World Journal of Economics and Business Research
Print ISSN: 2960-0073
Online ISSN: 2960-0081
DOI: https://doi.org/10.61784/wjebr3123

PRACTICAL CHALLENGES AND IMPROVEMENT PATHWAYS IN THE PERFORMANCE EVALUATION OF GOVERNMENT-PURCHASED PUBLIC SERVICES

Nan Li
Institute of Law, Shandong Academy of Social Sciences, Jinan 250000, Shandong, China.

Abstract: Performance evaluation of government-purchased public services is not merely a matter of project acceptance or financial verification. It is an institutional process that begins with public needs, operates through purchase contracts, and centers on service quality and public value, thereby enabling an integrated assessment of project decisions, contract performance, fiscal expenditure, and actual outcomes. In practice, mechanisms for setting performance objectives, conducting third-party evaluation, and providing feedback have gradually been established. Yet performance objectives often remain abstract; indicators overemphasize measurable outputs; evaluative authority is concentrated in purchasing authorities while service users participate only marginally; and evaluation findings rarely exert substantive influence on payments, provider selection, or budget allocation. These difficulties stem from the disconnection between needs assessment and project approval, inconsistencies among budget objectives, procurement requirements, contractual provisions, and evaluation indicators, the structural dependence and limited professional capacity of third-party evaluators, and unclear rules and accountability boundaries for using performance information. Drawing on a review of the literature and normative analysis, this article argues for an evaluation foundation that links needs, objectives, contracts, and indicators around actual user benefits and public value; a pluralistic evaluation structure in which purchasing authorities remain accountable, service users participate effectively, and third parties operate under clear professional rules; and the lawful, contract-based incorporation of evaluation results into payment, contract management, budget adjustment, corrective-action verification, and information disclosure. The fundamental directions for reform are a shift from counting outputs to assessing quality and outcomes, from single-actor judgment to triangulation across multiple sources of evidence, and from merely producing evaluation reports to substantively using performance information.
Keywords: Government purchase of public services; Performance evaluation; Public value; Service-user participation; Use of evaluation results

1 INTRODUCTION

Government purchasing of public services is an institutional arrangement through which public authorities, while retaining ultimate responsibility for service provision, use contractual mechanisms to involve social organizations and other nongovernmental actors in service delivery. As such projects have expanded in eldercare, disability assistance, employment services, social assistance, and community governance, attention has shifted from whether and how services should be purchased to whether the purchase is effective and whether citizens actually benefit. Both the policy of comprehensive budget performance management and the Measures for the Administration of Government Purchase of Services require performance evaluation and the use of its results to be embedded in budget and project management[1-2]. Existing scholarship explains this issue from several angles. Studies of government transformation and the "smart purchaser" emphasize capacity for needs assessment, provider selection, and contract management[3-4]; research on whole-process management and effectiveness examines the relationship among project design, implementation oversight, and service performance[5-6]; studies of indicator systems advocate multidimensional measures tailored to the attributes of public services[7]; research on public participation stresses preference expression and the responsiveness of purchasing authorities[8]; and studies of third-party evaluation show that evaluation reliability, organizational independence, and professional competence still require institutional safeguards[9-10].
These studies provide an important foundation for understanding the performance of government-purchased public services, but they generally address indicator construction, actor configurations, or third-party operation separately. Less attention has been paid to how these problems jointly distort performance information and how evaluation evidence enters payment, contract management, and budget allocation. In practice, evaluation often concentrates on service frequency, numbers of people covered, budget execution rates, and documentary completeness, while service quality, problem resolution, and the actual benefits received by citizens are insufficiently captured. This can produce a form of "formalistic performance" in which a project is completed without generating meaningful effects. Accordingly, this article uses literature synthesis and normative analysis to examine three connected stages: performance objectives and indicators, evaluators and information sources, and the use of evaluation results. It argues that current difficulties arise not only from technical weaknesses in indicator design, but from the absence of a closed institutional loop linking needs assessment, contract design, the allocation of evaluative authority, and the use of results. Reform should therefore

move from quantitative accounting toward quality and outcome assessment, from unilateral judgment by purchasing authorities toward multi-actor participation, and toward stronger linkages between evaluation results and payment, contract management, and budget decisions.

