International Journal of Organizational Leadership 14(First Special Issue-2025)689-704





Content Validity Assessment of Malaysian Teachers' Professional Judgement Instrument

Adila Athirah Abdul Hadi¹, Mohd Effendi Ewan Mohd Matore^{2*}

¹Faculty of Education, Universiti Kebangsaan Malaysia (UKM), Selangor, Malaysia ²Research Centre of Education Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia, Selangor, Malaysia

ABSTRACT

Keywords:

Classroom-based assessment, Content Validity Ratio, Professional judgement, Teacher

Received 29 April 2025 Received in revised form 20 May 2025 Accepted 23 May 2025

*Correspondence: effendi@ukm.edu.my Professional judgement needs to be practiced by teachers in providing a holistic approach to assessing their students. However, the validation of the items to measure teachers' professional judgement is always questionable. Hence, this study examines the content validity of an instrument developed to assess professional judgement, a skill that needs to be mastered by Malaysian teachers, in parallel with the implementation of classroom-based assessment. Eight experts, consisting of four professionals and four lay experts, were selected to be involved in this study, which used the Content Validity Ratio (CVR) method. Only experts with certain criteria were being selected, such as experience, expertise and relevance to this study. The instrument consists of five constructs; each represents one of the elements of professional judgement respectively. Starting with teachers' knowledge, followed by teaching experience, student's input, teachers' professional responsibility and teachers' intuition. A total of 102 items were assessed and 93 items met the critical CVR threshold of 0.75. This indicates strong agreement among experts regarding their suitability. The other nine items that did not meet this criterion were revised in terms of clarity, relevance and agreement with the theoretical constructs. This study makes a novel contribution by developing a culturally responsive instrument that is tailored to the Malaysian classroom assessment system and fills a significant gap in the measurement of teachers' professional judgement in local education policy and practice.

©CIKD Publishing

Validating the content of a newly developed instrument is a fundamental step to be taken, especially if we want it to be considered as a high-quality instrument. It is important to undergo this process to ensure the constructs of the instrument adequately reflect what they are intended

to measure as it helps us to evaluate the definition of the dimensions and elements of a concept (Aguirre et al., 2024). Content validity relies on skills in a specific area, which is meaningful in developing a reliable instrument (Yaghmaie, 2003). Although content validity is subjective in nature, the validity still can be assured by making it more objective, using structured evaluation frameworks and having expert consensus (Rubio et al., 2003).

In educational measurement, validity refers to whether an instrument measures what it is supposed to measure and in what context, which includes both the theoretical and evidential support for interpretations of the instrument in the uses for which it is used (Aguirre et al., 2024). However, an instrument validated for a certain use or population does not necessarily work in a different context (Knekta et al., 2019). Content validation, or, for instance, consensus of experts, ensures that the items of the instrument correspond to the construct that is being measured. One way of implementing this is through a Content Validity Ratio (CVR) that involves experts to review and input the quality of the instrument (Mohd Effendi Ewan et al., 2021). Likert-scale questionnaires are a popular tool for social sciences despite issues such as a mismatch between questions and answers, highlighting the need for content validation to fill these potential gaps (Pozzo et al., 2019).

As an instrument developed to measure complex constructs, such as professional judgement, content validation is required to ensure that items reflect all aspects of the intended construct. Professional judgement, especially in an educational context, is highly needed nowadays as part of student potential development in schools. This involves a human-centered holistic approach covering physical, intellectual, emotional and spiritual aspect of an individual (Mohd Effendi Ewan et al., 2018). This high-quality education is definitely not an option nowadays in preparing our future generation to face volatility, uncertainty, complexity and ambiguity (VUCA) in the fourth industrial revolution (IR4) phase (Melor et al., 2021).

The process of developing an instrument aligned with current needs includes bringing together the results of the review of literature, the perceptions of the representative populations, and the views of expert panels (Mohd Effendi Ewan et al., 2021). The views of experts are subjective but important as they are the link between the construct of interest and the actual representation of that construct with the items (Aguirre et al., 2024). Although the items can be refined without involving experts to assess them, bias cannot be eliminated, and the items may not be able to reflect the actual construct of interest (Creswell, 2014). Therefore, content validation is important for studies of teachers' professional judgment to ensure the soundness of the measurement of this important competency.

In the field of social science research and education specifically, the most widely used instrument is a questionnaire with a Likert scale (Memmedova & Ertuna, 2024). However, few instruments exist to measure teachers' professional judgment in Malaysia, as existing instruments often measure different constructs or populations. Therefore, this study develops an instrument to measure Malaysian teachers' professional judgment, which is an important skill if teachers are to make ethical and context-sensitive decisions in the classroom.

To ensure that the instrument has good quality, this study uses the Content Validity Ratio (CVR) method, where experts are used to validate the instrument such that it accurately reflects the construct of professional judgment. While the CVR method is a well-established technique used to validate content, this study offers a unique contextual adaptation specific to Malaysia's educational system. The development of the instrument was based on the Malaysian policy

framework on teaching assessment and was guided by experts from various institutions, including the Teacher Training Institutes, Aminuddin Baki Institute, which specialized in educational leadership, and the Curriculum Departments of the Ministry of Education. The construct of professional judgement in this study goes beyond conventional dimensions by incorporating culturally significant elements such as intuition and ethical responsibility, which reflect holistic educational values rooted in Malaysian pedagogy. This approach puts the instrument not as a direct replication of existing instruments but as a tailored innovation that responds to local needs and contributes to the under-researched discourse on educational assessment in Southeast Asia. This study aims to answer the following research questions:

1. What is the content validity of the developed instrument in measuring Malaysian teachers' professional judgment based on expert evaluation?

