



From Instrument Reform to Systemic Alignment Reform: Comparative Insights from Educational Evaluation Systems in Developed Countries

Ambyah Harjanto^{1,2*}, Fitriana Rahmawati^{1,2}, Rabiyatul Andawiyah Siregar²,
Handoko Handoko²

¹STKIP PGRI Bandar Lampung, Lampung, Indonesia

¹Universitas Lampung, Lampung, Indonesia

cambyasoul@gmail.com*

ABSTRACT

Educational evaluation has become a central instrument in contemporary educational reform, yet its effectiveness remains uneven across countries despite increasing investments in assessment systems and accountability mechanisms. This study examines educational evaluation systems in Singapore, Finland, Japan, England, Australia, the United States, and Germany from the perspective of educational policy evaluation. Using a qualitative comparative approach based on a Systematic Literature Review (SLR), the study synthesized 34 selected studies and policy documents through the analytical framework of Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). The findings reveal four distinct educational evaluation typologies: High-Alignment Systems, Trust-Based Adaptive Systems, Accountability-Driven Systems, and Coordinated/Fragmented Hybrid Systems. The analysis demonstrates that educational evaluation effectiveness is not primarily determined by the sophistication of assessment instruments but by the degree of alignment among policy objectives, institutional capacities, and implementation processes. Based on these findings, the study develops the Systemic Alignment Framework and proposes the concept of Systemic Alignment Reform as a new perspective for understanding educational evaluation effectiveness. The study contributes to comparative educational policy scholarship by integrating policy coherence, system capacity, and implementation fidelity into a unified explanatory framework and offers policy insights for strengthening sustainable educational reform across diverse governance contexts.

ARTICLE INFO

Keywords:

Comparative education;
Educational evaluation;
Educational policy evaluation;
Policy coherence;
System capacity

How to Cite in APA Style:

Harjanto, A., Rahmawati, F., Siregar, R. A., & Handoko, H. (2026). From Instrument Reform to Systemic Alignment Reform: Comparative Insights from Educational Evaluation Systems in Developed Countries. *IJLHE: International Journal of Language, Humanities, and Education*, 9(1), 219-234. <https://doi.org/10.52217/rp5bv398>

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



INTRODUCTION

Educational evaluation has become one of the most influential policy instruments in contemporary educational reform. Governments worldwide increasingly rely on assessment systems, accountability frameworks, and evaluation mechanisms to monitor educational quality, improve learning outcomes, and support evidence-based policymaking. International organizations such as the Organisation for Economic Co-operation and Development (OECD), UNESCO, and the World Bank have consistently emphasized the strategic role of educational evaluation in strengthening education systems and promoting sustainable educational development (OECD, 2021; UNESCO, 2021). The growing prominence of international assessments such as the Programme for International Student Assessment (PISA) has further intensified global attention toward educational evaluation as a key mechanism for measuring system performance and guiding policy reform (OECD, 2023).

The expansion of educational evaluation has been accompanied by substantial investments in assessment technologies, accountability systems, and performance measurement frameworks. Many countries have introduced standardized testing programs, national quality assurance systems, and data-driven decision-making mechanisms in an effort to improve educational effectiveness. Despite these developments, evidence suggests that improvements in assessment instruments do not automatically translate into better educational outcomes. Educational reforms frequently encounter implementation challenges, institutional resistance, policy fragmentation, and uneven organizational capacity, resulting in significant gaps between policy intentions and practical outcomes (Fullan, 2020; OECD, 2024).

Existing scholarship has provided valuable insights into various dimensions of educational evaluation. Several studies have focused on accountability systems and performance measurement as mechanisms for improving educational quality (Ball, 2003; Perryman et al., 2018; Kim, 2023). Other studies have examined the role of educational governance, policy implementation, and institutional capacity in shaping reform outcomes (Datnow & Park, 2021; Yan et al., 2023). Comparative education scholars have also investigated how countries borrow, adapt, and implement educational policies across different socio-political contexts (Steiner-Khamisi, 2014; Baek & Nordin, 2024). Furthermore, research on high-performing education systems has highlighted the importance of policy coherence, professional capacity, and organizational learning in sustaining educational improvement (Ng, 2020; Sahlberg, 2021; Hopkins, 2022).

