Multidisciplinary Output Research For Actual and International Issue (MORFAI Journal) ISSN (e): 2808-6635
Volumes 6 No. 2 (2026)

ADAPTATION TEACHERS' PERCEPTION GRADING PRACTICE SCALE (TPGPS) INDONESIAN VERSION

Fadillahi Arief[1], Farah Aulia[2]
[1,2]Universitas Negeri Padang
Email: fadillahiarief@student.unp.ac.id, farah_aulia@fip.unp.ac.id

Received	: 29 October 2025	Published	: 28 December 2025
Revised	: 20 November 2025	DOI	: https://doi.org/10.54443/morfai.v6i2.4776
Accepted	: 15 December 2025	Publish Link	: https://radjapublika.com/index.php/MORFAI/article/view/4776

Abstract

This study aims to adapt and conduct the initial validation of the Teachers' Perceptions of Grading Practices Scale (TPGPS) into the Indonesian language. The scale measures teachers' perceptions of grading practices across six components: importance, usefulness, student effort, student ability, teachers' grading habits, and perceived self-efficacy of the grading process. Participants consisted of 130 Indonesian high school teachers who completed the questionnaire online. Data were analyzed using Confirmatory Factor Analysis (CFA) using JAMOVI version 2.3.28 to assess the model fit of the six-factor structure. The results indicated an acceptable model fit with RMSEA = 0.0361 (≤ 0.08), although CFI (0.775) and TLI (0.758) values were slightly below the recommended threshold (≥ 0.90). The overall reliability of the scale was high (Cronbach's α = 0.827), indicating good internal consistency. Several items with low factor loadings are recommended for revision in future validation stages. Overall, the Indonesian version of TPGPS is valid and reliable for assessing teachers' perceptions of grading practices within the Indonesian educational context.

Keywords: Scale Adaptation; Assessment; Teacher's Perception

INTRODUCTION

Reforms in the field of education continue to evolve along with changes in academic policies and standards. In Indonesia, it is reflected in Government Regulation Number 4 of 2022 concerning National Education Standards (SNP). Diverse education delivery practices are influenced by a number of factors, such as the quality of teachers and students, infrastructure, and assessment systems (Totalia et al., 2024). As the main actor in learning, teachers are required to understand and apply SNP, especially in linking the curriculum to an effective learning process (Sharma, 2015). However, some teachers do not understand SNP well because teachers still need a lot of learning to be more skilled and easy to apply assessments (Widyastuti, 2020). One of the efforts made by the government to control the quality of education is to carry out assessments (Ilyafi & Salehudin, 2023). Assessment is a strategic step towards improving the quality of learning outcomes at all levels of education (Andini et al., 2022). Assessment or assessment is all activities that include methods and decision-making on student learning outcomes in a learning environment (Kusainun, 2020).

One of the functions of teachers is as an evaluator, namely conducting an assessment of the results that have been achieved so that they can determine the success of achieving goals, students' mastery of the lesson, and the accuracy or effectiveness of teaching methods (Ulfa, 2023). Education in Indonesia uses the Educational Assessment Standards as a guideline in carrying out student learning outcomes assessments in the form of formative assessments and summative assessments (Noptario et al., 2023). Teachers are the perpetrators of assessment practices that are greatly influenced by their perceptions and beliefs. Adijaya and Zain (2024) explained that teacher perception is an attitude or action taken from the results of a perspective or assessment taken from the results of observation of a certain phenomenon. However, the measurement of teachers' perception of assessment practices in Indonesia is still limited, even though this understanding is important to ensure that the implementation of assessments runs according to standards. This condition requires educational institutions to evaluate and strengthen training for teachers so that their perceptions are aligned with educational goals. Most studies in Indonesia focus more on student learning outcomes or the effectiveness of teaching methods, so aspects

of teachers' perceptions of the assessment system are rarely studied in depth using standardized instruments. In fact, understanding teachers' perceptions of assessment practices is very important as a basis for policy development and teacher training. One of the instruments developed to measure teachers' perceptions of grading practices is the *Teachers' Perceptions of Grading Practices Scale* (TPGPS) compiled by Liu et al. (2006) which has proven to be valid and reliable in various countries. However, this instrument does not yet have an Indonesian version that is culturally adapted. In the context of cross-cultural studies, adaptation is very important because differences in the meaning of items can affect the factorial structure of the instrument and potentially reduce the validity of measurement results. Based on the above, this study aims to adapt linguistics and validate the initial content of the *Teachers' Perceptions of Grading Practices Scale* (TPGPS) into Indonesian.

