

**INTERNATIONAL JOURNAL OF
MODERN EDUCATION
(IJMOE)**www.gaexcellence.com/ijmoe

A CONTENT VALIDATION STUDY OF THE TEACHER JOB PERFORMANCE INSTRUMENT FOR PRIMARY SCHOOLS

Goh Kok Ming¹, Fariha Diyana Awang Ali², Rosinah Mahmood^{3*}

¹Faculty of Education, Open University Malaysia (OUM), Malaysia

 kokming888@gmail.com

 <https://orcid.org/0009-0002-6666-2042>

²Faculty of Education, Open University Malaysia (OUM), Malaysia

 fariha_diyana@oum.edu.my

 <https://orcid.org/0009-0006-9795-5605>

³Faculty of Education, Open University Malaysia (OUM), Malaysia

 rosinah_mahmood@oum.edu.my

 <https://orcid.org/0009-0008-3391-5983>

*Corresponding Author

Article Info:

Article history:

Received date: 13.08.2026

Revised date: 30.08.2026

Accepted date: 14.09.2026

Published date: 23.09.2026

To cite this document:

Goh, K. M., Ali, F. D. A., & Mahmood, R. (2026). A Content Validation Study Of The Teacher Job Performance Instrument For Primary Schools. *International Journal of Modern Education*, 8(31), 1236-1252.

Abstract:

Teacher job performance is widely acknowledged to have a significant impact on the level of instruction quality, school performance, and student learning outcomes. With the increasing importance placed upon accountability and evidence-based evaluation in education systems around the world, there is a rising urgency to find a sound and valid method to assess teacher performance within a certain context. Although numerous studies have been carried out to study the factors affecting teacher performance, most of them heavily rely on the existing performance scales that are either modified or not applicable to the context in question. In this respect, the study aims to develop and validate the Teacher Job Performance Instrument (TJPI), which comprises four dimensions: instructional effectiveness, classroom management, student engagement, and professional development. To establish the content validity of the TJPI, the quantitative method was employed. The research was conducted with the help of seven experts in educational leadership and teacher evaluation. To determine the content validity of the measures employed, the Item-Level Content Validity Index (I-CVI), Scale-Level Content Validity Index (S-CVI/Ave), and modified kappa (κ^*) were utilized to account for chance occurrence. Results showed that all the 24 items had excellent content validity (I-CVI = 0.86–1.00), with an overall S-CVI/Ave being more than 0.90 which is the acceptable cut-off. The modified kappa results ranged from 0.74 to 1.00, which indicates substantial to excellent agreement above chance level. The results provide strong support for the theoretical representativeness, theoretical consistency, and expert

agreement regarding the construct of the TJPI. The developed measure serves as a critical tool for any future steps related to validation and testing of the construct.

DOI: 10.35631/IJMOE.831072

Keyword:

Teacher Job Performance, Content Validity, Teacher Professional Development



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact ijmoe@gaexcellence.com.

Introduction

The job performance of teachers is seen as an important factor in the success of a school and the results of student education (Ramos-Villagrasa et al., 2019). In academic writing, the assessment of teacher job performance has a multidimensional nature that includes features such as task-oriented performance, contextual performance, and professional involvement (Campbell & Wiernik, 2015). At the same time, modern research studies show how leadership, information management, and organizational practices influence the job performance of teachers (Oni & Avcı, 2024; Berhanu, 2025). In Malaysia, teacher job performance is acquiring greater importance for the national educational transformation process among the educational reforms set out in the document called the Malaysian Education Blueprint (2013-2025), which declares the importance of quality in instruction, the effectiveness of school leaders, and performance evaluation based on the principles of accountability.

The Ministry of Education Malaysia (MOE) has established structured systems of evaluation such as PBPPP (Penilaian Bersepadu Pegawai Perkhidmatan Pendidikan) that are meant to evaluate the degree of effectiveness of teachers based on the Standard Kualiti Pendidikan Malaysia Gelombang 2 (SKPMg2), especially Standard 4 that highlights the evaluation of quality in the teaching-learning process. Standard 4 views teacher performance assessment as multidimensional and includes teaching planning, classroom management, student involvement, assessment implementation, and professional practice related to reflection. Even though this structured framework has attracted significant scholarly interest, many research studies implement modified and intact cases of performance evaluation schemes not tailored to the context of schools. Therefore, the importance of an instrument that has a theoretical foundation and through which effective performance evaluation can be enacted appears to be highly needed. Before empirical use and implementation, the instrument's content should be validated to ensure its items adequately cover the construct domain (Polit & Beck, 2006).

Literature Review

Conceptual Foundations of Teacher Job Performance

In the studies related to educational effectiveness, the concept of teacher job performance assumes critical significance as it has an impact on students' performance, classroom environment, and general quality of an educational institution. In the field of organizational psychology, it refers to observable actions controlled by the employee and directly contributing to the organization's goals (Kumar, 2023). There is also a behavioral approach to this concept, allowing differentiation between the term performance and that of results, as the first one refers to actions while the second one describes the results of those actions (Parveen et al., 2022). In pedagogical practice, the element of job performance involves more than only the performance of ordinary duties. The notion covers everything starting from effective delivery of lessons and classroom management to working with students and cooperation with fellow teachers (Ali et al., 2024; Rashid et al., 2024; Suhaimi et al., 2024).

