



Smartphone application for assessing teacher performance

Hesty Marlianawati*; Suranto; Badrun Kartowagiran

Universitas Negeri Yogyakarta, Indonesia

*Corresponding Author. E-mail: hestymw@gmail.com

ARTICLE INFO

Article History

Submitted:

4 August 2022

Revised:

2 March 2023

Accepted:

8 September 2023

Keywords

development; application; evaluation; teacher performance; smartphone

Scan Me:



ABSTRACT

This work seeks to design a smartphone application for assessing teacher performance utilizing a 4D development model modified with the Mardapi instrument development model of (1) define (defining); (2) design (deciding instrument requirements, determining application media specifications, creating flowcharts, and creating storyboards); (3) develop (assemble instruments, code, execute tests, evaluate test results, and interpret measurement findings); and (4) disseminate (spread). Proof of the application's viability was established through product testing with 26 principals acting as application users and evaluations by material and media specialists. The content validity of the acquired results was evaluated using the Aiken method, and the construct validity was evaluated using PLS (Partial Least Squares) analysis because the sample size does not have to be large, in accordance with this study where the number of samples is not so large. Using Cronbach's alpha to evaluate the reliability. The results demonstrated that the built application was "possible" as proven by the existing instruments in the valid and trustworthy application. This is proven by the fact that the content validity test utilizing the Aiken method yielded good validity results, with a mean score of 0.80 for material experts and 0.75 for media experts. The construct validity test using PLS (Partial Least Squares) yielded positive results for validity. The estimation of reliability using Cronbach's alpha yields positive findings. Therefore, it can be concluded that the developed application for teacher performance evaluation possesses the features of a valid and trustworthy assessment instrument.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



To cite this article (in APA style):

Marlianawati, H., Suranto, S., & Kartowagiran, B. (2023). Smartphone application for assessing teacher performance. *REID (Research and Evaluation in Education)*, 9(2), 118–129. <https://doi.org/10.21831/reid.v9i2.52384>

INTRODUCTION

It is essential to conduct teacher performance evaluations in order to generate a stronger education. Evaluation of teacher performance is necessary for controlling and monitoring teacher growth. Teacher performance assessment is the evaluation of instructors in schools based on each action as the principal responsibility of being a teacher for rank, position, and career advancement ([Regulation of the Minister of Empowerment of State Apparatuses and Bureaucratic Reform No. 16 of 2009](#)). Teachers' ability to demonstrate their professionalism as educators will be bolstered by optimal teacher performance. Due to the importance of assessing teacher performance in improving the quality of education, it is necessary to conduct continuous teacher performance assessments in order to obtain up-to-date information that can describe the level of teacher performance professionally so that the results are immediately known, thereby accelerating the necessary follow-up. In accordance with [Sinaga et al. \(2021\)](#), the teacher performance appraisal system is an evidence-based assessment system for evaluating teachers' professional performance ([Sinaga et al., 2021](#)). The principle is the authorized party for evaluating a teacher's performance and accomplishments, with daily access to the teacher's performance data.

As a manager, the principal is responsible for overseeing not just the curriculum and lesson materials, but also teacher resources, administrative personnel, assets, and agency finances.

According to the Technical Instructions for Assigning Teachers as Principals, school principals are responsible for implementing enhancements in the areas of entrepreneurship, management, and supervision of education personnel, including teachers. Therefore, the principal must possess at least three professional, personal, and managerial skills ([Regulation of the Director General of Teachers and Educational Personnel of the Ministry of Education, Culture, Research, and Technology No. 5958/B/HK.03.01/2022](#)). Leadership is the most essential aspect of management, hence the ability to lead influences the achievement of agency objectives ([Saleh, 2016](#)). Thus, the function of the principal is crucial in determining the success of an agency/organization.

Assessment is one of the principal's primary responsibilities in determining the effectiveness of an institution. Assessment is an integral component of determining the efficacy of the ongoing learning process. Evaluation is the process of assigning a value to an object based on a set of criteria ([Sudjana, 2010](#)). Effective evaluation must be conducted in accordance with the criteria of the object being evaluated. In order to achieve an effective teacher performance evaluation, an assessment tool or instrument is required. The goodness of the tool is determined by its reliability and validity. In addition, an instrument is seen to be of high quality if it is simple, effective, and efficient to use. An instrument is a device used to collect data on a variable to evaluate a specific object. Instruments can measure an object if they have been tested for validity and reliability and are arranged systematically in accordance with existing norms ([Djaali et al., 2000](#)). The instrument will be utilized effectively to assess teacher performance if it can be used effectively and efficiently and if the assessment procedure can be conducted with flexibility in terms of time and location.

