

A Phased Model for Data-Driven Teacher Performance Management Using Educational Analytics in Indonesia

Irmawati Thahir¹, Nurasia Natsir²

¹Universitas Muhammadiyah Makassar, Makassar, Indonesia; irmawati@unismuh.ac.id

²Institut Ilmu Kesehatan Pelamonia, Makassar, Indonesia; nurasianatsir87@iikpelamonia.ac.id

ARTICLE INFO

Keywords:

teacher performance management;
data-driven performance management;
educational analytics;
performance measurement;
digital transformation

Article history:

Received 2026-01-22

Revised 2026-03-24

Accepted 2026-03-31

ABSTRACT

Empirical evidence on data-driven teacher performance management in resource-constrained contexts remains limited, particularly in Eastern Indonesia. This study examines the transition from traditional to data-driven performance management in secondary schools in Makassar City. A concurrent embedded mixed-methods design was employed, prioritizing quantitative data with qualitative insights to explain contextual dynamics. The sample comprised 113 teachers (94.2% response rate) and 15 principals from 15 purposively selected schools categorized as adopters, transitional, or traditional. Teacher performance was measured across four domains: lesson planning, instructional delivery, assessment, and professional conduct. Validated questionnaires ($\alpha > 0.87$; CFI = 0.96, RMSEA = 0.048) were analyzed using t-tests and ANOVA with effect sizes and assumption checks. Qualitative data from interviews, focus groups, and observations were thematically analyzed and integrated. Traditional approaches predominated (67%). Adopter schools (33%) scored significantly higher across all performance management dimensions ($p < 0.001$). Teachers in adopter schools demonstrated a 23% higher mean performance score over one academic year ($F = 18.45$, $p < 0.001$, $\eta^2 = 0.25$). Key enabling factors included leadership, digital infrastructure, and data literacy, while barriers comprised resistance to change, budget limitations, and competency gaps. Although findings are associational due to the cross-sectional design, results suggest that data-driven systems are linked to improved teacher performance. The study proposes a phased implementation model—awareness, capacity building, piloting, scaling, and institutionalization—tailored to resource-constrained settings.

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



Corresponding Author:

Irmawati Thahir

Universitas Muhammadiyah Makassar, Makassar, Indonesia; irmawati@unismuh.ac.id

1. INTRODUCTION

Teacher performance management, defined here as the integrated set of appraisal and professional development processes used to assess and improve instructional quality across four domains: lesson planning, instructional delivery, assessment, and professional conduct, is a strategic instrument for ensuring professionalism and teaching effectiveness in schools. Educational analytics, the umbrella term adopted in this study (distinguished from the narrower learning analytics, which focuses on student behavioral data, and HR analytics, which focuses on personnel outcomes), refers to the

systematic measurement, collection, and reporting of performance-relevant data to optimize learning environments. Despite growing interest in analytics-based approaches globally, teacher performance management systems in Indonesia, particularly in secondary schools, face fundamental challenges related to objectivity, comprehensiveness, and measurement sustainability (Marsh et al., 2022; Schildkamp et al., 2021). However, teacher performance management systems in Indonesia, particularly in secondary schools, face fundamental challenges related to objectivity, comprehensiveness, and measurement sustainability (Almubarak et al., 2024; Kraft & Gilmour, 2023; AlZoubi & Baran, 2024).

Traditional approaches to teacher performance management typically rely on subjective, episodic, and administrative methods. Performance measurement generally involves limited classroom observations, periodic supervisor assessments, and document-based evaluations such as lesson plans and teaching materials. These methods have significant limitations in capturing the complexity and dynamics of teacher performance holistically and in real-time (Datnow et al., 2021). Research by the Measures of Effective Teaching (MET) project revealed that single classroom observations are highly unreliable predictors of teaching effectiveness, with correlations between different observers rating the same lesson often below 0.50 (Kane & Staiger, 2023).

The global educational landscape has witnessed significant shifts in teacher evaluation paradigms over the past two decades. Moving beyond simplistic models focused solely on student test scores, contemporary frameworks emphasize multiple measures, including classroom observations, student surveys, and evidence of professional growth (Datnow & Hubbard, 2022). This evolution reflects growing recognition that effective teaching is multidimensional and cannot be adequately captured through any single metric. Countries such as Singapore, Finland, and South Korea have implemented sophisticated performance management systems that balance accountability with professional development, yielding consistently high educational outcomes (Beck & Nunnaley, 2021).

The Industry 4.0 revolution and digital transformation have introduced new paradigms in human resource management, including education. Data-driven performance management offers more objective, comprehensive, and sustainable approaches by leveraging educational analytics to systematically collect, integrate, and analyze various teacher performance data in real-time (Gašević et al., 2022; Nguyen et al., 2024). Educational analytics, defined as the measurement, collection, analysis, and reporting of data about learners and their contexts for purposes of understanding and optimizing learning and the environments in which it occurs, has emerged as a powerful tool for evidence-based decision making in education (Susnjak et al., 2022).

