

# Developing an Indicator System for Assessing the Effectiveness of Public Expenditure on Education in Vietnam

THI DOAN TRANG HA, THU HUYEN PHAM, THU TRANG PHAM  
Faculty of Public Finance,  
Academy of Finance,  
Hanoi,  
VIETNAM

**Abstract:** - This study aims to develop and validate an indicator system for assessing the effectiveness of public expenditure on education in Vietnam, conceptualizing effectiveness as a multi-layered policy transformation process rather than a linear relationship between spending levels and educational outcomes. Public expenditure effectiveness is conceptualized across five pillars: budgetary resources; allocation and operational mechanisms; direct manifestations of budget utilization; medium-term systemic changes; and long-term social impacts. Based on this framework, the study designs an indicator system and conducts a three-round Delphi process involving 62 experts in public finance, public administration, education economics, and policy evaluation. The indicator system demonstrates high structural consistency and strong inter-expert reliability, while content validity reflects the contested and context-dependent nature of policy measurement. The standardized system is refined to 27 indicators, designed for policy diagnosis rather than causal inference or performance ranking. The study provides a valuable measurement tool for analyzing and governing public expenditure on education in transitional institutional contexts.

**Key-Words:** - public expenditure effectiveness; education finance governance; policy-oriented indicator system; Delphi-based validation; performance-informed budgeting; public financial management; fiscal decentralization; content validity; Vietnam.

Received: April 9, 2026. Revised: May 25, 2026. Accepted: June 29, 2026. Published: September 22, 2026.

## 1 Introduction

In contemporary public policy scholarship, the effectiveness of public expenditure has shifted from an accounting-oriented concept toward a complex analytical construct that reflects the relationship between resource allocation, institutional capacity, and long-term social outcomes [1], [2], [3]. This shift is particularly salient in education – a domain of public expenditure that simultaneously represents a strategic investment, exhibits long impact lags, generates significant social externalities, and resists quantification through conventional fiscal indicators [4], [5]. In this context, the central question of education budget governance no longer concerns the scale of spending alone, but rather a more fundamental issue: whether public expenditure on education is transformed into commensurate public value, and how such transformation can be measured systematically.

In Vietnam, education has long been established as a priority sector within national

development policy, as reflected in strategic documents and the prevailing legal framework. Nevertheless, publicly available budget data reveal a complex and multi-layered picture. According to data from the World Bank and the UNESCO Institute for Statistics, Vietnam's public expenditure on education as a share of GDP in recent years has fluctuated around 2.8–3.0%, which is considerably lower than the 4–6% of GDP commonly regarded as an international benchmark for developing countries [6], [7]. When measured as a proportion of total state budget expenditure, education spending in Vietnam has accounted for approximately 15–16% in recent years, falling short of the 20% target articulated in long-term policy orientations [8], [9]. These figures indicate that while education remains a major expenditure item within the budget structure, it is not necessarily exceptional in comparative international terms.

More importantly, assessments of the effectiveness of public expenditure on education in Vietnam continue to rely predominantly on input-

based indicators or budget compliance criteria, rather than on measures that capture the relationships among resources, processes, outputs, and impacts [2], [10]. Budgetary and expenditure finalization reports show that recurrent spending on education has continued to increase in absolute terms – reaching hundreds of trillions of Vietnamese dong annually – yet improvements in quality, equity, and systemic effectiveness remain uneven across regions, education levels, and population groups [9], [11]. This situation highlights a policy paradox: increased spending does not necessarily translate into increased effectiveness, and the absence of a reliable measurement framework allows debates over education budgets to devolve into subjective or purely qualitative arguments.

From the perspective of public finance theory, expenditure effectiveness encompasses not only technical efficiency but also allocative efficiency and dynamic efficiency, particularly in human capital investment sectors such as education [1], [12]. Meanwhile, education economics emphasizes that educational outcomes are shaped jointly by policy design, implementation capacity, and institutional context, rendering direct attribution from spending levels to learning outcomes inaccurate in the absence of appropriate intermediate indicators [13], [14]. This divergence in measurement logics has created a significant scholarly gap: the lack of an indicator system for assessing the effectiveness of public expenditure on education that is grounded in public finance principles yet sufficiently sensitive to the distinctive characteristics of education as a special public good.

Recent international studies increasingly underscore the role of effectiveness indicators in performance-based budgeting reforms and evidence-based policymaking [10], [15]. However, most existing indicator frameworks have been developed within OECD country contexts, characterized by rich data environments, stable institutions, and relatively high levels of autonomy among public sector entities. Directly transferring these frameworks to Vietnam – a transitional economy with strong fiscal decentralization, substantial inter-regional disparities, and uneven levels of educational autonomy – poses considerable challenges in terms of contextual relevance and operational feasibility [9]. Consequently, the need to develop an indicator system that is institutionally “localized” while remaining aligned with international academic standards has become particularly pressing.

Against this backdrop, the central objective of this study is to develop an indicator system for assessing the effectiveness of public expenditure on education in Vietnam, designed as a policy analysis tool rather than a purely technical measurement instrument. Specifically, the study pursues three closely interconnected objectives: (i) *to conceptualize public expenditure effectiveness in education as a multidimensional construct that fully reflects the layers of the budgetary cycle*; (ii) *to design and standardize an indicator system with content validity, measurability, and feasibility under Vietnamese data conditions*; and (iii) *to assess, through structured expert validation, the operational feasibility of this indicator system as a tool for analyzing and comparing expenditure effectiveness across units and over time*. In doing so, the study seeks to contribute on two fronts: academically, by enriching the discourse on public expenditure effectiveness in education; and in policy terms, by offering an evidence-informed decision-support tool aligned with contemporary public financial management reform requirements.

In sum, this study approaches the effectiveness of public expenditure on education not as a policy slogan or a simple calculation, but as a complex issue of institutional design and policy measurement. As Vietnam faces the dual pressures of expanding educational access, improving quality, and maintaining fiscal discipline, developing an indicator system to assess the effectiveness of public expenditure is not merely an academic endeavor, but a necessary condition for strengthening long-term state governance capacity.

## **2 Theoretical Foundations and Analytical Framework**

### **2.1 The Theory of Public Expenditure Effectiveness in Public Finance**

In modern public finance, public expenditure effectiveness is no longer understood as a unilinear technical concept, but rather as a multidimensional construct associated with the creation of public value through the allocation and utilization of budgetary resources. The classical theoretical foundation of this approach originates from the three-function framework of the state budget proposed by Musgrave and Musgrave, in which expenditure effectiveness is closely linked to the allocative function of public finance [1]. Accordingly, public expenditure is considered effective when budgetary resources are allocated to

sectors and programs that generate the highest marginal social benefits under fiscal constraints.

Alongside public management reforms since the late twentieth century – particularly under the influence of New Public Management – the concept of public expenditure effectiveness has been expanded to emphasize outcomes and performance rather than mere input compliance [16], [17]. The literature commonly distinguishes three fundamental forms of effectiveness: technical effectiveness, reflecting the capacity to transform inputs into outputs; allocative effectiveness, reflecting the alignment between expenditure structures and social priorities; and dynamic effectiveness, reflecting the capacity to generate sustainable benefits over the long term [12], [18]. In social investment sectors such as education, dynamic effectiveness plays a central role due to long impact lags and strong institutional dependence.

In parallel, the value-for-money approach conceptualizes expenditure effectiveness as a combination of economy, efficiency, and effectiveness, emphasizing the relationship between costs and the public value generated [2]. However, this conceptual expansion has also intensified measurement challenges, requiring indicator systems that capture the full layers of the budget cycle rather than relying on isolated indicators. This constitutes a critical theoretical premise for designing expenditure effectiveness assessment frameworks in complex public sectors such as education.

## 2.2 Education as a Distinctive Domain of Public Expenditure

Education is regarded as a distinctive domain of public expenditure because it simultaneously exhibits characteristics of both public and private goods, generates substantial social externalities, and cannot be fully internalized through market mechanisms [3]. The economics of education demonstrates that public investment in education produces benefits extending beyond individual learners, including economic growth, labor productivity, and social cohesion, thereby reinforcing the central role of the state in education financing [4], [5].

At the same time, these characteristics complicate the assessment of expenditure effectiveness. First, long impact lags render the relationship between spending and outcomes indirect and highly susceptible to non-budgetary influences [13]. Second, educational objectives are inherently multidimensional – encompassing

quality, equity, access, and labor market relevance – making it impractical to reduce outcomes to a single indicator [14]. Third, institutional fragmentation in education governance, particularly in highly decentralized systems, causes expenditure effectiveness to depend heavily on allocative and implementation capacity across different levels of government [9].

A substantial body of empirical research shows that the relationship between education spending levels and learning outcomes is non-linear: higher spending does not automatically yield better outcomes, while resource utilization efficiency and governance quality play decisive roles [15], [19]. Accordingly, in this study, education is conceptualized as a complex policy space in which public expenditure effectiveness emerges from the interaction between budgetary design, allocation mechanisms, and institutional capacity.

## 2.3 Approaches to Measuring Effectiveness in Policy Evaluation

In public policy evaluation, effectiveness measurement has evolved through multiple approaches to reflect the diversity of policy objectives and contexts. The Input–Process–Output–Outcome–Impact model is among the most widely used analytical frameworks, allowing the differentiation of policy intervention layers and avoiding the conflation of short-term results with long-term impacts [20]. This framework is particularly well suited to education, where the effects of public expenditure often materialize only in the medium and long term.

In addition, quantitative methods such as Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) are widely employed to measure the technical efficiency of public expenditure by comparing input–output relationships across units [12], [21]. However, these methods primarily capture technical efficiency and face limitations in accounting for multidimensional policy objectives and institutional factors, thereby constraining their applicability as tools for evaluating expenditure effectiveness at the policy level.

More recently, performance-based budgeting and evidence-based governance approaches have stimulated the development of composite indicator systems that integrate both quantitative and qualitative measures [2], [15]. Nevertheless, the effectiveness of such systems depends heavily on institutional fit and their capacity to be embedded within actual decision-making processes [10]. It can therefore be argued that assessing the

effectiveness of public expenditure on education requires a hybrid measurement framework that both inherits core principles of public finance and is adapted to reflect the specific objectives and data realities of education – forming the direct foundation for the analytical framework proposed in this study.

## **2.4 Conceptual Framework for Assessing the Effectiveness of Public Expenditure on Education**

Drawing on theories of public expenditure effectiveness, the economics of education, and policy evaluation, this study develops a conceptual framework (Figure 1, Appendix) that approaches the effectiveness of public expenditure on education as a multi-layered policy phenomenon shaped through interactions among fiscal resources, operational mechanisms, and institutional context. This perspective aligns with the arguments in [22], [23], which contend that budgets are not merely financial allocation tools but policy institutions that reflect state priorities, capacities, and constraints.