2 CONCEPTUAL FOUNDATIONS OF PERFORMANCE EVALUATION FOR GOVERNMENT-PURCHASED PUBLIC SERVICES

2.1 Basic Meaning of Performance Evaluation

Performance evaluation of government-purchased public services is an institutional process that uses public needs, project objectives, purchase contracts, and relevant management rules to assess systematically the rationality of project approval, the economy of fiscal inputs, the efficiency of implementation, the quality of contract performance, and social outcomes. It is related to, yet distinct from, financial inspection, contract acceptance, and ordinary satisfaction surveys. Financial inspection primarily verifies the authenticity and compliance of expenditure; contract acceptance determines whether promised deliverables have been supplied; and satisfaction surveys capture service users' subjective experiences. Performance evaluation must go further by asking whether the project merits continued purchase, whether the provider has performed effectively, and whether public expenditure has generated service value commensurate with the resources invested. Its object should therefore extend beyond the service provider to include whether the project itself was appropriately designed, whether the purchasing authority fulfilled its responsibilities for needs management and contract oversight, whether service users received fair access, and whether the project generated its intended public outcomes.

Theoretically, government purchasing of public services rests on a principal-agent relationship mediated by contract. The professional quality and long-term effects of services such as eldercare, psychological support, and employment assistance cannot be fully specified when a contract is signed, nor can they be established solely through observable outputs. Providers generally possess more process information than purchasing authorities and service users, and the purchaser's capacity for continuous monitoring and contractual adjustment directly affects performance quality[11]. Public services also embody public values such as equity, responsiveness, accessibility, and trust. An evaluation focused only on unit costs and quantitative efficiency may allow measurable "efficiency" to conceal declining service quality or the exclusion of disadvantaged groups. Public value management holds that the legitimacy of public management should emerge through dialogue, responsiveness, and the balancing of multiple values[12], while performance measurement research shows that different managerial purposes require different measures and that no single "universal indicator" can simultaneously satisfy the needs of accountability, control, budgeting, learning, and improvement[13]. Performance evaluation of government-purchased public services should therefore encompass at least economy, efficiency, effectiveness, equity, responsiveness, and procedural compliance, with differentiated judgments tailored to the service type.

2.2 Principal Functions of Performance Evaluation

First, performance evaluation serves resource-allocation and accountability functions. Government-purchased services are financed primarily through public funds, and evaluation should determine whether the public need addressed by a project is genuine, whether purchasing is an appropriate delivery mechanism, whether inputs and outputs are proportionate, and whether the project should be retained, adjusted, or discontinued. Purchasing changes the producer of a service, but it does not transfer the government's ultimate responsibility for ensuring public service provision. By translating contractual requirements concerning staff qualifications, service procedures, quality standards, target-group coverage, and risk response into verifiable items, evaluation can identify fraudulent service delivery, insufficient outputs, professional misconduct, biased selection of service users, and inefficient expenditure. It can prevent purchasing authorities from relaxing oversight after completing procurement procedures and can require providers to remain accountable to both public funds and service users through traceable records.

Performance evaluation also supports organizational learning and service improvement. The needs addressed by government-purchased services are often dynamic and context-dependent, and initial project design cannot fully anticipate changes in target populations, resource constraints, or implementation barriers. By combining contract data, on-site observation, service-user experience, complaint records, and professional judgment, evaluation can identify bottlenecks in service processes, omissions in target-group coverage, and design errors, and feed those findings into the next round of needs assessment, procurement requirements, and budget preparation. The core value of evaluation therefore lies not in generating a score or ranking, but in providing evidence for public decisions about whether to continue purchasing, what to purchase, who should provide it, and how delivery should be improved. Only when accountability and learning are combined can performance evaluation avoid becoming a redundant extension of contract acceptance.

