International Journal of Evaluation and Research in Education (IJERE)
Vol. 14, No. 6, December 2025, pp. 4307~4319
ISSN: 2252-8822, DOI: 10.11591/ijere.v14i6.33370

Content validity of an innovative behavior rubric for polytechnic engineering students

Nor Aisyah Che Derasid[1], Aede Hatib Musta'amal @ Jamal[1], Nornazira Suhairom[1], Hary Suswanto[2]
[1]Department of Advanced Technical and Vocational Education and Training, Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia, Skudai, Malaysia
[2]Department of Electrical and Informatics Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia

Article Info

Article history:

Received Oct 24, 2024
Revised May 9, 2025
Accepted Jul 28, 2025

Keywords:

Assessment rubric
Content validity
I-CVI
Innovative behavior
Kappa statistic

ABSTRACT

In engineering education, fostering innovative behaviors is crucial for preparing students to tackle complex, real-world challenges. Developing an assessment tool or rubric that accurately measures innovative behavior is essential to provide educators with the means to systematically evaluate students' innovative potential. This article mainly focuses on assessing the content validity of the innovative behavior assessment rubric, which is designed to measure the innovative behavior of engineering students. The rubric was designed around three core dimensions: problem recognition, idea generation, and idea implementation. Content validity was assessed using the item-level content validity index (I-CVI), scale-level content validity index (S-CVI), and modified kappa statistic. Expert evaluations resulted in a final 35-item rubric, with an overall S-CVI of 0.85, indicating high content validity. Items with I-CVI values below 0.70 were either revised or removed to ensure relevance and clarity. The study highlights the importance of expert judgment in the validation process and underscores the utility of both I-CVI and kappa in refining assessment tools. Future research will focus on construct and criterion validation, as well as practical application across diverse educational contexts.

Corresponding Author:

Nor Aisyah Che Derasid
Department of Advanced Technical and Vocational Education and Training
Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia
81310 Skudai, Johor, Malaysia
Email: noraisyahchederasid@gmail.com

1. INTRODUCTION

In the field of educational evaluation, especially in polytechnic engineering programs, the importance of effective assessment tools cannot be overstated. In engineering education, fostering innovative behaviors is crucial for preparing students to tackle complex, real-world challenges. As industries progress and technology advances rapidly, the ability to think creatively and execute innovative solutions becomes more essential. Nonetheless, conventional assessment methods frequently fail to adequately reflect students' innovative skills. Many existing instruments rely heavily on self-perception, which can introduce bias and undermine the validity of assessments [1], [2].

The reliance on self-reported measures for assessing innovative behavior can lead to significant discrepancies between perceived and actual competencies. Previous researches [3], [4] indicate that self-assessments are often influenced by cognitive biases, such as overconfidence or lack of awareness of one's limitations. This situation creates a pressing need for more objective assessment tools that can reliably evaluate innovative behavior in engineering students. Without such tools, educators may struggle to identify students' true capabilities, potentially hindering their development and preparedness for future challenges.

Previous studies [5], [6] underscore the significance of utilizing structured assessment instruments such as rubrics to boost the consistency and accuracy of evaluations. Other research [7], [8] emphasize that rubrics offer explicit criteria that aid both students and instructors in understanding expectations, thus minimizing subjectivity in grading. Additionally, Panadero *et al.* [9] discovered that rubrics have a beneficial effect on student performance by encouraging self-regulation and more defined goal setting. Rubrics have been demonstrated to support formative assessment practices, enabling students to track their progress and enhance their work based on targeted feedback [10].

Well-structured rubrics are essential for evaluating intricate skills like innovation and creativity, as they offer clear and observable standards for assessment. These tools facilitate objective evaluation and enhance the learning experience by aligning assessment with educational goals. Research by Navarrete [11] designed rubrics for capstone projects in engineering, which improved student performance and aligned with accreditation standards. In addition, research has shown that reliable and valid assessment tools can effectively evaluate innovative thinking. A study developed a five-factor model with 25 elements to measure innovative thinking in vocational students, demonstrating strong validity and reliability [12]. Rubrics can bridge the gap between subjective self-assessment and objective evaluation by articulating observable behaviors associated with innovative thinking.

To address the identified gaps in assessing innovative behavior among engineering students, this study proposes the development of a comprehensive rubric specifically designed for this purpose. Developing an assessment tool or rubric that accurately measures innovative behavior is essential to provide educators with the means to systematically evaluate students' innovative potential. Innovation is a complex skill [13] that includes creativity, problem-solving, critical thinking, and applying knowledge in original ways. This rubric aims to provide a robust framework for evaluating student performance objectively by utilizing a multi-dimensional approach that includes both qualitative and quantitative measures. Without a standardized way to assess these qualities, educators may struggle to identify and nurture the innovative potential within their students. The proposed rubric will allow educators to systematically evaluate the outcomes of students' innovative projects and the processes they employ, focusing on criteria such as original thought, problem recognition, idea generation, and implementation. By concentrating on these dimensions, educators can gain deeper insights into how students approach challenges and generate new ideas.

The rubric will undergo a rigorous content validity process involving expert reviews to ensure that it accurately reflects the construct of innovative behavior. Content validity is the degree to which an assessment instrument accurately reflects the construct. It aims to measure is a foundational element in ensuring that educational evaluations are reliable and meaningful. This process will include soliciting feedback from faculty members and industry professionals to align the rubric with current standards and expectations in engineering education. As highlighted in existing literature, content validity is crucial for the reliability of instruments and the overall quality of educational assessments [14], [15]. Indeed, Ferraz and Pereira-Guizzo [16] highlighted that content validity is crucial for ensuring that assessment items are relevant and representative of the target constructs, such as transferable skills in engineering students. This study focuses on developing a behavior rubric specifically tailored for polytechnic engineering students, aiming to enhance the assessment of innovative behaviors that are increasingly vital in today's dynamic technological landscape.

This study offers a novel solution to the lack of reliable instruments for evaluating innovative behavior in engineering students by developing and validating an individually tailored rubric designed for Malaysian polytechnic engineering education. Unlike existing assessment tools, this rubric is uniquely grounded in established content validity principles and specifically designed to capture the distinct dimensions of innovative behavior relevant to technical and vocational education. The novelty of this study lies in its integrated methodological approach, which combines both qualitative expert reviews and quantitative empirical testing to ensure the instrument's reliability and validity. This dual-validation process represents a rigorous and comprehensive strategy that is rarely applied in similar educational assessment tools. By doing so, the study not only addresses a significant gap in the literature but also contributes a practical, academically sound instrument that can be adopted across polytechnic institutions. The validated rubric offers new insights into how structured behavioral assessment can enhance student engagement, creativity, and innovation skills in engineering education, ultimately supporting more effective teaching and learning practices in line with 21st-century competency demands.

