

A Conceptual Model of Multiple-Measures Assessment for Evaluating Indonesian Teachers: A Literature Review

Mahurni

Indonesian International Islamic University, Depok, Indonesia

*Correspondence Author Email: mahurni@uiii.ac.id

Abstrak: Kualitas guru merupakan faktor kunci yang menentukan capaian akademik peserta didik dan efektivitas sistem pendidikan secara keseluruhan. Di Indonesia, meskipun program sertifikasi guru telah diimplementasikan secara luas, dampaknya terhadap peningkatan kinerja pengajaran masih dipertanyakan. Kondisi ini menunjukkan perlunya kerangka evaluasi yang lebih komprehensif dan akuntabel. Penelitian ini bertujuan untuk meninjau dan mensintesis secara sistematis literatur ilmiah mengenai *Multiple-Measures Evaluation* dalam konteks pendidikan, mengkaji kontribusinya terhadap peningkatan kinerja instruksional guru di Indonesia, serta mengidentifikasi model pengukuran dan pendekatan administrasi yang mendukung akuntabilitas dan pengembangan profesional. Penelitian ini menggunakan desain kajian pustaka dengan menganalisis 60 sumber akademik yang diperoleh dari berbagai basis data pendidikan. Hasil kajian menunjukkan bahwa integrasi berbagai instrumen evaluasi, seperti *Value-Added Measures*, observasi kelas, survei siswa dan orang tua, serta portofolio, memiliki tingkat reliabilitas, validitas konstruk, dan validitas keputusan yang lebih kuat dibandingkan sistem evaluasi berbasis satu ukuran. Selain itu, pendekatan administratif secara konjungtif dinilai lebih sesuai dalam menjamin akuntabilitas yang seimbang dalam konteks pendidikan Indonesia. Meskipun setiap model pengukuran memiliki keterbatasan, penerapan secara terpadu memberikan kerangka evaluasi yang lebih komprehensif dan kredibel. Dengan demikian, sistem *Multiple-Measures Evaluation* berpotensi meningkatkan efektivitas pengajaran, memperkuat akuntabilitas, serta mendorong pengembangan profesional guru secara berkelanjutan.

Kata kunci: Evaluasi Guru, Pendekatan Multi-Ukuran, Kinerja Mengajar

Abstract: *Teacher quality remains a decisive factor in determining students' academic achievement and the overall effectiveness of educational systems. In Indonesia, despite the large-scale implementation of the teacher certification program, concerns persist regarding its limited impact on improving instructional performance. This situation highlights the need for a more comprehensive and accountable evaluation framework. This study aims to systematically review and synthesize existing scholarly literature on Multiple-Measures Evaluation in educational settings, examine how this approach contributes to improving teachers' instructional performance in Indonesia, and identify the measurement models and administrative approaches that support accountability and professional development. Employing a literature review design, this research analyzed 60 selected academic sources retrieved from major educational databases. The findings reveal that integrating multiple evaluation measures, such as Value-Added Measures, classroom observation, student and parent surveys, and portfolio assessment, provides stronger reliability, construct validity, and decision validity compared to single-measure systems. Furthermore, the literature suggests that a conjunctive administrative approach is more suitable for ensuring balanced accountability within the Indonesian educational context. Although each measurement model has inherent limitations, their combined implementation offers a more comprehensive and credible framework for evaluating teaching performance. The study concludes that a structured Multiple-Measures Evaluation system has significant potential to enhance instructional effectiveness, strengthen accountability, and foster sustainable professional development among teachers in Indonesia.*

Keywords: *Teacher Evaluation, Multiple-Measures Approach, Teaching Performance*

Submission History:

Submitted: October 7, 2025

INTRODUCTION

The certification initiative that was initiated in Indonesia in 2005 is widely regarded as one of the most extensive endeavours ever undertaken in underdeveloped countries. Since its inception, an estimated three million educators in Indonesia have actively engaged in the project and successfully acquired certification across the nation's vast archipelago. The implementation of this plan is necessary in compliance with the provisions outlined in the Teacher and Lecturer Law No. 4 of 2005. According to The Centre for Development and Enterprise (2015, p.4), teacher certification functions as a mechanism to augment teacher responsibility and promote professional growth. As a result, the main aim of this programme is to augment the knowledge and skills of educators, while concurrently endeavouring to improve their general welfare. The substandard outcomes of the national civic service examination in Indonesia (Jalal, 2009, p.8) underscore the main contributors to the inadequate quality of teacher candidates. Currently, among the entire population of 2.78 million instructors, an estimated 60% do not possess a college degree. According to a statement made by the Ministry of National Education (MONA, 2008), it was revealed that around 90% of primary school instructors lack a bachelor's degree. Furthermore, insufficiency in remuneration emerged as an additional influential element in the deterioration of teacher efficacy. According to Dolton and Gutierrez (2011, p.3), the implementation of supplemental incentives for present teachers, which are contingent upon their percentile performance and level of experience, is a viable strategy for improving the overall quality of instruction. Therefore, in order to meet the requirements for certification as a proficient educator and get unlimited professional compensation, teachers must satisfactorily fulfil the assessments and evaluations of their portfolio, along with successfully passing the PLPG.

However, after being in existence for over a decade, the certification programme has consumed a significant portion of the allocated state money, totaling IDR 69.7 trillion (www.apbn.org, 2016), without demonstrating any noticeable impact on educational achievements or the competitive position of the country. In contrast, as per the data published by the OECD in 2015, Indonesia was found to have the lowest ranking among the 45 nations surveyed in terms of student reading and science performances, and second to last in terms of student mathematics performances. Numerous research studies have presented empirical findings that lend support to the assertion that the aforementioned programme encountered notable deficiencies and led to a considerable misallocation of governmental resources (Fahmi et al., 2011; Hastuti et al., 2009; PISA, 2009; World Bank, 2012; The Jakarta Post, 2013; OECD, 2015; Haryanto et al., 2016). The ongoing discussion revolves around the causal relationship between the outcomes of certified teacher programmes and their quality. To effectively tackle this regrettable circumstance, it is crucial that the measurement system presently employed in the project receives a comprehensive and prompt assessment.

Hence, it is crucial to consider the following questions: what are the reasons behind the lack of significant improvement in teaching performance resulting from the current teacher certification programme in Indonesia? What is the underlying justification for the requirement to evaluate a teacher's proficiency in teaching? Which measures can be classified as accountable? Hanover Research (2012) raises a worry regarding the utilisation of a unique measure, specifically the implementation of the portfolio in the Indonesia Teacher Certification Programme. It is widely acknowledged that this particular strategy is prone to introducing a notable degree of subjectivity and often exhibits a lack of consistency. Moreover, Hastuti et al. (2009) posited that the aforementioned programme failed to adequately enhance the pedagogical competencies of the educators within their respective disciplinary domains. The researchers additionally discovered a positive association between the program's lack of success and the existence of fraudulent practises employed during the portfolio preparation procedure. The study revealed that a considerable proportion of teachers who were involved in the project received passing certifications, notwithstanding their failure in the test and subsequent need for corrective interventions. The failure can be further attributed to the intentions of the professors. Kusumawardhani (2017) posits that the notion of acquiring dual pay income arises as a noteworthy aspect that serves as motivation for instructors to engage in certification programmes. In the study conducted by Kusumawardhani (2017), it was found that a minority of educators hold the belief that the primary objective of teaching training within certification programmes is to augment teacher accountability and facilitate avenues for professional growth.

According to Peterson (1987, p.1), the assessment of educators is of paramount importance when it comes to understanding successful instructional methods, recognising outstanding achievements, and improving educational training initiatives. Moreover, as stated by Warring (2015), the incorporation of efficient assessment methods is crucial for improving the instructional efficacy of educators. Darling-Hammond (2010) asserts that the establishment of a dependable performance evaluation system is imperative for the attainment of a highly proficient teaching practicum. Hence, the objective of this study is to perform a comprehensive review of the existing literature. The purpose of this literature review is to provide evidence in favour of conducting more study on the inquiry question: "What is the effect of Multiple-Measure Evaluation on teacher teaching performance in Indonesia?" The study is expected to yield a range of outcomes, encompassing many dimensions such as self-reflection and professional development within the context of education.