At the core of the conceptual framework is the positioning of public expenditure on education within the entire lifecycle of budgetary policy, rather than treating it as an input flow or a direct explanatory factor for educational outcomes. Public financial management research indicates that effectiveness does not arise at the moment of disbursement, but rather accumulates through successive policy decisions – from budget formulation and allocation to implementation, monitoring, and adjustment [24], [25]. Accordingly, public expenditure effectiveness in education should be understood as a gradual transformation process in which public value is created or eroded at multiple points along the operational chain.

At the foundational level, the framework establishes the initial fiscal context of public expenditure on education, including the scale, composition, and stability of budgetary resources. These factors reflect policy commitment and state resource mobilization capacity, but are not regarded as intrinsic manifestations of effectiveness. This view is consistent with the arguments in [23], [26], which caution that an excessive focus on spending levels can generate “input illusions,” whereby increased expenditure is mistakenly equated with improved effectiveness.

The subsequent level captures the operational functioning of public expenditure, where allocation rules, transfer mechanisms, degrees of flexibility and transparency, and the implementation capacity of public actors begin to shape actual resource use.

Numerous studies emphasize that this intermediate space is decisive, as it is at this stage that expenditure effectiveness is often created or eroded, largely independently of budget size [10], [27]. In education, these operational mechanisms are particularly critical due to decentralized governance structures and substantial variation in institutional capacity across localities and service providers.

The conceptual framework further distinguishes direct manifestations of budget utilization from systemic changes. These direct manifestations reflect the extent to which public resources are translated into operational conditions of the education system, such as service provision capacity or program coverage. However, consistent with the cautions in [28], [29], such manifestations are not equated with education quality or learning outcomes, as they represent only intermediate layers within the public value chain.

At higher levels, the framework considers medium- and long-term changes associated with public expenditure on education, including system capacity, equity, and broader social benefits. These changes align with the perspective in [15], which treats them as reference points for assessing policy coherence rather than as direct measures of expenditure effectiveness. As documented in [30], public policy reforms frequently generate unintended consequences, and maintaining this analytical distance helps avoid reductive causal reasoning, whereby all improvements or declines in social outcomes are directly attributed to budgetary spending.

The core of the conceptual framework lies in the alignment among three analytical axes: the orientation of education policy objectives, the design and operation of fiscal mechanisms, and institutional capacity for implementation. This approach directly inherits the concepts of policy coherence and institutional alignment in policy studies, which posit that effectiveness emerges only when objectives, instruments, and implementation capacities are mutually consistent [31], [32]. Conversely, misalignment among these axes can result in low effectiveness even under conditions of high nominal expenditure – a phenomenon widely documented in education and public finance research in developing economies.

In sum, the proposed conceptual framework does not aim to produce a simple *ex post* metric, but rather repositions the effectiveness of public expenditure on education as a matter of fiscal policy design and alignment within a specific institutional context. This framework provides the

foundational structure for organizing the subsequent analysis of the study, delineating the boundaries of causal inference and establishing a coherent analytical logic for the development of measurement components in the following sections.

### 3 Design of the Indicator System

#### 3.1 Principles for Constructing Indicators to Assess Public Expenditure Effectiveness

In this study, the construction of indicators for assessing the effectiveness of public expenditure on education is not approached as a technical exercise of selecting available measurement variables, but rather as a normative policy design problem. Accordingly, indicators are required not only to be “measurable,” but also to reflect accurately the logic through which effectiveness is generated in public expenditure, consistent with the nature of the budget as a policy institution rather than a purely accounting instrument [22], [23]. The set of principles proposed here thus functions as a “conceptual filter,” ensuring that each indicator included in the system has a clear theoretical position and meaningful policy-explanatory value.

The first principle is alignment with the transformation chain of public expenditure policy. Indicators must be designed to capture different layers of the process through which fiscal resources are transformed into public value, rather than focusing on a single analytical cut. This principle directly builds on the arguments in [33] and on public financial management scholarship on the budget cycle [24], [34], which emphasize that policy effectiveness is not the immediate outcome of spending decisions, but the cumulative product of successive operational stages. Constructing indicators that overlook intermediate layers risks reducing expenditure effectiveness to a linear relationship between spending levels and final outcomes – an approach that has been widely criticized in public finance and comparative education research.

The second principle is the conceptual distinction between expenditure effectiveness and educational outcomes. The indicators are not designed to measure learning quality, student achievement, or training outcomes directly, but to reflect the extent to which public expenditure is organized, allocated, and utilized in ways that can generate such outcomes. This principle is consistent with the views in [35], [36], which argue

that equating effective management with the measurement of final outcomes can lead to “management by numbers,” obscuring the role of policy design and institutional capacity. In education, where outcomes are influenced by many factors beyond budgetary inputs, this distinction is particularly necessary to avoid excessive causal inference.

The third principle is institutional sensitivity. Indicators must reflect the role of institutional context, particularly fiscal decentralization arrangements, the degree of autonomy of budget-using units, and the coordination capacity of state management agencies. Numerous studies demonstrate that identical spending levels can yield markedly different effectiveness outcomes across institutional contexts, due to variations in operational rules and implementation capacity [25], [37], [38]. Accordingly, the indicators must allow identification not only of “how much is spent,” but also of “under what institutional conditions spending occurs,” thereby enhancing the policy-diagnostic value of the indicator system.

The fourth principle is usability within the budgetary and public governance cycle. Indicators assessing public expenditure effectiveness must be capable of integration across different stages of the budget cycle – from budget formulation and allocation to implementation, monitoring, and policy adjustment. This principle is emphasized in studies on performance-based budgeting and evidence-informed governance, which argue that indicators are meaningful only when they can be practically used in decision-making processes, rather than existing solely as *ex post* analytical tools [34], [39]. In education, where budgets are typically recurrent and long-term, such usability is especially important to support policy adjustment over time.

The fifth principle is balance between international standards and local adaptability. The indicator system must reference international measurement frameworks and practices to ensure comparability and academic value, while also being adapted to Vietnam's data conditions, statistical capacity, and governance context. Research on policy transfer shows that the mechanical adoption of international indicators without conceptual and institutional adjustment often leads to failure or the production of purely symbolic measures [40], [41]. This principle should therefore not be understood as a compromise driven by data limitations, but as a theoretically grounded choice to ensure the operational relevance and policy significance of the indicator system.

Taken together, these principles demonstrate that constructing indicators for assessing public expenditure effectiveness in education in this study is a process of reverse reasoning from the established conceptual framework and policy philosophy, rather than a mechanical selection of measurement variables. This set of principles provides the foundation for structuring the multidimensional indicator system presented in subsequent subsections, ensuring that each indicator contributes to clarifying how public expenditure on education is transformed into public value within a specific institutional context.

### 3.2 Structure of the Multidimensional Indicator System

Building on the established conceptual framework, the design of the indicator system in this study is approached as a problem of measurement architecture rather than an exercise in selecting or aggregating individual indicators. The structure of the indicator system is intended to organize measurement information across different analytical dimensions of public expenditure effectiveness in education, thereby enabling effectiveness to be interpreted as a complex policy phenomenon that cannot be reduced to a single metric.

The adoption of a multidimensional orientation stems from longstanding critiques in policy studies and public finance of single-indicator approaches, which tend to obscure intermediate mechanisms and impoverish the information available for decision-making [42], [43]. In the context of public expenditure on education, where effectiveness is shaped simultaneously by budget design, institutional capacity, and implementation conditions, reliance on a composite index would not only lack explanatory power but also risk generating misleading policy interpretations.

The multidimensional indicator system is designed according to a principle of purposeful informational complementarity. Each indicator group is not intended to reflect the same aspect of expenditure effectiveness, but rather to address distinct policy questions regarding how public resources are organized, operated, and transformed. Under this approach, indicator groups do not compete with or substitute for one another, but coexist within a single analytical structure, allowing users to examine expenditure effectiveness from multiple perspectives simultaneously. This organizational logic is consistent with the argument in [44] that performance measurement in the public sector

should prioritize polycentric information over technical simplification.

A key feature of the indicator system structure is its policy-diagnostic orientation rather than scoring or ranking. Indicator groups are organized to support the identification of bottlenecks within the operational chain of public expenditure on education, such as problems arising in allocation, implementation, or institutional coordination. As noted in [45], performance measurement systems are valuable only insofar as they help managers and policymakers understand why effectiveness is achieved or not achieved, rather than merely indicating effectiveness levels at a given point in time. Accordingly, the measurement structure in this study is designed to support causal reasoning and policy adjustment, rather than to produce ex post aggregate judgments.

In addition, the structure of the indicator system is designed to be open to comparative and dynamic analysis. Organizing indicator groups according to measurement logic allows the system to be flexibly applied in comparisons across administrative units, governance levels, or time periods without altering the underlying structure. This approach reflects recommendations in public policy evaluation research, which emphasize the need for indicator systems to maintain conceptual stability while remaining adaptable to different data contexts and analytical objectives [28], [46].

Importantly, the multidimensional structure does not imply a rigid causal hierarchy or evaluative ordering. Indicator groups are placed in analytical rather than linear dependency relationships. Interpretation of measurement results therefore requires cross-reading across groups, in conjunction with institutional context and specific policy objectives. This organizational logic helps avoid the tendency toward “governing by numbers,” in which quantitative indicators are used in isolation from policy context and system logic [47], [48].

To render this measurement architecture analytically explicit, the indicator system can be formalized as follows. Let  $P = \{P_1, P_2, \dots, P_5\}$  denote the set of conceptual pillars corresponding to the five stages of the public expenditure transformation chain, and let each pillar  $P_k$  contain a finite set of indicators  $I_k = \{i_{k1}, i_{k2}, \dots, i_{km(k)}\}$ , where  $m(k)$  is the number of indicators retained under pillar  $k$ . The complete system is the union  $I = I_1 \cup \dots \cup I_5$ , with  $\sum m(k) = 27$  in its final standardized form. Each indicator is associated with a measurement function  $f_{kj}$  mapping administrative and budgetary data onto an indicator

value, and the diagnostic reading of the system operates on the pillar-level profile vector  $v = (v_1, \dots, v_5)$  rather than on any scalar aggregate.

Two formal properties follow directly from the design principles set out in Section 3.1. First, non-aggregation: no composite function  $g(v)$  is defined over the pillar profile, since any such aggregation would impose implicit weights and obscure precisely the diagnostic information the system is intended to preserve [43]. Second, conditional interpretability: the informational value of any indicator  $i_{kj}$  is defined relative to the values of indicators in adjacent pillars, reflecting the sequential logic of the expenditure transformation chain rather than a set of independent measurements.

### 3.3 Description of Indicator Groups and Measurement Logic

#### *Indicator Group Reflecting the Context and Structure of Public Expenditure on Education*

The first indicator group reflects the context and structure of public expenditure on education, providing baseline information on how fiscal resources are allocated and sustained within the budget system. Indicators in this group do not aim to assess effectiveness in outcome terms, but rather to clarify the analytical starting point, including budgetary priority, resource stability, and allocation structures across levels, sectors, or target groups.

