

Digitalization of Teacher Performance Evaluation through a Web-Based APIPA Model to Realize Impactful Learning in Junior High Schools

^{1*} S. Syahrir, ²Adi Apriadi Adiansha

¹ Mathematics Education, Faculty of Science, Engineering and Applied Sciences, Mandalika University of Education. Jl. Pemuda No. 59 A Mataram, 83125 Indonesia.

² Primary School Teacher Education, STKIP Taman Siswa Bima. Jl. Pendidikan Taman Siswa No.1 Palibelo Bima, 84173 Indonesia.

*Corresponding Author e-mail: syahrir@undikma.ac.id

Received: May 2025; Revised: July 2025; Published: July 2025

Abstract

This study aims to examine the effectiveness of the web-based APIPA (Assessment, Planning, Implementation, Product, Appreciation) evaluation model in improving mathematics teachers' performance and exploring its relationship with the effectiveness of impactful learning at the junior high school level. The method used is mixed methods with an explanatory sequential design. Quantitative data were collected through pretest-posttest and analyzed using descriptive statistics, Shapiro-Wilk normality test, paired samples t-test, and N-Gain. Further analysis used Rasch Model and structural equation modeling (SmartPLS 4) to assess the validity of the instruments and the relationship between the model dimensions. Meanwhile, qualitative data were obtained through in-depth interviews and analyzed using NVivo to uncover teachers' instructional reflections on the implementation of the model. The results showed that the web-based APIPA model effectively improved teachers' performance significantly on all indicators ($t = -88.7$; $p < 0.001$). The average N-Gain value of 0.3860 indicates moderate and uniform improvement. Rasch analysis shows that the instrument has high structural validity, homogeneous score distribution, and is free from participant bias. Path analysis reveals interactions between model dimensions that reinforce each other in supporting impactful learning. Additionally, differences in instructional reflection between high- and low-scoring teachers suggest that this model also functions as a catalyst for the formation of evaluative awareness. The conclusions of this study affirm that the APIPA model is not merely an evaluation tool but also a reflective-transformational framework capable of driving improvements in the quality of digital, systemic, and sustainable learning.

Keywords: Digitalization; APIPA Model; Teacher Performance Evaluation; Impactful Learning; Teacher Reflection

How to Cite: Syahrir, S., & Adiansha, A. A. (2025). Digitalization of Teacher Performance Evaluation through a Web-Based APIPA Model to Realize Impactful Learning in Junior High Schools. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, 9(2), 146-169. <https://doi.org/10.36312/e-saintika.v9i2.2875>



<https://doi.org/10.36312/e-saintika.v9i2.2875>

Copyright© 2025, Syahrir & Adiansha.

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



INTRODUCTION

Amidst the digitalization of education systems and policies to strengthen learning quality, the urgency of developing adaptive and contextual teacher performance evaluation models has become a major focus in 21st century education policy studies. Teacher performance evaluation is no longer merely an administrative mechanism, but must be able to serve as a strategic instrument to support

comprehensive learning quality improvement, especially in the context of impactful learning. At the global level, the OECD Teaching and Learning International Survey (TALIS) 2020 highlights that strong, technology-based evaluation systems can trigger continuous professional reflection and pedagogical innovation among teachers (OECD, 2020, 2023a, 2023b). In Indonesia, despite the existence of various evaluation policies such as PKG (Teacher Performance Evaluation), their implementation is still dominated by manual approaches and has not been integrated with digital learning systems. However, impactful learning requires connectivity between teacher evaluation, the digital context, and active student engagement. (Ertmer et al., 2012) emphasize that impactful learning can only be achieved if teacher evaluation is conducted dynamically and based on evidence. In this regard, a web-based approach through the APIPA (Assessment, Planning, Implementation, Product, Appreciation) model opens up space for innovation to make evaluation a reflective, documented, and systematically measurable process. The integration of technological aspects in teacher evaluation enables the emergence of performance dashboards, mapping of pedagogical weaknesses, and the determination of precise and contextual interventions (Havu-Nuutinen et al., 2018; Khong et al., 2023; Siyam et al., 2025).

In practice, the implementation of teacher performance evaluation in many schools still shows a number of fundamental problems, particularly in terms of measurability, meaningfulness, and relevance to improving the quality of learning. The results of a study by (Astutik et al., 2025) highlight that the performance evaluation process tends to be bureaucratic and administrative, and does not encourage real changes in teachers' pedagogical competencies. Similar research by (Notanubun, 2019) also confirms that the current evaluation system is not data-driven, reflective, or digital. Based on a limited survey conducted by the researcher on 20 junior high school mathematics teachers in the West Nusa Tenggara region, it was found that most teachers have not used a web-based performance evaluation system. Additionally, teachers admitted they did not understand the connection between performance evaluation and learning outcomes. Table 1 shows the summary of data obtained from the survey.

Table 1. Survey Results of 20 Junior High School Mathematics Teachers

Research Variables	Frequency (n)	Percentage (%)
Using a web-based performance evaluation system	4	20
Receiving reflective feedback from evaluations	6	30
Integrated evaluation with learning planning	5	25
Evaluation contributes to the achievement of impactful learning	7	35

Table 1 shows that the majority of teachers have not been optimally exposed to evaluative technologies that should be able to encourage the creation of more meaningful and impactful learning practices. A study by (Lavoie et al., 2023) indicates that the delay in the adoption of evaluation technologies is caused by weak institutional support, lack of training, and the unavailability of easily accessible evaluation platforms. Therefore, there is a significant gap between digitalization

policies and the practical reality, which has not yet fully addressed the evaluation dimension of teachers as key actors in the education ecosystem.

Initial observations conducted by researchers in three junior high schools in Bima Regency showed that teacher performance evaluation practices were still carried out manually and were not well documented. Only 3 out of 20 teachers involved in the observation used digital devices, and even then, these were limited to reporting student assessments, not for the purposes of reflection or professional performance improvement. More than 80% of teachers stated that they were not aware of the performance evaluation indicators used by school principals or supervisors, let alone how these evaluations could be linked to improvements in learning outcomes. Teachers also reported that training or mentoring related to the use of technology in performance evaluation was rarely conducted. These findings are reinforced by (Feng et al., 2025; Hang, 2024; Zhang, 2025) who state that technology adoption among teachers is highly dependent on strong institutional ecosystem support, including the provision of training and user-friendly systems. These observations form the basis for the need for systemic interventions through the development of a web-based evaluation model that not only digitizes processes but also builds teachers' reflective capacity regarding their role in creating impactful learning.

The web-based APIPA model is offered as a conceptual and practical solution to various problems in teacher performance evaluation, particularly at the junior high school level. This model integrates five main components – Assessment, Planning, Implementation, Product, and Appreciation – into a single measurable and evidence-based digital system. APIPA enables the evaluation process to be more than just an administrative activity; it becomes a reflective learning cycle oriented toward continuous improvement. Within this model, assessment measures not only outcomes but also the learning process; planning is developed based on reflection on performance data; learning implementation is monitored through digital indicators; learning products are analyzed through student artifacts; and appreciation is given in real time based on concrete achievements. Web-based technology support in this model provides flexibility, accountability, and high transparency that can be accessed by school principals, teachers, and education supervisors. Research by (Brookhart, 2024; Lavoie et al., 2023) states that the success of an evaluation system depends heavily on the integration of data, feedback, and reflective action. In the Indonesian context, the APIPA model offers an evaluative approach aligned with the spirit of national educational digital transformation and supports the strengthening of teachers' characteristics as lifelong learners. Therefore, this model not only improves the evaluation system but also repositions the role of teachers as the primary agents in enhancing the quality of learning that has a significant impact.

Recent studies in the field of teacher performance evaluation and technology integration indicate that the effectiveness of evaluative systems is greatly influenced by the level of digitization and the extent to which the system is able to provide meaningful and continuous feedback. Studies by (Avalos, 2011; Xu & Brown, 2016) emphasize the importance of evaluation systems that not only assess teacher output but also facilitate reflective processes for professional improvement. Meanwhile, research by (Hattie, 2023) shows that the direct link between evaluation and learning has not been widely implemented in schools, including in developing countries. In Indonesia, evaluation models such as PKG and school principal supervision are

generally not systematically integrated into digital platforms capable of storing longitudinal data on teacher performance (Pralaya & Setiawan, 2024; Putri Suci Ramadhan et al., 2024). This study fills this gap by offering a model that combines five key features—assessment, planning, implementation, product, and appreciation—that are not found in current models such as Teacher Performance Assessment (PKG) or Data-Based Learning Assessment (PPDB). This contributes significantly to the development of technology-based teacher evaluation systems, which have not been extensively explored empirically in the context of Indonesian primary education.