2. Which items meet the minimum Content Validity Ratio (CVR) threshold of .75 and are deemed suitable for inclusion?

Literature Review

The Professional Judgment in Teaching

Professional judgment is one of the main components of teaching. It includes the ability to make decisions based on context, subject and pedagogical knowledge. In recent research, professional judgement is seen as an interdisciplinary and context-dependent construct. An ecological model has been proposed to explain how psychological, emotional and social factors interact in professional judgement, particularly in fields such as education, social work and medicine (Helm & Roesch-Marsh, 2017). Situational Judgement Tests (SJTs), increasingly used in assessments, highlight the influence of cultural background and personal belief systems on judgement decisions, emphasizing the importance of reflective practice in a professional context (Iqbal et al., 2025).

In practicing professional judgment, teachers need to face classroom complexity, curriculum and student assessment. It is the process of making decisions on the information that is interpreted, reflected and applied to make sense of the information within the experience. It is viewed by Coles (2002) as a process of judgment based on practical wisdom developed through reflective practice and discussion with colleagues. This judgement requires a balance between theoretical knowledge and situational adaptability to decide the most appropriate pedagogical context. In other words, professional judgment reflects the estimation of what is most appropriate to do in different pedagogical contexts through a balance between theoretical knowledge and classroom situation. Respecting and valuing the professional judgement of teachers is crucial to their professional standing and the overall quality of education (Wharton, 2022).

In the field of professional development, professional judgment is formed through training, experience, and the deliberate application of knowledge (Winch, 2022). This shows that professional judgment is not a basic ability but one that is formed over time. The two studies also maintain that professional judgment requires cognitive and reflective skills that are related to experience-based judgment in dealing with specific pedagogical problems.

On a theoretical level, professional judgment is the framework that connects the abstract to the situation. This kind of judgment is a theoretical tool for changing abstract knowledge into concrete knowledge, especially in non-academic areas where contextual demands are highly needed, such as in vocational training (Winch, 2022). Professional judgment must go beyond the application of protocols, as teaching often involves solving dilemmas for which there are no clear-cut solutions (Coles, 2002). These theoretical perceptions tell us that it is highly needed to integrate reflective practice and contextual understanding into teacher education and professional development.

The Gaps Between Professional Judgment and Standardized Assessment Practices

Standardized assessment is based on a set protocol and predetermined criteria of implementation to ensure consistency and comparability. In contrast, professional judgment is dynamic and considers the complexity of the classroom and individual learners. Allal (2013) illustrates this difference in summative assessment, where teachers combine test results with observations to capture a picture of student performance. Unlike standardized applications that focus on uniformity, professional judgment allows teachers to tailor their decisions to local conditions, ensuring that assessments are fair and relevant in different educational contexts.

Professional judgment also involves interpreting data when making high-stakes decisions. Teachers often rely on their intuition to make sense of complex information, even though the data-driven decision-making process did provide valuable insights (Vanlommel & Schildkamp, 2019). This dual reliance on logical data interpretation and self-instinct further distinguishes professional judgment from the standardized assessment pathway.

Professional Judgement in Malaysian Context

In Malaysia, the concept of professional judgment is of great importance due to the country's evolving education policy and the move towards more autonomous, classroom-based assessments. The Malaysian education system is increasingly emphasizing the role of teachers' professional judgment, especially in response to the shift of assessment to the decentralized classrooms. Adila Athirah and Mohd Effendi Ewan (2024) identify key challenges in this transition, including limited assessment literacy, insufficient time for reflective practice and a lack of robust decision-making frameworks. These challenges highlight the need to strengthen professional judgment through targeted training and policy reform.

Another dimension of professional judgment in Malaysia is focusing on how gender and class size affect student outcomes (Surianshah, 2022). For example, smaller class sizes benefit higher performing female students more than their male classmates, suggesting that teachers need to consider gender differences in performance when designing instruction and assessment (Surianshah, 2022). This complexity stresses the importance of situational adaptability and reflective practices in professional judgment.

However, to date, there is limited study assessing teachers' quality, especially in terms of their professional judgement practice. Teachers' professional judgment is highly needed in ensuring high quality education is provided to all stakeholders, as it reflects the quality of an educator. While various assessment tools like Six Sigma have been implemented in educational settings, particularly in countries like India (Siti Hannah & Mohd Effendi Ewan, 2024), Malaysia currently lacks specific instruments to measure this professional practice among

teachers. This gap in assessing teachers requires a valid and reliable tool to help evaluate teachers' professional judgement in the context of the Malaysian education system.