This article mainly focuses on assessing the content validity of the innovative behavior assessment rubric, which is designed to measure the innovative behavior of engineering students. Following a thorough content validity process and incorporating expert feedback, all accepted items will be utilized to develop a comprehensive rubric for evaluating innovative behavior in this context. The specific techniques employed in the item development, which includes conducting literature reviews, consulting experts, and analyzing metadata, as well as the fundamental dimensions that underpin their design and functionality, fall outside the scope of this paper. A comprehensive analysis of these aspects would require a separate investigation, given their complexity and importance in understanding the overall nature of the items in question. Instead, the article seeks to confirm that the instrument created is both valid and reliable, without elaborating on the specifics of how the constructs were originally developed.

2. METHOD

2.1. The innovative behavior instruments

The study aimed to measure innovative behavior levels among engineering students in polytechnics through the innovative behavior assessment rubric. The main tool for this study is an analytical rubric. This scoring system allows lecturers to evaluate students' work against defined criteria. The creation of this tool was based on insights from the learning outcomes found in the compilation of learning outcome references and the learning taxonomy (JPPKK, 2019), adhering to the Malaysian Polytechnic Standards (MyPOLYStandards) established by the Department of Polytechnic Education (DPE), Ministry of Education (MOE), alongside research from past journal articles focused on innovative behaviors. A thorough examination of these documents and a literature review was undertaken to pinpoint the innovative behavior's content domain. Through metadata analysis and the grounded theory method, the researcher pinpointed three key components of innovative behavior: problem recognition, idea generation, and idea implementation. Table 1 shows how items are distributed across the instrument.

Expert panels were consulted to assess the instrument's comprehensive coverage of the construct it is designed to measure, by reviewing the list of items. Experts are tasked with assessing 46 items using a relevance scale that ranges from 1 to 4 (1 indicates not relevant, 2 signifies somewhat relevant, 3 represents quite relevant, and 4 denotes highly relevant), as shown in Table 2. These items are categorized into three primary constructs: problem recognition, idea generation, and idea implementation. Problem recognition emphasizes a student's ability to identify and articulate problems effectively. Idea generation evaluates a student's creativity and capacity to produce viable solutions, while idea implementation measures how well students can convert ideas into actionable plans.

Table 1. List of item distribution

Dimensions	Sub-dimensions	Total of items
Problem recognition	Problem finding	6
	Fact finding	5
	Problem definition	5
Idea generation	Idea exploration	8
	Solution evaluation	8
Idea implementation	Action plan	7
	Prototyping	7
Overall		46

2.2. Identification of validation panels

Experts are highly knowledgeable and skilled individuals who have extensive expertise in a particular field, allowing them to provide valuable insights and solutions within that domain [17], [18]. Determining the optimal number of experts for panel studies varies significantly across the literature, reflecting different methodological approaches and contextual needs. Amirzadeh *et al.* [15] advocates for a minimum of five experts to achieve reliable consensus, emphasizing the need for diverse perspectives to reduce bias. Conversely, that three experts may be sufficient for specialized topics [19], [20]. Nevertheless, the Lawshe content validation method just needs four experts at the very least. Research by [21], [22] suggest a flexible range of 3 to 10 experts per category to balance inclusivity with manageability. Overall, the ideal expert panel size is influenced by topic complexity, expert availability, and study objectives, generally recommending at least 6 and no more than 10 experts for effective content validation. Therefore, the eight experts in this study are based on their professional and field experts, respectively.

Professional experts typically have the necessary job experience and expertise to collaborate on the study, as well as knowledge of the research field, publications on the subject, and presentations at national or worldwide seminars [23]–[25]. These experts assist in determining the degree to which the instrument was designed with psychometric testing to be considered [26]. The criteria used to select the experts were compulsory to have at least 10 years of working experience. Experts were also chosen based on their experience supervising final-year student projects and their knowledge of the product design process, innovation, and design thinking. Each expert panel comprises faculty members from different institutions, evaluated based on their qualifications. Furthermore, experts in curriculum development, instrument design, and psychometrics were chosen to participate in the study's expert panel. This opinion that the chosen experts ought to be specialists in both the subject field and instrument development was reinforced by additional researches [27], [28]. These experts were selected due to their diverse backgrounds and experiences, guaranteeing a well-rounded view on the topic. They contribute to refining the study's item list by providing insights on the most relevant, necessary, or effective items. This may involve suggesting new items to include, recommending the removal of less relevant ones, or modifying existing items to improve clarity and effectiveness. The list of expert panels' expertise and years of experience are shown in Table 3.

Table 2. Content validation form

Judgments			How relevant is this item?			
			1	2	3	4
Problem recognition						
Problem finding	1.	Analyze empathy (understand feeling of another) on users' needs				
	2.	Select promising prospects with relevant justification				
	3.	Recognize underlying issues that may not be immediately apparent				
	4.	Distinguish between symptoms and root causes of problems				
	5.	Identifying issues, changes, and chances for improvement				
	6.	Select problems worth investigating				
Fact finding	7.	Gather relevant data to understand the problem context				
	8.	Analyze existing information to identify gaps				
	9.	Summarize findings in a clear and concise manner				
	10.	Evaluate the credibility of sources when collecting facts				
	11.	Consult stakeholders to gain different perspectives on the problem				
Problem definition	12.	Develop a concise, measurable, and systematic objectives				
	13.	Construct detailed problem statement				
	14.	Demonstrate innovative solution approach for continual improvement				
	15.	Analyze the problem and fully describe the problem in own words				
	16.	Relate the problem to a real-world need				
Idea generation						
Idea exploration	17.	Generates a wide range of diverse and unconventional ideas before narrowing down the options				
	18.	Demonstrates skill in selecting a smaller number of promising solutions for further evaluation and development				
	19.	Brainstorm multiple solutions to a given problem				
	20.	Record all generated ideas without immediate judgment				
	21.	Identify potential barriers to idea generation and address them				
	22.	Encourage unconventional thinking during idea generation sessions				
	23.	Collaborate with peers to expand idea generation efforts				
	24.	Utilize feedback from others to enhance idea exploration				
Solution evaluation	25.	Assess the feasibility of proposed solutions				
	26.	Compare potential solutions against established criteria				
	27.	Analyze risks associated with each solution option				
	28.	Use decision-making frameworks to guide evaluations				
	29.	Seek input from experts when evaluating solutions				
	30.	Prioritize solutions based on effectiveness and efficiency				
	31.	Develop a pros and cons list for each potential solution				
	32.	Document the rationale for selecting specific solutions				
Idea implementation						
Action plan	33.	Explain clear prospect of solutions to achieve goals in well-defined steps				
	34.	Develop a step-by-step action plan for implementing a solution				
	35.	Identify necessary resources for successful implementation				
	36.	Assign timeliness for each task in the plan				
	37.	Explain clear prospect of solutions to achieve goal in well-defined steps				
	38.	Communicate the action plan clearly to all stakeholders involved				
	39.	Review and adjust the action plan as needed based on feedback				
Prototyping	40.	Create a prototype or model of the proposed solution				
	41.	Test the prototype under real-world conditions to gather data				
	42.	Document lessons learned during the prototyping process				
	43.	Solicit feedback on the prototype from users and stakeholders				
	44.	Refine the prototype based on testing results and feedback received				
	45.	Present findings from prototyping efforts to relevant stakeholders				
	46.	Prepare for full-scale implementation based on prototype insights				

2.3. Sampling of expert panels

In this study, experts were deliberately selected based on their relevant expertise in the domain constructs of the instrument. Therefore, purposive sampling was employed in this study. Following their consent to participate, they were sent appointment letters and validation materials through email. The experts involved in the study were widely dispersed geographically, so an online approach was utilized to carry out the evaluation and validation processes. The instrument consists of 46 items, and each panel of experts was given 14 days to review and validate the items to streamline the procedure.

