JIPPG
Published by AEDUCIA

Jurnal Indonesia Pendidikan Profesi Guru
Volume 2, Issue 2, 2025, 103-114
Available online at: https://ojs.aeducia.org/index.php/jippg

E- ISSN 3063-0797
P- ISSN 3063-0827
Open Access Journals

The Exploration of Psychometric Models in Teachers Performance Evaluation

Josephine Obiageli Okafor
Nwafor Orizu College of Education Nsugbe, Nigeria

A B S T R A C T

Background: Teacher performance evaluation plays a vital role in improving educational quality. However, achieving accuracy, fairness, and effectiveness in such evaluations remains a challenge. Psychometric models offer promising tools to enhance the validity and reliability of teacher assessments. **Objective:** This study aims to explore the application of psychometric models in evaluating teacher performance and to investigate how these models can improve the overall assessment process. **Method:** The research involves an in-depth analysis of key psychometric models, including Classical Test Theory (CTT), Item Response Theory (IRT), Generalizability Theory (GT), Structural Equation Modeling (SEM), and Multilevel Modeling (MLM). The integration of these models into teacher evaluation systems is critically examined. **Result:** Findings indicate that each model offers unique advantages for improving assessment accuracy and reliability. However, limitations such as potential biases, data interpretation challenges, and ethical concerns must be carefully addressed. Contextual factors such as teacher characteristics, institutional policies, and opportunities for professional development significantly affect the outcomes of psychometric-based evaluations. **Conclusion:** Psychometric models can strengthen teacher evaluation systems when applied thoughtfully and ethically. Their effective use requires alignment with institutional goals and attention to contextual influences. **Contribution:** This study contributes to the field of educational assessment by synthesizing current knowledge on psychometric models and offering insights into their practical, ethical, and policy-related implications in the context of teacher performance evaluation.

K E Y W O R D S

Exploration; Psychometric Models; Teachers Performance; Evaluation

A R T I C L E H I S T O R Y

Received: April 09, 2025
Revised: April 28, 2025
Accepted: May 21, 2025
Published: May 26, 2025

C O N T E N T

1. INTRODUCTION

Teachers' performance evaluation is an essential aspect of education systems, aimed at ensuring the quality of instruction and fostering professional development. Evaluating teachers effectively remains a persistent challenge due to the complexities of teaching as a profession and the diverse contexts in which teachers operate. Modern evaluation frameworks have increasingly adopted psychometric models to address these complexities, leveraging their ability to provide reliable, valid, and objective measures of performance (Kasman & Lubis, 2022). Teacher evaluation has evolved significantly, transitioning from traditional approaches to contemporary data-driven methods. Traditional methods, such as classroom observations and supervisor appraisals, were often criticized for their subjectivity and inconsistency. They relied heavily on qualitative feedback, which lacked standardization and reproducibility (Ghaffarian-Asl & Osam, 2021).

* **Corresponding Author:** Josephine Obiageli Okafor, josephine.obiageli.okafor@nocen.edu.ng
Department of French, School of languages, Nwafor Orizu College of Education Nsugbe Anambra State Nigeria
Address: No. 1 college road, Abata, Nsugbe 432108, Anambra, Nigeria

How to Cite (APA Style 7th Edition):

Okafor J. O. (2025). Exploration of Psychometric Models in Teachers Performance Evaluation. *Jurnal Indonesia Pendidikan Profesi Guru, 2*(2), 97-108. https://ojs.aeducia.org/index.php/jippg/article/view/259

Contemporary evaluation systems aim to balance qualitative and quantitative measures by incorporating advanced assessment tools and frameworks. Current trends include competency-based assessments, student performance metrics, and feedback mechanisms that integrate self-assessments and peer reviews (Crespi et al, 2022). Despite these advancements, challenges persist, including cultural biases, resource constraints, and resistance to adopting new evaluation methods. The integration of psychometric models addresses some of these issues by offering robust statistical methods to evaluate teacher performance objectively and consistently (Foley et al, 2023).

Psychometric models have become instrumental in education, particularly in teacher evaluations. These models provide a systematic approach to measuring latent traits, such as teaching competence and effectiveness, using statistical methods to ensure reliability, validity, and fairness (Pellert et al, 2024). Psychometric models contribute to a more comprehensive understanding of teaching practices and their impact on student outcomes. The reliability of psychometric models lies in their ability to minimize subjective biases. For instance, Classical Test Theory (CTT) evaluates overall performance scores, while Item Response Theory (IRT) analyzes individual test items to provide diagnostic feedback (Zhang et al, 2023). Similarly, Generalizability Theory (GT) explores multiple sources of error variance, offering more precise reliability estimates. The adoption of these models ensures that evaluations are not only consistent but also sensitive to the diverse contexts in which teachers operate.

Moreover, psychometric models support professional development by identifying specific areas where teachers require improvement. For example, Structural Equation Modeling (SEM) and Multilevel Modeling (MLM) enable evaluators to identify the direct and indirect factors influencing teacher effectiveness, fostering targeted interventions (Hall & Malmberg, 2020). This emphasis on data-driven insights enhances accountability, equity, and the overall quality of education systems. The purpose of this essay is to critically examine the role of psychometric models in teachers' performance evaluation, focusing on their strengths, limitations, and implications for educational practice and policy.

2. METHOD

2.1 Research Design

The research method used was a Systematic Literature Review (SLR); this design was used to conduct a structured and comprehensive review based on a specific protocol. The focus of the literature review includes (1) What psychometric models are used in teacher performance evaluation?; (2) How is the effectiveness of these models?; (3) What are the advantages and disadvantages of the psychometric models used?

2.2 Inclusion and Exclusion Criteria

Inclusion criteria include (1) Articles focusing on psychometric models in teacher performance evaluation; (2) Articles published in the last 10-15 years; (3) Articles in indexed journals (Scopus, WoS, or others); (4) Relevant quantitative, qualitative, or mixed studies.