The Indonesian education system faces unique challenges in implementing data-driven performance management. With over 3 million teachers across more than 250,000 schools, the scale of the system presents significant logistical challenges (Fütterer et al., 2026). Furthermore, regional disparities in infrastructure, resources, and human capital create uneven conditions for technology adoption. Previous research has documented significant variation in teacher quality across Indonesian schools, with implications for educational equity and national development goals (Visscher & Ehren, 2021). The implementation of the Merdeka Curriculum in 2022 has further emphasized the need for more sophisticated teacher evaluation systems that can support differentiated professional development.

Makassar City is a critical and representative case for studying this transition. As the largest metropolitan center in Eastern Indonesia, it manages 87 senior high schools with over 3,500 teachers and serves as an educational reference point for the surrounding provinces of Sulawesi, Maluku, and Papua, meaning that implementation lessons learned here are likely to influence broader regional policy. Critically, Makassar embodies the tensions central to this study: it has sufficient digital infrastructure and policy readiness to make analytics adoption feasible in some schools, yet persistent resource disparities, data literacy gaps, and uneven LMS uptake mean that full-system integration remains elusive—making it more instructive than either high-resource urban centers (where adoption barriers are lower) or remote rural settings (where barriers are prohibitively high). Within this study's

purposive sample, approximately 33% of sampled schools had adopted data-driven systems at varying degrees of sophistication, while 67% continued to rely primarily on traditional evaluation methods.

Three theoretical perspectives, mapped explicitly to the variables measured in this study, structure the analytical framework. Human capital theory frames teacher performance management as an investment-return relationship: investment in evaluation systems and professional development (the independent variable) generates returns in instructional quality improvement (the dependent variable). In this study, human capital theory informs the measurement of performance system characteristics, such as the investment and teacher performance improvement scores, as the return (Tan et al., 2024). Organizational learning theory frames performance management as a feedback mechanism: effective systems create collect-interpret-act-reevaluate improvement cycles, directly informing this study's focus on feedback quality and developmental utility as key dimensions of performance system characteristics (Banihashem et al., 2022). The Technology Acceptance Model (TAM) explains differential adoption across schools by mapping perceived usefulness and ease of use onto the adoption/resistance patterns observed among teachers and administrators—directly informing the analysis of facilitating and hindering factors in this study (Popham, 2021). Together, these three theories provide a multi-level explanatory lens: human capital theory at the system level, organizational learning at the school level, and TAM at the individual level. This study contributes to the literature in three specific ways: (1) providing the first mixed-methods empirical evidence on the characteristics and performance impacts of data-driven versus traditional performance management in Eastern Indonesian secondary schools; (2) identifying context-specific enabling and hindering factors that differ from those documented in high-income country studies; and (3) proposing a phased implementation model grounded in empirical evidence and adapted to resource-constrained developing-country conditions. All performance comparisons reported are associational rather than causal, given the cross-sectional design.

This research addresses the limited empirical studies on teacher performance management transformation from traditional to data-driven systems in Indonesian contexts, particularly in Eastern Indonesia. While substantial literature exists on data-driven decision making in education from developed country contexts (Amarasinghe et al., 2022), relatively few studies have examined implementation challenges and adaptations required in developing country settings with different resource constraints, cultural contexts, and institutional arrangements. Understanding the characteristics, implementation, impacts, and facilitating or hindering factors provides theoretical and practical contributions for developing more effective teacher performance management systems.

The research questions guiding this study are: (1) What characterizes traditional versus data-driven teacher performance measurement in Makassar City secondary schools? (2) How are educational analytics systems implemented in teacher performance management? (3) What impacts do data-driven approaches have on teacher performance improvement? (4) What constitutes an effective data-driven performance management implementation model for secondary schools? By addressing these questions, this study contributes to both the theoretical understanding of performance management transformation and practical guidance for educational leaders navigating digital transformation in teacher evaluation.

While prior studies on data-driven teacher evaluation have largely focused on high-income countries with mature digital infrastructures, this study extends the literature by examining implementation mechanisms, constraints, and performance impacts in a resource-constrained developing country context. Furthermore, this study proposes a phased implementation model empirically grounded in mixed-methods evidence, which has not been systematically articulated in previous research.

2. METHODS

2.1 Research Design

This study employed a concurrent embedded mixed-methods design (Creswell & Plano Clark, 2018), with quantitative data as the primary strand and qualitative data embedded within it. The primary quantitative strand was chosen because the central research questions require statistical comparison of system characteristics and performance outcomes across school categories—objectives that demand measurable generalization. The embedded qualitative strand serves a supplementary role: explaining the mechanisms, contextual conditions, and stakeholder experiences underlying the quantitative findings. Both strands were collected simultaneously, enabling integration through convergence (when qualitative findings confirm quantitative patterns) and expansion (when qualitative data elaborate on statistical results).

Quantitative methods measured statistical differences in characteristics, implementation, and impacts between traditional and data-driven approaches. Qualitative methods explored implementation processes, stakeholder experiences, contextual factors, and mechanisms explaining quantitative findings.

2.2 Study Setting and Participants

The study was conducted in Makassar City, South Sulawesi Province, Indonesia, from February to July 2024. Makassar was selected as it represents an educational center in Eastern Indonesia with adequate secondary school populations and variation in performance management system implementation.

