Uniglobal of Journal Social Sciences and Humanities

Performance Evaluation Systems for School Leaders in China: Policies and Practices

Zhang, Kai[1] & Muhammad, Mazni[2*]

[1,2] Faculty of Education, Universiti Islam Melaka, 78200 Kuala Sungai Baru, Malacca, Malaysia

*Corresponding author: dr.mazni@unimel.edu.my

Abstract: This study examines China's performance evaluation systems for school leaders, analyzing their effectiveness, implementation challenges, and potential areas for reform. Through quantitative analysis of survey data from 500 school leaders across urban and suburban regions, the research identifies a strong emphasis on standardized test scores (mean=4.6) and policy compliance (mean=4.4) as dominant evaluation criteria, while revealing comparatively lower prioritization of teacher professional development (mean=3.8) and pedagogical innovation (mean=3.2). The findings highlight systemic rigidity in current evaluation frameworks, with modest but consistent urban-suburban variations suggesting limited adaptation to local contexts. The study critically evaluates how this over-reliance on quantifiable metrics may inadvertently constrain educational innovation and leadership autonomy. It proposes concrete recommendations for policy implementation, including the development of more balanced assessment criteria, enhanced stakeholder participation in evaluation processes, and regionally differentiated approaches to account for resource disparities. The paper concludes by outlining key directions for future research, emphasizing the need for longitudinal studies on evaluation impacts, qualitative investigations of leadership experiences, and cross-national comparisons with high-performing education systems. By bridging empirical findings with practical policy suggestions, this study contributes to ongoing discussions about optimizing school leadership evaluation to support educational quality and equity in China's rapidly evolving education landscape.

Keywords: School leadership evaluation, Educational policy in China, Performance assessment, Standardized testing, Educational innovation

1. Introduction

The bedrock of any thriving education system lies in the efficacy of its school leadership. Principals, headteachers, and other school administrators are widely acknowledged as pivotal figures in shaping school culture, driving pedagogical innovation, fostering teacher development, and ultimately, improving student outcomes (Ahmad, 2021). Their influence extends beyond mere administrative duties, encompassing strategic vision-setting, instructional leadership, community engagement, and resource management, all of which are critical for navigating the complexities of modern educational environments (Alzoraiki et al., 2023). Recognizing this profound impact, governments and educational authorities worldwide have increasingly focused on mechanisms to enhance leader effectiveness, accountability, and continuous professional growth. Among these mechanisms, performance evaluation systems have emerged as a cornerstone, designed to systematically assess leaders' contributions, identify areas for improvement, and ensure alignment with broader educational goals (Chatzipanagiotou & Katsarou, 2023).

The concept of performance evaluation, rooted in human resource management, has evolved significantly over time, transitioning from rudimentary assessment tools to sophisticated frameworks aimed at fostering development rather than merely imposing sanctions. Globally, there is a discernable shift from traditional top-down, input-focused evaluations to more comprehensive, multi-source, and outcome-oriented approaches that often incorporate elements of self-reflection, peer feedback, and 360-degree assessments (Alzoraiki et al., 2023). The primary rationale behind these systems is multi-faceted. Firstly, they serve an accountability function, ensuring that school leaders are answerable for their performance against established professional standards and institutional objectives. Secondly, and increasingly emphasized, is the developmental purpose, where evaluations provide constructive feedback, identify professional learning needs, and guide tailored professional development initiatives (Chu et al., 2021). Thirdly, evaluation systems can facilitate strategic

alignment, linking individual leader performance to school-wide improvement plans and national educational priorities, thereby ensuring a cohesive effort towards common goals (Haiyan & Allan, 2020).

Despite the clear theoretical benefits and widespread adoption, the design and implementation of effective performance evaluation systems for school leaders are fraught with challenges. Issues such as the subjectivity inherent in assessing leadership qualities, the difficulty in isolating a leader's impact on student learning, the potential for bias, and the administrative burden of comprehensive systems often plague their efficacy (Li & Liu, 2020). Moreover, for evaluation to be truly developmental, it requires clear criteria, transparent processes, skilled evaluators, and a culture of trust and continuous improvement rather than one dominated by fear or compliance (Alzoraiki et al., 2023). These challenges highlight the necessity for context-specific research to understand how evaluation systems operate in diverse educational landscapes, including those undergoing rapid transformation.