Exclusion criteria included (1) Literature not directly related to teacher performance evaluation, (2) Articles that only discussed psychometric theory without application, and (3) Articles in languages inaccessible to researchers.

2.3 Data Source

The source of data for this research is from Electronic Databases, Scopus, Web of Science, ERIC, Google Scholar, ResearchGate, The keywords used in the data collection process included “psychometric models” AND “teacher performance evaluation,” “teacher assessment” AND “psychometric techniques,” “item response theory” OR “generalizability theory” AND “teacher evaluation.”

2.4 Literature Search and Selection Process

The stages of the literature search and selection process included (1) Identification, which involved using keywords to search for articles in various databases; (2) Screening, reading titles and abstracts to determine the relevance of the articles; (3) Eligibility evaluation, which involved reading full texts for articles that passed the screening stage and evaluating whether they met the inclusion criteria.

2.5 Data Extraction

The data extraction process included (1) the psychometric model used (Item Response Theory, Classical Test Theory, Generalizability Theory); (2) the educational context (e.g., primary, secondary, or higher education); (3) key findings related to the effectiveness of the model.

2.5 Data Analysis and Synthesis

The data analysis process included five stages: (1) Thematic analysis; (2) Identification of key themes emerging from the literature; (3) Types of psychometric models used; (4) Criteria for validity and reliability applied; (5) Application in the context of teacher performance evaluation.

3. RESULT AND DISCUSSION

3.1 Result

1) The Concept of Psychometric Models

Psychometric models provide statistical and mathematical frameworks to measure latent psychological constructs such as abilities, attitudes, and performance (Pellert et al, 2024). These models are central in educational research and practice, particularly in designing and analyzing assessments for evaluating teachers, students, and educational systems. Psychometric models enable researchers to quantify abstract traits that cannot be directly observed, ensuring objectivity and precision in evaluations. The application of psychometric models in education emphasizes understanding and improving teaching effectiveness. Using these models will help educators to identify specific areas of strength and weakness, thereby tailoring interventions to address gaps in teaching practices. Additionally, psychometric models allow for the comparison of performance across different contexts, ensuring fairness and consistency (Ferretti-Rebustini, 2023).

Psychometric models are defined as statistical tools that facilitate the measurement of psychological attributes through standardized tests and assessments (Swan et al, 2023). These models rely on foundational principles of reliability, validity, and fairness to ensure accurate and meaningful results. Reliability refers to the consistency of measurement outcomes, while validity ensures that the assessment accurately measures what it intends to. Fairness ensures that results are unbiased and equitable for all test-takers (Ibrahim, 2023). Key features of psychometric models include their ability to handle measurement errors, provide diagnostic insights, and establish generalizability. For instance, models such as Item Response Theory (IRT) evaluate individual test items to identify their difficulty and discrimination levels, enabling a more detailed understanding of performance (Ayanwale et al, 2024). Another critical feature is their scalability, allowing these models to be applied across diverse educational settings, from classroom-level assessments to large-scale standardized testing programs.

2) Psychometric Principles and Standards

Psychometric models operate based on established principles and standards that guide their development, implementation, and interpretation. These principles ensure that assessments are reliable, valid, and fair, minimizing biases and errors in measurement. Reliability is achieved through repeated testing and analysis to confirm consistency across different conditions and populations (Izah et al, 2024). Validity, on the other hand, is established through careful test design and empirical evidence demonstrating that the assessment accurately measures the intended construct.

Fairness is a critical standard in psychometric modeling, particularly in education, where assessments significantly impact students, teachers, and institutions. Ensuring fairness involves designing tests that are culturally sensitive and free from discriminatory content (Alordiah & Oji, 2024). Adherence to international guidelines, such as those provided by the Standards for Educational and Psychological Testing, ensures that psychometric models meet ethical and technical requirements. Another vital principle is the ability of psychometric models to address measurement errors. GT, for example, extends beyond traditional reliability concepts by analyzing multiple sources of error variance, such as differences in test administration or scoring conditions (Olose, 2024). This approach provides a more comprehensive understanding of assessment quality, enabling evaluators to make more informed decisions.

3) Teachers' Performance Evaluation

Teachers' performance evaluation is a systematic process aimed at assessing the effectiveness of teachers in achieving educational goals and improving students' learning outcomes. It serves multiple purposes, including

accountability, professional development, and resource allocation (Edem & Edem, 2023). Effective evaluation frameworks ensure that teachers receive constructive feedback to enhance their instructional practices while identifying areas for improvement. The evaluation process also helps educational institutions monitor teaching quality, align performance with institutional objectives, and recognize outstanding contributions (Adigun et al, 2022). Historically, performance evaluations primarily focused on classroom observations and student academic achievements. However, contemporary approaches incorporate a range of data sources, such as teacher self-assessments, peer reviews, and feedback from students and parents (Ehinola & Akomolafe, 2022). These approaches aim to provide a comprehensive understanding of teaching effectiveness. Despite their significance, teacher evaluations remain a complex undertaking, influenced by various contextual, professional, and systemic factors.

4) Traditional Methods of Performance Evaluation

Traditional methods of teacher performance evaluation were largely subjective and often relied on limited assessment tools. The most common method involved classroom observations conducted by school administrators. In these observations, principals or supervisors would rate teachers based on their instructional delivery, classroom management, and adherence to institutional guidelines (Ozigi & Onyeukwu, 2022). However, the reliability and validity of such evaluations have been questioned due to biases, inconsistent criteria, and a lack of standardized protocols. Another traditional approach involved evaluating teachers based on student outcomes, particularly test scores. While this method provided a measurable indicator of teacher effectiveness, it often failed to account for external factors influencing student performance, such as socioeconomic status, parental involvement, and learning disparities (Unegbu et al, 2022). Additionally, reliance on standardized test scores placed undue pressure on teachers to focus on test preparation rather than holistic learning.

