MDB REFORM ADVISORY

mdbreform.com | pbrar@mdbreform.com

THE MISSING MIRROR

Twenty-Five Years of World Bank Service Delivery Guidance — and the IEG Evidence It Does Not Cite

July 2026

EXECUTIVE SUMMARY

The World Bank's June 2026 synthesis paper — Supporting Service Delivery in Protracted Crisis and Violent Contexts: Navigating Tradeoffs (P505313) — is the most operationally sophisticated guidance the Bank has produced on service delivery in fragile settings. It is also a milestone in a twenty-five year trajectory, beginning with the Wappenhans Report (1992) and running through successive frameworks on governance, political economy, adaptive management, and fragility. Each generation has added analytical sophistication.

The research programme that produced this paper is itself substantial: Phase 1 delivered 290 pages across three documents — a 190-page synthesis, an 82-page evidence review covering 200+ peer-reviewed papers, and an 18-page practitioner operational note — supported by 37 structured interviews, a 42-respondent survey, a Washington workshop with over 100 participants, and country-level analytics across 26 PCV settings. Phase 2 is planned.

But the paper rests on an assumption it never tests. Its three-step framework — analyse the PCV context, calibrate objectives, adapt implementation — begins after the project has already been approved, the PDO set, and the financing committed. This paper argues that the IEG's own evaluation database, drawn from 10,453 evaluated projects including 1,432 in FCS-classified countries, identifies a different and more fundamental constraint: not the quality of implementation, but the quality of project design at entry.

KEY FINDINGS	Evidence
Service delivery in FCS is structurally weak — and flat across two decades	FCS countries achieved S+ outcomes 23.1% of the time across the five sectors this paper covers — 420 projects, $21.8bn committed, 2000–2026. Three in four dollars did not produce a Satisfactory outcome. The decade of peak analytical activity (2010–2019) produced the lowest S+ rate in the database: 19.2%.
Quality at Entry is one of the strongest predictors of outcomes in FCS settings	Projects with good QAE (S or HS) achieve S+ outcomes 61–81% of the time even in FCS environments. Projects rated Moderately Unsatisfactory at entry achieve 1.4%; Unsatisfactory or Highly Unsatisfactory achieve 0.0% across 77 projects and $2.8bn. Adaptive implementation cannot compensate for poor project architecture.
One in three service delivery projects in FCS enters with poor design	33.7% of FCS service delivery projects received MU or worse QAE ratings; in those 141 projects S+ outcomes = 0.0%. Over half of FCS projects (55.6%) have M&E frameworks rated Negligible or Modest — yet the PCV

	paper recommends adaptive management approaches that require a functioning monitoring system.
The Bank applies institutional diagnostics to borrowing governments but not to itself	The PCV paper diagnoses capacity gaps, misaligned incentives and weak accountability in borrowing governments — but applies none of these questions to the Bank itself. The IEG data show that TTL incentives, preparation timelines, supervision budgets and matrix management all predict project quality. They are absent from the PCV paper.
The programme's own costs are not publicly accessible	P505313's three Phase 1 deliverables are publicly disclosed on the Bank's Documents & Reports portal, each with standard metadata. None carries a budget, trust fund number, or any financial information. A lending project of equivalent scale would have a publicly disclosed PAD, a commitment amount, and a Finances One profile. The accountability asymmetry is structural.

The paper concludes that improving implementation guidance alone will not materially improve service delivery outcomes in fragile settings. It proposes four institutional reforms aimed at strengthening project design, evaluating knowledge products, measuring operational uptake, and increasing transparency.

I. The Report and Its Place in a Twenty-Five Year Trajectory

Supporting Service Delivery in Protracted Crisis and Violent Contexts: Navigating Tradeoffs (World Bank, June 2026, P505313) is a product of a two-year action-research programme, financed by a grant from the UK Foreign, Commonwealth and Development Office (FCDO), covering twenty-six countries classified as experiencing protracted crisis and violence. Its framework rests on three structural conditions — rivalry, fragmentation, and scarcity — and proposes a three-step decision tool: analyse context, calibrate objectives, and adjust project design to manage the trade-offs between effectiveness, do no harm, and potential to last.

P505313 is not an isolated publication. It forms part of a sustained World Bank knowledge architecture that has produced successive frameworks on service delivery, governance, political economy, adaptive management, and fragility over more than two decades. Each generation has added analytical sophistication. The table below maps that trajectory against a single outcome figure.