Instruments will be effective if a medium that can be used for the assessment process is created. Instruments are described as instruments used to facilitate the effective and efficient completion of activities ([Utami et al., 2021](#)). However, the use of instruments for evaluating teacher performance is still inadequate, making it ineffective and inefficient. Typically, teacher evaluations are still conducted traditionally using paper (paper-based tests). The implementation of traditional paper-based assessment has numerous flaws ([Istiyono et al., 2020](#)). One of the drawbacks of paper as a media instrument is that it necessitates the procurement of a great deal of paper logistics, resulting in significant prices. Reducing paper consumption can cut expenditures on paper and other supporting materials ([Hamid, 2016](#)). Occasionally, printed or photocopied evaluation findings are less legible and of poor print quality. In contrast to the use of display technology, the alternative will be more appealing and readily apparent, and duplication will be free of charge, so that the findings of teacher performance evaluations can be displayed and implemented in a more efficient and practical manner.

By developing and implementing applications that make it easier to analyze teacher performance, cellphones can be utilized as one of the supporting instruments for evaluating teacher effectiveness. Smartphone technology is currently highly helpful in terms of speed, accuracy, ease, and cost ([Suranto, 2018](#)). Smartphones offer the benefit of providing users with convenience. The use of teacher performance appraisal applications on the Android operating system has several benefits, including: (1) easing the process of conducting teacher performance evaluations; (2) allowing teachers to view the results of performance evaluations obtained; and (3) facilitating the monitoring process that is conducted by the head of the performance evaluation procedure for all teachers ([Pradana & Kuswinardi, 2020](#)). Thus, the principal will find it more convenient if the teacher performance evaluation is packaged as a smartphone application. The evaluation process may be conducted anywhere and at any time, making it more effective, efficient, and adaptable. Through the application's available capabilities, the findings of the collected teacher performance evaluation will also be readily available. Based on the topic of the problem, it is necessary to develop a performance appraisal system for teachers using a smartphone-based application that can make the teacher performance appraisal process more effective and efficient, thereby making it easier for school principals to use as users or assessors.

METHOD

This study used the 4D (four-D) development model modified by the instrument development model created by Mardapi. After modifications, the development flow consists of the following stages: (1) define stage, (2) design stage consisting of determining instrument specifications, determining application media specifications, creating flowcharts, and creating storyboards, (3) developing stage consisting of assembling instruments, coding, testing, analyzing test results, and interpreting measurement results, and (4) disseminate stage.

This survey included 26 principals of public elementary schools in the district of Gringsing. The sample method for this study was chosen based on the basic random sampling method. In this study, two lecturers in Educational Research and Evaluation and two lecturers in Electrical Engineering Education served as validators to validate the developed content and product instruments (expert judgment). This study's data collection method was a questionnaire. A questionnaire is a collection of written questions used to do research (Arikunto, 2016). Two questionnaires are required for the research: (1) a questionnaire for material experts, which generates data in the form of product feasibility on the contents of the instrument according to the standard; and (2) a questionnaire for media experts, which generates data from the product feasibility of a smartphone-based teacher performance assessment.

This study's data collection method was a questionnaire. Two questionnaires are required for the research: (1) a questionnaire for material experts, which generates data in the form of product feasibility on the contents of the instrument according to the standard; and (2) a questionnaire for media experts, which generates data from the product feasibility of a smartphone-based teacher performance assessment:

$$V = \frac{\sum s}{[n(c - 1)]} \quad (1)$$

The information in Equation (1) shows that the value of s is obtained from the result of subtracting r by l_0 , the value of l_0 is obtained from the lowest validity rating, the value of c is obtained from the highest grade for reliability, the value of r is obtained from the score awarded by the evaluator, and the value of n is expert judgment.

Two material specialists and two media specialists conducted the content validity tests. The validity procedure is conducted by assessing the questionnaire items that have been produced; if there are flaws or defects, adjustments are then made. The validation outcomes from the experts serve as an evaluation and input. The inputs and ideas were analyzed qualitatively as material for the enhancement of the developed instrument. The quantitative assessment findings generate test items using the Aiken validity index. According to Aiken (1985), the content validity criteria is met by instruments scoring 0.73 (Aiken, 1985). The validity coefficient of Aiken is calculated using the raw scores of n experts (Bashooir & Supahar, 2018). Aiken's validity analysis is performed to determine whether the generated instrument's items are consistent with the identified indicators or factors (Andrian, 2019).

