Journal of Creative Practices in Language Learning and Teaching (CPLT)
Volume 14, Number 1, 2026

TABLE OF CONTENT

Journal of Creative Practices in Language Learning and Teaching (CPLT)
Volume 14, Number 1, 2026

Content Validity of a Survey on Knowledge, Skills, and Readiness for Generative AI

Julazzmie Kambutong
julazzmie_mp2411071t@iluv.ums.edu.my
Faculty of Education and Sports Studies
Universiti Malaysia Sabah, Malaysia

Nur Farha Shaafi*
farhashaafi@ums.edu.my
Faculty of Education and Sports Studies
Universiti Malaysia Sabah, Malaysia

Corresponding author*

Received: 25 November 2025
Accepted: 21 February 2026
Published: 25 May 2026

CITE THIS ARTICLE:

Kambutong, J., & Shaafi, N. F. (2026). Content validity of a survey 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

ABSTRACT

This study aims to evaluate the content validity of the questionnaire instrument for the level of knowledge, skills, and readiness of secondary school Science teachers in Sabah in using generative AI for teaching and learning, via the Content Validity Index (CVI) assessed based on the views of four experts. The assessment of content validity was conducted through the application of the Item-Level Content Validity Index (I-CVI) and the Scale-Level Content Validity Index (S-CVI). The study findings show that the overall S-CVI (average) score for the dimensions of level of knowledge, skills, and readiness to use generative AI among secondary schools Science teachers in the questionnaire instrument was 1.00. The S-CVI (UA) for each dimension was also at a score of 1.00, meeting the required criteria for the overall S-CVI value of the questionnaire instrument for the level of knowledge, skills, and readiness of secondary school Science teachers. In conclusion, the 34 items in the questionnaire were accepted and retained to assess the level of knowledge, skills, and readiness of secondary school Science teachers in the pilot study for the purposes of construct validity and instrument reliability.

Keywords: content validity, generative AI, knowledge, skills, readiness

INTRODUCTION

The increasing availability of generative artificial intelligence (AI) tools has prompted growing interest in their potential applications within educational contexts. In particular, tools such as ChatGPT, Gemini, and Copilot have been highlighted for their capacity to support lesson planning, instructional material development, formative assessment, and student engagement. In Science education, where abstract concepts and complex representations often challenge learners, generative AI has been positioned as a supportive technology that may enhance conceptual understanding and instructional efficiency (Aliya et al., 2024; El Fathi et al., 2025). However, meaningful integration of generative AI into classroom practice depends not only on technological availability, but also on teachers' knowledge, skills, and readiness to use these tools appropriately and confidently.

Despite increasing attention to generative AI in education, empirical studies consistently report that teachers' adoption of AI remains limited. Existing research suggests that many teachers possess only surface-level awareness of AI, with limited understanding of how generative AI can be applied pedagogically (Casal-Otero et al., 2023; Samsudin et al., 2023). In addition, teachers often report insufficient skills related to prompt construction, critical evaluation of AI-generated outputs, and the alignment of AI use with instructional objectives. These limitations are compounded by uncertainty regarding ethical considerations, data security, and institutional support, all of which influence teachers' readiness to adopt generative AI in sustained and meaningful ways (Al Darayseh, 2023; Cooper, 2023).

Within the Malaysian context, and particularly in Sabah, these challenges are especially salient. Although Science teachers generally demonstrate basic technological literacy, specialised competencies related to generative AI remain underdeveloped (Lowa, 2024; Juwait & Siew, 2022). National initiatives such as the Digital Education Policy 2023 and the Digital Competency Standards for Educators emphasise the importance of advanced digital capabilities, yet evidence suggests that most teachers continue to operate at foundational levels of digital practice. This raises concerns about teachers' preparedness to engage with generative AI, which requires higher-order technological, pedagogical, and evaluative competencies (Ishmuradova et al., 2025).

Importantly, challenges related to generative AI readiness are not confined to Science instruction alone but have clear implications for language learning and creative classroom practices. Generative AI tools are increasingly used to support language pedagogy through automated feedback on writing, scaffolding of speaking practice, generation of reading materials, and assistance in lesson design. Teachers who lack adequate knowledge, skills, or readiness may be unable to leverage these affordances effectively, thereby limiting opportunities for interdisciplinary and language-rich instructional approaches in Science classrooms (Abdullah et al., 2020; Ishmuradova et al., 2025). Consequently, understanding and measuring teachers' readiness to use generative AI has relevance beyond content delivery, extending to communication, scientific literacy, and creative pedagogical practices (Al Darayseh & Mersin, 2025).

Although prior studies have explored teachers' perceptions, acceptance, and readiness to use AI, there remains a methodological gap concerning the availability of validated instruments that

accurately capture these constructs in the context of generative AI. Many existing questionnaires lack clear construct boundaries or systematic validation procedures, raising concerns about the validity and interpretability of findings. In particular, the constructs of knowledge, skills, and readiness are often treated as overlapping or self-evident, despite representing conceptually distinct dimensions of teachers' engagement with generative AI.

In this study, knowledge refers to teachers' conceptual understanding of generative AI, including its functions, limitations, ethical considerations, and pedagogical potential. Skills denote teachers' practical and technical abilities to use generative AI tools, such as designing effective prompts, evaluating AI-generated content, and integrating AI outputs into instructional activities. Readiness reflects teachers' psychological and behavioural disposition to adopt generative AI, encompassing confidence, perceived usefulness, and perceived ease of use. These constructs are theoretically grounded through the integration of the Technological Pedagogical Content Knowledge (TPACK) framework and the Technology Acceptance Model (TAM). TPACK explains teachers' knowledge and skills as the interaction of technological, pedagogical, and content domains, while TAM accounts for readiness through perceived usefulness and perceived ease of use.

Accordingly, the purpose of this study is not to measure teachers' levels of AI adoption or to evaluate instructional outcomes, but to establish the content validity of a questionnaire designed to measure secondary school Science teachers' knowledge, skills, and readiness to use generative AI. Content validation is a critical preliminary step in instrument development, ensuring that questionnaire items are relevant, clear, and representative of the intended constructs. This study employs the Content Validity Index (CVI) as a systematic and rigorous method to evaluate the adequacy of the instrument items. By providing a validated measurement tool, this study contributes methodologically to future empirical research on generative AI integration in education and supports the development of evidence-based professional development initiatives for teachers.