Although these studies have contributed significantly to understanding educational reform processes, the literature remains fragmented in several respects. First, many studies examine assessment systems, educational governance, policy implementation, or institutional capacity as separate domains of inquiry. Second, comparative studies often focus on individual countries or specific policy instruments without exploring the systemic interactions that influence educational evaluation effectiveness. Third, existing research rarely integrates policy coherence, system capacity, and implementation fidelity within a single analytical framework capable of explaining why some educational

evaluation systems generate sustainable improvement while others produce limited policy impact. As a result, the mechanisms through which educational evaluation contributes to educational effectiveness remain insufficiently understood.

This limitation becomes particularly important in light of growing evidence that educational reform success depends not only on the quality of policy design or assessment instruments but also on broader governance conditions. OECD (2024) emphasizes that educational policies often fail when implementation processes, institutional capacities, and policy objectives are insufficiently aligned. Similarly, Yan et al. (2023) demonstrate that policy capacity significantly influences the success of educational reform initiatives, while Datnow and Park (2021) argue that the effective use of evaluation data depends heavily on organizational learning capacities. These findings suggest that educational evaluation should be understood as a systemic governance process rather than merely a technical measurement activity.

Accordingly, this study addresses an important research gap by examining educational evaluation systems through the integrated perspectives of Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). Policy coherence refers to the degree of alignment among educational goals, governance structures, and policy instruments (OECD, 2020). System capacity encompasses the institutional, organizational, and professional resources necessary to implement educational reforms effectively (Yan et al., 2023). Implementation fidelity refers to the extent to which educational policies are translated into practice while maintaining their intended objectives (Viennet & Pont, 2017). Together, these dimensions provide a comprehensive framework for understanding how educational evaluation systems function within broader policy environments.

The novelty of this study lies in the development of a Systemic Alignment Framework that conceptualizes educational evaluation effectiveness as the product of coordinated interactions among policy coherence, system capacity, and implementation fidelity. Unlike previous studies that have examined these dimensions separately, this research integrates them into a unified explanatory model for comparative educational evaluation analysis. The framework further extends existing comparative education literature by identifying patterns of systemic alignment across different educational governance traditions and evaluation regimes.

To examine these relationships, this study focuses on seven developed countries: Singapore, Finland, Japan, England, Australia, the United States, and Germany. These countries were selected because they represent distinct educational governance models and evaluation traditions. Singapore and Japan represent highly coordinated systems characterized by strong policy alignment and implementation consistency. Finland represents a trust-based model emphasizing professional autonomy and institutional capacity. England illustrates an accountability-driven governance system with extensive performance monitoring mechanisms. Australia, Germany, and the United States represent various forms of decentralized and federal educational governance, providing

valuable perspectives on the challenges of policy coordination and implementation across multiple administrative levels.

Based on these considerations, this study aims to analyze educational evaluation systems in seven developed countries from the perspective of educational policy evaluation. Specifically, the study seeks to: (1) examine the characteristics of educational evaluation systems across different governance contexts; (2) analyze the relationships among policy coherence, system capacity, and implementation fidelity; (3) identify comparative patterns of educational evaluation effectiveness; and (4) develop a systemic alignment perspective that can inform future educational policy and reform initiatives.

METHOD

This study employed a qualitative comparative approach using a Systematic Literature Review (SLR) design to examine educational evaluation systems in seven developed countries: Singapore, Finland, Japan, England, Australia, the United States, and Germany. The seven countries were purposively selected to represent diverse educational governance traditions and evaluation regimes. Singapore and Japan represent highly coordinated systems, Finland reflects a trust-based governance model, England represents an accountability-driven system, while Australia, Germany, and the United States illustrate decentralized and federal governance arrangements. The diversity of these governance contexts provides a robust comparative basis for examining the interaction among policy coherence, system capacity, and implementation fidelity. The review was conducted to identify patterns, similarities, differences, and policy implications of educational evaluation practices across diverse educational governance contexts. A systematic review approach was selected because it enables the synthesis of evidence from multiple studies and facilitates the development of broader theoretical insights regarding educational policy evaluation (Page et al., 2021).

Data Sources and Search Strategy

The literature search was conducted between January and March 2025 using major academic databases, including Scopus, Web of Science, ERIC, ScienceDirect, Taylor & Francis Online, SpringerLink, and OECD iLibrary. Additional documents were obtained from reports published by OECD, UNESCO, and national education agencies. The search was limited to publications published between 2020 and 2025 to ensure the inclusion of recent evidence regarding educational evaluation and educational policy reforms. Seminal works published before 2020 were included selectively when they provided foundational theoretical perspectives relevant to policy evaluation, implementation studies, or comparative education.

