Content Validity and Reliability of an Instrument Measuring AI Acceptance, AI Self-Efficacy and AI Competency among Computer Science and Basic Computer Science Teachers: A Pilot Study

Dg. Norlina Junit*, Norhasyimah Hamzah

Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia

DOI: https://doi.org/10.47772/IJRISS.2026.101100152

Received: 05 November 2026; Accepted: 10 November 2026; Published: 23 November 2026

ABSTRACT

Artificial intelligence (AI) is increasingly used in teaching and learning, yet there remains a lack of contextually validated instruments for measuring teachers' AI acceptance, AI self-efficacy and AI competency in Malaysia. This pilot study examined the content validity and internal consistency of a 48-item questionnaire measuring AI acceptance, AI self-efficacy and AI competency among secondary school Computer Science (CS) and Basic Computer Science (BCS) teachers. Items were drawn from three established sources, namely the Unified Theory of Acceptance and Use of Technology (UTAUT), the Computer Self-Efficacy scale and Intelligent-TPACK, and were answered on a six-point agreement scale. Five experts were invited, and four returned complete evaluations of face and content validity. The revised instrument was then administered online to 30 CS and BCS teachers from three districts in Sabah. All experts agreed with the face validity criteria. The scale-level content validity index was high (S-CVI/Ave = .984; S-CVI/UA = .938), with 45 of 48 items achieving an item-level CVI of 1.00. Fifteen items were revised based on expert feedback. Cronbach's alpha was .962 for AI acceptance, .918 for AI self-efficacy and .986 for AI competency, while dimension-level alphas ranged from .919 to .970, and McDonald's omega values were similar (.920–.987). A sensitivity analysis excluding five respondents with uniform response patterns showed that alphas remained satisfactory (constructs: .879–.981; dimensions: .893–.960). A ceiling effect was observed at the item level in this pilot sample. These findings provide preliminary evidence of content validity and internal consistency in the pilot sample, while the instrument's dimensional structure and discriminant validity remain to be evaluated through partial least squares structural equation modelling (PLS-SEM) in the main study.

Keywords: artificial intelligence, self-efficacy, teacher AI competency, content validity, pilot study

INTRODUCTION

Artificial intelligence (AI) is increasingly used in education to support the development of learning materials, learning analytics and feedback provision (UNESCO, 2023). In Malaysia, the Digital Education Policy aims to produce teachers capable of embedding digital technology purposefully in their teaching (Ministry of Education Malaysia, 2023). However, access to technology and positive attitudes alone do not guarantee that teachers use AI meaningfully in the classroom, particularly when training and professional support are limited (Fundi et al., 2024).

Under the Standard Curriculum for Secondary Schools (KSSM) introduced in 2017, Basic Computer Science (BCS; *Asas Sains Komputer*) is taught in Forms 1–3 and Computer Science (CS; *Sains Komputer*) in Forms 4–5 (Curriculum Development Division, 2020). The authors' main study investigates whether AI self-efficacy explains how AI acceptance relates to AI competency among CS and BCS teachers in Sabah. AI acceptance is measured using three UTAUT constructs (Venkatesh et al., 2003), AI self-efficacy is adapted from the Computer Self-Efficacy scale (Compeau & Higgins, 1995), and AI competency is adapted from the Intelligent-TPACK framework (Çelik, 2023). Since the original scales were developed in different contexts and languages, their validity and reliability must be re-established before use among Malaysian teachers.

A pilot study plays an important role in assessing item comprehension, administration feasibility and preliminary reliability before the main study (Hertzog, 2008). Accordingly, this study had the following three objectives.

1. To evaluate the face and content validity of the instrument based on expert review.
2. To identify item revisions required based on expert and respondent feedback.
3. To evaluate the internal consistency reliability of the instrument for each construct and dimension.

LITERATURE REVIEW

AI acceptance and the Unified Theory of Acceptance and Use of Technology

Explanations of why teachers embrace or avoid new technologies have relied heavily on acceptance models. UTAUT, developed by Venkatesh et al. (2003), synthesised eight earlier models and proposed four key predictors, namely performance expectancy, effort expectancy, social influence and facilitating conditions. Put simply, performance expectancy captures the belief that a technology will help one do the job better, effort expectancy reflects how easy the technology seems to use, and facilitating conditions describe whether the necessary organisational and technical support is available.

