

Article

Toward a Refined Framework for Qualitative VfM Assessment in Chinese PPPs: A Comparative Approach

Chuan Chen ^{1,*}  and Lin Huang ²¹ Business School, Sichuan University, Chengdu 610064, China² Roca Infrastructure Data and Analytics Co., Ltd., Chengdu 610021, China; huanglin@rocapp.com

* Correspondence: chenchuan@scu.edu.cn

Abstract

China's extensive experience with Public–Private Partnerships (PPPs) over the past decade offers valuable lessons for international infrastructure governance. Central to the PPP concept is the imperative to deliver Value for Money (VfM), and VfM assessment serves as a critical tool to determine whether the PPP modality is appropriate for a given project. While China's Ministry of Finance (MoF) introduced a qualitative VfM assessment framework in 2015, its conceptual ambiguity, indicator overlap, and procedural inconsistencies have limited its effectiveness in guiding PPP decisions. This paper critically examines the existing Chinese qualitative VfM framework through comparative analysis with two international approaches: the UK HM Treasury's three-dimensional VfM approach and the UN ESCAP's developmental framework for Asia–Pacific countries. Through pair-wise and structural comparison, the study identifies key gaps in the Chinese model, including the absence of a preliminary applicability screening stage and insufficiently specified treatment of output clarity, scope stability, interface coordination, private-sector capability, revenue sustainability, and asset-life/contract-term alignment. Drawing on international insights, a refined VfM framework is proposed, featuring a pre-screening mechanism, a refined indicator system, and enhanced scoring methodology. The framework is applied to a complex, large-scale PPP case in China through a retrospective reassessment to illustrate its operability and diagnostic usefulness. Results show that the refined approach preserves the original positive VfM judgement while improving analytical clarity, procedural transparency, and the traceability of expert scoring. The paper concludes by discussing how a more structured qualitative VfM procedure can contribute to evaluation governance in PPP decision-making, while also identifying the need for further testing across additional PPP cases in China and beyond.

Keywords: public–private partnerships (PPPs); value for money (VfM); China; qualitative assessment; comparative analysis; infrastructure



Academic Editor: Ali Keyvanfar

Received: 25 May 2026

Revised: 10 July 2026

Accepted: 14 July 2026

Published: 15 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Public–Private Partnerships (PPPs) have become a defining feature of contemporary infrastructure governance, widely adopted across both advanced and emerging economies as a means to bridge infrastructure gaps, enhance service delivery, and foster innovation [1]. At the heart of PPP practice lies a dual imperative: effective risk transfer and Value for Money (VfM). These were succinctly articulated by UK Chancellor Kenneth Clarke in 1994 when he reaffirmed the foundational principles of the Private Finance Initiative (PFI):

private capital must bear genuine risk, and public funds must not be expended unless VfM can be credibly demonstrated [2].

VfM, however, should not be understood as a static accounting metric or a narrow search for the lowest-cost procurement option. As defined by HM Treasury [3] (p. 7), VfM refers to “the optimum combined whole-life costing and quality (or fitness of purpose) to meet the user’s requirement.” This definition positions VfM as a governance judgement about whether a procurement and delivery arrangement can meet public objectives over the project life cycle. It therefore implicates not only efficiency, but also risk allocation, service quality, institutional accountability, and the public interest in infrastructure delivery [4–6].

Yet, as the institutionalization of VfM has progressed, so too has its contestation. Numerous critiques highlight the methodological fragility and political plasticity of prevailing approaches: quantitative assessments are often overly reliant on stylized assumptions, susceptible to strategic manipulation, and blind to distributional and environmental consequences [7–10]. Broader critiques further warn that, when assumptions and judgement are insufficiently disclosed, VfM may become proceduralized as a legitimacy device rather than functioning as substantive scrutiny [11,12]. These concerns do not negate the importance of VfM assessment, but they highlight the need to make its judgement structure more explicit and reviewable.

Within this contested terrain, qualitative VfM analysis plays an important but still undertheorized role. Unlike its quantitative counterpart, which benchmarks PPPs against a public sector comparator (PSC), qualitative VfM focuses on strategic fit, institutional capacity, market maturity, and stakeholder alignment—factors that resist monetization yet decisively shape project viability [13,14]. Often conducted at an early stage of project development, qualitative assessments function as a form of institutional sense-checking [3,15]. In this paper, it is further conceptualized as a staged evaluation-governance mechanism: it can screen whether a project has the basic conditions for PPP delivery, diagnose project suitability across multiple qualitative criteria, and make expert judgement, scoring logic, and uncertainty more transparent.

China provides a particularly important setting for examining this issue. Since the Ministry of Finance issued the *Guideline on Value for Money Evaluation (Trial)* in 2015, qualitative VfM assessment has become a formal step in PPP project appraisal. The framework has helped standardize ex ante evaluation across jurisdictions, but its application over the past decade has also revealed limitations: the absence of an explicit preliminary applicability screen, overlapping indicator boundaries, insufficient justification for supplementary indicators and weights, and limited disclosure of expert scoring and aggregation procedures. These problems are not merely operational defects. They raise a broader research question about how qualitative VfM assessment should be designed when public authorities rely on expert judgement to evaluate complex PPP suitability under institutional and market uncertainty [16].

This study addresses that question through a comparative and case-based refinement of China’s qualitative VfM framework. It compares the Chinese MoF framework with the UK HM Treasury approach and the UN ESCAP framework, not to transplant foreign models mechanically, but to identify assessment logics that may be institutionally translated into China’s PPP appraisal context [17,18]. Concepts such as pre-screening, output specification, and market-capability assessment are not entirely new in international PPP practice. The contribution of this study lies instead in examining how such logics can be adapted to the institutional setting of China’s MoF expert-panel-based VfM system, which has been implemented in the world’s largest PPP market since 2015. In this sense, this paper is positioned as an institutionally grounded framework-refinement study rather than as a proposal for a universal replacement of existing international approaches.

The study makes three interrelated and bounded contributions. First, it conceptualizes qualitative VfM assessment as a staged evaluation-governance mechanism that separates non-compensatory PPP applicability screening from compensatory diagnostic scoring. Second, it clarifies indicator boundaries among related constructs such as bankability, market contestability, private-sector capability, public-sector capacity, revenue sustainability, and interface complexity. Third, it advances a transparency-oriented approach to expert-based qualitative assessment by emphasizing expert-panel composition, indicator-weight justification, score aggregation, expert-score dispersion, and robustness checks. Together, these contributions seek to improve the reviewability and practical usefulness of qualitative VfM assessment without abandoning the basic institutional structure of the existing MoF framework.

The empirical component is a retrospective reassessment of a large, complex PPP project in China. The case is used to illustrate feasibility, diagnostic usefulness, and score-transparency improvements under the revised framework. It is not treated as independent external validation, nor as evidence that the revised framework is universally superior to the existing MoF framework. Rather, the case demonstrates how a more explicit qualitative VfM procedure can support reviewable and institutionally grounded PPP suitability judgement, while identifying the need for future independent and multi-case testing.

2. China's MoF Qualitative VfM Assessment Framework

Although the intellectual origins of PPPs are rooted in developed economies, the center of gravity has decisively shifted toward emerging markets in recent decades. Nowhere is this more evident than in China, which launched an ambitious national PPP initiative in 2014. By December 2022, China had registered over 14,000 PPP projects, with a cumulative investment exceeding RMB 20.9 trillion (approximately USD 2.89 trillion), making it the world's largest and most state-directed PPP ecosystem.

To institutionalize project appraisal, China's Ministry of Finance [19] issued the Guideline on Value for Money Evaluation (Trial) [Cai Jin (2015) No. 167] on December 18, 2015 (hereinafter referred to as the "Guideline"). This document introduced a formal framework for qualitative VfM assessment tailored to China's infrastructure and public service delivery context. Under this framework, all PPP projects are required to undergo structured qualitative evaluation as a prerequisite for procurement approval.

At its core, the Chinese model is organized around six primary indicators: Lifecycle integration, Risk identification and allocation, Performance orientation and innovation, Potential for competition, Public sector capacity, and Bankability. These criteria reflect a blend of fiscal discipline and developmental pragmatism. In addition to the six core indicators—which account for 80% of the total score—supplementary indicators may be introduced to capture project-specific attributes, such as Project scale, Expected asset life, Revenue growth potential, and Sectoral demonstration effects. Each core indicator typically carries a weight of no more than 20%, and each supplementary indicator no more than 10%. Scoring follows a five-tier scale: "Highly Favorable" (100–81), "Favorable" (80–61), "Neutral" (60–41), "Unfavorable" (40–21), and "Highly Unfavorable" (20–0). This structure combines standardized mandatory assessment with a limited space for contextual adaptation.

The qualitative assessment is conducted by an expert panel comprising professionals from a broad range of fields, including public finance, finance, accounting, asset valuation, law, engineering and technology, project management, and operation and maintenance. The evaluation meeting is convened by the local department of finance (or the PPP unit), in collaboration with the relevant sectoral government agencies. To ensure informed judgment, all required project materials must be provided to the experts in advance of the

meeting. During the meeting, experts engage in structured discussion before independently scoring each evaluation indicator. The final score is calculated as a weighted average of all indicator scores. A project is deemed to have passed the qualitative VfM assessment if the composite score reaches 60 or above.

The MoF framework therefore operates as an expert-panel-based, compensatory weighted-scoring model. It provides a standardized structure for qualitative appraisal while leaving room for project-specific judgement through supplementary indicators and weight adjustment. In practice, the specific weights of indicators are often determined by the expert panel according to project characteristics and local policy considerations. This flexibility is important for accommodating project heterogeneity, but it also makes the transparency of indicator selection, weighting justification, expert scoring, and score aggregation central to the credibility of the assessment.

3. Critical Appraisal of the Qualitative VfM Framework in China

Since its promulgation in 2015, China's qualitative VfM assessment framework has provided an essential scaffolding for formalizing PPP project selection across subnational jurisdictions. As a normative device, it has served to enhance procedural uniformity and institutionalize ex ante accountability. However, as PPP implementation has evolved in both scope and complexity, the framework's underlying design reveals several critical limitations that merit reassessment—not merely as a matter of technical refinement, but as a question of institutional rationality and evaluative integrity. In particular, the current design tends to accommodate different types of judgement within a single weighted scoring process, leaving open the question of whether some threshold conditions for PPP suitability should be considered before projects are assessed through compensatory indicator scoring.

First, the framework exhibits structural interdependencies among key indicators, which undermines their analytical independence. For example, the indicator Risk identification and allocation directly influences the Accuracy of lifecycle cost estimation—a foundational input for evaluating both Bankability and the Potential for competition. These latter indicators are further shaped by variables such as Project scale and Revenue growth potential. While such interlinkages may be justifiable within a qualitative logic model, their application in numerical scoring introduces risks of collinearity among indicators, thereby weakening the discriminative power of the evaluation and compromising its reliability [9,11].

Second, the treatment of supplementary indicators is procedurally weak. Although intended to enhance contextual sensitivity, their selection often lacks transparent justification or empirical grounding. In practice, many evaluations reveal arbitrary inclusion of such criteria, with weight assignments governed more by administrative discretion than by methodological rigor. This absence of a normative framework for indicator customization opens space for opportunistic adaptation [12], weakening both credibility and replicability.