3 MAJOR PROBLEMS AND THEIR CAUSES

3.1 Vague Performance Objectives and Overreliance on Quantitative Indicators

The vagueness of performance objectives is first evident in the failure to operationalize the target population and the change a project is expected to produce. Project documents often adopt broad goals such as "improving eldercare services," "promoting employment among disadvantaged job seekers," or "strengthening community governance" without specifying whether priority should be given to older people with functional impairments, those living alone, or the general older population. They may fail to distinguish unique service users from repeated service contacts or to specify quality requirements concerning response times, staff qualifications, risk management, and service continuity. At the evaluation stage, these broad goals are then replaced by data that can be read directly from administrative records, such as service contacts, numbers of activities, budget execution rates, and documentary completeness. In a home-based eldercare project, for example, repeated meal deliveries to the same older person within one month may be counted as multiple service contacts, but the resulting completion figure does not show whether very old, functionally impaired, or geographically isolated older people were accurately reached. Similarly, the number of trainees completing an employment-assistance program says little about job matching, employment retention, or income improvement. Such indicators can prove that a provider "carried out activities," but not that the project "changed the underlying problem." Zhang's case study demonstrates that formal goal completion may become detached from actual public needs[14], and research on indicator systems likewise shows that diverse public services cannot be evaluated through a single generic template[7].

The reason vague objectives become a bias toward quantity is that front-end needs assessment, contract design, and back-end evaluation do not form a continuous evidentiary chain. Some projects begin with the previous year's budget ceiling and an inherited project label rather than with a roster of prospective service users, a needs baseline, and an identified service gap. Performance objectives may be drafted by budget staff, procurement requirements by the responsible program unit, contracts reviewed by procurement or legal personnel, and the evaluation plan assembled only shortly before project closeout. The four stages then use different language. A budget objective may promise to "improve service satisfaction," while the procurement document merely requires a specified number of annual activities, the contract permits monthly payment upon submission of attendance sheets, and the final evaluation has no basis for examining actual changes in service users' circumstances. If evaluators become involved only after implementation, they cannot obtain pre-intervention baselines, information on eligible persons who were not served, or records of participants who dropped out. They are left to verify the files, photographs, and financial documents supplied by the provider. Research on the "smart purchaser" identifies needs assessment and contract-management capacity as preconditions for effective government purchasing[4], while contract-management research shows that weak monitorability raises the cost of controlling performance[11].

Under these conditions, quantitative indicators not only carry limited information but may also reshape provider behavior. To protect completion rates, providers have stronger incentives to select service users who are easy to reach, communicate with, and likely to report satisfaction, while reducing investment in groups whose needs are complex, service costs are high, or outcomes are unlikely to emerge quickly. The same activity may be inflated through repeated sign-ins, artificial subdivision into multiple sessions, or counting under several categories, whereas professional tasks that leave fewer auditable traces, such as individual assessment, sustained follow-up, resource referral, and crisis intervention, are compressed. Purchasing authorities may likewise treat high budget execution rates, complete files, and an absence of complaints as safe evidence of performance, producing the substitute judgment that "the more complete the paperwork, the better the project." Research on "bureaucratic outsourcing" shows that administrative tasks and documentation requirements can crowd out the professional autonomy of social organizations[15], while studies of project attributes and process management indicate that performance depends on the fit between management arrangements and project characteristics rather than the mechanical attainment of a single indicator[16]. The essential problem is therefore not the existence of quantitative measures, but their displacement of indicators that should capture quality, outcomes, and equitable coverage.

3.2 Limited Evaluator Diversity and Insufficient Participation by Service Users

The concentration of evaluators is evident first in the simultaneous concentration of evaluative roles and information sources. Purchasing authorities commonly control project approval, procurement, contract administration, acceptance, and payment, while performance evaluation is organized by the same functional unit or administrative chain. A low evaluation score may discredit not only the provider but also the purchaser's own project design and routine oversight, creating incentives to defend prior decisions. Even where a third party is commissioned, the evaluation questions, sample frame, and document checklist are often still determined by the purchasing authority, and the evaluator may enter only once near the end of the project. Service-user participation is frequently reduced to an end-of-project satisfaction questionnaire. When the provider supplies the contact list, users who are easiest to reach, participated most often, and have favorable relations with staff are more likely to be selected, while applicants who failed to obtain service, participants who withdrew, complainants, and eligible persons whom the project should have reached but did not are often excluded. A questionnaire that asks only whether staff were friendly or whether an activity was satisfactory cannot establish whether service was timely, convenient, or effective in resolving actual difficulties, even when it produces a high score. Research on public participation shows that participation becomes effective only when citizens can form and express preferences and purchasing authorities respond to them; the mere presence of a participatory form is insufficient[8].