This study aims to:

- (1). Systematically review and synthesize existing scholarly literature on Multiple-Measures Evaluation in educational settings.
- (2). Examine the extent to which Multiple-Measures Evaluation contributes to the improvement of teachers' instructional performance in the Indonesian context.
- (3). Identify and analyze the measurement models and administrative approaches discussed in the literature that enhance accountability and promote teachers' professional development.

METHOD

The primary objective of this literature review is to analyse and clarify the study inquiry, which pertains to the degree to which the adoption of a multiple-measure assessment system contributes to the improvement of teaching effectiveness among educators in Indonesia. The literature review encompasses a thorough assessment and examination of previously published academic literature that pertains to the understanding and comprehension of a certain field of study. In light of the aforementioned inquiry, the Monash Library Databases were employed to perform a comprehensive search for scholarly articles using the specified search terms, including "Multiple-Measure Definitions," "Multiple-Measure in Education," "Multiple-Measure Evaluation," "Multiple-Measure Challenges," and "Multiple-Measure Use for Improving Teaching Performance," among others. The examination was limited in its focus to the primary research inquiries, specifically "Multiple-Measures" and "Teaching" or "Teachers' Performance," across different situations. The imposition of this constraint was imperative to concentrate on the application of these particular and widely used phrases across various datasets. The selection of the Monash Library Database was made due to its comprehensive compilation of data, encompassing a vast number of resources such as journal articles, e-books, and conference proceedings, amounting to billions in quantity. The following paragraphs include a comprehensive examination of the procedures entailed in conducting data searches and synthesising information.

In the course of conducting the research, the author made the original decision to exploit the educational databases that are accessible through the Monash Host Library. The study has combined many databases, such as ERIC, ProQuest, ELSEVIER, Academic Search Complete, SAGE, EBSCO, CAULT Taylor & Francis, and others. In order to refine the search for scholarly articles, the Monash Library Database was employed, employing the subsequent search parameters: (1) utilisation of advanced search databases, (2) emphasis on applications related to language, (3) incorporation of both peer-reviewed and non-peer-reviewed articles, and (4) prioritisation of academic journals as the primary source category. A comprehensive examination of the inquiry's key phrases resulted in the identification of a total of 4300 journal articles that were deemed pertinent. Nevertheless, a significant proportion over 50% of the papers exhibit a deficiency in direct pertinence to educational subjects. The data synthesis procedure involved conducting a systematic and detailed review of a total of 60 papers, which were categorised into three distinct phases. The articles are classified into one to three categories based on their topic. Moreover, the selection of papers was based on the inclusion of phrases that provided a comprehensive description of the study issue. The data descriptions related to the inquiry phases were obtained from journal articles to aid in further synthesis. Additional material from other sources was included where needed.

RESULT AND DISCUSSION

Multiple-Measures: An Evaluation for Accountability

According to Henderson-Montero et al. (2003), several-Measures refers to the use of several assessment and evaluation measures across different content domains. These measurements encompass a range of methods, such as test scores, classroom observation, and portfolios, among others. Brookhart (2009) employs the term MMES to refer to the application of several assessment measures, including those that evaluate distinct constructs, those that evaluate the same construct but through diverse approaches, and the administration of many iterations of a single test. The main objectives of the MMES encompass the validation of information's precision and reliability, along with the mitigation of personal bias's impact throughout the assessment and evaluation procedures (Hanover Research, 2012). The adoption of a multiple measures evaluation system (MMES) is considered a reliable and effective method for evaluating the instructional practises of teachers in countries recognised for their strong academic performance (Taylor & Tyler, 2012; Strunk et al., 2014; Muñoz & Dossett, 2016).

The justification for its implementation is based on the idea that decisions about a student's educational progress, such as retention, tracking, or graduation, should not be based solely on the results of a single examination. Instead, it is recommended to consider additional relevant and reliable data (NEA Education Policy and Practise Department, n.d., p.1). Muñoz and Dossett (2016) conducted a study wherein they identified three distinct advantages linked to the use of various metrics during the evaluation process. First and foremost, it demonstrates an increased ability to make predictions, subsequently showcasing improved dependability, and eventually holds the potential to be utilised as a diagnostic instrument, thus providing significant insights for the improvement of pedagogical approaches. Moreover, as stated by Brookhart (2009), there has been a contention that the use of these metrics can augment both the construct validity and choice validity.