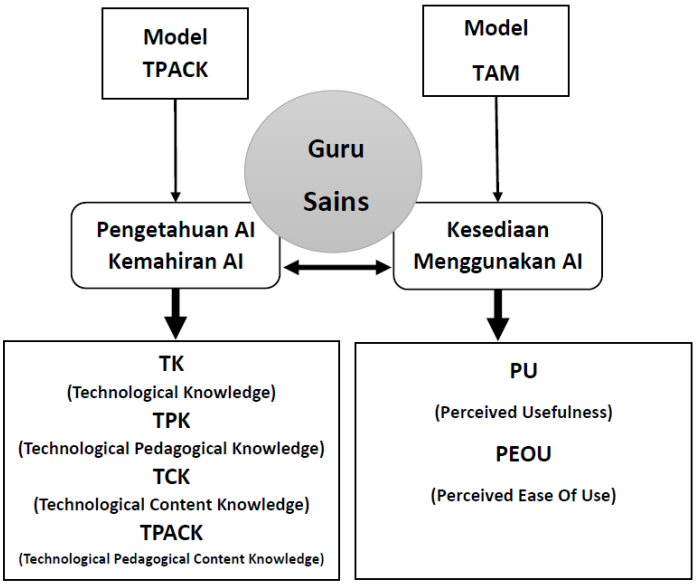


Figure 1. Conceptual Framework of the Study

METHODOLOGY

Content validity ensures that the operationalisation of a construct is based on items that accurately represent a specific content domain relevant to the measurement context. To achieve this, it is generally recommended that at least three experts participate in the content validation process (Shrotryia & Dhanda, 2019). There are two forms of the Content Validity Index, namely the Item level CVI (I CVI) and the Scale level CVI (S CVI) (Yusoff, 2019). In determining the CVI, experts are invited to review each item in the needs analysis questionnaire and provide their professional judgement. The CVI is then calculated both for individual items and for the overall scale. During this evaluation, experts assess the degree to which each item reflects the underlying construct. To prevent neutral responses, a four point rating scale is used, with categories defined as follows: [1] not relevant, [2] somewhat relevant, [3] quite relevant, and [4] highly relevant (Lynn, 1986; Polit & Beck, 2006). An I CVI should not be lower than 0.78, while the minimum acceptable S CVI is 0.80 to demonstrate adequate content validity (Lynn, 1986; Shrotryia & Dhanda, 2019). Table 1 summarises the number of experts and the corresponding acceptable CVI threshold. For two experts, the acceptable CVI is at least 0.80. For three to five experts, it must be 1.00. For six to eight experts, it should be at least 0.83. For nine or more experts, a minimum value of 0.78 is required. As four experts were involved in evaluating the questionnaire in this study, the acceptable CVI threshold is must be 1.00.

Table 1. Number of Experts and the Corresponding Acceptable Content Validity Index (CVI) Threshold

Number of experts	Acceptable CVI values	Source
Two experts	At least 0.80	Davis (1992)
Three to five experts	Exactly 1.00	Polit & Beck (2006), Polit et al. (2007)
At least six experts	At least 0.83	Polit & Beck (2006), Polit et al. (2007)
Six to eight experts	At least 0.83	Lynn (1986)
At least nine experts	At least 0.78	Lynn (1986)

This study employed a quantitative approach using a survey questionnaire as the primary data collection instrument. To ensure the content validity of the instrument, the researcher followed the six-step procedure proposed by Yusoff (2019) in determining the level of content validity, as outlined below:

i. Step 1: Prepare the content validity form
ii. Step 2: Select the panel of experts
iii. Step 3: Initiate the content validity process
iv. Step 4: Review the constructs and items
v. Step 5: Rate each item
vi. Step 6: Calculate the CVI

Step 1: Prepare the Content Validity Form

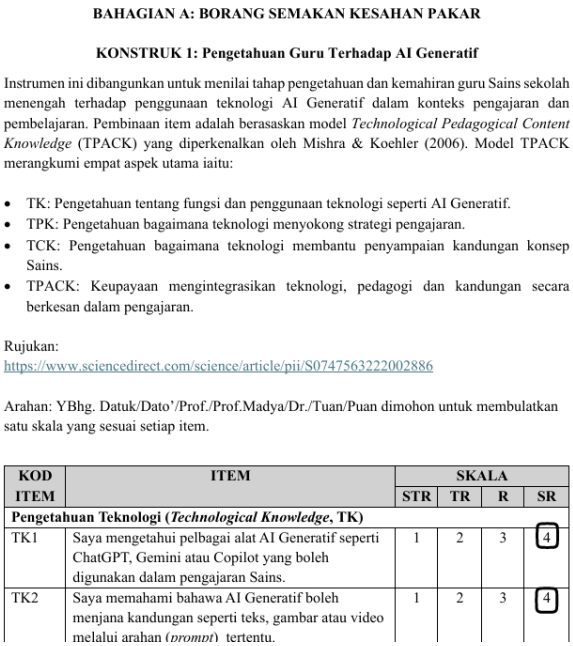
The survey instrument in this study was developed to collect feedback on the level of knowledge, skills, and readiness among secondary school Science teachers in Sabah. A survey approach was selected due to its practical benefits, particularly its ease of administration and the efficiency it provides in processing and analysing data (Hussin et al., 2014). The questionnaire items were adapted, rather than directly adopted, from two established instruments to ensure contextual relevance and construct alignment. Items measuring knowledge and skills were adapted from Celik (2023), whose instrument focused on teachers' professional knowledge and ethical integration of AI-based tools in educational settings. Items measuring readiness were adapted from Lewis and Sauro (2019), which operationalised technology acceptance constructs grounded in the Technology Acceptance Model (TAM). The adaptation process involved three key modifications. First, item wording was contextualised to reflect Science teaching practices, such as lesson planning, explanation of scientific concepts, and use of visual representations. Second, references to general AI tools were refined to explicitly emphasise generative AI functionalities, including content generation, feedback provision, and adaptive instructional support. Third, contextual localisation was applied to ensure relevance to secondary school Science teachers in Sabah, including alignment with local instructional practices and curriculum expectations. Although the original instruments had undergone prior validation, adaptation to a new educational context and construct focus necessitates renewed content validation. Content validation at this stage is appropriate because the purpose of the study is to establish item relevance and representativeness rather than construct equivalence or factor structure, which will be addressed in subsequent psychometric analyses.