Year	Framework / Initiative	What It Added
1992	**Wappenhans Report**	Named the approval culture and weak project design as core institutional pathologies. Recommended structural reform of preparation incentives.
2004	**WDR: Making Services Work for Poor People**	Governance, accountability and citizen voice as the primary levers for improving service delivery. Launched a decade of decentralisation and social accountability programming.
2007–14	**Governance & Institutional Capacity Agenda**	Political economy, state capability and implementation capacity added to the analytical toolkit. Projects grew more complex; results frameworks longer; safeguards multiplied.
2011	**WDR: Conflict, Security and Development**	Recognised that fragility and conflict required a different development model. Security, justice and jobs as the entry points. Laid groundwork for the FCV Strategy.
2015	**Adaptive Management & Doing Development Differently**	Iteration, learning and adaptation became the operational mantras. Flexibility and course-correction positioned as the response to implementation complexity.
2020	**FCV Strategy 2020–2025**	Protracted crisis, displacement and conflict-sensitive programming given a dedicated strategic framework. IDA FCV envelope scaled to $30 billion under IDA20.
2024–26	**Human Development Initiative & Outcome Scorecards**	Shift from measuring disbursements to measuring people reached. New service delivery scorecards and human development metrics presented at 2026 Spring Meetings.
2026	**PCV Synthesis Paper: Navigating Tradeoffs (P505313)**	Most sophisticated operational guidance the Bank has produced on service delivery in fragile settings. Funded by FCDO. No mandatory evaluation of its use or impact.
Result	**FCS service delivery (health, education, urban, water, governance): 23.1% S+ across 420 evaluated projects, $21.8bn committed, 2000–2026. The analytical frameworks have become more sophisticated. The outcomes have not.**	

The trajectory raises a question the PCV paper does not ask: if twenty-five years of increasingly sophisticated analytical frameworks have not moved the FCS service delivery S+ rate materially above 23%, what does that tell us about where the binding constraint actually lies?

The latest step in this evolution — the Human Development Initiative and outcome scorecards presented at the 2026 Spring Meetings — shifts attention toward measuring people reached and service delivery results during implementation. The initiative is directionally welcome. But the IEG evidence presented in this paper suggests that measurement during implementation cannot substitute for stronger project architecture before approval. If the projects entering implementation are poorly designed, more accurate measurement of their execution will document the shortfall more precisely — it will not reverse it.

The PCV paper's answer, implicit in its structure, is that teams need better tools for managing implementation trade-offs. This paper argues that the IEG evidence points toward a different answer: the constraint is not in implementation, but in project architecture — and in the institutional incentives that shape it.

The Question the Report Does Not Ask

How many of the trade-offs that operational teams are asked to manage during implementation were embedded in the project design at preparation? If the majority originate at entry — in unrealistic PDOs, weak institutional analysis, inadequate risk assessment — then improving implementation guidance addresses the symptom, not the cause.

II. The IEG Evidence: What the Bank's Own Data Show

This analysis draws from the IEG's ICRR database (March 2026 update, 10,453 evaluated projects). The performance threshold used throughout is S+ (Satisfactory or Highly Satisfactory) — the minimum standard IFC and MIGA apply to their own corporate performance, and the appropriate benchmark for assessing whether a project has delivered meaningful development outcomes. The IDA21 Deputies Report states that IEG ratings show 91% satisfactory performance in IDA-financed operations — using the MS+ (Moderately Satisfactory or above) threshold. The analysis below uses S+. The gap between the two figures — 68 percentage points for IDA+FCS service delivery sectors — is itself a finding.

Finding 1: Service delivery in FCS countries performs at 23.1% S+ — flat across two decades

Metric	IEG Finding	Context
Health in FCS countries (eval 2000–2026)	**27.4% S+**	84 projects evaluated
Education in FCS countries (eval 2000–2026)	**24.7% S+**	81 projects
Urban/Resilience in FCS countries (eval 2000–2026)	**25.0% S+**	116 projects
Water in FCS countries (eval 2000–2026)	**18.0% S+**	50 projects
Governance in FCS countries (eval 2000–2026)	**18.0% S+**	89 projects
All 5 SD sectors combined, FCS	**23.1% S+**	420 projects, $21.8bn committed
Below-Satisfactory commitment in SD+FCS	**$16.2bn**	74.3% of total SD commitment in FCS countries

The decade breakdown is equally striking. The 2010–2019 period — when the governance, political economy, and adaptive management agenda was at its most active — produced the lowest S+ rate in the database (19.2%), against 496 projects and $20.8bn committed. Analytical sophistication increased steadily across this period. Measured outcomes did not improve commensurately. This is a pattern that deserves direct management attention, and does not receive it in the PCV paper.