The concept of enhancing construct validity pertains to the degree to which the results derived from a measurement accurately and faithfully represent the fundamental construct that is being assessed. Brookhart's study provides evidence to support the notion that it is possible to perceive a certain achievement within a particular discipline. In contrast, the process of strengthening decision validity involves the improvement of one measure's outcome in relation to the others, or the mutual enhancement of performance measurements. When evaluating an individual's eligibility for a scholarship, it is crucial to conduct a comprehensive evaluation of both their academic and non-academic achievements. In addition, it is imperative to take into account many aspects related to evaluation and accountability, such as reliability, credibility, alignment, and efficiency, during the process of constructing numerous measures (Schafer, 2003). Despite the availability of well-designed assessment tools such as Danielson's Framework for Teaching for classroom observations and Ferguson's Tripod Survey (often referred to as 7Cs) for student surveys, this issue remains significant.

Teacher Measurement Models

One of the main challenges faced in the context of utilising numerous measures is the task of selecting suitable measurement models and effectively integrating them in a manner that is both efficient and easily understandable. However, according to the available literature, it may be argued that Valued-Added Measures (VAMs), Classroom Observation (CO), Students and Student-Parents' Survey (SS), and Portfolio are the four often employed measures. In industrialised nations like the United States, multiple measure designs are frequently utilised to evaluate teacher teaching efficiency or performance, employing a range of assessment methodologies. The subsequent statements provide their justifications.

1. Valued-Added Measures

The integration of value-added indicators into a teacher assessment system has emerged as a key educational issue. The Valued-Added Measures (VAMs) are widely recognised as a highly efficacious and often employed evaluation system for assessing the efficiency of teachers. According to Mangiante (2011), the categorization of Value-Added Models (VAMs) relies on the evaluation of instructors' efficacy within the specific setting of classroom teaching, without taking into account variables such as students' socioeconomic status, racial or cultural background, and level of motivation. Warring (2015) posits that the utilisation of Value-Added Models (VAMs) is accompanied by a multitude of benefits. In a theoretical context, the use of Value-Added Models (VAMs) allows educational practitioners to evaluate the influence of VAMs on the academic advancement of pupils. Furthermore, the implementation of Value-Added Measures (VAMs) can function as a valuable instrument for instigating conversations pertaining to the efficacy of the curriculum, instructional approaches, and programme structures. Moreover, through the utilisation of Value-Added Measures (VAMs), educators are able to effectively identify the strengths, shortcomings, and prospective areas for enhancement in students' academic achievements. Moreover, the utilisation of value-added models (VAMs) data is employed to determine the teachers who require professional training.

Nevertheless, it is worth considering the potential implications if the assessment of Value-Added Models (VAMs) were to encompass supplementary factors that hold significance, such as students' attributes, class size, linguistic and cultural aspects, economic variables, and other pertinent elements. Does the influence on teacher evaluation remain consistent across kids from affluent and economically disadvantaged backgrounds? Jensen and Minke (2017) suggest that the role of the family is a crucial factor in shaping the academic achievement of students. In the present-day context of Indonesia, it has become a prevalent practise for students belonging to moderate to upper socioeconomic strata to participate in additional educational endeavours outside of their usual school hours, both within and outside of formal educational settings. The study into the validity, stability, and precision of Value-Added Models (VAMs), which are utilised for the assessment of teacher effectiveness, remains ongoing (Murphy, 2013).

To address the issues of bias, inconsistency, and inaccuracy in the application of Value-Added Models (VAMs) in educational contexts, particularly in Indonesia, it is crucial

to do a comprehensive analysis of the accountability aspect. Hence, it is crucial that all relevant parties engaged in the procedure, encompassing educators, pupils, and administrators, have a shared agreement on the pedagogical objectives prior to commencing the execution (Lee & Waimer, 2002). This represents a significant progress in the techniques of the evaluation system, accompanied by the focused quality measurements and subsequent plans for execution.