Method

The study adopts a quantitative approach using a questionnaire. It involves two types of experts which are the professional experts and lay experts, and they are known as Content Evaluation Panel (Lawshe, 1975). Eight subject matter experts have been selected to answer the questionnaire based on their expertise in the subject being studied, as it aims to obtain the necessary information from knowledgeable individuals (Aguirre et al., 2024). They are considered to be suitable for this study as the expert panel is chosen based on specific criteria to validate the content of items related to teachers' professional judgment.

Professional experts are those who are professionally involved or work in the field being studied. After being selected based on stringent criteria, they helped to determine whether the developed measurements were well-constructed for psychometric testing (Aguirre et al., 2024). They have been chosen based on their field of expertise, number of publications, or work experience (Rubio et al., 2003). In this study, professional experts are lecturers specializing in education and teacher training, and officers in the Ministry of Education with a doctorate qualification. This selection process aligns with the concept of content validity, which states that individuals working in the relevant field can act as experts or judges to provide ratings on the instrument's content and evaluate whether the items can represent the sample being studied (Mohd Effendi Ewan et al., 2017). As for this study, four professional experts were involved in the validation process.

Two out of four were from the Malaysia Teacher Training Institute and are the lecturers for student teachers. The other two were experts from the Aminuddin Baki Institute and the Division of Sports, Co-Curriculum and Arts, Malaysia Ministry of Education, respectively. Experts from the Malaysia Teacher Training Institute were chosen for their expertise in teacher education and pedagogy. They ensure the instrument aligns with the practical realities of teaching and the competencies required in the Malaysian education context. An expert from Aminuddin Baki Institute was selected for their expertise in education responsible for designing and formulating assessment instruments for the needs of educational leaders to achieve set quality objectives and standards. Choosing an expert from this institute ensures the instrument uses clear, appropriate terminology aligned with Malaysian educational standards, making it easily understood by teachers. An expert from the Division of Sports, Co-Curriculum, and Arts, Malaysia Ministry of Education, ensures the instrument aligns with current policies and standards for co-curricular activities, validating its relevance to teachers' professional judgment.

The selection criteria for the professional expert panel include, (a) holding a Doctor of Philosophy (PhD) degree in the educational field, (b) having more than ten years of experience as an officer in the ministry or teaching under MoE institutions, (c) currently serving an institute under MoE such as teachers training college and educational leaders' training institute (d) actively involved in writing articles on mainstream media platform, giving expert opinions publicly, or teaching future teachers, and (e) provided their consent to participate in the study. As for lay experts, four of them are involved in this study. Lay experts are indeed very relevant

to the topics being measured, as they represent the population to be studied, the teachers. The distinction between the selection of lay and professional experts was intentional to ensure a balanced validation process. Professional experts were selected based on their academic and institutional authority to ensure consistency with policy and pedagogical standards, while lay experts, in this study, teachers with more than 10 years of experience, were included to verify the practical relevance and applicability of the instrument. This combination ensures both theoretical integrity and practical applicability of the instrument in the Malaysian educational context.

This study utilizes online methods (email or online form), which have their own strengths. The researcher also contacted the expert panel through phone calls, letters, and emails to explain the study's objectives and procedures and seek permission for their involvement. The expert panel was given two to four weeks to complete their evaluations. Official appointment letters for the expert panel were issued by the Faculty of Education, Universiti Kebangsaan Malaysia. Table 1 provides the detailed information of the respective experts.

Table 1

Details of Experts

| No | Panel Experts | Position | Institution |
|----|---------------|----------|--|
| 1 | A1 | Officer | Division of Sports, Co-Curriculum and Arts, Malaysia Ministry of Education |
| 2 | A2 | Lecturer | Teacher Training Institute |
| 3 | A3 | Lecturer | Teacher Training Institute |
| 4 | A4 | Officer | Aminuddin Baki Institute |
| 5 | A5 | Teacher | Secondary School |
| 0 | A6 | Teacher | Secondary School |
| 7 | A7 | Teacher | Secondary School |
| 8 | A8 | Teacher | Secondary School |

Content Validity Ratio (CVR) Metrics of Measurement

The Content Validity Ratio (CVR) is a metric used to assess the content validity of items in a survey or instrument. This study utilised this method to evaluate its content validity by analysing expert consensus. In addition to CVR, the inter-rater reliability between two experts was analysed using Cohen's Kappa to assess the consistency of their evaluations. The analysis yielded a Cohen's Kappa value of .49, indicating moderate agreement between the two raters based on the interpretation of Landis and Koch (1977). This additional metric supports the reliability of the content validation process by quantifying the level of agreement between the experts beyond chance agreement. The moderate agreement between raters suggested that the evaluation process reflected a reasonable level of consistency in expert judgement.

The CVR method requires experts to assess whether each item is suitable and accurate for measuring the construct of interest (Lawshe, 1975; Mohd Effendi Ewan et al., 2021). It is a widely recognized quantitative approach for determining content validity through empirical evaluation and is commonly used in questionnaire validation due to its practicality, cost-effectiveness, and ease of administration (Angraini et al., 2021).

In this study, the CVR analysis is conducted to evaluate the content validity of the questionnaire, with critical CVR values determined based on the number of experts, as suggested by Lawshe (1975). Each expert panel is required to state their level of agreement in validating each item using a 4-point Likert scale, which represents the suitability of the item. Table 2 provides the detailed information on the scale used.