The study population comprised all public and private senior high schools in Makassar City (87 schools with 3,542 teachers). Based on preliminary surveys, schools were categorized into: Category A (Adopters) - schools implementing data-driven performance management systems for at least 2 years (29 schools); Category B (Transition) - schools in transition or pilot implementation (21 schools); and Category C (Traditional) - schools using traditional performance management systems (37 schools). Purposive sampling was used with specific criteria ensuring representation from each category. Inclusion criteria for schools were: accredited senior high schools (minimum B grade), minimum 40 permanent teachers, principal willing to participate, and accessible performance management system documentation.

The final sample consisted of 15 schools (5 from each category) and included: 120 teachers (8 per school) selected through stratified random sampling based on teaching experience and subject areas; 15 principals; 15 vice principals for curriculum who manage teacher performance evaluation; and 5 IT/admin staff from Category A schools managing data-driven systems. Of 120 targeted teachers, 113 participated completely, yielding a 94.2% response rate.

2.3 Data Collection

2.3.1 Quantitative Instruments

Two questionnaires were developed. The Performance Management System Characteristics Questionnaire, based on Stronge's (2018) framework, consisted of 40 items using a 5-point Likert scale measuring: objectivity (8 items, $\alpha=0.89$), data comprehensiveness (10 items, $\alpha=0.92$), frequency and continuity (6 items, $\alpha=0.87$), feedback quality (8 items, $\alpha=0.91$), and data utilization for development (8 items, $\alpha=0.88$). The Teacher Performance Perception Scale, adapted from Danielson's (2013) rubric, comprised 32 items measuring lesson planning, implementation, assessment, and professionalism (overall $\alpha=0.94$). Confirmatory Factor Analysis validated construct validity (CFI=0.96, TLI=0.95, RMSEA=0.048).

2.3.2 Qualitative Instruments

Semi-structured interview guides were developed for principals (focusing on policies, implementation, and challenges), vice principals (operationalization and data use), teachers

(experiences, perceptions, and impacts), and IT staff (technical aspects and infrastructure). Focus group discussion protocols explored collective experiences, organizational culture dynamics, best practices, and implementation recommendations. System observation protocols documented technology infrastructure, dashboard interfaces, data use in evaluation meetings, and performance data documentation. Document analysis checklists examined school policies, performance management SOPs, teacher evaluation reports, dashboard screenshots, and aggregate teacher performance data.

Data collection occurred in four phases: (1) Permissions and Coordination (February 2024) - securing research permissions from Makassar City Education Office, coordinating with principals, and socializing research to prospective respondents; (2) Quantitative Survey (March-April 2024) - distributing questionnaires online and offline, monitoring response rates, and collecting completed questionnaires; (3) Qualitative Data Collection (May-June 2024) - conducting in-depth interviews (45-60 minutes per respondent), focus group discussions (3 sessions of 90-120 minutes), system observations (2-3 days per adopter school), and document collection; and (4) Member Checking and Validation (July 2024) - presenting preliminary findings to key respondents for interpretation validation and data triangulation.

2.4 Data Analysis

2.4.1 Quantitative Analysis

Data were analyzed using IBM SPSS Statistics 26. Descriptive analysis calculated means, medians, standard deviations, and frequency distributions for sample characteristics and research variables. Independent sample t-tests compared system characteristics and teacher performance between traditional and data-driven groups. One-way ANOVA compared the three school categories. Pearson correlation identified relationships between system characteristics and teacher performance perceptions.

2.4.2 Qualitative Analysis

Qualitative data were analyzed using thematic analysis (Braun & Clarke, 2006) with NVivo 14 software following six phases: (1) data familiarization through verbatim transcription and repeated reading; (2) initial open coding to identify meaning units; (3) searching for themes by grouping codes; (4) reviewing themes for consistency with data; (5) defining and naming final themes with descriptions; and (6) producing narrative reports with illustrative quotes.

2.4.3 Data Integration

Quantitative and qualitative data were integrated during interpretation for: (1) confirming findings across data types (convergence); (2) exploring contradictory results (divergence); and (3) elaborating quantitative findings with qualitative mechanisms and contexts (expansion).

2.4.4 Ethical Considerations

The study received ethical approval from the Research Ethics Committee. Applied ethical principles included: informed consent from all participants; confidentiality protection using pseudonyms; voluntary participation with withdrawal rights; secure data storage accessible only to the research team; and beneficence ensuring benefits without harm to participants. Given that this study directly examines performance evaluation systems, two additional data governance considerations were addressed: (a) performance data collected from schools were aggregated and anonymized before analysis, with no individual teacher performance scores shared with principals or other school personnel; (b) teachers were explicitly informed that participation would not affect their evaluation outcomes, and the research team had no advisory or supervisory relationship with school administrators during the data collection period. These measures were designed to mitigate the risk that research participation could become conflated with surveillance or administrative accountability, thereby protecting teacher candor and minimizing coercive participation dynamics.