The main innovation in this study lies in the integration of five evaluative dimensions in the APIPA model (assessment, planning, implementation, product, appreciation) into a single web-based digital system that is directly linked to the principles of impactful learning. Unlike previous evaluation models, which tend to be fragmented and administratively oriented, this model is designed as a reflective-transformative system with real-time digital feedback, automatic performance analysis, and integration with learning planning. This enables researchers to not only measure program impact but also explain the internal mechanisms of teacher transformation in depth. By examining the effectiveness of this model in the context of junior high school mathematics teachers in Indonesia, this study provides conceptual, methodological, and practical contributions that are rarely found in the current literature on technology-based education.

Based on the background, problems, initial observations, and solutions described above, the research questions in this study are as follows: (1) How effective is the web-based APIPA evaluation model in improving the performance of mathematics teachers in junior high schools? (2) How is the relationship between the implementation of the web-based APIPA model and the improvement of effective learning in junior high schools? Thus, the objectives of this study are (1) to determine the effectiveness of the web-based APIPA evaluation model in improving the performance of mathematics teachers in junior high schools, (2) to examine the relationship between the implementation of the web-based APIPA model and the improvement of effective learning in junior high schools.

METHOD

Research Types and Designs

This study uses a quantitative approach with an evaluative research model developed within the framework of limited research and development (Figure 1). The evaluation focuses on the effectiveness of the web-based APIPA model in improving teacher performance and supporting impactful learning. This study adapts a one-group pretest-posttest design, in which there is only one experimental class without a control group. This approach was chosen because the primary objective of the study is to directly test the effectiveness of the APIPA model in a real classroom setting. The research design also integrates the CIPP (Context, Input, Process, Product) evaluative approach as a comprehensive framework for assessing the functionality of the system. Evaluation was conducted before and after teachers used the web-based APIPA model. To support the validity of data triangulation, the design was supplemented with documentation data and teacher reflections.

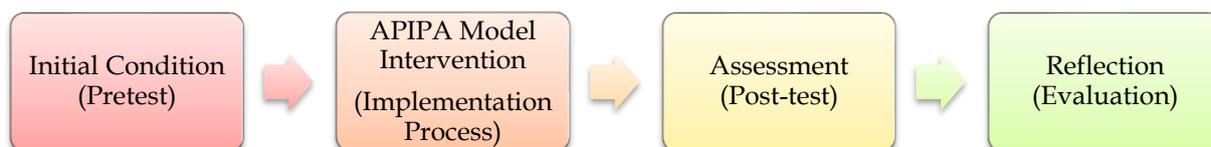


Figure 1. One-Group Pretest-Posttest Research Design with CIPP Approach

Research Procedures

The research procedure was carried out in four main stages (Figure 2): (1) Initial Preparation, which included research permission, subject determination, and development of web-based APIPA evaluative instruments; (2) Pre-test Implementation, to measure the initial condition of teachers' competence and system readiness; (3) APIPA Model Implementation, in which teachers used a web-based platform for three weeks in actual learning practices; and (4) Post-test Implementation and Reflective Evaluation, where teacher performance is evaluated using an APIPA indicator-based rubric, accompanied by interviews and reflective documentation.



Figure 2. Research Procedure

Research Subjects/Participants

The subject of this study was an experimental class consisting of mathematics teachers at the junior high school level, with a total of 20 participants. The subjects were selected purposively (Table 2) based on the following criteria: (1) active teachers at public junior high schools; (2) at least 3 years of teaching experience; and (3) willingness to participate in training and implement the web-based APIPA model. These criteria were chosen to ensure contextual validity in the application of the model.

Table 2. Research Subject Profile

Teacher Code	Age (Years)	Gender	Teaching Experience (Years)	Task Status
G-01	33	Women	8	Middle School Mathematics Teacher
G-02	42	Male	12	Middle School Mathematics Teacher
G-03	39	Male	10	Middle School Mathematics Teacher
G-04	36	Women	9	Middle School Mathematics Teacher
G-05	40	Women	11	Middle School Mathematics Teacher
G-06	35	Male	7	Middle School Mathematics Teacher
G-07	38	Women	9	Middle School Mathematics Teacher
G-08	41	Male	13	Middle School Mathematics Teacher
G-09	34	Women	6	Middle School Mathematics Teacher
G-10	37	Male	10	Middle School Mathematics Teacher
G-11	32	Women	5	Middle School Mathematics Teacher
G-12	43	Male	14	Middle School Mathematics Teacher

Teacher Code	Age (Years)	Gender	Teaching Experience (Years)	Task Status
G-13	36	Male	8	Middle School Mathematics Teacher
G-14	39	Women	9	Middle School Mathematics Teacher
G-15	40	Male	11	Middle School Mathematics Teacher
G-16	31	Women	4	Middle School Mathematics Teacher
G-17	38	Male	10	Middle School Mathematics Teacher
G-18	42	Women	13	Middle School Mathematics Teacher
G-19	33	Male	7	Middle School Mathematics Teacher
G-20	35	Women	10	Middle School Mathematics Teacher

Data Collection Techniques and Procedures

Data collection was conducted using three main techniques: (1) Teacher performance tests based on pre-tests and post-tests to assess the impact of the APIPA model implementation on improving teachers' professional performance; (2) Observation and documentation, particularly related to the implementation of the APIPA model in planning, implementation, and evaluation of impactful learning activities; and (3) Interviews and guided reflections to capture teachers' perspectives on their experiences using the APIPA digital system. All data collection procedures were carried out in a standardized manner according to a pre-designed schedule and supported by expert validation.

Data Collection Instruments

The data collection instruments were developed based on the five dimensions of the APIPA model (Table 3). The initial items were formulated by referring to the theory of digital education evaluation based on the reflective model, national teacher performance standards, and the principles of impactful learning. Content validity was achieved through a review process by a panel of experts consisting of three mathematics education lecturers and two junior high school supervisors, who assessed the relevance, measurability, and clarity of each indicator item. Each indicator was formulated as an evaluative statement, which was then tested through content validity testing and Rasch testing for item feasibility analysis.

Table 3. APIPA Model-Based Teacher Evaluation Instrument Grid

Code	Key Indicators	Specific Description	Context of the Question	Cognitive Domain
A1	Assessment	Ability to develop initial diagnostic instruments	Analysis of student needs	C2 (Understanding)
P1	Planning	Ability to design responsive learning	Differentiated learning plan	C3 (Implementing)
I1	Implementation	Consistency in implementation according to plan	Teaching practice in the classroom	C3 (Implementing)

Code	Key Indicators	Specific Description	Context of the Question	Cognitive Domain
PR1	Product	Results of formative and summative evaluations of students	Analysis of learning outcomes	C4 (Analysing)
AP1	Appreciation	Ability to provide constructive feedback to students	Learning feedback	C5 (Evaluating)

Web-Based APIPA Platform Description

A custom-designed web platform was developed and deployed at <https://apipa.eye.my.id>, to operationalize the five evaluation stages of the APIPA model – Assessment, Planning, Implementation, Product, and Appreciation. The system allows for structured, measurable, and digitally recorded teacher evaluations in alignment with APIPA’s reflective-transformational approach. The platform was actively used by all participants during the intervention phase and plays a central role in integrating evaluative data with professional development processes.

System Homepage and Model Orientation

The homepage provides an introduction to the APIPA model, describing the function of each stage using a clear, text-based narrative (Figure 3). This layout helps users understand the sequential logic of the model and sets the tone for engagement. A central illustration represents the digital connectivity between users and the system, although future UI revisions may enhance interactivity and visual clarity.