Third, the so-called interdisciplinary design of the expert-based scoring mechanism masks a deeper epistemic blind spot. While the Guideline emphasizes disciplinary diversity in panel composition, it remains agnostic—if not indifferent—toward expertise in the very core of PPP governance: contract structures and long-term risk allocation. This disconnect breeds a dangerous proceduralism, where assessments are performed in form but hollow in substance. What emerges is not robust scrutiny, but a ritualized performance of oversight—what Power [20] aptly called the “audit society,” where the appearance of evaluation supplants its analytical depth.

These challenges are not unique to China. Internationally, qualitative VfM frameworks reflect divergent institutional logics and procurement cultures. The UK's HM Treasury framework, for example, embeds VfM into a broader policy gateway model, emphasizing

strategic fit, market testing, and public interest rationale [3]. ESCAP, by contrast, advances a more developmentalist lens, foregrounding innovation potential and inclusive service delivery [21]. Both offer instructive contrasts to China's administratively driven and fiscally anchored model.

What is at stake is not the abandonment of qualitative assessment, but its repoliticization and recontextualization—to treat it not as a checklist exercise, but as an evolving governance tool situated within a dynamic policy environment. The subsequent section offers a comparative synthesis of international approaches to inform the next generation of qualitative VfM assessment in China.

4. Method

To develop and demonstrate an improved qualitative VfM framework, this study adopted a comparative and design-science-informed methodology. The revised framework is treated as an assessment artifact developed in response to a practical governance problem and demonstrated through case-based application, rather than as a statistically validated predictive model [22,23]. The methodological objective is therefore not to test causal relationships or to prove the universal superiority of the revised framework, but to identify functional gaps in China's existing qualitative VfM assessment system, translate relevant international assessment logics into the Chinese institutional context, and examine whether the refined procedure can be operationalized within a realistic PPP evaluation setting.

A distinctive feature of the study is the authors' practitioner positionality. Author 1 contributed to the original design of China's MoF VfM framework, and both authors co-founded a consultancy that has advised on VfM evaluations for over 100 PPP projects across sectors. This background provided access to the policy intent, field-level implementation experience, and practical limitations of the existing framework. At the same time, it also creates a need for caution in interpreting the proposed refinements and the case reassessment. Accordingly, the study relies on documented framework items for comparative mapping, reports the authors' role transparently, discloses the potential conflicts of interest, and treats the case reassessment as a feasibility and diagnostic demonstration rather than independent external validation.

The comparative analysis proceeded in three steps. First, the existing indicators from China's MoF VfM Guideline were coded as A1–A12 (see Table 1) and compared with the question sets used by HM Treasury in the UK and the UN ESCAP framework. Because the three frameworks are formal assessment instruments with clearly specified indicators or questions, the mapping was not designed as open-ended qualitative coding of unstructured textual data. Instead, it followed a functional-equivalence logic. An item was considered comparable when it addressed the same assessment function, even if different terminology was used across frameworks. For example, questions concerning output definition, performance standards, and contractibility were grouped under output specification clarity. Where an international item addressed a function only partly covered by an MoF indicator, it was treated as an indirect or partial match rather than as a direct equivalent. Where an international item addressed a function not represented in the MoF framework, it was treated as a potential gap only when the issue was also relevant to China's PPP practice and could be incorporated without disrupting the MoF expert-panel workflow.

Table 1. Coded indicators in China’s MoF qualitative VfM assessment framework.

Core Evaluation Indicators	
A1	Lifecycle integration
A2	Risk identification and allocation
A3	Performance orientation and innovation
A4	Potential for competition
A5	Public sector capacity
A6	Bankability
Supplementary evaluation indicators	
A7	Project scale
A8	Expected asset life
A9	Major asset types
A10	Lifecycle cost estimation accuracy
A11	Revenue growth potential
A12	Sectoral demonstration effects

Source: Authors’ own work.

Second, the mapping results were interpreted through two structured comparisons. The MoF framework was first compared with the UK HM Treasury approach, which is widely recognized for its emphasis on output-based contracting, contractual feasibility, market testing, and public-sector readiness. This comparison helped identify assessment functions that are explicit in the UK framework but implicit, under-specified, or absent in the Chinese framework, including output specification clarity, implementation flexibility, interface coordination complexity, and private-sector capability. The MoF framework was then compared with ESCAP’s framework, which reflects the institutional conditions of developing countries in Asia and the Pacific. This comparison further highlighted the value of distinguishing threshold eligibility questions from detailed suitability scoring, and supported refinements such as aligning asset life with contract term and shifting from revenue growth potential to revenue sustainability.

Third, the comparative insights were translated into a refined assessment framework compatible with China’s existing MoF-style expert-panel procedure. The purpose was not to mechanically transplant the UK or ESCAP models, but to adapt selected assessment logics to the Chinese institutional setting. The resulting framework contains three linked components: a preliminary applicability screening mechanism, an optimized indicator system, and improved scoring and reporting procedures. During this synthesis, particular attention was paid to construct boundaries among related indicators, including potential for competition, bankability, private-sector capability, public-sector capacity, revenue sustainability, and interface coordination complexity. This was intended to reduce conceptual overlap and make the scoring logic more reviewable.

Finally, the refined framework was illustrated through a retrospective reassessment of a real-world Chinese PPP case using an actual expert panel. The panel employed a redesigned evaluation toolkit consisting of a pre-screening checklist, an updated indicator handbook, and structured scoring templates. The reassessment documented the full scoring pathway from indicator selection and weighting to individual expert scores, trimmed means, weighted scores, and final judgement. The original MoF-based result and the revised result are reported together so that the comparison does not rest only on two aggregate scores, but also on the procedural choices through which those scores were produced.

Because qualitative VfM assessment relies substantially on expert judgement, the revised procedure emphasized expert-panel composition, advance provision of project information, written scoring rationales, transparent aggregation rules, and sensitivity checks. Kendall's coefficient of concordance and, where appropriate, the intra-class correlation coefficient (ICC), were used as descriptive indicators of expert agreement rather than as stand-alone validation tests. In line with the reporting cautions noted by Koo and Li [24], ICC results were interpreted with reference to the rating design and model assumptions. These procedures were intended to make the case reassessment more traceable and reproducible, while preserving the appropriate boundary of a single-case retrospective analysis. The case evidence is therefore interpreted cautiously: it demonstrates feasibility, procedural operability, and diagnostic usefulness, but it does not establish predictive validity, statistical reliability, or universal superiority over the existing MoF framework.

5. Comparative Analysis I: HM Treasury's Framework

5.1. Overview

The United Kingdom has long been regarded as a pioneer in the development of PPPs, particularly through its Private Finance Initiative (PFI) and its later iteration, PF2. To ensure sound decision-making in project selection, HM Treasury issued the Value for Money Assessment Guidance [3], which outlines a structured framework for qualitative VfM assessments. The approach rests on three fundamental dimensions: Viability, Desirability, and Achievability. Each dimension is broken down into targeted questions that guide the evaluator in determining whether a proposed PPP project can deliver superior value compared to traditional public procurement.

This framework was designed not merely as a checklist, but as a thinking tool to encourage rigorous, context-sensitive analysis. For instance, Viability examines whether the project is economically and financially sound, Desirability tests whether PPP offers clear benefits over conventional procurement in terms of service delivery and risk transfer, and Achievability assesses whether the necessary institutional, contractual, and market conditions exist to deliver the project successfully under a PPP model. Together, these dimensions promote a holistic understanding of the PPP value proposition. In this study, the HM Treasury framework is used as a functional reference rather than as a model for direct transplantation. Its value lies in showing how qualitative VfM assessment can organize early judgement around project outputs, contractual feasibility, risk allocation, market response, and public-sector readiness.

5.2. Mapping Analysis and Structural Comparison

To understand how China might benefit from the UK experience, we conducted a systematic mapping between the indicators used in the MoF Guideline and HM Treasury's qualitative VfM question set. To improve readability, Figure 1 provides a graphical synthesis of the mapping results, highlighting the main correspondences, transferable additions, and comparative insights, while the full item-by-item mapping is retained in Appendix A. More precisely, the mapping was used to identify which assessment functions are explicit in the UK framework but only implicit, under-specified, or absent in the Chinese framework. The mapping reveals three key patterns:

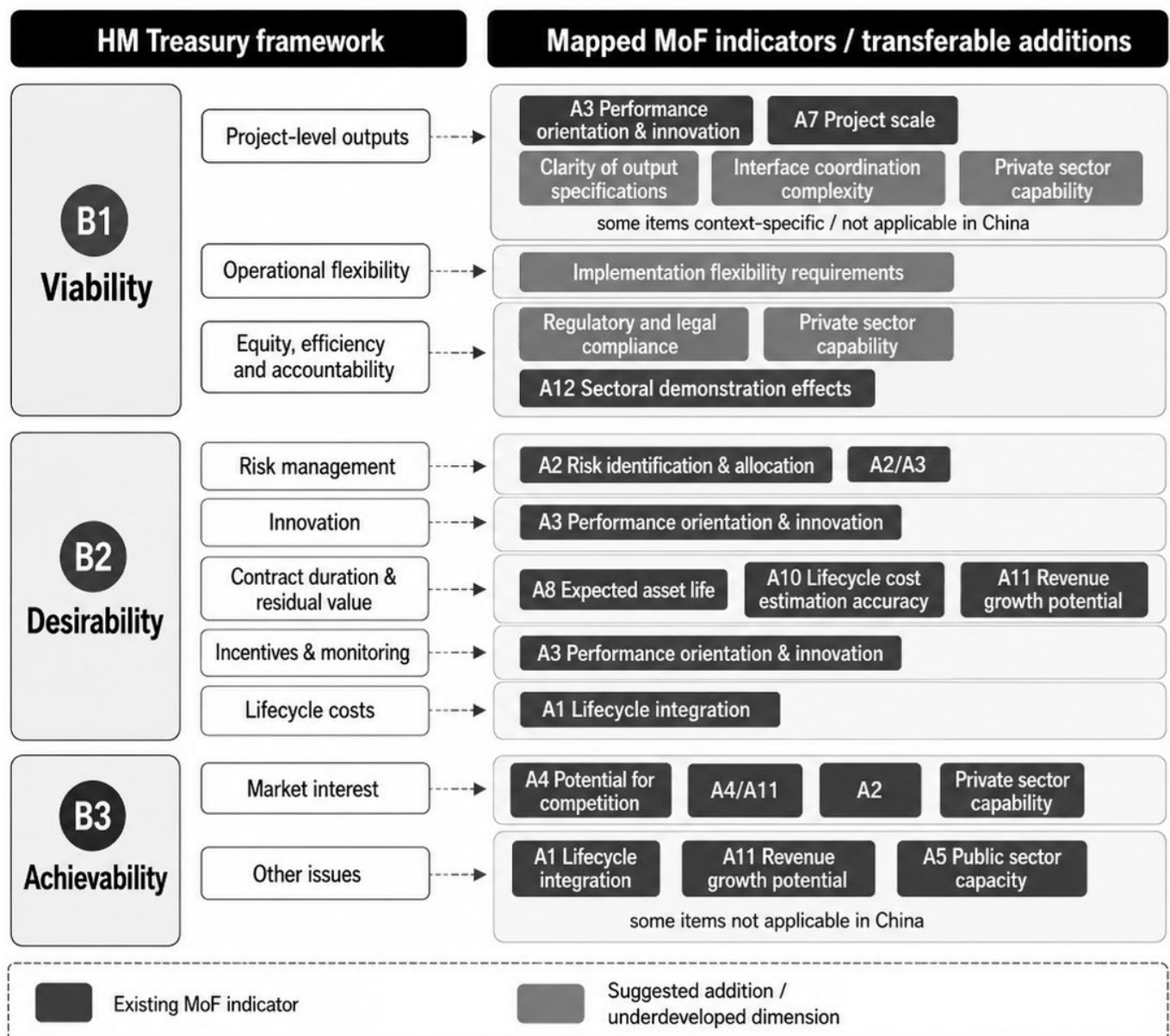


Figure 1. Comparative mapping of qualitative VfM frameworks: HM Treasury vs. China MoF.