Other methods included teacher portfolios and annual performance appraisals. Although portfolios allowed teachers to showcase their achievements, they were time-consuming to prepare and assess. Performance appraisals, on the other hand, lacked depth and failed to provide actionable feedback for teachers' professional growth (Lawal et al, 2022). These limitations highlighted the need for more robust and comprehensive evaluation systems.

5) Current Trends in Performance Evaluation Systems

Modern teacher performance evaluation systems emphasize a multi-faceted approach, integrating both qualitative and quantitative measures. One prominent trend is the use of multiple evaluation sources, including classroom observations, student performance data, teacher self-assessments, and peer reviews. This approach ensures a balanced and holistic evaluation of teaching effectiveness (Adigun et al, 2022). Another significant trend is the incorporation of technology-driven assessments. Digital tools and learning management systems now provide real-time data on teaching practices and student engagement, enabling more accurate performance analysis. Online platforms facilitate continuous monitoring and feedback, reducing the administrative burden associated with traditional evaluations (Falola et al, 2022).

The adoption of student surveys as a component of teacher evaluation is also gaining traction. These surveys capture students' perceptions of teaching practices, offering valuable insights into teacher-student relationships and instructional quality (Bada & Jita, 2023). Additionally, professional development is increasingly integrated into evaluation systems, linking performance outcomes to training opportunities that enhance teachers' skills. Data-driven decision-making has further refined evaluation systems. Educational institutions now analyze large datasets to identify trends in teacher performance, enabling evidence-based interventions and policy decisions. These trends reflect a shift toward transparency, accountability, and teacher-centered improvement (Hurley, 2024).

6) Challenges in Evaluating Teacher Performance

Despite advancements in performance evaluation systems, several challenges persist. One key issue is the subjectivity of evaluation criteria. Classroom observations, while valuable, remain prone to personal biases and inconsistencies among evaluators. Ensuring standardized and objective criteria across evaluators remains a significant challenge. Another challenge is the over-reliance on student achievement scores as a measure of teacher effectiveness. While test scores provide quantifiable outcomes, they fail to account for contextual factors beyond teachers' control, such as resource availability, class size, and student demographics (Babayemi, 2022). This approach risks misrepresenting teacher performance and fostering a culture of "teaching to the test." Resource constraints also impede effective teacher evaluations. Many educational institutions lack the infrastructure, funding, and personnel needed to implement comprehensive and technology-driven evaluation systems. In such cases, evaluations may be inconsistent, incomplete, or poorly executed. Additionally, the lack of teacher buy-in presents a challenge. Teachers often perceive evaluations as punitive rather than developmental, leading to resistance and

mistrust. For evaluation systems to succeed, teachers must view them as tools for growth rather than judgment (Murwaningsih, 2024).

3.2. Discussion

1) Psychometric Models and Their Impact on Educational Policy

Psychometric models play a crucial role in ensuring the reliability, validity, and fairness of teacher performance evaluation systems. Different models are employed to measure teacher effectiveness, each offering unique strengths and addressing specific limitations of traditional evaluation methods. Prominent models include Classical Test Theory (CTT), Item Response Theory (IRT), Generalizability Theory (GT), Structural Equation Modeling (SEM), and Multilevel Modeling (MLM).

Classical Test Theory (CTT): Classical Test Theory (CTT) is one of the foundational psychometric models used in educational evaluation. It focuses on understanding the relationship between an observed score and a true score by assuming that any observed score consists of a true score and measurement error (Ekpenyong et al, 2022). In teacher evaluation, CTT is applied to analyze test scores and other performance indicators, providing insights into the reliability and consistency of evaluation tools. For example, CTT examines internal consistency, split-half reliability, and test-retest reliability to ensure that assessments yield stable and accurate results. While CTT is widely used due to its simplicity and ease of interpretation, it has notable limitations. The model assumes that errors are random and consistent across all test items, which may not hold in real-world evaluations. Additionally, CTT's reliance on test scores restricts its ability to address item-level analysis, leading to limited diagnostic information about teacher performance (Ayanwale et al, 2022).

Item Response Theory (IRT): Item Response Theory (IRT) improves upon CTT by focusing on the relationship between an individual's ability and their probability of correctly responding to test items. Unlike CTT, IRT considers the difficulty of each item and the ability of the test taker, offering a more precise evaluation of performance (Ayanwale et al, 2024). In teacher evaluation, IRT enables the development of assessments that are tailored to accurately measure specific teaching competencies. For instance, IRT identifies poorly performing test items, ensuring that evaluations are both fair and valid. IRT models, such as the one-parameter (Rasch), two-parameter, and three-parameter models, allow evaluators to analyze test items based on difficulty, discrimination, and guessing factors (Oyiborhoro, 2023). This granularity makes IRT particularly useful for large-scale teacher evaluations where robust and detailed information is required. However, implementing IRT is complex and requires sophisticated statistical tools and expertise, which may limit its adoption in resource-constrained educational systems.

Generalizability Theory (GT): Generalizability Theory (GT) extends beyond CTT by providing a framework to analyze multiple sources of measurement error. GT considers various factors such as raters, test items, and contexts that contribute to score variability, offering a more holistic approach to reliability (Imasuen & Adeosun, 2023). In teacher evaluations, GT helps identify and quantify the effects of different error sources, ensuring that the evaluation outcomes are dependable and generalizable across contexts. For example, GT allows evaluators to assess the consistency of classroom observations conducted by multiple raters, thereby addressing inter-rater reliability issues (Ojedokun et al, 2022). Modeling multiple facets of the evaluation process gives GT ability to guide improvements in assessment design. However, GT's complexity and requirement for extensive data may pose challenges in its practical application.

