Sukma : Jurnal Penelitian Psikologi
Juni 2026, Vol. 7, No. 01, hal 203-212

Adaptation and validation of the teacher job performance scale for indonesian teachers

Liliza Agustin[1], Wan Nurul Izza Wan Husin[2]

[1]Faculty of Psychology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia; Faculty of Human Development, Sultan Idris University of Education, Malaysia
[2]Faculty of Human Development, Sultan Idris University of Education, Malaysia
Corresponding author: wannurulizza@fpm.upsi.edu.my

Abstract

This study aimed to adapt and validate the Teacher Job Performance Scale (TJPS) for use among Indonesian teachers. The original TJPS developed by Limon and Sezgin-Nartgün (2020) was selected because it offers a multidimensional framework for assessing teacher performance through task performance, contextual performance, and adaptive performance. The adaptation process involved forward translation, expert review, back-translation, focus group discussion, readability testing, and psychometric validation. Data were collected from 202 in-service teachers in Indonesia and analyzed using Confirmatory Factor Analysis (CFA) with AMOS. The initial scale consisted of 37 items, including 16 task performance items, 9 contextual performance items, and 12 adaptive performance items. Four items (TP1, TP2, TP10, and CP9) showed standardized loadings below .50 and were removed. The final 33-item model demonstrated acceptable factor loadings (.511-.960), high construct reliability (CR = .982), and adequate convergent validity (AVE = .625). The overall model fit was acceptable, with χ^2 = 1993.450, df = 492, χ^2/df = 4.052, RMSEA = .078, and CFI = .900. These findings indicate that the Indonesian adaptation of the TJPS is a psychometrically acceptable instrument for assessing teacher job performance, although further validation using broader and more diverse samples is recommended.

Keywords: Teacher Job Performance Scale; Confirmatory Factor Analysis; scale adaptation; construct validity; Indonesian teachers

Abstrak

Penelitian ini bertujuan untuk mengadaptasi dan memvalidasi Teacher Job Performance Scale (TJPS) agar dapat digunakan pada guru di Indonesia. TJPS yang dikembangkan oleh Limon dan Sezgin-Nartgün (2020) dipilih karena memiliki kerangka multidimensi yang mengukur kinerja guru melalui tiga dimensi, yaitu task performance, contextual performance, dan adaptive performance. Proses adaptasi mencakup penerjemahan, telaah pakar, penerjemahan balik, diskusi kelompok terfokus, uji keterbacaan, dan validasi psikometrik. Data diperoleh dari 202 guru di Indonesia dan dianalisis menggunakan Confirmatory Factor Analysis (CFA) melalui AMOS. Skala awal terdiri atas 37 item, yaitu 16 item task performance, 9 item contextual performance, dan 12 item adaptive performance. Empat item, yaitu TP1, TP2, TP10, dan CP9, memiliki nilai loading terstandar di bawah 0,50 sehingga dikeluarkan dari model. Model akhir dengan 33 item menunjukkan nilai loading yang memadai, yaitu antara 0,511 hingga 0,960, reliabilitas konstruk yang tinggi (CR = 0,982), serta validitas konvergen yang memadai (AVE = 0,625). Indeks kesesuaian model berada pada tingkat yang dapat diterima, yaitu χ^2 = 1993,450, df = 492, χ^2/df = 4,052, RMSEA = 0,078, dan CFI = 0,900. Temuan ini menunjukkan bahwa TJPS versi Indonesia merupakan instrumen yang dapat diterima secara psikometrik untuk mengukur kinerja guru, meskipun validasi lanjutan dengan sampel yang lebih luas masih diperlukan.
Kata kunci: Skala Kinerja Guru; Analisis Faktor Konfirmatori; adaptasi skala; validitas konstruk; guru Indonesia.

Introduction

Teacher job performance is a central determinant of educational quality because teachers directly influence students' learning experiences, classroom engagement, academic progress, and wider school effectiveness. Contemporary education systems increasingly require teachers not only to deliver instruction effectively, but also to collaborate with colleagues, contribute to school development, and adapt to changes in curriculum, technology, and student diversity. Consequently, teacher performance should be measured using instruments that capture the multidimensional nature of teachers' professional roles.