Acceptance models have repeatedly accounted for a meaningful proportion of teachers' intentions to use technology. Synthesising 45 studies, Scherer and Teo (2019) reported that Technology Acceptance Model variables explained about 39% of the variance in teachers' intentions, although the weight of each predictor differed between samples. A larger meta-analytic study likewise identified perceived usefulness and perceived ease of use as key drivers of teachers' adoption of digital technology (Scherer et al., 2019).

Three UTAUT predictors, namely performance expectancy, effort expectancy and facilitating conditions, were retained to represent AI acceptance in this study. AI acceptance therefore refers to teachers' perceptions of the usefulness and ease of use of AI-based tools and of the support available for using them, and it does not directly measure behavioural intention. Social influence and behavioural intention were omitted because the main study is concerned with how teachers themselves view AI and the support around it, rather than with forecasting their intention to use it. Moreover, the use of AI in teaching is currently voluntary for Malaysian teachers, and Venkatesh et al. (2003) found that social influence was significant only in mandatory usage contexts. AI acceptance is positioned as an antecedent of teachers' AI self-efficacy and AI competency.

AI self-efficacy

Bandura (1977, 1997) described self-efficacy as a person's confidence in being able to plan and carry out the actions needed to reach a particular goal. This confidence shapes which tasks people take on, how much effort they make and how long they persist when obstacles arise. Such beliefs are fed by four sources, namely personal mastery, observing others, encouragement from others, and one's physical and emotional reactions (Bandura, 1997).

Compeau and Higgins (1995) extended this idea to computing, framing computer self-efficacy as how capable people judge themselves to be when working with computers. Their measure asked respondents to rate their confidence in completing a task with an unfamiliar software package under different levels of support, such as having no one to help, having only manuals for reference, or having someone show them how to do it first. This approach recognises that self-efficacy depends on the level of support available. The AI self-efficacy items in this study follow the same logic, asking teachers about their confidence in using AI, integrating AI into teaching and solving AI-related problems without guidance, with guides or references, and with guidance from others.

Teacher AI competency and Intelligent-TPACK

The Technological Pedagogical Content Knowledge (TPACK) framework (Mishra & Koehler, 2006) is widely used to describe what teachers need to know to teach well with technology, stressing how technological,

pedagogical and content knowledge work together. With AI now entering schools, scholars contend that teachers also need knowledge specific to AI. Ng et al. (2021), for instance, framed AI literacy in terms of understanding AI, applying it, evaluating and creating with it, and dealing with its ethical issues.

Extending TPACK, Çelik (2023) introduced Intelligent-TPACK, which brings ethical considerations into the knowledge teachers need when using AI-based tools. UNESCO's (2024) framework of AI competencies for teachers similarly sets out five areas, namely a human-centred mindset, AI ethics, AI foundations and applications, AI pedagogy, and AI for professional learning. In this study, teacher AI competency was operationalised through four domains derived from these frameworks, namely AI knowledge, AI pedagogy, AI integration in teaching and AI ethics.

Relationships among acceptance, self-efficacy and competency

Earlier work indicates that these constructs are interconnected. In a study of Korean pre-service teachers, Joo et al. (2018) showed that TPACK predicted both teacher self-efficacy and perceived ease of use, which, together with perceived usefulness, predicted intention to use technology. Self-efficacy may therefore serve as the psychological link between how teachers view a technology and how capable they are of using it. Because these constructs are conceptually close, instruments that measure them together must demonstrate that each construct is represented by relevant items and can be distinguished from the others.

Content validation and pilot testing

Content validity reflects how well a set of items covers the construct it is meant to capture and is usually judged by experts. Expert agreement on relevance can be summarised with the content validity index (CVI) for individual items (I-CVI) and for the whole scale (S-CVI) (Lynn, 1986; Polit & Beck, 2006). Acceptable values depend on panel size. With three to five experts, each item should reach an I-CVI of 1.00, and an S-CVI/Ave of .90 or above signals excellent content validity (Lynn, 1986; Polit et al., 2007; Yusoff, 2019).

Once experts have reviewed the items, a pilot study lets researchers check item clarity, procedures and initial reliability ahead of the main study (Hertzog, 2008). For early scale development, Johanson and Brooks (2010) suggested piloting with about 30 people from the target population. Cronbach's alpha is the usual index of internal consistency, but excessively high values may signal overlapping items rather than better measurement (Streiner, 2003; Tavakol & Dennick, 2011). Reporting both expert-based and respondent-based evidence therefore provides a stronger foundation for instruments adapted to new contexts.