Finding 2: QAE is among the strongest predictors of outcomes in FCS settings

The relationship between Quality at Entry ratings and development outcomes in FCS countries is one of the strongest predictors in the IEG dataset. The gradient is not a correlation — it is one of the strongest and most consistent patterns in the IEG dataset.

Quality at Entry Rating	N (FCS)	S+ Outcome Rate	Committed ($)
Highly Satisfactory	16	81.2%	$2.2bn
Satisfactory	315	61.0%	$22.6bn
Moderately Satisfactory	397	17.4%	$29.3bn
Moderately Unsatisfactory	216	**1.4%**	$11.3bn
Unsatisfactory	71	**0.0%**	$2.7bn
Highly Unsatisfactory	6	**0.0%**	$0.1bn

Projects with Satisfactory or Highly Satisfactory QAE achieve S+ outcomes 61–81% of the time — even in FCS environments that rank among the most difficult operating contexts on earth. Projects rated Moderately Unsatisfactory at entry achieve 1.4%. Projects rated Unsatisfactory or Highly Unsatisfactory achieve 0.0% across 77 projects and $2.8bn. The practical implication is direct: adaptive implementation has very limited ability to compensate for serious weaknesses in project architecture. The PCV paper's three-step framework — analyse context, calibrate objectives, adapt — is a framework for the 61% of well-designed projects that can still be rescued by good supervision. For the 39% that entered poorly designed, the data show adaptation cannot compensate.

The Adaptation Paradox

IDA+FCS projects with poor QAE (Moderately Unsatisfactory or worse): 206 projects, S+ outcomes = 2 (1.0%), committed financing = $10.9bn.

IDA+FCS projects with Satisfactory or Highly Satisfactory QAE: 236 projects, S+ outcomes = 144 (61.0%), committed financing = $20.5bn.

The Bank is committing $10.9bn to projects in its most difficult operating environments with a demonstrated 1% chance of a Satisfactory outcome — and responding with 190 pages of better guidance on adapting those same projects during implementation.

Finding 3: Poor QAE is endemic in the PCV paper's target sectors

Across the five service delivery sectors the PCV paper covers, in FCS countries evaluated since 2000, 33.7% of projects received Moderately Unsatisfactory or worse QAE ratings from IEG. One in three service delivery projects in the most difficult development environments on earth entered implementation already carrying design weaknesses that the QAE gradient shows are strongly associated with below-S+ outcomes. In those 141 poorly-designed FCS service delivery projects, the S+ outcome rate is 0.0%. This is the most actionable variable in the entire dataset — the variable the Bank directly controls. It does not appear in the PCV paper.

Finding 4: Bank Performance ratings and M&E confirm the governance dimension

Metric	IEG Finding	Context
FCS projects with Poor Bank Performance (MU/U/HU)	**23.0%**	238 of 1,034 rated FCS projects
FCS projects with Poor Quality at Entry (MU/U/HU)	**28.7%**	293 of 1,021 rated FCS projects
FCS projects with S+ Bank Performance	**31.3%**	324 projects
FCS projects with S+ Quality at Entry	**32.4%**	331 projects
FCS projects: Poor M&E Quality (Negligible/Modest)	**55.6%**	473 of 850 M&E-rated projects — the majority

Over half of FCS-country projects have monitoring and evaluation frameworks rated Negligible or Modest — the bottom two tiers. The PCV paper recommends embedding ongoing context analysis to support adaptive implementation. That recommendation requires a functioning monitoring system. The data show the majority of FCS projects do not have one.

III. The Countries the PCV Paper Covers: What the IEG Record Shows

The PCV paper draws evidence from twenty-six countries experiencing protracted crisis and violence. The IEG database provides a direct performance baseline for the core PCV countries named in the report — a baseline the paper does not present.