In Indonesia, the measurement of teacher performance requires particular attention because teaching practices are shaped by cultural, institutional, geographical, and policy contexts. Existing evaluation practices often emphasize administrative compliance or general professional appraisal, but they may not sufficiently capture the behavioral dimensions of teacher performance in a psychometrically rigorous manner. This gap is important because instruments that are not culturally and statistically validated may produce inaccurate conclusions about teacher effectiveness and professional development needs.

The Teacher Job Performance Scale (TJPS), developed by Limon and Sezgin-Nartgün (2020), provides a useful framework for teacher performance assessment because it conceptualizes performance through three complementary dimensions: task performance, contextual performance, and adaptive performance. Task performance refers to teachers' effectiveness in carrying out core instructional duties. Contextual performance reflects extra-role behaviors that support the school environment, such as cooperation and participation in school activities. Adaptive performance concerns teachers' capacity to respond flexibly to changes, challenges, and new educational demands.

The TJPS was selected for adaptation because it offers a structured and multidimensional measure that is aligned with contemporary views of job performance, including task-related, social-contextual, and adaptive components. However, direct use of an instrument developed in another linguistic and cultural setting may lead to construct bias or measurement error. Therefore, translation alone is insufficient. A valid adaptation requires linguistic equivalence, conceptual equivalence, cultural relevance, and empirical evidence of reliability and validity.

Accordingly, the present study aimed to adapt and validate the Teacher Job Performance Scale for Indonesian teachers. Specifically, this study examined the factor structure, convergent validity, construct reliability, and model fit of the Indonesian version of the TJPS using Confirmatory Factor Analysis (CFA).

Methods

Research Design

This study employed a quantitative psychometric validation design to adapt and validate the Indonesian version of the Teacher Job Performance Scale. This design was considered appropriate because the main objective of the study was not to test causal relationships, but to examine the validity, reliability, and factor structure of an adapted measurement instrument.

Participants and Sampling

The study involved 202 in-service teachers in Indonesia. Participants were recruited through an online survey using Google Forms. Because the available documentation did not provide evidence of a formal randomization procedure, the sampling procedure is more appropriately described as a non-probability purposive or voluntary sampling approach. The inclusion criteria were that participants were active teachers in Indonesia, were able to understand the Indonesian version of the instrument, and agreed to participate voluntarily.

Table 1 presents the demographic profile of the respondents included in the validation analysis. The demographic distribution covers gender, age, school level taught, highest educational qualification, employment status, teaching experience, and educator certification.

Table 1. Demographic profile of respondents (n = 202)

Variable	Category	n	%
Gender	Male	82	40.6
	Female	120	59.4
Age	20-27 years	31	15.3
	28-35 years	150	74.3
	36-57 years	21	10.4
School level taught	Elementary school (SD/MI)	83	41.1
	Senior high school (SMA/MA)	65	32.2
	Junior high school (SMP/MTs)	54	26.7
Highest educational qualification	Bachelor degree	197	97.5
	Master degree	5	2.5
Employment status	Government employee	12	5.9
	Private employee	190	94.1
Teaching experience	More than 5 years	81	40.1
	5 years	32	15.8
	2-4 years	89	44.1
Educator certification	No	17	8.4
	Yes	185	91.6

Instrument

The primary instrument was the Teacher Job Performance Scale (TJPS), originally developed by Limon and Sezgin-Nartgün (2020). In the present study, the adapted TJPS initially consisted of 37 items distributed across three dimensions. Task Performance (TP) comprised 16 items assessing teachers' core instructional duties, such as lesson planning, classroom implementation, and student assessment. Contextual Performance (CP) comprised 9 items measuring extra-role contributions that support the school environment, such as collaboration with colleagues and participation in school activities. Adaptive Performance (AP) comprised 12 items measuring teachers' flexibility and responsiveness in facing curricular, technological, and contextual changes. The scale has demonstrated strong reliability, with Cronbach's alpha values of .881 for TP, .889 for CP, .937 for AP, and .937 for the overall scale.

All TJPS items were measured using a five-point Likert frequency scale with the following response options: 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Usually, and 5 = Always. Higher scores indicate higher perceived teacher job performance. The instrument does not contain reverse-coded items, and the questionnaire is relatively brief, requiring approximately five minutes to complete. This response format was selected because it enables respondents to indicate the frequency of performance-related behaviors in a clear and graduated manner.