First, the UK framework covers most of the Chinese indicators, often with more detailed questions. For example, China's "Public sector capacity" is mirrored by the UK's questions on governance, interface management, and legal frameworks (e.g., B1.1.1–B1.1.12). However, the UK guidance places a stronger emphasis on contract readiness and clarity of service outputs, issues which are often underarticulated in Chinese assessments. This difference is important because output clarity is not merely a drafting issue; it determines whether performance obligations, payment mechanisms, and risk allocation can be made enforceable over the contract term.

Second, several MoF indicators lack direct UK counterparts, as follows:

- A6 (Bankability):** While China isolates this as a separate indicator, the UK approach treats bankability as a function of risk allocation, market appetite, and contract clarity. Indirectly, questions such as B1.1.5 and B1.3.2 touch on financing feasibility but do not elevate it to an independent criterion. This suggests that bankability should not be interpreted only as access to finance, but also as an outcome shaped by contract design, risk allocation, payment certainty, and market confidence.

- **A9 (Major asset types):** This reflects China's emphasis on integrated, multi-functional infrastructure projects. UK guidance rarely addresses this directly, although B1.3.4 hints at scale and integration benefits through service bundling. The absence of a direct counterpart does not necessarily indicate a weakness in either framework. Rather, it reflects different project portfolios: China's local PPP practice has included many integrated area-development and multi-asset projects, while the UK framework places stronger emphasis on whether the bundled service scope can be specified, governed, and delivered through a long-term contract.
- **A10 (Lifecycle cost estimation accuracy):** The UK embeds this issue within its treatment of contract design, boundary definition, and performance metrics, but does not consider it a standalone metric. This difference again points to a construct-boundary issue: lifecycle cost accuracy is not only a technical estimation matter, but is also affected by output specification, risk allocation, asset condition, performance standards, and the availability of reliable project information.

Third, the UK framework places greater weight on certain dimensions that are underdeveloped in the Chinese model:

- **Output specification:** Clear, quantifiable outputs are seen as prerequisites for successful PPP implementation. UK questions such as B1.1.1 and B1.1.2 stress the importance of contractible service definitions.
- **Interface and enabling conditions:** UK guidance includes questions on adjacent systems, environmental preconditions, and boundary clarity (B1.1.9-B1.1.12), all of which contribute to smoother implementation and risk management. These questions are particularly relevant for complex Chinese PPP projects involving multiple sub-projects, land development, public facilities, municipal infrastructure, and inter-agency coordination.
- **Private sector capability:** The UK evaluates the qualifications, track record, and delivery capacity of private partners (B1.3.1-B1.3.6), a consideration not formally embedded in China's indicator system. This indicates a potential gap between assessing whether competition may exist and assessing whether prospective private participants have the technical, financial, managerial, and operational capacity to deliver the required outputs.

5.3. Propositions for the Chinese System

Based on this comparison, we identify six concrete recommendations to improve the Chinese qualitative VfM assessment framework:

- **Include an "Output specification clarity" indicator.** In many Chinese PPPs, vague output definitions lead to contract ambiguities, performance disputes, and high transaction costs. The UK model insists that without contractible outputs, VfM cannot be achieved. A dedicated "output clarity" indicator would encourage project owners to articulate precise service goals and performance metrics.
- **Include a "Stability of project scope for long-term contracting" indicator.** UK guidance warns against rigid PPP contracts for projects facing high policy, demand, or technological uncertainty. Chinese projects often suffer from excessive scope adjustments during implementation. A new indicator could evaluate whether a project's content is sufficiently stable to justify a long-term PPP contract.
- **Include an "Interface complexity" indicator.** Interface problems are a frequent cause of delay and cost escalation in China. The UK model systematically assesses environmental, technical, and administrative interdependencies. Institutionalizing this in the Chinese framework would promote early identification and mitigation of coordination risks. This is especially relevant for area-based development projects, where the

value of PPP delivery depends not only on individual assets but also on the coordination among public facilities, municipal infrastructure, land development, industrial introduction, and government commitments.

- **Include a “Private sector capability” Indicator.** China currently lacks an explicit mechanism to assess the operational, technical, and financial capacity of prospective private partners. Introducing this indicator would allow for early-stage filtering of projects with incapable potential bidders and promote sustainable contract performance. At the same time, this indicator should be clearly distinguished from “Potential for competition” and “Bankability.” Potential for competition concerns whether a credible market contest can occur; bankability concerns whether the project’s payment, revenue, and risk structure can attract finance; private sector capability concerns whether prospective private participants can actually deliver, operate, coordinate, and manage the project over the contract life.
- **Modify “Expected asset lifespan” to emphasize matching with contract tenure.** The UK framework assesses whether the expected lifespan of the asset is appropriately aligned with the contract duration, ensuring that the contract does not outlast the asset or require premature replacement—both of which could undermine value for money (VfM). In contrast, China’s current A8 indicator focuses solely on the asset’s lifespan in isolation, without considering its compatibility with the proposed contractual term. This disconnection can lead to ill-matched project designs and misjudged VfM. The indicator should therefore be revised to evaluate the fit between asset durability and contract tenure, rather than assessing lifespan in abstract.
- **Include a “Compliance” indicator.** The UK framework includes explicit checks on legal, regulatory, and ethical compliance as part of VfM appraisal. The absence of an embedded compliance indicator risks overlooking fundamental eligibility constraints. Introducing a dedicated compliance item into the VfM framework of China would help institutionalize these boundaries, enhance procedural legitimacy, and reinforce alignment with national policy priorities.

However, in the Chinese context, compliance may be better treated as a front-end eligibility consideration rather than as a fully compensatory weighted scoring item. This point is further examined in the comparison with ESCAP’s two-stage screening logic in the next section.

6. Comparative Analysis II: ESCAP’s Framework

6.1. Overview

The United Nations Economic and Social Commission for Asia and the Pacific (ESCAP) has developed a structured framework for the qualitative assessment of VfM in PPPs, designed specifically for the institutional realities and developmental priorities of emerging Asia-Pacific economies. As outlined in the Guidelines and Toolkits for Qualitative Value for Money Assessment of PPP Projects in Developing Countries in Asia and the Pacific [21], the ESCAP framework adopts a two-stage screening process to determine PPP suitability, tailored to the institutional realities of developing Asia-Pacific countries. The Elimination Stage serves as an initial filter based on four pass/fail criteria: (1) the project must involve public infrastructure or services under government responsibility; (2) it must comply with national PPP policy or legal eligibility; (3) it must be identified through legitimate public-sector procedures; and (4) it must align with national or sectoral planning frameworks. Only projects that pass all four criteria proceed to the Selection Stage, where they are assessed against four more refined dimensions: (1) Sector readiness—whether the sector is institutionally prepared for PPP; (2) Private sector interest and readiness—likelihood of attracting capable bidders; (3) Government agency readiness and public support—

institutional capacity and stakeholder buy-in; and (4) Project feasibility and readiness—technical, financial, and legal viability. This two-tiered approach balances efficiency with analytical rigor, ensuring that only eligible and prepared projects enter the PPP pipeline.

For the purpose of this study, ESCAP is not used as a model to be directly transplanted into China's MoF framework. Rather, it provides a regional reference for how qualitative VfM assessment can distinguish threshold eligibility questions from more detailed suitability assessment. This distinction is particularly relevant for examining whether China's existing weighted-scoring approach should be complemented by a front-end judgement of basic PPP applicability.

6.2. Mapping Analysis and Structural Comparison

This section presents a structured mapping analysis between the ESCAP framework and the MoF's Guideline to examine their alignment and divergence in VfM assessment criteria. As the MoF framework does not contain an equivalent to ESCAP's elimination stage, the comparison focuses solely on the selection-stage criteria defined in the ESCAP system.

The analysis reveals a significant degree of alignment between the two frameworks, particularly in core dimensions such as Public sector capacity, Risk identification and allocation, and Bankability. However, several of China's current indicators, notably A3 (Performance orientation and innovation), A8 (Expected asset life), A9 (Major asset types), and A12 (Sectoral demonstration effects), are not explicitly addressed in the ESCAP framework, possibly reflecting China's emphasis on industrial policy alignment and asset diversification.

Conversely, ESCAP includes several dimensions not presently embedded in the MoF's assessment structure, which merit close attention. Three clusters stand out:

First, ESCAP emphasizes the enabling environment for PPP implementation, including legal and institutional compatibility (e.g., C1.1, C1.2), market maturity (C4.1.6), and regulatory adequacy (C4.2.1–C4.2.3). Second, it addresses government support in terms of land acquisition, data availability, and project preparation (e.g., C4.1.1, C4.4.1, C4.4.4). Third, it considers political commitment and public acceptance (C3.4, C3.5). While these indicators are essential in many jurisdictions, their relative absence from the MoF of China framework can be attributed to the structural features of China's policy system—where legal and institutional readiness is generally high, and PPP projects typically receive strong top-down political endorsement. As such, the lack of these dimensions in China's VfM system may reflect contextual fit rather than analytical omission.

That said, the graphical synthesis in Figure 2 shows that several ESCAP indicators not directly addressed in the MoF framework—particularly C2.2, C4.1.5, C4.4.3, C5.1, and C5.5—point to structural weaknesses in China's current approach. The full item-by-item mapping is retained in Appendix B. These indicators highlight clear opportunities for conceptual and procedural refinement. In particular, they reinforce three themes already identified in the HM Treasury comparison: private-sector delivery capability, output specification clarity, and interface or performance independence. They also introduce a more explicit staged-screening logic, through which basic eligibility questions are separated from more detailed suitability scoring.

