ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖӘНЕ ҚҰҚЫҚ ҒЫЛЫМДАРЫ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИЕ И ЮРИДИЧЕСКИЕ НАУКИ
SOCIO-ECONOMIC AND LEGAL SCIENCES

CENTRAL ASIAN SCIENTIFIC JOURNAL
ISSN 3135-3061
2026. No. 4 (32), Vol. 2. pp. 191–200
DOI: https://doi.org/10.5281/zenodo.24272894
Journal website: https://cajournal.kz

UDC 336.71:330.34

MEASURING THE ADDITIONALITY OF NATIONAL DEVELOPMENT BANKS: A CONCEPTUAL FRAMEWORK FOR EMERGING ECONOMIES

Buketova Aigul
Development Bank of Kazakhstan, Astana, Kazakhstan

Buketova Aigul — Development Bank of Kazakhstan, Astana, Kazakhstan.

© Buketova Aigul, 2026

Abstract: National development banks are often assessed by the volume of loans they provide, even though disbursement alone does not show what economic activity occurred because of public support. Drawing on the harmonized additionality framework of multilateral development banks, research on national development banks and counterfactual evaluation methods, this article proposes a layered framework that distinguishes financial, non-financial, firm-level and systemic additionality. Each dimension is linked to observable indicators, feasible counterfactual strategies and minimum data requirements. The article addresses practical problems arising from heterogeneous projects, small samples of large borrowers, selection bias, delayed development effects and public cost, and proposes reporting labels that separate documented, plausibly causal and causally estimated claims. The framework is designed as a research agenda rather than an empirical evaluation and may be adapted to resource-dependent emerging economies.

Keywords: additionality; counterfactual evaluation; development banks; causal inference; emerging economies; impact measurement

Introduction

A development bank can expand its loan portfolio without increasing aggregate investment. When public finance replaces a commercial loan that was already available, the intervention changes the source or cost of funding rather than the existence of the activity. Conversely, a comparatively small intervention can have significant effects when it completes an otherwise unfinanceable project or attracts new lenders. Evaluation should therefore ask what happened because of the institution, not only what happened after its participation.

The proposed framework begins with a counterfactual principle: additionality is a difference between an observed outcome under public intervention and a credible estimate of the outcome without that intervention. The potential-outcomes formulation is useful even when exact estimation is impossible. For project i, additionality for an outcome Y can be described as Y_i(1) minus Y_i(0), where Y_i(1) is the outcome with public support and Y_i(0) the outcome without it. Only one is observed; the other requires assumptions and evidence.

This paper develops a practical typology based on the harmonized multilateral development bank framework [5], research on national development banks [4; 6] and established empirical evaluation logic. It does not claim that multilateral rules automatically govern national institutions or that any particular bank meets the proposed standards.

The article proceeds as follows. The literature review summarizes the concepts of additionality used by multilateral and national development finance institutions and the empirical research on which the framework draws. The methodology section explains how the framework was

constructed. The results present indicators, a measurement architecture for different instruments, counterfactual strategies, data requirements and an evaluation protocol, and the discussion examines the application of the framework in emerging economies and its limitations.

Literature Review

The concept of additionality has developed mainly in the practice of multilateral development banks, whose harmonized framework distinguishes financial from non-financial contributions and asks institutions to document why their participation is needed [5]. Evaluation syntheses based on this practice draw attention to the strength of the evidence behind additionality claims and to the difference between additionality expected at approval and additionality demonstrated after implementation [1].

Research on national development banks adds two further insights. Descriptive surveys show that national institutions differ widely in mandate, funding and operating model, so an evaluation framework must be adaptable rather than uniform [4]. Firm-level evidence shows that public support can change financing conditions without changing real investment: in the Brazilian case, subsidized development-bank loans reduced financial expenses but did not consistently raise investment or performance [6].

Studies of credit guarantees and small-business lending show how research design shapes what can be learned. Exploiting variation in borrowers' access to lenders experienced with a public loan program, one influential study estimates employment effects of guaranteed credit in the United States [3]. Comparative work on partial credit guarantee schemes shows that coverage, pricing, eligibility and claims rules differ widely, which suggests that design features may explain differences in effects across programs [2]. Building on these strands, the framework developed here organizes additionality into four layers.