Within this global discourse, China presents a particularly compelling and complex case study. As the world's most populous nation, China operates the largest education system, catering to hundreds of millions of students across a vast and diverse geographical expanse. Education is not merely a social service in China; it is a strategic national priority, viewed as indispensable for economic development, social harmony, and technological advancement (Liu et al., 2023). Over the past four decades, China has undergone monumental educational reforms, characterized by rapid expansion, quality enhancement initiatives, and a gradual shift towards more diversified and localized governance structures, albeit still under the overarching guidance of the central government (Sun et al., 2022). In this colossal system, school leaders serve as crucial intermediaries, responsible for implementing national and provincial policies, managing school operations, leading curriculum reforms, and nurturing a conducive learning environment for both teachers and students (Wang & Wang, 2020). They are the direct link between policy rhetoric and practical reality, making their effectiveness paramount to the success of China's ambitious educational agenda.

The role of school leaders in China has evolved considerably from primarily administrative functions to encompassing more complex pedagogical, managerial, and even entrepreneurial responsibilities (Chu et al., 2021). Faced with increasing demands for quality education, equitable access, and innovative pedagogical practices, Chinese school leaders are expected to be transformational figures who can navigate intricate bureaucratic structures while fostering localized initiatives (Gao et al., 2022). Historically, the evaluation of school leaders in China might have been less formalized, relying more on personal relationships, political loyalty, and informal assessments by superiors(Yan et al., 2021). However, with the deepening of educational reforms and the national push for greater accountability and professionalization across public sectors, the need for more systematic and objective methods of evaluating school leaders has become increasingly apparent.

In response to these demands, the Chinese government has progressively introduced and refined policies pertaining to the performance evaluation of school leaders. These policies often align with broader national objectives such as "Educational Modernization 2035," which emphasizes high-quality education and professional development for educators (Yan et al., 2021). The stated goals of these evaluation systems are typically multifaceted: to promote professional growth, ensure policy compliance, enhance school performance, and maintain a meritocratic promotion system (Chu et al., 2021). Such policies usually outline criteria related to leadership effectiveness, teaching quality, student development, school management, and community engagement. However, due to China's vastness and decentralized implementation at provincial and municipal levels, the specific criteria, procedures, and the emphasis of these evaluations can vary significantly across different regions and even within different types of schools (Haiyan & Allan, 2020).

Despite the clear policy directives and the critical importance of school leaders in China's education system, there remains a notable gap in comprehensive research specifically focusing on the policies and practices of performance evaluation systems for school leaders in China. While some studies may touch upon aspects of school leadership or educational policy, in-depth empirical investigations that bridge the stated policy intentions with the actual lived experiences and perceptions of school leaders regarding these evaluation systems are scarce (Alzoraiki et al., 2023). Existing literature often provides a broad overview of Chinese education or focuses narrowly on specific reform initiatives, leaving a lacuna concerning the practical implementation and perceived effectiveness of leader evaluation mechanisms.

Filling this research gap is of paramount significance. Understanding the interplay between policy formulation and practical implementation provides invaluable insights for policymakers seeking to refine existing systems, ensuring they are fair, transparent, and genuinely developmental. For school leaders themselves, such research can highlight best practices and areas for improvement in their professional growth journey (Maheshwari, 2021). Furthermore, given China's unique socio-political context, this study can contribute to the broader comparative education discourse by offering a non-Western perspective on performance management in the public sector, potentially informing other developing nations grappling with similar challenges in enhancing educational leadership (Ahmad, 2021). This study, therefore, aims to systematically analyze both the formal policies and the lived realities of performance evaluation for school leaders in China, providing a nuanced understanding of its complexities and implications.