METHODS

This research was attended by 130 high school/vocational school teachers in West Sumatra. The respondents consisted of 23 males and 107 females with varied fields of study such as English, Indonesian, mathematics, physics, biology, chemistry, geography, sociology, economics, sports, ICT, and PAI. The average age of respondents was 38 years. This study uses a *cross-cultural adaptation study research design*, which is a type of research that aims to adjust the instrument or scale from one culture/country to another, so that the instrument is equivalent in language, equivalent in meaning, culturally equivalent, psychometrically equivalent, until finally feasible for use in new populations without changing the original construct being measured. The instrument of this research is *the Teachers' Perception Grading Practice Scale* (TPGPS) developed by Liu (2006), which is the views, beliefs, and assessments of teachers about how the grading process should be carried out, including the purpose, fairness, method of awarding grades, consideration factors, and their ability to apply the assessment. TPGPS consists of 40 items with six components, namely *importance*, *usefulness*, *student effort*, *student ability*, *teachers' grading habits*, and *perceived self-efficacy of grading process* (perception of self-efficacy in the assessment process). The researcher made adaptations that refer to the steps according to Beaton (1999) starting from translating the original language to the target language to the trial stage in a small number of participants.

The researcher adapts to the steps according to Beaton (1999), the first is to translate the scale into the target language where the original scale is English and will be translated into Indonesian by two translators who have good skills in both languages. After the translation process, a synthesis is carried out to identify the similarities and differences so that a translation is obtained that will be used later. The next stage is *back-translation* to the original language which aims to compare with the original scale whether there is a difference in meaning or not. After that, conduct *expert judgment* with experts such as lecturers and educational psychologists who master the concept of the scale to be prepared. The next stage is a scale test in a small number of subjects to find out whether or not the scale item has been understood by the respondents. After the adaptation process, the researcher will conduct a study with a larger number of respondents to see how the validity, reliability, and CFA are comparable. Data analysis was carried out with CFA (*Confirmatory Factor Analysis*) to examine the factor structure of the TPGP instrument using *the JAMOVI application version 2.3.28.* Data analysis is performed to see if the scale items have matched the specified components. Values that reflect a good construct are *the Goodness of Fit Index* (GFI) $\geq$ 0.90, *the Comparative Fit Index* (CFI) $\geq$ 0.90, and *the Root Mean Square Error of Approximation* (RMSEA) $\leq$ 0.08 (Byrne, 2013). The results of the CFA will show whether the model being tested is acceptable and how well the data obtained reflects the desired structure.

RESULTS AND DISCUSSION

The description of the respondents used in this study is attached in table 1. Based on the table, out of a total of 130 respondents, there were 84 respondents aged 20-40 years (early adult), and 46 respondents aged 41-60 (middle adult). The age division is based on Hurlock's theory (2000). In addition, there were 23 male and 107 female respondents. Furthermore, there are 29 teachers who teach in grade 10, 59 teachers who teach in grade 11, and 42 teachers who teach in grade 12. The subjects taught are English, Indonesian, mathematics, physics, biology, chemistry, geography, sociology, history, economics, PAI, cultural arts, and others.

Table 1. Respondent Demographic Data

Categories	Quantity	Introduce yourself
Age		
20-40	84	64.6%
41-60	46	35.4%
Gender		
Male	23	17.7%
Women	107	82.3%
Classes taught		
10	29	22.3%
11	59	45.4%
12	42	32.3%
Field of Study		
Indonesian Language	16	12.3%
English	16	12.3%
Math	14	10.8%
Biology	19	14.6%
Physics	10	7.7%
Chemistry	8	6.2%
Geography	4	3.1%
Sociology	3	2.3%
Economy	2	1.5%
History	5	3.8%
Cultural Arts	10	7.7%
GOOD	8	6.2%
and more	15	11.5%