When evaluating the innovative behavior rubric, it is important to consider how well the rubric items reflect the specific characteristics and competencies relevant to the content area [7], [29]. The detailed instructions given to the evaluators were crucial in guiding them through this process. They carefully reviewed each rubric item to assess its alignment with the behaviors and skills important for innovative thinking and problem-solving in that field. Focusing on making each rubric item relevant and easy to score improves its practicality. Experts are encouraged to check if each item accurately reflects important aspects of innovative behavior in their specific field. This ensures that the items are based on theory and applicable in real-life situations, making the rubric more relevant to actual practices [30].

Table 3. Experts panel list

Code	Expertise	Current institutions	Experience (years)
E1	A doctorate holder: rubric development for product design innovation in field of technical and vocational education.	Deputy Director of the Innovation Unit at the Department of Polytechnic Education and Community College.	18
E2	A lecturer: currently teaching, supervising, and assessing students' final-year projects.	Department of Mechanical Engineering, in polytechnic.	17
E3	A lecturer: currently teaching, supervising, and assessing students' final-year projects.	Department of Mechanical Engineering, in polytechnic.	16
E4	A lecturer: currently teaching, supervising, and assessing students' final-year projects.	Department of Electrical Engineering, in polytechnic.	15
E5	A lecturer: currently teaching, supervising, and assessing students' final-year projects.	Department of Civil Engineering, in polytechnic.	15
E6	A doctorate holder and lecturer: facilitating industry collaboration and innovation competitions, particularly in innovation-driven fields.	Head of the unit of research, innovation, and commercialization in polytechnic.	18
E7	A doctorate holder: psychometrics and educational measurement with expertise in rubric development.	Senior lecturer in public university.	18
E8	A civil engineering lecturer with extended experience in curriculum development and student assessment.	Head of student projects, unit of research, innovation, and commercialization, in polytechnic.	18

2.4. Quantification of content validity

This section outlines the methods used to quantify content validity for the innovative behavior assessment rubric developed for polytechnic engineering students. The quantification involved two key metrics: the content validity index (CVI) and kappa statistics, which together provide a comprehensive evaluation of item relevance and expert agreement.

2.4.1. Content validity index

The content validity of an instrument can be determined by consulting a panel of experts for their perspectives. The CVI approach is the most used method to assess content validity by many researchers, including those in studies [31]–[33]. CVI measures agreement among experts on items' relevance. The CVI can be evaluated either at the individual item-level content validity index (I-CVI) or for the entire scale-level content validity index (S-CVI). The I-CVI calculates the percentage of experts who consider an item relevant, while the S-CVI is determined by taking the average of all the I-CVIs for the items in the scale. The main goal of the CVI is to thoroughly assess how effectively the items, as a group, capture the intended construct. To eliminate the possibility of neutral responses, a 4-point scale is utilized, consisting of the ratings: 1=not relevant, 2=somewhat relevant, 3=quite relevant, and 4=highly relevant. Before calculating the CVI, it is essential to reformat the relevance ratings by assigning a value of 1 to scores of 3 or 4, and a value of 0 to scores of 1 or 2. The calculation of the CVI indices, as presented in Table 4, is exemplified by data provided in Table 5 (results).

A CVI value, which ranges from 0 to 1, indicates the strength of content validity, with higher values denoting better validity. The minimum acceptable S-CVI is any value between 0.80 to 0.90 [18], [34], [35]. According to previous studies [36], [37], it is recommended that judges should have at least an 80% agreement for newly developed instruments. Thus, the evaluation criteria for each item are as: an I-CVI greater than 0.79 signifies that the item is deemed suitable. Should the I-CVI fall within the range of 0.70 to 0.79, the item necessitates revision. An I-CVI below 0.70 leads to the item being removed. Several studies [37], [38] recommended setting the I-CVI at 1.00 for cases involving five or fewer judges. In contrast, the I-CVI should not fall below 0.78 for studies with 6 or more experts. However, the I-CVI should be at least

0.78 for research involving six or more experts. This methodical approach not only increases the rigor of the evaluation but also lays out a definitive guide for refining the instrument in response to expert critiques.

Table 4. CVI calculation [20], [36]

CVI indices	Definition	Formula
I-CVI	The proportion of content experts giving item a relevance rating 3 or 4.	I-CVI=(agreed item)/(number of expert)
S-CVI/Ave (scale-level content validity index based on the average method)	The average of the I-CVI scores for all items on the scale or the average of proportion relevance judged by all experts. The proportion relevant is the average of relevance rating by individual expert.	S-CVI/Ave=(sum of I-CVI scores)/(number of item) S-CVI/Ave=(sum of proportion relevance rating)/(number of expert)

Table 5. Interpretation of kappa statistic coefficient [18], [36]

Kappa values	Interpretation
0.74 and above	Excellent
0.6 to 0.74	Good
0.40 to 0.59	Fair

2.4.2. Kappa

The CVI is a widely used method for evaluating content validity. However, it overlooks the potential for chance agreement to influence the outcomes. To address this limitation, it is advisable to employ both the CVI and the multi-rater kappa statistic in content validity research [39]. The kappa statistic, unlike the CVI, compensates for chance agreement, a point highlighted in previous research [40], [41]. As an index measuring the agreement among raters and adjusting for chance, the kappa statistic offers crucial insights into the extent of agreement that surpasses mere chance, making it an essential complement to the CVI.

First, it is necessary to determine the probability of random concordance for every item to determine the modified kappa statistic. This is done using the (1):

$$PC = [N!/A!\ (N - A)!] * 5N \qquad (1)$$

here, N represents the total number of experts in the panel, while A denotes the number of panelists who believe the item is relevant. After calculating the I-CVI for each item in the instrument, the kappa value was obtained using the (2):

$$K = (I - CVI - Pc)/(1 - Pc) \qquad (2)$$

where, probability of chance agreement (Pc) represents the numerical values of the Pc and I-CVI signifies the CVI for each item. The kappa evaluation criteria classify values as seen in Table 5.

3. RESULTS AND DISCUSSION

The findings of this study provide significant insights into the development and validation of the innovative behavior assessment rubric for polytechnic engineering students. The analysis of the CVI and modified kappa statistics revealed three critical dimensions of innovative behavior: problem recognition, idea generation, and idea implementation, encompassing a total of 46 assessment items. Notably, the cumulative S-CVI for the rubric was established at 0.85, indicating a high level of content validity as assessed by expert panels. The average S-CVI scores for the dimensions of problem recognition, idea generation, and idea implementation in innovative behavior were 0.83, refer to Table 6; 0.80, refer to Table 7; and 0.91, refer to Table 8, respectively. The results demonstrated that most items within the dimensions exhibited strong expert consensus, with I-CVI values predominantly above the acceptable threshold of 0.70. For all items across the three dimensions, the I-CVI values varied between 0.38 and 1. Any item displaying an I-CVI below 0.7 suggests the necessity for refinement.