The measurement logic of this group is grounded in public finance arguments that effectiveness cannot be interpreted independently of resource allocation context and budgetary constraints [24], [26]. In policy analysis, this group addresses the question: under what fiscal conditions does observed effectiveness occur, thereby avoiding the tendency to attribute all effectiveness problems to implementation while ignoring structural constraints.

However, measurement results from this group are not used as direct evidence of effectiveness or ineffectiveness, since spending levels or expenditure structures alone do not imply effectiveness without considering operational mechanisms and resource utilization. This indicator group therefore acquires meaning only when interpreted in conjunction with subsequent analytical layers.

#### *Indicator Group Reflecting Allocation and Operational Mechanisms of Public Expenditure*

The second indicator group focuses on the operational functioning of public expenditure on education, including allocation mechanisms, budget flexibility, transparency, and adaptive capacity

during implementation. This group plays a central role in the diagnostic logic of effectiveness, as it directly reflects how fiscal resources are transformed from policy decisions into concrete managerial actions.

The measurement logic of this group aligns with public financial management approaches emphasizing that many effectiveness problems arise not from resource scarcity, but from poorly designed or inadequately operated allocation mechanisms [34]. Through this group, analysis can identify bottlenecks such as allocation rigidity, weak linkage between policy objectives and fiscal instruments, or limitations in coordination capacity across governance levels.

It is important to note that indicators in this group do not directly reflect educational outcomes, but rather provide information on the enabling conditions for expenditure effectiveness. Interpretation must therefore avoid assuming that well-functioning operational mechanisms automatically yield positive educational outcomes, and instead consider these indicators in relation to other groups within the system.

#### *Indicator Group Reflecting Direct Manifestations of Budget Utilization*

The third indicator group captures the direct manifestations of education budget utilization, allowing observation of how public expenditure is translated into concrete operational conditions of the public education system. Indicators in this group focus on aspects such as service provision capacity, coverage levels, or organizational conditions, without extending to learning quality or final educational achievements.

The measurement logic of this group is consistent with policy evaluation arguments emphasizing the need for intermediate indicators to bridge operational mechanisms and systemic changes, while avoiding conflation between organizational outputs and social outcomes [20], [29]. This group provides critical information for diagnosing organizational performance and implementation capacity among budget-using entities.

Nevertheless, indicators in this group have clear interpretive limitations. They are not used to draw conclusions about expenditure effectiveness in terms of ultimate public value, but only to reflect the immediate transformation of fiscal resources. Interpreting this group in isolation may lead to biased assessments if not considered alongside indicators of mechanisms and context.

### ***Indicator Group Reflecting Medium-Term Systemic Changes***

The fourth indicator group reflects medium-term systemic changes associated with public expenditure on education, such as improvements in system capacity, equity, or stability of service provision. Indicators in this group are designed to capture trends and system shifts rather than immediate results or long-term impacts.

The measurement logic of this group is grounded in the view that public expenditure effectiveness should be evaluated dynamically, with attention to the capacity to generate sustainable improvements over time [15] and to the slow-moving, cumulative nature of institutional change [49]. This group serves as an analytical bridge, linking direct manifestations of expenditure to broader social objectives without resorting to reductive causal inference.

It should be emphasized that indicators in this group are not used as sole evidence of policy impact, as they are influenced by numerous factors beyond public expenditure. Instead, they provide contextual signals for interpreting effectiveness in relation to other indicator groups and specific institutional conditions.

### ***Indicator Group Reflecting Long-Term Social Impact***

The fifth indicator group reflects the long-term social impacts to which public expenditure on education is expected to contribute, including human capital accumulation, equity of access and outcomes, and broader societal returns. Indicators in this group are deliberately few in number, as they describe the direction in which expenditure policy is oriented rather than results attributable to it.

The measurement logic of this group differs from that of the preceding four. Rather than measuring how resources are allocated or used, it situates the diagnosis within the societal objectives that justify public financing of education, drawing on evidence that the returns to education operate through cognitive skills and accumulate over long horizons [4], [5]. The indicators therefore function as normative reference points against which patterns observed in the other groups are read [29].

The interpretive limits of this group are correspondingly strict. Its indicators are not used to attribute social outcomes to education budgets, since such outcomes are shaped by demographic, labor market, and macroeconomic conditions lying well outside the budget cycle [20]. Their function is to keep the diagnosis anchored to public value and

to signal when expenditure policy has drifted from stated societal objectives.

### ***Combined Reading Logic and Policy-Diagnostic Value***

All indicator groups are designed to be read in combination rather than independently. The analytical value of the indicator system lies in its capacity to enable cross-group comparisons, thereby identifying patterns of effectiveness or ineffectiveness in public expenditure on education. This approach aligns with evidence-informed governance research, which emphasizes that measurement information becomes meaningful only when embedded within a policy-oriented interpretive structure [28], [39].

Accordingly, the indicator groups do not aim to produce a single aggregated conclusion regarding public expenditure effectiveness in education, but function as distinct analytical lenses that support policy diagnosis and adjustment in budget governance practice.

### **3.4 Comparison with International Expenditure Frameworks and Localization Capacity**

Over recent decades, numerous international organizations and national governments have developed indicator frameworks to assess public expenditure effectiveness in general and education spending in particular. These frameworks reflect different analytical logics, closely tied to the institutional contexts, policy objectives, and data capacities of the governance systems in which they were developed. Consequently, the comparison undertaken in this study does not seek to establish a normative hierarchy of “standard versus deviation,” but rather to clarify differences in measurement approaches, thereby positioning the theoretical contribution of the proposed indicator system.

A prominent feature of many international frameworks – particularly those developed by OECD and the World Bank – is their strong emphasis on outcomes and performance, often operationalized through output and outcome indicators that facilitate cross-country or inter-jurisdictional comparison. Such approaches are well suited to contexts with rich micro-level data, high statistical capacity, and relatively stable institutional environments [14]. Within these frameworks, intermediate layers of public expenditure are often treated as “black boxes” or considered only as auxiliary explanatory factors rather than as central measurement objects.

In contrast, the indicator system proposed in this study approaches the effectiveness of public expenditure on education through a policy transformation logic, in which intermediate layers – particularly allocation mechanisms, operational processes, and institutional capacity – occupy the central analytical role. This distinction does not arise from data constraints, but from a deliberate theoretical choice, grounded in research demonstrating that in highly decentralized and institutionally transitional contexts, expenditure effectiveness is more strongly determined by how policies are operated than by spending levels or final outcomes [10], [50]. Accordingly, rather than maximizing international comparability through standardized outcome indicators, this study prioritizes policy-diagnostic capacity within a specific context.

Another important difference concerns the role of institutions in measurement structure. Many international frameworks implicitly assume the existence of foundational institutional conditions, such as strong fiscal discipline, clear accountability mechanisms, and relatively uniform implementation capacity across units. Under such assumptions, indicators can focus primarily on measuring the degree of goal attainment. By contrast, the indicator system developed in this study treats institutions not merely as contextual background but as indirect objects of analysis, captured through indicator groups that reflect operational mechanisms and implementation capacity. This approach aligns with arguments in development political economy that differences in public sector effectiveness across countries and localities stem largely from variations in institutional capacity rather than from formal policy choices alone [37], [51].

Moreover, international indicator frameworks often rely heavily on micro-level or standardized survey data, enabling the construction of composite indices and rankings. The indicator system proposed here, by contrast, is designed to be compatible with existing administrative and budgetary data, while remaining open to integrating additional data sources when conditions permit. This difference does not reflect a trade-off driven by data limitations, but a theoretically grounded choice to ensure operational feasibility in public governance practice, particularly in contexts where micro-level data are incomplete or inconsistent across units.

From a localization perspective, this study does not approach the adaptation of international frameworks as a process of “simplification” or

“reduction,” but as a theoretically informed restructuring. Core normative principles of public expenditure effectiveness – such as results orientation, accountability, and evidence use – are retained, but implemented through a measurement architecture better aligned with Vietnam’s institutional context and budget cycle. This approach is consistent with policy transfer research emphasizing that successful adaptation requires adjustment at the level of logic and design, not merely at the level of instruments [40], [41].

In summary, compared with existing international indicator frameworks, the system proposed in this study does not seek to compete in terms of coverage or ranking capacity, but rather contributes an alternative measurement logic focused on explanatory power and policy-diagnostic effectiveness in diverse institutional contexts. This approach enables a close connection between international academic standards and the practical requirements of education budget governance in Vietnam, while also offering potential reference value for countries with similar conditions.

## **4 Research Design and Validation of the Indicator System**

### **4.1 Methodological Positioning and Validation Logic**

This study is positioned within the tradition of policy evaluation and the measurement of public expenditure effectiveness, with its central focus not on testing causal hypotheses or estimating the impact of public spending on educational outcomes, but on validating the legitimacy and operational capacity of a policy-oriented indicator system. Accordingly, the methodology adopted in this research does not belong to impact evaluation approaches, nor does it fall within the category of technical efficiency measurement based on production functions or efficiency frontiers. Instead, it is designed as a study on indicator system validation.

The primary object of validation in this study is the indicator system for assessing the effectiveness of public expenditure on education, developed on the basis of the conceptual framework and measurement architecture presented in previous sections. The validation undertaken here does not seek to answer the question of whether public expenditure improves educational outcomes, but rather to address more fundamental methodological

questions, including: (i) *whether the proposed indicator system consistently reflects the conceptualized logic through which public expenditure effectiveness is formed*; (ii) *whether the indicators within the system achieve content and structural validity*; and (iii) *whether the indicator system can be operationalized using real-world data to support policy analysis and diagnosis*.

On this basis, the study adopts a validation-oriented logic rather than a hypothesis-testing logic. This approach is consistent with research on measurement tool development in policy studies, where scientific value lies not in demonstrating specific causal relationships, but in ensuring that the measurement instrument rests on a sound theoretical foundation, exhibits internal structural coherence, and can be meaningfully applied in practical public governance contexts. Under this logic, empirical validation is understood as a systematic confirmation process, rather than a statistical test designed to reject or accept hypotheses.

A key methodological positioning of the study is the deliberate separation between indicator system validation and policy outcome evaluation. While impact evaluation studies typically use public expenditure as an explanatory variable to analyze changes in outcome indicators, this study treats public expenditure on education as an object of policy design, and the indicator system as a tool for analyzing how that object is organized and operated. Consequently, the empirical analyses conducted do not aim to infer policy success or failure, but to examine whether the proposed indicator system possesses sufficient capacity to reflect and diagnose effectiveness-related issues across different contexts.

The validation logic of the study is constructed as a multi-layered approach, reflecting the complex nature of public expenditure effectiveness. Specifically, validation is conducted across three mutually reinforcing levels. The first level is content validity, which seeks to confirm that the indicators and indicator groups adequately reflect the core dimensions of public expenditure effectiveness in education as established by the theoretical framework. The second level is structural consistency, which ensures that the indicator system avoids conceptual overlap and maintains clear analytical boundaries among measurement groups. The third level is empirical operability, which evaluates whether the indicator system can be computed stably using real-world data and whether it yields information that is meaningful for policy analysis.