The literature has increasingly acknowledged the multidimensional nature of teacher performance, which is intrinsically linked to the domains of instruction, organization, and relationships. Numerous empirical studies have recently validated this multidimensionality. For instance, research has shown that instructional leadership has a significant impact on teachers' professional performance since it involves systematic pedagogical guidance and an accountability system (Berhanu, 2025). Besides, knowledge management practices such as knowledge sharing, support systems, and competency improvement have also demonstrated significant links with teachers' performance (Oni & Avci, 2024). Thus, it becomes obvious that the idea of teacher performance cannot be understood in terms of one or two instruction-related measures. Nevertheless, despite wide consensus over the importance of teacher performance, there remains no clear definition of the concept in the studies on educational performance. Some studies depend on general criteria of employee performance borrowed from the business sector, while others confine its meaning to instruction-related measures only. This suggests that accurate methodological tools for measuring teacher performance should be developed in a theory-based and context-oriented manner.

Multidimensional Models of Job Performance

The main theoretical framework behind the study of performance is based on Campbell's performance theory (1990) that views job performance as a latent variable that consists of various behavioural dimensions. Campbell and Wiernik (2015) classify performance into task performance, contextual performance, and counterproductive work behaviour. Ramos-Villagrasa et al. (2019) support the existence of a multidimensional nature of performance by means of empirical studies with the Individual Work Performance Questionnaire.

- a) Task performance refers to behaviours directly related to core job responsibilities. In teaching, this includes lesson planning, instructional delivery, curriculum alignment, and assessment implementation (López-Cabarcos, Vázquez-Rodríguez & Quiñoá-Piñero, 2022)
- b) Contextual performance involves behaviours that support the broader organizational and social environment, such as collaboration, adaptability, and initiative. In educational settings, this may manifest as fostering a positive classroom climate,

mentoring colleagues, or contributing to school improvement initiatives (Akanpaadgi, Kuuyelleh & Adam, 2024)

- c) Counterproductive work behaviour includes actions that hinder organizational functioning. In teaching contexts, such behaviours might involve negligence, resistance to improvement, or unprofessional conduct (Ghasemi & Herman, 2024)

Though this tripartite model has impressive theoretical potential, it lacks the precision needed for straightforward applicability in education. Educators should deal with teaching-related issues that require a combination of didactic expertise and classroom management skills, interpersonal communication, and professional development. As a result, a growing number of researchers advocate the need for tailored approaches to measuring teachers' performance that consider the specifics of teaching practice. In this context, the literature suggests that pedagogical leadership influences instructional practices (Berhanu, 2025). Furthermore, school knowledge-sharing programs help enhance performance through professional development (Oni & Avcı, 2024). All the mentioned contributions suggest the importance of adding the development and motivation aspects to the assessment of teacher performance.

Teacher Performance Evaluation in Malaysia

Teacher performance in Malaysia can be evaluated through the National Quality Assurance Instrument (SKMPg2) and Individual Teacher Performance Appraisal (PBPPP). The coexistence of SKPMg2 and PBPPP reflects Malaysia's structured approach to teacher evaluation (Mukhtar & Noor, 2024)

National Quality Assurance Instrument: SKPMg2

In Malaysia, teacher job performance is formally embedded within the National Quality Assurance Framework (Standard Kualiti Pendidikan Malaysia Gelombang 2, SKPMg2) instrument. SKPMg2 serves as the national quality assurance system guiding school evaluation and continuous improvement. Within this instrument, Standard 4: Pengajaran dan Pembelajaran (PdP) specifically evaluates teaching quality. Standard 4 conceptualizes teacher performance across multiple domains, including: (i) Instructional planning and preparation, (ii) Pedagogical delivery, (iii) Classroom management, (iv) Student engagement, (v) Assessment practices, and (vi) Reflective improvement and professional development. Notably, this structure mirrors multidimensional performance theory (Campbell & Wiernik, 2015) while emphasizing pedagogical and relational competencies specific to teaching.

PBPPP: Individual Teacher Performance Appraisal

Individual Teacher Performance Appraisal (PBPPP) is Malaysia's formal individual teacher performance appraisal system, implemented by the Ministry of Education to evaluate teachers annually. PBPPP integrates multiple evaluation components, including: (i) Professional competence, (ii) Instructional effectiveness, (iii) Organizational contribution, (iv) Professional values and ethics, (v) Leadership (for senior teachers), and (vi) Co-curricular and extra-role involvement. Therefore, PBPPP serves a dual function as an administrative evaluation, such as promotion, confirmation, and performance rating, and a professional accountability mechanism. Although PBPPP provides structured appraisal indicators, it primarily serves administrative and accountability purposes.

Theoretical Integration for the Present Study

Four theoretical foundations are used in this research as a basis to create the Teacher Job Performance Instrument (TJPI). The first is Campbell's theory of multidimensional performance that understands performance in the workplace as a complex of observable behaviors toward attaining work-related goals and controlled by a worker (Sriadmitum, 2023). Within this framework, a distinction is made between task performance and contextual performance, creating a basis for understanding teacher performance in broader terms than solely the performance outcome. Findings support the validity of applying a domain approach to performance measurement (Wartenberg et al., 2023).