Structural Equation Modeling (SEM): Structural Equation Modeling (SEM) is a powerful statistical technique that combines factor analysis and regression modeling to evaluate relationships among observed and latent variables. In teacher evaluations, SEM is used to assess the impact of various teaching attributes such as instructional methods, classroom management, and student outcomes on overall performance (Kineber et al, 2022). SEM provides a comprehensive understanding of how underlying factors contribute to teacher effectiveness, enabling the identification of strengths and areas for improvement. A key advantage of SEM is its ability to test complex models involving multiple variables simultaneously. For example, it can analyze the relationships between teacher qualifycations, teaching practices, and student achievement outcomes. However, SEM requires large sample sizes and advanced statistical tools, which may limit its applicability in smaller educational settings.

Multilevel Modeling (MLM): Multilevel Modeling (MLM) is particularly valuable in teacher evaluations because it accounts for the hierarchical structure of educational data. Teachers operate within classrooms, schools, and districts, creating nested data that traditional statistical models cannot handle effectively (Akindejia et al, 2024). MLM allows evaluators to analyze teacher performance while considering contextual factors, such as school resources, class size, and student demographics. For instance, MLM distinguishes between teacher-level and school-level effects, offering insights into how institutional factors influence teacher performance. This capability is

essential for understanding the broader systemic influences on evaluation outcomes. Despite its advantages, MLM requires significant computational resources and technical expertise, which can be a barrier for widespread adoption.

2) Application of Psychometric Models in Teacher Performance Evaluation

The integration of psychometric models into teacher performance evaluation systems has revolutionized the process of measuring teacher effectiveness. These models ensure that evaluations are objective, reliable, and valid, addressing limitations of traditional assessment approaches (Kenea et al, 2024). Utilizing advanced statistical techniques makes it possible for psychometric models to provide a framework for assessing teachers' competencies, instructional practices, and their impact on student learning outcomes. Psychometric models are embedded in teacher evaluation systems through standardized assessments, observational tools, and self-assessment frameworks. For instance, models like Item Response Theory (IRT) and Generalizability Theory (GT) are used to design teacher evaluations that measure both observable behaviors and latent constructs, such as pedagogical knowledge and student engagement (Kasman & Lubis, 2022). These models improve the accuracy of evaluation tools by analyzing the difficulty and reliability of assessment items and identifying sources of error.

In practical implementation, IRT helps refine evaluation instruments to ensure that items accurately differentiate between varying levels of teacher competence. Similarly, Multilevel Modeling (MLM) addresses the hierarchical nature of educational data, accounting for contextual factors like class size, school infrastructure, and student demographics (Pellert et al, 2024). Through this integration, teacher evaluation systems become fairer and more adaptable to diverse teaching environments Observation protocols also incorporate psychometric principles. Tools like the Framework for Teaching (FFT) and Classroom Assessment Scoring System (CLASS) rely on psychometric methods to establish inter-rater reliability and consistency in evaluator judgments. These tools provide quantitative insights into instructional quality while minimizing subjective biases (Ferretti-Rebustini, 2023).

The adoption of psychometric models in teacher evaluation offers several benefits. First, these models improve the validity and reliability of assessment tools by minimizing measurement errors and biases (Hall & Malmberg, 2020). Unlike traditional methods, psychometric approaches provide a scientific basis for evaluating teacher performance, ensuring fairness and consistency across diverse educational contexts. Second, psychometric models facilitate data-driven decision-making in teacher development. Identifying specific areas of strength and weakness will help evaluation systems inform targeted professional development programs, ultimately enhancing teaching quality and student outcomes (Zhang et al, 2023). Third, psychometric models promote accountability and transparency in evaluation processes. Tools based on IRT, GT, and MLM provide objective measures of teacher effectiveness, reducing reliance on subjective assessments and fostering trust among educators and stakeholders.

3) Strengths and Limitations of Psychometric Models in Teacher Evaluation

The use of psychometric models in teacher evaluation offers significant advantages, although they are not without limitations. While these models enhance the objectivity, reliability, and validity of performance assessments, they may also introduce technical, ethical, and contextual challenges that need to be addressed for effective implementation (Pellert et al, 2024). Psychometric models bring notable strengths to teacher evaluation systems. One primary advantage lies in their ability to ensure reliability and validity in assessments. Unlike traditional evaluation methods, psychometric approaches such as Item Response Theory (IRT), Generalizability Theory (GT), and Structural Equation Modeling (SEM)-provide precise and objective measures of teacher performance by accounting for measurement errors and external factors (Foley et al, 2023). This precision helps to produce consistent results across different evaluators and settings, fostering trust and fairness in the evaluation process.

Furthermore, psychometric models enable the identification of specific strengths and weaknesses in teacher competencies. For example, IRT can pinpoint particular areas where teachers excel or require development, thereby supporting targeted professional growth initiatives (Crespi et al, 2022). This data-driven approach facilitates personalized interventions and professional development programs that ultimately enhance teaching quality. Another strength of psychometric models is their adaptability to diverse educational contexts. Models like Multilevel Modeling (MLM) account for contextual influences, such as school location, class size, and socio-economic factors, ensuring that teacher evaluations are equitable and representative of real-world conditions (Falola et al, 2022). Similarly, these models offer scalable solutions for large-scale evaluations, making them suitable for policy-driven assessments across districts or states.

Despite their strengths, psychometric models face several criticisms and limitations. One major critique is their complexity and technical nature, which can pose challenges for implementation, particularly in under-resourced educational systems. Advanced models like SEM and MLM require significant expertise in statistical analysis, which

may be lacking in many institutions (Adigun et al, 2022). Another limitation relates to the reliance on quantitative data. Psychometric approaches often prioritize measurable variables, potentially overlooking qualitative aspects of teaching, such as creativity, classroom relationships, and emotional intelligence (Lawal et al, 2022). As a result, these models may fail to capture the full scope of teacher effectiveness.