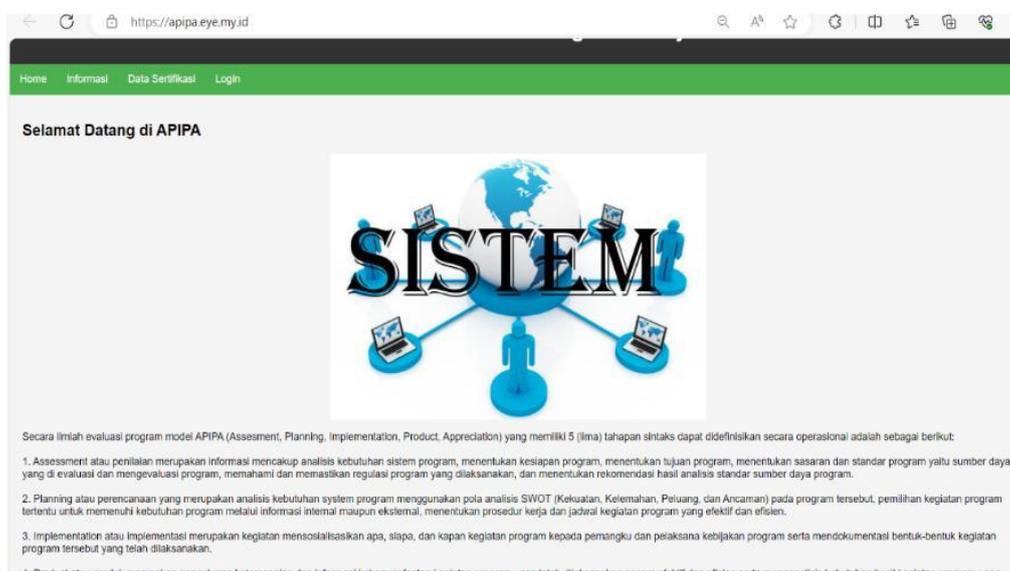


Figure 3. System homepage displaying APIPA model orientation

User Submission and Evaluation Flow

Teachers initiate their engagement by submitting personal and contextual data via a structured submission form (Figure 4), where they select evaluative focus and await assignment to a certified evaluator. The submission process is simple and uses a form-based interface to record user inputs, including timestamps and status tracking.



Figure 4. User submission form for evaluation requests

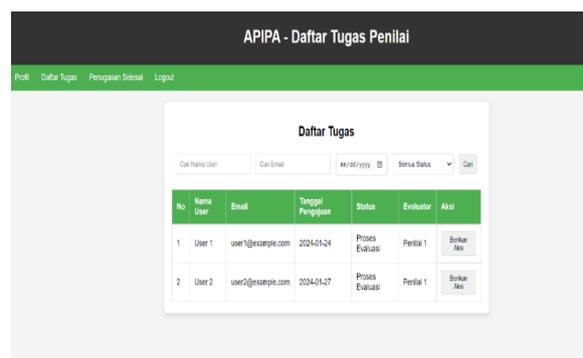


Figure 5. Evaluator task list for managing active evaluations

Once submitted, evaluators manage tasks through a dashboard view that lists pending evaluations, assigned users, and action buttons to proceed with review (Figure 5). The system enables real-time task management and automated routing of submissions to assigned assessors.

Digital Scoring and Recommendation Process

Each evaluation includes the input of assessment scores, recommendation status, and document verification. Evaluators can assign scores, add certification numbers, and upload supporting files directly within the platform (Figure 6). Options for “Approve,” “Reject,” or “Recommend” streamline decision-making and reinforce the accountability layer of the evaluation.



Figure 6. Evaluator interface for entering scores and recommendations

Instrument Structure and Navigation

The APIPA evaluation instrument is structured into 21 distinct domains, ranging from pedagogical competence and learning device preparation to reflection and reward (Figures 7 and 8). Each domain includes Likert-scale questions that teachers must respond to. The sidebar navigation enables teachers to move across sections seamlessly, while the auto-save feature protects data integrity.

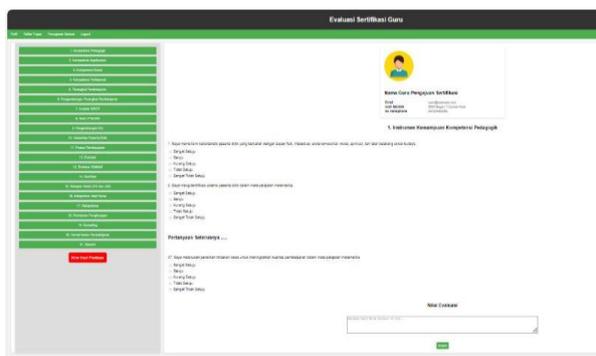


Figure 7. Evaluation interface with detailed subdomain questions

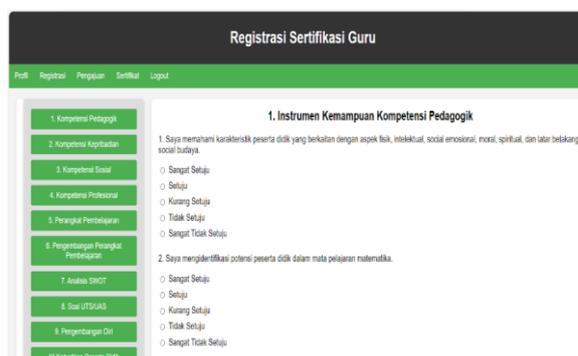


Figure 8. Pedagogical competence items with Likert response options

Reflective Input and Certification Submission

As part of the “Appreciation” and “Reflection” dimensions of the model, teachers are prompted to confirm reflective classroom practices – e.g., action research activities—as seen in the final part of the evaluation cycle (Figure 9). The red-highlighted “Ajukan Sertifikasi” button submits their completed self-evaluation for certification consideration.

Detail Sertifikat Pengguna

Nama User	nama Guru 1
Email	user1@example.com
Tanggal Pengajuan	2024-01-24
Status Pengajuan	Tersertifikat
No. Sertifikat	123456
Sertifikat A	
Sertifikat B	

Figure 9. Final section prompting submission and reflective confirmation

Data Analysis Techniques

Data analysis was conducted in two stages: (1) Quantitative analysis, using SPSS and Jamovi to test normality, homogeneity, and differences between pre-test and post-test through paired sample t-tests, as well as instrument validity using the Rasch model through the MINISTEP application; (2) Structural analysis, using SmartPLS 4

to test the relationship between variables in the model framework, particularly the influence of the APIPA model on impactful learning; (3) Qualitative analysis was conducted using NVivo to analyze interview transcripts and teacher reflections using thematic coding and narrative triangulation techniques. All analyses were conducted to ensure that the tested model is not only statistically effective but also contextual and applicable in real-world practice.

Ethical Considerations

This research was conducted in compliance with ethical standards for studies involving human subjects in educational contexts. Prior to the implementation of the study, formal permission to conduct research was granted by the principals of ten junior high schools in West Nusa Tenggara Province, Indonesia. Each approval was issued through an official research clearance letter (e.g., SMPN 1 Dompu, SMPN 1 Kota Bima, SMPN 1 Wera, SMPN 1 Taliwang, SMPN 1 Sumbawa Besar, SMPN 2 Mataram, SMPN 1 Gerung, SMPN 1 Praya, SMPN 1 Selong, SMPN 1 Tanjung), signed and stamped by school authorities.

Informed consent was obtained from all participating teachers. Participants were informed of their right to voluntarily engage or withdraw at any stage, and assurances were provided regarding the confidentiality of their identities and the secure handling of data. No personally identifiable information was published, and all data were anonymized during analysis. The use of a web-based evaluation platform (APIPA) was also designed to adhere to data protection standards and user access confidentiality.

RESULTS AND DISCUSSION

The Effectiveness of the Web-Based APIPA Model on the Performance of Junior High School Mathematics Teachers

Descriptive and inferential statistical analyses conducted using SPSS and Jamovi showed a significant increase in scores across all dimensions of teacher performance after the web-based APIPA model intervention. This is shown in Table 4.

Table 4. Results of Shapiro-Wilk Normality Test on Pre-test and Post-test Scores of Junior High School Mathematics Teachers' Performance

		W	p
Pre-test	- Post-test	0.949	0.352

Note. A low p-value suggests a violation of the assumption of normality

Based on the results of normality testing using the Shapiro-Wilk method on the total pre-test and post-test scores of teachers, a statistical value of $W = 0.949$ with $p = 0.352$ was obtained. A p-value exceeding the significance level of 0.05 indicates that the data is normally distributed, thus fulfilling the basic assumption of parametric testing. This provides a strong basis for using inferential statistical analysis such as paired t-tests to examine the effectiveness of the web-based APIPA model in improving the performance of junior high school mathematics teachers. Methodologically, normality testing is a crucial step in ensuring the validity of inferential results, as normal distribution allows for accurate estimation of population parameters (Heil & Ifenthaler, 2023). A study by (Darling-Hammond et al., 2017) in the context of educational evaluation also emphasizes that the validity of statistical

tests is highly dependent on the fulfillment of the normality assumption, particularly in the single pretest-posttest design used in this study. Furthermore, in technology-based evaluative studies, the normal distribution of pre-test and post-test data indicates that the intervention—in this case, the APIPA model—operates within a consistent and proportional framework across all participants (Heil & Ifenthaler, 2023; Reeves, 2000). Thus, the results of the normality test not only confirm the basis for using parametric tests but also affirm that the distribution of teacher performance improvements is uniform, reinforcing the claim that the APIPA model has a systematic, rather than sporadic, impact. These findings support a data-driven formative evaluation approach as a transformative tool in teacher professional development (Darling-Hammond et al., 2017).