The first layer is financial additionality at transaction level [5]. The evaluator asks whether public support supplied a needed amount, tenor, currency, risk-sharing arrangement or flexible repayment profile unavailable on reasonable market terms. Evidence may include lender term sheets, declined applications, independent pricing benchmarks and the financing plan at decision time. Financial additionality may be strong even when the nominal loan interest rate is not subsidized.

The second layer is non-financial additionality. Public institutions may improve environmental safeguards, governance standards, project preparation, procurement or disclosure. These contributions should be documented through explicit requirements, milestones and demonstrable changes in borrower behavior. A claim that expertise was provided is insufficient if no observable practice, capability or standard changed.

The third layer is firm or project additionality. Relevant outcomes include probability of project initiation, construction completion, fixed investment, productivity, employment, exports and technological upgrading. Evaluators need to distinguish a change in financing conditions from a resulting change in economic activity. This distinction also protects against double-counting a single project as both mobilized financing and separately created investment.

The fourth layer is systemic additionality. A program may influence lending practices, crowd in new investors, establish demonstration effects or support supplier networks. System-wide benefits are often diffuse, delayed and difficult to assign to one project. They require a theory of change and evidence beyond the directly financed client.

Methodology

This article follows a conceptual and methodological research design. It synthesizes three bodies of work: the harmonized framework for additionality adopted by multilateral development banks [5], institutional and empirical research on national development banks and credit guarantees [2–4; 6] together with evaluation syntheses [1], and the general logic of counterfactual impact evaluation. Rather than estimating effects, the article derives a typology of additionality claims, links each claim to observable indicators and identifies the minimum evidence required to support it.

© Central Asian Scientific Journal, 2026

The framework was developed in three steps. First, definitions of additionality in the selected sources were compared in order to separate transaction-level, firm-level and system-level claims. Second, each claim was matched to the financing instruments through which it could arise, including direct lending, guarantees, subsidies and co-financing. Third, standard identification strategies were assessed against the practical conditions of national development banks, such as small numbers of large borrowers, discretionary selection and long project horizons. The resulting framework is normative and has not been validated on institutional data.

The scope of the framework is limited to the evaluation of financing operations of national development banks and related public instruments, such as guarantees and interest-rate subsidies. It does not address the broader macroeconomic effects of development banking, the optimal size of public financial institutions or the political economy of credit allocation. These questions are important, but they require different methods and data. The framework is intended to be applicable with information that an institution can reasonably collect in the course of appraisal, monitoring and ex post review.

Results

A useful scorecard separates inputs, mechanisms, intermediate outcomes and final outcomes. The amount of a public loan is an input. A longer maturity or a lower collateral requirement is a mechanism. Project commencement is an intermediate outcome. Sustained additional productive capacity is a potential final outcome. Indicators should be chosen according to the market failure invoked at approval, not retrospectively selected to make the project appear successful.

Financial indicators include changes in maturity, pricing, financing volume, risk-sharing and the number of genuinely new private financiers. Project indicators include actual capital expenditures relative to planned and counterfactual investment, time to completion, output, export activity and debt-service resilience. Non-financial indicators include independently verified adoption of governance, environmental or technical practices. System indicators may include sector lending depth, entry of new financiers and spillovers to suppliers.

These indicators must be interpreted with denominators and baselines. A percentage of private co-financing in the final funding package measures participation, not necessarily additional private money. Similarly, counting new jobs without identifying displacement or the economic cycle may overstate the net effect. Where quantitative measurement is weak, transparent qualitative evidence is preferable to a false numerical precision.

A measurement system should begin with the nature of the financing instrument. A direct project loan may be additional because of maturity, financing scale or currency. A partial guarantee may be additional because it changes the lender’s willingness to accept borrower risk. A subsidy may be additional because the price of capital determines investment feasibility. A co-investment arrangement may be additional by resolving a coordination problem. Each case requires its own observable mechanism and alternative scenario.

The distinction between a financing input and a developmental outcome can be operationalized in a sequence. First, the institution records the binding constraint alleged at approval. Second, it documents how the proposed financing terms address that constraint. Third, it identifies an expected behavioral response, such as initiating construction or expanding equipment purchases. Fourth, it records outcomes over an appropriate horizon. Finally, an evaluator considers which changes are plausibly attributable to the intervention.