The purpose of the construct validity test is to determine whether the instrument can be utilized to measure and observe a research or evaluation. This developer's construct validity is evaluated using PLS (Partial Least Squares). In addition to being used to confirm the theory, PLS can also be used to explain whether or not there is a relationship between latent variables. PLS can examine constructs produced using both reflexive and formative indicators, but SEM (Structural Equation Model) cannot. In addition to the aforementioned reasons, the construct test employing PLS has the advantage of being able to handle many independent variables despite multi-collinearity between the independent variables. In addition, the sample size does not need to be vast or many, consistent with this study's small sample size.

Reliability is referred to as the consistency of outcomes, which indicates that the results will remain the same when applied to the same variable at different times. Reliability refers to an instrument that can be relied upon and utilized consistently or again (Cohen et al., 2007). This indicator's

reliability measures the dependability of the measurement indication for the latent variable. The stages are determined by analyzing the outcomes of each indicator's outer loading. A loading score of 0.7 suggests that the construct can explain more than fifty 50% of the variance in the indicator (Kwong & Wong, 2013; Sarstedt et al., 2017). Cronbach's alpha is used as the approach for measuring reliability. A satisfactory Cronbach's alpha value is more than 0.7 (Ghozali & Laten, 2015), considering 0.85 to be a respectable dependability value (Mehrens & Lehmann, 1973). The greater the reliability rating (closer to 1), the more convincing it is that the instrument is dependable or consistent. The values of dependability can be classified in Table 1.

Table 1. Reliability Value Range

Value	Information
0.00 – 0.39	Low
0.40 – 0.69	Medium
> 0.70	High

Testing the effectiveness is done by giving a questionnaire to application users. The questionnaire uses a Likert scale with a choice of five response scales. Users are asked to provide answers from the choices of items that are already available. The data obtained were analyzed to determine the level of effectiveness of the developed application. The scores obtained are totaled from each indicator score. The data obtained were described using data frequency analysis techniques and set criteria by categorizing scores.

FINDINGS AND DISCUSSION

Define

Define is a stage of need analysis in which literature studies and research observations are used to establish what products are generated. The researchers were tasked with gathering the necessary information by analyzing the theory and prior research to determine the characteristics of the teacher performance assessment instrument. This is in line with the opinion of Fauziyah and Triyono (2020), who stated that the formulation of indicators for learning outcomes or performance is the basis for determining what should emerge in the developed product.

Design

Specification of Instrumentation

Table 2. Blueprint of Performance Evaluation Device

Aspect	Indicator	Item
Pedagogic	a. Mastering the learning conditions of students	1
	b. Mastering instructional materials in accordance with educational learning principles	2
	c. Curriculum development	1
	d. The educational learning activities process	1
	e. Development of student potential	1
	f. Relationships with students	1
	g. Evaluation and evaluation	3
Personality	a. Conform to country religious, legal, social, and cultural standards	3
	b. Exhibit a mature and outstanding demeanor	3
	c. Being a teacher requires a strong work ethic, a high level of responsibility, and a sense of pride	4
Social	a. Be inclusive, act objectively, and do not engage in discrimination	3
	b. Communication with fellow educators, education personnel, parents, students, and the community	7
Professional	a. Mastery of the materials, structures, concepts, and scientific perspectives supporting the taught subjects	6
	b. Develop your professionalism through introspection	4

The designed instrument for the teacher performance evaluation application consists of four aspects: pedagogical, personality, social, and professional. In accordance with the study, four areas of teacher performance were examined: personality, pedagogical, social, and professional (Yuning-sih & Sunhaji, 2021). A Likert scale is employed as the measuring device. There are five answer options on the Likert scale, with the highest score being 5 and the lowest being 1. There were 14 indicators and 40 statement items generated for the four factors. The pedagogical statement components include ten statements, ten statements for personality, ten statements for social, and ten statements for professionals. Table 2 depicts a blueprint of smartphone-based instruments for assessing the performance of teachers. This instrument has four components.

Defining Application Media Specs

Smartphones with the following characteristics can run this application: screens ranging from four to six inches, a minimum of 500 MB of RAM, a minimum of 1 GHz processor capacity, a minimum operating system of Android 5.0 (Jelly Bean), and a minimum of 50 MB of storage space.

Developing Flowcharts

The flowchart is a diagram with visual symbols that describes the direction of data activities that will be generated as part of a software process. At this stage, researchers create a flowchart to describe the application development process.

Constructing Storyboards

A storyboard is a way of describing an outline in the form of a narrative, consisting of a sequence of images, sketches, and words that create a tale.

Develop

Putting Together the Instrument

Using 14 specified indications, the researchers compiled 40 items for the teacher performance evaluation application tools. Those items are presented in Table 3.