## 4.2 Expert-Based Indicator Development and Validation Process

The process of indicator development and validation in this study is designed according to a theory-driven approach with structured expert validation, in which theory and the conceptual framework guide indicator formulation, while experts serve as a structured validation mechanism to confirm content validity, conceptual clarity, and the operational feasibility of the indicator system within real policy contexts. This approach combines the strengths of theoretical deduction with practitioners' experiential knowledge, while avoiding the tendency toward unstructured "consensual co-construction" that is often criticized in poorly designed Delphi studies.

Experts were selected based on functional relevance rather than academic rank or narrowly defined disciplinary affiliation. The selection process was explicitly anchored to the five conceptual pillars of the indicator system, ensuring that each pillar was reviewed by experts possessing relevant expertise and experience corresponding to that stage of the public expenditure transformation chain in education (Table 1, Appendix). A total of 62 experts participated in the study, providing broad disciplinary coverage and a high degree of stability in consensus indicators.

Experts were classified along four primary dimensions: (i) *core area of expertise*; (ii) *professional role*; (iii) *years of professional experience*; and (iv) *institutional work context*. This structure was designed to ensure institutional and role diversity, thereby reducing bias arising from purely academic or purely managerial perspectives. Importantly, linking each expert group to one or more main pillars of the indicator system enabled a focused validation process, rather than requiring all experts to assess the entire system in a diffuse manner.

The study employed a three-round Delphi process, which is appropriate for validating a multidimensional indicator system while allowing a high level of opinion convergence without imposing excessive burden on participants.

Round 1 (Exploratory validation) focused on assessing the conceptual relevance and coverage of the initial indicator set, which was deduced from the theoretical framework and measurement architecture presented earlier. In this round, experts were asked to evaluate the relevance of each indicator to its corresponding pillar and to propose revisions in wording or the consolidation of overlapping indicators. The objective of Round 1 was not aggressive elimination, but preliminary

screening to remove indicators that were ambiguous or lacked operational grounding.

Round 2 (Refinement and convergence) concentrated on filtering and convergence. Indicators revised in Round 1 were further evaluated for necessity, data feasibility, and conceptual distinctiveness. Results from Round 2 enabled the removal of indicators with low consensus or substantial conceptual overlap, while refining the system structure toward greater parsimony without compromising coverage of the five pillars.

Round 3 (Final validation) served to finalize the system structure. In this round, experts assessed the reduced indicator set in terms of overall appropriateness, consistency across pillars, and usability for policy analysis. No new indicators were introduced at this stage; the focus was on confirming the stability of evaluations and the final level of consensus before proceeding to empirical validation.

Throughout the process, experts were not asked to independently generate new indicators, but rather to validate, revise, and confirm indicators that had been theoretically derived. This design ensures that the legitimacy of the indicator system originates from the conceptual framework, with the Delphi process functioning as a mechanism for confirmation and refinement rather than a substitute for theoretical logic.

The three-round Delphi process was implemented alongside a controlled adjustment trajectory for indicator quantity, designed to avoid the twin extremes of excessive complexity and excessive simplification. At the initial stage, the number of indicators was relatively large to ensure comprehensive coverage of relevant dimensions; across subsequent rounds, indicator numbers were progressively reduced through consensus-based filtering and operability assessment. The distribution of indicators across pillars reflects the study's underlying philosophy: pillars related to allocation mechanisms and operational functioning play a central role and thus retain a larger number of indicators, while the pillar addressing long-term impacts is maintained at a moderate level with a primarily normative reference function.

All expert consultations were conducted under conditions of complete anonymity, ensuring that evaluations were not influenced by participants' academic or administrative status. Personal information was not disclosed; instead, the study reports only the aggregated structure and composition of the expert panel, ensuring

methodological transparency while adhering to research ethics standards.

### 4.3 Data Sources and Empirical Validation Strategy

In this study, data are not employed as instruments for testing policy impacts or estimating causal relationships between public expenditure and educational outcomes. Rather, data are mobilized to confirm the usability and diagnostic capacity of the indicator system developed and validated in previous sections. Accordingly, the data strategy is designed around a clear distinction between two layers of data with distinct methodological roles, corresponding to two objectives: (i) *ensuring that the indicator system can be operationalized in policy practice*; and (ii) *empirically validating the stability and analytical value of the system*.

The first data layer consists of indicator base data – that is, the data sources that the indicator system is designed to utilize when implemented in actual governance and policy analysis. These data primarily comprise administrative records and publicly available budgetary data, reflecting core aspects of the public expenditure cycle for education, including expenditure scale, structure, allocation mechanisms, and operational manifestations.

The role of this data layer is not to generate empirical evidence for a specific research hypothesis, but to confirm the computability and accessibility of each indicator within the system. The reliance on administrative data as the primary axis aligns with recommendations in public finance research, which emphasize that expenditure effectiveness indicators hold policy value only when they can be computed regularly from official data sources, rather than depending on ad hoc surveys or micro-level datasets that are difficult to sustain over time [15], [26].

This approach also allows the indicator system to maintain methodological neutrality, avoiding dependence on a specific analytical model, while enabling integration into existing budgetary and policy evaluation cycles used by public authorities.

The second data layer consists of research data, which are used directly in the article to empirically validate the operational capacity of the indicator system. Unlike indicator base data, this layer is organized and processed for analytical purposes, enabling examination of three core aspects: (i) *the feasibility of computing indicators*; (ii) *the stability of indicators across spatial or temporal comparisons*; and (iii) *the capacity of the indicator*

*system to reflect meaningful differences in policy contexts.*

The data strategy employed here is confirmatory and illustrative rather than causally inferential. Indicators are not treated as independent or dependent variables within statistical models, but as units of observation used to assess the consistency and usefulness of the measurement system. This approach is consistent with research on measurement tool development in public policy, which emphasizes that the scientific value of an indicator system lies in its capacity to provide reliable diagnostic information, rather than in optimizing model fit statistics [28], [29].

Data use in this study serves three specific validation objectives. First, data are used to assess the computability and data completeness of each indicator, enabling the removal or adjustment of indicators that cannot be stably operationalized under current data conditions. Second, data allow evaluation of indicator stability and consistency across different contexts, avoiding indicators that are excessively sensitive to short-term fluctuations or technical statistical changes. Third, data are used to test policy-discriminatory value – that is, the ability of the indicator system to reflect meaningful differences in how public expenditure on education is organized and operated, without inferring policy success or failure.

It should be emphasized that data analyses in this study do not aim to estimate the impact of public expenditure, nor to construct rankings or performance comparisons across units. Instead, data function as an empirical test of the usefulness of the indicator system for policy analysis, consistent with the methodological positioning established in Section 4.1.

The research data are selected with sufficient breadth to test operational feasibility, but not to maximize statistical coverage. Spatially, data are organized at governance levels appropriate to the education budget cycle, allowing observation of contextual and operational variation. Temporally, data are collected across multiple consecutive years to assess indicator stability and repeatability, rather than representing a single cross-sectional snapshot.

Limiting data scope in this manner enables the study to maintain its methodological focus on indicator system validation, while avoiding expansion into questions beyond the article's scope, such as evaluating education policy effectiveness at the level of final outcomes.

#### 4.4 Formal Specification of Consensus and Validity Statistics

The statistical apparatus used to validate the indicator system is specified formally as follows. Let  $n$  denote the number of indicators evaluated and  $m$  the number of experts ( $m = 62$ ). Inter-expert concordance is measured by Kendall's coefficient of concordance,  $W = 12S / [m^2(n^3 - n)]$ , where  $S$  is the sum of squared deviations of the rank totals from their mean; values of  $W$  above approximately 0.7 are conventionally interpreted as strong consensus in Delphi applications [52]. Agreement beyond chance is assessed through Fleiss' kappa,  $\kappa = (\bar{P} - \bar{P}_e) / (1 - \bar{P}_e)$ , where  $\bar{P}$  is the observed mean proportion of agreement across rating categories and  $\bar{P}_e$  its expectation under chance [53]. The intraclass correlation coefficient is estimated under a two-way random-effects specification following [54], capturing the reliability of mean expert ratings.

Content validity is quantified at two levels. For each indicator, the item-level content validity index is  $I-CVI = n_r/m$ , the proportion of experts rating the indicator as relevant, and the scale-level index is computed as the average of item-level values [55]. The content validity ratio follows [56]:  $CVR = (n_e - m/2) / (m/2)$ , where  $n_e$  is the number of experts judging an indicator essential;  $CVR$  ranges from  $-1$  to  $+1$  and equals zero when exactly half the panel deems the indicator essential. Dispersion is summarized by the coefficient of variation,  $CV = (\sigma/\mu) \times 100$ , and the stability of relative orderings across rounds by Spearman's rank correlation coefficient  $\rho$ . Reporting these statistics jointly allows the distinct dimensions of convergence, agreement, validity, and stability to be separated rather than conflated in a single index.

This specification also clarifies the mathematical boundaries of the present validation exercise: the quantities above characterize the measurement properties of the indicator system, not the efficiency of expenditure itself. Frontier-based techniques such as data envelopment analysis and stochastic frontier analysis, which estimate technical efficiency from input–output relationships [12], [21], constitute a complementary analytical layer that the standardized system is designed to feed once indicator data are compiled, as discussed further in Section 6.4.

## 5 Research Findings

### 5.1 Results of Content Validity and Structural Consistency Validation

The descriptive statistics presented in Table 2 indicate a consistent improvement in the level of convergence among expert evaluations across the Delphi rounds. The mean score increased from 3.335 in Round 1 to 3.869 in Round 2 and remained virtually unchanged in Round 3 (3.874), while the median shifted from 3.0 to 4.0 and stabilized thereafter. This pattern suggests that the convergence process was largely completed in Round 2, with Round 3 functioning primarily as a confirmatory stage rather than generating substantial additional improvement.

Table 2. Descriptive Statistics of Expert Rating across Delphi Rounds

|                       | Round 1 | Round 2 | Round 3 |
|-----------------------|---------|---------|---------|
| <b>Mean</b>           | 3.335   | 3.869   | 3.874   |
| <b>Median</b>         | 3.0     | 4.0     | 4.0     |
| <b>Std. Deviation</b> | 0.981   | 0.760   | 0.739   |
| <b>IQR</b>            | 2.0     | 1.0     | 1.0     |
| <b>CV (%)</b>         | 29.414  | 19.631  | 19.084  |

Source: created by the authors

However, although the reductions in standard deviation and interquartile range (IQR) are evident, they do not reach the very low dispersion levels commonly observed in Delphi studies focusing on psychological scales or individual competencies. The coefficient of variation (CV) declined from 29.4% to approximately 19% in the final two rounds, indicating that differences in expert assessments were narrowed but not eliminated entirely. This outcome reflects the reality that experts, despite reaching a relatively high level of agreement, continue to hold certain differences in their interpretations of public expenditure effectiveness – a phenomenon that can be considered reasonable given the inherently contested nature of policy-related concepts.