For direct lending, relevant baseline information includes other available lender offers, maturity requirements, capital expenditure plans and the cost of financing. Evidence should be dated before approval to avoid post hoc reconstruction of a financing gap. A firm’s statement that it needed public support is useful context but not sufficient by itself; it may have an incentive to portray financing as indispensable. Independent term sheets, documented rejections and comparable transactions can strengthen the inference.

For guarantees, the relevant counterfactual may be a bank’s lending decision without risk transfer [2]. The assessment should therefore observe the loan approval margin, interest rate,

© Central Asian Scientific Journal, 2026

collateral, amount and tenor, as well as the guarantee share and residual risk retained by the lender. Subsequent firm investment must be measured separately. A higher number of guaranteed loans can coexist with weak real additionality if lenders simply substitute guaranteed credit for ordinary facilities.

For non-financial contributions, observable behavioral change is central. A bank may require a stronger procurement process, environmental management system or governance policy. The evaluation should specify the original practice, the imposed requirement, compliance milestones and independent verification. Merely attaching a checklist to a loan does not show that the borrower acquired lasting capability. The duration and persistence of change may matter more than initial formal adoption.

System effects require yet another unit of measurement. New commercial lenders may enter a sector after observing a successful transaction, but their entry could reflect commodity prices or macroeconomic reforms. A credible systemic claim would triangulate market-level data, lender interviews, timing and alternative explanations. Where evidence is incomplete, the bank can report a plausible contribution and its uncertainty without asserting a measured causal impact.

For interest-rate subsidies, the relevant question is whether the price of capital was the binding constraint. Evidence may include the project financial model at market and subsidized rates, the sensitivity of the investment decision to financing cost and the timing of the subsidy relative to the investment decision. For co-financing, the evaluator should record which financiers were committed before the public institution joined, which joined afterwards and on what terms. Only the latter group can plausibly be counted as mobilized, and even then a counterfactual is needed to determine whether their participation depended on the public institution.

The central identification threat is the non-random selection of supported projects. Credit analysts are expected to choose projects based on viability and risk, while policymakers may prioritize strategic sectors or disadvantaged regions. Both selection criteria correlate with future outcomes. Even a rich regression model cannot remove bias from unobserved managerial quality, political uncertainty or business expectations unless its design supplies additional identifying variation.

Matching methods compare recipients with observationally similar untreated units on variables measured before financing. They can reduce imbalance in observable firm size, leverage, sector, region and growth trajectories. The method is credible only when the necessary selection variables are observed and common support exists. Large industrial projects often lack close comparators, so a superficially matched control may be economically incomparable despite a similar statistical score.

Difference-in-differences uses changes for supported and comparison units. Its identifying assumption is not simply that the groups differ by a constant amount; absent treatment, their trends should have evolved similarly. Analysts should inspect pre-treatment trends and consider treatment anticipation, staggered adoption and sector-specific shocks. Where the timing of project approval responds to changing performance, dynamic treatment selection can undermine a simple before-and-after comparison.

A regression-discontinuity design may be feasible if a transparent eligibility score or threshold determines access to support. It compares observations near the boundary, where projects may be similar except for treatment probability. The resulting estimate is local to that threshold and may not describe the large projects most central to development-bank portfolios. Manipulation of the score or discretionary overrides would also complicate interpretation.

Instrumental-variable strategies use variation in treatment exposure that is relevant to financing but otherwise unrelated to outcomes. An example in credit research is variation in the local presence of lenders experienced with a program [3]. However, lender presence may itself be correlated with regional growth prospects. An instrument therefore requires a specific and defended exclusion restriction, not merely a strong first-stage relationship.

A credible design for very large unique projects may instead be a transparent counterfactual case assessment. Analysts can reconstruct the financing package available at decision time, identify

the binding requirement, consult independent market benchmarks and examine whether the project changed scope or timing after intervention. This yields richer mechanism evidence than a poorly powered regression. Its limits should be stated: the conclusion is case-specific and may depend on imperfect reconstruction of alternative offers.