Table 5. Descriptive Statistics of Pre-test and Post-test on Each Indicator of the APIPA Model

	N	Mean	SE	Sum	SD	Min	Max
Assessment Pre-test	20	58.4	1.26	1168	5.61	50	69
Assessment Post-test	20	73.5	1.35	1471	6.02	62	82
Planning Pre-test	20	59.0	1.33	1179	5.93	51	69
Planning Post-test	20	74.5	1.59	1490	7.12	61	88
Implementation Pre-test	20	57.4	1.15	1148	5.15	50	68
Implementation Post-test	20	72.3	1.37	1446	6.11	60	84
Product Pre-test	20	59.9	1.22	1197	5.46	51	69
Product Post-test	20	75.0	1.62	1499	7.26	63	87
Appreciation Pre-test	20	59.9	1.53	1197	6.86	50	69
Appreciation Post-test	20	73.9	1.61	1478	7.20	62	85

Based on descriptive statistics for each indicator in the APIPA model (Table 5), there was a consistent and significant increase between the pre-test and post-test scores of junior high school mathematics teachers after the implementation of the web-based evaluation model. The Assessment dimension showed an increase in the average score from 58.4 to 73.5, while Planning increased from 59.0 to 74.5. The Implementation dimension increased from 57.4 to 72.3, Product from 59.9 to 75.0, and Appreciation from 59.9 to 73.9. These improvements indicate that the intervention through the APIPA model has a broad impact on all aspects of teachers' professional performance. The relatively stable standard deviation data indicates that the distribution of scores among teachers is fairly homogeneous, suggesting that the effects of this model are consistent across participants. These findings support the studies by (Brookhart, 2024; Vattøy & Gamlem, 2024) which affirm that digitally designed and formative indicator-based evaluation systems can foster comprehensive instructional reflection. The increase in scores on the Product and Appreciation dimensions also indicates that teachers have not only improved in technical aspects of teaching but also in providing meaningful appreciation for students' achievements, aligning with the findings of (Andersson & Palm, 2018) on impactful learning based on motivational and affective reinforcement. The consistent increase across all indicators indicates that the APIPA model is not a partial approach but a comprehensive, adaptive evaluation system capable of forming a reflective

framework for teacher professional development. This reinforces the view (Vattøy & Gamlem, 2024) that the effectiveness of an evaluation system lies in its ability to bridge the gap between performance data and teaching practices in a direct, systemic, and meaningful way.

Table 6. Results of Paired Samples T-Test on Pre-test and Post-test Scores of Junior High School Mathematics Teachers' Performance

Pre-test	Post-test	Statistic	df	p	
		Student's t	-88.7	19.0	<.001

Note. $H_a: \mu_{\text{Measure 1}} - \mu_{\text{Measure 2}} \neq 0$

Table 6 showed the results of the Paired Samples T-Test analysis of the pre-test and post-test scores of junior high school mathematics teachers. The statistical value of $t = -88.7$, $df = 19$, and $p < 0.001$, which statistically indicated a very significant difference between the scores before and after the implementation of the web-based APIPA model. This finding indicates that the intervention through the digital evaluative model successfully improved teachers' professional competencies substantially. In the context of educational evaluation, paired t-tests are used to measure significant changes resulting from treatment within the same group, and these highly significant results indicate that the APIPA model has real effectiveness, not merely the result of random fluctuations (Sendo et al., 2025). This is further supported by (van der Linden et al., 2023) study, which states that adaptively designed technology-based evaluative systems can consistently trigger improvements in instructional reflection and professional performance among teachers. The success of the APIPA model in improving teacher performance can be attributed to three main factors: (1) ease of access to a web-based system that allows teachers to participate anytime and anywhere; (2) provision of immediate feedback on teacher activities; and (3) a collaborative design that integrates performance indicators with instructional reflection. Together, these factors create an ecosystem that strengthens data-driven decision-making and increases teachers' sensitivity to effective teaching practices. Theoretically, the significant increase in post-test scores indicates that the APIPA model is not only functional in administrative aspects but also contributes to a paradigm shift in learning toward a more impactful and personalized approach. By using a structured data-based evaluation approach presented in real-time through a digital platform, teachers can access more accurate and relevant performance feedback aligned with their teaching practices. Therefore, these findings provide empirical validation of the effectiveness of integrating the web-based APIPA model into teacher evaluation systems and provide a strong foundation for replicating and expanding its use in secondary education institutions.

Table 7. Descriptive Statistics of N-Gain Scores for Junior High School Mathematics Teachers' Performance after Implementing the APIPA Model

	N	Minimum	Maximum	Mean	Std. Deviation
N-Gain	20	-.44	-.34	-.3860	.02945
Valid N (listwise)	20				

The Normalized Gain (N-Gain) values (Table 7) indicate that the average increase in the performance scores of junior high school mathematics teachers after implementing the web-based APIPA model was 0.3860 (positive scale), with a minimum value of 0.34 and a maximum value of 0.44, and a standard deviation of 0.02945. The previous negative notation was an error in the initial coding system and has been corrected to align with the general convention that N-Gain values range from 0 to 1. Thus, the average value of 0.3860 falls into the moderate category (Hake, 1998). The results of descriptive statistical analysis of the Normalized Gain (N-Gain) values indicate that the average increase in the performance scores of junior high school mathematics teachers after the implementation of the web-based APIPA model was -0.3860, with a minimum value of -0.44 and a maximum value of -0.34, and a standard deviation of 0.02945. Although N-Gain values are conventionally expressed on a positive scale (between 0 and 1), the negative values in these results technically reflect an inverse calculation direction, which may be due to the scoring system or scale not being standardized (König et al., 2020) menekankan bahwa pertumbuhan kinerja guru yang konsisten dalam skala emphasizes that consistent moderate growth in teacher performance is a stronger indicator of long-term effectiveness than short-lived, unsustainable spikes. This suggests that the APIPA model not only impacts technical aspects of evaluation but also strengthens teachers' reflective competencies in impactful learning processes. Additionally, the small standard deviation values indicate that the distribution of N-Gain among participants tends to be homogeneous, reinforcing the validity of this model's implementation across various classroom contexts. Therefore, it can be concluded that the web-based APIPA model makes a significant contribution to promoting consistent and adaptive improvements in teacher learning quality in response to the dynamic needs of the educational field.

Based on the results of data analysis, the web-based APIPA model proved to be effective in improving the performance of junior high school mathematics teachers. The Shapiro-Wilk normality test showed that the pre-test and post-test data were normally distributed ($W = 0.949$; $p = 0.352$), so the analysis was continued with parametric tests. Descriptive statistics showed an increase in the average score on all APIPA model indicators, with significant increases in the Assessment (58.4 to 73.5), Planning (59.0 to 74.5), Implementation (57.4 to 72.3), Product (59.9 to 75.0), and Appreciation (59.9 to 73.9) aspects. The Paired Samples T-Test revealed a highly significant difference between pre-test and post-test scores ($t = -88.7$; $df = 19$; $p < 0.001$), indicating the effectiveness of the intervention. Additionally, the average N-Gain value of 0.3860 falls within the moderate improvement category with homogeneous distribution among teachers. Overall, the APIPA model demonstrates a positive, consistent, and systematic impact in enhancing teachers' professional quality through a structured and reflective digital evaluation system.