The operating practices of third-party evaluation can further amplify information bias. Under limited time and budget, some evaluators must complete document review, on-site verification, surveys, and comprehensive scoring in a single exercise. To control costs, they may rely on a standardized routine of office-based file review, a small number of telephone callbacks, and one site visit. Similar indicator templates and questionnaires may then be applied to highly differentiated fields such as eldercare, psychological services, legal aid, and employment assistance. Evaluators without sector-specific knowledge may be unable to determine whether service records meet professional standards or to detect problems such as records without substantive service, repeated counting of users, or the exclusion of high-risk cases. At the same time, reports often revolve around an aggregate contract-completion score. To avoid tension within the commissioning relationship, ratings may cluster in the "good" or "relatively good" range, while findings are confined to low-risk phrases such as "documentation could be further improved" or "satisfaction requires enhancement." Studies of third-party performance evaluation show that evaluations of similar programs do not necessarily yield consistent or reliable results[9], and that the evaluator's structural position, organizational autonomy, and sources of authority shape its capacity for independent judgment[10].

The difficulty of incorporating service-user views also arises from participation barriers and the risks associated with speaking candidly. Older people, persons with disabilities, and disadvantaged households may face obstacles in communication, mobility, and digital access. Some users depend on the provider for continuing support and may fear that negative feedback will affect future service, producing courtesy-biased high scores. Satisfaction is also shaped by personal expectations, the receipt of benefits, and staff demeanor, and therefore cannot be equated directly with professional quality or problem resolution. Simply increasing the number of questionnaires will not automatically improve evaluation quality. Co-production theory emphasizes that service users are not passive recipients; they possess irreplaceable information about changing needs, accessibility, and actual effects[17]. Evaluation should therefore include randomly selected registered users, unsuccessful applicants, participants who withdrew, complainants, and, where appropriate, family members or guardians, and should triangulate their views against administrative data, on-site observation, and professional judgment. The core difficulty is that purchasing authorities and providers retain control over project definition, document supply, and the dominant interpretation of results, while service users' experiential knowledge has not yet been converted into stable and verifiable evaluation evidence.

3.3 Weak Use of Evaluation Results and Limited Performance Accountability

Weak use of evaluation results is not merely an abstract tendency to "emphasize evaluation but neglect application"; it can be observed concretely in payment timing, budget cycles, and provider management. Project expenses are commonly paid monthly or quarterly on the basis of invoices, service records, or interim acceptance, whereas performance evaluation is scheduled at year-end or after project completion. By the time the report is issued, the final payment may already have been made, leaving the evaluation unable to affect payment under the current contract. Preparation of the following year's budget and procurement may also begin in advance. If the prior year's evaluation arrives after budget submission or project approval, the project scale and contract terms have largely been fixed and the findings can serve only as archival material. Some projects assign ratings of excellent, good, qualified, or unqualified without specifying the corrective actions, payment deductions, restrictions on future contracting, or exit measures attached to each grade. Scores also tend to cluster at the upper end, making it difficult to translate the difference between 90 and 85 points into distinct managerial actions. Whole-process performance research stresses that evaluation should connect ex ante objectives, in-process monitoring, and ex post use[5], yet misaligned timing and missing rules cause the chain to break at the results stage.

The direct cause of this disconnection is that contracts and organizational procedures do not provide an operational interface for performance information. Many contracts specify service quantities, duration, invoicing, and payment schedules in detail but do not state which quality or outcome indicators are conditions of payment, what evidence can trigger payment suspension, verification of corrective action, or termination, or how responsibility is to be determined. When an evaluation finds weak outcomes, the purchasing authority may lack a clear contractual basis for reducing payment; paying in full, however, undermines the authority of evaluation. In organizational terms, program units oversee implementation, finance staff process payments, procurement staff manage contracts, and budget staff determine future funding. Evaluation reports often circulate among these units only as a procedural formality. The reports themselves may provide little more than an aggregate score and general recommendations, without distinguishing whether a problem arose from provider performance, purchaser design, or changes in the external environment, and without specifying severity, responsible actor, deadline, or the contractual clause involved. Decision makers are therefore unable to convert "low performance" into a concrete response. Research on the use of performance information shows that information enters decision making only when organizational motivation, institutional arrangements, and information quality support its use; the number of evaluation reports alone is insufficient[18].