Mixed-method triangulation can improve confidence. Quantitative outcome series establish what changed; original records show what decision was contemplated; interviews clarify mechanisms; and external market information tests alternative explanations. Triangulation is not a substitute for identification, but it helps reveal whether a causal story is coherent. Reporting contradictory evidence is as important as reporting mutually supportive observations.

A mixed portfolio evaluation can therefore use different evidence standards. Numerous comparable SME loans may support quasi-experimental designs; a few unique industrial projects may require counterfactual financial modeling, external benchmarking and independently audited qualitative assessments. Results should be reported separately before any aggregation.

Data collection should begin at application rather than disbursement. Each application should ideally preserve the project's objective, estimated financing gap, terms offered by other lenders, reason for intervention, pre-treatment financial indicators and expected development outcomes. Rejected and withdrawn applications are particularly valuable for constructing comparison groups, although their use requires strong privacy protection and attention to selection.

At approval, an evaluator should record the pre-specified additionality claim and the evidence available to the committee. During implementation, the system should track changes to the financing structure, investment milestones, borrower outcomes and any private-financier participation. Follow-up should extend beyond the first disbursement and distinguish construction from operation. A timestamped data dictionary helps prevent later redefinition of measures.

Public reporting can aggregate results without disclosing individual borrowers. Data governance should include access controls, clear separation between operational underwriting and independent evaluation, and a policy for correcting missing or inconsistent observations. Where some data cannot be collected, the limitation should be stated rather than filled by assumptions.

Stage one is mandate alignment: articulate the specific public problem, beneficiary, instrument and development objective before approval. Stage two is the financing counterfactual: document market options and identify what feature of the public intervention is indispensable. Stage three is the monitoring plan: select a small set of outcomes, baseline measures, time points and plausible comparison units. Stage four is independent ex post assessment: distinguish observed outputs from causal effects and disclose uncertainty.

A disciplined decision rule can classify claims by evidence strength. "Documented" means that a financing or non-financial contribution is verifiable from contemporaneous records. "Plausibly causal" means that competing explanations have been considered in a transparent comparison or case study. "Causally estimated" should be reserved for a research design with defensible identifying assumptions and suitable data. These are proposed reporting labels, not official classifications of any institution.

The protocol also includes a public-cost dimension. Incremental investment may be valuable, but a net assessment must consider subsidy expense, losses, administrative cost and opportunity cost. This does not require every outcome to be reduced to a single monetary number; it does require a clear presentation of benefits and resource commitments.

Taken together, the results describe an evaluation chain with four links: the market failure invoked at approval, the financing or non-financial mechanism used to address it, the behavioral response of the borrower or other financiers, and the resulting change in economic activity and public cost. Each link has its own indicators and evidence standards, and a weakness at one link cannot be offset by strong evidence at another. A project with well-documented financing additionality but no information on investment outcomes, for example, supports a claim about financing terms but not about real additionality.

Table 1 brings the elements of the framework together by linking each type of additionality claim to a potential indicator and to the minimum evidence required to support it.

© Central Asian Scientific Journal, 2026

Table 1 – Additionality claims, indicators and minimum credible evidence

Additionality claim	Potential indicator	Minimum credible evidence
Amount additionality	Financing gap filled; change in feasible project scale.	Contemporaneous competing financing offers and documented funding requirements.
Maturity additionality	Tenor, grace period and repayment profile relative to alternatives.	Comparable market offers or independently defensible benchmarks.
Risk absorption	Residual private lender exposure and change in lending decision.	Guarantee terms, bank decision records and comparison with unsupported cases.
Currency additionality	Availability of local-currency financing where alternatives are foreign-currency.	Terms of available alternatives and the project's currency mismatch.
Non-financial expertise	Verified improvement in project design or operating practice.	Baseline, imposed requirement and independently checked implementation.
Project initiation	Probability and timing of construction start.	Eligible comparison projects or defensible documented counterfactual.
Incremental capital formation	Fixed investment attributable to support.	Longitudinal outcomes and a credible estimation strategy.
Private-capital mobilization	Incremental private commitments or disbursement.	Attribution rules; evidence private funds would not otherwise have participated.
Systemic spillovers	Entry of lenders, supplier upgrades or sector capability.	Market-level trend evidence with competing explanations examined.
Public cost	Subsidy, expected credit loss and administrative expense.	Transparent costing methods and separation of realized from contingent liabilities.