Table 2Level of Expert Agreement on the Item Measured

| Scale | Level of Agreement |
|-------|---------------------|
| 1 | Not Suitable at All |
| 2 | Not Suitable |
| 3 | Suitable |
| 4 | Very Suitable |

The CVR method ensures the selection of the most critical and accurate items for an instrument (Zamanzadeh et al., 2015) and can also be applied to assess content validity. The CVR score, which ranges from -1 to 1, reflects the level of agreement among experts, with higher scores indicating greater consensus on the importance of an item. The CVR is calculated using the formula below; ...(1)

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

where

ne = the number of experts rating an item as 3 and 4 N = total number of experts.

These scores are then compared to critical values derived from Lawshe's table. CVR values range from -1 to +1 (Lawshe, 1975). A CVR value of +1 indicates that all experts agree the evaluated element is suitable or very suitable for the content's validity. A CVR value of less than 0 means less than half of the expert panel rated the item as suitable or very suitable, while a CVR value of 0 indicates that the expert panel was divided, with some rated the item as not suitable and others as suitable. A CVR value of more than 0 implies that more than half of the experts rated the item as 3 and 4. Specifically, a CVR value of 1 reflects unanimous agreement among the experts that the item is suitable or very suitable. Based on feedback from the panel of eight experts, the critical CVR value for this study is set at .75, as suggested by Lawshe (1975) for an expert panel of eight members.

Instrument

The instrument developed contains five constructs measuring the elements of teachers' professional judgement: (1) Knowledge; (2) Experience; (3) Students' Input; (4) Professional Responsibility; (5) Intuition.

Knowledge refers to professional judgement based on a deep understanding of the subject matter, teaching strategies and the ability to assess students fairly. Teachers gather information from a variety of sources and use this knowledge to evaluate tasks and make informed decisions (Bahagian Pembangunan Kurikulum, 2019). Experience shows how teachers rely on their teaching experience to identify strengths and potential when interacting with students, enabling them to make spontaneous decisions without analysing new data (Haidi et al., 2020; Vanlommel & Pepermans, 2021). Their prior experience and discussions increase accuracy and confidence in their professional judgement (Allal, 2013).

Student Input emphasises the importance of analysing and evaluating students through observation, two-way discussions and various sources such as worksheets, portfolios, logbooks, tests and classroom performance (Bahagian Pembangunan Kurikulum, 2019). This information, combined with observations of student behaviour, enables teachers to make informed judgements in real time. Professional responsibility focuses on teachers conducting assessments responsibly, ethically, efficiently, transparently and holistically, incorporating their knowledge, student input and experience. These judgements are informed by the practises of the setting, the expectations of the community and the teachers' ability to address specific student issues, such as special talents, learning difficulties or mental health issues, while maintaining high ethical standards. Finally, intuition refers to professional judgement based on the recognition of patterns stored in memory (Vanlommel & Pepermans, 2021). Teachers use their intuition effortlessly and automatically, without much conscious effort, to make decisions (Ben Knight, 2023; Kahneman, 2011).

Results and Discussion

Based on this study, the objective is to examine the content validity of an instrument developed to assess professional judgement. Table 3 shows the number of constructs and items developed to assess teachers' professional judgement.

Table 3

Constructs and Number of Items of Professional Judgement Instrument

| | 5 | 5 | 5 | 0 | |
|-----------------------------|---|---|---|---|----------------|
| Construct | | | | | Number of Item |
| Knowledge | | | | | 18 |
| Experience | | | | | 15 |
| Students Input | | | | | 30 |
| Professional responsibility | | | | | 21 |
| Intuition | | | | | 17 |

To measure the content validity of this instrument, the CVR values for 102 items in the professional judgment survey for teachers in Malaysia were calculated. Overall, the results indicate that three items had a CVR value of .25, six items scored .50, 23 items achieved .75, and 70 items obtained a perfect CVR value of 1.00.

As explained in section 3.1, the critical value of CVR is .75 as there were eight experts. Therefore, 93 items were accepted while the remaining nine items should be revised as they did not reach the critical value. A systematic item revision strategy was used to refine the items that did not meet the CVR threshold. This involved analysing the panel's feedback to identify problems related to the vagueness of items in avoiding being open to more than one interpretation for each item, contextual misalignment or irrelevance of the construct. Items were rewritten to enhance clarity, achieve better alignment with the theoretical definitions of professional judgement, and reflect real-world classroom scenarios. Particular attention was given to simplifying technical terms, rephrasing vague sentences and aligning response options with Malaysian teaching assessment standards. These improvements aim to upgrade the validity of the items so that they can be retained in the instrument.

Based on Lawshe (1975)'s concept, the CVR values of the items in this study were analysed to determine their alignment with the content validity requirements. In this study, the critical CVR value for an expert panel of eight members is set at .75, as suggested by Lawshe (1975).

This critical threshold ensures that items rated as "essential" or "very suitable" by at least 75% of the panel are retained. Items with CVR values of .75 or higher are considered valid and accepted without revision, as they demonstrate strong agreement among experts regarding their suitability.

The CVR values recorded vary from .25, .5, .75 to 1. There are items that recorded CVR value as .25 and .5. This means that less than 75% of the panel decided the item as suitable or very suitable. In other words, the items have a lower agreement. Thus, the items should be revised to match the construct. Based on Table 3, CVR values below .75 from the construct knowledge and student input must be revised in terms of clarity or relevance to the content. This will increase the expert agreement score in the next revision.