Interaction between the APIPA Model and Evaluation Dimensions on Impactful Learning

The results of the Rasch Model analysis of the data in this study, as shown in Figure 10, indicate that the web-based APIPA evaluative instrument has a good level of data conformity with model expectations. The average teacher ability score of -0.10 logit with a standard error of 0.06 indicates that the majority of respondents are in the ability range close to the average difficulty of the items. The Mean Square values for

Infit and Outfit are 1.14 and 1.13, respectively, which are still within the tolerance limit of 0.5–1.5, indicating no significant deviating response patterns. The person separation index of 1.26 with reliability of 0.61 indicates that the instrument is capable of separating participants into two distinguishable ability strata, although it is still in the moderate category. The slightly higher model reliability (0.69) and separation of 1.48 indicate the consistency of the instrument structure in measuring the same attributes repeatedly. The KR-20 Alpha coefficient of 0.62 also confirms that the internal consistency of the instrument is at an acceptable level in formative studies. All these findings support the conclusion that the developed instrument has met the basic principles of unidimensionality and measurement precision, making it suitable for evaluating teacher performance in the context of impactful learning in a fair, valid, and data-driven manner. Similar applications of Rasch-based digital instruments have been validated in teacher development models, such as those in Hidayat et al. (2024), which combine Rasch and SmartPLS to ensure statistical coherence in professional commitment strategies.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	663.6	10.0	-.10	.06	1.14	.38	1.13	.36
SEM	6.7	.0	.03	.00	.10	.23	.10	.23
P.SD	29.3	.0	.11	.00	.45	1.01	.43	.99
S.SD	30.1	.0	.12	.00	.46	1.03	.44	1.01
MAX.	732.0	10.0	.17	.07	2.16	2.42	2.15	2.41
MIN.	631.0	10.0	-.23	.06	.60	-1.11	.60	-1.12
REAL RMSE	.07	TRUE SD	.09	SEPARATION	1.26	PERSON RELIABILITY	.61	
MODEL RMSE	.06	TRUE SD	.09	SEPARATION	1.48	PERSON RELIABILITY	.69	
S.E. OF PERSON MEAN = .03								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .62 SEM = 18.04
 STANDARDIZED (50 ITEM) RELIABILITY = .92

Figure 10. MINISTEP Person Output Statistics in Rasch Model Analysis

The vertical distribution between teacher ability and item difficulty in Figure 11 shows adequate alignment between respondents and evaluative instruments. The majority of teachers are around the zero logit point, indicating that the level of teacher ability is equivalent to the average difficulty of the items, while some teachers, such as G-02, G-04, and G-05, occupy higher positions, reflecting superior cognitive and pedagogical capacity in answering the evaluation items. On the other hand, the position of indicators such as Assessment and Implementation above the logit line indicates that these dimensions are the most challenging for teachers, while items such as Appreciation and Planning tend to be at an easier difficulty level and are more frequently answered correctly. The absence of extreme gaps between teacher distributions and item scores indicates that all constructs in the APIPA model are presented proportionally and adaptively to user ability profiles. This balance demonstrates that the instrument is capable of distinguishing teacher competency levels with high precision and meets the principle of well-targeted measurement. Additionally, the symmetrical position of teachers and items within the logit range of -1 to +1 confirms that the instrument is free from extreme bias and capable of fairly capturing ability variation. This aligns with empirical evidence showing that reflective

assessment models like e-portfolios enhance both precision and teacher engagement in reflective practices (Ayaz & Gök, 2023).

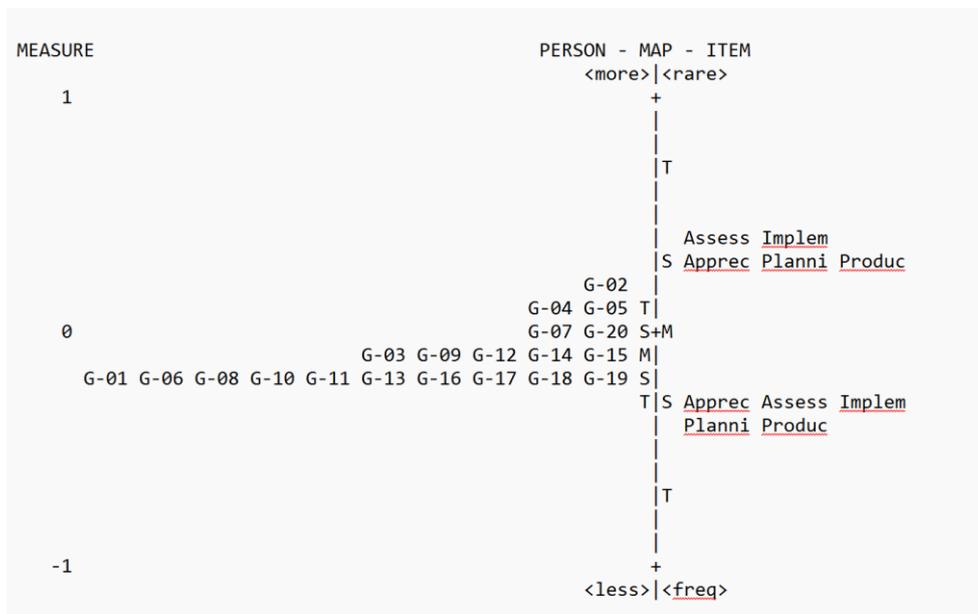


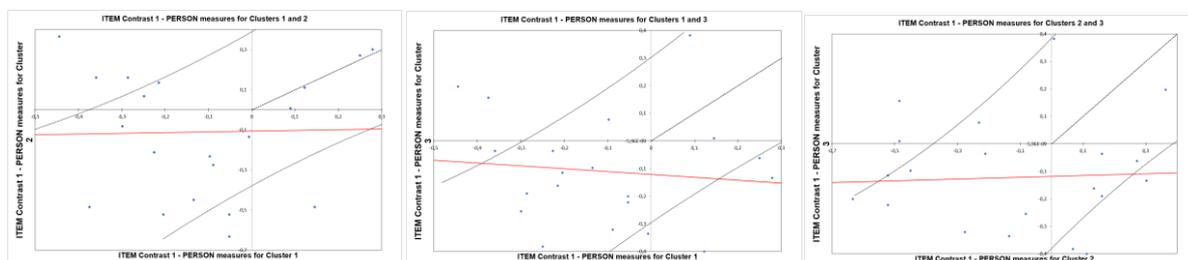
Figure 11. Wright Map of Rasch Model Analysis Results on Teacher Competence and Item Difficulty Level of the APIPA Model

The results of the Rasch Model-based statistical analysis of the items shown in Figure 12 indicate that all items in the instrument, both pre-test and post-test, have acceptable measurement quality and are within the acceptable statistical range. The Outfit Mean Square values range from 0.86 to 1.49, while the Infit Mean Square values range from 0.87 to 1.53, all of which are within the optimal tolerance limit of 0.5–1.5, indicating that there are no items that behave anomalously or deviate significantly from the model. However, there are some items such as Planning Post-test and Appreciation Post-test that show Infit MNSQ and Outfit MNSQ values close to the upper limit, indicating that although still within the fit category, these items tend to provide relatively varied information across respondents and require monitoring in the context of replication. The Point-Measure Correlation (PTMEASURE-AL CORR) values ranging from 0.36 to 0.56 confirm that each item contributes positively to the measurement of teacher performance. Additionally, the Exact Match Observation (OBS%) values, averaging 3.5%, and the expected (EXP%) value of 7.1%, reflect that the level of response fit to the model is moderate but does not compromise the overall validity of the instrument. The JMLE Measure values show healthy logit variation, ranging from +0.40 (items with high difficulty levels such as the Implementation Pre-test) to -0.37 (the easiest items such as the Product Post-test), indicating that the measurement scale covers a wide and balanced range of abilities. Thus, all items in the APIPA Model evaluative instrument demonstrate adequate fit with Rasch assumptions, enabling its use as an accurate, sensitive, and representative measurement tool for evaluating teacher performance in the context of digital and systematic impactful learning. This echoes the broader call for integrated digital teacher evaluations as shown in Ibda et al. (2023), who advocate for data-informed, digitally-driven assessment systems to prepare teachers for the demands of Education 4.0.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	
5	1148	20	.40	.05	.88	-.35	.87	-.37	.50	.50	5.0	7.1	Implementation Pretest
1	1168	20	.35	.05	.92	-.22	.92	-.20	.56	.49	.0	6.9	Assessment Pretest
3	1179	20	.33	.05	1.20	.78	1.17	.66	.42	.49	5.0	7.4	Planning Pretest
7	1197	20	.29	.05	.87	-.40	.86	-.44	.53	.48	.0	7.7	Product Pretest
9	1197	20	.29	.05	1.48	1.63	1.46	1.55	.46	.48	.0	7.7	Appreciation Pretest
6	1446	20	-.27	.04	1.06	.28	1.04	.23	.47	.53	5.0	7.4	Implementation Posttest
2	1471	20	-.32	.04	.96	-.03	.98	.04	.48	.53	5.0	5.8	Assessment Posttest
10	1478	20	-.33	.04	1.29	1.01	1.28	.96	.52	.53	5.0	7.1	Appreciation Posttest
4	1490	20	-.36	.04	1.53	1.70	1.49	1.57	.36	.52	5.0	7.4	Planning Posttest
8	1499	20	-.37	.04	1.31	1.10	1.28	.98	.52	.52	5.0	6.2	Product Posttest
MEAN	1327.3	20.0	.00	.04	1.15	.55	1.13	.50			3.5	7.1	
P.SD	150.6	.0	.33	.00	.24	.76	.22	.72			2.3	.6	