Inclusion and Exclusion Criteria

The selection of studies followed predefined inclusion and exclusion criteria. Publications were included if they: (a) discussed educational evaluation, assessment systems, educational governance, or policy implementation; (b) focused on one or more

of the seven selected countries; (c) were published in peer-reviewed international journals, books from reputable academic publishers, or reports issued by recognized international organizations; and (d) provided empirical, conceptual, or policy-related evidence relevant to educational evaluation. Publications were excluded if they: (a) focused exclusively on classroom assessment techniques without policy relevance; (b) lacked sufficient methodological transparency; (c) were conference abstracts, opinion pieces, or unpublished manuscripts; or (d) did not provide information relevant to policy coherence, system capacity, or implementation fidelity.

Study Selection Process

The study selection process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page et al., 2021). The initial database search identified 168 records. After removing 21 duplicate records, 147 publications remained for title and abstract screening. The screening process excluded studies that did not meet the inclusion criteria, resulting in 52 potentially relevant publications. A full-text review was subsequently conducted, leading to the final inclusion of 34 studies and policy documents for qualitative synthesis.

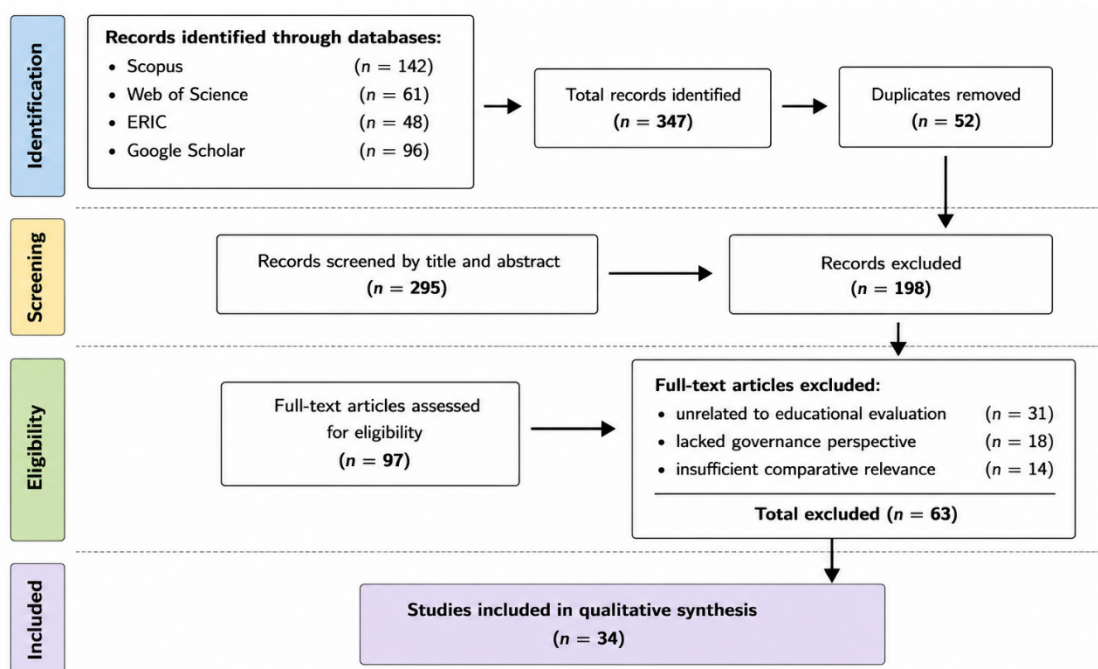


Figure 1. PRISMA Flow Diagram of Study Selection Process

Analytical Framework

Data analysis was conducted using thematic comparative analysis. The selected studies were systematically coded and categorized according to three analytical dimensions derived from educational policy and implementation literature: Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). The PC–SC–

IF framework was used to compare educational evaluation systems across the seven countries and to identify recurring patterns of alignment, fragmentation, and adaptation. Comparative findings were subsequently synthesized to develop the concept of Systemic Alignment Reform, which serves as the principal analytical contribution of this study.

Trustworthiness and Analytical Rigor

Several strategies were employed to enhance the credibility and rigor of the review. First, data were collected from multiple international databases and policy repositories to ensure source triangulation. Second, explicit inclusion and exclusion criteria were applied throughout the review process. Third, comparative coding procedures were conducted consistently across all selected studies. Finally, interpretations were continuously compared with existing theoretical and empirical literature to strengthen analytical validity and reduce potential bias.