The high CVR values for most items, especially for those rated 1, reflect the unanimous agreement of the experts and show that the content of the instrument is largely consistent with the intended constructs. However, those items that need to be revised underscore the step-by-step process of instrument development needed. The opinion of experts is needed to rework the items' wording to reinforce the overall validity of the instrument. This systematic approach strengthens the instrument's ability to accurately measure teachers' professional judgment. From the knowledge construct, as in Table 4, most items have a high CVR value, and they are accepted without further modification. However, Items 5 and 15, with a CVR value of .5, are considered good and need to be revised. A possible explanation for these results may be due to the wording of these items, which lacks clarity and cannot explain the construct correctly. Some experts suggested the items should be rephrased, and the response choices should be aligned with the theoretical knowledge.

| No. | Construct | Item | CVR Value | Item Status |
|-----|-----------|------|-----------|-------------|
| 1 | Knowledge | 1 | 1 | Accepted |
| | | 2 | 1 | Accepted |
| | | 3 | 1 | Accepted |
| | | 4 | 1 | Accepted |
| | | 5 | .5 | Revised |
| | | 6 | 1 | Accepted |
| | | 7 | .75 | Accepted |
| | | 8 | 1 | Accepted |
| | | 9 | 1 | Accepted |
| | | 10 | 1 | Accepted |
| | | 11 | 1 | Accepted |
| | | 12 | 1 | Accepted |
| | | 13 | .75 | Accepted |
| | | 14 | .75 | Accepted |
| | | 15 | .5 | Revised |
| | | 16 | 1 | Accepted |
| | | 17 | 1 | Accepted |
| | | 18 | 1 | Accepted |

Table 4

Construct 1 CVR Values and Item Status

For Construct 2, the experience construct, as presented in Table 5, most of the construct items are agreed by the experts, which is evident in the CVR values of the items. Items 23, 24, 26 and 30 are considered good with a CVR value of .75 and can be revised slightly. Possible explanations for this result may be that certain response options were inappropriately put into context and contributed to redundancy. However, it is worth highlighting this construct as

teachers' experience is a very important factor in influencing teachers' judgment. Results have been shown in the previous study that the rating quality among teachers was different, depending on teachers' rating experience and teaching experience (Muhamad Firdaus & Mohd Effendi Ewan, 2022).

| No. | Construct | Item | CVR Value | Item Status |
|-----|------------|------|-----------|-------------|
| 2 | Experience | 19 | 1 | Accepted |
| | | 20 | 1 | Accepted |
| | | 21 | 1 | Accepted |
| | | 22 | 1 | Accepted |
| | | 23 | .75 | Accepted |
| | | 24 | .75 | Accepted |
| | | 25 | 1 | Accepted |
| | | 26 | .75 | Accepted |
| | | 27 | 1 | Accepted |
| | | 28 | 1 | Accepted |
| | | 29 | 1 | Accepted |
| | | 30 | .75 | Accepted |
| | | 31 | 1 | Accepted |
| | | 32 | 1 | Accepted |
| | | 33 | 1 | Accepted |

 Table 5

 Construct 2 CVR Values and Item Status

However, the items in the student input construct presented in Table 6 have a wide range of CVR values and need to be revised. Items 38, 40, 43, 52, and 55 are not good with CVR values below .75, and the items are revised. Experts agreed the response choices are unclear and not in the context of student input. This is probably due to the number of items in this construct, which is twice as many as in other constructs. Experts suggest using more descriptive and clear words, and the response choices should reflect the real situation in the classroom so that the number of items can be reduced to help respondents answer them better. For instance, one suggestion involved replacing generic terms with more specific descriptors to enhance relevance.

As for Construct 4 in Table 7, most items in the professional responsibility construct have a CVR value of 1, which means most items are agreed upon by the expert. Surprisingly, even though this construct has the second highest number of items, experts seemed to agree to retain all of the items. However, items with a CVR value of .75, such as Items 68 and 78, even though they did not reach the CVR value of 1, are considered good but can be revised. Experts suggested the wording to be as precise and ethical as possible since all items in this construct reflect the belief to be morally right.

Table 6

Construct 3 CVR Values And Item Status

| No. | Construct | No. of Item | CVR Value | Item Status |
|-----|-----------------|-------------|-----------|-------------|
| 3 | Student's Input | 34 | 1 | Accepted |
| | | 35 | 1 | Accepted |
| | | 36 | 1 | Accepted |
| | | 37 | 1 | Accepted |
| | | 38 | .5 | Revised |
| | | 39 | .75 | Accepted |
| | | 40 | .25 | Revised |
| | | 41 | .75 | Accepted |
| | | 42 | .75 | Accepted |
| | | 43 | .5 | Revised |
| | | 44 | 1 | Accepted |
| | | 45 | 1 | Accepted |
| | | 46 | 1 | Accepted |
| | | 47 | 1 | Accepted |
| | | 48 | 1 | Accepted |
| | | 49 | 1 | Accepted |
| | | 50 | .75 | Accepted |
| | | 51 | .75 | Accepted |
| | | 52 | .25 | Revised |
| | | 53 | 1 | Accepted |
| | | 54 | .75 | Accepted |
| | | 55 | .25 | Revised |
| | | 56 | 1 | Accepted |
| | | 57 | 1 | Accepted |
| | | 58 | 1 | Accepted |
| | | 59 | 1 | Accepted |
| | | 60 | 1 | Accepted |
| | | 61 | 1 | Accepted |
| | | 62 | 1 | Accepted |
| | | 63 | .75 | Accepted |