METHODOLOGY

This study was conducted in two phases, namely (1) validity evaluation by an expert panel and (2) a pilot study among CS and BCS teachers.

Instrument

The core instrument comprised seven demographic items and 48 measurement items. The items were translated from English into Malay by the first author, and the Malay version was reviewed for linguistic accuracy and contextual suitability by the expert panel. Back-translation was not conducted. The instrument was administered in Malay, while the item statements presented in this article are English translations prepared for reporting purposes. The pilot form also contained location questions (district and school) and five instrument usability questions. Responses were recorded on a six-point scale ranging from 1 (strongly disagree) to 6 (strongly agree). An even number of options was used so that respondents could not select a neutral midpoint, which may reduce social desirability bias (Garland, 1991). Six-point scales have also shown slightly higher discrimination and reliability than five-point scales (Chomeya, 2010). All items were positively worded. The facilitating conditions item on compatibility, which was negatively worded in the original UTAUT (Venkatesh et al., 2003), followed the positively worded version in UTAUT2 (Venkatesh et al., 2012). The original Computer Self-Efficacy scale asked respondents whether they could complete a task and how confident they were on a 10-point scale (Compeau & Higgins, 1995). In this study, the nine items were

rewritten as statements answered on the same six-point agreement scale as the other sections, to keep the response format consistent across the instrument and to reduce respondent burden. Social influence and behavioural intention were not included in AI acceptance, for the reasons outlined in the literature review. Table 1 presents the instrument structure.

Table 1: Structure of the research instrument

Section	Construct	Dimensions (items)	Source
A	Demographics	7 items	—
B	AI Acceptance	Performance Expectancy (B1–B4), Effort Expectancy (B5–B8), Facilitating Conditions (B9–B12)	Venkatesh et al. (2003, 2012)
C	AI Self-Efficacy	Single construct (C1–C9) covering three aspects, namely use, integration and problem-solving	Compeau & Higgins (1995)
D	Teacher AI Competency	AI Knowledge (D1–D5), AI Pedagogy (D6–D12), AI Integration in Teaching (D13–D23), AI Ethics (D24–D27)	Çelik (2023) and UNESCO (2024)

Expert review

Five experts were invited to evaluate the instrument, and four returned complete evaluation forms between July and August 2026. The panel comprised two academics (computer science; technical and vocational education and instructional design), a school administrator holding a doctorate in educational technology and AI, and a teacher. This number meets the recommended minimum of three experts (Lynn, 1986). The experts evaluated the Malay version of the instrument.

Face validity was assessed through 11 statements on format, language, terminology, scale and completion time using a four-point agreement scale. For content validity, experts rated the relevance of every item from 1 (not relevant) to 4 (highly relevant), and ratings of 3 or 4 were counted as relevant (Polit & Beck, 2006). Experts also wrote comments on items they felt needed revision.

Pilot study

The pilot study was conducted online via Google Forms in October 2026. Thirty CS and BCS teachers from three districts in Sabah participated, namely Keningau (13), Beaufort (9) and Papar (8). The three districts were selected purposively. Invitations were distributed to CS and BCS teachers in these districts through their schools, and the 30 teachers who voluntarily completed the questionnaire formed the pilot sample. This sample size falls within the range of 10 to 40 respondents recommended for pilot studies (Hertzog, 2008; Johanson & Brooks, 2010). Pilot respondents will be excluded from the main study. At the end of the questionnaire, respondents were asked about completion time, ease of viewing the response scale and any items that were difficult to understand.

Data analysis

- **Content validity:** I-CVI was computed for each item, S-CVI was computed by averaging (S-CVI/Ave) and by universal agreement (S-CVI/UA), and modified kappa (k*) was calculated for items with I-CVI below 1.00 (Polit et al., 2007).

- **Reliability:** Cronbach's alpha and corrected item-total correlation (CITC) for each construct and dimension. AI self-efficacy was analysed as a single nine-item construct, consistent with the original scale by Compeau and Higgins (1995). Ninety-five per cent confidence intervals for alpha were computed using Feldt's (1965) method, and McDonald's omega was estimated as an additional index of internal consistency from a single-factor maximum likelihood factor model for each construct and dimension (Hayes & Coutts, 2020; McDonald, 1999). Because the factor structure was not tested, neither coefficient establishes unidimensionality. Item-level means, standard deviations, skewness, kurtosis and floor and ceiling percentages were also examined, with more than 15% of responses at the lowest or

highest score indicating floor or ceiling effects (Terwee et al., 2007). The main analysis used all 30 respondents.