Country (PCV/FCS)	N Evaluated	S+ Rate (IEG)	Total Committed
Afghanistan	109	**16.5%**	$10.2bn
Congo, Dem. Rep.	49	**12.2%**	$6.7bn
Iraq	25	**16.0%**	$4.2bn
Haiti	43	**11.6%**	$1.4bn
Mali	70	**18.6%**	$3.8bn
Central African Republic	31	**19.4%**	$0.8bn
South Sudan	25	**24.0%**	$0.8bn
Myanmar	14	**28.6%**	$2.4bn
Sudan	21	**28.6%**	$2.3bn
Nigeria	89	**24.7%**	$14.1bn

These are not countries of shallow engagement. They represent the sustained multi-decade organisational presence of the world's largest development institution, producing outcome rates below 20% by that institution's own independent evaluation standard.

The pattern is not explained by a single difficult decade or a specific sectoral weakness. It persists across the full range of analytical frameworks the Bank has deployed since 2000 — and across the full range of preparation and supervision efforts those frameworks have generated. The PCV paper's three-step operational framework does not address this track record. It does not ask why it exists.

IV. Who Diagnoses the Diagnosticians?

The PCV paper applies the Bank's institutional diagnostic framework — with considerable sophistication — to borrowing governments. It asks in careful detail whether those governments have adequate implementation capacity, whether accountability systems function, whether institutional incentives are aligned with service delivery outcomes. These are exactly the right questions. They are also questions the Bank has been asking of client governments since the Wappenhans Report was commissioned in 1992.

What is absent is any equivalent analysis directed at the Bank itself. The table below makes the asymmetry explicit. It is not rhetorical: each question in the right column corresponds to a variable that IEG data shows has a direct, measurable relationship with project outcomes in FCS settings.

What the Bank studies in borrowing governments	What the Bank rarely studies in itself
Is there political commitment to reform?	**Is management rewarding lending volume over design quality?**
Is implementation capacity adequate for the task?	**Are TTLs given adequate preparation time and resources?**
Are accountability systems functioning?	**Who is accountable when a poorly-designed project fails?**
Are incentives aligned with development outcomes?	**Are Bank incentives aligned with outcomes or disbursement targets?**
Is the monitoring and evaluation framework credible?	**Why do weak Results Frameworks continue to be approved by QER?**
Is there risk of elite capture in project management?	**Who bears responsibility for over-ambitious PDOs in FCS contexts?**
Is institutional capacity realistic for the task at hand?	**Do preparation budgets reflect the complexity of FCV environments?**
Are accountability mechanisms for programme performance in place?	**Why is ASA — $783M/year — not subject to independent evaluation of operational uptake or development results?**

The final row is particularly important. P505313 is itself an ASA product — part of the approximately $783 million the Bank spends annually on advisory services and analytics. Unlike lending operations, ASA products are generally not subject to independent IEG validation, and systematic public reporting on costs, operational uptake, and development results is limited. The contrast with the lending portfolio is precise: every completed lending operation produces a mandatory Implementation Completion Report, which IEG independently reviews for 100% of projects. Knowledge products produce a publication and dissemination — but no public evidence of systematic evaluation of operational influence or uptake.

The scope of P505313 illustrates the point. Phase 1 alone produced three publicly available documents: a 190-page Synthesis Paper, an 82-page Evidence Review (authored by Mcloughlin, Grandvoinnet, and Lemay-Ort, covering 200+ peer-reviewed papers and 28 World Bank evaluations), and an 18-page practitioner Operational Note. The research methodology covered 37 structured TTL and Country Manager interviews, a 42-respondent task team survey, a two-day Washington workshop with over 100 participants, three sector analyses (health, education, urban services), and country-level analytics across 26 PCV settings. An extended team of over fifteen World Bank staff and consultants, two University of Birmingham academics, six sector focal points, and six peer reviewers contributed to the work. Phase 2 — action labs and further refinement — is planned. This is a two-year programmatic ASA, not a single publication. No cost figure is publicly accessible in any World Bank financial system.

None of that cost is disclosed in any publicly accessible World Bank financial record. The three deliverables are all listed on the Bank's Documents & Reports portal: each carries standard metadata — title, authors, disclosure date, P-code. None carries a budget figure, a trust fund number, an FCDO grant reference, or any cost information of any kind. The Bank's Finances One platform explicitly excludes Trust Fund data. The Projects & Operations portal covers lending projects only. External stakeholders — including FCDO, the programme's own funder — cannot use any public World Bank financial system to determine what P505313 cost, what the three deliverables cost individually, or how the grant was structured. This is not a claim about internal management information. It is an observation about external accountability — the standard the Bank holds its borrowers to in their own public expenditure management systems.