Doing Coding

At this stage of coding, the researchers were aided by the development team, which consisted of two people: the author as a drafter, and a developer who was an expert in creating Android Studio applications in accordance with the author's concept. During the coding process, the development team coordinates regarding the application being created. There were a number of features that could not be executed as intended, which necessitated a number of process enhancements and modifications. Occasionally, debugging occurs throughout the process of developing an application. Debugging is the process of locating and removing mistakes or bugs in code that might cause a system to malfunction or crash.

Doing the Test

The instrument in the created teacher performance assessment application was subjected to a feasibility test via a content validity test conducted by subject matter experts and media experts. The purpose of the validation performed by material and media experts is to examine the validity of the content. Through validation by material and media specialists, improvement-related recommendations and inputs are received. Two lecturers from the Education Research and Evaluation Study Program as material experts and two lecturers from the Education Technology Study Program as media experts validated the content of the smartphone-based teacher performance assessment application. After validation by specialists, a product trial was conducted with 26 elementary school principals from the Gringsing District as users. Principals were asked to evaluate

the performance of teachers in schools using the developed teacher performance appraisal application during the product trial.

Table 3. Instrument for Teacher Performance Evaluation Application Tools

Aspect	Item
Pedagogic	1. Teachers are able to condition learning students in class
	2. Teachers teach according to their fields
	3. The teacher teaches the material without reading directly from the textbook
	4. The teacher associates learning with real examples in everyday life through learning media
	5. Teacher makes an Implementation Plan Learning
	6. The teacher guides students who have difficulty in the process Learning
	7. The teacher communicates with spoken language politely.
	8. Teachers have complete instruments (questions, keys, and guidelines, scoring)
	9. The teacher reflects on the learning that has been implemented
	10. The teacher gives feedback on the assigned job
Personality	11. Teachers have noble characters that can be used as an example
	12. Teachers guide students to be obedient to worship
	13. Teachers follow the rules applied at school
	14. The teacher has authority in front of the participants Educate
	15. The teacher puts the public interest first in decision-making
	16. The teacher is calm in dealing with conflicts among coworkers
	17. The teacher comes to school on time
	18. Teachers can complete tasks assigned by the principal
	19. Teachers can complete the work according to their duties well
	20. Teachers have an exemplary attitude as educators
Social	21. The teacher shows a positive attitude towards differences of opinion
	22. Teachers behave and behave honestly
	23. The teacher does not differentiate treatment towards students and between co-workers
	24. The teacher helps colleagues in solving problems
	25. The teacher builds a good relationship with fellow teacher
	26. The teacher builds a good relationship with the principal
	27. Teachers build good relationships with parents of students through discussion forums between teachers and parents
	28. Teachers build good relationships with students at school
	29. Teachers build good relationships with the community
	30. Teachers follow community organizations in the environment
Professional	31. Teachers master the learning material
	32. The teacher chooses the material according to the learning objectives
	33. The teacher carries out the learning process coherently
	34. Teachers carry out learning according to the time allocation
	35. The teacher teaches according to the concept of the material lessons to be taught
	36. The teacher includes a summary of the material in the lesson plans that have been made
	37. Teachers use the internet to support learning materials
	38. The teacher makes teaching aids/media
	39. Teachers attend seminars/training/ workshops to increase knowledge
	40. Teachers make writings in the field of Education

Analyzing the Test Results

In this feasibility test, the Aiken formula is utilized for content validity analysis. The content validity criteria are met by instruments scoring 0.73 (Aiken, 1985). The analytical results from material specialists are 0.80, while those from media experts are 0.75. The analysis revealed that the instrument satisfied the minimum requirement of validity. According to the findings of the material expert's investigation, the developed instrument was suitable for use after undergoing many revisions based on the material expert's suggestions. The following changes must be made: "evaluation" should be replaced with "assessment"; point 1 of the pedagogical section should read "Teachers are able to condition the learning of students in their class"; student writing must be consistent; point 27 of the social section should read "Teachers foster good relations with parents of students through discussion forums between teachers and parents"; Enhancements to conceptual and operational definitions; references from the preceding five years included. When analyzing an individual's identification by writing, "competence" becomes "aspect", and "full name" be-

comes “name.” Regarding the pedagogical component, point 4 is altered to read “The instructor connects learning to real-world experiences.”; In the pedagogical aspect, point 5 was modified from “Teachers create lesson plans” to “Learning Implementation Plans”; in the professional component, seminar was added to “seminar/training/workshop” in point 39. The out-comes of this investigation are then used to enhance the design of the manufactured instrument. Start writing and developing applications using the *Android Studio Software* once the instrument has been deemed fit for usage.

The generated application is next reviewed by media specialists to determine the viability of content validity. In accordance with the validator’s recommendations, a number of modifications must be done to the produced application before it is suitable for use, as indicated by the study’s findings. The following changes must be implemented: alternative answers must be made consistent; each item of the statement must be assigned a number; the assessment results must be presented once the “complete” button is clicked; developer identification must be added. The out-comes of these studies are utilized to enhance applications created prior to product testing.