1.1 Research Gap and Significance

The pivotal role of effective school leadership in driving educational quality and achieving systemic reforms is widely recognized across diverse educational contexts globally (Chatzipanagiotou & Katsarou, 2023). Concurrently, performance evaluation systems have gained prominence as critical tools for accountability, professional development, and strategic alignment in educational leadership (Chiu, 2023). In China, a nation undergoing profound educational transformations aimed at fostering a high-quality, modernized education system, the imperative to professionalize and enhance the performance of school leaders is a central policy focus (Li & Liu, 2020). Existing scholarship provides valuable insights into general aspects of Chinese education policy (Liu et al., 2023), the evolving roles of school leaders within this context (Haiyan & Allan, 2020), and the broader trends in performance management within public administration (Maheshwari, 2021). However, despite these contributions, a crucial research gap persists concerning the specific dynamics of performance evaluation systems for school leaders in China.

Specifically, while official documents and policy statements outline the frameworks and intentions for evaluating school leaders, there is a notable paucity of comprehensive empirical research that critically examines the interplay between these stated policies and their actual implementation practices at the local level (Liu et al., 2023). Many studies tend to describe policy prescriptions without delving deeply into the nuanced realities of how these policies are enacted, interpreted, and experienced in diverse school settings across China's vast and varied educational landscape (Ahmad, 2021). Such empirical insights into the practicalities and subjective experiences are essential for a holistic understanding but remain largely underexplored in the current literature. Furthermore, the variation in implementation across different regions or levels of governance within China, which influences the effectiveness and equity of these systems, has not been sufficiently documented or analyzed (Gao et al., 2022).

Addressing this distinct research gap carries substantial significance for multiple stakeholders. Theoretically, this study will enrich the discourse on educational leadership and performance management, offering a unique perspective from a large, rapidly developing, and centrally-governed yet regionally diverse non-Western context. It can contribute to refining existing theoretical models of evaluation by incorporating the specific challenges and opportunities inherent in the Chinese educational bureaucracy and cultural environment (Chiu, 2023). Practically, the findings will provide crucial, evidence-based insights for Chinese policymakers at national, provincial, and municipal levels. By highlighting convergences and divergences between policy intentions and implementation realities, the study can inform the refinement of existing evaluation systems, making them more effective, equitable, and genuinely developmental for school leaders (Maheshwari, 2021). For school leaders themselves, the research will shed light on common experiences, challenges, and effective strategies for engaging with evaluation processes, potentially fostering a more supportive professional environment. Lastly, from a comparative education standpoint, China's experiences in structuring and implementing leader evaluation systems can offer valuable lessons and cautionary tales for other countries, particularly those in similar stages of educational development or those navigating complex centralized-decentralized governance dynamics in their own pursuit of educational excellence.

1.2 Research Objectives
This study has two primary research objectives:
1) To systematically analyze and document the current policies and regulations governing the performance evaluation of school leaders in China at various administrative levels.
2) To investigate and describe the actual practices and experiences of implementing these performance evaluation systems among school leaders in selected regions of China.

1.3 Research Questions
This study has two primary research questions:
1) What are the key characteristics, criteria, and procedures stipulated in the current national and regional policies for the performance evaluation of school leaders in China?
2) How are performance evaluation policies for school leaders implemented in practice, and what are the perceptions and experiences of school leaders regarding these evaluation systems in selected regions of China?

2. Literature Review
Performance evaluation systems for school leaders have gained increasing attention in global educational governance, particularly as nations strive to enhance school effectiveness and leadership accountability (Wang et al., 2022). In China, the evaluation of school principals and other educational leaders is deeply embedded in the broader framework of national education reforms, which emphasize quality improvement, equity, and administrative efficiency. The Chinese government has implemented various policies to refine leadership evaluation mechanisms, aligning them with socialist educational values and modernization goals (Yan & Brown, 2021). This literature review examines the theoretical foundations, policy evolution, and practical implementations of performance evaluation systems for school leaders in China, while also exploring comparative international perspectives and emerging challenges.