Accountability Stage	Lending Operations: Standard	ASA / Knowledge Products: Standard
Project Appraisal Document (PAD)	Before Board approval. Public. Includes theory of change, risk assessment, results framework, implementation arrangements, institutional capacity analysis.	ASA concept note, if any. Generally internal. No standardised public disclosure of scope, budget, or expected influence pathway.
Implementation Status Reports (ISR)	Mandatory at regular intervals throughout implementation. Tracks progress against PDO indicators. Disclosed publicly.	No equivalent. Internal task management tools exist but are not subject to systematic public disclosure.
Implementation Completion Report (ICR)	Mandatory at project closure. Self-evaluates outcomes against PDO. Rates Bank performance, Borrower performance, M&E quality.	No equivalent mandatory product. Some ASA tasks produce completion summaries for internal management purposes.
IEG ICRR / PPAR	Independent IEG validation of 100% of ICRs. Provides independent rating of outcome, efficiency, Bank performance. Published. Enters public database.	Not subject to independent IEG validation. No equivalent public database of ASA outcomes, influence, or operational uptake.
Cost disclosure	Project costs, trust fund co-financing, and Bank budget contributions disclosed in PAD and financial statements. Accessible via Finances One.	Trust Fund ASA activities excluded from Finances One. Individual ASA task budgets not systematically disclosed. P505313 cost: not publicly accessible.

There is an irony in this that the Evidence Review almost names. It observes that aid actors in PCV settings suffer from 'fragmentation, competing priorities and chronic resource gaps' and that 'internal competition and silos create problems for effective delivery.' These are accurate observations about the operating environment in fragile states. They are also an accurate description of the World Bank's own project preparation system — matrix management, GP silos, competing lending targets, chronic supervision budget constraints. The Bank's own analytical framework, applied to itself, would identify the same structural problems it identifies in its client countries.

The Missing Political Economy

The Bank has applied political economy analysis to borrower governments for over two decades — identifying how elite capture, patronage, fragmentation, and misaligned incentives undermine service delivery. It has not applied the same framework to itself.

The measurable institutional variables are not mysterious: TTL incentives weighted toward Board approval rather than implementation success; preparation time compressed from approximately 19 months to 12 months under current management; supervision budgets running at 4.5 cents per lending dollar in FY2026; matrix management that diffuses accountability for project design failures; and an ASA portfolio of $783M/year that has never been systematically evaluated against outcomes. These are not hypothetical constraints. They predict project quality in the IEG data. They are absent from the PCV paper.

V. The Missing Step: Design Before Adaptation

The PCV paper proposes a three-step decision framework: Analyse Context → Calibrate Objectives → Adjust Design and Manage Trade-offs. The IEG data indicate a Step 0 that must precede all three.

A Revised Framework for Service Delivery in PCV Contexts

Step 0 — Design a realistic and implementable operation: rigorous political economy analysis calibrated to what is actually available in the FCV environment; institutional capacity assessment grounded in evidence rather than assumption; realistic PDO; M&E framework built for adaptive management under insecurity; explicit risk assessment at preparation, not at restructuring.

Step 1 — Analyse the PCV context: rivalry, fragmentation, scarcity (as the PCV paper recommends).

Step 2 — Calibrate development objectives: effectiveness, do no harm, potential to last.

Step 3 — Adjust design and manage trade-offs through implementation.

Without Step 0, Steps 1–3 are guidance for managing the consequences of inadequate project architecture. The IEG data show that in FCS environments, those consequences are rarely reversed.

The shareholders who fund IDA replenishments are told by the IDA21 Deputies Report that IEG ratings show 91% satisfactory performance in IDA-financed operations. That figure uses MS+ as the threshold. Using S+ — the standard IFC and MIGA apply to their own investments, and the standard the Bank established in 1960 as its original pass/fail benchmark — the IDA all-period figure is 40.5%. For IDA+FCS, 35.9%. For IDA+FCS service delivery sectors specifically, the outcome record documented in Section II: 23.1%.

The communities in Afghanistan, DRC, Haiti and Mali experience the 23%. The 68-percentage-point gap between those two numbers is not a rounding error. It is the governance mechanism that has obscured IDA's delivery performance from accountability for four decades — and the gap the PCV paper, for all its sophistication, does not address.