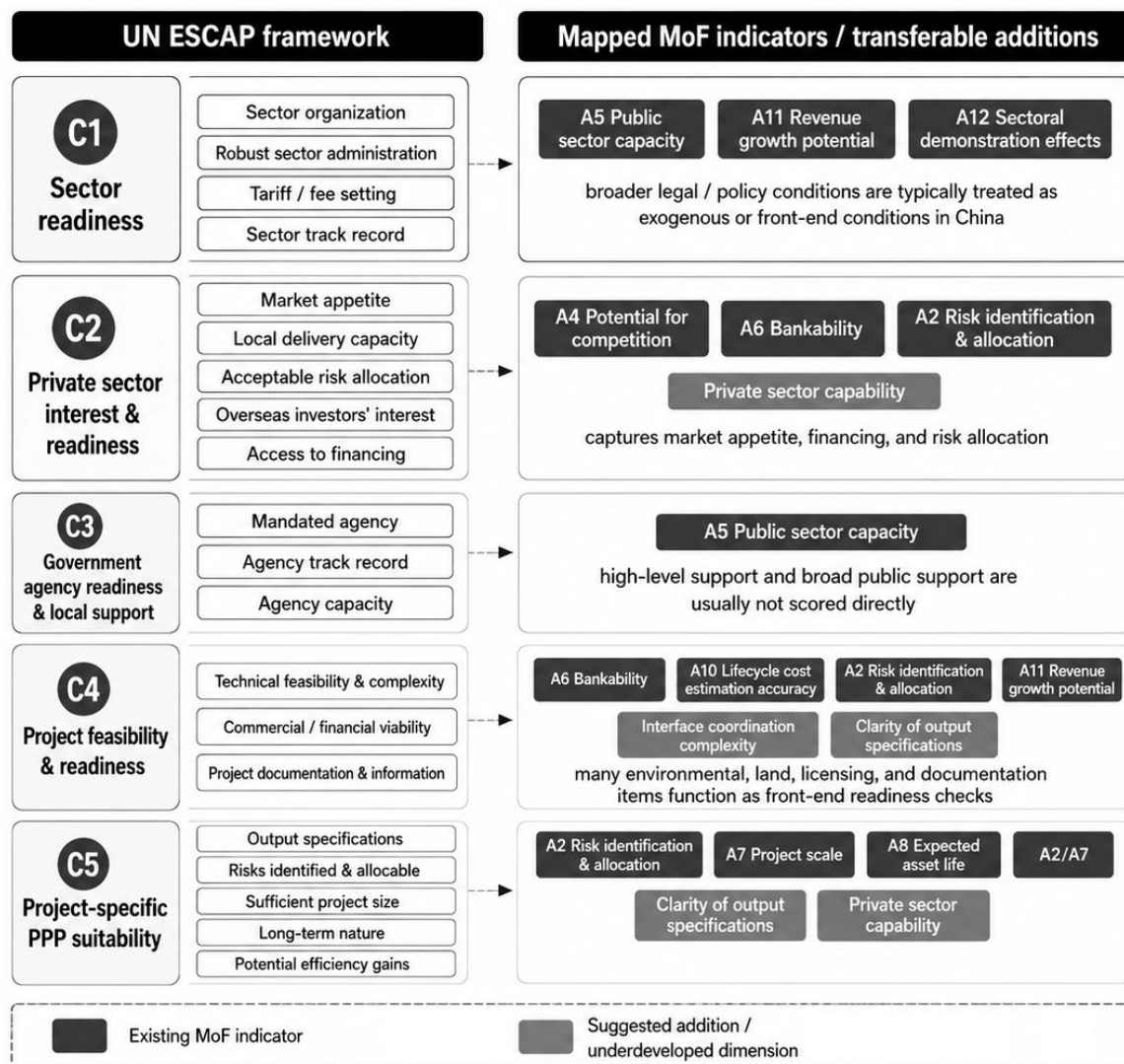


Figure 2. Comparative mapping of qualitative VfM frameworks: UN ESCAP vs. China MoF.

6.3. Propositions for the Chinese System

- Include a “Private sector capability” indicator.** ESCAP explicitly incorporates assessments of Private sector capability under C2.2 and C5.5, including the ability to fulfill performance obligations, assume risk, and mobilize resources over the long term. In contrast, the MoF framework evaluates outcomes like bankability or risk allocation but lacks a standalone indicator on the fitness of prospective private partners. This omission creates blind spots in assessing feasibility, particularly for complex infrastructure requiring advanced technical, financial, or operational expertise. In line with recommendations derived from the UK HM Treasury model (see Section 5.3), the inclusion of a dedicated “Private sector capability” indicator in China’s VfM framework is therefore supported by both UK and ESCAP comparisons. This indicator should be distinguished from bankability and potential competition; it concerns the actual delivery, coordination, financing, and operational capacity of prospective private participants over the contract life.
- Include an “Interface complexity” indicator.** ESCAP’s C4.1.5 draws attention to whether the project scope is clearly defined—a factor critical to contract enforcement, accountability, and transaction structuring. Ambiguities in project boundaries or interface management have been recurrent sources of inefficiency in Chinese PPP practice, especially for integrated urban development or cross-sectoral initiatives. As also noted

in the UK framework (B1.1.9–B1.1.12), institutional coordination challenges at the planning stage can significantly erode VfM during implementation. This reinforces the case for introducing an “Interface coordination complexity” indicator into the MoF’s evaluation structure.

- **Include an “Output specification clarity” indicator.** ESCAP emphasizes the clarity and measurability of output definitions through indicators C4.4.3 and C5.1. This is consistent with the UK model’s prioritization of output-based contracting. In contrast, Chinese practice has witnessed frequent contractual disputes stemming from poorly specified service levels or ambiguous performance targets. The current MoF framework does not directly assess whether a project’s intended outputs are sufficiently defined to serve as the basis for contract enforcement and performance incentives. This gap justifies the addition of a “Clarity of output specifications” indicator as a prerequisite for PPP suitability.
- **Modify “Expected asset lifespan” to emphasize matching with contract tenure.** ESCAP’s C5.4 emphasizes the importance of aligning contract duration with private investment recovery periods, which serves to stabilize financial expectations and optimize risk-sharing. While China’s A8 indicator addresses asset life, it does not account for this crucial aspect of temporal congruence between asset depreciation and contract obligations. As discussed in Section 5.3, a more precise formulation—such as “Alignment of asset life and contract term”—would better reflect VfM logic and the underlying financial structure of PPPs.
- **Modify “Revenue growth potential” to emphasize revenue sustainability.** ESCAP places notable emphasis on the sustainability of revenue streams rather than their long-term growth trajectory. This diverges from the MoF’s A11 indicator, Revenue growth potential. Drawing from ESCAP (C4.3 series) and the UK model (B2.2.3), a conceptual shift from growth-centric to sustainability-centric revenue expectations is warranted. For long-term, service-oriented infrastructure, revenue predictability is often more valuable than aggressive growth assumptions, which can create perverse incentives or distort project risk assessments. The MoF should consider revising A11 to “Revenue sustainability,” thereby aligning with best practices in fiscal prudence and risk management.
- **Deprioritize “efficiency gains” as a standalone indicator.** ESCAP’s C5.5 includes efficiency gains as an evaluative dimension. However, efficiency gains are better understood as an emergent outcome, not an ex ante criterion. Their realization depends on upstream factors such as clearly specified outputs, appropriate risk allocation, private sector capability, and adequate project scale—all of which are already addressed through other indicators in both ESCAP and the MoF frameworks. Retaining “efficiency gains” as a standalone indicator risks redundancy and conceptual confusion. It should instead be treated as a composite result, not an input to VfM assessment.
- **Institutionalize a preliminary screening stage.** Finally, the two-tier screening approach of the ESCAP framework offers critical procedural value. The Elimination Stage identifies projects fundamentally unfit for PPP implementation—such as those lacking public service character, policy alignment, or regulatory clarity—thereby preventing costly selection-stage assessment on inherently unsuitable candidates. This upfront filtration enhances efficiency, ensures resource discipline, and safeguards the legitimacy of the PPP pipeline. Incorporating a similar pre-screening mechanism into China’s VfM system would substantially reduce misjudgments at the source and improve overall efficiency. For China, however, such a mechanism should not simply duplicate ESCAP’s elimination criteria. Some compliance and planning issues are already addressed through existing administrative procedures. The more rele-

vant lesson is the staged logic itself: before a project is assessed through weighted qualitative scoring, evaluators should first consider whether it satisfies basic PPP applicability conditions, including public-service character, policy and legal eligibility, output definability, scope stability, and manageable implementation interfaces. This staged logic provides the procedural basis for the refined framework proposed in the following section.

7. Refined Qualitative VfM Assessment Framework for China

Based upon the comparative analysis of qualitative VfM assessment frameworks adopted by the MoF of China, the UK HM Treasury, and the UN ESCAP, this section proposes a refined VfM assessment framework for PPP projects in China. The proposed improvements are structured around three dimensions: (1) a preliminary applicability screening mechanism, (2) an optimized indicator system, and (3) enhanced scoring methodology. The refinement does not replace the MoF framework; rather, it reorganizes and supplements it so that qualitative VfM assessment can better reflect the staged nature of PPP decision-making. In this revised logic, basic PPP applicability is considered before detailed scoring, project suitability is then diagnosed through more clearly bounded indicators, and expert-based scoring is made more traceable through improved procedural design.

7.1. Preliminary Applicability Screening Mechanism

Given that PPPs are long-term contractual arrangements for the provision of public services, it is crucial to assess, at an early stage, whether a project is fundamentally suitable for PPP implementation before undertaking a full-scale VfM evaluation. As ESCAP emphasizes, its pre-selection screening mechanism “allows one to determine whether the project is eligible and merits further assessment. This ensures that time and resources are not wasted on developing ineligible project ideas”.

Building on the nature of the indicators or questions embedded in the three previously reviewed frameworks, along with the pre-selection criteria defined in ESCAP’s elimination stage, we propose a set of four preliminary indicators. These indicators are designed to serve as a gatekeeping mechanism—filtering out unsuitable projects and ensuring that only those with the fundamental prerequisites for PPP success proceed to formal evaluation.

- **Clarity of output specifications.** Projects should demonstrate explicit definitions regarding service content, quantity, quality, standards, and performance indicators. Where output clarity is lacking, PPPs should be deemed inappropriate.
- **Stability of project scope for long-term contracting.** PPP contracts are inherently rigid due to their long-term nature. Projects that are likely to experience frequent scope changes, regulatory shifts, or technological disruption during implementation are ill-suited to such arrangements. This new indicator should assess whether a project’s volatility aligns with PPP’s contractual rigidity. Projects requiring high adaptability should be filtered out at this stage.
- **Interface coordination complexity.** This indicator is used to assess whether project interfaces are clearly defined, responsibility allocation is transparent, and coordination is feasible. Projects with high coordination difficulty should be cautiously considered for PPP structuring.
- **Legal and policy compliance.** Certain projects involving national security, sensitive public interests, or activities prohibited by law—are inherently unsuitable for PPP implementation.

In China, all four ESCAP pre-selection criteria (see Section 6.1) can be interpreted as compliance-related, given the country's comprehensive and centralized legal-policy regime.

- Criterion 1 is addressed through detailed regulations defining public infrastructure or services under government responsibility. For example, the MoF's Notice on Regulating the Management of the PPP Project Database (No. 92 [2017]) excludes projects outside the public service domain or beyond government delivery obligations—such as commercial real estate, and also those involving national security or major public interests that are not appropriate for private sector participation. Notice on Further Standardizing the Implementation of PPP Projects under the New Mechanism (No. 1013 [2024]) issued by the National Development and Reform Commission (NDRC) further prohibits highly market-oriented commercial and industrial projects from applying the PPP modality.
- Criterion 2—compliance with PPP laws and policies—is self-evident.
- Criterion 3 is reflected in formal project identification procedures, such as the NDRC's basic construction protocols and the MoF's PPP operational guidelines.
- Criterion 4 is embedded in China's planning-led development model, where national and regional plans serve as binding references for infrastructure investment and project approval.

Together, these elements form a robust compliance architecture that warrants explicit incorporation into China's VfM framework as a front-end filter. At the same time, the proposed screening mechanism should be understood as a PPP-specific extension of China's existing administrative review system, not as a duplication of all planning, approval, or fiscal-affordability procedures. Its distinctive contribution is to connect formal eligibility review with the substantive conditions for long-term PPP contracting, especially output definability, scope stability, interface manageability, and legal-policy suitability.