2.1 Theoretical Foundations of Leadership Evaluation

The evaluation of school leaders is rooted in multiple theoretical frameworks, including transformational leadership, instructional leadership, and distributed leadership. Transformational leadership theory (Sun et al., 2022) emphasizes visionary leadership, motivation, and organizational change, which are highly relevant in the Chinese context, where principals are expected to drive school improvement under national policy directives. Instructional leadership (Chu et al., 2021) focuses on the principal's role in shaping teaching and learning outcomes, a key aspect of China's performance evaluation criteria. Additionally, distributed leadership (Haiyan & Allan, 2020) highlights shared leadership responsibilities, which aligns with China's collective approach to school management. In China, leadership evaluation also incorporates socialist political ideologies, where school leaders are assessed not only on educational outcomes but also on their adherence to national policies, moral integrity, and contributions to social stability (Li & Liu, 2020). This dual emphasis on professional competence and ideological conformity distinguishes China's approach from Western models, which tend to prioritize measurable academic outcomes over political alignment.

2.2 Policy Evolution of School Leader Evaluation in China

China's performance evaluation systems for school leaders have undergone significant transformations alongside broader educational reforms. Historically, evaluations were largely administrative and based on compliance with government directives rather than leadership effectiveness (Gao et al., 2022). However, since the early 2000s, reforms have introduced more structured and multi-dimensional evaluation frameworks. The *National Medium- and Long-Term Educational Reform and Development Plan (2010-2020) marked a pivotal shift by emphasizing principal professionalism and school autonomy (Sun et al., 2022). Subsequent policies, such as the Guidelines on Deepening the Reform of Educational Evaluation in the (Liu et al., 2023) further refined evaluation criteria to include student development, teacher satisfaction, and community engagement (Liu et al., 2023). These reforms reflect a gradual move from bureaucratic assessments to more holistic, evidence-based evaluations. Provincial and municipal governments have also developed localized evaluation systems. For instance, Shanghai's Comprehensive Evaluation System for Principals incorporates student academic performance, teacher professional development, and school innovation metrics (Maheshwari, 2021). Such regional adaptations demonstrate China's decentralized yet policy-aligned approach to leadership evaluation.

2.3 Current Practices in Evaluating School Leaders

Contemporary evaluation practices in China typically involve multiple stakeholders, including government education bureaus, school staff, students, and parents. The process often combines quantitative metrics and qualitative assessments. A notable feature is the *360-degree feedback* mechanism, where principals receive input from superiors, subordinates, and community members (Li & Liu, 2020). However, challenges persist in balancing accountability and autonomy. While standardized metrics ensure transparency, critics argue that excessive emphasis on exam scores may discourage innovative leadership (Haiyan & Allan, 2020). Additionally, the heavy administrative burden of evaluation processes can divert leaders' attention from pedagogical improvement (Gao et al., 2022).

2.4 International Comparisons and Lessons for China

Comparative studies reveal that China's leadership evaluation system shares similarities with those in high-performing education systems like Singapore and Finland, where leadership quality is closely tied to school success (Alzoraiki et al., 2023). However, unlike China's centralized model, Finland employs a trust-based system with minimal standardized testing, relying instead on professional discretion. Singapore's Enhanced Performance Management System (EPMS) integrates career development with evaluation, offering insights into how China might link assessment with leadership growth (Chu et al., 2021). These comparisons suggest that China could benefit from further decentralizing evaluations, reducing bureaucratic rigidity, and fostering a culture of trust and professional development.

2.5 Emerging Challenges and Future Directions

Despite progress, China's evaluation systems face challenges, including regional disparities, resistance to change, and the need for more dynamic assessment tools. Future reforms may incorporate artificial intelligence and big data analytics to enhance objectivity (Chiu, 2023). Additionally, greater emphasis on ethical leadership and social-emotional learning outcomes could align evaluations with global education trends. China's performance evaluation systems for school leaders reflect a complex interplay of policy mandates, theoretical frameworks, and practical constraints (Alzoraiki et al., 2023). While significant advancements have been made, ongoing reforms must address structural inefficiencies and adapt to evolving educational demands. By integrating international best practices and leveraging technological innovations, China can further refine its leadership evaluation mechanisms to foster sustainable school improvement.