Additionally, psychometric models are vulnerable to biases and measurement errors. For example, evaluations based on standardized tests may disadvantage teachers working with students from marginalized or low-income backgrounds, where external factors impact learning outcomes (Unegbu et al, 2022). In such cases, evaluations may unfairly reflect contextual challenges rather than true teaching effectiveness. Furthermore, the ethical implications of psychometric models raise concerns regarding teacher accountability and stress. Over-reliance on data-driven evaluations can create a high-pressure environment, where teachers prioritize performance metrics over holistic education.

To address the limitations and biases inherent in psychometric models, several strategies can be employed. First, integrating mixed-method approaches-combining quantitative psychometric models with qualitative evaluations-can provide a more comprehensive assessment of teacher performance (Ozigi et al, 2022). For example, classroom observations and teacher self-assessments can complement psychometric data to capture non-measurable aspects of teaching. Second, models like Multilevel Modeling (MLM) and Generalizability Theory (GT) can mitigate biases by accounting for contextual factors, such as school resources, socio-economic challenges, and cultural influences (Ehinola & Akomolafe, 2022). Psychometric evaluations has become more equitable and reflective of diverse teaching environments.

Third, ensuring regular updates and refinements of evaluation tools can help reduce errors and biases. Ongoing validation studies and pilot testing can improve the accuracy and fairness of psychometric instruments over time (Adigun et al, 2022). Lastly, professional training for educators and evaluators is essential to address technical challenges associated with psychometric models. Providing training on interpreting evaluation results and understanding statistical methods can enhance the utility and effectiveness of these models in practice (Edem & Edem, 2023).

4) Factors Influencing Teachers' Performance Assessment

The assessment of teachers' performance is influenced by a combination of contextual, personal, institutional, and policy-related factors as shown in Figure 1. These elements play significant roles in determining the validity, reliability, and overall effectiveness of evaluation systems. Addressing these factors is essential for achieving fair and meaningful assessments that contribute to improved educational outcomes.

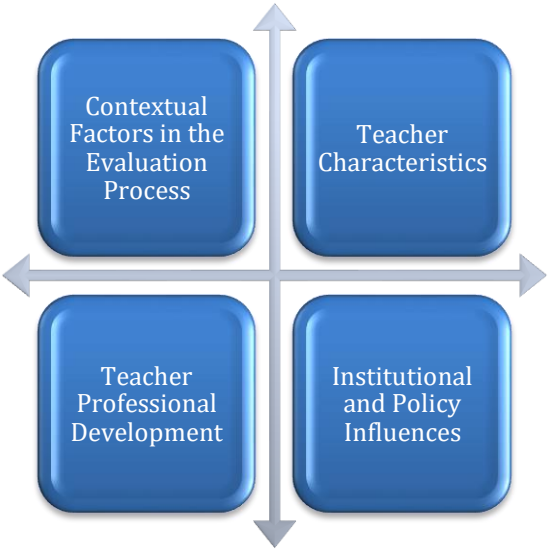

Figure 1. Factors Influencing Teachers' Performance Assessment

Contextual factors significantly impact teacher performance assessments, as the teaching environment often shapes the effectiveness of instructional delivery. Socio-economic status, geographical location, school resources, and class size all influence student performance, which is often used as a proxy for evaluating teachers (Olose, 2024). For instance, teachers in under-resourced schools or rural areas may face challenges such as limited access to instructional materials, overcrowded classrooms, and low student readiness for learning. These conditions can

unfairly reflect on teacher performance when assessments rely solely on student outcomes (Alordiah & Oji, 2024). Cultural and social factors also affect teacher evaluations. In diverse settings, cultural expectations regarding teaching styles and teacher-student relationships can create variations in assessment outcomes. For example, teachers working in multicultural classrooms may adopt differentiated teaching approaches, which might not align with standardized evaluation frameworks. As such, effective teacher evaluation systems must account for contextual differences to ensure fairness and equity.

Teacher-specific attributes, including experience, qualifications, and professional development, are critical determinants of performance assessment outcomes. Experienced teachers are often more skilled in managing classrooms, delivering content, and engaging students effectively, which translates into higher performance ratings (Izah et al, 2024). However, newly qualified teachers, despite their potential, may lack the practical experience to achieve similar outcomes within rigid evaluation systems. Professional development plays a central role in improving teacher performance and influencing evaluation outcomes. Continuous training programs provide opportunities for teachers to enhance their instructional strategies, adapt to new technologies, and address individual weaknesses (Ayanwale et al, 2024). Evaluations that incorporate measures of professional growth can highlight the dynamic nature of teacher effectiveness over time, moving beyond static assessments of competence. Teacher motivation and job satisfaction also influence performance assessments. Teachers who receive adequate support, recognition, and incentives are more likely to excel in their roles. Conversely, teachers operating under stressful conditions or experiencing burnout may underperform, resulting in lower evaluation scores. Thus, understanding teacher characteristics alongside their professional development journey is crucial for a holistic assessment.

Institutional structures and educational policies are significant drivers of teacher performance evaluation. Institutional factors, such as school leadership, organizational culture, and resource availability, create enabling or constraining environments that impact teaching effectiveness. Supportive school leadership, for instance, fosters a culture of collaboration, mentorship, and accountability, which positively influences teacher performance (Ibrahim, 2023). Schools with limited funding or outdated resources, however, hinder teacher effectiveness, regardless of individual effort. Educational policies governing teacher evaluation systems also play a pivotal role in shaping assessment processes. Policies that emphasize student achievement as the primary metric for evaluating teachers can inadvertently create high-stakes environments, where teachers focus on test preparation rather than holistic learning (Kenea et al, 2024). Alternatively, policies that incorporate multiple assessment methods such as classroom observations, peer reviews, and professional portfolios allow for more comprehensive and fair evaluations.

