3% and 15.0%) — too few for thematic analysis. Several pre-test comments raised the accessibility of technical and English-language vocabulary, one post-test comment repeated that concern, and other post-test comments expressed positive affect or partial comprehension. These remarks are consistent with the language-access barrier noted in Section 2.3 but carry no evidential weight.

DISCUSSION

Overall change (RQ1)

The overall composite rose from 3.08 to 3.64 ($d_z = 0.82$; 95% CI [0.38, 0.74]). Respondents answered the same items more favourably after the activity, and the shift was consistent enough across individuals to produce a large standardised effect. The magnitude sits between the intra-group effects reported by Merino-Armero et al. (2022; $g = 1.044$) and Hu (2024; $g = 0.621$), although these comparisons offer orientation rather than benchmarking, since the meta-analyses aggregate skill and achievement measures whereas the present instrument measures self-report.

Differential change across constructs (RQ2)

P-SE changed most ($d_z = 0.94$), CT-SE substantially ($d_z = 0.69$), and LAT-STEM least ($d_z = 0.36$). Two explanations are consistent with this ordering, and the named constructs now make it possible to weigh them. The first is **substantive and theory-predicted**. The activity gave pupils direct mastery experiences in exactly the tasks P-SE asks about — running code, simulating the device's behaviour, finding and fixing errors, and trying again. Bandura (1997) identifies mastery experience as the strongest source of self-efficacy, and the largest item-level gains (Q21, Q23, Q25) are the tasks pupils actually performed. CT-SE items describe more general problem-solving practices that the activity exercised less directly, and they moved less. LAT-STEM items describe dispositions — interest, career aspiration, confidence in discussion — that are generally more stable and less likely to shift in a single session.

The second explanation is **measurement-based**: LAT-STEM began at 3.69 of 5, while CT-SE and P-SE began below the midpoint, so a ceiling constraint limited how much LAT-STEM could register. The two explanations are not mutually exclusive, and the data suggest both operate: pupils arrived enthusiastic about STEM but unsure of their own ability, and the activity mainly addressed the latter. This pattern mirrors Zhang et al. (2021), who found robotics interventions affect computational thinking more consistently than STEM attitudes, and Sun et al. (2021), who treat attitude as a relatively stable predictor rather than a rapidly changing outcome.

Items showing no robust change (RQ3)

After Holm correction, no LAT-STEM item remained significant, and even under the more lenient false-discovery-rate control, Q11 (enjoyment), Q14 (tools make STEM more interesting), Q15 (confidence in discussing STEM), Q17 (tools help learn deeply), and Q18 (STEM career interest) did not. These items share two features: all had high pre-test means (3.75–3.92), and all concern affective or aspirational orientations rather than task capability. Career interest in particular (Q18) is unlikely to shift in a single same-day session. Reporting these results matters: the widespread change in self-efficacy is robust to correction, whereas the evidence for attitudinal change is fragile, and the exceptions cluster systematically rather than randomly. Full item-level reporting with adjustment, including nulls, addresses a recognised weakness of the IoT-education literature, in which perception measures are frequently reported without reliability evidence or complete item-level results (Abichandani et al., 2022).

Gains across groups (RQ4)

Neither respondent group nor gender was associated with significantly different gains, and the pupil-only analysis replicates the pooled result. The absence of a gender difference in self-reported gain is encouraging for

inclusive practice, though it concerns self-report rather than performance. For adult observers, the result should be read with particular caution: the items are worded for learners describing their own capability, and adults who observed rather than led the activity may have interpreted them differently. The underpowered comparison ($n = 15$) cannot show that the activity "works equally well" for both groups.

Measurement quality (RQ5)

Internal consistency was high at both occasions ($\alpha = .92$ pre, $.95$ post for the full scale; $.85$ – $.91$ for constructs), supporting the use of composites. However, the weak cross-occasion correlations — particularly for LAT-STEM ($r = .092$) — indicate that the instrument's rank ordering of respondents was not stable across occasions. A scale can be internally consistent at each administration while ordering people differently at each; this is a caution about the instrument that reliability coefficients alone would have hidden.

Educational Implications

These implications apply to settings comparable to the study sample and should not be generalised further.

Primary-school teachers. A short IoT-based activity was associated with substantial increases in pupils' confidence in programming and computational thinking — domains in which pupils began below the scale midpoint. Teachers can capitalise on this by structuring activities around repeated cycles of build–run–debug, which provide the mastery experiences associated with the largest gains. Positive attitudes towards STEM, by contrast, were already high and did not change robustly; sustaining interest and career aspiration likely requires repeated exposure, role models, and links to real-world careers rather than a single session. The recurring references to technical vocabulary in the open-ended comments, although evidentially weak, warrant vocabulary scaffolding, including Malay-language glossaries for programming terms.