Using the Aiken formula, test the validity of the content in this development. According to Aiken, the content validity criteria is met by instruments scoring 0.73 (Aiken, 1985). According to the findings of the computation, the coefficient of validity for each expert is as shown in Table 4.

Table 4. Validity of Content Calculation

No.	Expert	Aiken V	Information
1.	Material	0.80	valid
2.	Media	0.75	valid

According to the analysis, the score from material experts is 0.80 while the score from media experts is 0.75. Referring to the Aiken table, where items are evaluated by two raters using a 5-point scale and a 5% error rate, the minimal v value is 0.040. Therefore, depending on the established criteria, the minimal V Aiken standard for this study is 0.73 with a probability of 0.040. Therefore, it can be stated that the content validity of the instrument on the smartphone-based teacher performance assessment application can be used for assessment.

It was assessed with positive results based on the results of validation utilizing Aiken from material specialists and media professionals. The outcomes of the product testing were then evaluated for their validity and reliability utilizing PLS. The PLS test on the *smartphone*-based application for evaluating teacher performance analyzes four characteristics of teacher performance in phases. The construct validity test is determined using PLS, which is used to determine if the construct has adequate validity, i.e., the loading value of the target construct criteria must be greater than other constructs. As a result, the indicator is deemed reliable for measuring the pertinent variables.

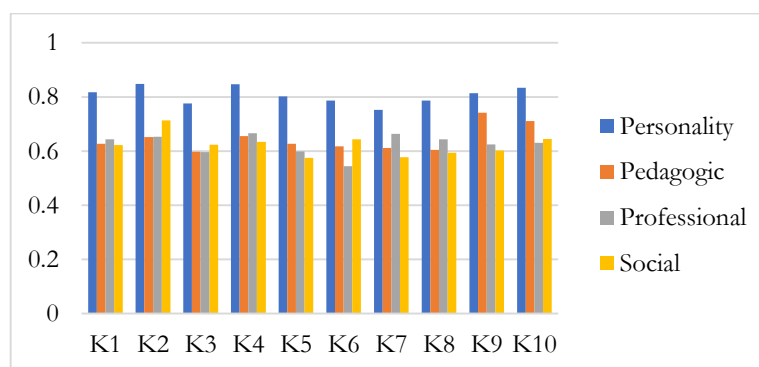


Figure 1. Results of Personality Variable Tests

On the basis of [Figure 1](#), it can be seen that each sub-indicators variables have a stronger cross loading on their own sub-variables than on other sub-variables; this indicates that all indicators meet the validity criteria. In addition, based on [Figure 2](#), each sub-indicators variable has a higher cross loading on their own sub-variables than on other sub-variables; this implies that the indicator has met the validity criteria. Meanwhile, according to [Figure 3](#), the indicators for each sub variable have a stronger cross loading on their own sub variables than on other sub variables, indicating that all indicators meet the validity standards.