Based on the results of the CFA (*Confirmatory Factor Analysis*) conducted on the Indonesian version of TPGPS, the six-component model with 40 items shows various match index values. The chi-squared value (χ^2 = 848, df = 725, $p < 0.001$) is statistically significant due to the large sample size of 130 subjects. However, the χ^2/df ratio of 1.169 (< 3) indicates that the model is acceptable which can be seen in the table below:

Table 2. *Chi-square Test*

X2	d	p
848	725	0.001

In addition, the *values of the Comparative Fit Index* (CFI = 0.775) and *the Tucker-Lewis Index* (TLI = 0.758) were below the recommended threshold (≥ 0.90), indicating that the model is not yet fully compliant with the data, but is still acceptable for the early stages of adaptation scale validation. Then, the *Root Mean Square Error of Approximation* value (RMSEA = 0.0361; 90% CI [0.0241, 0.0460]) met the good match criteria (≤ 0.08), indicating that the model had a low approximation error which can be seen in the table below:

Table 3. *Fit Measures*

			RMSEA 90% CI	
CFI	**TLI**	**RMSEA**	**Lower**	**Upper**
0.775	0.758	0.0361	0.0241	0.0460

Analysis of the loading factor showed that most of the items were significant ($p < 0.05$) with varying estimates, ranging from 0.135 to 0.557. Factor 6 (*Perceived Self-Efficacy*) had the highest loading (e.g., item 38: 0.557) which read "Sorting students by merit when taking assessments was the easiest thing for me", while some

items such as item 17 (Factor 3: *Student Effort*) read "I consider student effort. When I did the assessment," had a relatively low (0.135) and insignificant loading (p = 0.060), so it needed to be considered for revision or removal. It can be seen in the table below:

Table 4. *Loading Factor*

Factor	Indicator	Estimate	OR	95% Confidence Interval		Z	p
				Lower	Upper		
Importance	1	0.248	0.0693	0.11232	0.384	3.58	<.001
	2	0.276	0.0650	0.14904	0.404	4.25	<.001
	3	0.336	0.0700	0.19910	0.473	4.81	<.001
	4	0.230	0.0584	0.11566	0.345	3.94	<.001
	5	0.331	0.0858	0.16256	0.499	3.85	<.001
	6	0.221	0.0685	0.08706	0.355	3.23	0.001
Usefulness	7	0.254	0.0757	0.10538	0.402	3.35	<.001
	8	0.184	0.0706	0.04556	0.322	2.61	0.009
	9	0.341	0.0754	0.19361	0.489	4.53	<.001
	10	0.280	0.0948	0.09383	0.465	2.95	0.003
	11	0.295	0.0631	0.17184	0.419	4.68	<.001
	12	0.304	0.0781	0.15120	0.457	3.90	<.001
	13	0.230	0.0663	0.09984	0.360	3.47	<.001
	14	0.355	0.0773	0.20320	0.506	4.59	<.001
	15	0.316	0.0748	0.16915	0.462	4.22	<.001
	16	0.192	0.0624	0.06946	0.314	3.07	0.002
Student Effort	17	0.135	0.0717	-0.00554	0.276	1.88	0.060
	18	0.332	0.0913	0.15349	0.511	3.64	<.001

Table 4. ***Loading Factor***

Factor	Indicator	Estimate	OR	95% Confidence Interval		Z	p
				Lower	Upper		
	19	0.206	0.0827	0.04434	0.369	2.50	0.013
	20	0.318	0.0885	0.14414	0.491	3.59	<.001
	21	0.250	0.0893	0.07497	0.425	2.80	0.005
	22	0.359	0.0820	0.19793	0.520	4.37	<.001
Student Ability	23	0.373	0.0750	0.22574	0.520	4.97	<.001
	24	0.358	0.0771	0.20670	0.509	4.64	<.001
	25	0.341	0.0626	0.21808	0.463	5.45	<.001
	26	0.353	0.0678	0.22024	0.486	5.21	<.001
	27	0.385	0.0698	0.24849	0.522	5.52	<.001
	28	0.408	0.0958	0.22011	0.596	4.26	<.001
Teachers Grading Habit	29	0.191	0.1259	-0.05568	0.438	1.52	0.129
	30	0.207	0.0906	0.02953	0.385	2.29	0.022
	31	0.276	0.0674	0.14397	0.408	4.10	<.001
	32	0.424	0.1155	0.19788	0.651	3.67	<.001
	33	0.331	0.0678	0.19759	0.464	4.87	<.001
	34	0.274	0.0849	0.10812	0.441	3.23	0.001
Perceived self-efficacy of grading process	35	0.530	0.1179	0.29941	0.762	4.50	<.001
	36	0.420	0.1093	0.20631	0.635	3.85	<.001
	37	0.509	0.1038	0.30519	0.712	4.90	<.001