Consensus-related indicators reported in Table 3 demonstrate generally high levels of inter-expert consistency. Kendall's *W* remained around 0.80–0.85, exceeding the threshold commonly accepted as strong consensus in Delphi research (approximately 0.70). Both Kappa and ICC values exceeded 0.90, indicating very high inter-rater reliability, particularly noteworthy given the relatively large expert panel ( $n = 62$ ).

Nevertheless, an important observation is that Kendall's *W* declined slightly from Round 1 to Round 2 before stabilizing again in Round 3. This

apparent paradox suggests that the indicator refinement process in Round 2, while clarifying the structure of the system, may also have reactivated latent disagreements among expert groups as indicators became more conceptually precise. The non-monotonic trajectory of *W* across rounds indicates that consensus in this study is conditional rather than absolute.

Table 3. Consensus and Inter-Rater Reliability Indices across Delphi Rounds

|                           | Round 1 | Round 2 | Round 3 |
|---------------------------|---------|---------|---------|
| <b>Kendall's <i>W</i></b> | 0.853   | 0.797   | 0.800   |
| <b>Kappa</b>              | 0.954   | 0.913   | 0.932   |
| <b>ICC</b>                | 0.929   | 0.912   | 0.925   |

Source: created by the authors

Results related to content validity, presented in Table 4, most clearly reveal the methodological limits of the indicator system. The CVI increased from 0.482 in Round 1 to 0.656 in Round 3, indicating a marked improvement in content acceptance following the filtering process. However, the final CVI remains below the commonly recommended threshold of 0.70–0.80 in studies developing standardized measurement scales.

Most notably, the CVR value in Round 3 reached 0.312. Although this figure appears modest against the levels above 0.70 that would signal near-unanimous expert judgments, the critical value of the CVR declines with panel size and lies near 0.25 for a panel of this magnitude [56], so the final indicator set clears the size-adjusted threshold of “essentiality”. This result indicates that although many indicators were judged to be relevant, not all were considered “essential” in a strict sense by all experts. This constitutes a key paradox: the indicator system achieved overall acceptance, yet consensus regarding the non-substitutability of individual indicators remained limited.

Rather than interpreting this outcome as purely negative, the study treats the moderate CVR level as a defining characteristic of policy-oriented indicator systems, in which indicators are often complementary and diagnostic rather than minimalistic or mandatory, as is typically expected in psychological measurement scales. This distinction reflects the multidimensional and context-dependent nature of public expenditure effectiveness in education.

Table 4. Content Validity and Rank-Order Stability across Delphi Rounds

|                              | <i>Round 1</i> | <i>Round 2</i> | <i>Round 3</i> |
|------------------------------|----------------|----------------|----------------|
| <b><i>CVI</i></b>            | 0.482          | 0.641          | 0.656          |
| <b><i>CVR</i></b>            | -0.036         | 0.281          | 0.312          |
| <b><i>Spearman's Rho</i></b> | 0.932          | 0.909          | 0.918          |

Source: created by the authors

Despite content validity not reaching idealized thresholds, Spearman's Rho remained above 0.90 across all three rounds, indicating a high degree of stability in experts' relative rankings over time. This finding implies that even where disagreement persists regarding the absolute "essentiality" of certain indicators, experts nonetheless share a consistent evaluative logic concerning the relative importance of indicators and conceptual pillars.

The combination of high structural consensus and content validity that is acceptable but not optimal suggests that the indicator system occupies a position of being "sufficiently reliable for policy analysis, yet normatively open". This is precisely the analytical space targeted by the study: to provide a measurement tool that is operational and diagnostically meaningful, while remaining adaptable and open to recalibration as policy contexts evolve.

## 5.2 Indicator System for Evaluating Public Expenditure on Education

Based on the validation results regarding content legitimacy and structural consistency presented in Section 5.1, this study establishes the final standardized indicator system for evaluating public expenditure on education in Vietnam. The system comprises 27 indicators organized into five conceptual pillars that collectively reflect the full transformation chain of public expenditure from fiscal resources to long-term social impacts (as illustrated in Figure 1, Appendix). This indicator system is the outcome of a controlled refinement process implemented through three Delphi rounds, in which each adjustment step was grounded in quantitative evidence of convergence, consensus, and operational feasibility rather than subjective judgment by the research team.

Compared to the initial indicator system consisting of 49 indicators (Figure 2, Appendix), the final standardized system retains 27 indicators, representing a reduction of approximately 45% from the original set. This reduction was not applied uniformly across pillars but instead reflects the differentiated analytical roles of each stage within the public expenditure chain for education. In particular, pillars addressing fiscal allocation

mechanisms and operational functioning retain a larger number of indicators, while the pillar reflecting long-term social impacts is more strongly streamlined and assigned a primarily normative reference role.

The final distribution of indicators in the standardized system is as follows (Figure 3, Appendix):

- (1) *Resources for the Education Budget: 5 indicators*
- (2) *Fiscal Allocation and Operational Mechanisms: 7 indicators*
- (3) *Direct Manifestation of Budget Utilization: 6 indicators*
- (4) *Medium-Term Changes to the System: 5 indicators*
- (5) *Long-Term Social Impact: 4 indicators*

This structure ensures that each pillar is represented by a sufficient set of indicators to capture its analytical essence, while avoiding conceptual redundancy and unnecessary measurement burden.

The transition from the initial indicator system to the final standardized system follows three overarching principles. First is the principle of expert convergence, whereby indicators were retained only if they achieved an acceptable level of expert consensus and demonstrated stability of evaluation across Delphi rounds. Indicators with low content validity – particularly those not widely regarded as essential – were eliminated or merged with conceptually adjacent indicators.

Second is the principle of non-duplication in diagnostic function. In the initial system, certain indicators reflected closely related aspects of the same policy phenomenon (e.g., budget prioritization, budget stability, and budget protection). Through the refinement process, such indicators were consolidated into higher-level indicators capable of capturing the core phenomenon without sacrificing analytical information.

Third is the principle of data operability. Indicators requiring hard-to-access micro-level data, unstable survey-based data, or purely qualitative information were not retained in the final standardized system. Instead, priority was given to indicators that can be computed regularly using existing budgetary and administrative data, consistent with Vietnam's institutional context and data capacity.

Despite significant reduction in indicator quantity, the final standardized system preserves the conceptual integrity of the original analytical framework. Each pillar corresponds to a distinct analytical stage in the public expenditure chain for education, and indicators within each pillar are designed to perform a diagnostic function rather than to measure outputs or final impacts in a causal sense.

Importantly, the reduction in the number of indicators under the Long-Term Social Impact pillar does not imply a diminished emphasis on social outcomes. Rather, it reflects a deliberate methodological choice: indicators within this pillar serve as normative anchors guiding policy interpretation, rather than as instruments for evaluating the success or failure of public expenditure on education. This approach helps the indicator system avoid over-inference beyond available data and preserves its intended role as a policy analysis tool.

For ease of use in analysis and visualization, Table 5 (Appendix) summarizes the structure of the standardized indicator system by pillar and indicator count. From an academic standpoint, however, the core value of the system lies not in the enumerated list itself, but in the organizational logic and refinement principles that have been empirically validated and made transparent in the preceding sections.

## 6 Discussion and Policy Implications

### 6.1 Contributions of the Indicator System to Research on Public Expenditure for Education

A core scholarly contribution of this study lies in its reorientation of how the effectiveness of public expenditure on education is assessed – from scale- and outcome-based metrics toward a structural-institutional approach, in which effectiveness is understood as a multi-layered policy transformation process rather than a linear relationship between spending levels and educational outcomes. In much of the existing literature, public expenditure on education is commonly evaluated through simplified indicators such as expenditure-to-GDP ratios, per capita spending, or through models of technical efficiency and production frontiers [19], [57]. While these approaches are valuable for comparing scale or relative efficiency, they often implicitly assume that institutional structures and operational mechanisms of spending are either

homogeneous or secondary, thereby overlooking critical differences in policy design and implementation.

The indicator system proposed in this study contributes to the literature by shifting the analytical focus from “how much is spent” to “how spending is structured,” conceptualizing the effectiveness of public expenditure on education as the outcome of interactions among fiscal resources, allocation mechanisms, operational manifestations, and system capacity over time. This approach aligns with arguments in contemporary political economy and public governance, which emphasize that the core problem of public effectiveness lies not in spending levels per se, but in how spending is executed, through which mechanisms, and within what institutional contexts [25], [37]. By structuring these dimensions into an ordered indicator system, the study provides an intermediate measurement framework capable of explaining why education systems with similar spending levels may generate markedly different outcomes.

Methodologically, the study contributes by positioning the validation of an indicator system as an independent object of inquiry, rather than merely a preparatory step for causal analysis. Whereas most studies on education expenditure focus on impact evaluation – employing regression models or experimental and quasi-experimental designs to estimate causal effects – this study argues that a reliable “measurement layer” is a necessary but often neglected prerequisite in the policy research chain. The development and validation of an indicator system with content legitimacy, structural consistency, and operational feasibility therefore do not compete with impact evaluation, but instead complement it by addressing a missing analytical layer in the literature [28], [29].

This approach is directly connected to research on policy diagnostics and performance-informed budgeting, where the primary objective is not to declare policies successful or unsuccessful, but to identify bottlenecks within the policy design–implementation–adjustment chain [9], [26]. The indicator system developed in this study is explicitly designed to serve this diagnostic function, enabling analysts to distinguish problems arising at the level of resources, allocation mechanisms, or operational execution, rather than attributing all educational shortcomings to insufficient spending.

The fact that content validity indicators such as CVR do not reach the ideal thresholds typically observed in psychological scale development,

while structural consensus remains high, reflects the nature of policy concepts: they are contested, context-dependent, and cannot be reduced to a single set of universally “essential” indicators [58], [59]. This interpretation helps avoid the mechanical application of micro-level measurement standards to macro-level analytical tools and reinforces the argument that the value of the indicator system lies in its capacity to open a structured space for policy dialogue, rather than in achieving absolute consensus.

Overall, the contribution of the proposed indicator system does not lie in offering an “optimal” or “universal” metric, but in establishing a shared measurement language for research and analysis of public expenditure on education – bridging public finance, institutional governance, and education policy. In doing so, the study enriches the international literature by shifting attention from reductive outcome evaluations toward structural diagnostic approaches better suited to diverse and transitioning institutional contexts such as Vietnam.

## 6.2 Policy Diagnostic Value of the Indicator System

Unlike output- or impact-based evaluation approaches, which tend to reduce effectiveness to a single composite index, the indicator system provides a multi-layered analytical map in which each pillar performs a distinct diagnostic function while remaining interpretable within a systemic relationship.

First, the indicator system enables a clear distinction between resource-related issues and mechanism-related issues – an ambiguity that frequently characterizes policy debates. In many contexts, stagnation or decline in educational outcomes is often attributed to “insufficient funding,” even when fiscal analyses reveal relatively stable expenditure levels. By placing resource indicators alongside allocation and operational mechanism indicators, the system allows analysts to determine whether the problem lies in fiscal commitment, allocation design, or governance capacity, rather than collapsing all explanations into a single cause [34], [60].