According to Table 6, 13 out of 16 items in the problem recognition dimension met the standard, while 10 out of 16 items in the idea generation dimension were deemed appropriate, as seen in Table 7. The analysis presented in Table 8 indicates that most of the assessed items in idea implementation demonstrate perfect agreement beyond chance, as evidenced by nine out of fourteen items achieving both an I-CVI and a kappa statistic of 1.00. However, several items across all dimensions fell below this threshold, necessitating refinement or removal to enhance the rubric's overall effectiveness [18]. This highlights a critical aspect of the validation process: not only does it confirm which items are relevant, but it also identifies those that may compromise the integrity of the assessment tool. The I-CVI and kappa scores play a critical role in the refinement of assessment items in developing a rubric for innovative behavior.

Table 6. Ratings for problem recognition dimension of engineering students' innovative behavior

Items	E1	E2	E3	E4	E5	E6	E7	E8	Experts	Agree	I-CVI	Pc	K
1	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
2	-	X	-	X	X	X	-	X	8	5	0.63	0.218750	0.52
3	X	X	X	X	X	-	X	X	8	7	0.88	0.031250	0.87
4	X	-	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
5	X	X	X	X	X	X	X	-	8	7	0.88	0.031250	0.87
6	X	-	-	X	-	-	X	-	8	3	0.38	0.218750	0.20
7	X	-	X	X	-	X	X	X	8	7	0.88	0.031250	0.87
8	X	-	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
9	X	X	-	X	X	X	X	X	8	8	1.00	0.003906	1.00
10	X	X	X	X	X	X	X	-	8	7	0.88	0.031250	0.87
11	-	X	X	X	-	X	-	-	8	4	0.50	0.273438	0.31
12	X	X	-	X	X	X	X	X	8	7	0.88	0.031250	0.87
13	X	X	-	X	X	X	X	X	8	7	0.88	0.031250	0.87
14	X	X	-	X	X	X	X	X	8	7	0.88	0.031250	0.87
15	X	-	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
16	X	X	X	X	X	X	X	-	8	7	0.88	0.031250	0.87

Note: S-CVI (average)=0.83 (accepted). I-CVI=item content validity index; Pc=probability of chance agreement; S-CVI=scale content validity index.

Table 7. Ratings for idea generation dimension of engineering students' innovative behavior

Items	E1	E2	E3	E4	E5	E6	E7	E8	Experts	Agree	I-CVI	Pc	K
17	X	X	-	X	X	X	X	-	8	6	0.75	0.109375	0.72
18	X	X	-	X	X	X	X	-	8	6	0.75	0.109375	0.72
19	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
20	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
21	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
22	-	-	-	X	X	-	X	X	8	3	0.38	0.218750	0.20
23	-	X	-	X	X	X	-	-	8	5	0.63	0.218750	0.52
24	X	X	X	X	-	X	X	X	8	7	0.88	0.031250	0.87
25	X	X	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
26	X	X	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
27	X	X	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
28	X	X	-	X	-	-	X	-	8	4	0.50	0.273438	0.31
29	-	X	-	-	-	X	X	-	8	3	0.38	0.218750	0.20
30	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
31	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
32	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00

Note: S-CVI (average)=0.80 (accepted). I-CVI=item content validity index; Pc=probability of chance agreement; S-CVI=scale content validity index.

Table 8. Ratings for idea implementation dimension of engineering students' innovative behavior

Items	E1	E2	E3	E4	E5	E6	E7	E8	Experts	Agree	I-CVI	Pc	K
33	X	X	X	X	-	X	X	-	8	6	0.75	0.109375	0.72
34	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
35	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
36	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
37	X	X	X	X	-	-	-	X	8	5	0.63	0.218750	0.52
38	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
39	X	X	X	-	X	X	X	X	8	7	0.88	0.031250	0.87
40	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
41	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
42	X	X	-	X	X	X	-	-	8	5	0.63	0.218750	0.52
43	-	X	X	X	X	X	X	X	8	7	0.88	0.031250	0.87
44	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
45	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00
46	X	X	X	X	X	X	X	X	8	8	1.00	0.003906	1.00

Note: S-CVI (average)=0.90 (accepted). I-CVI=item content validity index; Pc=probability of chance agreement; S-CVI=scale content validity index.

The I-CVI provides a clear metric for determining item relevance based on expert consensus, while kappa scores evaluate the reliability of that consensus beyond chance agreement [42]. Together, these metrics guide researchers in making informed decisions about which items to retain, revise, or remove, ultimately leading to a more valid and reliable assessment tool for measuring innovative behavior among polytechnic engineering students. This rigorous refinement process is essential for ensuring that the rubric accurately captures the multifaceted nature of innovation and effectively supports educational objectives in engineering education [43], [44].

Items with an I-CVI of 1.00 (items 1, 4, 9, and 14) also have a kappa value of 1.00, indicating perfect agreement beyond chance. However, item 6 has a very low kappa value (0.20) due to both its low I-CVI and higher chance agreement. The kappa statistic for most items was within the range of 0.20 to 1. The evaluation of the rubric revealed significant insights regarding its content validity, particularly within the idea generation dimension. Out of 16 items assessed, 10 demonstrated strong relevance with I-CVI values above 0.79 and favorable kappa statistics, indicating a high level of expert consensus. However, 6 items (17, 18, 22, 23, 28, and 29) exhibited lower I-CVI scores ranging from 0.38 to 0.75, necessitating further revisions. Notably, 3 of these items had kappa statistics significantly below 0.4, suggesting that their inclusion could compromise the overall validity of the rubric.

The S-CVI for this dimension was acceptable but highlighted potential issues due to the presence of several low-scoring items. A higher S-CVI indicates that the rubric effectively captures the intended constructs of innovative behavior; however, problematic items can detract from this overall measure [45]. The findings underscore the importance of maintaining high standards for item inclusion to ensure that the rubric remains a reliable tool for assessing innovative behavior in engineering students.

The analysis also revealed that items with high I-CVI scores, but low kappa values warrant further scrutiny. Such discrepancies suggest that while there may be agreement on an item’s relevance among experts, this consensus might not reflect genuine alignment with the construct being measured as highlighted by previous research [46], [47]. This dual assessment using both I-CVI and kappa statistics is crucial for refining the rubric and ensuring that it accurately reflects innovative behaviors.

Based on these findings, 9 items (2, 6, 11, 22, 23, 28, 29, 37, and 42) were ultimately removed from the instrument due to insufficient expert support (less than 80% consensus). This decision is critical because retaining items lacking robust validation could lead to inaccuracies in measuring students’ innovative capabilities. The removal process highlights the necessity of a rigorous validation framework that prioritizes expert judgment and statistical analysis to enhance content validity. The underlying reason for their removal is that items lacking support from at least 80% of experts could lead to inaccuracies or irrelevant data in the instrument, compromising its overall content validity [20], [48].