- **Response patterns:** Respondents who gave identical responses to all 48 items (straight-lining) were flagged (Meade & Craig, 2012). Because all items were positively worded, this pattern does not necessarily indicate carelessness, as respondents may genuinely agree strongly with all statements. These respondents were therefore not excluded. A reliability analysis without them (n = 25) was conducted as a sensitivity analysis only.

RESULTS

The instrument achieved excellent content validity (S-CVI/Ave = .984) and high reliability for all three constructs (α = .918–.986).

Face validity

All four experts agreed (rating 3 or 4) with all 11 face validity statements, with an overall mean of 3.61 out of 4. The lowest ratings were for grammar (M = 3.00) and terminology (M = 3.25). Experts recommended simplifying the sentence structure of the AI competency items, clarifying terms such as "adaptive feedback" and "personalised learning", and aligning the estimated completion time.

Content validity

Table 2: Content validity index by construct and dimension (four experts)

Construct / Dimension	Items	S-CVI/Ave	S-CVI/UA	Items with I-CVI < 1.00
AI Acceptance	12	0.979	0.917	B5
Performance Expectancy	4	1	1	—
Effort Expectancy	4	0.938	0.75	B5
Facilitating Conditions	4	1	1	—
AI Self-Efficacy	9	1	1	—
Teacher AI Competency	27	0.981	0.926	D11, D27
AI Knowledge	5	1	1	—
AI Pedagogy	7	0.964	0.857	D11
AI Integration in Teaching	11	1	1	—
AI Ethics	4	0.938	0.75	D27
Overall	48	**0.984**	**0.938**	3 items

Forty-five of the 48 items (93.8%) were rated relevant by all experts (I-CVI = 1.00). Three items (B5, D11 and D27) obtained I-CVI = .75 and k* = .67, which is considered "good" (Polit et al., 2007). However, for panels of three to five experts, I-CVI should reach 1.00 (Lynn, 1986). These three items were therefore revised based on expert comments rather than removed.

The overall S-CVI/Ave (.984) exceeded the .90 threshold, and the S-CVI/UA (.938) exceeded the .80 threshold (Polit & Beck, 2006). AI self-efficacy was the only construct with universal agreement on all items.

Item revisions

Fifteen items were revised based on expert comments. Table 3 presents the main revisions. The item wording shown is an English translation of the Malay items.

Table 3: Main item revisions based on expert comments

Item	Issue raised by experts	Revised statement (translated)
B4	"Overall performance" too broad	Using AI-based tools improves my performance in carrying out teaching tasks.
B5	"Interaction" ambiguous and unclear whether it measures ease of use	My interaction with AI-based tools, including prompts, responses and navigation, is clear and understandable.
D1	Referred to "daily life" instead of teaching tasks	I know how to interact with AI-based tools when carrying out teaching-related tasks.
D3	Phrase "initiate a task for the technology" unnatural	I know how to give instructions to AI-based tools through text or voice to initiate a task.
D9	Phrase "regardless of form or curriculum level" obscured the focus	I know how to use AI-based tools to monitor students' learning.
D10	"Messages from AI tools" too general	I can interpret outputs generated by AI-based tools to provide real-time feedback to students.
D11	"Notifications" vague and not specific to AI	I can understand outputs or automated feedback generated by AI-based tools to support students' learning.
D24	Examples "race and gender" too narrow	I can evaluate whether AI-based tools fairly consider students' diversity and individual differences.
D26	Should refer to the explainability of AI outputs	I can understand the explanation or justification for outputs or decisions produced by AI-based tools.
D27	Identifying developers does not measure accountability	I understand the importance of identifying the parties responsible for the design, development and decisions of AI-based tools.

The other revised items were D2, D16, D17, D18 and D19 (restructured sentences). The term "AI tools" was also standardised to "AI-based tools" throughout the instrument. Items D14 and D22 were retained even though experts raised concerns about dimensional overlap. Their placement will be re-examined through the measurement model in the main study.