Weak results use is also reflected in the absence of corrective verification, the limited sharing of low-performance records, and institutional inertia surrounding project exit. After an evaluation identifies deficiencies, a written explanation from the provider may be treated as sufficient evidence that remediation is complete, while the purchasing authority seldom conducts renewed sampling, callbacks, or data checks. The same problem may therefore recur in the following year. Poor performance in one project is rarely consolidated into a unified provider record and may not be available to other program units or subsequent procurement panels, preventing evaluation from creating a cross-project reputational constraint. For continuing eldercare, disability, and social-assistance programs, purchasing authorities may also fear that withholding funds, terminating a contract, or replacing a provider will disrupt service continuity. They consequently tend to preserve existing arrangements even when performance is weak. Case research shows that the effectiveness of government-purchased public services depends jointly on project design, oversight, and interactions among actors[6]. Results use therefore cannot be reduced to automatic deductions based on scores, but neither should service continuity justify indefinite inaction. What is needed is a graded response system grounded in contract, capable of allocating responsibility, requiring verified remediation, and feeding into the next budget and procurement cycle.

4 PATHWAYS FOR IMPROVING PERFORMANCE EVALUATION MECHANISMS

4.1 Improve Performance Objectives and Indicator Design

Improving indicator design should begin with needs identification before project approval rather than with the retroactive preparation of a scoring sheet at project closeout. Before launching a project, the purchasing authority should use questionnaires, consultation meetings, home visits, complaint data, administrative files, and existing service data to determine the size, distribution, and differentiated needs of the target population, identify gaps in coverage and weaknesses in service quality, and answer five questions: what public problem is the project intended to address, whom will it serve, through what activities, with what outputs, and with what expected changes? Needs identification and purchasing-management capacity are central elements of the "smart purchaser"[4]. On this basis, a logical chain can be constructed linking the public problem, target users, project activities, service outputs, service quality, and expected outcomes, together with verifiable baselines, interim targets, and final targets. Where outcomes emerge only over a long period or are strongly influenced by external factors, unrealistic causal commitments should be avoided. Intermediate outcomes, changes experienced by service users, and reasonable proxy indicators may instead be used, with the boundary of the project's contribution made explicit. This approach prevents both vague objectives and the attribution of broad social outcomes entirely to a single purchasing project.

Performance objectives should also be substantively embedded in procurement requirements and purchase contracts so that budget objectives, procurement documents, contractual clauses, and evaluation indicators are designed from a common source. Whole-process performance evaluation requires ex ante objectives, in-process monitoring, and ex post evaluation to be mutually connected[5], while contract-management research shows that quality standards create stable constraints only when they are translated into monitorable and adjustable contractual obligations[11]. In addition to service quantities, duration, and funding, contracts should specify rules for identifying target users, service coverage, staff qualifications, quality standards, timeliness, risk response, data reporting, satisfaction surveys, evaluation methods, and the use of results. Multi-year or complex projects should include midterm monitoring and stage-acceptance points so that problems can be corrected while adjustment remains possible. Indicator systems need not be excessively elaborate. A basic framework can be organized around project management, service outputs, service quality, and service outcomes. Management and output indicators verify whether a project was genuinely and properly implemented, while quality and outcome indicators determine whether it generated public value. Table 1 presents this structure.

Table 1 A Basic Indicator Framework for Performance Evaluation of Government-Purchased Public Services

Indicator category	Main evaluation content	Evaluation focus
Project management indicators	Project progress, fund use, staffing, contract performance, and risk control	Whether implementation is standardized and orderly
Service output indicators	Number of users served, service frequency, coverage, and completion time	Whether the basic outputs specified in the contract have been delivered
Service quality indicators	Standardization, timeliness, professionalism, continuity, safety, and accessibility	Whether the agreed quality standards have been met
Service outcome indicators	Problem resolution, actual benefits, equitable coverage, satisfaction, and sustained effects	Whether the project has generated the expected public value