Second, the system supports pillar-based diagnostics rather than unilinear aggregate assessments. This approach is consistent with research on diagnostic governance, which emphasizes that effective policy reform requires identifying the correct “intervention layer,” thereby avoiding diffuse or misdirected solutions [10], [50]. In the context of public expenditure on education,

this distinction is particularly important because reforms to allocation mechanisms, financial governance, or implementation structures often exhibit different trajectories and time lags that cannot be captured by a single indicator.

Third, the indicator system is designed to differentiate between short-term operational manifestations and medium-term systemic changes, thereby preventing confusion between “implementation glitches” and “structural problems.” Indicators under the Direct Manifestation of Budget Utilization pillar reflect how budgets are translated into operational conditions in the short term, while indicators under Medium-Term Changes to the System capture cumulative adjustments in governance capacity and allocation effectiveness. This separation helps policymakers avoid overreacting to short-term fluctuations while correctly identifying longer-term structural trends requiring strategic intervention [15].

Fourth, the inclusion of the Long-Term Social Impact pillar as a normative reference – rather than as an outcome variable – creates a guiding but non-prescriptive interpretive framework. Indicators in this pillar are not used to declare policy success or failure, but to situate diagnoses related to resources, mechanisms, and operations within broader societal objectives such as equity, human capital development, and sustainability. This approach aligns with contemporary policy evaluation literature emphasizing the need to separate operational diagnostics from normative judgments, thereby avoiding conclusions that exceed available evidence [61], [62].

Finally, the diagnostic value of the indicator system also lies in its capacity to facilitate structured policy dialogue among stakeholders. By providing a validated and standardized set of indicators, the system establishes a shared analytical language for financial managers, education administrators, and policy researchers. Rather than debating effectiveness at an abstract level, stakeholders can focus discussions on specific pillars and indicators, thereby improving decision quality and reducing interpretive conflict [63], [64].

Taken together, the proposed indicator system enhances policy diagnostic capacity by clarifying problem structures, correctly positioning intervention layers, and expanding evidence-based dialogue – without falling into the traps of causal over-inference or reductive performance assessment. This constitutes its substantive contribution to research and practice in governing

public expenditure on education within complex and transitioning institutional contexts.

### **6.3 Policy Implications for Governing Public Expenditure on Education in Vietnam**

From a policy governance perspective, the proposed indicator system is not intended to replace existing budgetary instruments or oversight mechanisms, but to provide an additional analytical layer that helps public authorities correctly identify the nature of problems within the public expenditure chain for education. Policy implications derived from this study therefore focus on how the indicator system can be used as a governance tool, rather than as a mechanical performance metric or a basis for automatic budget allocation.

#### ***From Monitoring Spending Levels to Supervising Spending Structure and Operations***

The most significant policy implication is the need to shift governance focus from monitoring expenditure scale to supervising expenditure structure and operational functioning. In Vietnam's current practice, budget reports and expenditure oversight in education continue to emphasize spending shares, growth rates, or per capita expenditure, while paying limited attention to allocation design, coordination, and implementation mechanisms. The proposed indicator system complements this practice by clearly distinguishing among resource issues, allocation mechanisms, and operational manifestations, thereby helping authorities avoid simplistic policy responses – such as across-the-board spending increases – when the core problem lies in allocation or governance.

This approach aligns with budget governance reforms observed in many middle-income countries, where recent shifts have moved from “spending expansion” toward “spending quality and governance” [65]. Under conditions of fiscal constraint, improving spending quality through better mechanisms and operational capacity often yields higher marginal returns than expanding expenditure scale.

#### ***Supporting Decentralization Reform and Strengthening Local Governance Capacity***

Another important implication concerns decentralized governance of education expenditure, a defining feature of Vietnam's public finance system. While strong decentralization enables flexibility and responsiveness to local needs, it also

amplifies disparities in governance capacity and expenditure effectiveness across localities. The multi-pillar indicator system allows these differences to be identified without relying on direct ranking, thereby supporting the design of targeted capacity-building interventions.

Experiences from countries with similar decentralization structures, such as Indonesia and the Philippines, demonstrate that diagnostic tools focusing on expenditure structure and operational capacity play a crucial role in enabling “conditional decentralization,” where autonomy is linked to accountability and technical support [66], [67]. In this context, the indicator system proposed in this study can serve as a foundational tool for identifying which localities require governance support, rather than applying uniform administrative measures.

#### ***Strengthening the Link Between Budgeting, Policy Implementation, and Oversight***

The indicator system also offers important implications for strengthening the linkage between budgeting, policy implementation, and oversight – an area that remains underdeveloped in Vietnam's education expenditure governance. The absence of intermediate measurement tools often confines oversight to ex post audits or inspections, limiting opportunities for timely policy adjustment during implementation.

International experience with performance-informed budgeting suggests that diagnostic indicators are most effective when used not for direct reward–punishment mechanisms, but for creating effective policy feedback loops across budget cycle stages [39], [68]. In Vietnam's context, the indicator system could be integrated into sectoral budget analysis reports for education, thereby facilitating structured policy dialogue among financial authorities, education agencies, and subnational governments.

#### ***Expanding Institutionally Adjusted International Comparisons***

Finally, the proposed indicator system opens opportunities for institutionally adjusted international comparison, particularly with countries at similar development stages and with comparable governance structures. Rather than directly comparing educational outcomes or spending levels – both of which are heavily influenced by socioeconomic context – the system enables comparison of expenditure structures and operational logic, an approach increasingly

recommended in comparative education policy research [69].

For Vietnam, this approach situates challenges in governing education expenditure within a broader comparative space, where issues of decentralization, institutional capacity, and accountability are recognized as shared challenges among transitioning economies rather than as country-specific deficiencies. This perspective not only enhances the international relevance of the study but also positions Vietnam as a reference case with broader analytical value in the literature on public expenditure governance in education.

#### **6.4 Scope of Use and Limitations of the Indicator System**

Clearly defining appropriate scope of use and inherent limitations is essential to ensuring that the indicator system is applied cautiously and in alignment with policy objectives. Articulating these limits does not weaken the study's contribution; rather, it precisely positions the academic and practical role of the indicator system within the broader research and policy landscape.

##### ***Appropriate Scope of Application***

First, the indicator system is designed as a system-level policy diagnostic tool, suitable for analyzing expenditure structures, reviewing allocation mechanisms, and assessing operational capacity in public expenditure on education. It is particularly useful in contexts requiring identification of institutional bottlenecks or comparison of governance structures across time periods, sectors, or administrative units – without necessitating causal inference or final impact measurement [17].

Second, the system is well suited to complement existing budget governance instruments, such as sectoral budget analysis reports, public expenditure reviews, or results-based monitoring mechanisms. In this role, it does not replace financial audits or impact evaluations, but provides an intermediate analytical layer that helps link fiscal information to policy adjustment decisions [26].

Third, in international contexts, the indicator system may serve as a reference framework for institutionally adjusted comparison, particularly among middle-income countries with similar decentralization structures. Using the system in this way helps avoid reductive comparisons based solely on spending levels or isolated education outcomes, which often ignore institutional and implementation differences [50].

##### ***Inherent Limitations of the Indicator System***

Alongside its appropriate applications, the indicator system has inherent limitations that must be explicitly acknowledged. First, it is not designed to measure the causal impact of public expenditure on education outcomes or socioeconomic indicators. Using the system to infer policy effectiveness in an impact sense would exceed its methodological capacity and risk generating misleading conclusions [70].

Second, the indicator system is not intended for ranking or performance comparison among localities or educational units. Although some indicators may be computed at subnational levels, using them for ranking would disregard contextual differences in resources, mandates, and policy objectives, and would contradict the system's diagnostic orientation [71].

Third, the operational value of the system depends to some extent on the quality and consistency of administrative and budgetary data. In contexts where data are fragmented or classification standards change over time, application of the system should be accompanied by appropriate data validation and harmonization procedures. This limitation reflects a general characteristic of policy analysis tools based on administrative data rather than a specific weakness of the proposed indicator system [72].

##### ***Significance of Defining Use Boundaries***

Explicitly defining the scope and limitations of the indicator system carries important scholarly significance. It prevents inferential overreach and ensures that the system is understood and applied as a tool for structural analysis and policy diagnosis rather than as a metric of success or failure in public expenditure on education. This approach is consistent with recommendations in contemporary policy evaluation literature, which emphasize that the value of measurement tools lies in their fitness for purpose, not in their ability to answer all policy questions [61].

Finally, the present validation opens a clearly defined quantitative research agenda. Once the standardized indicators are compiled across provinces and years, the resulting dataset lends itself to formal efficiency modeling: data envelopment analysis and stochastic frontier analysis can estimate the technical efficiency of education spending conditional on the mechanism and context pillars [12], [57]; constrained optimization can be applied to the problem of allocating resources across pillars under fiscal ceilings; and machine learning methods can exploit

large administrative datasets to detect nonlinear associations between indicator profiles and medium-term systemic changes. These extensions would transform the diagnostic architecture developed here into a fully quantitative decision-support system and constitute the immediate next stage of this research program.

In conclusion, by clearly delineating its scope and limitations, this study reinforces the argument that the proposed indicator system does not seek to replace existing evaluation approaches, but instead complements them by providing a missing analytical perspective in research and practice on governing public expenditure for education.