The instrument uses a Likert scale to measure respondents' perceptions, views, and attitudes toward specific issues or phenomena (Padzil et al., 2021). It consists of two types of items: positively worded items scored as 5, 4, 3, 2, 1, and negatively worded items scored in reverse order, namely 1, 2, 3, 4, 5 (Tahar & Zaid, 2019). For the knowledge and skills components, the researchers adapted items from Celik (2023), who investigated teachers' professional knowledge regarding the ethical integration of artificial intelligence based tools in education. The readiness component was adapted from Lewis (2019), who developed a questionnaire based on the Technology Acceptance Model. The questionnaire comprises four sections. Section A gathers demographic information, while Sections B, C, and D consist of five point Likert scale items labelled as (1) strongly disagree, (2) disagree, (3) unsure, (4) agree, and (5) strongly agree. These sections focus on assessing the levels of knowledge, skills, and readiness among secondary school Science teachers in Sabah. During the development of the content validation form, the researchers ensured that the expert panel clearly understood the instrument being evaluated. Detailed explanations of the rating scale were provided in the validation form to guide their assessment. Clear definitions of each construct were also included to support the experts in reviewing the items, as illustrated in Figure 2. Table 2 presents the score definitions used to evaluate the Item Level Content Validity Index within the expert assessment form.

Table 2. Instrument Evaluation Score of I-CVI

Definition	Score
Not relevant	1
Less relevant	2

Relevant	3
Very relevant	4



BAHAGIAN A: BORANG SEMAKAN KESAHAN PAKAR

KONSTRUK 1: Pengetahuan Guru Terhadap AI Generatif

Instrumen ini dibangunkan untuk menilai tahap pengetahuan dan kemahiran guru Sains sekolah menengah terhadap penggunaan teknologi AI Generatif dalam konteks pengajaran dan pembelajaran. Pembinaan item adalah berasaskan model *Technological Pedagogical Content Knowledge* (TPACK) yang diperkenalkan oleh Mishra & Koehler (2006). Model TPACK merangkumi empat aspek utama iaitu:

- TK: Pengetahuan tentang fungsi dan penggunaan teknologi seperti AI Generatif.
- TPK: Pengetahuan bagaimana teknologi menyokong strategi pengajaran.
- TCK: Pengetahuan bagaimana teknologi membantu penyampaian kandungan konsep Sains.
- TPACK: Keupayaan mengintegrasikan teknologi, pedagogi dan kandungan secara berkesan dalam pengajaran.

Rujukan:
https://www.sciencedirect.com/science/article/pii/S0747563222002886

Arahan: YBhg. Datuk/Dato'/Prof./Prof.Madya/Dr./Tuan/Puan dimohon untuk membulatkan satu skala yang sesuai setiap item.

KOD ITEM	ITEM	SKALA			
		STR	TR	R	SR
Pengetahuan Teknologi (*Technological Knowledge*, TK)					
TK1	Saya mengetahui pelbagai alat AI Generatif seperti ChatGPT, Gemini atau Copilot yang boleh digunakan dalam pengajaran Sains.	1	2	3	4
TK2	Saya memahami bahawa AI Generatif boleh menjana kandungan seperti teks, gambar atau video melalui arahan (*prompt*) tertentu.	1	2	3	4

Figure 2. Content Validation Form

Step 2: Select the Panel of Experts

The content validity questionnaire was distributed to experts working in different locations. Selecting experts for the content validation process is a crucial step, requiring clear and rigorous selection criteria that consider professional experience, academic qualifications, and prior involvement in evaluation activities. These criteria help ensure the credibility and suitability of the validation process (Yusoff, 2019; Almanasreh et al., 2018). This study required experts whose areas of specialisation were directly related to the research focus (Taherdoost, 2016). An expert in this context is defined as an individual with specialised experience or skills in the relevant field (Rubio et al., 2003; Yusoff, 2019). This approach ensures that each panel member possesses the knowledge and practical experience needed for reliable and meaningful validation (Polit & Beck, 2006). Purposive sampling was used to ensure that only individuals with the appropriate level of expertise were selected. Four experts from the fields of Educational Technology and Science Education from various public universities in Malaysia participated in the content validation process. A panel consisting of four or more experts is generally considered adequate, as recommended by Almanasreh et al. (2018), Lynn (1986), and Polit et al. (2007). Larger expert panels can further enhance reliability, particularly in high risk or emerging areas of research. One of the key criteria used in this study was that each expert should have more than ten years of experience in their respective fields. Table 3 presents the panel of experts involved in validating the questionnaire on the level of knowledge, skills, and readiness of secondary school Science teachers in Sabah.

Table 3. Experts for the Content Validation of Instrument

No.	Expert	Profile	Years of Experience	Expertise
P1	Expert 1	Senior Lecturer	15	Educational Technology
P2	Expert 2	Senior Lecturer	16	Educational Technology
P3	Expert 3	Senior Lecturer	21	Science Education
P4	Expert 4	Senior Lecturer	20	Science Education

The researcher selected experts in the field of language education to evaluate the face validity of the questionnaire. To ensure a high level of face validity, specific selection criteria were established namely, holding a Doctor of Philosophy (PhD) in Malay Language and having more than five years of relevant professional experience. These criteria were set to ensure that the experts possessed the necessary expertise and credibility to assess the relevance and appropriateness of the instrument. The experts were asked to review the questionnaire and provide feedback on whether the items demonstrated correct, appropriate, and contextually suitable Malay grammar (Connell et al., 2018).