Respondent profile and pilot feedback

Respondents comprised 16 female (53.3%) and 14 male (46.7%) teachers. Half were aged 40–49 years (50.0%), 17 (56.7%) had more than 15 years of teaching experience, and 27 (90.0%) held a bachelor's degree. All respondents had used AI tools in teaching and learning, and 18 (60.0%) used them several times a week or daily.

Eighteen respondents (60.0%) completed the questionnaire in less than 15 minutes, 10 (33.3%) in 15–20 minutes and two (6.7%) in more than 30 minutes. All respondents found the response scale easy to view and select on their devices. Three respondents (10.0%) indicated that some statements were difficult to understand. Based on this feedback and the experts' recommendations, brief definitions of key terms (e.g., "using AI" and "integrating AI") will be added to the instructions for the main study.

Reliability

Table 4: Cronbach's alpha, 95% confidence interval, McDonald's omega, mean and standard deviation by construct and dimension

Construct / Dimension	Items	α (n = 30)	95% CI (α)	ω	α (n = 25)[a]	M	SD
AI Acceptance	12	0.962	.939–.979	0.962	0.949	5.36	0.63
Performance Expectancy	4	0.94	.895–.969	0.942	0.923	5.48	0.7
Effort Expectancy	4	0.919	.859–.958	0.92	0.893	5.32	0.67
Facilitating Conditions	4	0.92	.859–.958	0.926	0.893	5.27	0.67

AI Self-Efficacy	9	0.918	.865–.955	0.929	0.879	5.25	0.59
Teacher AI Competency	27	0.986	.978–.992	0.987	0.981	5.1	0.67
AI Knowledge	5	0.95	.915–.974	0.953	0.938	5.17	0.72
AI Pedagogy	7	0.97	.950–.984	0.971	0.96	5.1	0.71
AI Integration in Teaching	11	0.967	.946–.982	0.968	0.953	5.12	0.65
AI Ethics	4	0.931	.879–.964	0.931	0.908	4.95	0.76

[a] Sensitivity analysis excluding five respondents with uniform responses (straight-lining). M and SD are reported for n = 30 on a 1–6 scale. 95% CI computed using Feldt's (1965) method. ω = McDonald's omega from a single-factor model.

In the main analysis (n = 30), construct alphas ranged from .918 to .986 and the dimension alphas of AI acceptance and AI competency ranged from .919 to .970, exceeding the .70 threshold (Tavakol & Dennick, 2011). Given the small sample, the 95% confidence intervals were relatively wide, but all lower bounds exceeded .85. Omega values (.920–.987) were nearly identical to alpha (Table 4), although this similarity does not confirm that the scales are unidimensional. Corrected item-total correlations ranged from .54 to .92, exceeding the minimum value of .30. Deleting C3 or C6 would marginally increase the AI self-efficacy alpha (from .918 to .920 and .922, respectively). Both items were retained because the increase was negligible (≤ .004), their CITC values (.58 and .54) exceeded .30, and both are needed to represent the "guidance from others" support level in the original scale design. No other item increased the construct alpha if deleted.

In the sensitivity analysis (n = 25), alphas for all three constructs (.879–.981) and all dimensions (.893–.960) remained satisfactory. This indicates that the reliability of the instrument did not depend on the five respondents with uniform responses.

Item means for AI self-efficacy showed a pattern consistent with the scale design. Items without guidance (C1, C4 and C7, M = 4.97–5.13) recorded the lowest means, whereas items with guides or references (C2, C5 and C8, M = 5.27–5.43) and with guidance from others (C3, C6 and C9, M = 5.30–5.40) recorded higher means. This pattern suggests that respondents differentiated the levels of support stated in the items.

Response patterns

Five respondents (16.7%) gave identical responses to all 48 items, of whom three selected 6, one selected 5 and one selected 4. All three respondents who selected 6 reported using AI several times a week or daily. Responses were skewed towards the upper end of the scale, with 39.5% of responses at 6, 41.9% at 5, 17.0% at 4, 1.5% at 3 and only 0.1% (two responses) at 2. No respondent selected 1. At the item level, no item had any response at the lowest score, whereas all 48 items exceeded the 15% ceiling criterion (23.3%–70.0%). Skewness ranged from −2.02 to 0.04 (Appendix A). Correlations between construct mean scores were high, namely AI acceptance–AI self-efficacy (r = .76), AI acceptance–AI competency (r = .77) and AI self-efficacy–AI competency (r = .84).