3.   Research Method

This study adopts a quantitative research approach to systematically examine the performance evaluation systems for school leaders in China. Quantitative methods are suitable for this investigation as they allow for the collection of numerical data that can be statistically analyzed to identify patterns, correlations, and trends in leadership evaluation practices. By employing structured surveys and standardized assessment metrics, this research ensures objectivity,

reliability, and generalizability of findings. The quantitative approach aligns with the study's goal of evaluating the effectiveness of existing performance evaluation frameworks while providing empirical evidence for potential policy improvements.

3.1 Research Design

The research follows a cross-sectional survey design, which enables the collection of data from a diverse sample of school leaders, administrators, and education policymakers at a single point in time. This design is appropriate for capturing current perceptions and practices related to leadership evaluation without the need for longitudinal tracking. The study incorporates both descriptive and inferential statistics to summarize evaluation trends and test hypotheses regarding the relationship between evaluation criteria and leadership effectiveness. Additionally, the research design includes a comparative analysis of different regions in China to assess variations in implementation due to local policy adaptations. By combining these elements, the study provides a comprehensive understanding of how performance evaluation systems function across different educational contexts.

3.2 Population and Sample

The target population for this study consists of school principals, vice-principals, and district-level education administrators across primary and secondary schools in China. Given the vast size and regional diversity of China's education system, a stratified random sampling technique is employed to ensure representation from urban, suburban, and rural schools. The sample is drawn from multiple provinces, including economically developed regions and less developed areas, to account for disparities in resource allocation and policy implementation. A minimum sample size of 500 participants is targeted to achieve statistical power, with adjustments made based on response rates. This sampling strategy enhances the study's external validity, allowing findings to be generalized across different school leadership contexts in China.

3.3 Instrumentation

The quantitative methodology outlined in this study is designed to produce empirical, data-driven insights into the performance evaluation systems for school leaders in China. By employing a cross-sectional survey, stratified sampling, and validated instrumentation, the research ensures methodological rigor while addressing regional and institutional variations. The findings will contribute to evidence-based recommendations for optimizing leadership evaluation frameworks in alignment with China's educational goals.

4. Findings and Discussions

The Table 1 presents a quantitative analysis of six key evaluation criteria used in assessing school leaders in China, displaying their mean scores and standard deviations based on a 5-point Likert scale (where 1=Never and 5=Always). Student academic performance emerges as the highest-weighted criterion (mean score=4.6, SD=0.7), reflecting China's strong emphasis on exam-based outcomes in leadership evaluations. Compliance with national policies (mean=4.4, SD=0.6) and school safety and discipline (mean=4.1, SD=0.8) also score highly, underscoring the dual focus on policy adherence and institutional management. In contrast, teacher professional development (mean=3.8, SD=0.9) and community/parent satisfaction (mean=3.5, SD=1.1) receive moderate attention, suggesting these areas may be secondary priorities. The lowest-scoring criterion, innovation in teaching methods (mean=3.2, SD=1.0), highlights a potential gap in incentivizing pedagogical creativity, possibly due to the dominance of standardized testing. The small standard deviations (all ≤1.1) indicate consistent agreement among respondents about the prevalence of these criteria, though the wider spread for community/parent satisfaction (SD=1.1) implies less consensus in its implementation. Overall, the table reveals a performance evaluation system that prioritizes academic results and policy alignment over collaborative or innovative leadership practices.