Discussion

Commodity cycles complicate the interpretation of development-finance effects. When global prices rise, both private bank lending and industrial investment may increase independent of public support. Exchange-rate swings can affect debt-service capacity and the currency composition of financing. A credible evaluation should specify these macroeconomic factors, particularly for firms with foreign-currency obligations or input imports.

Structural diversification also creates long time horizons. A manufacturing project may generate measurable exports only after construction, certification and supplier learning. At the same time, intended spillovers should not be counted before they occur. The framework therefore proposes a timeline of expected mechanisms, each linked to a feasible observation date and evidence type.

A related problem is concentration. A small number of expensive projects can dominate portfolio-level indicators and make averages unstable. Reporting project-level ranges, sensitivity tests and explicit case selection can be more informative than a single blended additionality score. The ultimate objective is an auditable account of incremental economic activity, not an artificial ranking of incomparable projects.

These considerations are particularly relevant for institutions whose portfolios consist of a limited number of large industrial and infrastructure projects. In such portfolios, conventional statistical comparison groups are often unavailable, individual projects can dominate aggregate indicators and the decisive evidence about financing alternatives is held in application-stage records. The framework therefore places greater weight on documented counterfactual reasoning, transparent case selection and the separation of financing contributions from real-economy outcomes than on portfolio-wide averages. This emphasis does not lower evidentiary standards but adapts them to the structure of the portfolio being evaluated.

Evidence can be classified by strength without assigning arbitrary numerical scores [1]. A documented claim shows that a mechanism or contract term existed. A plausible additionality claim shows that the mechanism appears important relative to considered alternatives. An estimated causal effect arises from a research design with defensible assumptions and relevant data. These statements convey different levels of knowledge; they should not be compressed into one composite percentage without a transparent aggregation rule.

Measurement error is especially problematic in private-finance mobilization. Public and private institutions can each claim the same project financing as mobilized, and a refinancing

transaction may be counted despite no new investment. A defensible reporting protocol should disclose source, attribution method, timing, gross and net definitions, and whether the private funds were committed or actually disbursed. The precise terminology should remain consistent across annual reports.

Uncertainty also derives from the outcome horizon. Capital expenditures can occur quickly, while production, exports and productivity respond only after commissioning. Long follow-up introduces attrition, ownership changes and macroeconomic shocks. Rather than using a single arbitrary end date, the evaluation should specify short, medium and long horizons based on the project theory of change, with planned measures assigned to each stage.

Heterogeneity makes aggregate averages difficult to interpret. The effect of a guarantee for a small firm is not comparable in scale or mechanism to a syndicated loan for major infrastructure. Aggregation may still be useful for portfolio management, but it should group reasonably homogeneous products and disclose the weighting rule. Total investment, number of projects and average additionality can each privilege different categories of activity.

Sensitivity analyses are essential where counterfactual assumptions are not testable. A case study might compare a conservative scenario in which commercial financing was readily available with a stringent scenario in which no reasonable private alternative existed. An observational study might vary matching variables or time windows. The purpose is not to find the scenario with the most favorable result, but to identify which conclusions remain robust to reasonable alternatives.

Data gaps should be visible. Missing rejected applications, incomplete pre-treatment financial statements or inconsistent project codes reduce the ability to infer additionality. A monitoring framework should report the share of projects with usable baseline records and the reasons information is missing. It is preferable to publish a narrower, clearly defined assessment than a portfolio-wide score based on undocumented extrapolation.

The framework also has implications for how development banks communicate their results. Public reports that present loan volumes, numbers of projects and gross mobilized finance alongside claims of development impact invite readers to treat these figures as evidence of additionality. A clearer approach would present financing outputs, evidence-graded additionality claims and public costs in separate sections, with the evidence label attached to each claim. Such reporting would make it easier for researchers, supervisors and the public to understand what an institution has demonstrated and what remains to be tested.

Institutional incentives also affect the quality of additionality evidence. When staff performance is judged mainly by approval and disbursement volumes, there is little reason to document financing alternatives carefully or to report weak outcomes. Separating appraisal from evaluation, rewarding complete documentation and treating negative findings as inputs to product design can change these incentives. Evaluation is therefore not only a technical exercise but also part of institutional governance.