Table 7

Construct 4 CVR Values And Item Status

| No. | Construct | No. of Item | CVR Value | Item Status |
|-----|-----------------------------|-------------|-----------|-------------|
| 4 | Professional Responsibility | 64 | 1 | Accepted |
| | | 65 | 1 | Accepted |
| | | 66 | 1 | Accepted |
| | | 67 | 1 | Accepted |
| | | 68 | .75 | Accepted |
| | | 69 | 1 | Accepted |
| | | 70 | 1 | Accepted |
| | | 71 | 1 | Accepted |
| | | 72 | 1 | Accepted |
| | | 73 | 1 | Accepted |
| | | 74 | 1 | Accepted |
| | | 75 | 1 | Accepted |
| | | 76 | 1 | Accepted |
| | | 77 | 1 | Accepted |
| | | 78 | .75 | Accepted |
| | | 79 | 1 | Accepted |
| | | 80 | 1 | Accepted |
| | | 81 | 1 | Accepted |
| | | 82 | 1 | Accepted |
| | | 83 | 1 | Accepted |
| | | 84 | 1 | Accepted |

In Table 8, the intuition construct has high and medium CVR values. Items 88 and 100 are not good with CVR values of .5 and have to be revised. Experts agreed that these items are

unclear in representing intuition and not in the context of professional judgment. To address this, they recommended rephrasing the items to include specific examples of pattern recognition or decision-making based on intuition. Items with CVR values of .75 were accepted, but must be reviewed for potential refinements. A possible explanation for this might be the nature of a teacher, who would consider their intuition when making a judgement. This is agreed by previous research, where expert intuition is increasingly considered a valid form of knowledge and has proven its effectiveness in judgement and decision-making in a variety of fields (Hurteau et al., 2020; Vanlommel & Pepermans, 2021).

Table 8

| No. | Construct | No. of Item | CVR Value | Item Status |
|-----|-----------|-------------|-----------|-------------|
| 5 | Intuition | 85 | .75 | Accepted |
| | | 86 | 1 | Accepted |
| | | 87 | .75 | Accepted |
| | | 88 | .5 | Revised |
| | | 89 | .75 | Accepted |
| | | 90 | .75 | Accepted |
| | | 91 | 1 | Accepted |
| | | 92 | .75 | Accepted |
| | | 93 | 1 | Accepted |
| | | 94 | 1 | Accepted |
| | | 95 | .75 | Accepted |
| | | 96 | .75 | Accepted |
| | | 97 | 1 | Accepted |
| | | 98 | 1 | Accepted |
| | | 99 | 1 | Accepted |
| | | 100 | .5 | Revised |
| | | 101 | 1 | Accepted |
| | | 102 | 1 | Accented |

Construct 5 CVR Values And Item Status

Overall, the results show that feedback from experts is very important in improving the items and response options in instrument development. High CVR values for most items mean that the items match well with the intended constructs, which gives confidence in the validity of the instrument. However, items with a CVR value less than .75 have to be revised to make them clearer and to suit the context.

The CVR value is a reflection of how good the wording of items and response choices in the research instrument are. The instrument is good if the wording of items and response choice is clear and in the context. Items that scored particularly low (CVR = .25 or .50) were critically reviewed to identify recurring problems. The experts found that these items often lacked contextual clarity, were too abstract, or did not align well with real classroom experiences. For example, items 40 and 52 under the Student Input Construct used general terms that did not accurately reflect classroom dynamics, resulting in lower levels of agreement. In item 55, the wording implied a one-size-fits-all assumption which contradicted the different realities in Malaysian schools. These findings emphasize the importance of contextual sensitivity and specificity, comprehensible language in developing items intended for widespread use in professional judgement.

However, the overall results were very encouraging. The results of the step-by-step process of content validation show that the instrument is not only theory-based but also practical and can be used in practice. The overall high CVR values in this study are consistent with findings

701

from other educational instrument validation studies. For instance, Nurhafizah et al. (2024) also reported strong expert agreement in validating the Situational Judgement Test for Digital Leadership using a similar CVR approach. Patronella and Mohd Effendi Ewan (2024) also emphasized that most of their items achieved high agreement when evaluated by the experts, especially when the construct was clearly operationalized. These illustrate that high CVR values are common in validating constructs that require expert judgement, which reinforces the credibility of the results of this study. This, of course, needs to be followed with further study in assessing its reliability, using Rasch Model analysis, as has been conducted in previous studies in the Malaysian educational context (Muhamad Firdaus & Mohd Effendi Ewan, 2022; Saralah et al., 2022).