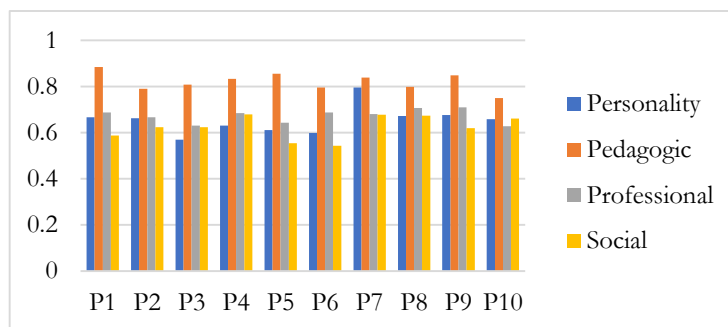


Figure 2. Pedagogic Variable Test Results

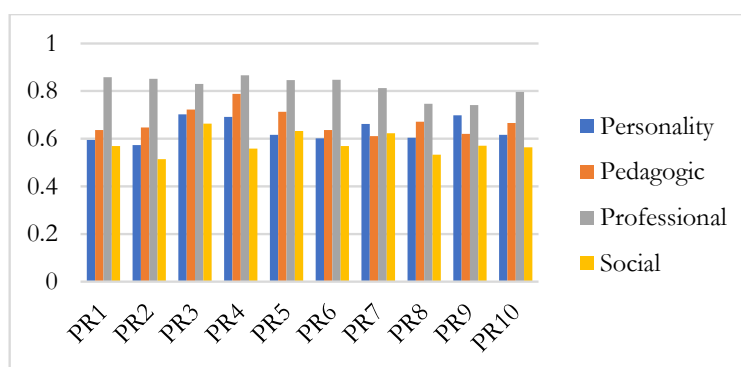


Figure 3. Variable Professional Test Results

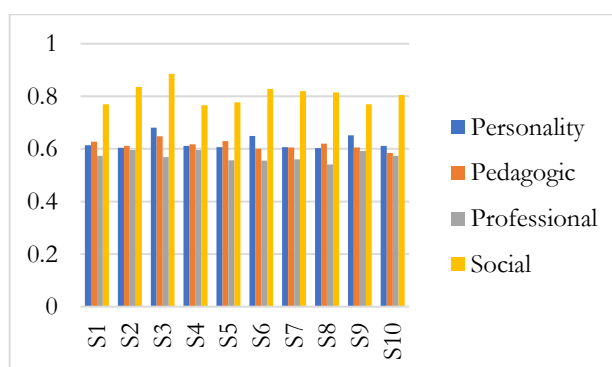


Figure 4. Social Variable Test Results

According to [Figure 4](#), the indicators for each sub variable have a stronger cross loading on their own sub variables compared to other sub variables, indicating that all indicators have good validity criteria. Validity can be determined using Fornel-Larcker criteria as well as loading and cross loading indicators. In the hypothetical model study, the Fornel-Larcker criteria were determined by comparing the AVE root of each construct to the correlation between each construct. If the results of the Fornel-Larcker criterion calculation indicate that the root of the AVE sign for each construct is bigger than the sign correlation between one construct and another, as shown in [Table 5](#), the research model is regarded to have very good validity.

Table 5. Fornel-Locker

	Personality	Pedagogic	Professional	Social
Personality	0.807			
Pedagogic	0.800	0.821		
Professional	0.777	0.820	0.821	
Social	0.773	0.761	0.707	0.808

Note. AVE's root value is the value in the diagonal direction.

According to Table 5, the findings of the PLS on the application for the teacher performance assessment indicate that all aspects of the teacher performance evaluation either meet the standards or already have a high *Goodness of Fit*. Comparing the AVE score and the correlation between the constructs reveals validity. The validity test yielded scores of 0.821 for the pedagogical aspect, 0.807 for the personality aspect, 0.808 for the social aspect, and 0.821 for the professional aspect. When the AVE root value is more than the correlation value of each concept, validity is regarded to be high. With a correlation value of 0.820 and an AVE root value of 0.821, it can be concluded that the pedagogic variable has high validity. With a correlation value of 0.800 and an AVE root value of 0.807%, the personality variable may be deemed to have good validity. With a correlation value of 0.773 and an AVE root value of 0.808, the validity of the social variable can be considered to be strong. Lastly, the professional variable has the greatest correlation value of 0.820 and the highest AVE root value of 0.821, indicating that it has good validity. This is in line with what was stated by Hair et al. (2021) that the variance of the constructs among all model constructs should not be greater than the value of AVE, and it indicates good discriminant validity. On the basis of the test findings, it can be concluded that the instrument in the teacher performance assessment application has good validity and is suitable for use as a pedagogical, personality, social, and professional sub variable teacher performance assessment instrument.

Cronbach's alpha is used to determine the test results after fit, in line with the result of research conducted by Retnowati et al. (2017) on evaluating reliability using *Cronbach's alpha*. It is possible to estimate dependability using *Cronbach's alpha*. *Cronbach's alpha* greater than 0.6 indicates that the requirements for a construct are dependable and have good reliability.

Table 6. Results of Reliability Estimation

Level	Cronbach's Alpha	Category
Personality	0.940	Good
Pedagogic	0.946	Good
Professional	0.946	Good
Social	0.940	Good
Teacher Effectiveness	0.979	Good

According to Table 6, the dependability coefficient for the personality component is 0.940 (> 0.6), for the pedagogical aspect it is 0.946 (> 0.6), for the social aspect it is 0.940 (> 0.6), and for the professional aspect it is 0.946 (> 0.6). Cronbach's alpha values greater than 0.6 indicate that the construct criteria are reliable. Each sub variable of personality, pedagogy, professional, social, and teacher performance aspects had Cronbach's alpha values greater than 0.6. This indicates that the manufactured instrument is reliable. This is in line with previous research findings that indicate that a reliability coefficient of an instrument greater than 0.6 can be considered reliable (Jong, 2019). Therefore, it can be argued that all indicators accurately measure the pedagogical, personality, social, and professional dependent factors.