Step 3: Initiating the Content Validity Process

In this study, the content validity process was conducted remotely via email. Although the process was carried out online, clear instructions and guidelines were provided to the expert panel within the distributed questionnaire materials. The online approach was selected for its cost-effectiveness; however, it typically requires a longer turnaround time to obtain expert feedback compared to face-to-face methods (Yusoff, 2019).

Step 4: Reviewing the Constructs and Items

In this study, **knowledge** refers to teachers' conceptual understanding of generative AI, including its functions, pedagogical affordances, ethical considerations, and instructional applications in Science education. **Skills** denote teachers' practical ability to use generative AI tools effectively, such as constructing prompts, evaluating AI-generated outputs, and integrating AI-supported activities into Science lessons. **Readiness** reflects teachers' perceived usefulness, perceived ease of use, and confidence in adopting generative AI for instructional purposes. The questionnaire consists of five sections: Sections A, B, C, D, and E. Section A contains five items related to respondents' demographic information. Sections B through E form the core components of the questionnaire used for content validity review, comprising a total of 34 items across three main dimensions: teachers' knowledge of generative AI, teachers' skills in using generative AI, and teachers' readiness for generative AI, as outlined in Table 4. At this stage, each expert was asked to review all items comprehensively before assigning their respective rating scores.

Table 4. Dimension and Items

Dimension	Item Code	Item Description
Knowledge of teacher towards the AI Generative	B1	I am aware of various Generative AI tools such as ChatGPT, Gemini, or Copilot that can be used in Science teaching.
	B2	I understand that Generative AI can generate content such as text, images, or videos based on specific prompts.

	B3	I understand that ethical issues, copyright concerns, and information accuracy when using Generative AI in education.
	B4	I understand that Generative AI can help me adapt Science teaching to meet students' learning needs.
	B5	I am aware that Generative AI can generate automated suggestions or feedback to support student learning.
	B6	I am aware that Generative AI can be used to design engaging and meaningful learning activities.
	B7	I am aware that Generative AI can automatically generate Science materials such as quizzes, notes, or exercises.
	B8	I am aware that Generative AI can help present scientific data in visual forms such as graphs or charts.
	B9	I am aware that Generative AI can produce explanations of scientific concepts that are easy for students to understand.
	B10	I am aware that Generative AI can be used to generate Daily Lesson Plans (RPH) based on topics and learning objectives.
	B11	I am aware that Generative AI can help tailor Science content and teaching approaches according to students' ability levels.
	B12	I am aware that Generative AI can help me plan problem-solving activities for teaching Science topics.
Skills of teacher towards the AI Generative	C1	I am skilled at using various Generative AI tools to obtain information and learning materials.
	C2	I am skilled at providing clear prompts to Generative AI to produce high-quality outputs.
	C3	I am able to evaluate whether the content generated by Generative AI is suitable for Science teaching or not.
	C4	I am skilled at adapting teaching activities using Generative AI according to students' levels.
	C5	I am skilled at guiding students to use Generative AI to support their self-directed learning in Science.
	C6	I am skilled at using Generative AI to help students understand scientific concepts more easily.
	C7	I am skilled at using Generative AI to explain scientific concepts visually, such as through images or videos.
	C8	I am skilled at using Generative AI to generate quizzes, questions, or exercises based on Science topics appropriate to students' levels.
	C9	I am skilled at using Generative AI to produce explanations that are easy for students to understand.
	C10	I am skilled at using Generative AI to support problem-based learning or project-based learning in Science.
	C11	I am skilled at guiding other teachers to use Generative AI effectively in Science teaching.
	C12	I am skilled at planning Science lessons by integrating content, pedagogy, and the use of Generative AI technology.

Readiness of teacher towards the acceptance of AI Generative	D1	Generative AI helps me prepare Science teaching materials and assignments more quickly.
	D2	Generative AI helps me improve the quality of my Science teaching.
	D3	Generative AI makes me more efficient and productive in teaching Science.
	D4	Generative AI makes it easier for me to deliver Science content clearly and effectively.
	D5	Generative AI simplifies my tasks in implementing the Science teaching and learning process.
	D6	I believe it is easy to learn how to use Generative AI for Science teaching.
	D7	I am ready to use Generative AI to produce the necessary Science teaching materials.
	D8	I believe that interacting with Generative AI makes daily Science teaching tasks easier.
	D9	I am confident that I can use Generative AI even without assistance from others.
	D10	I believe Generative AI is easy to use in Science teaching.

Step 5: Rate Each Item

After reviewing all the items, the expert panel was asked to evaluate and assign a score to each item using the four-point scale provided, along with the accompanying descriptions in the content validity form (Figure 1). Each item also included a comment section to allow the experts to provide remarks or suggestions for improvement. At the end of the questionnaire, an additional space was provided for the experts to summarise whether the instrument and its items were appropriate or inappropriate for use. Once the evaluation was completed, the experts were requested to return the completed feedback form to the researcher via email.

Step 6: Calculate the CVI

In this step, the CVI was calculated using Microsoft Excel. Two types of CVI were computed: the Item-level CVI (I-CVI) and the Scale-level CVI (S-CVI). The calculations were based on formulas adapted from Lynn (1986), Davis (1992), Polit and Beck (2006), and Polit et al. (2007) as outlined in the ABC of Content Validation and Content Validity Index Calculation. The CVI values were computed using I-CVI, S-CVI (average), and S-CVI (universal agreement), as shown in Table 5. Meanwhile, Table 6 presents the original scale and the converted scale used in calculating the I-CVI.

Table 5. Definition for Calculation of CVI Index Formula

CVI Index	Definition	Formula
I-CVI *(item-level content validity index)*	The proportion of content experts who rated an item as relevant, using the ratings relevant (3) or highly relevant (4).	$I-CVI = \frac{Agreed\ item}{No.of\ experts}$

S-CVI *(scale-level content validity index based on average method)*	The average I-CVI score for all items in the scale, or the mean proportion of relevance as rated by all experts.	$S-CVI/Ave = \frac{Total\ score\ I-CVI}{No.of\ item}$
S-CVI/UA *(scale-level content validity index based on universal agreement method)*	The percentage of items in the scale that achieved a relevance rating of 3 or 4 from all experts. The Universal Agreement (UA) score is assigned a value of 1 when an item receives 100% agreement from the experts; otherwise, the UA score is assigned a value of 0.	$S-CVI/UA = \frac{Total\ score\ UA}{No.of\ item}$

Source: Lynn (1986); Polit & Beck (2006); Polit et al. (2007); Yusoff (2019), Lukmanul Hakim et al. (2025).