Table 1. Frequency of Evaluation Criteria Used (5-point Likert Scale: 1=Never, 5=Always)

Evaluation Criteria	Mean Score	Standard Deviation
Student academic performance	4.6	0.7
Teacher professional development	3.8	0.9
Compliance with national policies	4.4	0.6
School safety and discipline	4.1	0.8
Community/parent satisfaction	3.5	1.1
Innovation in teaching methods	3.2	1.0

The data reveals a concerning imbalance in China's school leadership evaluation system that warrants critical examination. While the strong emphasis on student academic performance (4.6) and policy compliance (4.4) aligns with the nation's exam-oriented education tradition, this narrow focus comes at significant pedagogical costs. The system's

prioritization of quantifiable outcomes creates a perverse incentive structure that likely discourages educational innovation (3.2) and holistic development, as evidenced by the marginalization of teaching creativity and professional growth (3.8). The moderate attention given to community/parent satisfaction (3.5) and the notable standard deviation (1.1) in its implementation suggest an inconsistent and potentially tokenistic approach to stakeholder engagement, despite its recognized importance in educational quality. More troubling is how these evaluation priorities reflect and reinforce a bureaucratic, top-down management approach that values compliance over leadership. By heavily weighting easily measurable indicators like test scores and policy adherence, the system risks reducing school leadership to administrative box-ticking rather than fostering transformative educational vision. The consistently low scores for innovative teaching methods expose a fundamental tension between China's stated goals of educational modernization and an evaluation regime that effectively penalizes pedagogical experimentation.

The small standard deviations across most criteria indicate this is not just a prevalent pattern but a systemic one, suggesting these imbalances are institutionalized rather than incidental. While such standardization ensures consistency, it may also reflect an inflexible, one-size-fits-all approach ill-suited to China's vast regional disparities in educational resources and needs. For China to achieve its ambition of becoming an education powerhouse, it must reform these evaluation frameworks to better balance accountability with support for professional autonomy, innovation, and community-responsive leadership. The current overemphasis on measurable outputs threatens to stifle the very qualities needed for 21st-century educational excellence.

The Table 2 presents a comparative analysis of four key evaluation criteria for school leaders across urban and suburban areas in China, based on survey responses from 300 urban and 120 suburban participants. The data reveals several notable patterns in leadership evaluation practices. First, standardized test scores are heavily emphasized in both settings, though slightly more in urban areas (4.5) compared to suburban schools (4.2). This aligns with China's broader focus on academic achievement as a primary metric for school performance. Second, government policy compliance checks rank highest overall (urban: 4.6, suburban: 4.3), reflecting the centralized nature of China's education system and the importance placed on adherence to national directives. The inclusion of teacher feedback shows moderate implementation, with urban schools (3.9) incorporating it more consistently than suburban institutions (3.6). Similarly, professional development emphasis scores lower in both contexts (urban: 3.7, suburban: 3.4), suggesting it may be a secondary priority compared to test results and policy compliance. The table highlights an evaluation system that prioritizes measurable outcomes (test scores, policy adherence) over collaborative or developmental approaches (teacher input, professional growth). The urban-suburban differences, while modest, may reflect disparities in resources, training, or policy implementation intensity. These findings could inform discussions about balancing accountability with support for leadership development in China's education system.

Table 2. Regional Differences in Evaluation Practices (Mean Scores)

Evaluation Criteria	Urban (n=300)	Suburban (n=120)
Use of standardized test scores	4.5	4.2
Teacher feedback inclusion	3.9	3.6
Government policy compliance checks	4.6	4.3
Professional development emphasis	3.7	3.4