Table 4. *Loading Factor*

Factor	Indicator	Estimate	OR	95% Confidence Interval		Z	p
				Lower	Upper		
	38	0.557	0.1142	0.33325	0.781	4.88	<.001
	39	0.322	0.1285	0.07015	0.574	2.51	0.012
	40	0.287	0.1151	0.06100	0.512	2.49	0.013

In addition, the overall reliability of the TPGP scale (*Cronbach's α* = 0.827) is relatively good, although some items such as items 17 and 29 have a low item-total correlation (< 0.20), as in the table below:

Table 5. *Scale Reliability Statistics*

	Red	SD	Cronbach's α
Scale	3.98	0.300	0.0241

The results of the *Confirmatory Factor Analysis* (CFA) validation show that the Indonesian version of the six-factor TPGPS model has a fairly good level of compatibility. The results showed an RMSEA (0.0361) that met the match criteria (≤ 0.08), indicating that the model had a low rate of prediction errors. However, the values of CFI (0.775) and TLI (0.758) are still below the recommended standard (≥ 0.90). This indicates that the model may require modifications, such as the removal of items with low loading factors or the addition of residual covariance to improve compatibility (Byrne, 2013). The overall scale has good reliability (α = 0.827), but some items (e.g., Item 17 on *Student Effort* with a load of 0.135, p = 0.060) show a weak contribution to the construct. In addition, *the item-rest correlation* for some items (such as Item 29) is very low (<0.20), suggesting that they may not measure the same construct with the expected consistency. Revelle (2023) says that if an item has an item-total correlation below 0.30, then it needs to be considered for removal or revision.

DISCUSSION

The results of this study show that the adaptation of *the Teachers' Perceptions of Grading Practices Scale* (TPGPS) into Indonesian has adequate construct validity. An RMSEA value of 0.0361 indicates a low model approximation error rate, which according to Byrne (2013) is a strong indicator that the model structure is acceptable. These findings are in line with the research of Liu et al. (2006; 2011) which showed that TPGPS consistently has a stable six-factor structure in various cultural contexts such as the United States and China. Nonetheless, the CFI and TLI values (0.775 and 0.758) have not yet reached the ideal limit ≥ 0.90. This phenomenon is common in cross-cultural adaptation due to differences in the interpretation of items and assessment habits between countries (Beaton et al., 1999). This condition is consistent with the findings of Liu (2011), who reported that the adaptation of TPGPS in different cultural contexts can result in variations in fit models due to differences in item interpretation. Similar findings are also shown in the adaptation of TPGPS to Turkish where some items need to be revised to fit the local cultural context (Özkan et al., 2022).

Research by Zhang and Burry-Stock (2003) also found that teachers in Asia often respond differently to assessment items due to differences in norms, so the initial fit model tends to be lower. A similar thing emerged in the study of the adaptation of teacher assessment instruments by Ilyafi & Salehudin (2023) in Indonesia, which also reported the need for linguistic revisions on certain items to be more in line with the national education context Some of the items in the study, such as items 17 and 29, showed low *loading factors*. Revelle (2023) asserts that items with a low item-total correlation need to be revised as they have the potential to interfere with construct strength. This strengthens the need to revise items that are not in accordance with teacher assessment habits in Indonesia. In terms of reliability, *Cronbach's alpha* value of 0.827 indicates good internal consistency

and is above the minimum limit. Thus, the Indonesian version of TPGPS has the potential as a diagnostic instrument to map teacher assessment practices, help schools identify areas that need training, and support the implementation of more accountable assessments. However, this study has some limitations. The sample only involved teachers from West Sumatra, so the generalization of the findings needs to be done carefully. Hair et al. (2019) recommend a larger and more diverse sample to strengthen the validity of the construct. Based on these limitations, further research is recommended to involve a sample of teachers from various provinces and levels of education, conduct advanced validity analyses such as AVE, CR, and invariance measurements, and revise items that have low loading so that the structure of the Indonesian version of TPGPS factors becomes stronger and more in line with the cultural context of Indonesian teachers.