Curriculum developers. The clustering of null results among high-baseline attitude items indicates a measurement design issue as much as a curriculum one. Developers of evaluation instruments for primary IoT activities should consider wider response scales or behavioural indicators for constructs on which pupils are expected to start high, and should separate capability beliefs from attitudes, as this instrument does.

Educational policymakers. The Digital Education Policy's emphasis on digital content and competence at school level (Ministry of Education Malaysia, 2023) presumes that classroom digital activity produces measurable change; this study offers one small, uncontrolled data point consistent with that premise in a primary setting. It is not a basis for procurement or mandate decisions.

Teacher professional development. Adult observers also reported significant change. Because the instrument was designed for learners, this should not be over-interpreted, but it suggests that observing pupil-facing IoT activity may itself shift adults' confidence — a possibility worth examining with an instrument designed for teachers.

Limitations

1. **No comparison group.** The one-group pre-test–post-test design cannot rule out maturation, testing effects, instrumentation, regression to the mean, or history (Shadish et al., 2002). No causal attribution to the IoT activity is warranted.
2. **Testing, demand, and social-desirability effects.** The identical instrument was administered immediately before and after the activity on the same day. Recall of earlier answers, the expectation that a programme should produce improvement, and the wish to please the facilitators may all have inflated post-test scores.
3. **Self-report only.** No achievement, performance, or observational measure was collected. The study speaks to what respondents believed about their capability and attitudes, not to what they learned or can demonstrably do.

4. **Common-method bias.** All variables come from a single self-report instrument administered in the same way at both occasions.
5. **Sample size and composition.** $N = 60$ is modest. The adult-observer subgroup ($n = 15$) is too small for adequately powered comparison, and it responded to learner-oriented items, so pooling it with pupils assumes that the items meant the same thing to both groups. The pupil-only analysis mitigates but does not remove this concern.
6. **Generalisability.** Participants were drawn from selected schools in one district, and generalisation beyond comparable settings is not supported.
7. **Instrument validation.** The instrument's factor structure has not been confirmed, and the English item wordings are authors' translations that have not been independently back-translated.
8. **Scale compression.** The five-point scale limits the range available for change on high-baseline attitude items.
9. **Multiple comparisons.** Item-level tests were adjusted with Holm and Benjamini–Hochberg procedures. The Holm results are conservative, and some LAT-STEM items may reflect real but small changes that this sample was too small to detect after correction.
10. **Open-ended data.** With five and nine responses, Section E cannot support qualitative inference.

RECOMMENDATIONS FOR FUTURE RESEARCH

1. **Add a non-equivalent comparison group.** A quasi-experimental design in which a comparable class receives conventional STEM instruction over the same period would provide a defensible estimate of the IoT activity's contribution beyond testing and demand effects.
2. **Supplement self-report with objective measures.** Validated computational-thinking assessments, performance tasks (e.g., debugging a faulty program within a set time), structured classroom observation, and project-based achievement rubrics would test whether reported gains in self-efficacy correspond to demonstrated capability.
3. **Include a delayed post-test.** Re-administering the instrument several weeks after the activity would show whether gains persist beyond the immediate post-activity window, and would reduce the influence of same-day recall.
4. **Use differentiated instruments and separate analyses.** Pupils and adults should be assessed with instruments designed for each — for example, a learner self-efficacy scale for pupils and an observation or teacher-efficacy instrument for adults — and analysed separately with adequate subsamples.
5. **Validate and refine the instrument.** A formal validation study — expert content review, back-translation between Malay and English, and confirmatory factor analysis with an adequate sample (at least 250 respondents for 25 items) — would confirm the three-construct structure. Wider response scales for high-baseline attitude items would reduce ceiling effects.
6. **Plan for multiple comparisons.** Future studies should pre-specify composite-level primary outcomes and apply family-wise or false-discovery-rate correction to item-level analyses from the outset.
7. **Target STEM attitude with sustained designs.** Because attitudes did not shift robustly in a single session, longer multi-session IoT programmes should test whether repeated mastery experiences eventually carry over into interest and career aspiration.
8. **Investigate the language barrier.** The vocabulary concerns in the open-ended responses warrant dedicated mixed-methods study.

CONCLUSION

This study examined whether primary pupils' self-reported computational thinking self-efficacy, programming self-efficacy, and learning attitude towards STEM changed across participation in an IoT-based STEM activity. Using matched data from 60 respondents, it found a significant overall increase of 0.56 scale points ($t(59) = 6.38, p < .001, d_z = 0.82$), with significant increases on all three constructs. Change was not uniform: consistent with self-efficacy theory, it was largest for programming self-efficacy ($d_z = 0.94$), substantial for computational thinking self-efficacy ($d_z = 0.69$), and smallest for STEM attitude ($d_z = 0.36$). After correction for multiple comparisons, all self-efficacy items remained significant while no attitude item did. The main finding held among pupils alone, and gains did not differ significantly by group or gender.