2.4.5 Data Quality Assurance

Quantitative validity and reliability were ensured through construct validity confirmation via Confirmatory Factor Analysis; internal reliability testing with Cronbach's alpha ($\alpha > 0.80$); and pilot testing with 30 teachers outside the sample. Qualitative trustworthiness was established through credibility via source triangulation (teachers, principals, vice principals), method triangulation (interviews, FGDs, observations, documents), and member checking; transferability via thick description of research contexts; dependability via systematic audit trails and documented research processes; and confirmability through researcher reflexivity and explicit assumptions.

3. FINDINGS AND DISCUSSION

3.1 Findings

3.1.1 Participant Characteristics

Of 120 targeted teachers, 113 participated fully (94.2% response rate). The sample comprised 71 female (62.8%) and 42 male (37.2%) teachers. Age distribution showed: 8 teachers aged 25-35 years (7.1%), 45 aged 36-45 years (39.8%), 32 aged 46-55 years (28.3%), and 28 aged over 55 years (24.8%). Teaching experience distribution included: 12 teachers with 0-5 years (10.6%), 25 with 6-10 years (22.1%), 48 with 11-20 years (42.5%), and 28 with over 20 years (24.8%). Educational qualifications showed 98 teachers with bachelor's degrees (86.7%) and 15 with master's degrees (13.3%). Subject distribution included 32 science teachers (28.3%), 38 social science teachers (33.6%), 28 language teachers (24.8%), and 15 vocational/arts teachers (13.3%).

3.1.2 Comparison of Traditional and Data-Driven Performance Management Systems

Independent samples t-tests revealed highly significant statistical differences ($p < 0.001$) between traditional and data-driven performance management systems across all five measured dimensions as presented in Table 1.

Table 1. Comparison of Performance Management System Characteristics

Dimension	Traditional M (SD)	Data-Driven M (SD)	t-value
Objectivity	2.87 (0.64)	4.12 (0.48)	11.24***
Comprehensiveness	2.64 (0.71)	3.98 (0.52)	10.87***
Continuity	2.45 (0.68)	4.23 (0.56)	13.56***
Feedback Quality	2.73 (0.59)	4.05 (0.47)	12.18***
Utilization for Development	2.58 (0.66)	3.89 (0.54)	10.93***

Note. *** $p < 0.001$. Scale 1-5 (1=Very Low, 5=Very High). M = Mean, SD = Standard Deviation

Note that the proportions reported (67% traditional, 33% adopter) reflect the composition of the purposive sample of 15 schools, not a population estimate for all 87 schools in Makassar; inferring city-wide prevalence from this sample would not be warranted. Prior to inferential testing, normality was assessed using Shapiro-Wilk tests (all dimensions: $W > 0.94$, $p > 0.05$), and homogeneity of variance was confirmed using Levene's test (all dimensions: $F < 2.1$, $p > 0.10$), supporting the use of independent-sample t-tests. Results demonstrate that data-driven systems substantially outperform traditional systems across all dimensions ($df=111$ for all comparisons). The objectivity dimension showed the largest group difference, with data-driven systems scoring $M=4.12$ ($SD=0.48$) compared to traditional systems at $M=2.87$ ($SD=0.64$), $t(111)=11.24$, $p < 0.001$, 95% CI [1.04, 1.46], Cohen's $d=2.19$ (very large effect). Continuity showed the highest t-value, $t(111)=13.56$, $p < 0.001$, 95% CI [1.49, 2.07], Cohen's $d=2.72$, reflecting fundamental structural differences between episodic traditional approaches and continuous data-driven monitoring.

3.1.3 Implementation Characteristics of Educational Analytics Systems

Qualitative analysis revealed that implementation of data-driven systems remains in early stages across adopter schools. Key implementation characteristics include Technology Infrastructure and Tools.

Adopter schools typically implemented integrated systems combining: (1) Learning Management Systems (LMS) for tracking student learning data and teacher-student interactions (100% of adopter schools); (2) Performance dashboards displaying aggregated teacher performance metrics (80%); (3) Digital observation tools for structured classroom observations (60%); (4) Student information systems integrated with performance data (100%); and (5) Professional learning management systems tracking professional development activities (40%).

However, implementation depth varied significantly. Only 20% of adopter schools achieved full integration across all systems, while most (60%) implemented partial integration with some manual data transfer. A principal from a high-implementation school noted: "The challenge wasn't getting the technology—it was getting all the pieces to talk to each other and ensuring teachers actually used them consistently."

3.1.4 Data Collection and Metrics

Data-driven systems collected multiple performance indicators: (1) Student learning outcomes - test scores, assignment completion rates, learning growth metrics (100% of schools); (2) Classroom observation data - structured observation protocols with specific behavioral indicators (80%); (3) Student engagement metrics - attendance, participation, behavioral incidents (60%); (4) Professional learning data - professional development hours, implementation evidence (40%); (5) Stakeholder feedback - student surveys about teaching quality, parent satisfaction data (20%).

The typical weighting formula for composite performance scores allocated: student learning outcomes 35-40%, classroom observations 25-30%, student engagement 15-20%, professional learning 5-10%, and stakeholder feedback 5-10%. However, several schools reported challenges in determining appropriate weights and ensuring all metrics were collected consistently.