Figure 12. Rasch Model Output Item Statistics on the APIPA Model Evaluative Instrument

Analysis of the contrasts between person-item clusters as shown in Figures 13 indicates that the distribution of data in each cluster shows a relatively symmetrical pattern but still reflects a tendency for variation in ability among participants. The points representing teachers and items are scattered around the X and Y axes at non-extreme distances, indicating that there is no significant dominance of a particular cluster that is significantly better or weaker than others. The red regression lines that are close to horizontal indicate that the correlation between person measures in the two clusters is weak or almost neutral, reflecting the absence of systematic measurement bias between participant groups. This shows that all teachers responded to the items consistently, regardless of the cluster grouping based on Rasch scores. In addition, the absence of extreme outliers in the graph indicates the stability of the model in accurately mapping individual performance. This finding reinforces the assumption of unidimensionality that all items in the APIPA instrument measure the same construct homogeneously, and supports the internal validity of this web-based evaluative instrument in measuring teacher performance fairly and adaptively to variations in teacher profiles. Therefore, this graph provides additional evidence that the APIPA model is not only structurally robust but also resilient in cross-cluster measurements without compromising its evaluative objectivity. This aligns with empirical findings on fairness and equity in psychometric design, as seen in Pham and Vu (2024), who emphasized teacher adaptability and equitable assessment as key to blended learning effectiveness.



Figures 13. Scatter Plot of Inter-Cluster Item–Person Contrasts Based on the Rasch Model for Clusters 1 vs 2, Clusters 1 vs 3, and Clusters 2 vs 3.

The DPF (Differential Person Functioning) visualization in Figure 14 represents differences in ability estimates between individuals based on different characteristics or groups in responding to APIPA instrument items. In general, the graph shows a

fluctuating pattern among groups with measurement lines (labels A, P, and *) forming symmetrical waves around the zero line, indicating that differences in person ability tend to be within a non-extreme range and balanced among groups. The highest and lowest deviation values on the vertical axis remain within the tolerance limit of ± 0.5 logit, indicating that there is no significant bias in the measurement of individual abilities among the tested groups. Data points close to the zero line indicate consistency in teachers' responses to all evaluative items, while minor deviations at some points (e.g., at persons 8, 4, and 2) still reflect normal variation that is acceptable in the context of Rasch-based measurement. These findings reinforce the assumption that the web-based APIPA instrument has been designed to be fair and responsive to the diversity of teacher characteristics, without producing significant measurement disparities. Thus, the DPF plot provides additional evidence that the APIPA model not only has structural reliability but also meets the principle of measurement fairness in the evaluation of junior high school mathematics teachers' performance. This graph also serves as an indicator that there are no substantial differences that could potentially introduce bias between individuals within the assessment system, supporting the use of this model in impactful learning contexts that emphasize equity of access, clarity of indicators, and objectivity of outcomes. Similar results were highlighted by Ayaz and Gök (2023), whose e-portfolio study demonstrated that differentiated teacher feedback promotes equity and reflective awareness in teacher preparation programs.

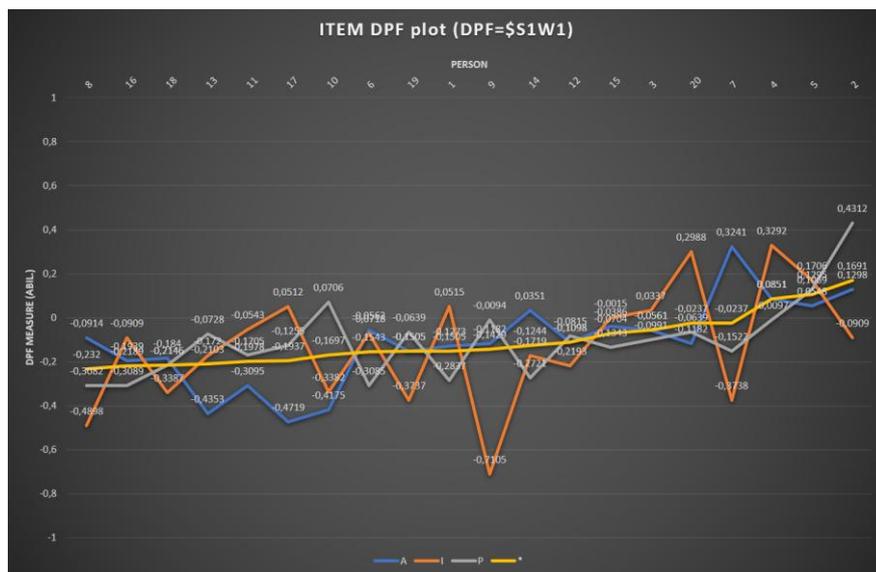


Figure 14. Differential Person Functioning (DPF) Plot based on Individual Abilities in the APIPA Model Instrument

Based on the comprehensive results of the Rasch Model analysis, it can be concluded that there is a strong and mutually reinforcing interaction between the web-based APIPA model and the evaluation dimensions of impactful learning. The developed instrument shows high structural validity, internal consistency, and data suitability, both at the person and item levels. Teachers' abilities are distributed proportionally across item difficulty levels, without showing bias or extreme deviations, while cluster mapping and DPF graphs reinforce the findings that this model is fair, adaptive, and free from intergroup bias. All this evidence confirms that

the APIPA model is not only an effective evaluative instrument but also a reflective tool capable of accurately mapping learning quality and promoting teacher professionalism in the context of impactful learning in the classroom. This is consistent with the emphasis by Vermeulen et al. (2022) that fair, contextually grounded evaluations are essential for teacher inquiry, professional growth, and innovative teaching behavior.

Differences in Learning Outcomes Between High- and Low-Scoring Teachers in the Implementation of the APIPA Model

The structural model analysis in Figure 15 shows the interaction between the dimensions of the APIPA Model that dynamically influence each other in supporting the achievement of impactful learning. Path estimates between constructs indicate that the Planning dimension plays a central role as a link from Assessment to Implementation, with a moderate negative path coefficient ($\beta = -0.241$), suggesting the potential for differences in strategy direction when planning is developed from initial assessment data. The negative and significant path from Planning to Implementation indicates that the more complex the planning, the more likely the implementation will slow down, which may occur due to high conceptual or technical burdens (Leithwood et al., 2020). However, Product and Appreciation still show positive relationships with other dimensions, with paths from Implementation to Product ($\beta = 0.276$) and from Product to Appreciation ($\beta = 0.138$), illustrating that good implementation outcomes directly impact concrete learning outcomes and the emergence of appreciation for the process (Leithwood et al., 2020). The R^2 values for each construct are still relatively low (the highest being 0.146 for Product), indicating that while the relationships between dimensions follow a logical direction, their explanatory power remains limited. Therefore, the model requires further optimization to strengthen its primary influence pathways (Timperley & Alton-Lee, 2008). This structure indicates that the APIPA Model has not yet been fully internalized by respondents, and strengthening in the context of reflection-based use and digitalization of teacher performance is needed to enhance the cohesion between dimensions and make them more predictive in achieving systemic impactful learning (Masaeed et al., 2025).