#### References:

- [1] Musgrave, R. A., & Musgrave, P. B. (1989). *Public finance in theory and practice* (5th ed.). McGraw-Hill. ISBN: 978-0-07-044127-9.
- [2] Robinson, M. (2013). *Program classification for performance-based budgeting: How to structure budgets to enable the use of evidence*. World Bank. ISBN: 978-1-60244-238-2. <https://hdl.handle.net/10986/21767>.
- [3] Stiglitz, J. E., & Rosengard, J. K. (2015). *Economics of the public sector* (4th ed.). W. W. Norton. ISBN: 978-0-393-92522-7.
- [4] Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668. <https://doi.org/10.1257/jel.46.3.607>.
- [5] Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics*, 26(5), 445–458. <https://doi.org/10.1080/09645292.2018.1484426>.
- [6] UNESCO Institute for Statistics. (2024). *SDG 4 data: Education expenditure indicators*. UNESCO, [Online]. <https://databrowser.uis.unesco.org/> (Accessed Date: January 27, 2026).
- [7] World Bank. (2023). *Government expenditure on education (% of GDP): Vietnam*. World Bank Open Data, [Online]. <https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?locations=VN> (Accessed Date: January 27, 2026).
- [8] National Assembly of Vietnam. (2019). *Law on Education* (Law No. 43/2019/QH14), [Online]. <https://datafiles.chinhphu.vn/cpp/files/vbpq/2019/07/43.signed.pdf> (Accessed Date: January 27, 2026).
- [9] World Bank. (2017). *Vietnam public expenditure review: Fiscal policies towards sustainability, efficiency, and equity*. World Bank. <https://hdl.handle.net/10986/28610>.
- [10] Andrews, M., Pritchett, L., & Woolcock, M. (2017). *Building state capability: Evidence, analysis, action*. Oxford University Press. ISBN: 978-0-19-874748-2. <https://doi.org/10.1093/acprof:oso/9780198747482.001.0001>.
- [11] Ministry of Finance of Vietnam. (2024). *State budget final accounts report*. Ministry of Finance, [Online]. <https://mof.gov.vn/bo-tai-chinh-en/state-budget/statebudgetfinal-account-settlement-for-the-fiscal-year-2023> (Accessed Date: January 27, 2026).
- [12] Afonso, A., Schuknecht, L., & Tanzi, V. (2010). Public sector efficiency: Evidence for new EU member states and emerging markets. *Applied Economics*, 42(17), 2147–2164. <https://doi.org/10.1080/00036840701765460>.
- [13] Hanushek, E. A., & Woessmann, L. (2010). Education and economic growth. In E. Baker, B. McGaw, & P. Peterson (Eds.), *International encyclopedia of education* (Vol. 2, pp. 245–252). Elsevier. ISBN: 978-0-08-044894-7. <https://doi.org/10.1016/B978-0-08-044894-7.01227-6>.
- [14] OECD. (2018). *Education at a glance 2018: OECD indicators*. OECD Publishing. ISBN: 978-92-64-30338-6. <https://doi.org/10.1787/eag-2018-en>.
- [15] OECD. (2019). *OECD good practices for performance budgeting*. OECD Publishing. ISBN: 978-92-64-11199-8. <https://doi.org/10.1787/c90b0305-en>.
- [16] Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3–19. <https://doi.org/10.1111/j.1467-9299.1991.tb00779.x>.
- [17] Pollitt, C., & Bouckaert, G. (2017). *Public management reform: A comparative analysis* (4th ed.). Oxford University Press. ISBN: 978-0-19-879518-6.
- [18] Tanzi, V. (2011). *Government versus markets: The changing economic role of the state*. Cambridge University Press. ISBN: 978-1-107-09653-0.
- [19] Hanushek, E. A., & Woessmann, L. (2011). How much do educational outcomes matter in OECD countries? *Economic Policy*,

- 26(67), 427–491.  
<https://doi.org/10.1111/j.1468-0327.2011.00265.x>.
- [20] Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach* (8th ed.). SAGE. ISBN: 978-1-5063-0788-6.
- [21] Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics*, 9(3), 245–268.  
<https://doi.org/10.1080/09645290110086126>.
- [22] Wildavsky, A. (1986). *Budgeting: A comparative theory of budgetary processes* (2nd ed.). Transaction Books. ISBN: 978-0-88738-604-6.
- [23] Schick, A. (1998). Why most developing countries should not try New Zealand reforms. *World Bank Research Observer*, 13(1), 123–131.  
<https://doi.org/10.1093/wbro/13.1.123>.
- [24] Allen, R., Hemming, R., & Potter, B. H. (Eds.). (2013). *The international handbook of public financial management*. Palgrave Macmillan. ISBN: 978-0-230-30024-8.  
<https://doi.org/10.1057/9781137315304>.
- [25] Andrews, M. (2013). *The limits of institutional reform in development: Changing rules for realistic solutions*. Cambridge University Press. ISBN: 978-1-107-01633-0.  
<https://doi.org/10.1017/CBO9781139060974>.
- [26] Robinson, M. (2014). Spending reviews. *OECD Journal on Budgeting*, 13(2), 81–122.  
<https://doi.org/10.1787/budget-13-5jz14bz8p2hd>.
- [27] Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174–184.  
<https://doi.org/10.1111/0033-3352.00168>.
- [28] Pollitt, C. (2013). The logics of performance management. *Evaluation*, 19(4), 346–363.  
<https://doi.org/10.1177/1356389013505040>.
- [29] Mayne, J. (2015). Useful theory of change models. *Canadian Journal of Program Evaluation*, 30(2), 119–142.  
<https://doi.org/10.3138/cjpe.230>.
- [30] Margetts, H., Perri 6, & Hood, C. (Eds.). (2010). *Paradoxes of modernization: Unintended consequences of public policy reform*. Oxford University Press. ISBN: 978-0-19-957354-7.
- <https://doi.org/10.1093/acprof:oso/9780199573547.001.0001>.
- [31] Howlett, M. (2019). *Designing public policies: Principles and instruments* (2nd ed.). Routledge. ISBN: 978-1-138-29364-9.
- [32] Peters, B. G. (2015). *Advanced introduction to public policy*. Edward Elgar. ISBN: 978-1-78195-577-2.
- [33] Sabatier, P. A., & Weible, C. M. (Eds.). (2014). *Theories of the policy process* (3rd ed.). Westview Press. ISBN: 978-0-8133-4926-8.
- [34] Schick, A. (2014). The metamorphoses of performance budgeting. *OECD Journal on Budgeting*, 13(2), 49–79.  
<https://doi.org/10.1787/budget-13-5jz2jw9szgs8>.
- [35] Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, 63(5), 586–606.  
<https://doi.org/10.1111/1540-6210.00322>.
- [36] Kelman, S., & Friedman, J. N. (2009). Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service. *Journal of Public Administration Research and Theory*, 19(4), 917–946.  
<https://doi.org/10.1093/jopart/mun028>.
- [37] Fukuyama, F. (2013). What is governance? *Governance*, 26(3), 347–368.  
<https://doi.org/10.1111/gove.12035>.
- [38] North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press. ISBN: 978-0-521-39734-6.  
<https://doi.org/10.1017/CBO9780511808678>.
- [39] Moynihan, D. P., & Beazley, I. (2016). *Toward next-generation performance budgeting: Lessons from the experiences of seven reforming countries*. World Bank. ISBN: 978-1-4648-0954-5.  
<https://doi.org/10.1596/978-1-4648-0954-5>.
- [40] Dolowitz, D. P., & Marsh, D. (2000). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance*, 13(1), 5–24.  
<https://doi.org/10.1111/0952-1895.00121>.
- [41] Peck, J., & Theodore, N. (2015). *Fast policy: Experimental statecraft at the thresholds of neoliberalism*. University of Minnesota Press. ISBN: 978-0-8166-7731-3.

- <https://doi.org/10.5749/minnesota/9780816677306.001.0001>.
- [42] Bouckaert, G., & Halligan, J. (2008). *Managing performance: International comparisons*. Routledge. ISBN: 978-0-415-42394-6.
- [43] Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1), 65–77. <https://doi.org/10.1007/s11205-006-0024-9>
- [44] Kuhlmann, S., & Bouckaert, G. (Eds.). (2016). *Local public sector reforms in times of crisis: National trajectories and international comparisons*. Palgrave Macmillan. ISBN: 978-1-137-52547-5. <https://doi.org/10.1057/978-1-137-52548-2>.
- [45] Moynihan, D. P., & Pandey, S. K. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory*, 20(4), 849–866. <https://doi.org/10.1093/jopart/muq004>
- [46] Hatry, H. P. (2006). *Performance measurement: Getting results* (2nd ed.). Urban Institute Press. ISBN: 978-0-87766-734-6.
- [47] Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517–538. <https://doi.org/10.1111/j.1467-9299.2006.00600.x>.
- [48] Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press. ISBN: 978-0-19-828947-0.
- [49] Pierson, P. (2004). *Politics in time: History, institutions, and social analysis*. Princeton University Press. ISBN: 978-0-691-11715-7. <https://doi.org/10.1515/9781400841080>.
- [50] Pritchett, L., Woolcock, M., & Andrews, M. (2013). Looking like a state: Techniques of persistent failure in state capability for implementation. *Journal of Development Studies*, 49(1), 1–18. <https://doi.org/10.1080/00220388.2012.709614>.
- [51] North, D. C., Wallis, J. J., & Weingast, B. R. (2009). *Violence and social orders: A conceptual framework for interpreting recorded human history*. Cambridge University Press. ISBN: 978-0-521-76173-4. <https://doi.org/10.1017/CBO9780511575839>.
- [52] Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3), 763–774. <https://doi.org/10.1111/j.1540-5915.1997.tb01330.x>.
- [53] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. <https://doi.org/10.1037/h0031619>.
- [54] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>.
- [55] Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>.
- [56] Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>.
- [57] De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339–363. <https://doi.org/10.1057/jors.2015.92>.
- [58] Bovens, M., Goodin, R. E., & Schillemans, T. (Eds.). (2014). *The Oxford handbook of public accountability*. Oxford University Press. ISBN: 978-0-19-964125-3. <https://doi.org/10.1093/oxfordhb/9780199641253.001.0001>.
- [59] Peters, B. G. (2019). *Institutional theory in political science: The new institutionalism* (4th ed.). Edward Elgar. ISBN: 978-1-78643-794-5.
- [60] Hood, C., & Dixon, R. (2015). *A government that worked better and cost less? Evaluating three decades of reform and change in UK central government*. Oxford University Press. ISBN: 978-0-19-968702-2. <https://doi.org/10.1093/acprof:oso/9780199687022.001.0001>.
- [61] Vedung, E. (2017). *Public policy and program evaluation*. Routledge. (Original work published 1997). ISBN: 978-0-7658-0687-1. <https://doi.org/10.4324/9781315127767>.
- [62] Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Prentice Hall. ISBN: 978-0-13-309725-2.
- [63] Ansell, C., & Gash, A. (2018). Collaborative platforms as a governance strategy. *Journal*

of Public Administration Research and Theory, 28(1), 16–32.

<https://doi.org/10.1093/jopart/mux030>.

- [64] Howlett, M., Ramesh, M., & Perl, A. (2020). *Studying public policy: Principles and processes* (4th ed.). Oxford University Press. ISBN: 978-0-19-902614-2.
- [65] OECD. (2019). *Government at a Glance 2019*. OECD Publishing. ISBN: 978-92-64-35156-1. <https://doi.org/10.1787/8ccf5c38-en>.
- [66] Lewis, B. D., & Smoke, P. (2017). Intergovernmental fiscal transfers and local incentives and responses: The case of Indonesia. *Fiscal Studies*, 38(1), 111–139. <https://doi.org/10.1111/1475-5890.12080>.
- [67] Martinez-Vazquez, J., Lago-Peñas, S., & Sacchi, A. (2017). The impact of fiscal decentralization: A survey. *Journal of Economic Surveys*, 31(4), 1095–1129. <https://doi.org/10.1111/joes.12182>.
- [68] Robinson, M., & Last, D. (2009). *A basic model of performance-based budgeting* (IMF Technical Notes and Manuals No. 09/01). International Monetary Fund. ISBN: 978-1-4623-2027-1. <https://doi.org/10.5089/9781462320271.005>.
- [69] Bruns, B., Filmer, D., & Patrinos, H. A. (2011). *Making schools work: New evidence on accountability reforms*. World Bank. ISBN: 978-0-8213-8679-8. <https://doi.org/10.1596/978-0-8213-8679-8>.
- [70] Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press. ISBN: 978-0-19-984162-2.
- [71] Hood, C. (2007). Public service management by numbers: Why does it vary? Where has it come from? What are the gaps and the puzzles? *Public Money & Management*, 27(2), 95–102. <https://doi.org/10.1111/j.1467-9302.2007.00564.x>.
- [72] Radnor, Z., & McGuire, M. (2004). Performance management in the public sector: Fact or fiction? *International Journal of Productivity and Performance Management*, 53(3), 245–260. <https://doi.org/10.1108/17410400410523783>

### **Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)**

Thi Doan Trang Ha conceptualized the study, developed the theoretical framework and the five-pillar measurement architecture, and drafted Sections 1, 2 and 3.