RESULTS AND DISCUSSION

Results

Typologies of Educational Evaluation Systems Across Developed Countries

Comparative analysis revealed that educational evaluation systems in the seven developed countries could be classified into four distinct governance typologies based on the interaction among Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). The typologies demonstrate that educational effectiveness is influenced less by the sophistication of assessment instruments and more by the degree of alignment among policy objectives, institutional capacities, and implementation processes.

Table 1. Typology of Educational Evaluation Systems Based on PC–SC–IF Configuration

Typology	Countries	Dominant Characteristics	Policy Outcome
High-Alignment System	Singapore, Japan	Strong policy coherence, high capacity, high implementation fidelity	Consistent policy implementation and sustainable improvement
Trust-Based Adaptive System	Finland	Strong professional autonomy, high capacity, adaptive implementation	Continuous improvement through trust and professional responsibility
Accountability-Driven System	England	Strong accountability mechanisms and performance monitoring	High transparency but risk of performativity pressures
Coordinated/Fragmented Hybrid System	Australia, Germany, United States	Mixed levels of coherence, capacity, and fidelity across jurisdictions	Variable implementation and uneven policy outcomes

The findings indicate that no single educational evaluation model is universally

applicable. Different governance arrangements can produce effective educational outcomes when supported by appropriate institutional capacities and implementation mechanisms. Similar observations have been reported in comparative education research emphasizing the contextual nature of policy effectiveness (Steiner-Khamsi, 2014; Baek & Nordin, 2024).

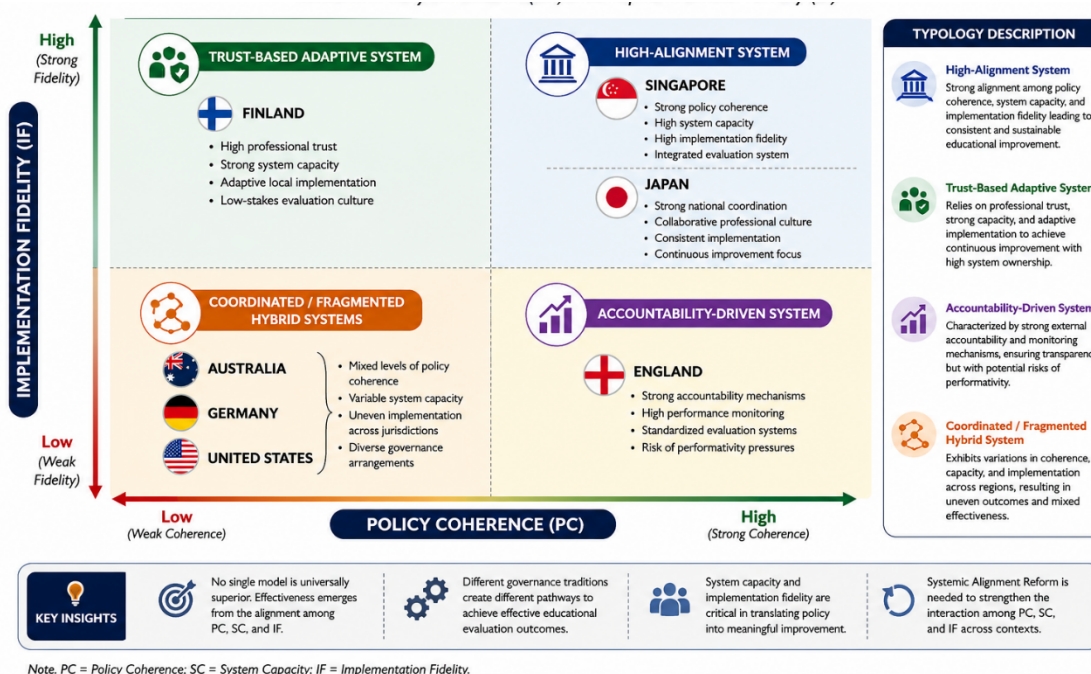


Figure 2. Typology of Educational Evaluation Systems Based on Policy Coherence and Implementation Fidelity

Policy Coherence as a Foundation for Educational Evaluation Effectiveness

The analysis demonstrates that policy coherence constitutes the foundational dimension of educational evaluation effectiveness. Countries exhibiting high levels of policy coherence tend to maintain stronger alignment among educational goals, curriculum frameworks, assessment systems, and accountability mechanisms.