A special note is warranted for complex area-based development projects that are unique to the Chinese context—such as industrial parks and urban redevelopment zones—which often constitute exceptions to Clarity of output specifications and Stability of project scope for long-term contracting. While these projects typically lack fully defined output specifications at the outset and exhibit a high degree of implementation flexibility, their institutional and transactional design compensates for such uncertainty through several key mechanisms: (1) The initial feasibility study and bidding documents outline a broad, strategic framework for overall development rather than fixed outputs; (2) project content is refined and implemented in a rolling, phased manner throughout execution; (3) the return mechanism is typically based on a cost-plus model, underpinned by endogenous revenues generated within the development zone (e.g., land sales and tax contributions from incoming enterprises). Importantly, if actual returns fall short, the project company bears the loss; if returns exceed expectations, the company is only entitled to a capped return based on the pre-agreed internal rate of return (e.g., 8%) determined at the time of bidding. Given this structured balance of risk and reward, and the unique nature of value creation in such projects, it is reasonable to consider applying the two aforementioned preliminary screening indicators through contextualized interpretation rather than mechanical exclusion. Their success depends less on precise early-stage specifications and more on adaptive implementation within a robust institutional and contractual framework. In such cases, the assessment report should explain how strategic planning, phased implementation, return regulation, performance monitoring, and government coordination mechanisms substitute for fixed early-stage specifications. This preserves the screening function while allowing the framework to accommodate governable flexibility in complex area-development PPPs.

7.2. Optimizing the Core and Supplementary Indicator System

Once projects pass the applicability screening, they may proceed to full-scale qualitative VfM evaluation. We propose a revised structure of core (80%) and supplementary (20%) indicators:

(a) Core Evaluation Indicators (Total Weight: 80%)

Building on the existing six indicators in the MoF Guideline, we propose the following refined structure with the addition of a new indicator:

1. Lifecycle integration (unchanged);
2. Risk identification and allocation (unchanged);
3. Performance orientation and innovation (unchanged);
4. Potential for competition (unchanged);
5. Public sector capacity (unchanged);
6. Private sector capability (new);
7. Bankability (unchanged).

The new indicator, Private sector capability, is intended to capture a dimension that is related to but analytically distinct from Potential for competition and Bankability. Potential for competition concerns whether a credible market contest can occur; Bankability concerns whether the project's revenue, payment, and risk structure can attract finance; Private sector capability concerns whether prospective private participants possess the technical, financial, managerial, operational, and risk-bearing capacity required for long-term delivery. This separation helps the indicator system distinguish market formation, financing feasibility, and delivery capability as related but different dimensions of PPP suitability.

(b) Supplementary Evaluation Indicators (Total Weight: 20%)

These indicators serve to provide contextual nuance and project-specific differentiation. Suggested revisions include:

1. Project scale (unchanged);
2. Alignment of asset life and contract term (revised);
3. Major asset type (unchanged);
4. Revenue sustainability (revised);
5. Sectoral demonstration effects (unchanged).

Two revisions are particularly important. First, "Expected asset life" should be reframed as "Alignment of asset life and contract term." The relevant VfM question is not whether the asset has a long lifespan in isolation, but whether asset durability, investment recovery, maintenance responsibility, residual value, and contract duration are mutually consistent. Second, "Revenue growth potential" should be reframed as "Revenue sustainability." For long-term infrastructure and public-service projects, credible and stable revenue or payment streams are generally more important than optimistic growth expectations.

The supplementary indicator system should therefore remain selective rather than automatic. The inclusion of supplementary indicators should depend on whether specific project characteristics materially affect contract design, financing logic, risk allocation, implementation complexity, or long-term performance. This preserves contextual flexibility while ensuring that each selected supplementary indicator has a clear diagnostic function.

7.3. Improving Scoring Mechanisms and Procedural Design

To enhance transparency, objectivity, and analytical robustness, the following improvements are proposed. Because qualitative VfM assessment involves structured judgement across multiple criteria, its credibility depends not only on the selected indicators, but also on the clarity of scoring rules, weighting logic, aggregation procedures, and the documentation of expert judgement [16].

1. Expert panel composition

Expert panels should maintain cross-disciplinary representation (finance, law, engineering, management), while further strengthening expertise in PPP contractual design, risk pricing, and performance management.

2. Use of trimmed mean to mitigate outlier effects

Given inherent subjectivity in expert scoring, we recommend applying a trimmed mean method—excluding the highest and lowest scores before calculating the average—to reduce bias and enhance result robustness.

3. Justification of indicator weighting and selection

Particularly for supplementary indicators, experts should provide structured justifications for the inclusion and weighting of indicators in evaluation reports to avoid ad hoc or inconsistent applications.

Although this study recommends maintaining the current 80:20 weight distribution for the core and supplementary indicators, it also encourages MoF to explore adaptive weighting schemes in the future—possibly informed by data-driven simulations or comparative logic—to establish a more refined and empirically grounded system over time.

Last but not least, the issue of indicator collinearity—as previously raised in Section 3—remains unresolved. This structural interdependence among indicators challenges the interpretability of a composite scoring system by inflating or distorting the final evaluation results. Because qualitative VfM indicators often describe related aspects of the same project configuration, some degree of interaction among indicators is unavoidable. The practical task is therefore not to eliminate all overlap, but to make the boundaries among related constructs clearer and to interpret composite scores with appropriate caution. For example, Risk identification and allocation, Bankability, Potential for competition, Private sector capability, and Revenue sustainability may influence one another, but they should not be allowed to collapse into a single undifferentiated judgement about project attractiveness.

In practice, scores substantially above the formal passing threshold may provide a more cautious basis for interpretation than marginal passing scores, especially where indicator interactions are evident. However, such reference points should be used descriptively and should not become additional formal thresholds without institutional authorization or empirical calibration. The more important methodological improvement is to accompany the composite score with a traceable explanation of indicator logic, weighting assumptions, score dispersion, and, where appropriate, sensitivity to alternative weighting assumptions. In this sense, the refined framework does not seek to convert qualitative judgement into mechanical calculation; rather, it seeks to make qualitative judgement more structured, transparent, and institutionally reviewable.

8. Case Study

8.1. Case Background and Reassessment Design

The case analyzed in this section involves a large-scale, area-based development PPP project (see Section 7.1), initiated by the local government of a Chinese city in March 2016. In November of the same year, a PPP advisory firm—co-founded by the authors—was engaged to provide technical support, marking the formal commencement of project preparation. The bidding process for selecting a concessionaire was launched in January 2018, followed by the signing of the PPP contract in April and the registration of the project company in June. Construction officially began in July 2018, accompanied by a ceremonial groundbreaking.

With an estimated total investment of approximately RMB 49 billion (around USD 7 billion), this project stands among the largest PPP initiatives ever undertaken in China.

Implemented under a Build–Operate–Transfer (BOT) model with a 20-year concession term, the project company was capitalized with RMB 500 million in registered equity—80% contributed by the consortium and 20% by the local government. The overall capital structure follows a typical PPP financing pattern, comprising 20% equity and 80% debt, aligning with prevailing norms for large infrastructure ventures.

This project provides a useful setting for applying the refined qualitative VfM framework because it combines several features that make early-stage PPP suitability judgement difficult: large investment scale, multiple asset categories, long-term contractual governance, phased implementation, government-payment commitments, and extensive coordination among planning, land development, infrastructure construction, public services, and industrial development. Its complexity is not confined to technical scale; it also derives from the differentiation and interdependence of organizational, contractual, financial, and environmental elements [25,26]. These characteristics make the case suitable for examining whether qualitative VfM assessment can move beyond aggregate scoring and provide a more explicit diagnosis of project suitability.

This section uses the project as a retrospective case reassessment to examine how the refined qualitative VfM framework operates when applied to a complex Chinese PPP project. The analysis returns to the original project context and assessment materials, but reorganizes the evaluation around the proposed staged logic: preliminary applicability screening, refined indicator selection, adjusted weighting, structured expert scoring, and more explicit score interpretation. In this sense, the case is used to compare the diagnostic profile generated by the original MoF-based assessment with that produced by the refined framework. The authors' involvement in the original project preparation provides access to project documentation, policy context, and implementation knowledge, which supports a more detailed reconstruction of the assessment process. At the same time, the case is interpreted as a context-rich reassessment of framework operability and diagnostic usefulness, rather than as a general test of the framework across all PPP settings.

8.2. Initial Qualitative VfM Assessment

The initial VfM assessment for the project commenced in August 2017, led by the advisory firm in accordance with the Guideline. In September 2017, a formal expert evaluation session was convened, employing the expert scoring method prescribed by the guideline.

A multidisciplinary panel of eight seasoned professionals was assembled. Each expert independently scored the project against the indicators, drawing upon their domain-specific insights. The panel covered project management, asset valuation, finance, public finance, sectoral development, accounting, law, and engineering technology, reflecting the broad expert-category arrangement commonly used in MoF-based qualitative VfM assessment. Three supplementary indicators were selected by the Bureau of Finance: Project scale, Major asset types, and Sectoral demonstration effects. Indicator weights were determined collaboratively by the expert panel.

The final VfM score was calculated as a weighted average of the individual indicators (see Table 2). For each indicator, the highest and lowest expert scores were removed before calculating the average score, consistent with the aggregation rule used in the original evaluation materials. As shown in Table 2, the project received strong evaluations in areas such as Lifecycle integration (86.17) and Major asset types (88.50), indicating high operational coherence and multi-asset synergies. The overall score of 83.50 suggested that the project met the qualitative VfM threshold for PPP suitability.

Table 2. Original VfM assessment results under the MoF framework.

No.	Evaluation Indicator	Weight	Average Score	Weighted Score
1	Lifecycle integration	15%	86.17	12.93
2	Risk identification and allocation	15%	84.67	12.7
3	Performance orientation and innovation	15%	81.5	12.23
4	Potential for competition	15%	80	12
5	Public sector capacity	10%	80.5	8.05
6	Bankability	10%	80	8
7	Project scale	6.66%	92.5	6.17
8	Major asset types	6.66%	88.5	5.9
9	Sectoral demonstration effects	6.66%	83.17	5.54
			Final Score:	83.5

Source: Authors' own work.

The detailed expert-score matrix underlying this reconstruction is reported in Appendix C, while Table 2 presents the corresponding summary results used for comparison in the main text. For the purpose of reporting, experts in the original assessment are anonymized as O1–O8. These codes are used only to present the original scoring matrix and do not imply matched identities or matched professional roles with the experts in the later reassessment.

This initial assessment establishes the MoF-based baseline for the case. It shows that the project obtained a clear positive qualitative VfM judgement under the formal scoring procedure. At the same time, the structure of the result also illustrates the analytical features of the original framework: the conclusion is expressed mainly through a composite score, while the diagnostic roles of supplementary indicators, expert-score dispersion, and indicator-weighting choices remain less visible in the published assessment summary. The following reassessment therefore does not revisit the project because the original conclusion was negative or procedurally invalid; rather, it uses the same case to examine how a refined framework may produce a more differentiated account of PPP suitability.

8.3. Qualitative VfM Assessment Using the Revised Framework

To apply the refined qualitative VfM framework to the case project, the advisory firm—previously involved in the original evaluation—conducted a retrospective reassessment using the same project context and available assessment materials. The purpose of the reassessment was to examine how the refined framework would organize the evaluation process differently, especially through preliminary applicability screening, adjusted indicator selection, revised weighting, and more traceable expert scoring. The following subsections present the reassessment process and results.

8.3.1. Pre-Meeting Preparations

(a) Project Information Compilation

To facilitate an informed and efficient assessment, the advisory firm compiled a comprehensive information package encompassing all relevant dimensions of the project. The package included an overview of the project's planning framework, locational profile, scope and scale of implementation, transaction structure, and return mechanisms. It also detailed the preparatory work undertaken by the local government, including lessons learned from similar past projects, institutional setup, and preliminary engagement with prospective investors and financial institutions. Additional components included the

proposed output specifications and performance targets, risk allocation plans, financial projections (including IRR and NPV), and fiscal affordability assessments.