Adaptation Procedure

The adaptation procedure was conducted to ensure linguistic, conceptual, and cultural equivalence between the original TJPS and the Indonesian version. The process included forward translation, expert review, back-translation, focus group discussion, readability testing, and psychometric evaluation.

Table 2

Adaptation and validation procedure

Stage	Activity	Implementer	Outcome
1. Forward translation	The original TJPS items were translated into Indonesian.	Two bilingual translators	Initial Indonesian version of the TJPS.
2. Expert review	The translated items were reviewed for content relevance, clarity, and cultural appropriateness.	Panel of psychological/education experts	Revised items with improved conceptual and cultural equivalence.
3. Back-translation	The Indonesian version was translated back into English and compared with the original version.	Certified translator / bilingual reviewer	Identification and correction of possible meaning shifts.
4. Focus group discussion	Problematic terms and culturally sensitive	Expert panel	Consensus version for readability testing.

	wording were discussed and refined.		
5. Readability test	The adapted instrument was tested for clarity and comprehensibility.	15 teacher respondents	Items were confirmed as understandable or revised when necessary.
6. CFA validation	The adapted instrument was administered to the main sample and analyzed using CFA.	Researcher using AMOS	Final 33-item Indonesian TJPS with acceptable validity and reliability.

Data Analysis

Data were analyzed using Confirmatory Factor Analysis (CFA) in AMOS to test whether the adapted TJPS retained the expected three-dimensional structure. The evaluation focused on standardized factor loadings, construct reliability (CR), average variance extracted (AVE), and model fit indices.

Indicators with standardized factor loadings below .50 were considered weak and were removed from the model. Convergent validity was evaluated using standardized factor loadings and AVE, whereas construct reliability was evaluated using CR. Model fit was evaluated using χ^2, degrees of freedom, χ^2/df, RMSEA, and CFI based on the available AMOS output. Additional indices such as SRMR, GFI, AGFI, and TLI should be reported when available.

Results

Initial CFA and Item Evaluation

The initial CFA was conducted on all 37 items. Most items demonstrated acceptable standardized loadings of .50 or above. However, four items had standardized loadings below the recommended threshold: TP1 (.331), TP2 (.458), TP10 (.085), and CP9 (.425). These items were therefore considered weak indicators of their respective latent constructs and were removed from the model before re-analysis.

Table 3

Initial standardized regression weights for the 37-item TJPS

Item	Factor	Loading	Item	Factor	Loading	Item	Factor	Loading
TP1	TP	.331	TP14	TP	.796	AP2	AP	.867
TP2	TP	.458	TP15	TP	.753	AP3	AP	.871
TP3	TP	.542	TP16	TP	.725	AP4	AP	.847
TP4	TP	.621	CP1	CP	.642	AP5	AP	.938
TP5	TP	.662	CP2	CP	.754	AP6	AP	.942
TP6	TP	.650	CP3	CP	.690	AP7	AP	.960
TP7	TP	.676	CP4	CP	.703	AP8	AP	.924
TP8	TP	.771	CP5	CP	.655	AP9	AP	.900
TP9	TP	.764	CP6	CP	.911	AP10	AP	.809
TP10	TP	.085	CP7	CP	.865	AP11	AP	.893
TP11	TP	.782	CP8	CP	.771	AP12	AP	.741
TP12	TP	.829	CP9	CP	.425			
TP13	TP	.793	AP1	AP	.857			

Note. Items TP1, TP2, TP10, and CP9 were removed because their standardized loadings were below .50.

Final CFA Model

After removing the four weak indicators, the CFA was rerun using 33 retained items. All remaining items showed standardized loadings of .50 or above, ranging from .511 to .960. This result supports the convergent validity of the retained indicators and indicates that the final set of items adequately represents the three dimensions of teacher job performance.