Singapore represents the clearest example of policy coherence. Educational policies, curriculum standards, assessment frameworks, and professional development systems operate within a highly integrated governance structure, enabling educational evaluation results to be translated directly into policy refinement and institutional improvement (Ng, 2020). Japan demonstrates a similar pattern through strong national coordination, although implementation relies more heavily on collective professional responsibility and school-level collaboration.

Finland achieves policy coherence through a different mechanism. Rather than relying on centralized accountability systems, coherence is maintained through shared professional norms, strong teacher preparation, and trust-based governance arrangements (Sahlberg, 2021; Niemi, 2020). In contrast, the United States, Australia, and Germany

experience greater challenges in maintaining policy coherence due to decentralized governance structures and multiple layers of educational authority. These findings support OECD (2024), which argues that educational reforms are more likely to succeed when policy objectives, governance arrangements, and implementation mechanisms remain mutually aligned.

System Capacity as the Enabling Condition of Reform

The findings further reveal that policy coherence alone is insufficient to ensure effective educational evaluation. Educational systems require adequate institutional and professional capacities to transform policy intentions into sustainable practices.

Across the seven countries, system capacity emerged as a critical determinant of reform effectiveness. High-performing systems consistently demonstrated substantial investments in teacher quality, leadership development, organizational learning, and institutional support structures. Singapore and Finland exemplify this pattern through sustained commitments to teacher professionalism and continuous professional development (Ng, 2020; Sahlberg, 2021).

Japan similarly benefits from strong professional learning cultures that facilitate collective responsibility for educational improvement. Conversely, decentralized systems such as Australia, Germany, and the United States often face uneven capacity distribution across regions, schools, and local authorities. The findings corroborate Yan et al. (2023), who emphasize that policy capacity significantly influences the success of educational reforms and the effectiveness of educational governance systems.

Implementation Fidelity and the Translation of Policy into Practice

Implementation fidelity emerged as the dimension most directly associated with the practical effectiveness of educational evaluation systems. Even well-designed policies and highly developed institutional capacities may fail to generate intended outcomes when implementation processes become fragmented or inconsistent.

Singapore and Japan exhibit relatively high levels of implementation fidelity due to clear policy communication, strong administrative coordination, and robust monitoring systems. Finland demonstrates a distinct form of implementation fidelity characterized by professional trust and adaptive local implementation. By contrast, decentralized systems frequently encounter challenges in maintaining implementation fidelity across diverse administrative units. The United States provides a notable example where state-level variations often produce significant differences in educational policy enactment and evaluation practices (OECD, 2024). These findings are consistent with implementation literature emphasizing that policy effectiveness depends not only on policy design but also on the quality of implementation processes and institutional coordination (Viennet & Pont, 2017; Honig, 2006).

Toward Systemic Alignment Reform

The comparative findings collectively suggest that educational evaluation

effectiveness is best understood as the product of systemic alignment among policy coherence, system capacity, and implementation fidelity. None of these dimensions independently explains educational success. Instead, educational effectiveness emerges when the three dimensions interact in a coordinated manner.