2. Classroom Observation

Classroom observation, also known as CO, is an extra assessment indicator that significantly impacts the enhancement of teaching effectiveness. The concept of the CO can be considered as a notable assessment framework that is characterised by its formal, direct, and equitable nature (Hull, 2013), particularly in situations where there are important consequences at stake (Lei, et al., 2018). According to Waxman and Huang (1999), the main purpose of these measurements is to evaluate the execution of classroom instruction or instructional designs after the identification of Value-Added measurements (VAMs). Munoz and Hossett (2016) conducted a study wherein they discovered three primary characteristics that play a significant role in facilitating the effective execution of CO. To ensure a thorough and consistent evaluation process, it is crucial to utilise a stringent instructional rubric that aligns with the prescribed terminology and performance criteria specified in the Danielson Framework. Furthermore, it is crucial to build a complete framework that efficiently aids institutional leaders or school administrators in carrying out their evaluation responsibilities. Finally, the curriculum provides school leaders with opportunities for professional development, allowing them to better their skills in instructional coaching and engage in relevant discussions regarding the improvement of instructional practises.

One noteworthy advantage of the CO is its capacity to enable observers, such as school principals, to directly engage in observation and provide constructive input that is impartial and unbiased. The purpose of this feedback is to improve instructional techniques in a classroom setting (Murphy, 2013). Moreover, a strong association can be observed between the classroom teacher's observation scores and the Value-added Score, as evidenced by Cingos' research findings (2015, p.5). An additional advantage of this measuring paradigm is in its primary emphasis on obtaining data pertaining to classroom instruction, encompassing instructor behaviours and activities (Hanover Research, 2012). Although there are certain limitations associated with the application of the CO, including its time-consuming nature, high expense, and potential lack of observer skill, it remains a significant tool for assessing classroom teaching ability. This aspect has significant importance due to the insufficiency of relying exclusively on student value-added measurements to comprehensively evaluate the efficacy and constraints of instructional methods. In summary, one can posit that the implementation of classroom observation as an evaluative tool has yielded substantial effects in improving teachers' instructional effectiveness. Moreover, it is argued that there exists a positive relationship between the frequency of conducting classroom observations and the reliability and validity of the evaluation outcomes.

3. Students' and Parents' Survey

Student surveys (SS) are widely regarded as a dependable assessment tool for evaluating the performance of teachers. The research findings suggest that the data collected from the SS regarding teacher performance is deemed accurate and trustworthy (Baker, 2011, as quoted in Hanover Research, 2013, MET project, 2011, Chan et al., 2014). The primary objective of this evaluation is to authenticate the two preceding evaluation models, specifically the value-added measures (VAMs) and the classroom observation (CO). In the event that the value-added measures (VAMs) and classroom observations (CO) produce unsatisfactory results in assessing teacher competence, it may be prudent to explore the implementation of the school system administration (SS) as an alternate complementary strategy. In 2013, Hanover Research conducted a study which found that students possess the capacity to offer feedback on their teachers' performance throughout instructional periods through the utilisation of the Student Survey (SS). The evaluation frequently use the tripod-Survey format, sometimes known as the 7 C's, which was created by Dr. Ron Ferguson.

The survey incorporates a range of teaching performance qualities, specifically caring, clarifying, captivating, consolidating, conferring, challenging, and controlling. When students have difficulties and confusion in understanding the content being taught, they have the chance to articulate their thoughts and feelings using the survey form. Therefore, this comments can be considered as valuable information for improving future instructional practises. Hence, as per Murphy's (2013) findings, the utilisation of surveys provides teachers with targeted and thorough feedback, thereby facilitating the improvement of their professional teaching practises. Unfortunately, the students involved in this assessment system are unable to offer comprehensive insights into the precise facets of the teacher's subject knowledge or other professional endeavours, as these lie beyond their domain of expertise. Overall, there is an ongoing improvement in the capabilities and effectiveness of educators. Therefore, it is necessary for me to claim that this technique significantly impacts the ongoing enhancement of teachers' instructional efficacy and students' educational accomplishments.

Another method for evaluating teacher performance that is valuable for improving teaching practises is gathering feedback through surveys distributed to parents or family members. The role of parents and family members in students' academic success has been acknowledged by researchers (Persada et al., 2016; Jensen & Minke, 2017). Furthermore, it is worth noting that parents and other relatives might offer significant insights for teacher evaluations, even though they are typically seen as members of the school system's audience (Hanover Research, 2012). Previous research has recognised the incorporation of a parental survey into the evaluation system as a means of implementing market accountability (Garn, 2001, as referenced in Williams & Engel, 2013). The researchers suggest that to optimise service quality and bolster instructor efficacy, educational institutions should demonstrate responsiveness towards their clientele, particularly the student-parent population.