The findings presented in the table reveal several critical issues in China's school leadership evaluation system that warrant deeper examination. While the heavy emphasis on standardized test scores (urban: 4.5; suburban: 4.2) and policy compliance (urban: 4.6; suburban: 4.3) aligns with China's centralized education governance, this approach raises concerns about its long-term sustainability and educational quality. The prioritization of quantifiable metrics may inadvertently encourage teaching-to-the-test behaviors and stifle pedagogical innovation, particularly as professional development receives notably lower emphasis (urban: 3.7; suburban: 3.4). The modest urban-suburban disparities in all criteria suggest systemic rigidity, where evaluation frameworks may be uniformly applied without sufficient adaptation to local contexts. While the slightly higher scores in urban areas could reflect better resources or stricter oversight, the minimal variation implies the system may be failing to address the unique challenges of suburban schools, such as resource gaps or diverse student needs. The marginal inclusion of teacher feedback (urban: 3.9; suburban: 3.6) is particularly problematic, as it sidelines the professional insights of educators in favor of top-down accountability measures. This imbalance risks undermining teacher morale and collaborative school improvement efforts. Furthermore, the near-identical patterns across regions suggest the evaluation system may be overly standardized, potentially neglecting the socioeconomic and cultural diversity within China's education landscape. To foster more equitable and effective school leadership, policymakers should consider rebalancing these criteria to value developmental inputs as much as measurable outputs, while allowing greater flexibility for regional adaptation. Without such reforms, the current system may perpetuate a narrow definition of leadership success that prioritizes compliance over meaningful educational transformation.

5. Conclusion

The findings of this study reveal that China's performance evaluation systems for school leaders prioritize measurable outcomes—particularly student academic performance (mean=4.6) and compliance with national policies (mean=4.4)—over developmental and innovative leadership practices. While this approach ensures accountability and standardization, it risks stifling pedagogical creativity (mean=3.2) and undervaluing teacher professional growth (mean=3.8) and community engagement (mean=3.5). The urban-suburban disparities, though modest, further highlight systemic rigidity, suggesting that evaluation frameworks may not adequately address regional inequities or local educational needs. These results underscore the need for reforms that balance high-stakes accountability with support for leadership autonomy, innovation, and stakeholder collaboration.

5.1 Implementation

The findings of this study present several critical opportunities for reforming China's school leadership evaluation system. To create a more balanced and effective framework, policymakers should first consider reducing the disproportionate emphasis on standardized test scores by introducing alternative metrics that capture holistic student development, such as social-emotional learning outcomes and creative problem-solving skills. Simultaneously, the evaluation process should be restructured to give greater weight to teacher professional development and innovative teaching practices, perhaps through dedicated incentive programs or recognition systems. The current system could also benefit from incorporating more robust mechanisms for gathering and integrating feedback from teachers, parents, and community stakeholders, ensuring evaluations reflect multiple perspectives on leadership effectiveness. Regional education authorities might develop differentiated evaluation rubrics that account for varying resource levels and challenges across urban, suburban, and rural contexts, while maintaining core national standards. Implementation should begin with pilot programs in select districts, accompanied by comprehensive training for evaluators and school leaders on the new assessment criteria. These changes should be rolled out gradually, with continuous monitoring and adjustment based on feedback from participating schools.

5.2 Future Research

This study opens several important avenues for further investigation that could deepen our understanding of school leadership evaluation. Future research should employ longitudinal designs to track how changes in evaluation systems affect leadership practices and school outcomes over time, particularly in terms of educational innovation and teacher retention. Comparative case studies between high-performing schools within China's system could reveal how different leaders navigate and respond to evaluation pressures, potentially identifying best practices. There is also a need for qualitative research exploring the lived experiences of school leaders working under the current evaluation regime, including how they perceive its strengths and limitations. Internationally, systematic comparisons with evaluation systems in countries like Finland and Singapore could yield valuable insights about alternative approaches to balancing accountability with professional autonomy. Emerging technologies present another promising research direction, particularly examining how data analytics and AI might enable more nuanced, real-time assessments of leadership effectiveness. Finally, equity-focused research should investigate how evaluation systems might be adapted to better serve schools in disadvantaged regions, potentially incorporating measures of resource equity and access to opportunities. These research directions could collectively inform the evolution of more effective, context-sensitive evaluation systems that support educational excellence across diverse settings.

Acknowledgement

Conflict of Interest

The authors declare no conflicts of interest.