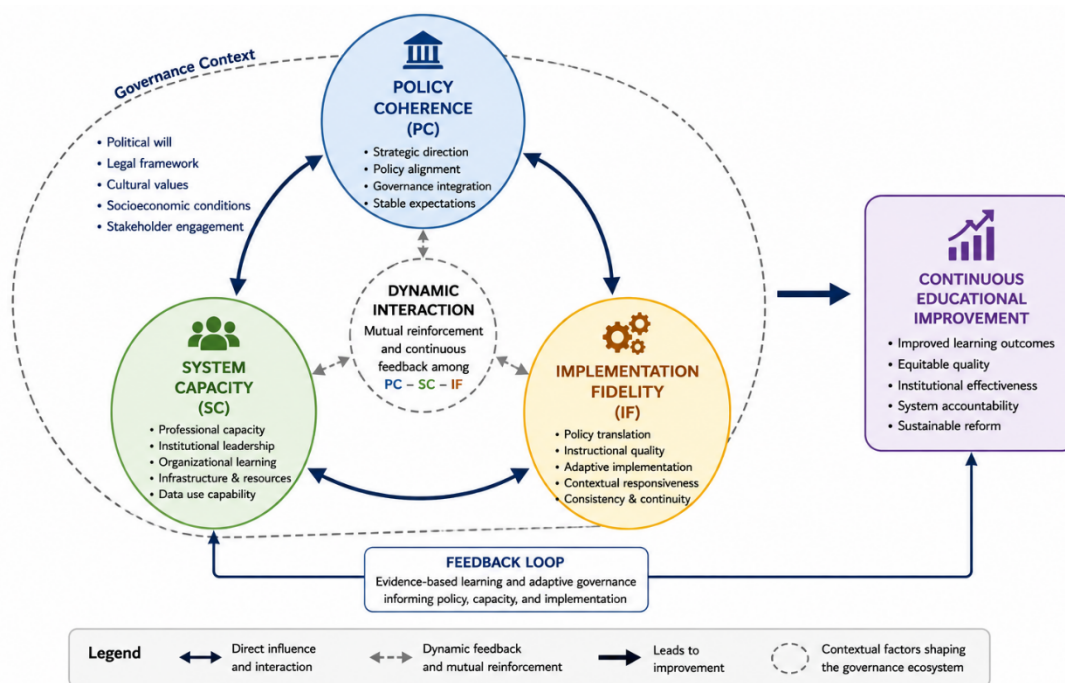


Figure 3. Systemic Alignment Framework of Educational Evaluation Effectiveness

Table 2. PC–SC–IF Comparative Matrix

Country	Policy Coherence (PC)	System Capacity (SC)	Implementation Fidelity (IF)	Alignment Level
Singapore	High	High	High	Very High
Japan	High	High	High	Very High
Finland	High	High	Moderate–High	High
England	High	Moderate	High	High
Australia	Moderate	High	Moderate	Moderate
Germany	Moderate	High	Moderate	Moderate
United States	Moderate	Moderate	Moderate	Moderate

The analysis therefore supports a transition from what may be termed instrument reform toward systemic alignment reform. Instrument reform focuses primarily on improving assessment tools, accountability mechanisms, or evaluation procedures. Systemic alignment reform, by contrast, emphasizes the coordinated development of governance structures, institutional capacities, and implementation systems that enable educational evaluation to contribute meaningfully to educational improvement. This perspective constitutes the principal analytical contribution of the study and provides a

foundation for understanding why educational reforms produce different outcomes across governance contexts despite employing similar evaluation instruments.

Discussion

Educational Evaluation Beyond Measurement and Accountability

The findings of this study suggest that educational evaluation should be understood as a governance mechanism rather than merely a measurement instrument. Countries with different assessment architectures often demonstrate comparable educational effectiveness when educational evaluation is embedded within broader systems of policy coordination and institutional support. Finland, for example, achieves strong educational outcomes despite relying less heavily on high-stakes accountability mechanisms than England or some states in the United States (Sahlberg, 2021). Similarly, Singapore and Japan demonstrate that educational evaluation can function effectively when assessment systems are integrated into coherent policy frameworks and supported by strong implementation structures (Ng, 2020).

Reinterpreting Educational Evaluation Through the PC–SC–IF Lens

A central finding of this study is that educational evaluation effectiveness is shaped by the interaction among Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). The comparative analysis demonstrates that these dimensions should not be treated as separate explanatory variables. Instead, they function as mutually reinforcing components of a broader educational governance system.



Figure 4. PC-SC-IF Interaction Model (the Synergetic Pathway to Effective Educational and Sustainable Improvement)

As illustrated in Figure 6, educational evaluation effectiveness emerges from the dynamic interaction among policy coherence, system capacity, and implementation fidelity rather than from any single dimension in isolation.. This finding contributes to ongoing debates within educational policy scholarship by emphasizing relational rather than isolated explanations of educational effectiveness.

Understanding Governance Diversity Through Evaluation Typologies

The typology developed in this study further demonstrates that educational effectiveness can emerge through multiple governance pathways. Singapore and Japan illustrate high-alignment systems characterized by strong coordination among policy design, institutional capacity, and implementation processes. Finland represents a trust-based adaptive system in which professional autonomy and institutional trust function as key mechanisms for maintaining coherence and implementation consistency. England

reflects an accountability-driven model where evaluation systems rely heavily on performance monitoring and external accountability structures. Meanwhile, Australia, Germany, and the United States demonstrate hybrid governance arrangements. These findings are consistent with comparative education research emphasizing the contextual nature of policy borrowing and policy transfer (Steiner-Khamsi, 2014; Baek & Nordin, 2024).