The results of the effectiveness of the assessment application are obtained through the assessment of the principal as a user. The principal was given a questionnaire to assess the effectiveness of the smartphone-based teacher performance appraisal application. The assessment includes four aspects, namely aspects of content feasibility, grammar, appearance, and practicality of the application. Each aspect is assessed with the criteria of very good, good, quite good, less good, and

very poor. The results of the analysis present the acquisition of the percentage value obtained from the assessment of the effectiveness of the principal as an application user of 95.3%. The scoring table shows the value of 95.3% is included in the good category, indicating that the smartphone-based teacher performance assessment application developed has a good effectiveness score.

Testing the effectiveness of the assessment application must meet several prerequisites that must be met. According to Kandak and Egen in [Kaluge et al. \(2004, p. 76\)](#), several criteria for the effectiveness of the assessment are: “Effective assessment in the real world of the classroom teacher has three interrelated features: It must be valid, systematic, and practical. To be valuable while remaining professionally sound, the assessment system must process all three features.”

Based on this opinion, it can be seen that the effectiveness of an assessment must meet three criteria: (1) valid, (2) systematic, and (3) practical. An assessment is said to be valid, if the assessment is able to assess and measure something that is being assessed. Assessment is said to be systematic, if the assessment activities are carried out regularly and well planned. It is said to be practical, if the assessment is easy to do, economical, and can achieve the expected results.

Based on the results of previous tests, it shows that the smartphone-based teacher performance assessment application meets the valid, systematic, and practical criteria. Thus, it can be said that the smartphone-based teacher performance appraisal application that was developed met the criteria for the effectiveness of the assessment. Thus, smartphone-based teacher performance appraisal applications can be used to facilitate teacher performance assessments. In line with the results of research by [Pradana and Kuswinardi \(2020\)](#) that teacher performance appraisal applications can make it easier to carry out teacher performance assessments, help teachers see the results of the assessments obtained, and make it easier for school principals to monitor the assessment results of all teachers.

Interpreting and Measuring Findings

Based on the test findings, it can be stated that the instrument in the teacher performance assessment application has good validity and reliability and is suitable for use as a pedagogical, personality, social, and professional sub variable teacher performance assessment instrument.

Disseminate

Disseminating is the process of making users aware of a newly developed product. The distribution process involves the mass deployment of apps. At this step, the produced application is distributed to every school principal in the Gringsing District, after which the principal can share it with his colleagues. WhatsApp and Google Drive are typically used to exchange smartphone-based teacher performance evaluation tools.

CONCLUSION

The development of smartphone-based teacher performance assessment applications includes four stages: (1) the define stage, (2) the design stage, which includes determining instrument specifications, determining application media specifications, creating flowcharts, and creating storyboards, (3) the developing stage, which includes assembling instruments, performing coding, conducting tests, analyzing test results, and interpreting measurement results, and (4) the disseminate stage. The findings of analyses of validity and reliability indicate that a smartphone-based teacher performance assessment application has the features of an instrument that meets the standards to be used in assessing. Through content validity and construct validity testing, it was determined that the smartphone application for assessing teacher performance was valid. Cronbach's alpha reliability estimation indicates that the instrument in the smartphone-based teacher performance assessment application has high reliability.

REFERENCES

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Andrian, D. (2019). Developing an instrument to evaluate the influential factors of the success of local curriculum. *REID (Research and Evaluation in Education)*, 5(1), 75–84. <https://doi.org/10.21831/reid.v5i1.23980>
- Arikunto, S. (2016). *Prosedur penelitian: Suatu pendekatan praktik [Research procedures: A practical approach]*. Rineka Cipta.
- Suranto, S. (2018). Evaluasi program Desa Mandiri Informasi berbasis aplikasi Android Sistem Informasi Desa (Sifordes) [Evaluation of the Independent Information Village program based on the Android Village Information System (Sifordes) application]. *Informasi*, 48(1), 1–13. <https://doi.org/10.21831/informasi.v48i1.20073>
- Bashooir, K., & Supahar, S. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran fisika berbasis STEM [Validity and reliability of scientific literacy performance assessment instruments in STEM-based physics lessons]. *Jurnal Penelitian dan Evaluasi Pendidikan*, 22(2), 219–230. <https://doi.org/10.21831/pep.v22i2.19590>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge Falmer.
- Djaali, D., Muljono, P., & Ramly, R. (2000). *Pengukuran dalam bidang pendidikan [Measurement in the field of education]*. Program Pascasarjana Universitas Negeri Jakarta.
- Fauziyah, S., & Triyono, M. B. (2020). Pengaruh e-learning Edmodo dengan model blended learning terhadap minat belajar [The influence of Edmodo e-learning with a blended learning model on interest in learning]. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, 4(1), 112–124. <https://doi.org/10.21831/jk.v4i1.27562>
- Ghozali, I., & Laten, H. (2015). *Partial Least Squares: Konsep, teknik dan aplikasi menggunakan program SmartPLS 3.0 [Partial Least Squares: Concepts, techniques and applications using the SmartPLS 3.0 program]* (2nd ed.). Badan Penerbit Universitas Diponegoro.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) using R: A workbook*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-80519-7>
- Hamid, M. A. (2016). Pengembangan instrumen penilaian hasil belajar siswa berbasis TIK pada pembelajaran dasar listrik elektronika [Development of an ICT-based instrument for assessing student learning outcomes in basic electrical and electronics learning]. *VOLT: Jurnal Ilmiah Pendidikan Teknik Elektro*, 1(1), 37–46. <https://jurnal.untirta.ac.id/index.php/VOLT/article/view/822>
- Istiyono, E., Dwandaru, W. S. B., Setiawan, R., & Megawati, I. (2020). Developing of computerized adaptive testing to measure physics higher order thinking skills of senior high school students and its feasibility of use. *European Journal of Educational Research*, 9(1), 91–101. <https://doi.org/10.12973/eu-jer.9.1.91>
- Jong, M. S. (2019). Sustaining the adoption of gamified outdoor social enquiry learning in high schools through addressing teachers' emerging concerns: A 3-year study. *British Journal of Educational Technology*, 50(3), 1275–1293. <https://doi.org/10.1111/bjet.12767>
- Kaluge, L., Setiasih, S., & Tjahjono, H. (2004). *The quality improvement of primary children learning through a school-based programme in Indonesia*. A Research Paper, Universitas Surabaya.