VI. Conclusion

Supporting Service Delivery in Protracted Crisis and Violent Contexts: Navigating Tradeoffs is the most sophisticated operational guidance the World Bank has produced on service delivery in conflict environments. The rivalry-fragmentation-scarcity analytical framework is sound. The three-step decision tool will help TTLs navigate genuinely difficult choices. The paper should be read, used, and built upon by practitioners working in the twenty-six PCV countries it covers.

But the paper also illustrates a pattern that has characterised the Bank's service delivery thinking across four analytical generations. The Wappenhans Report (1992) identified weak quality at entry, approval culture, and inadequate preparation incentives as core organisational pathologies. IEG has documented the same findings in every decade since. The Bank's response has been to produce better implementation guidance — community-driven development, governance, political economy, adaptive management, the FCV Strategy, and now the PCV paper. Each framework has been more analytically sophisticated than the last. The IEG data for FCS countries show that this pattern of response, repeated across twenty-five years, has not moved the outcome needle.

This is an observation about institutional learning, not a criticism of individual staff. The TTLs who participated in the PCV paper's research programme — interviewed by Grandvoinnet's team — demonstrate the operational creativity and dedication that exists within the institution. The constraint is not their capacity. It is the institutional architecture within which they operate: the approval culture, the preparation budgets, the disbursement incentives, the matrix management structure, and the ASA accountability vacuum that allows $783 million per year in knowledge expenditure to go unevaluated against any outcome standard.

The next step is not another framework — it is applying the same standards of evidence, accountability, and evaluation to the Bank's own operational and knowledge systems that it asks borrowing governments to meet. Three reforms toward that end are proposed in the concluding section.

Bottom Line

The PCV paper asks: How should teams manage difficult trade-offs during implementation?

The IEG data ask: Why do Bank-designed projects in FCS environments enter implementation with design weaknesses that a rigorous quality-at-entry process would have identified — and in many cases prevented?

For more than twenty-five years, the World Bank has become increasingly sophisticated at diagnosing why governments struggle to deliver services. The Missing Mirror is the absence of an equivalent diagnosis directed inward.

VII. From Better Frameworks to Better Learning

This paper documents a diagnostic asymmetry. The World Bank has continuously refined its analytical tools for understanding why borrowing governments struggle to deliver services — but has not applied the same discipline to understanding why its own knowledge and project preparation systems produce the outcomes documented here. The Bank currently lacks an evaluative framework for ASA; the IEG's own RAP 2025 concept note, covering 5,720 country-specific ASA activities since 2006, states precisely that. Twenty years earlier, Banerjee, Deaton, Lustig, and Rogoff found that Bank research was disconnected from operations, used selectively for advocacy purposes, and subject to insufficient quality control. The RAP 2025 is the first system-wide review since their panel reported. The diagnosis has not changed. That continuity is itself a data point about the Bank's learning culture.

Diagnosis without reform is what this paper criticises in the PCV paper itself. Three reforms follow directly from the evidence presented here. They are drawn in part from the platform's earlier analysis of the Bank's knowledge accountability gap (The Knowledge Bank, mdbreform.com, April 2026) and adapted to the specific argument of this paper.

Reform	What It Requires
1. Evaluate knowledge like lending	Every flagship ASA programme — including multi-year, multi-team research programmes of the scale of P505313 — should be subject to periodic independent review covering operational uptake, influence on project preparation, and evidence of use in PADs and restructuring papers. Not identical procedures to lending evaluation, but equivalent accountability. IFC has systematically evaluated all client-facing advisory projects since 2008, with IEG validating a 51% sample annually. The World Bank has no equivalent for ASA. If the Bank believes that better knowledge improves project design, it should be able to demonstrate that claim.
2. Apply quality-at-entry standards to knowledge products	The QAE evidence in this paper demonstrates that project architecture predicts outcomes more reliably than any contextual variable. The same logic applies to knowledge products: does the analytical architecture — the theory of change linking the research to operational influence — predict whether the product will be used? Flagship ASA should state, at inception, the pathway from research to operational uptake. At closure, that pathway should be assessed. If the Bank believes project QAE predicts outcomes, it should ask whether knowledge QAE predicts organisational learning.
3. Track operational uptake of guidance in lending	The Bank currently measures knowledge production — reports, downloads, dissemination events. It does not systematically measure whether guidance changes project preparation. A simple operational tracking mechanism — identifying whether PADs for projects in PCV countries reference the framework developed in P505313, or whether ISRs show design changes attributable to ASA guidance — would convert a publication metric into a learning metric. Annual reporting to the Board on Quality at Entry trends, disaggregated by FCS status and sector, would make the pattern documented in this paper visible to management and shareholders rather than analytically discoverable only from outside.
4. Increase transparency of flagship knowledge products	The scope of P505313 — described in Section IV — is a two-year programmatic ASA spanning 290 pages, 37 structured interviews, a 42-respondent survey, a Washington workshop with over 100 participants, and country-level analytics across 26 PCV settings. None of this cost is publicly accessible in any World Bank financial system. All three Phase 1 documents are listed on the Bank's Documents & Reports portal with standard metadata; none carries a budget figure, a trust fund number, or any financial information. A lending project of equivalent scale would have a publicly disclosed PAD, a commitment amount, and a Finances One profile. The accountability asymmetry is structural.