From Instrument Reform to Systemic Alignment Reform

The most important contribution of this study lies in its proposal of Systemic Alignment Reform as a conceptual framework for understanding educational evaluation effectiveness. Much contemporary educational reform remains dominated by instrument reform. Such reforms typically focus on improving assessment tools, introducing new accountability mechanisms, revising curriculum standards, or developing more sophisticated performance indicators. The comparative findings indicate that educational effectiveness emerges when policy coherence, system capacity, and implementation fidelity operate in a coordinated manner. The concept of Systemic Alignment Reform extends existing educational policy literature by integrating insights from governance theory, implementation studies, and educational evaluation research into a unified analytical framework.

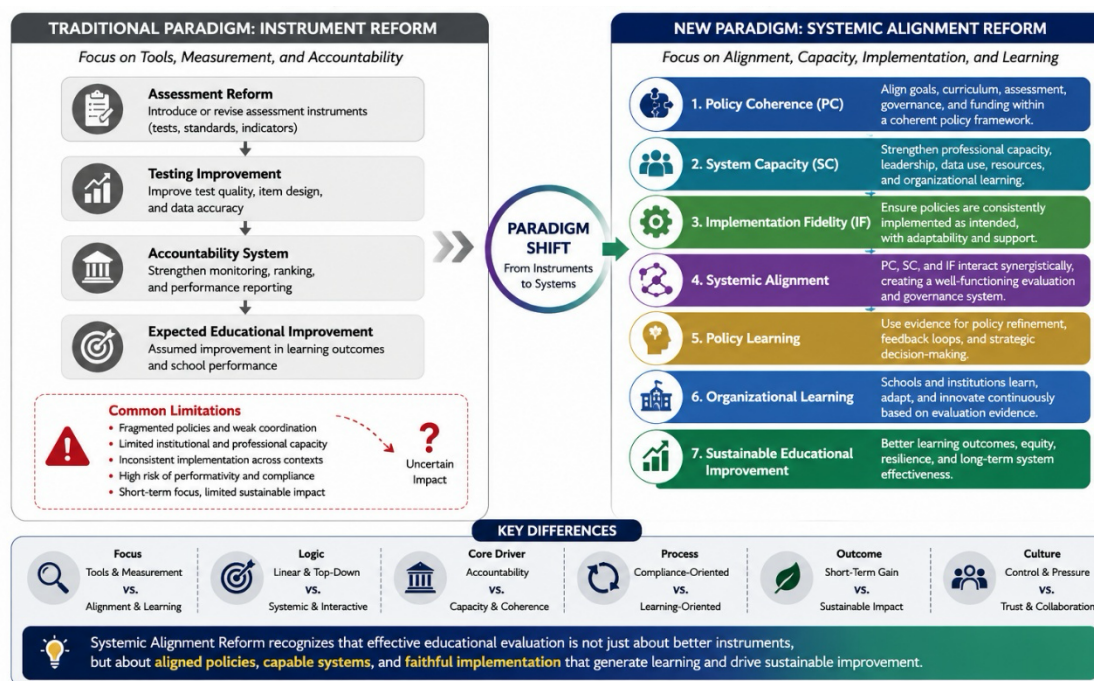


Figure 5. From Instrument Reform to Systemic Alignment Reform

Implications for Educational Policy and Reform

The findings also generate important implications for educational policymaking, particularly in developing countries seeking to strengthen educational quality through evaluation reforms. The comparative evidence suggests that policymakers should devote

equal attention to strengthening institutional capacity and implementation systems. Educational evaluation is most likely to support sustainable improvement when policy objectives are coherent, professional capacities are continuously developed, and implementation processes remain aligned with intended goals. For countries such as Indonesia, the findings imply that the principal challenge may not be the absence of evaluation instruments but the limited integration among policy design, organizational capacity, and implementation practice.

Table 3. Policy Implications for Developing Countries

Key Finding	Implication for Policy	Implication for Practice
Policy coherence matters	Align curriculum, assessment, and governance	Reduce policy inconsistency
System capacity matters	Invest in teachers and leaders	Strengthen professional learning
Implementation fidelity matters	Improve monitoring and feedback	Ensure consistent execution
Systemic alignment matters	Integrate reform initiatives	Build sustainable improvement systems

Theoretical Contribution

From a theoretical perspective, this study contributes to comparative educational policy scholarship in three ways. First, it synthesizes previously fragmented strands of literature concerning educational evaluation, governance, capacity building, and policy implementation. Second, it develops a comparative typology of educational evaluation systems based on PC–SC–IF configurations across seven developed countries. Third, it introduces the concept of Systemic Alignment Reform as a new explanatory framework for understanding why educational evaluation systems produce different outcomes despite employing similar assessment instruments.