(b) Design of Evaluation Toolkit

The assessment framework was supported by a three-part toolkit:

1. **Pre-screening Checklist:** A checklist of four binary (yes/no) questions regarding the preliminary applicability screening mechanism to establish the project's basic eligibility for VfM assessment.
2. **Indicator Handbook:** This manual provided refined definitions, scoring criteria, and recommended weight ranges for the seven core and five optional indicators used in the evaluation.
3. **Data Collection Templates:** The templates were designed to record indicator selection, assigned weights, individual expert scores, trimmed means, and weighted scores. This allowed the reassessment to retain the MoF-style expert scoring procedure while making the scoring pathway more explicit.

(c) Expert Panel Selection

Given the project's large investment scale, complex asset composition, and multi-phase implementation, the consulting team devised a rigorous expert selection strategy. Experts were evaluated against three criteria: (1) demonstrated expertise across PPP-related domains; (2) hands-on experience with large-scale infrastructure projects; and (3) direct involvement in key PPP functions such as performance management, risk allocation, and contract design. Following availability consultations and suitability reviews, seven experts were provided in advance with the full project dossier. One-on-one consultations were also offered to ensure familiarity with the project and alignment with the refined VfM framework.

For reporting purposes, the seven experts in the reassessment are anonymized as R1–R7 in the detailed scoring matrix. These codes are independent of the O1–O8 codes used for the original assessment and do not indicate matched expert identities or matched professional roles. Compared with the original eight-member panel, the reassessment did not include a separate asset-valuation expert. This adjustment reflected the greenfield BOT nature of the project and the refined diagnostic focus of the assessment. Asset valuation expertise is particularly central to TOT, brownfield asset-transfer, or existing-asset concession projects; in this case, asset-related judgement was mainly reflected through the alignment of asset life and contract term, supported by engineering, financial, legal, and project-management judgement.

8.3.2. Panel Meeting Proceedings

(a) Project Briefing and Clarification

The meeting began with a detailed presentation by the advisory firm, outlining the project's critical features. A structured Q&A session followed, during which panelists posed questions regarding data gaps and clarification points.

(b) Pre-screening Assessment

The project was unanimously deemed eligible for full VfM evaluation based on its solid performance against the key criteria outlined in the preliminary applicability screening. In particular, it satisfied the requirements for interface coordination and legal and policy compliance—two indicators closely linked to implementation feasibility. The presence of a dedicated implementing agency with clear institutional mandates and established inter-agency coordination mechanisms significantly reduced the risk of fragmentation, a concern particularly relevant to large-scale, multi-stakeholder initiatives. On the compliance front, the project aligned with national PPP priorities and conformed to all relevant legal and

regulatory frameworks. No statutory or policy-related red flags—such as ineligible project types or sectoral restrictions—were identified.

That said, it is important to acknowledge that the project constitutes a typical exception to two other screening dimensions: clarity of output specifications and stability of project scope over the long term. As discussed in Section 7.1, these limitations reflect the intrinsic characteristics of area-based development PPPs rather than deficiencies in project design. In this sense, the project passed the pre-screening not by rigidly meeting all criteria, but by demonstrating how context-sensitive mechanisms and institutional safeguards can support progressive specification, phased implementation, and controlled adjustment during the cooperation period—thereby justifying its advancement to the full VfM evaluation stage.

(c) Selection and Justification of supplementary indicators

The panel reviewed five candidate supplementary indicators to complement the seven mandatory core ones.

- **Project scale:** Experts acknowledged that project size can enhance economies of scale. However, in this case, the already substantial scale could distort the overall assessment if overemphasized. Hence, this indicator was excluded.
- **Alignment of asset life and contract term:** For this long-term, large-scale development, alignment between asset life and contract duration was seen as vital to avoid either resource underutilization or excessive reinvestment. This indicator was thus retained.
- **Major asset types:** While asset diversity exists, the experts found it hard to quantify this indicator's influence on VfM in a consistent manner. It was excluded to avoid confounding effects.
- **Revenue sustainability:** Given the project's reliance on a mix of user fees, viability gap funding, and fiscal revenues from induced development, sustainability of revenue flows was deemed critical and included.
- **Sectoral demonstration effects:** The project's complexity and innovative structure, combined with strong government–capital–finance collaboration, make it a valuable benchmark for similar future developments. This indicator was also retained.

In summary, three supplementary indicators were adopted: Alignment of asset life and contract term, Revenue sustainability, and Sectoral demonstration effects.

(d) Weight Assignment

Core Indicators: Weights were determined through deliberative consensus.

Lifecycle Integration and Risk Allocation were both assigned 10%, reflecting their foundational importance in shaping PPP structure and long-term viability.

Performance orientation and Innovation received a weight of 10% as well, following debate between its theoretical value and practical constraints due to execution limitations and institutional maturity.

Public sector capacity, Private sector capability, and Bankability each received 15%, identified by panelists as the most decisive factors in large-scale PPPs. These dimensions were seen as determinants of whether the project could navigate its complexity and fulfill its performance targets.

Potential for competition was assigned 5%, as the project was still in the pre-procurement phase, with some interest already expressed but competitive depth untested.

Supplementary indicators:

Among the three retained supplementary indicators, Revenue sustainability received the highest weight of 10% due to its pivotal role in ensuring financial continuity. Asset-Contract Alignment and Demonstration Potential were assigned 5% each to reflect their relevance without overshadowing core capacity indicators.

(e) Scoring and Aggregation

Each expert independently evaluated the project across the selected indicators using a structured scoring form, which also required written justifications for their ratings. To mitigate the influence of outliers, the advisory team applied a trimmed mean approach by discarding the highest and lowest scores for each indicator. The resulting values were then used to calculate a weighted average across all indicators. Table 3 presents the average scores for each indicator, the assigned weights, the corresponding weighted scores, and a summary of the experts' comments. The detailed expert-score matrix is provided in Appendix D.

Table 3. Indicator weights, trimmed mean scores, weighted scores, and expert comments.

Evaluation Indicator	Weight	Average Score	Weighted Score	Summary of Comments
Lifecycle Integration	10%	89	8.90	The project integrates design, construction, financing, and O&M, aligning with the full-cycle governance logic of PPPs. Some experts noted concerns regarding cross-phase coordination costs and management efficiency.
Risk identification and allocation	10%	87.6	8.76	Risks are comprehensively identified with a sound allocation mechanism. However, experts advised strengthening contingency plans for force majeure and government default risks.
Performance orientation and innovation	10%	86.4	8.64	The contract specifies key investment milestones and outcome-based performance indicators, providing room for private innovation. Experts suggested further enriching the performance tools and evaluation mechanisms.
Potential for competition	5%	87.4	4.37	Several leading contractors proactively approached the government, indicating strong market interest.
Public sector capacity	15%	87.8	13.17	Experts recognized the local government's strong organizational capacity and PPP experience in early-stage preparation, especially through establishing a dedicated task force and engaging consultants. However, experience in area-based development still needs to be strengthened.
Private sector capability	15%	84	12.60	Potential private partners include central and large state-owned enterprises with strong financial and implementation capabilities. However, they may lack experience in industrial operation and integrated area development, posing challenges in coordination and transformation.
Bankability	15%	85	12.75	The project has received preliminary support from several financial institutions, including policy banks, indicating favorable financing conditions and strong capital market appeal.
Alignment of asset life and contract term	5%	84	4.2	The 20-year contract covers most major assets' life spans, with mid-to-long-term maintenance mechanisms embedded. Some experts noted that responsibility for sub-project maintenance needs to be clarified.
Revenue sustainability	10%	81.6	8.16	The project relies mainly on government payments, with relatively stable and well-designed funding sources. However, certain subsidies rely heavily on budgeted funds and are sensitive to market fluctuations, creating fiscal uncertainty risks.
Sectoral demonstration effects	5%	87.4	4.37	Experts widely agreed that the project represents a typical model of "integrated development + park operation + industrial induction" in the region. It offers replicable and scalable value in terms of management, cooperation approaches, and financing.
Final Score	85.92			

Note: The average score refers to the trimmed mean calculated after excluding the highest and lowest expert scores for each indicator. Weighted score is calculated by multiplying the average score by the corresponding indicator weight.

The final score—85.92 out of 100—was well above the formal threshold of 60 and also above the 70-point conservative reference used in this study for descriptive interpretation. This result suggests that the project demonstrated strong alignment with PPP principles and was well-positioned to deliver value for money if implemented effectively. More

importantly, the refined assessment made the basis of that judgement more explicit by showing which dimensions supported the project's suitability and which areas still required attention during procurement, contract negotiation, and implementation monitoring.

8.4. Score Traceability and Interpretive Stability

The original MoF-based assessment and the reassessment reached the same qualitative judgement. The original assessment produced a rounded score of 83.50, while the reassessment produced a score of 85.92; both exceed the formal passing threshold of 60 and the conservative 70-point reference used for descriptive interpretation. The refined framework therefore did not reverse the project's PPP suitability judgement, but made the evaluative basis of that judgement more explicit.

The detailed scoring matrices in Appendices C and D show how individual expert scores were aggregated into trimmed means and weighted scores. Appendix E further reports expert-agreement measures and sensitivity checks. The original assessment showed stronger expert concordance, whereas the reassessment revealed greater dispersion, reflecting the more differentiated judgement required by indicators such as private-sector capability, Revenue sustainability, and asset-life/contract-term alignment. The sensitivity checks indicate that the positive judgement remained stable under alternative weighting assumptions. Taken together, the case comparison suggests that the main value of the refined framework lies less in the numerical increase from 83.50 to 85.92 than in making qualitative VfM judgement more traceable, diagnostic, and attentive to implementation risks.

9. Discussion

The refined qualitative VfM assessment framework for PPPs in China, as illustrated by the case application, should be understood as an incremental and institutionally compatible refinement of the existing MoF procedure. Its purpose is not to replace the MoF framework or to demonstrate numerical superiority over the original assessment. Rather, it seeks to strengthen the procedural transparency, diagnostic clarity, and reviewability of expert-based qualitative VfM judgement within China's existing expert-panel-based workflow.

First, the introduction of a preliminary applicability screening mechanism strengthens the gatekeeping function of qualitative VfM assessment. In the original MoF approach, threshold questions and diagnostic scoring are largely accommodated within a single compensatory weighted-score process. This may allow fundamental weaknesses in output definition, scope stability, interface coordination, or legal-policy eligibility to be offset by favorable scores in other dimensions. The proposed staged logic addresses this problem by separating non-compensatory applicability screening from subsequent compensatory diagnostic scoring. In practical terms, the pre-screening stage should function as a front-end pass/fail or pass/conditional-pass/fail mechanism. Projects that clearly fail fundamental requirements should not proceed to full qualitative scoring. Projects that receive a conditional pass should proceed only where the assessment report provides written justification and identifies mitigation measures. This operational distinction helps preserve the integrity of the screening function while retaining the flexibility needed for complex PPP projects.