Indicator weights and measurement methods should follow a classification principle combining common indicators with project-specific indicators. Research on indicator systems shows that public service evaluation must cover management, outputs, quality, and outcomes[7], while the varying purposes of performance measurement require different choices among accountability, control, and learning[13]. After necessary verification of expenditure, schedule, and quantity, greater weight should be given to timeliness, professionalism, continuity, accessibility, equitable coverage, problem resolution, and actual user benefits. Eldercare projects may emphasize coverage of high-need older people, response times, and improvements in daily convenience; employment services may focus on job matching and employment stability; and legal services may examine the professional quality of advice and the effectiveness of rights remedies. A single template should not be used across all project types. Where outcome indicators cannot be fully quantified, administrative data, random callbacks, on-site observation, case records, complaint analysis, and expert judgment should be combined so that multiple sources of evidence can offset the bias of any single data source. The number of indicators should also be proportionate to project scale and risk, preventing an overly complex evaluation system and excessive data collection from diverting resources away from service delivery.

4.2 Establish a Multi-Actor Participatory Evaluation Mechanism

Pluralistic evaluation must begin with the purchasing authority genuinely discharging its responsibilities. Commissioning a third party can supplement professional capacity and provide an external perspective, but it cannot transfer the purchaser's responsibility for objective setting, contract oversight, information verification, and results use. Whether public participation improves purchasing performance depends on the purchasing authority's capacity to absorb and respond to the needs expressed by service users[8]. The purchasing authority should define the purpose and core questions of evaluation, organize needs assessment, review indicators and sampling plans, provide complete administrative data, coordinate service-user research, examine the evidence supporting the report, and ensure that findings lead to remediation. To reduce role conflict, project implementation, contract acceptance, and performance evaluation can be separated internally where feasible. For projects involving substantial expenditure, high professional complexity, or broad social impact, staff from program, finance, procurement, and oversight functions, together with sector experts, should participate jointly. Evaluation reports should examine both provider performance and the purchaser's project management. Where low performance results from poorly designed objectives, inadequate interdepartmental coordination, or the purchaser's failure to provide required support, responsibility should not be shifted entirely to the provider.

Service-user participation should extend across needs assessment, process monitoring, and outcome evaluation rather than being confined to a single satisfaction questionnaire at project closeout. Co-production research shows that service users provide irreplaceable information in defining needs, delivering services, and judging outcomes[17]. Purchasing authorities should draw random samples from complete service rosters or administrative systems and combine telephone callbacks, face-to-face interviews, service-user forums, complaint analysis, follow-up of key cases, and assisted evaluation by family members or guardians. The sample should not be supplied exclusively by the provider. Questions should go beyond staff attitude to examine whether service was easy to obtain, waiting and response times were reasonable, personal information was protected, service was continuous, practical difficulties were alleviated, and users would choose to continue receiving it. Individuals who did not use the service, withdrew midway, or submitted complaints should also be studied, because these "silent cases" often reveal problems of accessibility and selective targeting. For older people, persons with disabilities, and disadvantaged households, paper, telephone, home-visit, and accessible communication channels should be available, and anonymity and the absence of provider staff during interviews should be used to reduce fear of speaking candidly.

Third-party evaluation should be governed by institutional constraints centered on independence, professional competence, and reviewability. Existing studies show that measurement quality, organizational position, and autonomy all shape third-party conclusions[9-10]; Ministry of Finance rules likewise require standardized selection of evaluators, conflict-of-interest avoidance, quality control, and management of commissioning contracts[19-20]. An institution that participated in the design or delivery of the same project, served as its procurement agent or acceptance consultant, or has a material interest in the provider should, in principle, not evaluate that project. Evaluation teams should include sector specialists suited to fields such as eldercare, employment, legal services, and psychological support rather than relying exclusively on staff with financial backgrounds. The commissioning contract should specify the evaluation scope, sample formation, methods and procedures, evidence retention, report review, confidentiality, and liability for fraudulent or misleading evaluation. Evaluation should also be differentiated by project type. Small, standardized, low-risk projects may be subject to a simplified evaluation organized by the purchasing authority, whereas major, complex, or socially sensitive projects should involve third-party institutions, sector experts, and more rigorous sampling. This can balance evaluation quality against evaluation cost.