The second theoretical foundation is found in empirical studies establishing links between leadership practices in education and teachers' performance in the working environment. Studies show that pedagogical leadership and supervision of instructional processes are substantial predictors of teachers' performance both directly and through mediation effects such as organizational commitment (Rokeman, Kob & Sobry, 2023). The findings support the view that teaching effectiveness consists of instructional activities as well as leadership and organizational elements. The research reports on the role of knowledge management and professional capacities suggest that teachers' performances can be enhanced through knowledge-sharing systems as well as the building of competencies and the establishment of a collaborative environment. Studies have established the significance of integrating professional development and collaborative learning behaviour in the assessment of teacher performance. Fourth, the study is placed within the context of the teacher evaluation framework in Malaysia, which is made up of SKPMg2 Standard 4, which defines teacher efficiency in the areas of instructional planning and execution, classroom management, student involvement in the teaching-learning process, and evaluation and self-improvement (Rahman et al, 2025; Masuwai, Zulkifli & Tamuri, 2022). It serves as a national standard for teacher appraisal at the school level, while PBPPP is the annual performance appraisal system for teachers in Malaysia.

The four foundations combined justify that the Teacher Job Performance Instrument (TJPI) connects with theoretical information and empirical evidence (Yusoff et al., 2025). It leads to the formulation of four dimensions: (i) Instructional Effectiveness, (ii) Classroom Management, (iii) Student Engagement, and (iv) Professional Development. Each dimension reflects the combination of Campbell's performance dimensions, predictors of teacher productivity, expectations set in SKPMg2 Standard 4, and competencies and professionalism requirements stated in PBPPP (Mustapa & Miskon 2022). The implementation of PBPPP and SKPMg2 Standard 4 in Malaysia has created a formal teacher evaluation system; however, it does not lessen the need for teacher evaluation instruments like the Teacher Job Performance Instrument (TJPI), which is validated. While the PBPPP is primarily an administrative performance evaluation instrument for accountability, staff evaluation, and career advancement purposes, SKPMg2 Standard 4 is a classroom evaluation tool (Mustapa & Miskon, 2022). Thus, the TJPI is an instrument that defines teacher performance as a complex behaviour based on performance theory without being duplicative of PBPPP and SKPMg2 domains.

Methodology

Research Design

This study employed a quantitative methodology focusing on content validation to determine the representativeness and theoretical appropriateness of the Teacher Job Performance Instrument (TJPI). The phase of content validation was chosen as the first stage in instrument development as per psychometric criteria, which demand domain representativeness prior to testing reliability and construct validity (Polit & Beck, 2006). The aim of this study was to determine how appropriate, clear, and relevant the items of the instrument are with respect to the theoretical construct of teacher job performance and the standards of Malaysia, such as SKPMg2 Standard 4 and PBPPP domains.

Instrument Formulation

The development of the TJPI followed a systematic multi-stage process: (i) Phase 1: Construct Identification

Phase 1: Construct Identification

The teacher's job performance was conceptualized as a multidimensional behavior as per Campbell's theory on performance, empirical studies on the performance of leaders, knowledge management-performance statistical analysis in the SKPMg2 Standard 4 fields, and dimensions of teacher competency covered by the PBPPP. From the information presented above, it can be inferred that several performance indicators have been determined, such as Instructional Effectiveness, Classroom Management, Student Engagement, and Professional Development (see Table 1). This information can also give the indication that teacher job performance dimensions are not based on arbitrary criteria but rather on Campbell's theory of performance, empirical analysis, and Malaysian policies. Instructional Effectiveness interprets task performance phenomenon by means of Campbell's theory, while Classroom Management is an example of contextual performance in educational settings. Student Engagement provides further elaboration on contextual performance in the process of pedagogy, and Professional Development is a result of the creation of dynamic capability.

Table 1: The Conceptualization of Teacher Job Performance

Theoretical Foundation	Core Constructs Identified	Alignment with Malaysian Framework	Synthesized Dimension in TJPI	Rationale for Inclusion
Campbell's Performance Theory (Behavioral Performance Model)	Task performance (core job behaviors)	SKPMg2: Standard 4 Planning and Implementation in Teaching and Learning	Instructional Effectiveness	Teaching requires structured lesson planning, delivery, and assessment behaviors that reflect task-oriented performance.
	Contextual performance	SKPMg2: Standard 4	Classroom Management	Effective teaching extends beyond

Theoretical Foundation	Core Constructs Identified	Alignment with Malaysian Framework	Synthesized Dimension in TJPI	Rationale for Inclusion
	(supportive behaviors)	Classroom Management		instruction to managing learning environments and behavioral regulation.
Leadership–Performance Research	Instructional leadership impact	SKPMg2: Standard 4 pedagogical supervision	Instructional Effectiveness	Empirical evidence shows instructional guidance predicts teacher instructional quality.
	Organizational commitment mediation	PBPPP: Professional responsibility	Professional Development	Leadership effects often operate through teacher growth and reflective practice.
Knowledge Management Research	Knowledge sharing, competence development	PBPPP: Professional competence	Professional Development	Teacher performance improves through continuous professional learning and collaborative practice.