Second, the case application shows that the refined framework can be incorporated into China's existing MoF-style expert-panel workflow without requiring a radical procedural redesign. The revised procedure retains the core structure of expert review, indicator weighting, and weighted scoring, but supplements it with a pre-screening checklist, clearer indicator definitions, written scoring rationales, and more transparent aggregation records. These additions are intended to make the evaluation process more traceable rather than more burdensome. Nevertheless, implementation challenges should not be ignored. Local

governments may differ substantially in PPP expertise, quality of project documentation, availability of qualified experts, and capacity to manage more structured reporting requirements. For the refined framework to be applied consistently, it would require standardized templates, clearer scoring anchors, expert training, and improved documentation discipline. The framework is therefore best viewed as a gradual institutional refinement that can be adopted within the existing system, rather than as a one-step procedural overhaul.

Third, the reassessment confirms that the main value of the revised framework lies in procedural transparency and diagnostic differentiation rather than in the small numerical difference between the original and revised scores. The original MoF-based assessment produced a score of 83.50, while the reassessment produced a score of 85.92. Both results support the same positive qualitative VfM judgement. This consistency is important because it indicates that the refined framework does not overturn the original conclusion. However, the reassessment makes the basis of the judgement more explicit by showing how indicator selection, weighting, expert scoring, trimmed means, and sensitivity checks contribute to the final conclusion. The value of the revised framework therefore lies less in changing the final score than in clarifying why the project is considered suitable for PPP delivery and where implementation risks remain.

Fourth, the expert-agreement results deserve careful interpretation. The original assessment showed stronger expert concordance, while the reassessment revealed lower concordance and weak absolute agreement. This result should not be treated as statistical validation of the refined framework, nor should it be ignored. Rather, it indicates that the revised indicators require experts to make more differentiated judgements about complex constructs, including private-sector capability, revenue sustainability, and asset-life/contract-term alignment. Greater dispersion may therefore reflect the increased diagnostic granularity of the revised framework, but it also signals that future applications need clearer scoring rubrics, more detailed indicator manuals, and stronger expert calibration before scoring. In this sense, expert-score dispersion is not merely a methodological weakness; it is also useful information about where the framework, expert preparation, or project documentation requires further refinement.

Fifth, the refined framework clarifies rather than eliminates construct overlap. Qualitative VfM assessment inevitably involves related constructs because PPP suitability depends on mutually reinforcing conditions, including risk allocation, market contestability, financing feasibility, institutional capacity, and long-term revenue stability. The practical task is therefore not to remove all conceptual interaction, but to prevent related indicators from collapsing into a single undifferentiated judgement of project attractiveness. In the revised framework, potential for competition concerns whether a credible market contest can occur; bankability concerns whether the project's payment, revenue, and risk structure can attract finance; private-sector capability concerns whether prospective private participants possess the technical, financial, managerial, operational, and risk-bearing capacity required for long-term delivery; revenue sustainability concerns the credibility and stability of user payments, government payments, or other revenue sources; and interface coordination complexity concerns the manageability of technical, administrative, and organizational interdependencies. These distinctions help make the scoring logic more reviewable, even though some interaction among indicators remains unavoidable.

Finally, the broader applicability of the framework should be interpreted cautiously. The case project is a large-scale, area-based development PPP with multiple asset categories, phased implementation, and extensive public–private coordination. Its complexity makes it a useful setting for demonstrating the operability of the staged assessment logic. However, a single retrospective case cannot establish generalizability across all sectors, regions, transaction structures, or project stages. The framework's screening emphasis, supplementary

indicators, expert composition, and weighting structure should be adapted to different PPP contexts, such as transport, water, environmental, energy, social infrastructure, and urban development projects. Further independent applications and multi-case comparisons are needed before stronger claims can be made about the framework's transferability or performance.

10. Conclusions

This study critically reviews China's qualitative VfM assessment framework for PPPs and proposes an institutionally grounded refinement informed by comparative analysis with the UK HM Treasury and UN ESCAP approaches. The paper does not claim that concepts such as pre-screening, output specification, or market-capability assessment are entirely new in international PPP practice. Its contribution lies in examining how these assessment logics can be translated into China's MoF expert-panel-based VfM system, which has been formally implemented in the world's largest PPP market since 2015.

The proposed framework contains three linked components. First, it introduces a preliminary applicability screening mechanism to distinguish non-compensatory threshold conditions from compensatory diagnostic scoring. Second, it refines the indicator system by clarifying the boundaries among related constructs such as potential for competition, bankability, private-sector capability, public-sector capacity, revenue sustainability, and interface coordination complexity. Third, it strengthens scoring and reporting procedures through clearer expert-panel requirements, written scoring rationales, trimmed-mean aggregation, expert-score dispersion reporting, and sensitivity checks. These refinements are intended to improve the transparency and reviewability of qualitative VfM judgement without abandoning the basic institutional structure of the existing MoF framework.

The retrospective reassessment of a complex, large-scale Chinese PPP project illustrates the feasibility and diagnostic usefulness of the refined framework. The revised assessment produced a score close to that of the original MoF-based assessment and did not alter the positive VfM conclusion. This is consistent with the incremental nature of the proposed refinement. The significance of the case application lies not in demonstrating a large improvement in numerical performance, but in showing how the revised procedure can make the basis of expert judgement more explicit. In particular, the refined framework makes visible the role of pre-screening, indicator selection, weighting justification, expert-score dispersion, and sensitivity analysis in supporting a more traceable qualitative VfM assessment.

The study also has important limitations. First, the retrospective case application is not independent external validation. The authors' involvement in both the original VfM framework and the case reassessment provides valuable contextual knowledge, but it also requires cautious interpretation. Second, the use of a single large-scale PPP case cannot establish generalizability across China's diverse PPP sectors and project types. Third, the expert-agreement results show that more differentiated indicators may generate greater dispersion in expert judgement, especially where constructs are complex and project information is incomplete. Fourth, the refined procedure may impose additional implementation demands on local governments, particularly where PPP expertise, project documentation, or expert-pool management capacity is limited.

Accordingly, the proposed framework should be regarded as a preliminary and institutionally adaptable refinement rather than a fully validated universal model. Future research should apply the framework to independent multi-case samples across different sectors, regions, transaction structures, and project stages. Longitudinal studies should also examine whether qualitative VfM assessments conducted under the refined framework are associated with subsequent project performance, fiscal sustainability, service quality,

and contract-management outcomes. In addition, further work is needed to develop more detailed scoring anchors, expert-calibration procedures, and digital decision-support tools. Such efforts would help determine whether the proposed framework can move beyond procedural improvement and contribute more directly to the long-term governance quality of PPP projects.

Author Contributions: Conceptualization, C.C.; methodology, C.C.; validation, L.H.; formal analysis, C.C. and L.H.; resources, C.C.; data curation, L.H.; writing—original draft preparation, C.C. and L.H.; writing—review and editing, C.C.; visualization, C.C. and L.H.; supervision, C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data used and analyzed during this study are available from the corresponding author by request.

Conflicts of Interest: The authors disclose that they are co-founders of a PPP advisory firm that provided technical support for the original VfM assessment of the case project. This involvement is acknowledged as a potential conflict of interest and interpretive limitation. To mitigate this concern, the case and expert information are anonymized, the scoring procedures are reported transparently, and the reassessment is presented as a demonstration of framework operability rather than independent external validation. The study received no external funding, and no external organization influenced the research design, analysis, interpretation, or manuscript preparation.

Appendix A. Full Item-by-Item Mapping Between HM Treasury’s Qualitative VfM Question Set and China’s MoF Indicators

HM Treasury	MoF of China	Note
B1 Viability		
B1.1 Project level outputs		
B1.1.1		Clarity of output specifications
B1.1.2		Clarity of output specifications
B1.1.3		Clarity of output specifications
B1.1.4		Clarity of output specifications
B1.1.5	A3	
B1.1.6	A7	
B1.1.7		Not applicable to China
B1.1.8		Not applicable to China
B1.1.9		Interface coordination complexity
B1.1.10		Not applicable to China
B1.1.11		Private sector capability
B1.1.12		Interface coordination complexity
B1.2 Operational flexibility		
B1.2.1		Implementation flexibility requirements
B1.2.2		Implementation flexibility requirements
B1.2.3		Implementation flexibility requirements

B1.2.4		Implementation flexibility requirements
B1.3 Equity, efficiency and accountability		
B1.3.1		Regulatory and legal compliance
B1.3.2		Regulatory and legal compliance
B1.3.3		Regulatory and legal compliance
B1.3.4		Private sector capability
B1.3.5		Private sector capability
B1.3.6	A12	
B1.4 Overall viability		
B2 Desirability		
B2.1 Risk management		
B2.1.1	A2	
B2.1.2	A2/A3	
B2.2 Innovation		
B2.2.1	A3	
B2.2.2	A3	
B2.2.3	A3	
B2.2.4	A3	
B2.3 Contract Duration and residual value		
B2.3.1	A8/A10	
B2.3.2	A8	
B2.3.3	A8/A11	
B2.4 Incentives and monitoring		
B2.4.1	A3	
B2.4.2	A3	
B2.4.3	A3	
B2.5 Lifecycle costs		
B2.5.1	A1	
B2.5.2	A1	
B2.6 Overall desirability		
B3 Achievability		
B3.1 Market Interest		
B3.1.1		Private sector capability
B3.1.2		Private sector capability
B3.1.3	A4	
B3.1.4	A4	
B3.1.5	A4/A11	
B3.1.6	A2	
B3.2 Other Issues		
B3.2.1		Not applicable to China
B3.2.2	A1	

B3.2.3	A11
B3.2.4	A5
B3.3 Overall Achievability	
Source: Authors' own work.	

Appendix B. Full Item-by-Item Mapping Between UN ESCAP's Qualitative VfM Criteria and China's MoF Indicators

	UN ESCAP	MoF of China	Note
C1	Sector Readiness		
	C1.1	Legal and Regulatory Framework	N/A
	C1.2	Supportive Policies for Private Sector Participation	N/A
	C1.3	Sector Organization	A5
	C1.4	Robust Sector Administration	A5
	C1.5	Tariff/Fee Setting Mechanisms	A11
	C1.6	Sector Track Record	A12
C2	Private Sector Interest and Readiness		
	C2.1	Market Appetite	A4/A6
	C2.2	Local Delivery Capacity	Private sector capability
	C2.3	Acceptable Risk Allocation	A2
	C2.4	Overseas Investors' Interest	A4
	C2.5	Access to Financing	A6
C3	Government Agency Readiness and Local Support		
	C3.1	Existence of a Mandated Agency to Undertake the Project	A5
	C3.2	Government Agency Track Record	A5
	C3.3	Government Agency Capacity	A5
	C3.4	High-level Support	N/A
	C3.5	Broad Public Support for the Project	N/A
C4	Project Feasibility and Readiness		
	C4.1	Project Implementation Complexity and Technical Feasibility	
	C4.1.1	Availability of Land and Key Resources	N/A
	C4.1.2	Proven Technology	A6
	C4.1.3	Cost Predictability	A6/A10
	C4.1.4	Manageable Construction and Operation Risks	A2/A10
	C4.1.5	Self-sustaining Performance / Performance Independent of Other Projects	Interface coordination complexity
	C4.1.6	Straightforward Licensing and Permitting Procedures	N/A
	C4.2	Socio-economic, Environmental, and Institutional Impacts	
	C4.2.1	Manageable Environmental, Health and Safety Issues	N/A

	C4.2.2	Limited Land Clearance and Resettlement Required		N/A
	C4.2.3	No Opposition Expected to Organizational / Institutional Restructuring		N/A
C4.3	Commercial/Financial Viability			
	C4.3.1	Acceptable Level of Fees Envisaged	A11	
	C4.3.2	Realistic Financial Feasibility Assessment	A10	
	C4.3.3	Stable Demand for Public Services (Limited Alternatives)	A11	
	C4.3.4	Identified Revenue Streams	A11	
C4.4	Availability of Project Documentation and Information			
	C4.4.1	Availability of Pre-feasibility Studies		N/A
	C4.4.2	Supporting Government Decisions Taken		N/A
	C4.4.3	Project Concept and Output Clearly Defined		Clarity of output specifications
	C4.4.4	Good Quality Information Available		N/A
C5	Project Specific PPP Suitability			
	C5.1	Output Specifications		Clarity of output specifications
	C5.2	Project Risks Identified and Allocable	A2	
	C5.3	Sufficient Project Size	A7	
	C5.4	Long-term Nature	A8	
	C5.5	Potential Efficiency Gains	A2/ A7	Private sector capability/Clarity of output specifications

Source: Authors' own work.