DISCUSSION

The findings provide preliminary evidence of content validity and internal consistency reliability of the instrument for use in the main study. Three issues warrant attention, namely very high alpha values, an item-level ceiling effect and uniform response patterns.

Content validity

An S-CVI/Ave of .984 indicates that the items represent the construct domains well according to the expert panel. This level of agreement is comparable to a recent content validation of an AI-related questionnaire for secondary school teachers in Sabah, which also used a four-expert panel (Kambutong & Shaafi, 2026). This is consistent with the robustness of the original UTAUT, Computer Self-Efficacy and Intelligent-TPACK scales

validated in previous studies (Çelik, 2023; Compeau & Higgins, 1995; Venkatesh et al., 2003, 2012). Most revisions arose from the adaptation process, namely original items that were too general (e.g., "daily life") or terms that became ambiguous in translation (e.g., "notifications" and "messages"). This underscores that translation alone is insufficient, and items must be adapted to the context of teachers' instructional tasks.

The combination of academic and practitioner experts provided complementary perspectives. Expert comments focused on two aspects, namely boundaries between dimensions (e.g., D14 was judged to overlap with AI knowledge, and D22 to reflect leadership rather than integration in teaching) and language clarity for teacher respondents.

Very high reliability

Alpha exceeded .95 for AI acceptance (.962) and AI competency (.986). Although commonly considered "excellent", such high values may indicate redundant items (Streiner, 2003; Tavakol & Dennick, 2011). Similarly high values have been reported in a recent pilot validation of an AI readiness instrument for Malaysian teachers, with alpha reaching .986 for one construct (Muniandy et al., 2026), which suggests that item redundancy deserves closer scrutiny in this area. For example, items D17, D18 and D19 differ only in a single phrase (adaptive feedback, personalised learning and real-time feedback), and their inter-item correlations were high (r = .88–.91). Nevertheless, all four experts rated each of these items as relevant (I-CVI = 1.00) and recommended only that the terms be clarified. This supports the relevance of each item, although it does not show that the three items contribute distinct information. The three items were therefore retained for now, definitions of the terms will be added for the main study, and their redundancy will be examined through inter-item correlations and outer loadings. Alpha also increases with the number of items, and AI competency comprises 27 items.

Ceiling effect and uniform response patterns

Only two responses (0.1%) were below 3, and 81.4% of responses were 5 or 6. Together with the item-level results (Appendix A), this indicates a ceiling effect in the pilot sample, although whether the pattern recurs in a larger and more diverse sample remains unknown. The high scores may relate to the characteristics of the CS and BCS teacher sample, as all respondents had used AI and 60% used it several times a week or daily. However, this pilot study did not collect information on respondents' field of specialisation, so this explanation should be interpreted with caution. A narrow score range may also slightly attenuate the relationships between constructs.

The five respondents with uniform responses were not automatically considered careless. Because all items were positively worded, genuinely proficient respondents might select the same score for all items. The sensitivity analysis showed that the main construct alphas remained high without these respondents, indicating that the reliability of the instrument did not depend on them.

The high inter-construct correlations (r = .76–.84), particularly between AI self-efficacy and AI competency, warrant attention. Both constructs capture teachers' confidence and ability to use AI. Therefore, the heterotrait–monotrait ratio (HTMT) should be examined in the main study's PLS-SEM analysis to confirm that the two constructs are empirically distinct (Hair et al., 2022).

Implications for the main study

1. **Retain all 48 items.** No item warranted removal based on CVI or CITC. Overlapping items (D17–D19) and dimension boundaries (D14, D22) will be evaluated in the PLS-SEM measurement model. Redundancy among D17–D19 will be examined through inter-item correlations and outer loadings, and the items will be merged or removed if redundancy is confirmed, while the heterotrait–monotrait ratio will be used to assess discriminant validity between constructs.

2. **Add brief term definitions.** Definitions of "using AI-based tools" and "integrating AI-based tools" will be added to the Section C instructions to help respondents distinguish between using AI to perform a task

and the planned application of AI as part of teaching activities. As recommended by the experts, a brief description of the support levels (without guidance, with guides or references, with guidance from others) will also be included, together with definitions before items D17–D19 (adaptive feedback, personalised learning, real-time feedback). Item wordings will remain unchanged.

3. **Add a demographic item on field of specialisation** (e.g., computer science/IT, multimedia, others) to help explain the high score distribution.