The alignment of teacher evaluation systems with broader educational reforms is essential for sustainable improvements in teaching quality. Policies must balance accountability with professional growth, ensuring that evaluations serve both formative and summative purposes (Akindejia & Madondo, 2024). Additionally, clear guidelines and standardized frameworks are necessary to maintain consistency and fairness in assessment processes across institutions.

5) Psychometric Models and Their Impact on Educational Policy

Psychometric models play a crucial role in shaping educational policy, particularly in the context of teacher performance evaluations. These models, including Classical Test Theory (CTT), Item Response Theory (IRT), and Generalizability Theory (GT), provide quantitative frameworks that inform policy decisions related to teacher quality, accountability, and professional development (Kineber et al, 2022). By employing these models, educational policymakers can ensure that teacher evaluations are more reliable, valid, and equitable, which directly influences the development of policies aimed at improving educational outcomes. Psychometric data offers valuable insights into teacher performance, allowing policymakers to make informed decisions about resource allocation, teacher training programs, and curriculum development. For instance, when teacher evaluations are integrated with psychometric models, policymakers can identify trends in teacher effectiveness across different educational contexts, such as urban and rural schools or high- and low-performing districts (Ojedokun et al, 2022). These findings enable the creation of targeted interventions and reforms that address specific needs within the education system.

Teacher evaluation is a critical tool in shaping educational policy as it serves as a measure of teaching quality, which directly affects student achievement. In many countries, educational policies prioritize teacher effectiveness as a key indicator of school success. Teacher evaluations that incorporate psychometric models help policymakers make data-driven decisions about teaching standards and performance expectations (Imasuen & Adeosun, 2023). The data derived from these evaluations can inform policy decisions about teacher recruitment, certification, and compensation, as well as the development of professional learning programs. For example, psychometric models can provide evidence that certain teacher characteristics—such as years of experience, subject expertise, or professional

development are linked to improved student outcomes. This information can influence policies on teacher training programs, professional development initiatives, and retention strategies (Oyiborhoro, 2023). In this way, teacher evaluation not only contributes to the improvement of individual teacher performance but also plays a broader role in shaping the educational landscape.

The integration of psychometric models into teacher evaluation systems provides evidence that informs educational decision-making at multiple levels. Findings from psychometric analyses offer objective data that can guide the formulation of policies aimed at improving teaching quality. For example, the application of IRT in large-scale teacher assessments allows policymakers to examine how individual items on performance evaluations contribute to the overall assessment of teacher effectiveness (Ayanwale et al, 2024). These insights can lead to changes in evaluation methods, ensuring that they more accurately reflect the skills and knowledge required for effective teaching. Moreover, psychometric models enable decision-makers to evaluate the impact of various educational interventions. If a policy or program is implemented to improve teacher performance, psychometric data can be used to assess its success and determine areas that need further improvement. The ability to monitor and measure the effectiveness of educational policies through psychometric evaluations is an essential component of evidence-based policymaking (Ayanwale et al, 2022).

While psychometric models provide valuable insights, their use in educational policy raises several ethical concerns. One major issue is the potential for bias in the evaluation process. Psychometric models, particularly those that rely heavily on standardized tests or quantitative measures, may inadvertently disadvantage teachers in schools with fewer resources or students from disadvantaged backgrounds. If these models are not properly adjusted to account for contextual differences, they may unfairly penalize teachers working in challenging environments (Ekpenyong et al, 2022). Additionally, the use of psychometric data for policy decision-making can contribute to high-stakes evaluations, where teachers may face undue pressure to perform based on limited metrics. This can create an environment where teaching to the test becomes the primary focus, overshadowing other important aspects of education, such as student engagement, creativity, and holistic development (Murwaningsih, 2024). To mitigate these ethical concerns, policymakers must ensure that psychometric data is used in conjunction with other forms of assessment, such as qualitative evaluations, to provide a well-rounded view of teacher effectiveness.

Another ethical consideration is the privacy and confidentiality of data. Teacher evaluations, particularly those based on psychometric data, often involve sensitive information about individual performance. Ensuring that this data is handled ethically, with respect for teachers' privacy, is essential to maintaining trust in the evaluation system (Babayemi, 2022). Furthermore, policymakers must be transparent about how evaluation data will be used, ensuring that teachers are aware of the potential consequences of their evaluation results.

4. IMPLICATIONS AND CONTRIBUTIONS

Research Implications. This research can expand the understanding of how psychometric models, such as Classical Test Theory (CTT), Item Response Theory (IRT), and Generalizability Theory (GT), can be applied specifically to evaluate teacher performance. Provide an empirical basis for policy formulation related to data-driven rewards, promotions, or interventions in educational contexts, and assist in the development of a more objective, valid, and reliable teacher performance evaluation system for use by educational institutions.

Research Contribution. The results of this study provide recommendations for developing more comprehensive psychometric-based evaluation instruments. Guide educational institutions in selecting and implementing data-based evaluation models and encouraging technology integration in developing and implementing teacher performance evaluation models using a psychometric approach.

5. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Some articles or related research may only be available through paid journals or databases that are not accessible to all researchers, limiting the scope of literature to be reviewed. The literature used in the review may not be uniform in quality. Some sources could have biases or weak methodologies that affect the validity of the review results.

Recommendation. Future researchers can develop models that simultaneously evaluate multiple dimensions of teacher performance, such as pedagogical skills, innovativeness, and interpersonal skills, and adapt psychometric models for various cultural and geographical contexts, making the evaluation results more relevant and valid.