Table 6. Original Scale and Converted Scale Used for the I-CVI

Original Scale	Converted Scale
1, 2	0
3, 4	1

Shi et al. (2012) and Ping and Shaafi (2026) pointed that the consistent agreement among two or more experts on the items relevance might occasionally result from random selection rather than true consensus. Relying solely on the I-CVI can therefore overestimate validity because it does not account for random agreement. To address this, Polit et al. (2007) proposed calculating the probability of chance agreement (Pc) using the binomial distribution, assuming each expert randomly selects "relevant" or "not relevant" with equal probability (0.5). Pc represents the likelihood that the observed agreement occurred by chance, with formula:

$$P_c = \left[\frac{n!}{A!\,(n-A)!}\right] \times 0.5^n$$

where *n* is the number of experts and *A* is the number who rated the item as relevant (score 3 or 4). The adjusted kappa statistic (κ) is then derived by combining I-CVI and Pc to correct for chance agreement using:

$$\kappa = \frac{(I-CVI) - P_c}{1 - Pc}$$

This adjustment provides a more accurate measure of expert consensus beyond chance. Interpretation thresholds by Shi et al. (2012) and Rodrigues et al. (2017) for κ are: <0.40 (poor), 0.40–0.59 (fair), 0.60–0.74 (good), and ≥0.74 (excellent) are being used in this study (refer Table 7). Items with higher κ values indicate stronger content validity and are more likely to represent the intended construct. Including Pc and κ in the analysis ensures that retained items in the final instrument are based on genuine expert agreement, rather than random agreement.

Table 7. Interpretation Guidelines for Adjusted Kappa (κ) Values (Rodrigues et al., 2017; Shi et al., 2012)

Adjusted Kappa (κ) Values Range	Interpretation
$\kappa \geq 0.74$	Excellent
$0.60 \leq \kappa < 0.74$	Good
$0.40 \leq \kappa < 0.59$	Fair
$\kappa < 0.40$	Poor

RESULTS AND DISCUSSION

In this study, content validity was examined using the four-point relevance rating scale introduced by Davis (1992), which is widely applied in prior research and is known for its simplicity and ease of interpretation. The scale consists of four levels: 1 for not relevant, 2 for somewhat relevant, 3 for relevant, and 4 for highly relevant. Each expert appointed for this study used this scale to evaluate the suitability of the items in the developed instrument. The Item-level Content Validity Index (I-CVI) was calculated as the proportion of experts who rated an item as relevant, while the Scale-level Content Validity Index (S-CVI/Ave) was computed as the mean of all I-CVI values to provide an overall indication of item relevance at the construct level. Consistent with commonly cited methodological guidance, S-CVI values of 0.80 or above were considered indicative of adequate content validity (Lynn, 1986; Rubio et al., 2003; Polit & Beck, 2006), and these thresholds were treated as guiding criteria rather than definitive indicators of overall instrument quality.

According to Polit and Beck (2006), an acceptable Item level value is 1.00, while the acceptable Scale level value must also be 1.00. Based on the analysis presented in Tables 8a to 8c for the content dimensions of the questionnaire measuring the levels of knowledge, skills, and readiness of secondary school Science teachers in Sabah regarding the use of generative artificial intelligence, the Scale level value obtained was 1.00. This value exceeds the recommended threshold, indicating that the items in these dimensions demonstrate a high level of content validity and may be accepted without reservation. Tables 8a to 8c show that all Scale level values for each constructed dimension achieved a perfect score of 1.00. The dimension assessing teachers perceptions of Science teaching and learning recorded a Scale level value of 1.00, as did the dimension assessing Science teachers level of knowledge regarding generative artificial intelligence, the dimension assessing Science teachers skills in using generative artificial intelligence, and the dimension assessing Science teachers readiness to use generative artificial intelligence. A minimum acceptable Scale level value of 0.80 is required to demonstrate adequate content validity (Lynn, 1986; Rubio et al., 2003; Polit & Beck, 2006). All values for the questionnaire dimensions assessing secondary school Science teachers knowledge, skills, and readiness to use generative artificial intelligence in teaching and learning in Sabah achieved or exceeded this requirement. The overall Scale level value for the entire questionnaire was 1.00, confirming the instrument's high level of content validity. Although perfect I-CVI and S-CVI values indicate strong expert agreement, such results require cautious interpretation rather than uncritical acceptance. One plausible explanation for this unanimous agreement is the relative homogeneity of the expert panel, whose members shared similar professional backgrounds in

Science Education and Educational Technology. Although this shared expertise supports informed judgment, it may also increase the likelihood of shared assumptions, potentially inflating agreement and limiting the identification of alternative conceptual perspectives. Thus, the perfect CVI values observed in this study should be interpreted as evidence of perceived relevance rather than as conclusive proof of comprehensive instrument validity.

Despite all items being rated as relevant, several items (B1, B3, C1, C4, C8, D1,D3 and D4) were revised based on qualitative expert comments that focused on wording precision, sentence structure, and contextual clarity rather than construct misalignment. These revisions were therefore undertaken to enhance comprehensibility and contextual appropriateness for secondary school Science teachers, rather than to address deficiencies in content validity. Following expert review, the revised items were further examined by a language specialist to ensure accuracy, clarity, and appropriateness of Malay grammar, in line with recommendations that linguistic refinement should complement content validation in educational instrument development (Connell et al., 2018). Explicitly distinguishing content relevance decisions from language and clarity refinements strengthens the conceptual coherence of the validation process.