These are not radical proposals. None requires new analytical frameworks. Each applies existing World Bank accountability principles to a part of the institution that has largely remained outside equivalent external scrutiny. They are extensions of principles the Bank already applies to its lending portfolio — mandatory evaluation, quality standards at entry, outcome tracking, financial transparency — into the knowledge domain. The asymmetry between lending and ASA accountability is not a design feature. It is an oversight that has persisted because no internal actor has had strong incentives to close it.

The Independent Evaluation Group has already recognised this gap. The RAP 2025 review, budgeted at approximately $1 million against an annual ASA spend of roughly $800 million, is a beginning. The Missing Mirror this paper documents is not only between what the Bank asks of borrowing governments and what it asks of itself. It is also between the quality of the Bank's external diagnostics and the quality of its internal learning systems.

The Central Challenge

The challenge facing the World Bank is no longer generating knowledge. Between the Wappenhans Report (1992) and P505313 (2026), it has generated thirty-five years of increasingly sophisticated diagnostics of why development projects fail in difficult environments — and devoted substantial and largely undisclosed resources to producing each new generation of guidance.

The RAP 2025 is the first system-wide review of the Bank's ASA portfolio in two decades. What it should establish — and what follows from the evidence in this paper — is a mandatory evaluative framework for programmatic ASA: a theory of change at inception, an assessment of operational uptake at closure, and public disclosure of cost on a par with lending. The IEG has the mandate and the methodology. What has been missing is the requirement.

The next challenge is ensuring that this knowledge changes the institution itself — and that the institution can demonstrate, publicly and transparently, that it has done so.

Data Note

All quantitative findings are from the IEG ICRR database (March 2026 update), 10,453 evaluated projects. The S+ (Satisfactory or Highly Satisfactory) threshold is used throughout — the minimum standard applied by IFC and MIGA to their own corporate performance, and the appropriate benchmark for meaningful development outcomes.

- FCS classification: IEG Country FCS Status field. Evaluation period 2000–2026 unless otherwise stated.
- QAE analysis: restricted to projects with both a rated outcome and a rated QAE. 'Not Rated' projects excluded. The relationship between QAE and outcomes reflects association, not demonstrated causation; other variables (country context, sector, project complexity) also contribute.
- Service delivery sectors: Health, Nutrition & Population; Education; Urban, Resilience and Land; Water; Governance Global Practices.
- MS+ vs S+ gap: IDA21 Deputies Report (World Bank, March 2025) states 91% satisfactory performance in IDA-financed operations using the MS+ threshold. The S+ figure for IDA+FCS service delivery is 23.1%. The 68-percentage-point gap reflects methodological choice, not data error.
- ASA expenditure: IEG RAP 2022 (annual ASA expenses $783M in FY22; peak $811M in FY19; 40–50% of total World Bank project expenses FY13–22).
- P505313 cost: Not disclosed in any publicly accessible World Bank financial system. Trust fund ASA activities are excluded from World Bank Finances One platform reporting.
- Related platform analysis: The Knowledge Bank — A $1 Billion Knowledge Portfolio. No Evaluative Framework (mdbreform.com, April 2026, mdbreform.com/the-knowledge-bank/). Documents the full ASA accountability gap, the Banerjee–Deaton benchmark, peer review capture, and five operational recommendations.

MDB Reform Advisory | Parminder Brar | mdbreform.com | pbrar@mdbreform.com | July 2026