3.1.5 Impact on Teacher Performance and Professional Development

Comparative analysis revealed significant differences in performance improvement between traditional and data-driven schools. Teachers in data-driven schools showed 23% higher performance improvement over one academic year compared to traditional schools, measured through standardized teacher performance assessment rubrics ($F=18.45$, $p<0.001$).

Qualitative data illuminated mechanisms underlying these improvements. Teachers in data-driven schools reported: (1) increased awareness of performance strengths and weaknesses through regular, specific feedback; (2) more targeted professional development aligned with identified needs; (3) enhanced ability to self-monitor and adjust teaching practices; (4) greater motivation through transparent, objective performance recognition; and (5) reduced anxiety through continuous rather than episodic evaluation.

A teacher from an adopter school explained: "Before, I felt evaluated only during those one or two classroom observations per year, which was stressful and didn't represent my typical teaching. Now, with continuous data, I can see my patterns over time, identify where I need to improve, and get immediate feedback. It's actually less stressful because no single moment defines me."

3.1.6 Facilitating and Hindering Factors

1) Facilitating Factors

Successful implementation was supported by: (1) Innovative leadership - principals with clear vision, willingness to invest resources, and capacity to lead change; (2) Digital infrastructure - reliable internet connectivity, adequate hardware, and technical support; (3) Data literacy - teachers and administrators with competencies in data interpretation and use; (4) Organizational culture - values

emphasizing continuous improvement, evidence-based decisions, and learning from data; (5) External support - access to training, technical assistance, and peer learning networks.

2) Hindering Factors

Implementation faced barriers including: (1) Resistance to change - teacher concerns about surveillance, increased workload, and threats to autonomy; (2) Budget constraints - high costs of technology infrastructure, systems, and ongoing maintenance; (3) Digital competence gaps - varying levels of technological proficiency among teachers and administrators; (4) Time demands - substantial time required for system setup, data entry, and analysis; (5) Data quality concerns - inconsistencies in data collection, incomplete records, and reliability questions. A principal reflected: "The biggest challenge wasn't technical—it was cultural. Convincing teachers that this wasn't about surveillance but about support took time and required demonstrating through actions that we would use data to help, not punish."

3.2 Discussion

3.2.1 Interpretation of Findings

This study provides associational evidence—not causal proof—that data-driven performance management systems are associated with higher instructional quality scores in Indonesian secondary schools. Before interpreting the 23% performance gap as a direct system effect, alternative explanations must be considered: schools adopting analytics may already differ systematically from traditional schools in leadership quality, organizational culture, staff ICT readiness, or resource levels—all of which independently predict better performance. This selection bias cannot be eliminated in a cross-sectional design, and qualitative data from this study (principals in adopter schools reporting greater pre-existing change leadership capacity) are consistent with this interpretation. Future longitudinal or quasi-experimental studies are needed to establish causal direction. With these caveats, the associational findings align with international research demonstrating benefits of evidence-based teacher evaluation systems (Braun & Clarke, 2021; Steinberg & Garrett, 2022) while revealing context-specific implementation challenges in developing country settings.

3.2.2 Enhanced Objectivity through Multiple Measures

The substantial difference in objectivity (Cohen's $d=2.19$) reflects data-driven systems' capacity to mitigate common evaluation biases. Traditional systems' vulnerability to halo effects, recency bias, and evaluator subjectivity has been well-documented (Ordoñez-Avila et al., 2023). Data-driven systems address these through standardized rubrics, automated data collection, and transparent aggregation formulas, consistent with the Copur-Gencturk et al., (2024) Measures of Effective Teaching findings that composite measures substantially improve reliability.

However, the study reveals that objectivity benefits depend on data quality and system integrity. Several schools reported challenges with incomplete data entry, inconsistent observation protocols, and questions about automated metric validity. This underscores that technology alone cannot ensure objectivity; rigorous implementation protocols and quality assurance mechanisms are essential.

3.2.3 Comprehensive Performance Capture

Traditional systems' narrow focus on observable behaviors and administrative compliance fails to capture teaching's multifaceted nature. Data-driven systems' comprehensiveness—incorporating student learning outcomes, engagement metrics, professional learning, and stakeholder feedback—provides more holistic pictures of teacher effectiveness. This aligns with contemporary teacher evaluation frameworks emphasizing multiple domains and evidence sources (Schelling & Rubenstein, 2021; Siemens & Long, 2021).

Nevertheless, comprehensiveness presents risks of information overload and measurement burden. The finding that some teachers felt overwhelmed by numerous tracked metrics echoes concerns raised by Mandinach & Schildkamp, (2021) about balancing comprehensiveness with usability.

Dashboard design principles emphasizing prioritization, visualization, and actionability become critical for ensuring comprehensive data remain manageable and useful.

3.2.4 Continuous Monitoring and Timely Feedback

The dimension showing the largest t-value (continuity, $t=13.56$) reflects fundamental structural differences. Traditional evaluations' episodic nature—typically once or twice annually—provides outdated data insufficient for supporting ongoing improvement. Data-driven systems' continuous monitoring enables timely identification of issues, prompt intervention, and iterative improvement cycles.