Appendix C. Original Expert Score Matrix and Score Reconstruction Under the MoF Framework

Category	Evaluation Indicator	Weight	O1	O2	O3	O4	O5	O6	O7	O8	Minimum	Maximum	Trimmed Mean	Weighted Score
Core indicators	Lifecycle integration	15%	85	84	85	83	90	78	90	90	78	90	86.17	12.93
	Risk identification and allocation	15%	85	82	70	81	95	80	95	85	70	95	84.67	12.70
	Performance orientation and innovation	15%	80	83	75	81	85	80	95	80	75	95	81.50	12.23
	Potential for competition	15%	75	72	80	82	80	78	90	85	72	90	80.00	12.00
	Public sector capacity	10%	80	74	70	82	85	82	85	80	70	85	80.50	8.05
Supplementary indicators	Bankability	10%	85	82	80	76	80	78	80	80	76	85	80.00	8.00
	Project scale	6.67%	95	81	95	92	100	85	98	90	81	100	92.50	6.17
	Major asset types	6.67%	85	82	90	89	95	80	95	90	80	95	88.50	5.90
	Sectoral demonstration effects	6.67%	85	78	60	88	90	78	98	80	60	98	83.17	5.54
Final score														83.51

Note: The trimmed mean is calculated by excluding one highest and one lowest expert score for each indicator and averaging the remaining six scores. The reconstructed final score is 83.51, which is consistent with the rounded score of 83.50 reported in the original MoF-based qualitative ViM assessment.

Appendix D. Reassessment Expert Score Matrix and Score Reconstruction Under the Refined Framework

Category	Evaluation indicator	Weight	R1	R2	R3	R4	R5	R6	R7	Minimum	Maximum	Trimmed mean	Weighted score
Core indicators	Lifecycle integration	10%	91	90	85	90	85	90	90	85	91	89.00	8.90
	Risk identification and allocation	10%	95	80	90	95	88	85	80	80	95	87.60	8.76
	Performance orientation and innovation	10%	90	90	90	80	82	90	75	75	90	86.40	8.64
	Potential for competition	5%	95	90	91	90	86	65	80	65	95	87.40	4.37
	Public sector capacity	15%	89	70	85	95	90	90	85	70	95	87.80	13.17
	Private sector capability	15%	80	85	88	80	87	70	95	70	95	84.00	12.60
	Bankability	15%	85	80	90	90	92	80	80	80	92	85.00	12.75
Supplementary indicators	Alignment of asset life and contract term	5%	80	90	85	90	75	55	90	55	90	84.00	4.20
	Revenue sustainability	10%	85	70	88	80	83	85	75	70	88	81.60	8.16
	Sectoral demonstration effects	5%	90	85	82	90	91	90	80	80	91	87.40	4.37
Final Score							85.92						

Note: R1–R7 denote anonymized experts in the retrospective reassessment. The R-codes are independent of the O-codes used in Appendix C and should not be interpreted as paired expert identities or paired professional roles. The trimmed mean is calculated by excluding one highest and one lowest expert score for each indicator and averaging the remaining five scores. The weighted score is calculated by multiplying the trimmed mean by the corresponding indicator weight. Minor rounding differences may occur because weighted scores are reported to two decimal places.

Appendix E. Score Traceability, Expert Agreement, and Sensitivity Checks

Table A1. Aggregate score comparison.

Measure	Value
Original reconstructed score	83.51
Original rounded score reported in the main text	83.50
Reassessment score under the refined framework	85.92
Absolute difference between reconstructed scores	2.41
Percentage difference relative to original reconstructed score	2.88%
Same qualitative judgement above the formal 60-point threshold	Yes
Same qualitative judgement above the 70-point conservative reference	Yes

Note: The two scores are compared as aggregate qualitative judgements rather than as paired indicator-level results. A matched indicator-level statistical test is not applied because the two frameworks differ in indicator composition, weights, scoring criteria, and expert-panel size.

Table A2. Expert agreement measures.

Framework	Kendall's W	Chi-Square	df	Approx. <i>p</i> -Value	ICC(2,1)	ICC(2,k)	Interpretation
Original MoF framework	0.4086	26.1500	8	0.0010	0.2503	0.7276	Moderate rank concordance and stronger average-rating consistency than the reassessment matrix.
Refined framework reassessment	0.1438	9.0623	9	0.4315	−0.0070	−0.0515	Low concordance and weak absolute agreement, indicating visible dispersion in expert judgement.

Note: Kendall's W and ICC are reported as descriptive agreement measures. They are used to show the structure and dispersion of expert judgement within each scoring exercise, not to establish predictive validity or general statistical reliability of the framework. ICC results should be interpreted with caution because their meaning depends on the rating design, model specification, and assumptions about the score scale [24].

Table A3. Sensitivity analysis of reassessment score under alternative weighting assumptions.

Scenario	Score	Above 60?	Above 70?	Threshold-Sensitive?
Reported reassessment weights	85.9200	Yes	Yes	No
Equal weights across all ten indicators	86.0200	Yes	Yes	No
Preserve core/supplementary group totals; equal weights within each group	86.2610	Yes	Yes	No
Single-indicator $\pm 10\%$ perturbations, normalized weights: minimum score	85.8772	Yes	Yes	No
Single-indicator $\pm 10\%$ perturbations, normalized weights: maximum score	85.9636	Yes	Yes	No
Single-indicator $\pm 20\%$ perturbations, normalized weights: minimum score	85.8353	Yes	Yes	No
Single-indicator $\pm 20\%$ perturbations, normalized weights: maximum score	86.0082	Yes	Yes	No
Emphasize lifecycle/risk/performance indicators at 50% total weight	86.4190	Yes	Yes	No
Emphasize market capacity and finance indicators at 50% total weight	86.0550	Yes	Yes	No
Emphasize supplementary context indicators at 50% total weight	85.4104	Yes	Yes	No

Note: The 70-point value is used only as a conservative descriptive reference, not as a formal threshold or a statistical correction for indicator interdependence. The sensitivity checks show that the reassessment conclusion remains above the tested reference points under the listed weighting assumptions. They do not remove conceptual interdependence among indicators or establish a universal correction rule.

References

1. Yescombe, E.R. *Public-Private Partnerships: Principles of Policy and Finance*; Elsevier: Oxford, UK, 2007.
2. Allen, G. *The Private Finance Initiative (PFI)*; Research Paper 03/79; House of Commons Library: London, UK, 2003.
3. HM Treasury. *Value for Money Assessment Guide*; HM Treasury: London, UK, 2006.
4. ADB. *Value for Money in PPPs: An Infrastructure Governance Approach*; Asian Development Bank: Manila, Philippines, 2022.
5. Hodge, G.A.; Greve, C. Public-Private Partnerships: An international performance review. *Public Adm. Rev.* **2007**, *67*, 545–558. [CrossRef]

6. Shaoul, J. A critical financial analysis of the Private Finance Initiative: Selecting a financing method or allocating economic wealth? *Crit. Perspect. Account.* **2005**, *16*, 441–471. [\[CrossRef\]](#)
7. Ball, R.; Heafey, M.; King, D. Private Finance Initiative-A good deal for the public purse or a drain on future generations? *Policy Politics* **2001**, *29*, 95–108. [\[CrossRef\]](#)
8. DeCorla-Souza, P.; Lee, D.; Mayer, J.; Timothy, D. Comparing Public-Private Partnerships with conventional procurement. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2346*, 32–39. [\[CrossRef\]](#)
9. Grout, P.A. Value-for-Money measurements in Public-Private Partnerships. *EIB Pap.* **2005**, *10*, 33–56.
10. Park, H.; Lee, S.; Kim, J. Do Public Private Partnership projects deliver value for money? An ex post Value for Money (VfM) Test on three road projects in Korea. *Int. J. Urban Sci.* **2018**, *22*, 579–591. [\[CrossRef\]](#)
11. Heald, D. Value for money tests and accounting treatment in PFI schemes. *Account. Audit. Account. J.* **2003**, *16*, 342–371. [\[CrossRef\]](#)
12. Zhao, J.; Liu, H.J.; Love, P.E.D.; Greenwood, D.; Sing, M.C.P. Value for Money assessments for Public-Private Partnerships: Characteristics, research directions, and policy implications. *Dev. Built Environ.* **2023**, *16*, 100246. [\[CrossRef\]](#)
13. PPIAF. *Module 5: Implementation and Monitoring in Toolkit for Public Private Partnerships in Roads and Highways*; Private Participation in Infrastructure Advisory Facility: Washington, DC, USA, 2009.
14. WBI; PPIAF. *Value-for-Money Analysis—Practices and Challenges: How Governments Choose When to Use PPP to Deliver Public Infrastructure and Services*; World Bank Institute and Private Participation in Infrastructure Advisory Facility: Washington, DC, USA, 2013.
15. EPEC. *Value for Money Assessment: Review of Approaches and Key Concepts*; European PPP Expertise Centre: Luxembourg, 2015.
16. Morgan, M.G. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7176–7184. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Dolowitz, D.P.; Marsh, D. Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance* **2000**, *13*, 5–24. [\[CrossRef\]](#)
18. Minkman, E.; van Buuren, M.W.; Bekkers, V.J.J.M. Policy transfer routes: An evidence-based conceptual model to explain policy adoption. *Policy Stud.* **2018**, *39*, 222–250. [\[CrossRef\]](#)
19. MoF. *Guideline on Value for Money Evaluation (Trial)*; Cai Jin (2015) No. 167; MoF: Beijing, China, 2015.
20. Power, M. *The Audit Society: Rituals of Verification*; Oxford University Press: Oxford, UK, 1997.
21. ESCAP. *Qualitative Value-For-Money Guidance & Toolkit for Assessing PPP Projects in Developing Countries in Asia and the Pacific*; United Nations Economic and Social Commission for Asia and the Pacific: Bangkok, Thailand, 2016.
22. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *Manag. Inf. Syst. Q.* **2004**, *28*, 75–105. [\[CrossRef\]](#)
23. Peffers, K.; Rothenberger, M.A.; Tuunanen, T.; Chatterjee, S. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [\[CrossRef\]](#)
24. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Baccarini, D. The concept of project complexity—A review. *Int. J. Proj. Manag.* **1996**, *14*, 201–204. [\[CrossRef\]](#)
26. Bosch-Rekveldt, M.; Jongkind, Y.; Mooi, H.; Bakker, H.; Verbraeck, A. Grasping project complexity in large engineering projects: The TOE (Technical, Organizational and Environmental) framework. *Int. J. Proj. Manag.* **2011**, *29*, 728–739. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.