4. CONCLUSION

Psychometric models are essential tools for evaluating teacher performance, providing data-driven insights that support educational policy decisions. Models like Classical Test Theory (CTT), Item Response Theory (IRT), and Generalizability Theory (GT) help assess teacher effectiveness, guide professional development, and inform curriculum adjustments. These models offer benefits such as increased accuracy and fairness in evaluations, but they also face challenges, including potential biases and ethical concerns, particularly in high-stakes assessments. Integrating psychometric models in teacher evaluations helps policymakers make informed decisions about recruitment, retention, and support. However, biases related to socio-economic and school context factors can affect fairness. Future research should explore mixed-methods approaches, combining psychometric models with qualitative data such as classroom observations, to provide a more holistic evaluation. Longitudinal studies could further examine the long-term impact of psychometric-based evaluations on teacher development and student outcomes. Ethical considerations, including transparency and teachers' privacy, must also be addressed. Clear guidelines for the ethical use of psychometric data are necessary. As technology advances, incorporating artificial intelligence (AI) into psychometric evaluations could improve efficiency and accuracy, but ethical standards must be maintained. In conclusion, while psychometric models are valuable in teacher evaluation, they must be used responsibly and in combination with other methods to ensure fairness and integrity in the evaluation process.

ACKNOWLEDGMENTS

I would like to thank colleagues in the Department of French, School of Languages, Nwafor Orizu College of Education Nsugbe Anambra State Nigeria; they provided much information and constructive suggestions.

AUTHOR CONTRIBUTION STATEMENT

This literature review research on Exploration of Psychometric Models in Teachers Performance Evaluation results from a systematic review conducted by the author. The author is fully responsible for all data in this study.

CONFLICT OF INTEREST STATEMENT

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICAL APPROVAL STATEMENT

The author has approved the article to be published in the Jurnal Indonesia Pendidikan Profesi Guru (JIPPG) by following the Publication Ethics and Journal Policies.

REFERENCES

Adigun, J. O., Irunokhai, E. A., Onihunwa, J. O., Yusuf, A. S., Caleb, A. J., Yetunde, M. A., & Ilori, A. O. (2022). Development and evaluation of a web based system for students' appraisal on teaching performance of lecturers. *International Journal of Information Engineering and Electronic Business*, *14*(1), 25. http://dx.doi.org/10.5815/ijieeb.2022.01.03

Akindejia, E. B., & Madondo, E. (2024). Motivational drivers for distributors engaging in multi-level marketing in KwaZulu-Natal, South Africa. *The Retail and Marketing Review*, *20*(1), 114-126. https://zenodo.org/badge/DOI/10.5281/zenodo.12506228.svg

Alordiah, C., & Oji, J. (2024). Test Equating in Educational Assessment: A Comprehensive Framework for Promoting Fairness, Validity, and Cross-Cultural Equity. *Asian Journal of Assessment in Teaching and Learning*, *14*(1), 70-84. https://doi.org/10.37134/ajatel.vol14.1.7.2024

Ayanwale, M. A., Chere-Masopha, J., & Morena, M. C. (2022). The classical test or item response measurement theory: The status of the framework at the Examination Council of Lesotho. *International Journal of Learning, Teaching and Educational Research*, *21*(8), 384-406. https://doi.org/10.26803/ijlter.21.8.22

Babayemi, J. O. (2022). Assessment of Teaching Skills of In-Service Science Teachers and their Teaching Performance in Secondary Schools in Uyo Senatorial District, Akwa Ibom State, Nigeria. *KIU Journal of Social Sciences*, *7*(4), 125-131. https://ijhumas.com/ojs/index.php/niujoss/article/view/1356

Bada, A. A., & Jita, L. C. (2023). Student's rating of secondary school physics teachers' classroom practice: Implications for teaching and learning. *International Journal of Evaluation and Research in Education*, *12*(1), 477-486. http://doi.org/10.11591/ijere.v12i1.24078

Crespi, G., Abbà, I., & Corgnati, S. P. (2022). Innovative metrics to evaluate HVAC systems performances for meeting contemporary loads in buildings. *Energy Reports*, *8*, 9221-9231. https://doi.org/10.1016/j.egyr.2022.07.011

Edem, O. A., & Edem, O. C. (2023). Competence evaluation by public secondary school teachers: a panacea for students improved performance in calabar zone, nigeria. *Global Journal of Educational Research*, *22*(1), 59-65. https://doi.org/10.4314/gjedr.v22i1.6

Ehinola, G. B., & Akomolafe, B. B. (2022). In-service training programmes and mentoring for improving teachers' job performance in north senatorial District of Ondo State, Nigeria. *Int. J. Educ. Res. Rev*, *3*, 1-7. https://spectacularjournals.org/pdf/1642022392Eniola%20and%20Akomolafe.pdf

Ekpenyong, J. A., Owan, V. J., Ogar, J. O., & Undie, J. A. (2022). Hierarchical linear modelling of educational outcomes in secondary schools: What matters–teachers' or administrators' input?. *Cogent Education*, *9*(1), 2133491. http://dx.doi.org/10.1080/2331186X.2022.2133491

Falola, H. O., Ogueyungbo, O. O., Adeniji, A. A., & Adesina, E. (2022). Exploring sustainable e-learning platforms for improved universities' faculty engagement in the new world of work. *Sustainability*, *14*(7), 3850. https://doi.org/10.3390/su14073850

Ferretti-Rebustini, R. E. D. L. (2023). Psychometrics: applications in Nursing. *Revista Latino-Americana de Enfermagem*, *31*, e3992. http://dx.doi.org/10.1590/1518-8345.0000.3993

Foley, H., Bugarcic, A., Adams, J., Wardle, J., Leach, M., & Steel, A. (2023). Criteria for the selection, evaluation and application of traditional knowledge in contemporary health practice, education, research and policy: a systematic review. *Health Information & Libraries Journal*, *40*(3), 233-261. https://doi.org/10.1111/hir.12499

Ghaffarian Asl, S., & Osam, N. (2021). A study of teacher performance in English for academic purposes course: Evaluating efficiency. *SAGE Open*, *11*(4), 21582440211050386. http://dx.doi.org/10.1177/21582440211050386