Table 4

Final standardized regression weights for the 33-item TJPS

Item	Factor	Loading	Item	Factor	Loading	Item	Factor	Loading
TP3	TP	.511	TP15	TP	.773	AP2	AP	.867
TP4	TP	.590	TP16	TP	.733	AP3	AP	.870
TP5	TP	.630	CP1	CP	.641	AP4	AP	.847
TP6	TP	.622	CP2	CP	.759	AP5	AP	.938
TP7	TP	.654	CP3	CP	.692	AP6	AP	.942
TP8	TP	.746	CP4	CP	.708	AP7	AP	.960
TP9	TP	.761	CP5	CP	.650	AP8	AP	.924
TP11	TP	.787	CP6	CP	.909	AP9	AP	.900
TP12	TP	.850	CP7	CP	.862	AP10	AP	.810
TP13	TP	.815	CP8	CP	.770	AP11	AP	.893
TP14	TP	.818	AP1	AP	.857	AP12	AP	.741

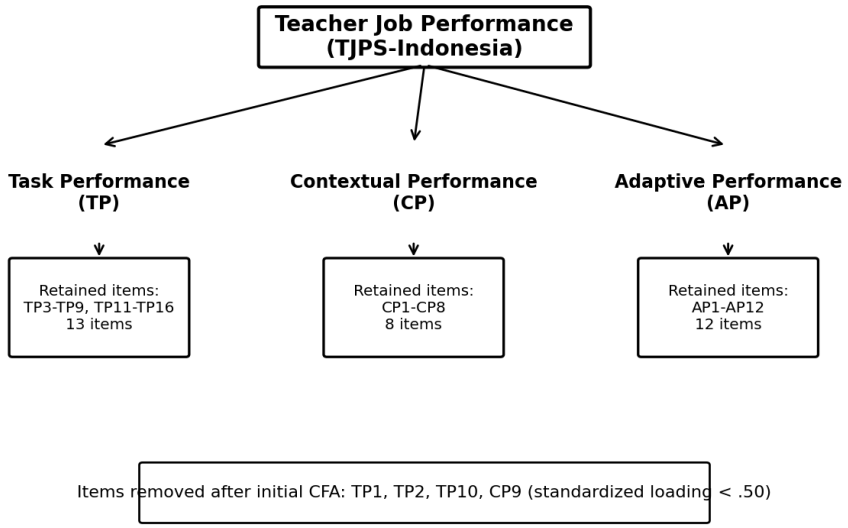


Figure 1. Schematic representation of the final CFA measurement model

Construct Reliability and Convergent Validity

The final scale demonstrated high construct reliability. The construct reliability value was .9818, exceeding the recommended threshold of .70. The AVE value was .6253, exceeding the recommended threshold of .50. These results indicate that the retained items provide consistent measurement and adequate convergent validity for the teacher job performance construct.

Table 5

Construct reliability and average variance extracted for the final TJPS

Statistic	Formula / Basis	Value	Interpretation
Total standardized loading	Sum of retained indicator loadings	25.830	Used to calculate CR
Measurement error	Σ(1 - squared loading)	12.3661	Used to calculate CR and AVE
Construct reliability (CR)	25.83^2 / (25.83^2 + 12.36613)	.9818	High reliability (≥ .70)
Total squared loading	Σ squared loading	20.6339	Used to calculate AVE
Average variance extracted (AVE)	20.63387 / (20.63387 + 12.36613)	.6253	Adequate convergent validity (≥ .50)

Goodness-of-Fit Indices

The final CFA model showed an acceptable level of model fit. Although the chi-square value was significant, this statistic is known to be sensitive to sample size and model complexity. The χ^2/df value of 4.052 was within the commonly accepted threshold of less than 5, the

RMSEA value of .078 indicated reasonable error of approximation, and the CFI value of .900 reached the minimum acceptable level. Overall, the model fit can be described as acceptable rather than excellent, and further refinement may be considered in future studies.

Table 6

Goodness-of-fit indices for the final TJPS model

Index	Recommended criterion	Obtained value	Decision
χ^2 / CMIN	Reported with df and p	1993.450	Significant; interpreted cautiously
df	> 0	492	Reported
p	Preferably > .05; often significant in CFA	.000	Significant
χ^2/df / CMIN/DF	< 5.00	4.052	Acceptable
RMSEA	< .08	.078	Acceptable
CFI	≥ .90	.900	Acceptable / minimum threshold

Discussion

This study adapted and validated the Teacher Job Performance Scale for Indonesian teachers. The findings support the use of a refined 33-item Indonesian TJPS consisting of three dimensions: task performance, contextual performance, and adaptive performance. The retention of these three dimensions indicates that the conceptual structure of the original TJPS is broadly applicable to the Indonesian educational context.