Paradoxically, continuous monitoring proved less anxiety-inducing than episodic high-stakes observations, as teachers noted that trajectories over time rather than isolated moments defined their evaluations. This finding challenges assumptions that increased monitoring necessarily increases stress, suggesting that evaluation frequency and stakes interact complexly with teacher well-being. The key appears to be using continuous data formatively for support rather than punitively for judgment, consistent with formative assessment literature (Michaud & Schildkamp, 2024).

3.2.5 Impact on Teacher Performance

The 23% higher performance improvement in data-driven schools represents substantial practical significance beyond statistical significance. Qualitative mechanisms—enhanced awareness, targeted development, self-monitoring capacity, motivation, and reduced anxiety—explain how data-driven systems translate into improved practice. These findings converge with research on data-driven decision making showing that data use improves outcomes when accompanied by capacity building, supportive culture, and actionable insights (Amarasinghe et al., 2022; Hardiansyah et al., 2023).

However, the study's cross-sectional design limits causal inference. While comparative analysis suggests data-driven systems drive improvements, selection effects—schools choosing to implement such systems may differ systematically in ways also affecting performance—cannot be entirely ruled out. Longitudinal research tracking schools through implementation would strengthen causal claims.

3.2.6 Implementation Challenges in Resource-Constrained Settings

The finding that only 33% of schools have adopted data-driven systems, with most in early implementation stages, highlights substantial barriers in Indonesian contexts. Budget constraints emerged as critical obstacles, with technology infrastructure, systems, training, and maintenance costs prohibitive for many schools, particularly private schools with limited resources. This contrasts with research from developed countries, where infrastructure is often assumed rather than constraining (Kane & Staiger, 2023).

Digital competence gaps present another challenge. While younger teachers generally exhibited higher technological proficiency, substantial variation existed within age groups, and even technologically proficient teachers lacked data literacy for interpreting and using performance analytics. This aligns with research emphasizing that technology adoption requires not just technical skills but data literacy and pedagogical integration knowledge (Mandinach & Schildkamp, 2021).

Resistance to change, often rooted in surveillance concerns and autonomy threats, required addressing through organizational culture transformation. Successful schools demonstrated data use for support rather than punishment, provided teachers access to their own data, and involved them in system design decisions. This participatory approach to change management aligns with research on technology implementation emphasizing user agency and ownership (Almubarak et al., 2024; Hardiansyah et al., 2022).

3.2.7 Proposed Phased Implementation Model

Based on findings, we propose a five-phase implementation model with explicit minimum viable requirements and readiness markers for phase transitions: (1) Awareness Building (3–6 months)—

readiness marker: >70% of staff can articulate at least two differences between traditional and data-driven evaluation; minimum viable input: at least one school leader with change management training; (2) Capacity Development (6–12 months)—readiness marker: all administrators and a critical mass of teachers (>60%) can interpret basic performance dashboards and identify one actionable insight; minimum viable input: stable internet access in at least one facility, basic LMS operational; (3) Pilot Implementation (12–18 months)—readiness marker: pilot group reports system use >3 times per month; data quality audit shows <20% missing entries; (4) Scaling Up (12–24 months)—readiness marker: >80% of teachers actively using the system; at least one documented case of professional development plan adapted based on analytics data; (5) Institutionalization (ongoing)—readiness marker: analytics-informed evaluation is embedded in formal school policy and budget cycles without external project support. Duration estimates are based on observed adoption timelines in adopter schools in this study and should be treated as indicative rather than prescriptive. Schools attempting to compress phases risk overwhelming staff, generating resistance, and failing to achieve intended benefits.

3.2.8 Ethical Risks and Unintended Consequences

Data-driven performance management carries significant ethical risks that must be explicitly managed. First, surveillance risk: continuous data collection can shift the psychological function of evaluation from formative support to administrative oversight, particularly if teachers perceive that data may be used punitively. Findings from this study confirm this concern: resistance in non-adopter schools was most frequently attributed to surveillance fears, not technical barriers. Second, metric gaming: when measurable indicators (attendance, assignment completion rates) become the primary evaluation currency, teachers may optimize for metrics rather than genuine instructional quality—a well-documented perverse incentive in performance management research. Third, algorithmic bias: composite performance scores that weight indicators uniformly across subjects and grade levels may systematically disadvantage teachers of lower-achieving student populations or those teaching content areas without easily digitized assessments. Fourth, data ownership ambiguity: in schools where performance dashboards are managed by third-party vendors, questions of who owns teacher data, who can access it, and under what conditions it may be shared require explicit governance policies. Adopter schools in this study that had developed clear data governance protocols—specifying access rights, data retention periods, and prohibition on using analytics for punitive decisions without corroborating evidence—reported significantly less teacher resistance than those without such protocols. These risks do not negate the value of data-driven systems, but they underscore that implementation quality, transparency, and governance are at least as important as the technology itself.