CONCLUSION

This study successfully adapted and tested the initial validity of *Teachers' Perceptions of Grading Practices* (TPGPS) developed by Liu (2006) into Indonesian. The results of *the Confirmatory Factor Analysis* (CFA) showed that the six-factor model consisting of *importance, usefulness, student effort, student ability, grading habits,* and *perceived self-efficacy* had a fairly good model match, characterized by a low RMSEA value and most items showed significant factor loading. The reliability of the scale is also in the high category with *Cronbach's value* α = 0.827, which indicates good internal consistency. Although there are two items that show low and insignificant contributions, in general, the Indonesian version of TPGPS can be declared constructively valid and reliable as a measure of teachers' perception of assessment practices in the context of Indonesian education. These findings provide a solid basis for the use of TPGPS in research and evaluative practice in education, and open up opportunities for the development of advanced instruments through external validity tests and trials in a wider population.

REFERENCES

Adijaya, N., & Zain, M. Y. (2024). Teachers' Perception of the Learning Process in the Post-Pandemic Era. *Lantern Scientific Journal of Education*, 17(1), 11–22. https://doi.org/10.52217/lentera.v17i1

Beaton, D. E. (1999). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Acta Odontologica Scandinavica*, 57(4), 225–230. https://doi.org/10.1080/000163599428823

Byrne, B. M. (2013). Structural Equation Modeling With AMOS. Structural Equation Modeling With AMOS, July. https://doi.org/10.4324/9781410600219

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). Multivariate Data Analysis (8th ed.). England: Pearson Prentice.

Ilyafi, M. N., & Salehudin, M. (2023). Teachers' Perceptions in the Implementation of National Assessment (AN) in Samarinda. 3(5), 224–232. *Journal of Instructional and Development Researches.* DOI: 10.53621/JIDER. V3I5.112

Kusainun, N. (2020). Analysis of Education Assessment Standards in Indonesia. 5, 1–7. *Journal of Education*. https://doi.org/10.26740/jp.v5n1.p%25p

Liu, X., O'Connell, A.A., & McCoach, D.B. (2006). The initial validation of teachers' perceptions of grading practices. Paper presented at the 2006 Annual Meeting of American Educational Research Association (AERA).

Liu, X. (2011). Assessing Measurement Invariance of the Teachers' Perceptions of Grading Practices Scale across Cultures. *International Journal of Humanities and Social Sciences*, 1(20), 70–80.

Noptario, Najib, M., Aisyah, S., & Shaleh. (2023). Analysis of Education Assessment Standards in Indonesia (Permendikkbud No. 21 of 2022). 9(June), 380–388. https://doi.org/10.5281/zenodo.8088817

Özkan, Y. O., Guvendir, M. A., Guvendir, E. (2022). Adaptation Of Teachers' Perceptions Of Grading Practices Scale To Turkish And Examination Of Measurement Invariance. *International Journal of Assessment Tools in Education*. 9, Special Issue, 300–316. https://doi.org/10.21449/ijate.1115317

Revelle, W. (2023). Package ' psych ' Procedures for Personality and Psychological Research. Northwestern University, Evanston.

Sharma, P. (2015). Standards-based Assessments in the Classroom : A Feasible Approach to Improving the Quality of Students' Learning. https://doi.org/10.1177/0973184914556864

Totalia, S. A., Sujadi, I., Aniek Hindrayani, Susantiningrum, Sigit Wahyudi. (2024). Teachers' Perception and Meaning of Process Standards and Educational Assessment Standards. *Journal of Education Policy Research*, 16(2), 81–92. https://doi.org/10.24832/jpkp.v16i2.726

Widyastuti, A. (2020). Teachers' Perceptions of the Concept of Independent Learning Minister of Education and Culture Nadiem Makarim in Islamic Religious Education at MTs Negeri 3 Sleman. Thesis. Islamic Religious Education Study Program, Faculty of Islamic Religious Sciences, Islamic University of Indonesia Yogyakarta

Zhang, Z & Burry-Stock, J.A. (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement In Education, 16(4), 323–342*. https://psycnet.apa.org/doi/10.1207/S15324818AME1604