Additionally, modifications were made to several items based on expert recommendations to improve clarity and reduce redundancy. Research by Lorenzo-Seva and Ferrando [49] emphasized that it is crucial to minimize redundancy while maintaining reliability in the development of the instrument. Thus, Izah *et al.* [50] recommended methods such as item analysis and optimization can minimize redundancy and improve homogeneity, ensuring that tests are both reliable and efficient. For instance, item 17 (generates a wide range of diverse and unconventional ideas before narrowing down the options) and item 18 (demonstrates skill in selecting a smaller number of promising solutions for further evaluation and development) were combined after experts noted their similar content. This streamlining process is essential for ensuring that each item distinctly assesses a specific aspect of innovative behavior without overlap [51]. Despite receiving a favorable CVI score, item 33 (explain clear prospect of solutions to achieve goals in well-defined steps) was flagged by expert feedback. This suggests that the item may not align well with the overall construct or could be redundant within the context of the assessment tool.

Overall, the iterative process of item refinement and expert validation demonstrates a commitment to methodological rigor in developing an effective assessment tool. The final 35-item rubric represents a validated measure capable of accurately assessing innovative behavior among polytechnic engineering students, as shown in Table 9. By employing both I-CVI and kappa analyses throughout this process, the study ensures that the rubric not only aligns with theoretical constructs but also meets practical validation standards. This combined strategy improves the reliability and usability of the rubric in different contexts [52], [53].

Table 9. Summary of item amendments

Dimensions	Item number			Total number items
	Remain	Revise (combine)	Remove	
Problem recognition	1, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16		2, 5, 11	13
Idea generation	19, 20, 21, 24, 25, 26, 27, 30, 31, 32	17 and 18	22, 23, 28, 29	11
Idea implementation	34, 35, 36, 38, 39, 40, 41, 43, 44, 45, 46		33, 37, 42	11
Total remaining items				35

Contextually, these findings align with existing literature emphasizing the importance of structured rubrics in educational assessments. Previous studies have indicated that traditional self-reported measures can introduce significant biases, undermining the validity of evaluations [54], [55]. By employing a rigorous validation process [56]; involving expert review and statistical analysis, this study addresses these concerns, providing a more objective framework for assessing innovative behaviors in engineering students.

The implications of these findings are profound for both educators and students. For educators, the validated rubric serves as a reliable tool for evaluating student performance in innovative behavior, fostering a more systematic approach to assessment that aligns with educational goals. It allows instructors to provide targeted feedback based on specific criteria, ultimately enhancing teaching practices and supporting student learning outcomes. For students, understanding how their innovative capabilities are assessed can promote self-reflection and encourage them to engage more deeply with their learning processes. Supporting literature reinforces these findings, Yusop *et al.* [57] highlights that clear assessment criteria are essential for promoting student engagement and improving learning outcomes. Transparent assessments evaluate student abilities and inform teaching methods, providing feedback that improves instruction and supports learning. Furthermore, Brookhart *et al.* [58] discusses that using rubrics in higher education boosts student motivation and academic performance. It also provides constructive feedback and increases interest in scientific literature.

Additionally, this research highlights the importance of ongoing enhancement of assessment instruments. The detection and elimination of items with low I-CVI scores not only simplifies the rubric but also guarantees that it stays concentrated on evaluating pertinent constructs. This process of iterative refinement is crucial for upholding high standards in educational evaluations and making certain that they truly represent students' abilities [59]. To sum up, this study makes a substantial contribution to the discussion surrounding the evaluation of innovative behavior in engineering education by introducing a validated tool that improves both teaching methods and student involvement. As industries progressively seek graduates capable of creative thinking and the application of innovative solutions, resources such as the innovative behavior assessment rubric will be essential in equipping students for upcoming challenges.

4. CONCLUSION

This study aimed to address the critical need for effective measurement tools in assessing innovative behavior among polytechnic engineering students. The findings confirm that the developed innovative behavior assessment rubric is a valid and reliable instrument for evaluating key dimensions of innovation—problem recognition, idea generation, and idea implementation. The rigorous content validation process, which included expert reviews and statistical analyses using the I-CVI and kappa statistics, demonstrated strong expert consensus on the relevance of most items, with an overall S-CVI of 0.85 indicating high content validity. By employing a robust validation process using S-CVI, I-CVI, and kappa analysis, we ensured that the items selected for the rubric align with expert consensus and reflect the essential components of the innovative behavior construct. The results align with the expectations set forth in the introduction, emphasizing the importance of structured rubrics in enhancing assessment quality and reducing biases associated with self-reported measures. By systematically evaluating students' innovative capabilities through clear criteria, this rubric not only supports educators in identifying and nurturing students' potential but also fosters an environment conducive to creativity and problem-solving. Furthermore, the iterative process of refining items based on expert feedback underscores the commitment to methodological rigor in developing assessment tools. The removal of nine items that did not meet the established validity thresholds ensures that the final rubric remains focused and effective. This careful curation enhances the overall reliability of the instrument, thereby contributing to its practical application in educational settings. Moving forward, once this rubric is finalized through additional phases like face validation, criterion validation, and practical testing, it is expected to provide educators with a clear, focused, and comprehensive tool. This tool will be vital for evaluating the different dimensions of innovative behavior in students, ensuring that the evaluation is accurate and aligned with the specific goals of educational programs aimed at fostering innovation. Looking ahead, the findings from this study pave the way for future research into the application and effectiveness of the rubric across diverse engineering contexts.

Further studies could explore how this assessment tool influences student learning outcomes and engagement in innovative projects. Additionally, expanding its use to other educational institutions could provide valuable insights into its adaptability and effectiveness in measuring innovative behavior beyond polytechnic settings. Despite the promising results, this study has some limitations that should be acknowledged. Firstly, relying on expert judgment for content validation introduces the potential for subjective bias. Even though efforts were made to ensure a diverse panel of experts, the opinions of a limited number of evaluators may not fully capture all perspectives. Additionally, while the S-CVI, I-CVI, and kappa analyses provide strong statistical support for content validity, they do not address other aspects of validity, such as criterion-related validity, face validity, and reliability testing. The study was limited to content validation, so the tool has yet to be tested for how well it measures innovative behavior in practice. Furthermore, the tool was developed specifically for polytechnic engineering students, which may limit its generalizability to other disciplines or educational settings. The constructs of innovative behavior—problem recognition, idea generation, and idea implementation may manifest differently across fields, which could require adjustments to the tool's content and structure. Additionally, the elimination of certain items, while statistically justified, might result in omitting potentially important aspects of innovative behavior that were not captured by the retained items.