A future evaluation should define treatment before examining outcomes. Approval of financing and actual disbursement are not identical interventions. The first may change expectations and project planning; the second provides liquidity. A clear definition determines when a project enters the treatment group, how delays are handled and what effect is estimated. Without this decision, observed improvements may be attributed to the wrong event.

The study population should also be fixed in advance. It could consist of applicants meeting an eligibility rule in a defined period. Withdrawn and rejected applications matter, because the experiences of successful borrowers alone do not reveal the untreated alternative. Excluding failed projects after the fact introduces survivorship bias. The plan must distinguish a new project entity from an established operating firm, since the availability of pre-treatment financial histories is different.

Primary outcomes should match the mechanism. If the purpose is to unlock construction, project initiation and actual capital expenditure are more direct outcomes than general revenue growth during the construction period. Secondary outcomes might include exports, employment and productivity after commissioning. Researchers should establish time horizons before seeing results

and explain deviations. An attractive coefficient at one arbitrary follow-up date is not a sufficient rationale for choosing it.

The comparison strategy must be justified. Matching requires credible observed selection variables and common support. Difference-in-differences requires a defensible comparison trend without treatment. Eligibility thresholds may support regression discontinuity, provided discretion and manipulation around the threshold are limited. Unique large projects may instead require a structured case-based counterfactual. No method should be presented as causal merely because it produces a numerical estimate.

A data-quality protocol should record missing lender offers, changes in ownership, incomplete financial statements and inconsistent project identifiers. Treating all missing outcomes as zero is not generally defensible. Sensitivity checks should show whether findings change with reasonable missing-data assumptions or alternate comparator sets. In public reporting, the share of supported projects with complete baseline and follow-up information is itself an important indicator of evaluation readiness.

Finally, the analytical plan should connect outcomes to public cost and ethics. Direct financing, an interest subsidy and a partial guarantee have different fiscal profiles. Benefits must be interpreted alongside expected loss, explicit transfers and administrative costs. Any later use of application records requires appropriate authorization, confidentiality safeguards and transparent reporting of methods. The present paper offers a protocol only; it makes no claim to have conducted such an empirical study.

The first implementation phase is institutional agreement on definitions. Credit, risk, strategy and evaluation teams should distinguish transaction additionality, real-economy outcomes and public costs in a shared data dictionary. A concise approval-time template can document the specific financing gap, alternatives sought and predicted development effects. The system should be designed around decisions already made in project appraisal to reduce reporting burdens.

The second phase is data linkage. Unique project and firm identifiers would connect applications, approvals, disbursements, changes in terms and periodic outcomes. It should preserve withdrawn and rejected applications when legally permissible, because they are analytically valuable comparators. Data access must be role-based and protect confidential financial information. Published evaluations can report aggregate results, anonymized methods and uncertainty without exposing clients.

The third phase is piloting evaluation designs in homogeneous product categories. A portfolio of many similar guarantees could support a quasi-experimental comparison, while a few strategic investment projects could receive independent structured case reviews. The purpose of piloting is to learn which indicators are actually obtainable, what their interpretation permits and where the burden of additional data collection exceeds its value.

The fourth phase is learning and feedback. Results should inform eligibility criteria, pricing, maturity design, guarantee coverage and the selection of outcome indicators. A finding that public finance improved terms without increasing investment might justify narrower targeting; a finding that non-financial assistance improved implementation could support complementary advisory services. A learning system should accommodate negative and mixed findings rather than rewarding only favorable narratives.

National development banks should also make the empirical boundaries of their claims explicit. A financing contribution is often directly observable, a counterfactual investment effect requires stronger inference, and aggregate diversification effects require still broader evidence. The conceptual framework does not eliminate these differences; it makes them visible. This transparency is itself a precondition for rigorous future research on development finance in emerging economies.

The proposed framework is normative and methodological rather than an empirically validated scoring instrument. Its categories draw on published additionality principles and general evaluation logic, but their practical feasibility depends on an institution’s mandate, data architecture and product mix. A national bank with a small number of unique borrowers may be unable to

© Central Asian Scientific Journal, 2026

generate credible conventional control groups. The framework therefore recommends evidence standards appropriate to the decision rather than pretending one estimator can serve every portfolio.