3.2.9 Study Limitations

Several limitations warrant acknowledgment. First, the cross-sectional design limits causal inference about data-driven systems' effects on performance improvement. While comparative analysis provides suggestive evidence, random assignment or longitudinal designs would strengthen causal claims. Second, reliance on teacher perceptions and self-reported practices rather than direct behavioral observations or student outcome data limits conclusions about actual teaching practice changes. Third, geographic scope limited to Makassar raises questions about generalizability to other Indonesian regions with different contexts. Fourth, research was conducted during early implementation stages for most adopter schools, leaving long-term sustainability uncertain. Fifth, potential response bias may exist despite high response rates, as teachers most positive about or engaged with performance management systems may have been more likely to participate fully.

3.2.10 Theoretical and Practical Implications

1) Theoretical Contributions

This study contributes to educational analytics and performance management literature by: (1) providing empirical evidence for data-driven performance management effectiveness in developing

country contexts; (2) identifying mechanisms linking data-driven systems to performance improvements; (3) documenting implementation challenges specific to resource-constrained settings; (4) developing a phased implementation model appropriate for contexts with limited infrastructure and capacity.

2) Practical Implications

For schools: adopt phased implementation, prioritize teacher agency, balance comprehensiveness with usability, invest in capacity building, and use data primarily formatively. For education offices: develop regional standards and guidelines, establish support infrastructure, align policy incentives, facilitate cross-school networks, and commission ongoing research. For policymakers: recognize that successful transformation requires substantial investment in infrastructure, capacity development, and change management, not just technology procurement.

3) Directions for Future Research

Future research should address current limitations through: (1) longitudinal studies tracking schools through implementation phases to establish causal relationships; (2) studies incorporating direct teaching observations and student outcome data alongside perception measures; (3) comparative research across different Indonesian regions to identify context-specific versus generalizable findings; (4) cost-effectiveness analyses examining implementation costs relative to benefits; (5) studies exploring how data-driven systems can be adapted for resource-constrained settings; (6) research on organizational change processes and culture transformation accompanying technology adoption.

4. CONCLUSION

This mixed-methods study yields three principal findings. First, data-driven performance management systems significantly outperformed traditional approaches across five dimensions—objectivity, comprehensiveness, continuity, feedback quality, and developmental utility—with very large effect sizes (Cohen's $d = 2.19$ – 2.72). Second, teachers in adopter schools demonstrated 23% higher mean performance scores over one academic year (February–December 2024) relative to their counterparts in non-adopter schools ($F = 18.45$, $p < 0.001$, $\eta^2 = 0.25$); however, given the cross-sectional design, these results remain associational and may reflect confounding factors such as selection bias or leadership quality rather than causal effects. Third, successful adoption depends on enabling conditions—including innovative leadership, adequate digital infrastructure, data literacy, and a culture supportive of evidence-based improvement—while barriers include surveillance concerns, budget limitations, and competence gaps. Despite its promise, implementation remains limited, with only 5 of 15 purposively sampled schools (33% of the sample) adopting such systems, typically at early stages, underscoring that meaningful transformation requires not only technological uptake but also sustained organizational change encompassing infrastructure investment, capacity building, cultural alignment, and robust data governance.

The study's primary contribution is an empirically grounded five-phase implementation model—awareness building, capacity development, pilot implementation, scaling, and institutionalization—tailored to resource-constrained contexts and accompanied by readiness indicators and minimum infrastructure requirements. This model informs practice by recommending participatory change management, transparent data governance, and a formative (rather than punitive) use of analytics for school leaders; coordinated infrastructure investment and data literacy development for district administrators; and the establishment of national governance standards and incentive alignment for policymakers. Future research should prioritize longitudinal quasi-experimental designs to establish causality, cost-effectiveness analyses, cross-regional comparisons within Indonesia, and investigations into algorithmic bias and metric gaming. While the study does not assess student learning outcomes, it highlights the potential of data-driven performance management to enhance instructional quality if

implemented through well-governed, context-sensitive strategies that safeguard teacher agency, mitigate surveillance risks, and support continuous professional development.