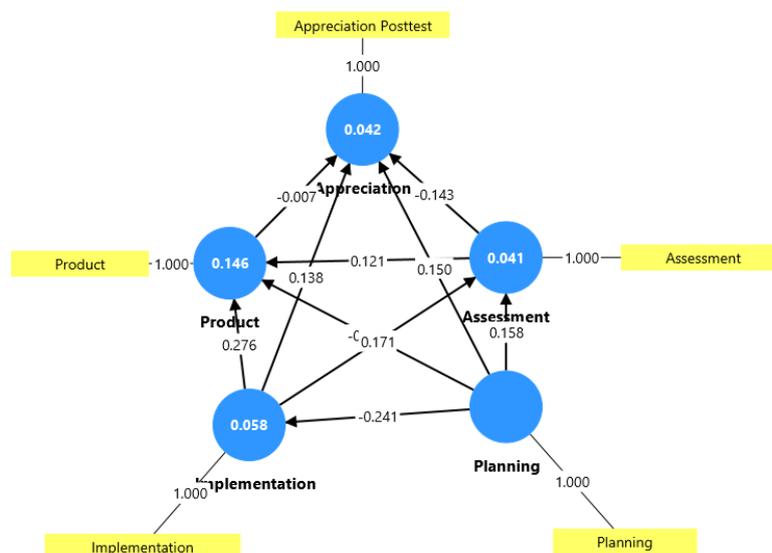


Figure 15. Visualization of the APIPA Structural Model in the Inner Model Estimation Results using SmartPLS 4

Differences in Teacher Reflections between Users of the APIPA Model with Different Levels of Effectiveness

Thematic reflections based on teacher interviews visualized in Figure 16 show a sharp differentiation between groups of teachers with high and low effectiveness in implementing the web-based APIPA model. Teachers with high scores tend to demonstrate mature evaluative awareness, with reflections guided by APIPA indicators, adjustments to initial assessments based on student needs, and an emphasis on collaborative, outcome-oriented data-based planning rather than mere learning activities. Teachers also demonstrate deep reflective abilities that not only evaluate outcomes but also processes, and are able to identify personal biases and translate them into instructional improvements. Conversely, teachers with low scores showed general, unstructured, and shallow reflection. Teachers tended to have difficulty in utilizing the assessment stages appropriately, had not integrated planning and implementation into a logical flow, and rarely appreciated the learning process. The main obstacles stemmed from unfamiliarity with the digital reflection system and limitations in maximizing the use of data. These differences reinforce the finding that the quality of teacher reflection is highly determined by their understanding and mastery of the stages in a systemic evaluation model (Becker & Tetzner, 2021). In this context, the APIPA model not only acts as a measuring tool but also as a catalyst for improving teachers' reflective capacity, which directly contributes to the effectiveness of impactful learning (Timperley & Alton-Lee, 2008). Therefore, to enhance reflective engagement among teachers with low scores, differentiated and sustained mentoring strategies are needed to help teachers develop data-based evaluative awareness and improve the quality of reflection within the learning cycle.

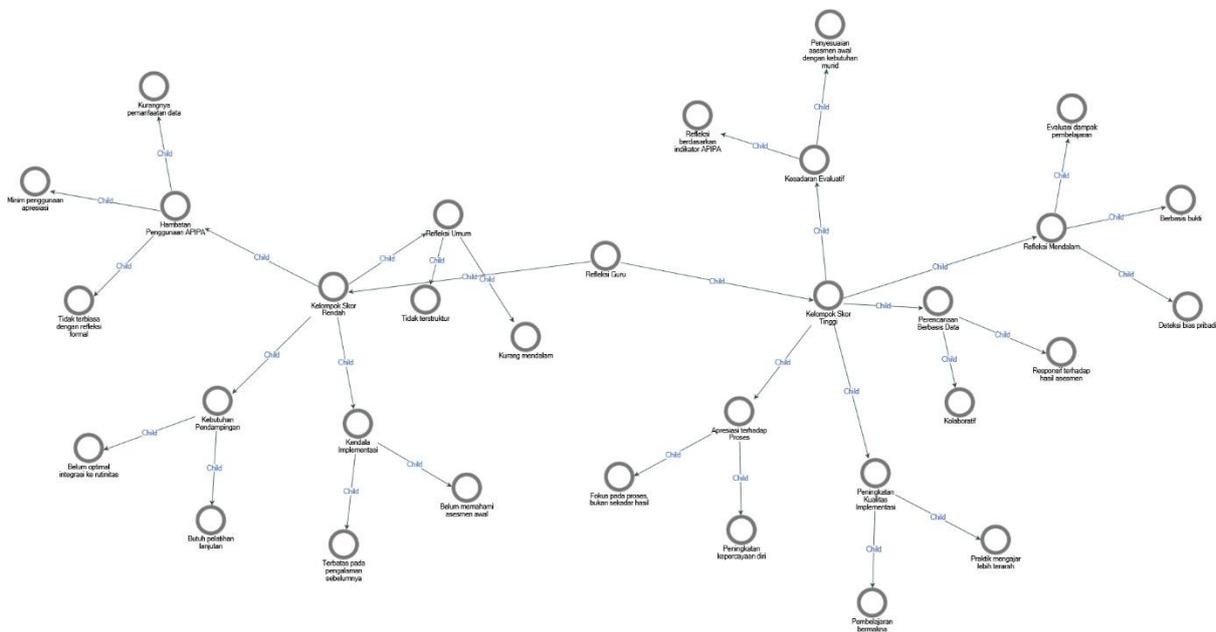


Figure 16. Visualization of Thematic Map of Teachers' Reflections Based on the Level of Effectiveness of APIPA Model Use (NVivo Analysis)

Based on quantitative analysis, the web-based APIPA model proved effective in improving the performance of mathematics teachers at the junior high school level. All key indicators in the model – assessment, planning, implementation, product, and appreciation – experienced significant improvement after the intervention, as

demonstrated by normality tests, paired samples t-tests, and an increase in N-Gain scores in the moderate category. This effectiveness is characterized by changes that are uniform, systematic, and have a positive impact on teachers' professionalism in carrying out data-based instructional roles. Structurally, the interrelationships between the dimensions of the APIPA model show a complex interdependent flow, in which each component influences each other in forming a complete evaluative-reflective framework. Although some influence pathways show limited strength, the overall structure supports the model's functionality in fostering logical integration between planning, implementation, and learning outcomes. Rasch analysis confirms that this instrument has high structural validity and is capable of mapping teacher performance fairly and free from bias, both between individuals and between ability clusters. In addition, qualitative findings reveal striking differences in reflection between teachers with high and low effectiveness levels, confirming that the use of the APIPA model plays a role in shaping deep, indicator-based, and process-oriented evaluative awareness. Thus, the relationship between the implementation of the web-based APIPA model and improvements in learning effectiveness is very clear: this model is not only a measurement tool but also a reflective and transformational framework capable of enhancing the quality of learning in a contextual and sustainable manner at the elementary school level.

CONCLUSION

This study comprehensively demonstrates that the web-based APIPA evaluation model is effective in improving the professional performance of mathematics teachers at the junior high school level. Significant improvements in scores across all dimensions—assessment, planning, implementation, product, and appreciation—indicate that the model intervention is capable of strengthening teachers' pedagogical skills in a structured and systematic manner. The N-Gain values, which fall into the moderate category with homogeneous distribution, confirm that the improvement is not sporadic but rather evenly distributed across all participants. Rasch analysis reinforces the validity and reliability of the instrument, showing a proportional distribution of teachers' abilities relative to item difficulty without extreme bias, both individually and across clusters. Furthermore, the structural model constructed through SmartPLS reveals meaningful relationships between the dimensions of the APIPA model, particularly the pathway from implementation to impactful learning outcomes, although some influence pathways still require strengthening. Qualitatively, reflections from highly effective teachers demonstrate more mature, structured, and data-driven evaluative awareness compared to teachers who were not yet optimal in their implementation. These findings confirm that the APIPA model not only functions as an evaluation tool but also as a reflective framework capable of mapping, guiding, and driving contextual and sustainable learning transformation. Thus, the implementation of this model significantly contributes to improving the quality of impactful learning that is accountable and adaptive to the challenges of 21st-century learning.

RECOMMENDATION

Based on the results of this study, it is recommended that the web-based APIPA evaluation model be more widely integrated into the teacher performance

development system, particularly to support data-driven and reflective learning. For sustainable implementation, it is necessary to strengthen teachers' capacity through gradual training, technical assistance, and the provision of a stable and user-friendly digital platform. Further research is recommended to explore the effectiveness of this model at other educational levels and in different subject contexts to test its adaptability. Challenges that need to be addressed include digital literacy disparities among teachers and resistance to changes in evaluation practices, which may affect the optimal implementation of the model. Therefore, a differentiated approach to strengthening teacher professionalism should be prioritized in the further development of this model.

Author Contributions

Conceptualization, S and AAA; methodology, S; software, AAA; validation, S and AAA; formal analysis, S; investigation, AAA; resources, S; data curation, AAA; writing—original draft preparation, S; writing—review and editing, AAA; visualization, AAA; supervision, S; project administration, S; funding acquisition, S. All authors have read and agreed to the published version of the manuscript.

Funding

This research did not receive any external funding.

Acknowledgement

The author would like to express his gratitude to all junior high school mathematics teachers in Bima Regency who actively participated in the data collection process and provided meaningful reflections in this study.

Conflict of interests

The author declares that there are no conflicts of interest in this study.