However, a notable obstacle in the implementation of parental surveys in Indonesia pertains to the restricted educational attainment of parents dwelling in distant regions. Assuming student-parents exhibit a respectable degree of intellectual attainment, it is reasonable to infer that they are capable of offering relevant input regarding the subject matter or issues under discussion. Nevertheless, their ability to evaluate the effectiveness of instructional methods is limited. Hence, one could contend that this survey holds a small degree of impact within the domain of educational advancement.

4. Portofolio

A potentially important source of information during the evaluation process is the submission of a compilation of papers that provide a comprehensive account of the professional activities carried out over different teaching sessions. The measurement technique discussed in this context has become widely recognised among the majority of teachers now employed, as portfolios were the exclusive basis for assessing teachers in Indonesia from 2006 to 2012. The portfolio offers a complete overview of the professional engagements and activities in which the teachers have effectively participated and actively involved themselves. These tasks involve multiple elements, such as the creation of meeting agendas, which may include semester programmes and lesson planning. Moreover, they encompass the provision of extensive explanations of finalised endeavours, such as educational resources. Moreover, teachers may also engage in activities that lie beyond their professional domain, such as taking on the responsibility of a coach in an extracurricular pursuit.

This evaluation model includes specific features that cannot be directly assessed using the four metrics stated above. Moreover, the analysis lacks the incorporation of comparison data such as the value-added score and students' survey. However, to a certain extent, the works produced by the students are integrated, providing the teachers with an opportunity to acquire understanding and insight into the process of student learning within the specified time periods. The extant literature has extensively documented the utilisation of portfolios. The extent of study evidence pertaining to the correlation between portfolios and student success is constrained, as indicated by a report conducted by Hanover study in 2012. When portfolios are implemented in accordance with defined protocols, it can be determined that they will successfully demonstrate the quality of teachers. Nevertheless, it is important to note that these evaluations may not offer a comprehensive understanding of the ongoing and future effectiveness of teaching practises.

Approach in Administering the Measurements

To optimise the effectiveness of various assessment methods, it is crucial to comprehend the simultaneous execution of these methods to guarantee that the assessment results maintain relevance, especially with regards to the professional development of educators. Brookhart (2009) delineated three separate methodologies for efficiently employing multiple metrics to collect data. There are three various ways that can be utilised, specifically conjunctive, compensatory, and complimentary. Nevertheless,

it is important to note that there may exist considerable disparities in terms of the precision of these three models when applied in different policy contexts, as shown by Martinez et al. (2016). The conjunctive technique necessitates that teachers effectively meet all essential criteria. If the evaluation system has three or four measures, it is essential for teachers to satisfactorily complete all three exams. According to Henderson-Montero et al. (2003), the main objective of the conjunctive technique is to integrate results from many measurements by setting a minimal threshold score for validation.

Conversely, under the compensating strategy, a diminished performance score on a particular measure might be mitigated by a superior score, so effectively neutralising the overall outcomes. In a study conducted by Martinez et al. (2016), it was shown that when a school concentrates exclusively on maximising one of the three specified model measures in teacher evaluation, it frequently results in the creation of imbalanced weights. Hence, one could contend that this phenomenon engenders a milieu that erodes the veracity of the aforementioned measures. On the other hand, the complementary approach allows an educator to deliver any of the designated assessments (Brookhart, 2009). Martinez et al. (2016) posited that the incorporation of data from one performance measure has the capacity to augment the credibility of information derived from another measure model.

After doing a thorough analysis of the three aforementioned ways, it is apparent that the conductive approach is in accordance with the multiple measure evaluation methodology. The aforementioned strategy seems to be more appropriate for the particular political and cultural circumstances of Indonesia. The data acquired from a single set of measures has the capacity to augment the credibility of the remaining measurements. Hence, it is advisable to employ it for the goal of evaluating the instructor's effectiveness.