Table 8(a). Values I-CVI and S-CVI for Teachers' Knowledge of Generative AI

Item	P1	P2	P3	P4	No. of Agreement	I-CVI	Item Status
B1	1	1	1	1	4	1.00	Accepted
B2	1	1	1	1	4	1.00	Accepted
B3	1	1	1	1	4	1.00	Accepted
B4	1	1	1	1	4	1.00	Accepted
B5	1	1	1	1	4	1.00	Accepted
B6	1	1	1	1	4	1.00	Accepted
B7	1	1	1	1	4	1.00	Accepted
B8	1	1	1	1	4	1.00	Accepted
B9	1	1	1	1	4	1.00	Accepted
B10	1	1	1	1	4	1.00	Accepted
					S-CVI	1.00	Accepted

Table 8(b). Values of I-CVI and S-CVI for Teachers' Skills in Using Generative AI

Item	P1	P2	P3	P4	No. of Agreement	I-CVI	Item Status
C1	1	1	1	1	4	1.00	Accepted
C2	1	1	1	1	4	1.00	Accepted
C3	1	1	1	1	4	1.00	Accepted
C4	1	1	1	1	4	1.00	Accepted
C5	1	1	1	1	4	1.00	Accepted
C6	1	1	1	1	4	1.00	Accepted
C7	1	1	1	1	4	1.00	Accepted
C8	1	1	1	1	4	1.00	Accepted
C9	1	1	1	1	4	1.00	Accepted
C10	1	1	1	1	4	1.00	Accepted
C11	1	1	1	1	4	1.00	Accepted
C12	1	1	1	1	4	1.00	Accepted
					S-CVI	1.00	Accepted

Table 8(c). Values of I-CVI and S-CVI of Teachers' Readiness for the Adoption of Generative AI

Item	P1	P2	P3	P4	No. of Agreement	I-CVI	Item Status
D1	1	1	1	1	4	1.00	Accepted
D2	1	1	1	1	4	1.00	Accepted
D3	1	1	1	1	4	1.00	Accepted
D4	1	1	1	1	4	1.00	Accepted
D5	1	1	1	1	4	1.00	Accepted
D6	1	1	1	1	4	1.00	Accepted
D7	1	1	1	1	4	1.00	Accepted
D8	1	1	1	1	4	1.00	Accepted
D9	1	1	1	1	4	1.00	Accepted
D10	1	1	1	1	4	1.00	Accepted
D11	1	1	1	1	4	1.00	Accepted
D12	1	1	1	1	4	1.00	Accepted
					S-CVI	1.00	Accepted

The study shows that the questionnaire was developed to assess secondary school Science teachers' knowledge, skills, and readiness to use Generative AI in teaching and learning in Sabah demonstrates a high level of content validity. The S-CVI (average) for all dimensions was 1.00, aligned with the figures suggested by Polit and Beck (2006) and Polit et al. (2007). Each dimension including teachers' knowledge of Generative AI, teachers' skills in using Generative AI, and teachers' readiness for Generative AI was achieving the maximum S-CVI score of 1.00. These findings indicate that all items in the instrument are highly relevant and fully represent the constructs being measured.

In terms of I-CVI values, nearly all items achieved a score of 1.00, reflecting unanimous agreement among experts regarding item relevance. However, several items, including B1, D15, E5, E6, E7, E9, and E12, were revised based on expert feedback to enhance clarity and precision of meaning. This refinement process aligns with prior research emphasizes the need to revise items according to expert recommendations to improve linguistic accuracy and contextual suitability (Connell et al., 2018). Additionally, language specialists reviewed the instrument to ensure adherence to proper Malay grammar and appropriateness for the secondary school context.

The perfect agreement observed (I-CVI and S-CVI = 1.00) may be attributed to the homogeneity of the expert panel's background and experience in Educational Technology and Science Education, which likely supported consistent interpretation of items. This is consistent with Fernández-Gómez et al. (2020), who found that expert panels with similar areas of specialization tend to reach higher levels of consensus. Furthermore, the instrument was theory-driven and had undergone preliminary reviews and refinements, reducing ambiguity and strengthening conceptual clarity. The clear presentation of constructs in the evaluation form further supported more uniform expert assessments.

The content validity results of this instrument have both theoretical and practical implications. Theoretically, they contribute to refining the measured constructs by ensuring that each item accurately reflects the intended dimensions. This strengthens confidence that the instrument

effectively measures the knowledge, skills, and readiness of secondary school Science teachers to apply Generative AI in teaching and learning in Sabah. Practically, the validated instrument offers a strong foundation for developing targeted educational interventions and provides useful insights for curriculum planners and teachers. Following the content validity analysis, several refinements were implemented to improve the research instrument, as presented in Figure 3. These modifications were made based on the experts' comments and annotations documented in the content validity assessment forms.

JADUAL PEMBETULAN ITEM SOAL SELIDIK

KONSTRUK 1: Pengetahuan Guru Terhadap AI Generatif

Bil.	Item Asal (IA)	Item Pembetulan (IP)	Catatan
	Pengetahuan Teknologi (Technological Knowledge,TK)	Pengetahuan Teknologi *Technological Knowledge* (TK)	Italic untuk tulisan Bahasa Inggeris
TK1	Saya mengetahui pelbagai alat AI Generatif seperti ChatGPT, Gemini atau Copilot yang boleh digunakan dalam pengajaran Sains.	Saya mengetahui pelbagai alat AI Generatif seperti ChatGPT, Gemini atau Copilot yang boleh digunakan dalam pengajaran Sains.	Kekal (P1,P2,P3 setuju item sesuai)
TK2	Saya memahami bahawa AI Generatif boleh menjana kandungan seperti teks, gambar atau video melalui arahan (prompt) tertentu.	Saya mengetahui bahawa AI Generatif dapat menjana kandungan seperti teks, gambar atau video melalui arahan (prompt) tertentu.	Kekal (P1,P2,P3 setuju item sesuai)
TK3	Saya mengetahui bahawa AI Generatif boleh menterjemah bahan pembelajaran ke dalam pelbagai bahasa secara automatik.	Saya mengetahui bahawa AI Generatif boleh menterjemah bahan pembelajaran ke dalam pelbagai bahasa secara automatik.	P2 -
TK4	Saya memahami batasan, ketepatan dan isu etika yang perlu dipertimbangkan apabila menggunakan AI Generatif dalam pendidikan.	Saya mengetahui batasan, ketepatan dan isu etika yang perlu dipertimbangkan apabila menggunakan AI Generatif dalam pendidikan.	Kekal (P1,P2,P3 setuju item sesuai)