There is an unavoidable trade-off between analytical ambition and reporting burden. Collecting every alternative lender offer, detailed project milestones and long-run firm outcomes may be costly or legally restricted. The suggested protocol should be piloted with a limited set of high-value indicators and revised after the quality and cost of collection are observed. A shorter reliable dataset is preferable to an extensive questionnaire that produces inconsistent records and a misleading sense of numerical precision.

Causal conclusions remain contingent on assumptions. Matching cannot remove unobserved selection, difference-in-differences may fail under divergent trends, and case reconstruction may depend on records preserved by interested parties. Reporting these limits is not a weakness of evaluation; it protects public decision-making against exaggerated certainty. Sensitivity analyses, independent review and explicit classification of evidence strength can help stakeholders distinguish a documented financing contribution from a demonstrated effect on real economic activity.

Finally, an evaluation system should be judged partly by whether it changes institutional learning. Indicators may be technically correct yet unused in product redesign or strategic decisions. A governance process should assign responsibility for reviewing negative and mixed findings and for communicating revisions to program rules. The conceptual contribution of this article is a coherent architecture connecting market failure, financing mechanism, observable outcome and public cost; application and validation remain tasks for future empirical research.

Conclusion

Additionality is best treated as a sequence of testable claims: a market gap existed, a public instrument changed financing or practice, economic behavior changed, and developmental value followed. Not every project can support a credible impact estimate, but every project can document its stated mechanism and evidence limitations. A layered evaluation framework allows national development banks to learn from heterogeneity rather than obscure it, while providing a research foundation for future empirical studies in emerging economies.

The framework makes three contributions. First, it separates financial, non-financial, firm-level and systemic additionality and links each to an observable mechanism and a minimum standard of evidence. Second, it matches evaluation strategies to portfolio structure, recommending quasi-experimental designs for large numbers of comparable loans or guarantees and structured counterfactual case assessments for a small number of unique projects. Third, it requires that benefits be reported together with public cost and that claims be labeled by evidence strength rather than compressed into a single composite score.

The framework is a research agenda rather than a validated instrument. Its next test is practical: piloting the proposed indicators and data protocol on homogeneous product categories, measuring how much reliable information can be collected at reasonable cost and revising the framework in light of negative and mixed findings. Such piloting would provide the evidence base needed for credible empirical studies of development-bank additionality in Kazakhstan and other resource-dependent emerging economies.

References:

1. Additionality as a catalyst for change: insights from evaluation [Electronic resource] / European Bank for Reconstruction and Development. London : EBRD, 2025. URL: https://www.ebrd.com/home/news-and-events/publications/evaluation/ctd-additionality.html (accessed: 15.11.2026).

2. Beck T., Klapper L. F., Mendoza J. C. The typology of partial credit guarantee funds around the world : Policy Research Working Paper No. 4771 [Electronic resource]. Washington, DC : World Bank, 2008. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/937071468314370786 (accessed: 15.11.2026).

3. Brown J. D., Earle J. S. Finance and growth at the firm level: evidence from SBA loans // The Journal of Finance. 2017. Vol. 72, No. 3. P. 1039–1080. DOI: 10.1111/jofi.12492.

4. De Luna-Martínez J., Vicente C. L., Arshad A. B., Tatucu R., Song J. 2017 survey of national development banks [Electronic resource]. Washington, DC : World Bank, 2018. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/977821525438071799 (accessed: 15.11.2026).

5. Harmonized framework for additionality in private sector operations [Electronic resource] / Multilateral Development Banks. 2018. URL: https://www.ebrd.com/content/dam/ebrd_dxp/assets/pdfs/partnering-for-impact/harmonized-framework-additionality-2018.pdf (accessed: 15.11.2026).

6. Lazzarini S. G., Musacchio A., Bandeira-de-Mello R., Marcon R. What do state-owned development banks do? Evidence from BNDES, 2002–09 // World Development. 2015. Vol. 66. P. 237–253. DOI: 10.1016/j.worlddev.2014.08.016.

© Central Asian Scientific Journal, 2026