Thu Huyen Pham designed the Delphi protocol, organized and administered the three-round expert consultation, and carried out the statistical analysis of consensus, reliability, and content validity reported in Sections 4 and 5.

Thu Trang Pham conducted the review of international indicator frameworks, interpreted the findings in relation to the policy implications discussed in Section 6, and was responsible for the final editing of the manuscript.

All authors have read and approved the final version of the manuscript.

### **Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself**

The authors received no specific funding, grant or financial support from any public, commercial or not-for-profit agency for the research presented in this article or for the article itself.

### **Conflict of Interest**

The authors have no conflict of interest to declare that is relevant to the content of this article.

### **Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)**

This article is published under the terms of the Creative Commons Attribution License 4.0

[https://creativecommons.org/licenses/by/4.0/deed.en\\_US](https://creativecommons.org/licenses/by/4.0/deed.en_US)

## APPENDIX

Table 1. Profile of Expert Panel Participating in the Delphi Study (n = 62)

| <b>Criteria</b>                          | <b>Classification</b>                      | <b>Quantity</b> | <b>Ratio (%)</b> | <b>Related Main Pillars</b>                 |
|------------------------------------------|--------------------------------------------|-----------------|------------------|---------------------------------------------|
| <b>Main Area of Expertise</b>            | Public Finance & Budgeting                 | 18              | 29.0             | Resources; Allocation                       |
|                                          | Public Policy & Public Administration      | 16              | 25.8             | Allocation; Medium-Term Changes             |
|                                          | Education Economics & Education Management | 17              | 27.4             | Direct Manifestation; Medium-Term Changes   |
|                                          | Policy Evaluation & Development Studies    | 11              | 17.7             | Long-Term Impact                            |
| <b>Occupational Role</b>                 | Academic Researcher                        | 24              | 38.7             | All Pillars                                 |
|                                          | Government Official / Public Manager       | 19              | 30.6             | Resources; Allocation; Direct Manifestation |
|                                          | Policy Advisor / Consultant                | 12              | 19.4             | Allocation; Medium-Term Changes             |
|                                          | Senior Education Manager                   | 7               | 11.3             | Direct Manifestation                        |
| <b>Professional Experience</b>           | 10 – 14 years                              | 21              | 33.9             | All Pillars                                 |
|                                          | 15 – 19 years                              | 24              | 38.7             | All Pillars                                 |
|                                          | ≥ 20 years                                 | 17              | 27.4             | All Pillars                                 |
| <b>Work Area</b>                         | Central Level Institutions                 | 26              | 41.9             | Resources; Allocation                       |
|                                          | Provincial/Local Level Institutions        | 17              | 27.4             | Allocation; Direct Manifestation            |
|                                          | Universities & Research Institutes         | 19              | 30.6             | Medium-Term; Long-Term                      |
|                                          |                                            |                 |                  |                                             |
| <b>Level of Participation by Pillar*</b> | Resources for the Education Budget         | 26              | 41.9             | —                                           |
|                                          | Fiscal Allocation & Operational Mechanisms | 34              | 54.8             | —                                           |
|                                          | Direct Manifestation of Budget Utilization | 29              | 46.8             | —                                           |
|                                          | Medium-Term Changes to the System          | 31              | 50.0             | —                                           |
|                                          | Long-Term Social Impact                    | 22              | 35.5             | —                                           |

Note (\*): One expert may contribute to more than one pillar; therefore, percentages in this category do not sum to 100%.

Source: created by the authors

Table 5. Structure and Diagnostic Orientation of the Standardized Indicator System

| <b>Pillar</b>                                       | <b>Analytical role</b> | <b>No. of indicators</b> | <b>Core diagnostic focus</b>                                 | <b>Data basis</b>                |
|-----------------------------------------------------|------------------------|--------------------------|--------------------------------------------------------------|----------------------------------|
| <b>Resources for the Education Budget</b>           | Fiscal input           | 5                        | Priority, stability, and predictability of education funding | State budget, MOF                |
| <b>Fiscal Allocation and Operational Mechanisms</b> | Institutional process  | 7                        | Transparency, coordination, and financial governance         | Budget allocation, audit reports |
| <b>Direct Manifestation of Budget Utilization</b>   | Operational execution  | 6                        | Translation of budgets into operational conditions           | Sectoral administrative data     |
| <b>Medium-Term Changes to the System</b>            | System adjustment      | 5                        | Governance capacity and allocative efficiency over time      | Time-series sectoral data        |
| <b>Long-Term Social Impact</b>                      | Normative reference    | 4                        | Equity, human capital, and societal returns                  | National statistics (contextual) |

Source: created by the authors

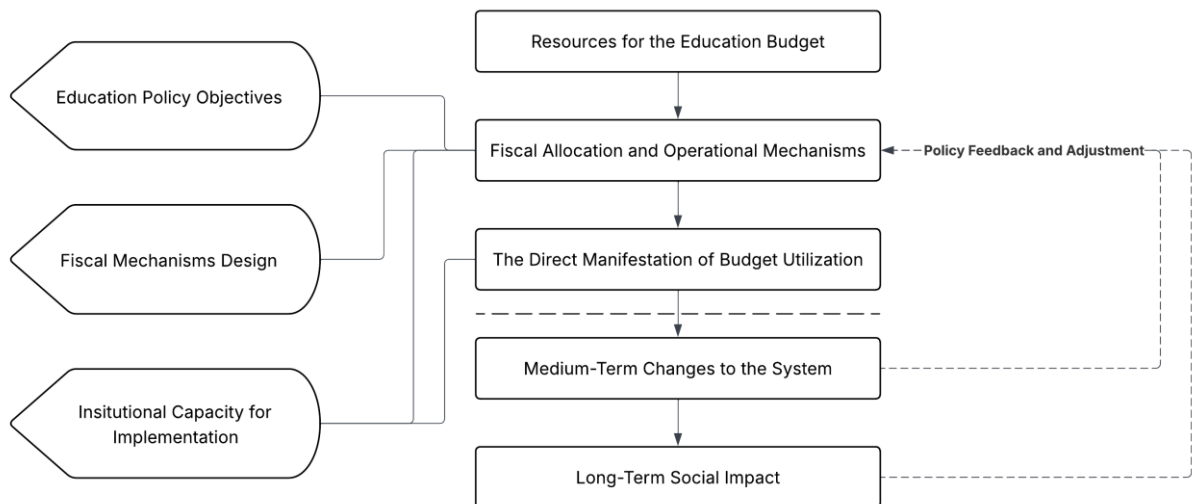


Fig. 1: Conceptual Framework for Assessing the Effectiveness of Public Expenditure on Education  
Source: created by the authors

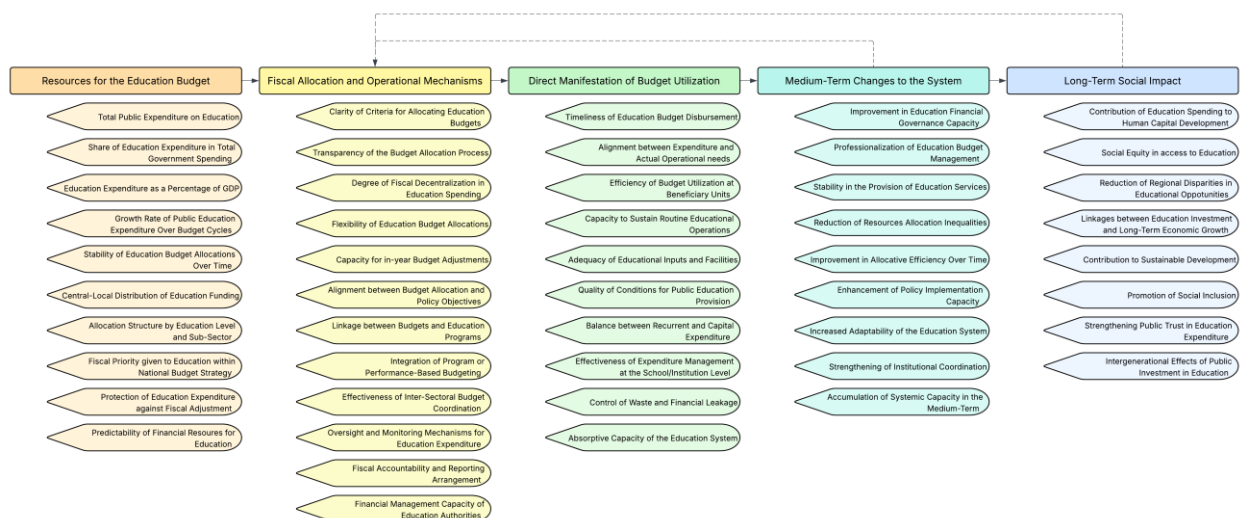


Fig. 2: Initial Indicator System for Assessing the Effectiveness of Public Expenditure on Education  
Source: created by the authors

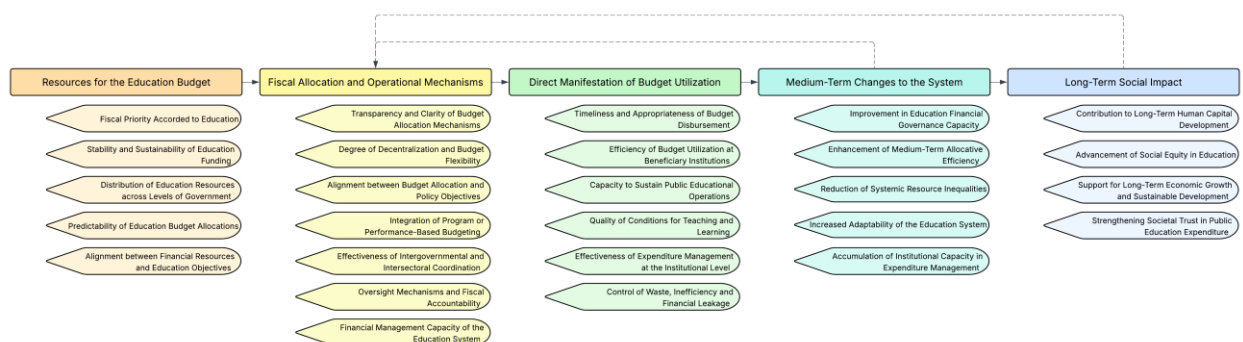


Fig. 3: Final Standardized Indicator System for Evaluating the Effectiveness of Public Expenditure on Education  
Source: created by the authors