Figure 3. Table of Revisions to the Questionnaire Items

Although all items achieved perfect I-CVI values, the modified kappa analysis was used as a complementary indicator to account for the possibility of chance agreement among experts. For several items (e.g., B1, B3, C1, C4, C8, D1,D3 and D4), the adjusted κ values for items that underwent revision indicated acceptable agreement beyond chance, suggesting that expert consensus was not solely attributable to random agreement. Expert comments associated with these items consistently highlighted grammatical or phrasing issues rather than redundancy or conceptual overlap, and all revised items were retained to preserve comprehensive coverage of the constructs of teachers' knowledge, skills, and readiness to use generative artificial intelligence in teaching and learning. Consistent expert feedback highlighted the need to refine wording to enhance precision and contextual appropriateness, particularly within the secondary school Science teaching context. Following these revisions, all items were retained to preserve comprehensive coverage of the intended constructs, namely teachers' knowledge, skills, and readiness to use generative artificial intelligence in teaching and learning. The acceptable kappa values observed indicate that expert agreement exceeded chance levels, thereby supporting the adequacy of the revised items. As this study represents an initial phase of instrument validation, re-validation after item revision was not undertaken at this stage. Instead, the revised items are intended to undergo further psychometric evaluation, including exploratory and confirmatory factor analyses, in subsequent large-scale studies to strengthen construct validity and reliability.

Table 9 showed the summary of the content validity evaluation for teachers' knowledge, skills and readiness of generative AI items.

Table 9. Summary of Content Validity Evaluation for Teachers' Knowledge, Skills, and Readiness of Generative AI Items (I-CVI, UA, Pc, κ, Comments, Decision, New item, S-CVI/Ave and S-CVI/UA)

Item	P1	P2	P3	P4	No. of Agreement	I-CVI	UA	Pc	κ	Comments	Decision	New item after revision
B1	1	1	1	1	4	1.00	Accepted	0.06	0.94	Minor phrasing	Revised & retained	I am aware of various Generative AI platforms that can be used in Science teaching.
B2	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B3	1	1	1	1	4	1.00	Accepted	0.06	0.94	Clarification needed	Revised & retained	I am aware that Generative AI can help me adapt Science teaching to meet students' learning needs.
B4	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B5	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B6	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B7	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B8	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B9	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
B10	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C1	1	1	1	1	4	1.00	Accepted	0.06	0.94	Wording clarity	Revised & retained	I am skilled at using various Generative AI tools to obtain Science learning materials.
C2	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C3	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C4	1	1	1	1	4	1.00	Accepted	0.06	0.94	Wording clarity	Revised & retained	I am able to evaluate whether the content generated by Generative AI is suitable for Science teaching.
C5	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C6	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C7	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C8	1	1	1	1	4	1.00	Accepted	0.06	0.94	Sentence structure	Revised & retained	I am skilled at using Generative AI to generate exercises based on Science topics appropriate to students' levels.
C9	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C10	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C11	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
C12	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D1	1	1	1	1	4	1.00	Accepted	0.06	0.94	Wording clarity	Revised & retained	Generative AI helps me prepare Science teaching materials more quickly.
D2	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D3	1	1	1	1	4	1.00	Accepted	0.06	0.94	Wording clarity	Revised & retained	Generative AI makes me more efficient in teaching Science.

D4	1	1	1	1	4	1.00	Accepted	0.06	0.94	Wording clarity	Revised & retained	Generative AI makes it easier for me to deliver Science content clearly
D5	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D6	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D7	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D8	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D9	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D10	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D11	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
D12	1	1	1	1	4	1.00	Accepted	0.06	0.94	No issues	Retained	-
				S-CVI/Ave			1.00					
				S-CVI/UA			1.00					

CONCLUSION AND RECOMMENDATION

In conclusion, all 34 items were retained, as they met the content validity criteria based on I-CVI and S-CVI values, indicating strong expert agreement on item relevance and representativeness. However, content validity alone does not provide comprehensive evidence of the instrument's psychometric quality. Therefore, it is recommended that the study proceed with a pilot test to examine internal consistency and reliability, using measures such as Cronbach's alpha (α) and McDonald's omega (ω). Furthermore, subsequent validation should include exploratory and confirmatory factor analyses (EFA and CFA) to assess the dimensionality and construct validity of the instrument. To enhance the generalizability and robustness of the instrument, future research should also consider cross-context validation across different subject areas, including language teachers, and potentially across different educational settings. These steps will ensure that the instrument not only reflects content relevance but also demonstrates reliable and valid measurement properties suitable for broader application.

REFERENCES

Abdullah, N., Maskur, H., & Abdul Mutalib, R. (2020). Penggunaan aplikasi kepintaran buatan (AI) dalam PDP: Satu kajian kepustakaan. *Proceeding International Multidisciplinary Conference (IMC 2020)*, 265–271. https://penerbit.unimap.edu.my/images/pdf/IMC%202020/265-272.pdf

Al Darayseh, A. (2023). Acceptance of artificial intelligence in teaching science: Science teachers' perspective. *Computers and Education: Artificial Intelligence, 4*, Article 100132. https://doi.org/10.1016/j.caeai.2023.100132

Al Darayseh, A., & Mersin, N. (2025). Integrating generative AI into STEM education: Insights from science and mathematics teachers. *International Electronic Journal of Mathematics Education*, *20*(3), 1–13. https://doi.org/10.29333/iejme/16232

Aliya, N., Suriani, A. B., Wong, K. T., Azzam, A. B., & Mohamad, A. (2024). Integrasi aplikasi berasaskan kecerdasan buatan dalam pembelajaran digital dan kaedah pengajaran dan pembelajaran fizik kuantum: Tinjauan literatur. *Jurnal Pendidikan Sains dan Matematik Malaysia*, *14*(2), 155–170. https://doi.org/10.37134/jpsmm.vol14.2.12.2024