Future research should focus on validating the tool in real-world settings to assess its practical applicability and ensure construct validity. Implementing the tool in diverse educational institutions, both within and beyond polytechnic contexts, will help determine its generalizability and reliability across different populations. Further studies could explore whether the current constructs of innovative behavior are sufficient or if additional dimensions, such as collaboration or adaptability, must be integrated into the tool. In addition, future research should consider conducting the rubric alongside existing innovation assessment tools in the same population to compare their results. This could include analyzing how well the rubric performs relative to established creativity or innovation measures regarding reliability, ease of use, and comprehensiveness. Gathering feedback from educators on the rubric's practicality and impact is encouraged. Conducting surveys to assess its ease of use and clarity will provide insights into improving it as a classroom assessment tool. By proposing these future studies, the paper can highlight a clear path for further development and refinement of the rubric. These studies will not only strengthen the theoretical and empirical foundation of the tool but also extend its applicability, making it a robust and versatile instrument for assessing innovative behavior in various educational and professional contexts. In summary, this research not only fills a critical gap in engineering education assessment but also offers a robust framework for evaluating innovative behaviors that are essential for success in today's dynamic technological landscape. The implications of this study extend beyond immediate educational practices, contributing to broader discussions on quality assurance and accountability in higher education.

ACKNOWLEDGMENTS

The authors sincerely thank the Ministry of Higher Education Malaysia and the Department of Polytechnic and Community College Education for their support and facilitation throughout this study. Sincere gratitude is hereby expressed to the esteemed lecturers and academic experts from various polytechnics and institutions affiliated with the Ministry of Higher Education. Their invaluable expertise and profound insights, as panel members during the content validation process of the innovative behavior rubric, were pivotal in the successful culmination of this research endeavor.

FUNDING INFORMATION

The authors would like to express our sincere appreciation to the Malaysia Ministry of Higher Education for the financial support provided under the Fundamental Research Grant Scheme (FRGS Ref: FRGS/1/2023/SS107/UTM/02/14) with Universiti Teknologi Malaysia Vot.No:5F639.

AUTHOR CONTRIBUTIONS STATEMENT)

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nor Aisyah Che Derasid	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Aede Hatib Musta'amal @ Jamal	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nornazira Suhairom	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		
Hary Suswanto	✓		✓	✓						✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author [NACD], upon reasonable request.

REFERENCES

[1] M. Akhtar, M. Bashir, K. Z. Gillani, and M. Abubakar, "The Influence of Perceived over Qualification on Innovative Behaviour of University Teachers: An Investigation based on Organizational Trust and Self Efficacy," *Pakistan Journal of Humanities and Social Sciences*, vol. 12, no. 1, pp. 304–321, Mar. 2024, doi: 10.52131/pjhss.2024.v12i1.2040.

[2] A. Azemi, "An Approach to Measure Innovation and Entrepreneurship Fostering of a Curriculum by Using ABET Assessment Data," in *2022 IEEE Frontiers in Education Conference (FIE)*, Oct. 2022, pp. 1–8, doi: 10.1109/FIE56618.2022.9962385.

[3] H. M. Ernst, J. Wittwer, and T. Voss, "Do they know what they know? Accuracy in teacher candidates' self-assessments and its influencing factors," *British Educational Research Journal*, vol. 49, no. 4, pp. 649–673, Aug. 2023, doi: 10.1002/berj.3860.

[4] S. H. Fraundorf, Z. A. Caddick, T. J. Nokes-Malach, and B. M. Rottman, "Cognitive perspectives on maintaining physicians' medical expertise: III. Strengths and weaknesses of self-assessment," *Cognitive Research: Principles and Implications*, vol. 8, no. 1, p. 58, Aug. 2023, doi: 10.1186/s41235-023-00511-z.

[5] Z. Ç. Edizer, M. Oryaşın, Ş. Dilidüzgün, and D. A. Başoğul, "Rubric for Evaluating Text Structure-Oriented Reading Tasks: A Study of Validity and Reliability," *International Journal of Progressive Education*, vol. 19, no. 2, pp. 118–134, Apr. 2023, doi: 10.29329/ijpe.2023.534.8.

[6] T. Farzana and A. Ilahi, "Quality in education: ascertaining the use of rubrics for successful implementation of OBE at the tertiary level in Bangladesh," *Education Quarterly Reviews*, vol. 6, no. 3, pp. 135–153, Sep. 2023, doi: 10.31014/aior.1993.06.03.769.

[7] S. M. Brookhart, "Appropriate Criteria: Key to Effective Rubrics," in *Frontiers in Education*, Apr. 2018, vol. 3, p. 22, doi: 10.3389/feduc.2018.00022.

[8] T. Hadibarata, T. Hidayat, and J. Kwabena, "The Use of Scoring Rubrics in University," *Acta Pedagogia Asiana*, vol. 3, no. 1, pp. 1–12, Jan. 2024, doi: 10.53623/apga.v3i1.312.

[9] E. Panadero, A. Jonsson, L. Pinedo, and B. Fernández-Castilla, "Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review," *Educational Psychology Review*, vol. 35, no. 4, p. 113, Dec. 2023, doi: 10.1007/s10648-023-09823-4.

[10] P. Karaman, "Effects of using rubrics in self-assessment with instructor feedback on pre-service teachers' academic performance, self-regulated learning and perceptions of self-assessment," *European Journal of Psychology of Education*, vol. 39, no. 3, pp. 2551–2574, Sep. 2024, doi: 10.1007/s10212-024-00867-w.

[11] P. F. H. Navarrete, "Design of Rubrics for the Evaluation of Research Work in the Capstone Project Subject of an Electronic Engineering Program," in *2024 IEEE World Engineering Education Conference (EDUNINE)*, Mar. 2024, pp. 1–6, doi: 10.1109/EDUNINE60625.2024.10500553.

[12] P. Sukkeewan, N. Songkram, and J. Nasongkhla, "Development and Validation of a Reliable and Valid Assessment Tool for Measuring Innovative Thinking in Vocational Students," *International Journal of Educational Methodology*, vol. 10, no. 1, pp. 35–44, Feb. 2024, doi: 10.12973/ijem.10.1.835.

[13] M. A. Adeoye and H. A. Jimoh, "Problem-Solving Skills Among 21st-Century Learners Toward Creativity and Innovation Ideas," *Thinking Skills and Creativity Journal*, vol. 6, no. 1, pp. 52–58, Jun. 2023, doi: 10.23887/tscj.v6i1.62708.

[14] M. S. B. Yusoff, "ABC of Content Validation and Content Validity Index Calculation," *Education in Medicine Journal*, vol. 11, no. 2, pp. 49–54, Jun. 2019, doi: 10.21315/eimj2019.11.2.6.

[15] S. Amirzadeh, D. Rasouli, and H. Dargahi, "Assessment of validity and reliability of the feedback quality instrument," *BMC Research Notes*, vol. 17, no. 1, p. 227, Aug. 2024, doi: 10.1186/s13104-024-06881-x.

[16] T. G. de A. Ferraz and C. de S. Pereira-Guizzo, "Developing a system for assessing engineering students' transferable skills: evidence for the content validity and replicability of the scales," *European Journal of Engineering Education*, vol. 48, no. 4, pp. 667–681, Jul. 2023, doi: 10.1080/03043797.2023.2195809.

[17] L. L. Davis, "Instrument review: Getting the most from a panel of experts," *Applied Nursing Research*, vol. 5, no. 4, pp. 194–197, Nov. 1992, doi: 10.1016/S0897-1897(05)80008-4.