4. **Establish a response pattern screening procedure.** Uniform response patterns will be flagged for further examination together with other indicators such as completion time, the longstring index, intra-individual response variability and overall response patterns. Respondents will not be excluded based on a single indicator (Ward & Meade, 2023). Negatively worded and attention-check items were not added, as reversed items can reduce reliability and introduce method effects (Suárez-Alvarez et al., 2018), and additional items would lengthen an already long questionnaire. These screening procedures help detect careless responding but do not address acquiescence bias, which is acknowledged as a limitation.

5. **Update the estimated completion time** to 15–20 minutes, based on the experience of 93% of respondents.

Limitations

The pilot sample was small (n = 30) and comprised volunteer respondents from three purposively selected districts. All respondents were CS and BCS teachers who had used AI, so the findings may not represent teachers with less exposure to AI. This sample size also does not permit stable factor analysis or Rasch analysis. Back-translation was not conducted, the experts rated item relevance only, the panel included a single computer science academic, and cognitive interviews with teachers were not carried out. Because all items were positively worded, acquiescence bias cannot be ruled out. These steps should be addressed in future validation work.

CONCLUSION

The instrument measuring AI acceptance, AI self-efficacy and teacher AI competency showed high content validity (S-CVI/Ave = .984) and high internal consistency (construct α = .918–.986; ω = .929–.987) among CS and BCS teachers in Sabah. Its main strength lies in the content validation and language adaptation process, in which 15 items were revised based on expert comments to improve clarity and suitability for the teaching context. Given the small pilot sample, these results are preliminary, but they support using the instrument in the main study with the addition of term definitions and a response pattern screening procedure. Future research should examine construct and discriminant validity through PLS-SEM with a larger sample.

ETHICAL APPROVAL

This study received approval from the Educational Planning and Research Division (EPRD), Ministry of Education Malaysia, through the e-Research Application System (eRAS) (Ref. No.: KPM.600-3/2/3-eras(30624), dated 15 September 2026), and from the Sabah State Education Department (Ref. No.: JPNSBH.SPS.600-1/1/1 Jld.54 (36), dated 5 October 2026). Permission to publish this article was granted by EPRD (Ref. No.: KPM.600-2/1/3-terbitan(10), dated 6 November 2026). Participation was voluntary. The introductory page of the online questionnaire explained the purpose of the study, the confidentiality of responses and the academic use of the data. Completion and submission of the questionnaire were treated as implied consent to participate. Responses were kept confidential and used for academic purposes only.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request, subject to the conditions of approval by the Ministry of Education Malaysia.

REFERENCES

1. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
2. Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
3. Çelik, İ. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, 107468. https://doi.org/10.1016/j.chb.2022.107468
4. Chomeya, R. (2010). Quality of psychology test between Likert scale 5 and 6 points. *Journal of Social Sciences, 6*(3), 399–403. https://doi.org/10.3844/jssp.2010.399.403
5. Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly, 19*(2), 189–211.
6. Curriculum Development Division. (2020). *Computational thinking in primary and secondary school curriculum* [Conference presentation]. APEC 2020 Conference, University of Tsukuba, Japan. Ministry of Education Malaysia. https://www.criced.tsukuba.ac.jp/math/apec/apec2020/presentations/12Feb/13-Mash_M-2020.pdf
7. Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30*(3), 357–370. https://doi.org/10.1007/BF02289499
8. Fundi, M., Sanusi, I. T., Oyelere, S. S., & Ayere, M. (2024). Advancing AI education: Assessing Kenyan in-service teachers' preparedness for integrating artificial intelligence in competence-based curriculum. *Computers in Human Behavior Reports, 14*, 100412. https://doi.org/10.1016/j.chbr.2024.100412
9. Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*, 66–70.
10. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.
11. Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629
12. Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health, 31*(2), 180–191.
13. Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394–400.
14. Joo, Y. J., Park, S., & Lim, E. (2018). Factors influencing preservice teachers' intention to use technology: TPACK, teacher self-efficacy, and technology acceptance model. *Educational Technology & Society, 21*(3), 48–59.
15. Kambutong, J., & Shaafi, N. F. (2026). Content validity of a survey on knowledge, skills, and readiness for generative AI. *Journal of Creative Practices in Language Learning and Teaching, 14*(1), 38–57. https://doi.org/10.24191/xnvcmw23
16. Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385.
17. McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
18. Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.
19. Ministry of Education Malaysia. (2023). *Digital education policy*. Ministry of Education Malaysia.
20. Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054.
21. Muniandy, P., Maraya, R., & Kandasamy, S. S. (2026). Development and pilot validation of an AI integration readiness instrument for Tamil language teachers. *International Journal of Academic Research in Business and Social Sciences, 16*(1), 173–187. https://doi.org/10.6007/IJARBSS/v16-i1/27385

22. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041
23. Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497.
24. Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467.
25. Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education, 128*, 13–35. https://doi.org/10.1016/j.compedu.2018.09.009
26. Scherer, R., & Teo, T. (2019). Unpacking teachers' intentions to integrate technology: A meta-analysis. *Educational Research Review, 27*, 90–109. https://doi.org/10.1016/j.edurev.2019.03.001
27. Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103.
28. Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33
29. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55.
30. Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
31. UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.
32. UNESCO. (2024). *AI competency framework for teachers*. UNESCO.
33. 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.
34. Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412
35. Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology, 74*, 577–596. https://doi.org/10.1146/annurev-psych-040422-045007
36. Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11*(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

APPENDIX A

Table A1: Item-level descriptive statistics (n = 30)

Item	M	SD	Skew	Kurt	Ceil. (%)	Item	M	SD	Skew	Kurt	Ceil. (%)
B1	5.53	0.73	−1.26	0.17	66.7	D4	5	0.91	−0.59	−0.36	33.3
B2	5.5	0.86	−1.56	1.36	70	D5	4.97	0.76	−0.44	0.15	23.3
B3	5.43	0.77	−0.96	−0.59	60	D6	5.1	0.8	−0.62	0.06	33.3
B4	5.47	0.68	−0.92	−0.24	56.7	D7	5.17	0.7	−0.89	2.01	30
B5	5.27	0.69	−0.41	−0.77	40	D8	5.13	0.78	−0.72	0.52	33.3
B6	5.3	0.7	−0.50	−0.78	43.3	D9	5.03	0.81	−0.48	−0.18	30
B7	5.33	0.76	−0.66	−0.91	50	D10	5.07	0.78	−0.12	−1.33	33.3
B8	5.4	0.81	−1.30	1.22	56.7	D11	5.13	0.73	−0.21	−1.02	33.3
B9	5.33	0.8	−1.13	0.99	50	D12	5.07	0.83	−0.52	−0.30	33.3
B10	5.3	0.7	−0.50	−0.78	43.3	D13	5.27	0.69	−0.41	−0.77	40

B11	5.23	0.73	−0.40	−0.96	40	D14	5.33	0.66	−0.48	−0.62	43.3
B12	5.2	0.76	−0.36	−1.14	40	D15	5.27	0.64	−0.29	−0.55	36.7
C1	5.1	0.84	−0.93	0.81	33.3	D16	5.13	0.73	−0.21	−1.02	33.3
C2	5.43	0.63	−0.64	−0.45	50.0	D17	5.03	0.76	−0.06	−1.23	30
C3	5.30	0.92	−1.82	4.57	50.0	D18	5.07	0.69	−0.09	−0.77	26.7
C4	4.97	0.89	−0.56	−0.27	30	D19	5.1	0.71	−0.15	−0.91	30
C5	5.33	0.61	−0.29	−0.55	40	D20	5.17	0.75	−0.29	−1.09	36.7
C6	5.33	0.88	−2.02	5.89	50.0	D21	5.13	0.86	−0.62	−0.44	40
C7	5.13	0.73	−0.21	−1.02	33.3	D22	4.83	0.83	−0.05	−0.78	23.3
C8	5.27	0.64	−0.29	−0.55	36.7	D23	4.97	0.85	−0.29	−0.72	30
C9	5.4	0.62	−0.52	−0.53	46.7	D24	4.93	0.87	−0.20	−0.92	30
D1	5.37	0.72	−0.69	−0.70	50	D25	4.8	0.85	0.04	−0.90	23.3
D2	5.33	0.71	−0.59	−0.76	46.7	D26	5.07	0.78	−0.12	−1.33	33.3
D3	5.17	0.83	−0.71	−0.08	40	D27	5	0.83	0	−1.55	33.3

Note. No item had any response at the lowest score (floor = 0% for all items). Ceil. = percentage of responses at the highest score (6). Kurt = excess kurtosis.