Constraint in the Administration

One prominent concern in Indonesia pertains to the recurrent revision of the curriculum, which has proven to be a considerable source of distraction for educators and has also presented difficulties for the assessment framework. Indonesia's education curriculum has experienced five adjustments since the 2000s, prompted by changes in educational paradigms or ideologies. The curriculum underwent a significant transition in its most recent edition, as it embraced a progressive educational philosophy. This transition encompassed a shift from a pedagogical model that prioritised the teacher's role to an educational strategy that placed greater emphasis on the student's involvement and agency. As per the OECD (2015), this particular model affords students the opportunity to freely demonstrate their ability without the necessity of rote memorization of subject information. Moreover, its purpose is to expedite modifications within the educational setting. It is noteworthy to emphasise that this particular approach can also be denoted as problem-based or experiential learning, as posited by Dewey (1986).

Muthim (2014) asserts that the modifications in instructional methods and content have adversely affected instructors' capacity to concentrate on the act of teaching. In

addition to the requirement for instructors to acquire new pedagogical abilities, the implementation of a new instructional delivery method also necessitates the acquisition of supplementary educational resources that are in line with their classroom instruction (Lizer, 2013). When educators frequently become diverted by solely these two activities, it leads to unsatisfactory instructional efficacy, so considerably affecting the evaluation results, particularly when taking into account value-added metrics and feedback from students and parents.

CONCLUSION

The present study has focused on examining the research topic pertaining to the degree of influence exerted by the implementation of various techniques on the improvement of teaching performance among high school educators in Indonesia. The findings suggest that the adoption of numerous assessment methods has a notable and beneficial influence on the instructional effectiveness of teachers. The usage of several measures has a significant potential for successful application within the teacher evaluation system in Indonesia. The integration of the four measures that were provided has shown that these measures exhibit adequate levels of reliability and validity, therefore making them suitable for a thorough review. Although it is indeed accurate that each individual measure has its own limitations, the advantages of utilising them combined within the framework of teacher evaluation outweigh the disadvantages.

Furthermore, with relation to the management of various assessments, the results suggest that it is preferable to utilise a comprehensive assessment strategy rather than only depending on a solitary metric for evaluating teachers. The reason for this is the substantial influence it exerts on the educational achievements of kids. Hence, it is advisable to employ a comprehensive set of assessment tools, encompassing the value-added score, classroom observation, student and parent questionnaires, and portfolio assessment. Moreover, the application of a conjunctive methodology in the field of measurement is closely connected to the characteristics of various measures. Hence, it is crucial that enhancements in teaching emphasise the integration of constructive feedback from diverse sources. When evaluating teaching performances, it is imperative to give priority to the utilisation of surveys administered to students and parents as the principal method of assessment. Supplementary components, such as classroom observation and portfolio assessment, should be included as additional measures.

REFERENCES

- Baker, D. (2011). Student perceptions of teachers: An analysis of the MET project. Hanover Research.
- Brookhart, S. M. (2009). Assessment theory for college classrooms. *New Directions for Teaching and Learning*, 118, 5–14. <https://doi.org/10.1002/tl.352>
- Centre for Development and Enterprise. (2015). Teacher evaluation: Improving performance and accountability. CDE.