[18] N. A. M. Said, S. M. Bujang, N. A. Buang, H. H. Siraj@Ramli, and M. N. A. Besar, "Critical Thinking Transfer Practice Instrument: A Content Validity Calculation Steps Based on Expert Panel Evaluation," *Education in Medicine Journal*, vol. 14, no. 3, pp. 61–74, Sep. 2022, doi: 10.21315/eimj2022.14.3.5.

[19] S. Handage and M. Chander, "Development of An Instrument for Measuring the Student Learning Outcomes: A Content Validation Process," *Indian Journal of Extension Education*, vol. 57, no. 3, pp. 1–7, 2021, doi: 10.48165/IJEE.2021.57302.

[20] M. R. Lynn, "Determination and Quantification of Content Validity," *Nursing Research*, vol. 35, no. 6, pp. 382–386, Nov. 1986, doi: 10.1097/00006199-198611000-00017.

[21] S. S. Guraya *et al.*, "Validating the Medical Education e-Professionalism Framework Using the Content Validity Index," *Education in Medicine Journal*, vol. 14, no. 3, pp. 31–47, Sep. 2022, doi: 10.21315/eimj2022.14.3.3.

[22] L. Sürücü and A. Maslakci, "Validity and reliability in quantitative research," *Business & Management Studies: An International Journal*, vol. 8, no. 3, pp. 2694–2726, Sep. 2020, doi: 10.15295/bmij.v8i3.1540.

[23] K. Beck, "Ensuring content validity of psychological and educational tests – the role of experts," *Frontline Learning Research*, vol. 8, no. 6, pp. 1–37, Sep. 2020, doi: 10.14786/flr.v8i6.517.

[24] J. Chong, S. E. Mokshein, and R. Mustapha, "A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL)," *International Journal of Academic Research in Business and Social Sciences*, vol. 11, no. 4, pp. 868–883, Apr. 2021, doi: 10.6007/IJARBSS/v11-i4/9547.

[25] C. B. Terwee *et al.*, *COSMIN methodology for assessing the content validity of PROMs–user manual*. Amsterdam: VU University Medical Center, 2018.

[26] E. Belita, J. Yost, J. E. Squires, R. Ganann, and M. Dobbins, "Development and content validation of a measure to assess evidence-informed decision-making competence in public health nursing," *PLoS One*, vol. 16, no. 3, p. e0248330, Mar. 2021, doi: 10.1371/journal.pone.0248330.

[27] A. Jemal, S. R. Bussey, and K. Davis, "Transformative Consciousness: Scale development and content validity," *Critical Social Work*, vol. 21, no. 1, pp. 30–53, Apr. 2020, doi: 10.22329/csw.v21i1.6226.

[28] A. Spoto, M. Nucci, E. Prunetti, and M. Vicovaro, "Improving content validity evaluation of assessment instruments through formal content validity analysis.," *Psychological Methods*, vol. 30, no. 2, pp. 203–222, Apr. 2025, doi: 10.1037/met0000545.

[29] K. Shively, K. M. Stith, and L. D. Rubenstein, "Measuring What Matters: Assessing Creativity, Critical Thinking, and the Design Process," *Gifted Child Today*, vol. 41, no. 3, pp. 149–158, Jul. 2018, doi: 10.1177/1076217518768361.

[30] Y. M. Sameer, "Innovative Behavior and Psychological Capital: Does Positivity Make any Difference?" *Journal of Economics and Management*, vol. 32, pp. 75–101, 2018, doi: 10.22367/jem.2018.32.06.

[31] R. Bhatnagar, C. Tanguay, C. Sullivan, and J. Many, "Observation of Field Practice Rubric: Establishing Content Validity and Reliability," *Georgia Educational Researcher*, vol. 18, no. 2, pp. 1–23, Jul. 2021, doi: 10.20429/ger.2021.180201.

[32] E. Gregori-Giralt and J.-L. Menéndez-Varela, "The content aspect of validity in a rubric-based assessment system for course syllabuses," *Studies in Educational Evaluation*, vol. 68, p. 100971, Mar. 2021, doi: 10.1016/j.stueduc.2020.100971.

[33] K. Nithia, F. D. Yusop, and C. Y. Piaw, "Development and Validation of an Authentic-Based Competency Assessment Rubric for Secondary School Multimedia Production Subject," *International Journal of Education and Pedagogy (IJEAP)*, vol. 2, no. 3, pp. 112–121, 2020.

[34] I. Dalawi, M. R. Isa, X. W. Chen, Z. I. Azhar, and N. Aimran, "Development of the Malay Language of understanding, attitude, practice and health literacy questionnaire on COVID-19 (MUAPHQ C-19): content validity & face validity analysis," *BMC Public Health*, vol. 23, no. 1, p. 1131, Jun. 2023, doi: 10.1186/s12889-023-16044-5.

[35] J. Shi, X. Mo, and Z. Sun, "Content validity index in scale development," *Journal of Central South University (Medical Sciences)*, vol. 37, no. 2, pp. 152–155, 2012, doi: 10.3969/j.issn.1672-7347.2012.02.007.

[36] D. F. Polit, C. T. Beck, and S. V. Owen, "Is the CVI an acceptable indicator of content validity? Appraisal and recommendations," *Research in Nursing & Health*, vol. 30, no. 4, pp. 459–467, Aug. 2007, doi: 10.1002/nur.20199.

[37] V. K. Shrotryia and U. Dhanda, "Content Validity of Assessment Instrument for Employee Engagement," *Sage Open*, vol. 9, no. 1, pp. 1–7, Jan. 2019, doi: 10.1177/2158244018821751.

[38] V. Zamanzadeh, A. Ghahramanian, M. Rassouli, A. Abbaszadeh, H. Alavi-Majd, and A.-R. Nikanfar, "Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication," *Journal of Caring Sciences*, vol. 4, no. 2, pp. 165–178, Jun. 2015, doi: 10.15171/jcs.2015.017.

[39] C. A. Wynd, B. Schmidt, and M. A. Schaefer, "Two Quantitative Approaches for Estimating Content Validity," *Western Journal of Nursing Research*, vol. 25, no. 5, pp. 508–518, Aug. 2003, doi: 10.1177/0193945903252998.

[40] E. Almanasreh, R. Moles, and T. F. Chen, "Evaluation of methods used for estimating content validity," *Research in Social and Administrative Pharmacy*, vol. 15, no. 2, pp. 214–221, Feb. 2019, doi: 10.1016/j.sapharm.2018.03.066.

[41] F. Madadizadeh and S. Bahariniya, "Tutorial on how to calculating content validity of scales in medical research," *Perioperative Care and Operating Room Management*, vol. 31, p. 100315, Jun. 2023, doi: 10.1016/j.pcorm.2023.100315.

[42] M. Salokivi, S. Salanterä, S. Saarikallio, and E. Ala-Ruona, "Validation of item pool for early adolescents' emotional skills assessment in music therapy," *Nordic Journal of Music Therapy*, vol. 33, no. 2, pp. 108–121, Mar. 2024, doi: 10.1080/08098131.2023.2245433.