Table 2: The Final Teacher Job Performance’s Dimensions

Final TJPI Dimension	Theoretical Source	Policy Alignment	Construct Description
Instructional Effectiveness	Task Performance Theory	SKPMg2: Planning & Delivery	Behavioral execution of curriculum-aligned instruction and assessment.
Classroom Management	Contextual Performance Theory	SKPMg2: Learning Environment	Regulation of classroom climate, discipline, and safety.
Student Engagement	Pedagogical Engagement Theory	SKPMg2: Student Participation	Facilitation of active, meaningful student involvement in learning.
Professional Development	Knowledge & Leadership Research	PBPPP: Professional Competence	Reflective practice, collaboration, and continuous instructional improvement.

Phase 2: Item Generation

The development of the Teacher Job Performance Instrument (TJPI) involved a systematic and theory-led process to achieve high levels of conceptual rigor and contextual importance. The first step in the process involved obtaining an initial list of 32 items through a synthesis of (i) national policy documents, which included, among other documents, SKPMg2 and PBPPP; (ii) empirical studies on teacher effectiveness, instructional leadership, and knowledge management; and (iii) performance theory mapping based on Campbell's multidimensional behavioural model of job performance. By using this integrative technique, researchers were able to obtain items that were theory-driven and not simply descriptive of policy standards. While SKPMg2 Standard 4 provided field-specific pedagogical indicators such as instructional planning, classroom delivery, management, student engagement, and reflection, PBPPP provided competency-based outcomes relating to professional accountability, instructional implementation, and development. The ideas developed in policy domains were conceptually mapped into structures of task and contextual performance as per the task performance theory, which ensured a high level of theoretical coherence and representation of the constructs.

All items were formulated in an observable manner, making use of behaviourally anchored statements in line with the definition of job performance as actions within the individual's control rather than pertaining to outcomes mandated by outside influences. A five-point Likert-type scale (1=Strongly Disagree; 5=Strongly Agree) was used to measure variability in self-reported professional performance while being effective from the psychometric standpoint for future validation purposes. After having undergone expert internal evaluation, redundancy analysis, and screening for conceptual relevance, 8 items were removed due to conceptual overlap or low distinctiveness. Finalization of the process led to a 24-item tool divided into 4 theoretical dimensions: Instructional Effectiveness (5 items), Classroom Management (7 items), Student Engagement (6 items) and Professional Development (6 items). Such structure represents a blend of performance theory earlier developed by scientists and Malaysian national assessment systems.

Expert Panel Selection

Content validation of the Teacher Job Performance Instrument (TJPI) was conducted by utilizing a purposely selected panel of seven specialist professionals in line with Lynn's (1986) belief that the engagement of six to ten specialists is adequate for efficient construction of Content Validity Index (CVI) estimates while decreasing the possibility of the occurrence of pure chance agreement. Amidst the selection of specialists, strict requirements were imposed to assure both theoretical richness and contextual relevance. In particular, the panel was required to possess the following: (i) a PhD degree in the area of education, educational administration, or a related field; (ii) a minimum of ten years of practical experience in the fields of education, educational administration, or education; (iii) proved experience in assessing teachers' performance, policy-making, or execution; (iv) in-depth knowledge of the system of teachers' evaluation in Malaysia; and (v) being the author of peer-reviewed publications. The choice of experts was made possible by the joint application of methodological knowledge and practical involvement in the evaluation of teachers in the country.

Content Validation Procedure

Every expert received a thorough validation document to guarantee that their judgments were both informed and contextually relevant. This package contained (a) conceptual definitions of the four dimensions of TJPI, Instructional Effectiveness, Classroom Management, Student Engagement and Professional Development, aimed at specifying the boundaries of the constructs; (b) the list of 24 items of the measure in its finalized behaviour statements format; (c) a mapping document showing how each item relates to the respective indicators of SKPMg2 Standard 4 (Teaching and Learning); and (d) the summary of references illustrating the alignment with PBPPP teacher competencies domains. All these provisions were very important to ensure consistency of ratings and reduce ambiguity of the concepts because the evaluation scope was based on theoretical as well as policy frameworks. Experts were asked to evaluate the relevance of each item independently by means of a four-point ordinal scale (1 = Not relevant, 2 = Somewhat relevant, 3 = Quite relevant, 4 = Highly relevant) in line with standard content validation procedures (Lynn, 1986). The four-point scale was utilized specifically to avoid the existence of a neutral midway point, thereby forcing decision-making concerning item representative proportion. Pursuant to the standard CVI process, ratings of 3 (Quite relevant) and 4 (Highly relevant) were classified as showing content validity. This classification allowed computing the Item-Level Content Validity Index (I-CVI) methodologically properly, with the possibility of differentiating between adequate and inadequate construct representation. All the evaluations were conducted independently to avoid the effect of peers on the results.

Data Analysis

To measure content validity, the study utilized the Content Validity Index (CVI) framework, which is one of the most widely recommended methods to establish domain representativeness in the development of scales (Polit & Beck, 2006; Polit et al., 2007; Zamanzadeh et al., 2015). At the item level, the Item-Level Content Validity Index (I-CVI) was calculated by dividing the number of experts who rated the item as either being "quite relevant" (3) or "highly relevant" (4) by the total number of experts. Therefore, I-CVI was calculated based on the level of agreement of experts regarding the relevance of each item. With a panel of experts that consisted of seven experts, an I-CVI threshold of ≥ 0.86 was used, meaning that the minimum number of experts that indicate the relevance of the item should be at least six (out of seven), conforming to some recommendations that the larger panels require higher levels of consensus to meet the requirement of content validity (Lynn, 1986; Polit et al., 2007). This threshold minimizes the chances that the items are retained based on inappropriate level of consensus.