Hall, J., & Malmberg, L. E. (2020). The contribution of Multilevel Structural Equation Modelling to contemporary trends in educational research. *International Journal of Research & Method in Education*, *43*(4), 339-347. https://doi.org/10.1080/1743727X.2020.1796066

Hurley, M. E. (2024). *Re-Centering Teachers in a New Ecology to Empower Teachers' Agency and School Change; Engaging Teachers in Critical Inquiry and Action Research Based in a Global Community of Practice Online* (Doctoral dissertation, Arizona State University). https://keep.lib.asu.edu/items/199166/metadata

Ibrahim, B. M. (2023). Psychometric properties analysis of mathematics West African Senior School Certificate Examination in Dala Education Zone, Kano State, Nigeria (2020-2022). *International Journal of Advanced Academic Research*, *9*(1), 199-212. https://www.ijaar.org/articles/v9n1/ijaar919.pdf

Imasuen, K., & Adeosun, P. K. (2023). Application of Generalizability Theory in Measurement Error in 2019 WAEC Mathematics Objective Examination in Benin Metropolis. *International Journal of Psychological and Brain Sciences*, *8*(2), 13-18. http://dx.doi.org/10.11648/j.ijpbs.20230802.11

Izah, S. C., Richard, G., Stanley, H. O., Ogwu, M. C., Sawyer, W. E., & Uwaeme, O. R. (2024). Prospects and application of multivariate and reliability analyses to one health risk assessments of toxic elements. *Toxicology and Environmental Health Sciences*, *16*(2), 127-134. http://dx.doi.org/10.1007/s13530-023-00199-0

Kasman, K., & Lubis, S. K. (2022). Teachers' performance evaluation instrument designs in the implementation of the new learning paradigm of the merdeka curriculum. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, *8*(3), 760-775. https://doi.org/10.33394/jk.v8i3.5674

Kenea, T. G., Mikire, F., & Negawo, Z. (2024). The relationship between the psychometric and performance properties of teacher-made tests and students' academic performance in Ethiopian public universities: a baseline survey study. *Cogent Education*, *11*(1), 2298049. http://dx.doi.org/10.1080/2331186X.2023.2298049

Kineber, A. F., Oke, A. E., Hamed, M. M., Rached, E. F., Elmansoury, A., & Alyanbaawi, A. (2022). A partial least squares structural equation modeling of robotics implementation for sustainable building projects: a case in Nigeria. *Sustainability*, *15*(1), 604. https://doi.org/10.3390/su15010604

Lawal, I.M., Kutty, S.R.M., Saleh, D., Kumar, V., Singh, P., Birniwa, A.H., Abubakar, S. and Jagaba, A.H., 2022, August. Performance evaluation of activated sugarcane bagasse for abattoir wastewater treatment. In *International Conference on Trends and Recent Advances in Civil Engineering* (pp. 495-507). Singapore: Springer Nature Singapore. http://dx.doi.org/10.1007/978-981-99-1886-7_42

Murwaningsih, T. (2024). The Influence of Teacher Professional Attitude, Welfare, Continuous SelfDevelopment, and Job Satisfaction on High School Teachers Performance. *International Journal of Instruction*, *17*(1), 229-252. https://e-iji.net/ats/index.php/pub/article/view/504

Ojedokun, O., Henschel, N., Arant, R., & Boehnke, K. (2022). Applying the theory of planned behaviour to littering prevention behaviour in a developing country (Nigeria). *Waste management*, *142*, 19-28. https://doi.org/10.1016/j.wasman.2022.02.006

Olose, E. O. (2024). *The Relationship Between Employee Perception of Performance Appraisal Fairness and Work Performance* (Doctoral dissertation, Grand Canyon University). http://dx.doi.org/10.5937/EkoIzazov1813017B

Oyiborhoro, A. V. (2023). Application of Item Response Theory in the Validation of Basic Science Test of Delta State Basic Education Certificate Examination. *International Journal of Research in Education and Sustainable Development*, *3*(7), 1-13. https://www.openjournals.ijaar.org/index.php/ijresd/article/view/27

Ozigi, O. A., & Onyeukwu, P. E. (2022). Performance Appraisal Fairness on Employees' Performance in the Nigeria Television Authority (Nta), Abuja, Nigeria. *Marketing*, *7*(3). http://dx.doi.org/10.18775/jibrm.1849-8558.2015.73.3003

Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., & Strohmaier, M. (2024). Ai psychometrics: Assessing the psychological profiles of large language models through psychometric inventories. *Perspectives on Psychological Science*, *19*(5), 808-826. https://doi.org/10.1177/17456916231214460

Swan, K., Speyer, R., Scharitzer, M., Farneti, D., Brown, T., Woisard, V., & Cordier, R. (2023). Measuring what matters in healthcare: a practical guide to psychometric principles and instrument development. *Frontiers in Psychology*, *14*, 1225850. http://dx.doi.org/10.3389/fpsyg.2023.1225850

Unegbu, H. C. O., Yawas, D. S., & Dan-Asabe, B. (2022). An investigation of the relationship between project performance measures and project management practices of construction projects for the construction industry in Nigeria. *Journal of King Saud University-Engineering Sciences*, *34*(4), 240-249. https://doi.org/10.1016/j.jksues.2020.10.001

Zhang, D., Wang, C., Yuan, T., Li, X., Yang, L., Huang, A., Li, J., Liu, M., Lei, Y., Sun, L. and Zhang, J., (2023). Psychometric properties of the Coronavirus Anxiety Scale based on Classical Test Theory (CTT) and Item Response Theory (IRT) models among Chinese front-line healthcare workers. *BMC psychology*, *11*(1), p.224. https://doi.org/10.1186/s40359-023-01251-x

Article Information

Article info:
https://ojs.aeducia.org/index.php/jippg/article/view/259

Word Count: 71-41