- Cingos, J. (2015). Classroom observation and teacher effectiveness: A study of value-added models. *Journal of Educational Assessment*, 22(3), 1–10.
- Chan, W., Marton, F., Tse, H., & King, R. (2014). Student evaluation of teaching: Teacher performance and student achievement. *Educational Psychology*, 34(5), 1–15. <https://doi.org/10.1080/01443410.2014.915929>
- Darling-Hammond, L. (2010). Performance assessment and accountability: Policy implications. Stanford Center for Opportunity Policy in Education.
- Dewey, J. (1986). *Experience and education*. Macmillan.
- Dolton, P., & Gutierrez, A. (2011). Performance-related pay and teacher effectiveness: Evidence from developing countries. World Bank Policy Research Paper.
- Fahmi, M., Maulana, A., & Yusuf, A. (2011). Teacher certification in Indonesia: A confusion of means and ends. *Education Economics*, 19(3), 341–364. <https://doi.org/10.1080/09645292.2011.585017>
- Garn, G. (2001). Market accountability and parental surveys in education. In Williams, A., & Engel, L. (2013). *Accountability in education systems*. Routledge.
- Hanover Research. (2012). *Best practices in teacher and principal evaluation*. Hanover Research.
- Hanover Research. (2013). *Student perception surveys and teacher assessments*. Hanover Research.
- Haryanto, E., Suryadarma, D., & Beatty, A. (2016). The effects of teacher certification on student outcomes in Indonesia. *Asian Development Review*, 33(1), 1–24.
- Hastuti, D., Toyamah, N., Suryadarma, D., & Reza, S. (2009). *Teacher certification in Indonesia: A reality check*. SMERU Research Institute.
- Henderson-Montero, D., Sharp, J., & Kimmel, H. (2003). Multiple measures in educational assessment. *Educational Assessment Journal*, 9(2), 47–65.
- Hull, J. (2013). *The principal perspective: Full report*. Center for Public Education.
- Jalal, F. (2009). *Teacher quality and certification program in Indonesia*. Indonesian Ministry of National Education Report.
- Jensen, E., & Minke, K. (2017). Family involvement in student academic achievement. *Journal of Educational Psychology*, 109(2), 234–250.
- Kusumawardhani, P. (2017). Teacher certification program in Indonesia: A promise unfulfilled? *Education Economics*, 25(6), 590–618. <https://doi.org/10.1080/09645292.2017.1315055>
- Lee, J., & Waimier, P. (2002). Accountability in teacher evaluation systems. *Educational Policy Analysis Archives*, 10(1), 1–21.
- Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, 46(3), 517–528.
- Lizer, K. (2013). Teacher preparation and curriculum challenges in developing countries. *International Journal of Education Development*, 33(2), 25–32.
- Mangiante, E. (2011). Teachers matter: Measures of teacher effectiveness in low-income schools. *Educational Policy Analysis Archives*, 19(10), 1–23.
- Martinez, J. F., Borko, H., & Stecher, B. (2016). Measuring instructional practice in science using classroom observation and teacher logs. *Educational Assessment*, 21(3), 1–18.
- MET Project. (2011). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill & Melinda Gates Foundation.
- Ministry of National Education (MONA). (2008). *Teacher qualification report*. Jakarta: Government of Indonesia.

- Muñoz, M., & Dossett, D. (2016). Multiple measures of teacher effectiveness and improvement of teaching practice. *Educational Assessment, Evaluation and Accountability*, 28(2), 1–20.
- Murphy, P. (2013). *Teacher evaluation in practice: Policy and evidence*. RAND Corporation.
- Muthim, R. (2014). Curriculum changes and teacher performance: Case study in Indonesia. *Journal of Curriculum Studies*, 46(4), 511–528.
- National Education Association (NEA) Education Policy and Practice Department. (n.d.). *Multiple measures of student learning*. NEA.
- OECD. (2015). *Education at a glance 2015: OECD indicators*. OECD Publishing.
- Persada, A., Yuliana, R., & Wahyudi, D. (2016). Parental involvement in student achievement: A case study from Indonesia. *Indonesian Journal of Education Review*, 3(2), 45–57.
- Peterson, K. (1987). Teacher evaluation: A comprehensive framework for supervisors. *Theory into Practice*, 26(1), 1–6.
- PISA. (2009). *PISA 2009 results: What students know and can do*. OECD.
- Schafer, W. D. (2003). Multiple measures and accountability. *Educational Measurement: Issues and Practice*, 22(2), 2–12.
- Strunk, K. O., Weinstein, T. L., & Makkonen, R. (2014). Choosing multiple measures for teacher evaluation. *Educational Policy*, 28(5), 1–28.
- Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, 12(4), 78–84.
- The Jakarta Post. (2013, November 27). Teachers' certification fails to improve education quality. *The Jakarta Post*.
- Williams, A., & Engel, L. (2013). Market accountability and education. *Comparative Education Review*, 57(4), 636–664.
- World Bank. (2012). *Teacher reform in Indonesia: The role of politics and evidence in policy making*. World Bank Publications.
- Warring, D. (2015). Teacher evaluation and value-added models: A policy analysis. *Education Policy Review*, 87(2), 123–135.
- Waxman, H. C., & Huang, S. L. (1999). Classroom observations of effective teaching in urban schools. *Journal of Educational Research*, 92(2), 95–103.
- www.apbn.org. (2016). *Laporan realisasi anggaran pendidikan 2016*.