At the scale level, the Scale-Level Content Validity Index was calculated by taking the average of the I-CVI for each individual item. The S-CVI/Ave provides an indication of how well the tool represents the construct being measured. According to well-established standard levels of psychometrics, the S-CVI/Ave value of 0.90 and above is indicative of an excellent level of content validity at the scale level (Polit & Beck, 2006; Polit et al., 2007). The S-CVI/Ave should be used instead of the S-CVI/UA when dealing with multi-dimensional measures because it is essentially stable and nonrestrictive as far as the estimation of scale adequacy is concerned (Zamanzadeh et al., 2015). To obtain a more cautious assessment of the degree of agreement above chance, which is provided by the CVI, the modified kappa statistic (represented as κ^*) was further calculated for each of the items. The modified kappa calculation is defined by the κ^* formula $\kappa^* = (I-CVI - P_c) / (1 - P_c)$, where P_c stands for the probability of chance

agreement. The use of the binomial probability formula, $P_c = [N! / (A!(N - A)!)] \times (0.5)^N$ makes it possible to calculate the probability of chance agreement, where N represents the total number of experts, and A stands for the number of experts who found the item important (Polit et al., 2007). Thus, we see that the κ^* adjusts the I-CVI to check whether agreement happened by chance or was true agreement. According to the established guidelines for interpretation, κ^* scores below 0.40 mean the existence of poor agreement, scores between 0.40 and 0.59 mean the existence of fair agreement, scores between 0.60 and 0.74 indicate good agreement, and scores that are 0.75 and above indicate the existence of excellent agreement (Polit et al., 2007). The use of κ^* contributes to the methodological rigor by excluding the possibility of high scores being obtained by means of chance agreement.

Qualitative feedback from professionals was subjected to a thorough review to spot instances of conceptual ambiguity, duplication, or incongruity with other definition concepts. Items that were deemed vague or overlapping in meaning were evaluated even if they complied with the statistical requirements, thus ensuring compliance with the most recent scale improvement methods (Zamanzadeh et al., 2015). Items that did not fulfill the quantitative criteria or that were pointed out as ideationally problematic were going to be altered or eliminated following the research completion. However, all the items appearing in the final version of this study were graded positively in terms of relevant statistical measures.

Ethical Considerations

Before data collection began, ethical permissions were acquired from necessary institutional review boards. The process involved validation of content by subject matter experts, who were informed about the purpose and methodology of the research. Each of them provided the necessary documentation showing their willingness to participate prior to taking part in the research. In order to make sure that confidentiality is maintained and the risk of evaluation bias is minimized, subject matter experts remained anonymous during the whole process of data analysis and data reporting. It should be noted that the validation exercise did not have any relation to assessment procedures applied in the framework of PBPPP or institutional performance measuring exercises. The separation was explained to the participants so that it would not result in any distortions in the participants' responses. Relevant data were stored properly for academic purposes only.

Findings

Content Validity of the Teacher Job Performance Instrument (TJPI)

Content validity of the Teacher Job Performance Instrument (TJPI) was established through systematic expert evaluation using the Item-Level Content Validity Index (I-CVI), Scale-Level Content Validity Index (S-CVI/Ave), and modified kappa statistic (κ^*), consistent with established psychometric standards (Polit & Beck, 2006; Polit et al., 2007). Across all four dimensions, the results demonstrated strong to perfect expert agreement, indicating excellent domain representativeness and conceptual clarity.

As presented in Table 3, all five items under the Instructional Effectiveness dimension achieved unanimous agreement among the seven experts, yielding I-CVI values of 1.00 for each item. The probability of chance agreement (P_c) was calculated at 0.03125, and the corresponding modified kappa values (κ^*) were 1.00 across all items, indicating perfect agreement beyond

chance. The S-CVI/Ave for the dimension was 1.00, exceeding the recommended threshold of 0.90 for excellent scale-level content validity (Polit et al., 2007). These findings suggest that the Instructional Effectiveness items were judged as highly relevant and conceptually aligned with the construct definition, as well as consistent with SKPMg2 Standard 4 instructional domains.

Table 3: Findings of Expert Validation on Instructional Effectiveness (IE) Dimension

Items	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI	Pc	Kappa Statistic
IE1	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
IE2	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
IE3	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
IE4	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
IE5	1	1	1	1	1	1	1	7	1.00	0.03125	1.00

Note. S-CVI/Ave = 1.00 (accepted). I-CVI = content validity index for items; S-CVI = content validity index for scales; Pc = Probability of chance agreement

For the Classroom Management dimension (Table 4), four of the seven items achieved unanimous agreement (I-CVI = 1.00; $\kappa^* = 1.00$), while three items demonstrated agreement from six of the seven experts (I-CVI = 0.86). For these items, the probability of chance agreement was 0.21875, resulting in modified kappa values of 0.82, which fall within the “excellent” range ($\kappa^* \geq 0.75$) according to established interpretive guidelines (Polit et al., 2007). The S-CVI/Ave for the Classroom Management dimension was 0.92, surpassing the recommended 0.90 threshold and indicating strong overall content validity at the dimension level. Although minor variations in agreement were observed for certain items, all retained items exceeded the acceptable I-CVI cut-off of 0.86 for a seven-expert panel (Lynn, 1986), thereby supporting their inclusion without revision.