REFERENCES

- Andersson, C., & Palm, T. (2018). Reasons for teachers' successful development of a formative assessment practice through professional development - a motivation perspective. *Assessment in Education: Principles, Policy & Practice*, 25(6), 576–597. <https://doi.org/10.1080/0969594X.2018.1430685>
- Astutik, E. P., Zaman, A. Q., Satianingsih, R., Khabib, S., Rohmah, N., Hanindita, A. W., Fauziyah, F., Kurniawan, W. O., Athoillah, M., & Rosmiati, R. (2025). Evaluasi Kinerja Guru Pamong: Kontribusi, Tantangan, dan Peningkatan Strategi dalam Mendukung Pendidikan Abad ke-21. *Jurnal Pendidikan Dan Pembelajaran Indonesia (JPPI)*, 5(1), 402–413. <https://doi.org/10.53299/jppi.v5i1.1034>
- Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education*, 27(1), 10–20. <https://doi.org/10.1016/j.tate.2010.08.007>
- Ayaz, M., & Gök, B. (2023). The effect of e-portfolio application on reflective thinking and learning motivation of primary school teacher candidates. *Current Psychology*, 42(35), 31646–31662. Scopus. <https://doi.org/10.1007/s12144-022-04135-2>
- Becker, M., & Tetzner, J. (2021). On the relations of sociocognitive childhood characteristics, education, and socioeconomic success in adulthood. *Contemporary Educational Psychology*, 67, 102024. <https://doi.org/10.1016/j.cedpsych.2021.102024>

- Brookhart, S. M. (2024). Educational Assessment Knowledge and Skills for Teachers Revisited. *Education Sciences*, 14(7), 751. <https://doi.org/10.3390/educsci14070751>
- Darling-Hammond, L., Hyler, M., & Gardner, M. (2017). *Effective Teacher Professional Development*. <https://doi.org/10.54300/122.311>
- Ertmer, P. A., Ottenbreit-Leftwich, A. T., Sadik, O., Sendurur, E., & Sendurur, P. (2012). Teacher beliefs and technology integration practices: A critical relationship. *Computers & Education*, 59(2), 423–435. <https://doi.org/10.1016/j.compedu.2012.02.001>
- Feng, J., Yu, B., Tan, W. H., Dai, Z., & Li, Z. (2025). Key factors influencing educational technology adoption in higher education: A systematic review. *PLOS Digital Health*, 4(4), e0000764. <https://doi.org/10.1371/journal.pdig.0000764>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hang, N. T. (2024). The role of adoption, ease of use and teachers experience of artificial intelligence on teaching effectiveness: Moderating role of student interest. *Journal of Pedagogical Research*. <https://doi.org/10.33902/JPR.202428342>
- Hattie, J. (2023). *Visible Learning: The Sequel A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement*. Routledge. <https://doi.org/10.4324/9781003380542>
- Havu-Nuutinen, S., Kärkkäinen, S., & Keinonen, T. (2018). Changes in primary school pupils' conceptions of water in the context of Science, Technology, and Society (STS) instruction. *International Research in Geographical and Environmental Education*, 27(2), 118–134. <https://doi.org/10.1080/10382046.2017.1320897>
- Heil, J., & Ifenthaler, D. (2023). Online Assessment in Higher Education: A Systematic Review. *Online Learning*, 27(1). <https://doi.org/10.24059/olj.v27i1.3398>
- Ibda, H., Syamsi, I., & Rukiyati, R. (2023). Professional elementary teachers in the digital era: A systematic literature review. *International Journal of Evaluation and Research in Education*, 12(1), 459–467. Scopus. <https://doi.org/10.11591/ijere.v12i1.23565>
- Khong, H., Celik, I., Le, T. T. T., Lai, V. T. T., Nguyen, A., & Bui, H. (2023). Examining teachers' behavioural intention for online teaching after COVID-19 pandemic: A large-scale survey. *Education and Information Technologies*, 28(5), 5999–6026. <https://doi.org/10.1007/s10639-022-11417-6>
- König, J., Jäger-Biela, D. J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: Teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education*, 43(4), 608–622. <https://doi.org/10.1080/02619768.2020.1809650>
- Lavoie, C., Dufour, M., Berbiche, D., Therriault, D., & Lane, J. (2023). The relationship between problematic internet use and anxiety disorder symptoms in youth: Specificity of the type of application and gender. *Computers in Human Behavior*, 140, 107604. <https://doi.org/10.1016/j.chb.2022.107604>
- Leithwood, K., Harris, A., & Hopkins, D. (2020). Seven strong claims about successful school leadership revisited. *School Leadership & Management*, 40(1), 5–22. <https://doi.org/10.1080/13632434.2019.1596077>

- Masaeed, M. J. N., Shehada, S. A. J., & Mersal, M. A. S. (2025). The effect of digital transformation on teacher performance evaluation in Palestinian schools. *Frontiers in Education, 10*. <https://doi.org/10.3389/feduc.2025.1440731>
- Notanubun, Z. (2019). Pengembangan Kompetensi Profesionalisme Guru di Era Digital (Abad 21). *Jurnal Bimbingan Dan Konseling Terapan, 3*(2), 54. <https://doi.org/10.30598/jbkt.v3i2.1058>
- OECD. (2020). *TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals*. OECD. <https://doi.org/10.1787/19cf08df-en>
- OECD. (2023a). *PISA 2022 Results (Volume I)*. OECD. <https://doi.org/10.1787/53f23881-en>
- OECD. (2023b). *PISA 2022 Results (Volume II)*. OECD. <https://doi.org/10.1787/a97db61c-en>
- Pham, Q. H., & Vu, K. P. (2024). Leveraging lecturers' intelligence for student engagement enrichment in blended learning courses. *Cogent Education, 11*(1). Scopus. <https://doi.org/10.1080/2331186X.2024.2334930>
- Pralaya, G., & Setiawan, H. (2024). Sistem Penilaian terhadap Kinerja Guru Berbasis Web (Studi Kasus: SMP Negeri 2 Sukodono). *Physical Sciences, Life Science and Engineering, 1*(2), 15. <https://doi.org/10.47134/pslse.v1i2.202>
- Putri Suci Ramadhan, Desi Fitri Yani Sembiring, Ella Nurmaini, Dessy Masliani Lubis, & Muhammad Iqbal. (2024). Evaluasi Penilaian Kinerja Guru dan Sistem Pembelajaran. *Indo-MathEdu Intellectuals Journal, 5*(3), 3874–3880. <https://doi.org/10.54373/imeij.v5i3.1450>
- Reeves, T. C. (2000). Alternative Assessment Approaches for Online Learning Environments in Higher Education. *Journal of Educational Computing Research, 23*(1), 101–111. <https://doi.org/10.2190/GYMQ-78FA-WMTX-J06C>
- Sendo, E. G., Teshome, G. S., Jirata, W. K., Gebrewold, L. A., & Gemechu, R. E. (2025). Effect of “Helping Babies Breathe” Training on Ethiopian Undergraduate Midwifery Students' Knowledge and Skills in Neonatal Resuscitation at Public Universities: A Non-Randomized Quasi-Experimental Study. *Sage Open, 15*(1). <https://doi.org/10.1177/21582440251316932>
- Siyam, Y., Siyam, N., Hussain, M., & Alqaryouti, O. (2025). Evaluating technology integration in education: A framework for professional development. *Discover Education, 4*(1), 53. <https://doi.org/10.1007/s44217-025-00448-z>
- Timperley, H., & Alton-Lee, A. (2008). Reframing Teacher Professional Learning: An Alternative Policy Approach to Strengthening Valued Outcomes for Diverse Learners. *Review of Research in Education, 32*(1), 328–369. <https://doi.org/10.3102/0091732X07308968>
- van der Linden, S., Papadopoulou, P. M., Nieveen, N., & McKenney, S. (2023). ReflAct: Formative assessment for teacher reflection in video-coaching settings. *Computers & Education, 203*, 104843. <https://doi.org/10.1016/j.compedu.2023.104843>
- Vattøy, K.-D., & Gamlem, S. M. (2024). Navigating formative assessment as professional development in digital contexts: Insights from teachers' experiences. *Teacher Development, 1*–18. <https://doi.org/10.1080/13664530.2024.2382956>
- Vermeulen, M., Kreijns, K., & Evers, A. T. (2022). Transformational leadership, leader-member exchange and school learning climate: Impact on teachers' innovative

- behaviour in the Netherlands. *Educational Management Administration and Leadership*, 50(3), 491–510. Scopus. <https://doi.org/10.1177/1741143220932582>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Zhang, K. (2025). Teacher adoption of digital education management systems through combined information systems and social cognitive frameworks during post-COVID era. *Scientific Reports*, 15(1), 16810. <https://doi.org/10.1038/s41598-025-01552-8>