The contribution is a transparent, fully reported estimate of pre-to-post change in a population the IoT-education literature documents thinly, including effect sizes, confidence intervals, reliability evidence, adjusted item-level results, and the null results that selective reporting would omit. The finding that hands-on IoT activity is most strongly associated with pupils' confidence in programming, and least with their already-positive attitudes, is directly useful for designing and evaluating comparable activities. Without a comparison group, the study establishes association with the activity period, not a programme effect; without performance measures, it establishes change in what respondents reported, not in what they learned. The recommended comparison-group, objective-measure, and delayed post-test designs are the next steps needed to settle that question.

ACKNOWLEDGEMENT

This study is part of the International Research Grant (Geran Penyelidikan Antarabangsa) ANTARABANGSA-ONSEMI/2025/FTMK/A00096, administered by the Centre for Research and Innovation Management (CRIM; Pusat Pengurusan Penyelidikan dan Inovasi), Universiti Teknikal Malaysia Melaka (UTeM). The authors thank CRIM and UTeM for their support, and the State Education Department, the participating schools, pupils, and adult observers for their cooperation.

REFERENCES

1. Abdullah, K. H., Gazali, N., Muzawi, R., Syam, E., Roslan, M. F., & Sofyan, D. (2024). Internet of Things (IoT) in education: A bibliometric review. *International Journal of Information Science and Management*, 22(1), 183–202. <https://doi.org/10.22034/ijism.2023.1977600.0>
2. Abichandani, P., Sivakumar, V., Lobo, D., Iaboni, C., & Shekhar, P. (2022). Internet-of-Things curriculum, pedagogy, and assessment for STEM education: A review of literature. *IEEE Access*, 10, 38351–38369. <https://doi.org/10.1109/ACCESS.2022.3164709>
3. Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
4. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
5. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
6. Dake, D. K., Bada, G. K., & Dadzie, A. E. (2023). Internet of Things (IoT) applications in education: Benefits and implementation challenges in Ghanaian tertiary institutions. *Journal of Information Technology Education: Research*, 22, 311–338. <https://doi.org/10.28945/5183>
7. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
8. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
9. Hu, L. (2024). Programming and 21st century skill development in K-12 schools: A multidimensional meta-analysis. *Journal of Computer Assisted Learning*, 40(2), 610–630. <https://doi.org/10.1111/jcal.12904>
10. Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

11. Li, F., Wang, X., He, X., Cheng, L., & Wang, Y. (2022). The effectiveness of unplugged activities and programming exercises in computational thinking education: A meta-analysis. *Education and Information Technologies*, 27(6), 7993–8013. <https://doi.org/10.1007/s10639-022-10915-x>
12. Merino-Armero, J. M., González-Calero, J. A., & Cózar-Gutiérrez, R. (2022). Computational thinking in K-12 education: An insight through meta-analysis. *Journal of Research on Technology in Education*, 54(3), 410–437. <https://doi.org/10.1080/15391523.2020.1870250>
13. Meylani, R. (2024). Transforming education with the Internet of Things: A journey into smarter learning environments. *International Journal of Research in Education and Science*, 10(1), 161–178. <https://doi.org/10.46328/ijres.3362>
14. Ministry of Education Malaysia. (2013). Malaysia education blueprint 2013–2025 (preschool to post-secondary education). Kementerian Pendidikan Malaysia.
15. Ministry of Education Malaysia. (2023). Digital education policy. Kementerian Pendidikan Malaysia.
16. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
17. Ouyang, F., & Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education*, 11(1), Article 7. <https://doi.org/10.1186/s40594-024-00469-4>
18. Ramalingam, V., & Wiedenbeck, S. (1998). Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *Journal of Educational Computing Research*, 19(4), 367–381. <https://doi.org/10.2190/C670-Y3C8-LTJ1-CT3P>
19. Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
20. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
21. Spyropoulou, N., Glaroudis, D., Iossifides, A., & Zaharakis, I. D. (2020). Fostering secondary students' STEM career awareness through IoT hands-on educational activities: Experiences and lessons learned. *IEEE Communications Magazine*, 58(2), 86–92. <https://doi.org/10.1109/MCOM.001.1900288>
22. Sun, L., Hu, L., Yang, W., Zhou, D., & Wang, X. (2021). STEM learning attitude predicts computational thinking skills among primary school students. *Journal of Computer Assisted Learning*, 37(2), 346–358. <https://doi.org/10.1111/jcal.12493>
23. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
24. Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
25. Yeh, C.-Y., Cheng, Y.-M., & Lou, S.-J. (2020). An Internet of Things (IoT) maker curriculum for primary school students: Develop and evaluate. *International Journal of Information and Education Technology*, 10(12), 889–896. <https://doi.org/10.18178/ijiet.2020.10.12.1475>
26. 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>