References

Ahmad, N. (2021). An Analysis of Instructional Leadership Practices of Primary School Head-Teachers on Teacher Effectiveness: A Qualitative Study of Teachers' Perceptions. *Pakistan Languages and Humanities Review*, *5*(2), 193–209. https://doi.org/10.47205/plhr.2021(5-ii)2.16

Alzoraiki, M., Ahmad, A. R., Ateeq, A. A., Naji, G. M. A., Almaamari, Q., & Beshr, B. A. H. (2023). Impact of Teachers' Commitment to the Relationship between Transformational Leadership and Sustainable Teaching Performance. *Sustainability*, *15*(5), 4620. https://doi.org/10.3390/su15054620

Chatzipanagiotou, P., & Katsarou, E. (2023). Crisis management, School Leadership in Disruptive Times and the Recovery of Schools in the Post COVID-19 era: a Systematic Literature Review. *Education Sciences*, *13*(2), 118. https://doi.org/10.3390/educsci13020118

Chiu, T. (2023). The Impact of Generative AI (GenAI) on practices, Policies and Research Direction in education: a Case of ChatGPT and Midjourney. *Interactive Learning Environments*, *32*(10), 1–17. https://doi.org/10.1080/10494820.2023.2253861

Chu, W., Liu, H., & Fang, F. (2021). A Tale of Three Excellent Chinese EFL Teachers: Unpacking Teacher Professional Qualities for Their Sustainable Career Trajectories from an Ecological Perspective. *Sustainability*, *13*(12), 6721. https://doi.org/10.3390/su13126721

Gao, Y., Zeng, G., Wang, Y., Khan, A. A., & Wang, X. (2022). Exploring Educational Planning, Teacher Beliefs, and Teacher Practices During the Pandemic: A Study of Science and Technology-Based Universities in China. *Frontiers in Psychology*, *13,* 903244. https://doi.org/10.3389/fpsyg.2022.903244

Haiyan, Q., & Allan, W. (2020). Creating conditions for professional learning communities (PLCs) in schools in China: the role of school principals. *Professional Development in Education*, *47*(4), 1–13. https://doi.org/10.1080/19415257.2020.1770839

Li, L., & Liu, Y. (2020). An integrated model of principal transformational leadership and teacher leadership that is related to teacher self-efficacy and student academic performance. *Asia Pacific Journal of Education*, *42*(4), 1–18. https://doi.org/10.1080/02188791.2020.1806036

Liu, M., Ren, Y., Lucy Michael Nyagoga, Stonier, F., Wu, Z., & Liang, Y. (2023). Future of education in the era of generative artificial intelligence: Consensus among Chinese scholars on applications of ChatGPT in schools. *Future in Educational Research*, *1*(1), 72-101. https://doi.org/10.1002/fer3.10

Maheshwari, G. (2021). Influence of Teacher-Perceived Transformational and Transactional School Leadership on Teachers' Job Satisfaction and Performance: A Case of Vietnam. *Leadership and Policy in Schools*, *21*(4), 1–15. https://doi.org/10.1080/15700763.2020.1866020

Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., & Weisz, J. D. (2022). Investigating Explainability of Generative AI for Code through Scenario-based Design. *27th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3490099.3511119

Wang, Q., Hou, H., & Li, Z. (2022). Participative leadership: A literature review and prospects for future research. *Frontiers in Psychology*, *13*(1), 1–12. NCBI. https://doi.org/10.3389/fpsyg.2022.924357

Wang, Q., & Utomo, S. B. (2024). Administrative Leadership and Performance Management: Enhancing School Outcomes in Chinese Urban Schools. *Peta International Journal of Social Science and Humanity.*, *3*(4), 1-13. https://doi.org/10.59088/pij.v2i4.65

Wang, X., & Wang, T. (2020). The mutability of pedagogical practice and space use: a case study of collaborative learning and classroom space in a Chinese primary school. *Compare: A Journal of Comparative and International Education*, 1–19. https://doi.org/10.1080/03057925.2020.1811640

Yan, Z., & Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985. https://doi.org/10.1016/j.stueduc.2021.100985

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 1–33. https://doi.org/10.1080/0969594x.2021.1884042