4.3 Strengthen the Practical Use of Evaluation Results

The linkage between evaluation results and payment should rest on contractual provisions and evidentiary rules. The requirements for standardized commissioning contracts in budget performance evaluation also provide an institutional reference for defining evaluation scope, evidentiary standards, and accountability boundaries[20]. Purchase contracts may divide payment into an advance, milestone payments, and a final payment, with part of the amount linked to interim targets, service quality, and verified remediation. Minor deficiencies may trigger time-limited correction, with payment released after successful review. Where a clearly agreed key performance target is not met and responsibility is attributable to the provider, the purchasing authority may, in accordance with the contract, suspend payment, reduce the corresponding fee, or require the service to be re-performed. Fraudulent service delivery, major safety risks, and serious breach should be handled further under the contract and applicable rules. Poor performance, however, does not necessarily constitute breach of contract. Project outcomes may be affected by changing needs, external policies, service users' baseline conditions, or insufficient cooperation from the purchasing authority. Results should therefore be applied only after distinguishing provider responsibility, purchaser design responsibility, and uncontrollable factors, and providers should be given opportunities to explain, contest findings, and seek dispute resolution. A general score should not become a basis for arbitrary deductions.

Evaluation results should also inform provider selection and budget allocation, but their use must not become mechanical. Both comprehensive budget performance management and the government service-purchasing system require stronger links between performance results, budgets, and project management[1-2]. For providers with stable contract performance, strong professional quality, and favorable service-user assessments, past performance may serve as an important reference in drafting future procurement requirements, qualification review, and bid evaluation. Providers with repeated poor evaluations, falsified records, or uncorrected deficiencies should face fewer future contracting opportunities or contract termination in accordance with law and applicable rules. Excellent performance does not justify bypassing procurement and fair-competition procedures through automatic renewal, and performance records should distinguish ordinary improvement needs from serious misconduct. In budgeting, projects that address genuine needs, perform well, and require continuity should receive reasonable and stable budgetary support; projects with average performance should have their target population, service content, or delivery method adjusted; and projects that remain inefficient, address limited needs, or duplicate other programs may be reduced, merged, or discontinued. Basic services for disadvantaged groups may carry high unit costs, so funding reductions should not be based solely on cost or aggregate scores; equity, statutory obligations, and service continuity must also be considered.

Finally, a traceable system of remediation and disclosure should make performance information an input into the next round of project design. Performance information moves from "information production" to "managerial use" only when designated officials receive it and translate it into changes in budgets, contracts, or procedures[18]. After evaluation, the purchasing authority should prepare lists of problems, responsible actors, and deadlines and manage a closed process of issue feedback, provider remediation, purchaser verification, and results recording. Verification should not rely only on a written explanation; sampling, callbacks, and data comparison should be used to determine whether the problem has actually been resolved. Recurrent systemic problems should lead to changes in needs assessment, contract clauses, and management procedures rather than the repetition of identical recommendations each year. For projects serving the public, the project name, fiscal input, evaluation conclusion, major problems, and remediation status may be disclosed in accordance with law, with personal privacy, trade secrets, and information concerning vulnerable groups anonymized. Internally, purchasing authorities should establish procedures for circulating performance reports among program, finance, procurement, and budget units and maintain a project performance ledger specifying who receives each finding, who acts on it, and when feedback is due. These arrangements can shift performance evaluation from "report production" toward "decision use."

5 CONCLUSION

The fundamental purpose of performance evaluation for government-purchased public services is not to add another layer of project assessment or to replace substantive public-management judgment with complex indicators. It is to determine whether public expenditure has been converted into public services that citizens can experience, that are sustainable, and that are equitably accessible. Current difficulties are concentrated in the disconnection between performance objectives and genuine needs, an overemphasis on quantities and documentary traces, the concentration of evaluation information in purchasing authorities and providers, and the failure of findings to enter payment, contract, and budget decisions effectively. These problems reflect an institutional rupture among needs assessment, contract design, the allocation of evaluative authority, and the use of performance information. Reform should therefore orient evaluation toward public value and actual user benefits, establish a coherent foundation linking needs, objectives, contracts, and indicators, create a pluralistic structure in which purchasing authorities remain responsible, service users participate effectively, and third parties operate under clear professional rules, and connect evaluation findings to payment, provider management, budget adjustment, verified remediation, and information disclosure. Only when evaluation supports both accountability and learning, and examines service effects as well as project completion, can government purchasing of public services improve both service quality and the efficiency of public expenditure.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Central Committee of the Communist Party of China, State Council. Opinions on Fully Implementing Budget Performance Management. 2018.