Table 4: Findings of Expert Validation on Classroom Management (CM) Dimension

Items	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI	Pc	Kappa Statistic
CM1	1	1	1	1	1	0	1	6	0.86	0.21875	0.82
CM2	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
CM3	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
CM4	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
CM5	1	1	0	1	1	1	1	6	0.86	0.21875	0.82
CM6	1	0	1	1	1	1	1	6	0.86	0.21875	0.82
CM7	1	0	1	1	1	1	1	6	0.86	0.21875	0.82

Note. S-CVI/Ave = 0.92 (accepted). I-CVI = content validity index for items; S-CVI = content validity index for scales; Pc = Probability of chance agreement

As shown in Table 5, the Student Engagement dimension demonstrated unanimous expert agreement across all six items (I-CVI = 1.00). Similar to the Instructional Effectiveness dimension, Pc was 0.03125 for each item, and κ^* values were 1.00, reflecting perfect agreement beyond chance. The S-CVI/Ave was calculated at 1.00, indicating exemplary scale-level content validity. These findings confirm that the Student Engagement items were perceived as

highly representative of the construct domain, reinforcing the theoretical inclusion of engagement as a core dimension of teacher job performance.

Table 5: Findings of Expert Validation on Student Engagement (SE) Dimension

Items	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI	Pc	Kappa Statistic
SE1	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
SE2	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
SE3	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
SE4	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
SE5	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
SE6	1	1	1	1	1	1	1	7	1.00	0.03125	1.00

Note. S-CVI/Ave = 1.00 (accepted). I-CVI = content validity index for items; S-CVI = content validity index for scales; Pc = Probability of chance agreement

The Professional Development dimension (Table 6) likewise achieved unanimous expert agreement across all six items, with I-CVI values of 1.00 and κ^* values of 1.00 for each item. The S-CVI/Ave was 1.00, indicating perfect scale-level content validity. The absence of disagreement suggests strong conceptual clarity and domain relevance, particularly in capturing reflective practice, collaborative learning, and continuous professional growth aligned with both PBPPP competency expectations and SKPMg2 reflective standards.

Table 6: Findings of Expert Validation on Professional Development (PD) Dimension

Items	E1	E2	E3	E4	E5	E6	E7	Experts in Agreement	I-CVI	Pc	Kappa Statistic
PD1	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
PD2	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
PD3	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
PD4	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
PD5	1	1	1	1	1	1	1	7	1.00	0.03125	1.00
PD6	1	1	1	1	1	1	1	7	1.00	0.03125	1.00

Note. S-CVI/Ave = 1.00 (accepted). I-CVI = content validity index for items; S-CVI = content validity index for scales; Pc = Probability of chance agreement

Overall Content Validity of the TJPI

As shown in Table 7, across all 24 items, I-CVI values ranged from 0.86 to 1.00, exceeding the recommended minimum threshold for acceptable content validity in panels of six or more experts (Polit et al., 2007). Modified kappa values ranged from 0.82 to 1.00, indicating excellent agreement beyond chance for all retained items. Dimension-level S-CVI/Ave values ranged from 0.92 to 1.00, consistently surpassing the recommended criterion of 0.90 for strong scale-level validity (Polit & Beck, 2006). In summary, these findings demonstrate that the TJPI exhibits robust content validity across all four dimensions. The integration of CVI and modified kappa statistics provides convergent evidence that the instrument items are not only theoretically coherent but also statistically defensible. The high level of expert consensus supports the conceptual representativeness of Instructional Effectiveness, Classroom

Management, Student Engagement, and Professional Development as core domains of teacher job performance within the Malaysian educational context.

Table 7: Summary of Content Validity Indices of TJPI

Dimension	No. of Items	I-CVI Range	κ^* Range	S-CVI/Ave	Interpretation
Instructional Effectiveness (IE)	5	1.00 – 1.00	1.00 – 1.00	1.00	Excellent content validity
Classroom Management (CM)	7	0.86 – 1.00	0.82 – 1.00	0.92	Excellent content validity
Student Engagement (SE)	6	1.00 – 1.00	1.00 – 1.00	1.00	Excellent content validity
Professional Development (PD)	6	1.00 – 1.00	1.00 – 1.00	1.00	Excellent content validity
Overall	24	0.86 – 1.00	0.82 – 1.00	0.98	Strong overall content validity

Note. I-CVI = Item-Level Content Validity Index; κ^* = Modified kappa statistic; S-CVI/Ave = Scale-Level Content Validity Index (average method). Acceptable thresholds: I-CVI ≥ 0.86 (for 7 experts), S-CVI/Ave ≥ 0.90 , $\kappa^* \geq 0.75$ indicates excellent agreement beyond chance (Polit et al., 2007).