Conclusion

This study succeeded in achieving consensus among the expert panel on items developed to measure teachers' professional judgment. Such steps are very important in determining whether the instrument developed is in line with the development of new instruments. Among 102 items created, nine items need to be refined, reflecting that most of the items are ready and wellconstructed based on the specification table of the instrument, clear conceptualization, and operationalization. The use of the Content Validity Ratio (CVR) in this study can directly show the strengths and weaknesses of every item created based on the level of agreement between the experts in the panel of experts. This method was very effective as the differences in expert opinion can be easily identified and refined. In addition to the CVR method, inter-rater reliability was supported through Cohen's Kappa analysis, which showed moderate agreement between the two experts. This statistical evidence adds further credibility to the expert consensus and strengthens the methodological rigor of the content validation process. In the final stage, all refined items need to be tested through a pilot study using the Rasch Model to ensure the validity and reliability of the instrument. Through Rasch Model analysis, every item will be critically examined, with only items meeting criteria such as mean square fit statistics (MNSQ), unidimensionality, and item polarity retained. Items that fail to meet these criteria will be eliminated based on statistical and conceptual considerations. This iterative process aims to produce a robust and reliable instrument for assessing teachers' professional judgment in Malaysia. Nevertheless, this study is not without limitations. While the size of the expert panel was acceptable for CVR analysis, it was relatively small and could limit generalizability. Furthermore, the study focuses exclusively on content validity. Future studies should include larger expert samples and apply additional psychometric analyses, such as using the Rasch Model and Confirmatory Factor Analysis (CFA), to determine construct and criterion validity. Beyond instrument development, this study gives a distinctive perception of teachers' decisionmaking processes by highlighting the role of professional judgement in classroom-based assessment practices. These findings are essential to ongoing reforms in teaching, learning and educational assessment, locally and globally.

Declarations Acknowledgements

Not applicable.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Ethics Approval

Not applicable.

Funding Acknowledgements

Not applicable.

Citation to this article

Abdul Hadi, A. A., & Mohd Matore, M. E. E. (2025). Content validity assessment of Malaysian teachers' professional judgement instrument. *International Journal of Organizational Leadership*, *14*(First Special Issue), 689-704. https://doi.org/10.33844/ijol.2025.60512

Rights and Permissions



 $\ensuremath{\textcircled{\sc 0}}$ 2024 Canadian Institute for Knowledge Development. All rights reserved.

International Journal of Organizational Leadership is published by the Canadian Institute for Knowledge Development (CIKD). This is an open-access article under the terms of the Creative Commons Attribution (CC BY) License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Adila Athirah, A. H., & Mohd Effendi Ewan, M. M. (2024). Teachers' professional judgment in Malaysia classroom-based assessment: The iconic runaway in education? *International Journal of Academic Research in Economics and Management Sciences*, 13(2), 172–183. https://doi.org/10.6007/IJAREMS/v13-i2/21246
- Aguirre, V. C. de S. P., Dória, J. P. da S., Melo, R. A. de, Fernandes, F. E. C. V., & Mola, R. (2024). Content Validity of an Instrument for Assessing Skin Lesions in a Mother and Child Hospital. *Revista de Enfermagem da UFPI*, *13*(1), 1–11. https://doi.org/10.26694/reufpi.v13i1.4579
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. Assessment in Education: Principles, Policy and Practice, 20(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364
- Angraini, Rose Alinda, A., & Okfalisa. (2021). Measuring information security policy compliance: Content validity of questionnaire. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(1), 469–475. https://doi.org/10.11591/ijeecs.v22.i1.pp469-475
- Bahagian Pembangunan Kurikulum. (2019). *Pentaksiran Bilik Darjah Edisi ke-2 Bahagian Pembangunan Kurikulum Kementerian Pendidikan Malaysia* (2nd ed.) [Classroom Based Assessment 2nd edition Curriculum Development Department Ministry of Education] Kementerian Pendidikan Malaysia.
- Ben Knight. (2023). Nurturing professional judgement (I. Menter, Ed.). Critical Publishing.
- Coles, C. (2002). Developing professional judgment. *Journal of Continuing Education in the Health Professions*, 22(1), 3–10. https://doi.org/10.1002/chp.1340220102
- Creswell, J. W. (2014). A concise introduction to mixed methods research. SAGE publications.
- Haidi, E., Mohd. Khairuddin, A., Mohd. Azizuan, A., & Mohd Rizal, L. (2020). Pengaruh Pengalaman Mengajar Terhadap Efikasi Guru dalam Kalangan Guru-guru Sekolah Rendah di Zon Pulau, Utara Sabah [The Influence of Teaching Experience on Teacher Efficacy among Primary School Teacher in the Northern Island Zone, Sabah]. E-Proceeding of the International Conference on Education, Social Sciences & Technology, 104–115.
- Helm, D., & Roesch-Marsh, A. (2017). The ecology of judgement: A model for understanding and improving social work judgements. *The British Journal of Social Work*, 47(5), 1361–1376. https://doi.org/10.1093/bjsw/bcw091