[2] Ministry of Finance. Measures for the Administration of Government Purchase of Services: Order No. 102 of the Ministry of Finance. 2020.

[3] Wang PQ. An Analysis of the Reform Mechanism of Government Purchasing Public Services from Social Forces. Journal of Peking University (Philosophy and Social Sciences), 2015, 52(4): 88-94.

[4] Zhan GB. Demand-Side Defects, Supply-Side Defects, and the Smart Purchaser: Dilemmas and Solutions in Government Purchasing of Public Services. Comparative Economic & Social Systems, 2013(5): 142-150.

[5] Jiang AH, Yang Q. Performance Evaluation across the Entire Process of Government Purchasing of Public Services. Journal of Central University of Finance & Economics, 2020(3): 3-9.

[6] Han QY, Sun T. The Effectiveness of Government-Purchased Public Services and Its Determinants: An Exploratory Study Based on 153 Cases. Journal of Public Management, 2019, 16(3): 62-72, 171.

[7] Fan WF, Xu Y. Constructing an Evaluation Indicator System for Government Purchasing of Public Services from Social Forces. Scientific Decision Making, 2020(5): 1-24.

[8] Fang GY, Shao JS, Jin X, et al. How Does Public Participation Affect the Effectiveness of Government-Purchased Public Services? A Case Analysis of Government Service-Purchasing Projects. Chinese Public Administration, 2022(4): 36-46.

[9] Ma L, Yu WX. Evaluating Third-Party Performance Evaluation of Public Services: A Comparative Case Study. Nanjing Journal of Social Sciences, 2013(5): 55-63.
[10] Xu XG, Huang Y. Separation and Solidarity between Government and Social Organizations: Risks and Governance of Third-Party Evaluation in Government-Purchased Social Services - Evidence from City S. Social Work and Management, 2017, 17(2): 70-77.
[11] Brown TL, Potoski M. Contract-Management Capacity in Municipal and County Governments. Public Administration Review, 2003, 63(2): 153-164.
[12] Stoker G. Public Value Management: A New Narrative for Networked Governance?. The American Review of Public Administration, 2006, 36(1): 41-57.
[13] Behn RD. Why Measure Performance? Different Purposes Require Different Measures. Public Administration Review, 2003, 63(5): 586-606.
[14] Zhang X. Formation Mechanisms and Remedies of Performance Dilemmas in Government-Purchased Public Services: The Case of Self-Service Libraries in District S. Journal of the Fujian Provincial Committee Party School of the CPC, 2019(1): 67-72.
[15] Jiang XP, Kang J. Bureaucratic Outsourcing: Stakeholders' Action Logics and Their Effects on Performance in Government Purchasing of Public Services. Administrative Tribune, 2019, 26(4): 63-68.
[16] Chen JG. Effects of Project Attributes and Process Management on the Performance of Government Service-Purchasing Projects: A Fuzzy-Set Qualitative Comparative Analysis of 20 Cases. Journal of North China Electric Power University (Social Sciences), 2022(1): 43-54.
[17] Bovaird T. Beyond Engagement and Participation: User and Community Coproduction of Public Services. Public Administration Review, 2007, 67(5): 846-860.
[18] Ma L. The Use of Government Performance Information: Theoretical Integration, Literature Review, and Research Prospects. Journal of University of Electronic Science and Technology of China (Social Sciences Edition), 2014, 16(5): 1-11.
[19] Ministry of Finance. Guiding Opinions on Commissioning Third-Party Institutions to Participate in Budget Performance Management: Cai Yu [2021] No. 6. 2021.
[20] General Office of the Ministry of Finance. Operational Guidelines for the Management of Commission Contracts for Budget Performance Evaluation (Trial): Cai Ban Jian [2024] No. 67. 2024.