Discussion

This study aimed to determine the validity of the Teacher Job Performance Instrument (TJPI) by means of systematic expert evaluation through I-CVI, S-CVI/Ave, and modified κ^* statistics. The results of the study indicate that the instrument has good domain representativeness and conceptual coherence across four dimensions. Specifically, I-CVI values ranged from 0.86 to 1.00 for all TJPI items, meaning that items met the recommended value of I-CVI for seven expert panelists (Lynn, 1986; Polit et al., 2007). To add, modified kappa values varied from 0.82 to 1.00, meaning that the agreement is excellent. It is worth mentioning that κ^* was applied in this study, which makes it methodologically sound, as it accounts for non-random consensus, which is often ignored in the development of educational tools. There was universal consensus in three out of four dimensions, which were Instructional Effectiveness, Student Engagement, and Professional Development, showing that the construct has been well operationalized. However, the level of consistency in the dimension of Classroom Management (I-CVI = 0.86 for items included in the category) is lower but still high and can be explained by a more complicated character of some practices connected to the issue of classroom management. Possible variations in terms of classroom management behaviors lead to variations in the expert's judgments.

The findings of this study also reaffirm the multidimensional deriving principle of teacher performance based on Campbell's performance theory, which considers two performance aspects: task and contextual (Campbell & Wiernik, 2015). Instructional Effectiveness can be ascribed to task performance since it includes planning based on curricula, teaching, and assessments. At the same time, Classroom Management and Student Engagement bring more of contextual performance issues to educational settings, representing behavioural aspects related to our relationships, environment, and participation in the educational activity

undertaken. Professional Development is the manifestation of different abilities, connecting both contextual performance and professional development.

Most notably, the agreement reached on the items related to Professional Development reflects the establishment of reflective practice and continuous learning as crucial components in determining teacher effectiveness. Modern researchers in education emphasize the fact that professional development is regarded as one of the predictors of maintaining effectiveness in teaching and adaptability of educational organization. The κ^* values calculated for the dimensions indicate how experts perceive professional development as a process that should be included in teacher performance. Therefore, the TJPI adds to the theory of performance by introducing different pedagogical, relational, and developmental elements into one performance measurement tool.

Moreover, the significant findings on content validity also have implications for educational policy frameworks in Malaysia. The standards specified in SKPMg2 Standard 4 and PBPPP standards shed light on several aspects of teacher job profile. However, it can be noted that these qualities have been defined for administrative and quality assurance purposes rather than for psychometric research objectives. The findings of this study show the possibility of converting a wide set of policy objectives into a research-based tool by employing different content validation techniques. By aligning TJPI items with those of SKPMg2 and PBPPP and using CVI and κ^* analysis, this study connects the two concepts: administrative evaluations and research measurement. This distinction is important because administrative evaluations can be influenced by structural relationships, accountability, and bias in rating. In contrast, research-based tools should focus on clear constructs and high-quality measurement, with no links to the evaluation process.

The high consistency across dimensions suggests that the concepts of SKPMg2 and PBPPP are conceptually compatible and can be transformed into quantifiable behavioural constructs. At the same time, the TJPI adds a layer of methodological clarity that is lacking in administrative methods by quantifying expert agreement. Therefore, rather than becoming another policy tool, the TJPI enriches the empirical basis for evaluating teachers' performance in Malaysia. At the same time, although the results demonstrate strong evidence of content validity, a few issues need to be discussed. First, content validity is just the first stage in the process of validating the instrument and does not ensure structural validity and predictive power. More empirical assessments must be conducted with larger teacher samples to check the validity of the instrument, its internal consistency, and criterion validity. Second, the expert group was reasonably large but deliberately chosen in the context of Malaysian education. The conduct of cross-cultural validation would enhance the generalizability of the instrument.

Conclusion

In conclusion, this study aimed to develop and establish the content validity of the Teacher Job Performance Instrument (TJPI) through a content validation process. Drawing upon multidimensional performance theory, empirical scholarship on educational leadership and professional competence, and alignment with Malaysian national evaluation frameworks of teacher performance, the study constructed a four-dimensional model of teacher job performance, including Instructional Effectiveness, Classroom Management, Student Engagement, and Professional Development. Through expert panel evaluation and application of the Content Validity Index (I-CVI, S-CVI/Ave) and modified kappa (κ^*) statistics, the

findings demonstrate strong evidence of domain representativeness and agreement beyond chance across all 24 items. The exceptionally high S-CVI/Ave values and robust κ^* coefficients provide compelling support for the conceptual clarity and contextual relevance of the instrument. Future research should proceed to exploratory and confirmatory factor analyses to establish structural validity, assess internal consistency reliability, and evaluate criterion-related and predictive validity using representative teacher samples. Cross-context validation beyond the Malaysian setting may further enhance generalizability.

Acknowledgements: The authors would like to express their sincere gratitude to Open University Malaysia for providing the necessary resources and support throughout the course of this research. Special appreciation is extended to colleagues and peers who contributed valuable insights and constructive feedback, which greatly enhanced the quality of this paper.

Funding Statement: No Funding

Conflict of Interest Statement: The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Modern Education (IJMOE).

Ethics Statement: This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards.

Author Contribution Statement: All authors contributed significantly to the development of this manuscript. Goh Kok Ming was responsible for the conceptualization, methodology, and overall supervision of the study. Fariha Diyana Awang Ali was responsible for data collection, analysis, and interpretation of the results. Rosinah Mahmood contributed to the literature review, drafting, and critical revision of the manuscript. All authors read and approved the final version of the manuscript prior to submission. All authors read and approved the final version of the manuscript before submission.