- Hurteau, M., Rahmanian, J., Houle, S., & Marchand, M.-P. (2020). The role of intuition in evaluative judgment and decision. *American Journal of Evaluation*, 41(3), 326–338. https://doi.org/10.1177/1098214020908211
- Iqbal, M. Z., Ivan, R., Robb, C., & Derby, J. (2025). Evaluating factors that impact scoring an open response situational judgment test: A mixed methods approach. Frontiers in Medicine, 11, 1–16. https://doi.org/10.3389/fmed.2024.1525156
- Kahneman, D. (2011). Thinking, fast and slow. Penguin Books.
- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sciences Education, 18(1), 1–17. https://doi.org/10.1187/cbe.18-04-0064
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310
- Lawshe, C. H. (1975). A Quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x
- Melor, M., Rosseni, D., & Mohd Effendi Ewan, M. M. (2021). Teaching 4.0 competency in higher learning institutions: A systematic mapping review. *International Journal of Learning, Teaching and Educational Research*, 20(10), 217–231. https://doi.org/10.26803/ijlter.20.10.12
- Memmedova, K., & Ertuna, B. (2024). Development of a fuzzy likert scales to measure variables in social sciences. *Information Sciences*, 654, 119792. https://doi.org/10.1016/J.INS.2023.119792
- Mohd Effendi Ewan, M. M., Ahmad Zamri, K., Siti Mistima, M., Nor Adila, A., & Effa Rina, M. M. (2018). The Influence of Intellectual Quotient (IQ), Emotional Quotient (EQ) and Spiritual Quotient (SQ) against Adversity Quotient (AQ) on Polytechnic Students in Malaysia. *Journal of Engineering Science and Technology*, 2018(Special Issue), 83–91.
- Mohd Effendi Ewan, M. M., Hisyamsani, I., Normawati, A. R., & Ahmad Zamri, K. (2017). Kesahan Kandungan Pakar Instrumen IKBAR Bagi Pengukuran AQ Menggunakan Nisbah Kesahan Kandungan [Expert Content Validaty of IKBAR Instrument for AQ measurement using Content Validity Ratio]. Proceeding of International Conference On Global Education V (ICGE V), 979–997.
- Mohd Effendi Ewan, M. M., Mohammed Afandi, Z., Muhamad Firdaus, M. N., Ahmad Zamri, K., & Nordin, A. R. (2021). The Development and Psychometric Assessment of Malaysian Youth Adversity Quotient Instrument (MY-AQi) by combining Rasch model and confirmatory factor analysis. *IEEE Access*, 9, 13314–13329. https://doi.org/10.1109/ACCESS.2021.3050311
- Muhamad Firdaus, M. N., & Mohd Effendi Ewan, M. M. (2022). Rater severity differences in english language as a second language speaking assessment based on rating experience, training experience, and teaching experience through manyfaceted Rasch measurement analysis. *Frontiers in Psychology*, 13. https://doi.org/10.3389/fpsyg.2022.941084
- Nurhafizah, A. M., Mohd Effendi Ewan, M. M., & Aida Hanim, A. H. (2024). Face validity of situational judgement test for digital leadership using content validity ratio. *Educational Administration: Theory and Practice*, 30(5), 12917–12927. https://doi.org/10.53555/kuey.v30i5.5567
- Patronella, W. Y., & Mohd Effendi Ewan, M. M. (2024). Development and validation of an instrument to measure science teachers' understanding of Fink's taxonomy. *Educational Administration: Theory and Practice*, 30(4), 8937–8949. https://doi.org/10.53555/kuey.v30i4.2932
- Pozzo, M. I., Borgobello, A., & Pierella, M. P. (2019). Using questionnaires in research on universities: Analysis of experiences from a situated perspective. *REIRE Revista d'Innovació i Recerca En Educació*, 12(2), 1–16. https://doi.org/10.1344/reire2019.12.227010
- Rubio, D. M. G., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifyng 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
- Saralah, S., Kamisah, O., & Mohd Effendi Ewan, M. M. (2022). Rasch analysis for disposition levels of computational thinking instrument among secondary school students. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(3), 2–15. https://doi.org/10.29333/ejmste/11794
- Siti Hannah, S., & Mohd Effendi Ewan, M. M. (2024). Systematic literature review on the implementation of the six sigma approach in education. *International Journal of Evaluation and Research in Education*, 13(1), 262–270. https://doi.org/10.11591/ijere.v13i1.26196
- Surianshah, S. (2022). Who gains from class size reduction? Another look at Malaysia's "lost boys phenomenon" in student achievement. *Jurnal Ekonomi Malaysia*, 56(3), 119–143. https://doi.org/10.17576/JEM-2022-5603-07
- Vanlommel, K., & Pepermans, E. (2021). Validation of the teacher decision-making inventory (TDMI): Measuring databased and intuitive dimensions in teachers' decision process. *Studia Paedagogica*, 26(4), 47–66. https://doi.org/10.5817/SP2021-4-3
- Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal*, 56(3), 792–821. https://doi.org/10.3102/0002831218803891

- Wharton, A. (2022). Trusting teacher professional judgement. In *Empowering Teachers and Democratising Schooling* (pp. 175–187). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-4464-2_12
- Winch, C. (2022). Forming and exercising professional judgement. *Journal of Vocational Education and Training*, 74(1), 12–30. https://doi.org/10.1080/13636820.2020.1860117
- Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1), 25–27. https://doi.org/10.22037/jme.v3i1.870
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, 4(2), 165–178. https://doi.org/10.15171/jcs.2015.017