The item refinement process showed that four items, namely TP1, TP2, TP10, and CP9, had factor loadings below .50 and were removed from the final model. Psychometrically, low loading values indicate that these items did not strongly represent their intended latent constructs in the Indonesian sample. Culturally and contextually, the low loadings may suggest that the wording or behavioral content of these items was less aligned with how Indonesian teachers interpret or enact job performance. For example, certain task-related or contextual behaviors may be influenced by local school policies, administrative expectations, communal norms, or differences in teacher role expectations across regions and school types. Therefore, the removal of these items was not merely a statistical decision, but a theoretically defensible refinement to improve construct clarity and cultural fit.

The adaptive performance dimension showed consistently high factor loadings. This may reflect the increasing importance of adaptability among Indonesian teachers, especially in relation to curriculum changes, digital learning demands, diverse student needs, and school-level innovation. In contemporary education, teachers are expected to adjust instructional strategies, respond to technological changes, and manage unpredictable classroom situations. Hence, adaptive performance appears to be a highly salient dimension of teacher performance in the Indonesian context.

The model fit indices indicated that the final 33-item TJPS reached an acceptable level of fit. The model should not be interpreted as perfect; rather, it provides an empirically adequate basis for measuring teacher job performance while leaving room for further refinement. The high CR and acceptable AVE indicate that the scale has strong internal consistency and adequate convergent validity. These findings are consistent with the multidimensional conceptualization of teacher job performance proposed in previous literature, including the distinction between task, contextual, and adaptive aspects of performance.

Compared with the original TJPS, the Indonesian adaptation maintained the core three-dimensional structure but required the removal of several items. This result reinforces the importance of cultural adaptation in measurement validation. Even when an instrument has been validated in its original context, some items may function differently when applied to another cultural and educational system. Therefore, local validation is essential before using an imported scale for research, evaluation, or policy purposes.

Practically, the validated Indonesian TJPS can assist researchers, school leaders, and policymakers in assessing teacher job performance more comprehensively. Rather than focusing only on classroom instruction, the instrument captures teachers' contribution to the school environment and their ability to adapt to change. This broader measurement framework can support professional development planning, teacher appraisal, and evidence-based educational improvement.

Several limitations should be acknowledged. First, the sample size of 202 teachers is acceptable for an initial CFA-based validation study, but larger samples are recommended for future validation, particularly for testing measurement invariance across gender, school level, region, and school status. Second, the data were collected through an online self-report survey, which may introduce response bias or social desirability bias. Third, further studies should calculate additional reliability coefficients such as McDonald's omega, test criterion-related validity, and examine whether the scale predicts external indicators of teacher effectiveness.

Future research should also compare the Indonesian TJPS across different provinces, school types, and teaching levels. Such analysis would strengthen the evidence for the scale's generalizability and help identify whether specific items function differently across cultural or institutional contexts.

Conclusion

The Indonesian adaptation of the Teacher Job Performance Scale demonstrated acceptable psychometric properties after CFA-based item refinement. From the initial 37 items, four items with low standardized loadings were removed, resulting in a final 33-item instrument. The final model showed acceptable factor loadings, high construct reliability, adequate convergent validity, and acceptable model fit. Therefore, the Indonesian TJPS can be considered a valid and reliable instrument for assessing teacher job performance in Indonesia, particularly when interpreted within the context of task, contextual, and adaptive performance. Further validation with broader samples and additional reliability/validity evidence is recommended before large-scale use.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), Field methods in cross-cultural research (pp. 137-164). Sage.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Sage.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. Psicothema, 20, 872-882.

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. International Journal of Testing, 10, 107-132.

Chang, M. C., Shaeffer, S., Al-Samarrai, S., Ragatz, A. B., de Ree, J., & Stevenson, R. (2014). Teacher reform in Indonesia: The role of politics and evidence in policy making. World Bank.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Education Policy Analysis Archives, 8(1), 1-44.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Prentice Hall.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). Multivariate data analysis (9th ed.). Cengage Learning.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. Journal of Applied Testing Technology, 1(1), 1-30.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1-55.

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

Limon, İ., & Sezgin-Nartgün, Ş. (2020). Adaptation of the Teacher Job Performance Scale to the Turkish language and culture. Journal of Education and Research, 34(2), 45-57.

Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. Human Performance, 10(2), 71-83.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. Journal of Applied Psychology, 79(4), 475-480.

OECD. (2018). Effective teacher policies: Insights from PISA. OECD Publishing.