References

- Akanpaadgi, E., Kuuyelleh, E. N., & Adam, A. (2024). Performance management and contextual performance in technical universities. *Social Sciences & Humanities Open*, 9, 100788. <https://doi.org/10.1016/j.ssaho.2023.100788>
- Ali, F. D. A., Idris, R., Shanmugam, J., Zaimi, F. H. M., Nazri, N. S., & Nizam, U. A. (2024). Unheard struggles: addressing mental health challenges in academia. *International Journal of Advanced Research in Education and Society*, 6(3), 780-789. <https://doi.org/10.55057/ijares.2024.6.3.65>
- Ali, F. D. A., Idris, R., Shanmugam, J., Zaimi, F. H. M., Nazri, N. S., & Nizam, U. A. (2024). The silent crisis: mental health challenges in the academic sphere. *4th International Virtual Conference on Education, Social Sciences and Technology 2024 (IVEST2024)* e-ISBN: 978-967-2476-65-8. Kuala Lumpur, Malaysia.
- Ghasemi, F., & Herman, K. C. (2024). Understanding teacher counterproductive work behavior: tracing individual, occupational, and organizational factors. *Psychology in the Schools*, 61(11), 4250-4272. <https://doi.org/10.1002/pits.23279>
- Kumar, J. (2023). Influence of motivation on teachers' job performance. *Humanities and social sciences communications*, 10(1), 158. <https://doi.org/10.1057/s41599-023-01662-6>
- Li, L., Kanchanapoom, K., Deeprasert, J., Duan, N., & Qi, Z. (2025). Unveiling the factors shaping teacher job performance: exploring the interplay of personality traits, perceived organizational support, self-efficacy, and job satisfaction. *BMC psychology*, 13(1), 14. <https://doi.org/10.1186/s40359-024-02324-1>
- López-Cabarcos, M. Á., Vázquez-Rodríguez, P., & Quiñoá-Piñeiro, L. M. (2022). An approach to employees' job performance through work environmental variables and leadership behaviours. *Journal of Business Research*, 140, 361-369. <https://doi.org/10.1016/j.jbusres.2021.11.006>
- Masuwai, A., Zulkifli, H., & Tamuri, A. H. (2022). Elements of teacher quality for Islamic education teacher in Malaysia. *International Journal of Special Education*, 37(3), 934-946.
- Mukhtar, N. A., & Noor, M. A. M. (2024). Teachers' performance appraisal: the reality of unified evaluation for education services officer implementation in Malaysian primary schools. *International Business Education Journal (IBEJ)*, 17(1), 26-37. <https://doi.org/10.37134/ibej.Vol17.1.3.2024>
- Mustapa, S. I. S., & Miskon, A. S. H. (2022). Level of teacher teaching and facilitation in one of Langat district secondary schools. *Journal Civics and Social Studies*, 6(1), 90-99. <https://doi.org/10.31980/civicos.v6i1.1735>
- Parveen, K., Quang Bao Tran, P., Kumar, T., & Shah, A. H. (2022). Impact of principal leadership styles on teacher job performance: an empirical investigation. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.814159>
- Rahman, S. M. A., Abd Ghafar, A. L., Nazri, N. A. A., Balian, S. R. C., Abdullah, F., Zawawi, S. A., & Ahmad, S. (2025). Exploring the experiences and challenges of science teachers in implementing practical assessment in Malaysian secondary schools. *International Journal of Academic Research in Progressive Education and Development*, 14(4), 630-643. <https://ijarped.com/index.php/journal/article/view/3881>
- Rashid, Y. K. A., Idris, R., Shukri, N. Z. M., Rabu, M., Ahmad Adha Mohd@Muhamad, S. H. M. N., Sapuan, N. A., Ali, F. D. A., Kamal, N. A., & Sulaiman, M. A. (2024). Embracing innovation: teacher insights and preparedness for i-think thinking maps in 21st century education. *International Journal of Academic Research in Progressive*

Education and Development, 13(3).
<https://ijarped.com/index.php/journal/article/view/2999>

- Rokeman, N. R. M., Kob, C. G. C., & Sobry, H. C. (2023). The role of reward in teachers' job satisfaction towards job performance: A literature review. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 8(11), e002591-e002591. <https://doi.org/10.47405/mjssh.v8i11.2591>
- Sriadmitum, I. (2023). Leadership style, work environment, and compensation on job satisfaction and teacher performance. *Journal of Applied Business and Technology*, 4(1), 79-92.
- Suhaimi, P. F. M., Abdullah, A., Zubairi, A. A., Ali, F. D. A., & Idris, R. (2024). Unveiling the nexus: organizational climate and psychological well-being of secondary school teachers in Kuala Selangor. *International Journal of Academic Research in Progressive Education and Development*, 13(3), 2350–2363.
- Wartenberg, G., Aldrup, K., Grund, S., & Klusmann, U. (2023). Satisfied and high performing? a meta-analysis and systematic review of the correlates of teachers' job satisfaction. *Educational Psychology Review*, 35(4), 114. <https://doi.org/10.1007/s10648-023-09831-4>
- Yusoff, S. M., Marzaini, A. F. M., Zakaria, N., & Lijie, H. (2025). Exploring the implementation of classroom-based assessment in Malaysian secondary schools: alignment with policy and teacher practices. *MOJEM: Malaysian Online Journal of Educational Management*, 13(1), 81-101. <https://doi.org/10.22452/mojem.vol13no1.5>