[43] A. Masuwai, H. Zulkifli, and M. I. Hamzah, "Evaluation of content validity and face validity of secondary school Islamic education teacher self-assessment instrument," *Cogent Education*, vol. 11, no. 1, 2024, doi: 10.1080/2331186X.2024.2308410.

[44] A. Yousef and A. Ayyoub, "Rubric development and validation for assessing educational robotics skills," in *Frontiers in Education*, Dec. 2024, vol. 9, p. 1496242, doi: 10.3389/feduc.2024.1496242.

[45] F. Wang and S. Sahid, "Content validation and content validity index calculation for entrepreneurial behavior instruments among vocational college students in China," *Multidisciplinary Reviews*, vol. 7, no. 9, May 2024, doi: 10.31893/multirev.2024187.

[46] I. J. Chikezie and J. U. Eme, "Content Validity Index of Basic Education Certificate Examination (BECE) Test Items in Basic Science and Technology (BST)," *British Journal of Education*, vol. 11, no. 6, pp. 7–24, May 2023, doi: 10.37745/bje.2013/vol11n6724.

[47] V. W. S. Yun, N. M. Ulang, and S. H. Husain, "Enhancing Content Validity of the Designed Hierarchy of Controls Instrument in Preventing Infectious Diseases on Construction Sites: A Multidimensional Approach," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 32, no. 2, pp. 508–521, Sep. 2023, doi: 10.37934/araset.32.2.508521.

[48] V. Swami and D. Barron, "Translation and validation of body image instruments: Challenges, good practice guidelines, and reporting recommendations for test adaptation," *Body Image*, vol. 31, pp. 204–220, 2019, doi: 10.1016/j.bodyim.2018.08.014.

[49] U. Lorenzo-Seva and P. J. Ferrando, "MSA: The forgotten index for identifying inappropriate items before computing exploratory item factor analysis," *Methodology*, vol. 17, no. 4, pp. 296–306, Dec. 2021, doi: 10.5964/meth.7185.

[50] S. C. Izah, L. Sylva, and M. Hait, "Cronbach's Alpha: A Cornerstone in Ensuring Reliability and Validity in Environmental Health Assessment," *ES Energy & Environment*, vol. 23, p. 1057, 2023, doi: 10.30919/esee1057.

[51] L. S. Lambert and D. A. Newman, "Construct Development and Validation in Three Practical Steps: Recommendations for Reviewers, Editors, and Authors," *Organizational Research Methods*, vol. 26, no. 4, pp. 574–607, Oct. 2023, doi: 10.1177/10944281221115374.

[52] M. H. A. Rahman, S. Sharanjeet-Kaur, S. Narayanasamy, H. F. M. Rasdi, and M.-L. C. Bastion, "Development of Knowledge, Attitude and Practice Questionnaire for age-related macular degeneration patients," *International Journal of Ophthalmology*, vol. 16, no. 4, pp. 589–600, Apr. 2023, doi: 10.18240/ijo.2023.04.13.

[53] M. K. Mace and D. Pearl, "Rubric Development and Validation for Assessing Comprehensive Internationalization in Higher Education," *Journal of Studies in International Education*, vol. 25, no. 1, pp. 51–65, Feb. 2021, doi: 10.1177/1028315319865790.

[54] L. Giromini, G. Young, and M. Sellbom, "Assessing Negative Response Bias Using Self-Report Measures: New Articles, New Issues," *Psychological Injury and Law*, vol. 15, no. 1, pp. 1–21, Mar. 2022, doi: 10.1007/s12207-022-09444-2.

[55] D. Tempelaar, B. Rienties, and Q. Nguyen, "Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application," *PLoS One*, vol. 15, no. 6, p. e0233977, Jun. 2020, doi: 10.1371/journal.pone.0233977.

[56] J. R. Cabigao, "Development and Validation of a Proposed Model Rubric in Rating Written Outputs in the Graduate School Level," *International Journal of Academic Multidisciplinary Research (IJAMR)*, vol. 5, no. 4, pp. 122–131, 2021.

[57] S. R. M. Yusop, M. S. Rasul, R. M. Yasin, H. U. Hashim, and N. A. Jalaludin, "An Assessment Approaches and Learning Outcomes in Technical and Vocational Education: A Systematic Review Using PRISMA," *Sustainability*, vol. 14, no. 9, p. 5225, Apr. 2022, doi: 10.3390/su14095225.
[58] S. M. Brookhart, T. Loureiro, R. T. Quevedo-Camargo "Using rubrics in basic education: A review and recommendations," (in Portuguese), *Studies in Educational Assessment*, vol. 35, p. 10803, 2024.
[59] D. S. Anggara and C. Abdillah, "Content validity analysis of literacy assessment instruments," *Jurnal Cakrawala Pendidikan*, vol. 42, no. 2, pp. 447–459, May 2023, doi: 10.21831/cp.v42i2.55900.

BIOGRAPHIES OF AUTHORS

Nor Aisyah Che Derasid is a Ph.D. candidate, Department of Advanced Technical and Vocational Education and Training at Universiti Teknologi Malaysia. Her research focuses on exploring and enhancing innovative behavior and innovation skills among TVET engineering students. She can be contacted at email: anor5@graduate.utm.my.

Aede Hatib Musta'amal @ Jamal is an associate professor at the Faculty of Educational Sciences and Technology, Universiti Teknologi Malaysia (UTM). Currently, he is deputy dean (research, innovation, and development) in faculty. He holds a Ph.D. in design and technology from Loughborough University. His research focuses on creativity and innovation in technical education. His research interests include technical and vocational education, creativity and innovation, design and technology education, green competencies, and technology in education. Dr. Aede has received multiple Excellence Service Awards from UTM and the faculty. He has also been recognized with teaching awards for his contributions to active blended learning. He can be contacted at email: aede@utm.my.

Nornazira Suhairom is an associate professor at Universiti Teknologi Malaysia (UTM), specializing in technical and vocational education and training (TVET). She completed her Ph.D. in technical and vocational education at UTM in 2016. Her research focuses on competency development, career progression in hospitality, and innovative teaching methods. She has authored numerous publications with over hundreds of citations, reflecting her influence in the field. Dr. Nornazira has received the Academic Excellence Award and has supervised multiple Ph.D. candidates, contributing significantly to the academic community. She can be contacted at email: p-nazira@utm.my.

Hary Suswanto is a lecturer at Universitas Negeri Malang, Indonesia, specializing in technical and vocational education and training (TVET). He holds a doctorate in engineering and currently serves as the secretary of the Institute for Research and Community Service. His research interests include educational technology, adaptive learning, gamification, vocational education, digital innovation in learning environments, and educational assessment. He has also been actively involved in technology integration for education and community empowerment. From 2020 to 2025, Dr. Suswanto secured 10 registered intellectual property rights (HaKI), completed more than 30 funded research projects, and led more than 19 community service initiatives supported by national and institutional grants. He published more than 6 scientific articles in national accredited journals and internationally reputable journals. His contributions include the development of mobile learning applications, smart adaptive educational systems, and innovative community-based digital platforms. He can be contacted at email: hary.suswanto.ft@um.ac.id.