CONCLUSIONS

This study examined educational evaluation systems in Singapore, Finland, Japan, England, Australia, the United States, and Germany from the perspective of educational policy evaluation. The comparative analysis reveals that educational effectiveness is not primarily determined by the sophistication of assessment instruments or accountability systems. What distinguishes effective educational evaluation systems is the extent to which policy objectives, institutional capacities, and implementation processes operate in a coordinated and mutually reinforcing manner.

A central contribution of this study is the identification of four educational evaluation typologies: High-Alignment Systems, Trust-Based Adaptive Systems, Accountability-Driven Systems, and Coordinated/Fragmented Hybrid Systems. These typologies demonstrate that no single model of educational evaluation can be considered universally superior.

The most significant theoretical contribution of this study lies in the development

of the Systemic Alignment Framework and the associated concept of Systemic Alignment Reform. The proposed framework conceptualizes educational evaluation effectiveness as the outcome of coordinated interactions among Policy Coherence (PC), System Capacity (SC), and Implementation Fidelity (IF). The concept of Systemic Alignment Reform represents a shift from conventional instrument-centered reform approaches toward a broader governance-oriented perspective.

Despite these contributions, the study has several limitations. The analysis relies primarily on secondary literature and policy documents. Furthermore, the study focuses on macro-level governance patterns. Future research may extend the Systemic Alignment Framework through empirical investigations involving quantitative, mixed-method, or multi-level comparative approaches.

In conclusion, this study argues that the future of educational evaluation lies not in the continuous expansion of assessment instruments but in the capacity of educational systems to achieve coherent relationships among policy design, institutional capacity, and implementation practice. Educational reform therefore requires a transition from instrument reform toward systemic alignment reform.

REFERENCES

- Baek, C., & Nordin, A. (2024). Understanding implicit reference societies in education policy. *Journal of Education Policy*, 39(5), 736–754. <https://doi.org/10.1080/02680939.2023.2282456>
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Datnow, A., & Park, V. (2021). *Data-driven leadership*. Jossey-Bass.
- Fullan, M. (2020). *Leading in a culture of change* (2nd ed.). Jossey-Bass.
- Honig, M. I. (2006). *New directions in education policy implementation: Confronting complexity*. State University of New York Press.
- Hopkins, D. (2022). *School and system improvement: A narrative state-of-the-art review*. Routledge.
- Kim, Y. (2023). Educational accountability policy for sustainable development: A comparative analysis across 30 countries. *Sustainability*, 15(18), 13883. <https://doi.org/10.3390/su151813883>
- Ng, P. T. (2020). *Learning from Singapore: The power of paradoxes*. Routledge.
- Niemi, H. (2020). Teacher professionalism and educational transformation in Finland. *Educational Research for Policy and Practice*, 19(3), 181–194. <https://doi.org/10.1007/s10671-019-09266-5>
- OECD. (2020). *Education policy evaluation: Surveying the OECD landscape*. OECD Publishing. <https://doi.org/10.1787/9f127490-en>
- OECD. (2021). *Education policy outlook 2021: Shaping responsive and resilient education in a changing world*. OECD Publishing.
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2024). *Education policy outlook 2024: Strengthening capacity through evaluation and monitoring*. OECD Publishing.

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Perryman, J., Ball, S. J., Braun, A., & Maguire, M. (2018). Can schools perform? Educational performativity and accountability. *Journal of Education Policy*, 33(3), 331–345. <https://doi.org/10.1080/02680939.2017.1388618>
- Sahlberg, P. (2021). *Finnish lessons 3.0: What can the world learn from educational change in Finland?* Teachers College Press.
- Steiner-Khamsi, G. (2014). *Cross-national policy borrowing in education: Politics of externalization and international policy attraction.* Teachers College Press.
- UNESCO. (2021). *Reimagining our futures together: A new social contract for education.* UNESCO Publishing.
- Viennet, R., & Pont, B. (2017). *Education policy implementation: A literature review and proposed framework* (OECD Education Working Papers No. 162). OECD Publishing.
- Yan, Y., Sano, H., & Sumiya, L. A. (2023). Policy capacity matters for education system reforms: A comparative study of two Brazilian states. *Journal of Comparative Policy Analysis*, 25(2), 253–281. <https://doi.org/10.1080/13876988.2022.2110472>