- Kwong, K., & Wong, K. (2013). Partial Least Squares Structural Equation Modeling (PLS-SEM) techniques using SmartPLS. *Marketing Bulletin*, 24, 1–32. <http://marketing-bulletin.massey.ac.nz>
- Mehrens, W. A., & Lehmann, I. J. (1973). *Measurement and evaluation in education and psychology*. Holt, Rinehart and Winston, Inc.
- Pradana, Y. F., & Kuswinardi, W. (2020). Rancang bangun aplikasi Android penilaian kinerja guru dengan metode agile development di MAN 1 Kabupaten Malang [Design and build an Android application for teacher performance assessment using the Agile Development method at MAN 1 Malang Regency]. *Rainstek: Jurnal Terapan Sains dan Teknologi*, 2(3), 182–191. <https://doi.org/10.21067/jtst.v2i3.4177>
- Regulation of the Director General of Teachers and Educational Personnels of the Ministry of Education, Culture, Research, and Technology No. 5958/B/HK.03.01/2022 concerning the Technical Guidance of Teacher Assignment as School Principal*. (2022).
- Regulation of the Minister of Empowerment of State Apparatuses and Bureaucratic Reform No. 16 of 2009 concerning Teachers' Functional Positions and Credit Points*. (2009).
- Retnowati, T. H., Mardapi, D., Kartowagiran, B., & Suranto, S. (2017). Model evaluasi kinerja dosen: Pengembangan instrumen untuk mengevaluasi kinerja dosen [Lecturer performance evaluation model: Development of instruments to evaluate lecturer performance]. *Jurnal Penelitian dan Evaluasi Pendidikan*, 21(2), 206–214. <https://doi.org/10.21831/pep.v21i2.16626>
- Saleh, A. M. (2016). *Komunikasi dalam kepemimpinan organisasi [Communication in organizational leadership]*. Universitas Brawijaya Press.
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial least square structural equation modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of market research* (Chapter 15). Springer.
- Sinaga, A. D., Lubis, J., & Sitanggang, N. (2021). Pengembangan model penilaian kinerja guru tetap SMP Perguruan Buddhist Manjusri Pematang Siantar [Development of a permanent teacher performance assessment model at SMP Perguruan Buddhist Manjusri Pematang Siantar]. *Jurnal Serunai Bahasa Indonesia*, 18(2), 95–104. <https://doi.org/10.37755/jsbi.v18i2.462>
- Sudjana, N. (2010). *Penilaian hasil proses belajar [Assessment of learning process results]*. Remaja Rosdakarya.
- Utami, N. R., Firdaus, E., Subakti, H., Purba, S., Salamun, S., Avicenna, A., Cecep, H., Cahyadin, W., Sakirman, S., Salim, N. A., Karwanto, K., & Tasrim, I. W. (2021). *Supervisi pendidikan [Educational supervision]*. Yayasan Kita Menulis.
- Yuningsih, I., & Sunhaji, D. (2021). Analisis hasil penilaian kinerja guru [Analysis of teacher performance assessment results]. *Educreative: Jurnal Pendidikan Kreativitas Anak*, 6(1), 338–346. <http://educreative.id/index.php/index>