Almanasreh, E., Moles, R., & Chen, T. F. (2018). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K–12: A systematic literature review. *International Journal of STEM Education*, *10*, 1–17. https://doi.org/10.1186/s40594-023-00418-7

Celik, I. (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*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

Connell, J., Carlton, J., Grundy, A., Buck, E. T., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: Lessons learnt from service users when developing the

Recovering Quality of Life measure (ReQoL). *Quality of Life Research, 27*(7), 1893–1902. https://doi.org/10.1007/s11136-018-1847-y

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Davis, K. A. (1992). Validity and reliability in qualitative research on second language acquisition and teaching. *TESOL Quarterly, 26*, 605–608. https://doi.org/10.2307/3587190

El Fathi, T., Saad, A., Larhzil, H., Lamri, D., & Al Ibrahmi, E. M. (2025). Integrating generative AI into STEM education: Enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance. *Disciplinary and Interdisciplinary Science Education Research, 7*, 1–21. https://doi.org/10.1186/s43031-025-00125-z

Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content validation through expert judgement of an instrument on the nutritional knowledge, beliefs, and habits of pregnant women. *Nutrients, 12*(4), 1–13. https://doi.org/10.3390/nu12041136

Hussin, F. B., Ali, J., & Noor, M. S. Z. (2014). *Kaedah penyelidikan & analisis data SPSS*. UUM Press.

Ishmuradova, I. I., Zhdanov, S. P., Kondrashev, S. V., Erokhova, N. S., Grishnova, E. E., & Volosova, N. Y. (2025). Pre-service science teachers' perception on using generative artificial intelligence in science education. *Contemporary Educational Technology, 17*(3), Article ep579. https://doi.org/10.30935/cedtech/16207

Juwait, S., & Siew, N. M. (2022). Pengaruh pengetahuan teknologi kandungan dan komitmen terhadap motivasi guru Fizik melaksanakan pengajaran dan pembelajaran dalam talian di Sabah. *International Journal of Education, Psychology and Counseling, 7*(45), 422–445. https://doi.org/10.35631/IJEPC.745033

Lewis, J. R. (2019). Comparison of four TAM item formats: Effect of response option labels and order. *Journal of Usability Studies, 14*(4), 224–236.

Lowa, M. (2024). Democratizing access to education: AI-driven solutions for inclusive learning in Malaysia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4813758

Lukmanul Hakim, N. A. M., Rasidi Pairan, M., & Ikram Zakaria, M. (2025). Step-by-step guide to calculating Content Validity Index (CVI) for single constructs using Excel. *International Journal of Research and Innovation in Social Science, IX*(III), 1717–1726. https://doi.org/10.47772/ijriss.2025.90300135

Lynn, M. R. (1986). Determination and quantification of content validity index. *Nursing Research, 35*, 382–386. https://doi.org/10.1097/00006199-198611000-00017

Padzil, M. R., Abd Karim, A., & Husnin, H. (2021). Employing DDR to design and develop a flipped classroom and project-based learning module applying design thinking in design and technology. *International Journal of Advanced Computer Science and Applications, 12*(9), 791–798. https://doi.org/10.14569/IJACSA.2021.0120988

Ping, L. C., & Shaafi, N. F. (2026). Content validity and reliability analysis of instruments on students' perceptions of flipped STEM learning. *Sains Humanika, 18*(2), 69–79. https://doi.org/10.11113/sh.v18n2.2284

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. https://doi.org/10.1002/nur.20147

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. https://doi.org/10.1002/nur.20199

Rodrigues, I. B., Adachi, J. D., Beattie, K. A., & MacDermid, J. C. (2017). Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. *BMC Musculoskeletal Disorders, 18*, 1–9. https://doi.org/10.1186/s12891-017-1914-5

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94–104. https://doi.org/10.1093/swr/27.2.94

Samsudin, Salbihana & Zarizi, Syed & Abdullah, Syed & Saleh, Md & Shukri, Ahmad. (2023). Pengetahuan Bakal Guru Tentang Ai Dalam Pendidikan. Kolokium Penyelidikan, Pengurusan Dan Kepimpinan & Pertandingan Inovasi 2023. https://www.researchgate.net/publication/374368631_PENGETAHUAN_BAKAL_GURU_TENTANG_AI_DALAM_PENDIDIKAN

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Journal of Central South University (Medical Sciences), 37*(2), 152–155. https://doi.org/10.3969/j.issn.1672-7347.2012.02.007

Shrotryia, V. K., & Dhanda, U. (2019). Content validity of assessment instrument for employee engagement. *SAGE Open, 9*(1). https://doi.org/10.1177/2158244018821751

Tahar, A. M., & Zaid, N. M. (2019). Persepsi pelajar terhadap penggunaan Frog VLE dalam matapelajaran Biologi Tingkatan 4 (Student's perception towards the usage of Frog VLE in form 4 Biology subject). *Innovative Teaching and Learning Journal, 2*(2), 43–58. http://161.139.21.34/itlj/index.php/itlj/article/view/24

Taherdoost, H. (2016). Sampling methods in research methodology: How to choose a sampling technique for research. *International Journal of Academic Research in Management, 5*, 18–27. https://doi.org/10.2139/ssrn.3205035

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

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

This manuscript was developed with the assistance of ChatGPT (OpenAI) and Gemini (Google) for the purpose of language editing and improving the clarity of writing. The tools supported the authors in reviewing and refining the content, and the authors assume full responsibility for the final version submitted for publication.

Conflict of Interest

The authors have no conflicts of interest to declare.

Acknowledgement

We thank Universiti Malaysia Sabah for providing Geran Penyelidikan Dana Kluster (DKP) with Research Code: DKP0105 to support the research publication, and Ministry of Higher Education (MOHE) Malaysia for providing a Fundamental Research Grant Scheme Early Career (FRGS-EC) with Research Code: FRGS/1/2024/SSI09/UMS/02/6.

Authors' Contributions

Nur Farha Shaafi: Conceptualization, Methodology, Data curation, Visualization, Investigation, Writing-Reviewing and Editing. Julazzmie Kambutong: Methodology, Data Collection and Writing-Original draft preparation.