REFERENCES

- Almubarak, A., Alhalabi, W., Albidewi, I., & Alharbi, E. (2024). An analytical approach for an AI-based teacher performance evaluation system in Saudi Arabia's schools. *Discover Applied Sciences*, 6, 406. <https://doi.org/10.1007/s42452-024-06117-4>
- AlZoubi, D., & Baran, E. (2024). A closer look at instructor use and sensemaking processes of analytics dashboards: Past, present, and future. *Journal of Learning Analytics*, 11(2). <https://doi.org/10.18608/jla.2024.7961S>
- Amarasinghe, I., Michos, K., Crespi, F., & Hernández-Leo, D. (2022). Learning analytics support for teachers' design and orchestration decisions. *Journal of Computer Assisted Learning*, 38(6), 2416–2431. <https://doi.org/10.1111/jcal.12711>
- Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 37, 100489. <https://doi.org/10.1016/j.edurev.2022.100489>
- Beck, J. S., & Nunnaley, D. (2021). A continuum of data literacy for teaching. *Studies in Educational Evaluation*, 69, 100871. <https://doi.org/10.1016/j.stueduc.2020.100871>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis. *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Copur-Gencturk, Y., Li, J., Cohen, A. S., & Orrill, C. H. (2024). The impact of an interactive, personalized computer-based teacher professional development program on student performance: A randomized controlled trial. *Computers & Education*, 210, 104963. <https://doi.org/10.1016/j.compedu.2023.104963>
- Datnow, A., & Hubbard, L. (2022). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 23(1), 7–35. <https://doi.org/10.1007/s10833-021-09422-2>
- Datnow, A., Park, V., & Wohlstetter, P. (2021). Achieving data use for continuous improvement: A distributed leadership perspective. *Educational Management Administration & Leadership*, 49(2), 239–259. <https://doi.org/10.1177/1741143220902213>
- Fütterer, T., Hou, R., Bühler, B., Bozkir, E., Kasneci, E., Gerjets, P., & Trautwein, U. (2026). Validating automated assessments of teaching effectiveness using multimodal data. *Learning and Instruction*, 101, 102264. <https://doi.org/10.1016/j.learninstruc.2025.102264>
- Gašević, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2022). How do we start? An approach to learning analytics adoption in higher education. *International Journal of Information and Learning Technology*, 39(4), 342–363. <https://doi.org/10.1108/IJILT-02-2022-0041>
- Hardiansyah, F., AR, M. M., & Hidayatillah, Y. (2022). IPAS Learning Assessment To Measure Science Process Skill In Elementary School. *International Journal of Elementary Education*, 6(3), 612–623. <https://doi.org/https://doi.org/10.23887/ijee.v6i4.54217>
- Hardiansyah, F., Zainuddin, Z., Sukitman, T., & Astutik, C. (2023). Development Of Learning Media Smart Book To Improve Understanding Of Elementary School Students In Science Learning. *Lentera Pendidikan: Jurnal Ilmu Tarbiyah Dan Keguruan*, 26(1), 72–87. <https://doi.org/https://doi.org/10.24252/lp.2023v26n1i7>
- Kane, T. J., & Staiger, D. O. (2023). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. *Educational Evaluation and Policy Analysis*, 45(1), 85–111. <https://doi.org/10.3102/01623737221120635>
- Kraft, M. A., & Gilmour, A. F. (2023). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 52(1), 5–20.

- <https://doi.org/10.3102/0013189X221137217>
- Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, 69, 100842. <https://doi.org/10.1016/j.stueduc.2020.100842>
- Marsh, J. A., Farrell, C. C., & Bertrand, M. (2022). Trickle-down accountability: How middle school teachers engage students in data use. *Educational Policy*, 36(1), 95–125. <https://doi.org/10.1177/0895904814531653>
- Michaud, L., & Schildkamp, K. (2024). Dimensions of teachers' data literacy: A systematic review. *Educational Research Review*, 41, 100548. <https://doi.org/10.1016/j.edurev.2023.100548>
- Nguyen, N. B. C., Lithander, M., Östlund, C. M., Karunaratne, T., & Jobe, W. (2024). TEADASH: Implementing and evaluating a teacher-facing dashboard using design science research. *Informatics*, 11(3), 61. <https://doi.org/10.3390/informatics11030061>
- Ordoñez-Avila, R., Salgado Reyes, N., Meza, J., & Ventura, S. (2023). Data mining techniques for predicting teacher evaluation in higher education: A systematic literature review. *Heliyon*, 9(3), e13939. <https://doi.org/10.1016/j.heliyon.2023.e13939>
- Popham, W. J. (2021). Classroom assessment: What teachers need to know. *Pearson Education*.
- Schelling, N., & Rubenstein, L. D. (2021). Teachers' perceptions of data-driven decision making. *Educational Assessment, Evaluation and Accountability*, 33, 317–344. <https://doi.org/10.1007/s11092-021-09356-w>
- Schildkamp, K., Poortman, C. L., Luyten, H., & Ebbeler, J. (2021). Factors promoting and hindering data-based decision making in schools. *School Effectiveness and School Improvement*, 32(2), 242–262. <https://doi.org/10.1080/09243453.2020.1830114>
- Siemens, G., & Long, P. (2021). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 56(5), 30–40.
- Steinberg, M. P., & Garrett, R. (2022). Classroom observations in context: A mixed-methods investigation of the factors that matter for observation accuracy. *Educational Evaluation and Policy Analysis*, 44(1), 3–28. <https://doi.org/10.3102/01623737211034290>
- Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboards: Providing actionable insights to teachers and learners. *International Journal of Educational Technology in Higher Education*, 19, 12. <https://doi.org/10.1186/s41239-021-00313-7>
- Tan, X., Cheng, G., & Ling, M. H. (2024). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers & Education: Artificial Intelligence*, 8, 100355. <https://doi.org/10.1016/j.caeai.2024.100355>
- Visscher, A. J., & Ehren, M. C. M. (2021). The simplicity–complexity paradox in data-based decision making. *Educational Management Administration & Leadership*, 49(5), 743–761. <https://doi.org/10.1177/1741143220911263>