Revised World Bank FCV Strategy:
Suggestions for Consideration

February 2026

Executive Summary

The World Bank has published on its website a presentation titled "Towards a refreshed WBG strategy for engaging in FCV settings". The Bank has asked for public comment on this document. This is Phase II of its consultation process with the last date for providing comment being 28th February 2026. This note has been prepared in response to that request.

The World Bank Group's strategic direction is sound. Security, justice, and jobs as the three pillars of development in fragile and conflict-affected settings; private sector and MSME development as the engine of employment; an enhanced FCV classification framework; and IDA21's $100 billion commitment authority — the largest in IDA's history — these are well-grounded, well-reasoned, and directionally correct. (The analytical basis for this framework — jobs, private sector, and stability in fragile settings — is reviewed in **Annex A**.)

The problem is not the strategy. The problem is the delivery platform on which the strategy will be executed. Ten years of matched IEG evaluation and World Bank commitment data — 1,823 rated IDA projects, $177.0 billion in IDA commitments, IEG Master Database March 2026 — tells a consistent story that the institution has not yet told itself with the directness it requires.

> $117.0bn of $177.0bn committed in IDA countries — 66.1% — was committed to projects that did not achieve Satisfactory outcomes (FY2015–2026). In FCS countries: $34.6bn of $47.1bn (74%). In the seven jobs-critical Global Practices: $47.2bn of $65.8bn (71.7%). This is not the exception. It is the dominant outcome.

This report is structured around four findings and a forward assessment:

- **Part I** establishes that theme is not the problem — strategic priorities are correct and do not require another revision.
- **Part II** presents the World Bank's IDA delivery record across lending instruments (IPF, DPF, PforR), regions, FCS countries, and quality dimensions, and identifies the four institutional root causes of persistent underperformance.
- **Part III** examines IFC's private sector record in the markets the strategy most relies upon — FCS and IDA — where investment outcome rates reached a nadir of 11% Satisfactory (CY2020–22) just as the strategy proposes to scale IFC's FCS presence, and sets out the seven structural constraints that no amount of PSW financing can resolve.

- **Part IV** assesses IDA21 — what it funds, what its architecture achieves, and where structural gaps remain in the accountability, additionality, and operational quality frameworks that will determine whether the $100 billion produces different results from the $93 billion before it.

The recommendations that follow are not strategic: they are institutional. They concern who is accountable, for what, measured how, with what consequence. Without this level of specificity, the next strategy cycle will produce a third generation of sound analysis applied to the same unreformed delivery machinery.

This note is prepared by **Parminder Brar**, former Country Manager of the World Bank, who worked extensively in fragile and conflict-affected (FCV) countries between 2003 and 2023. During that period, he served in four field postings in Africa and participated in over 600 operational missions.

This paper updates an earlier note produced by him in 2023 titled *“Review of IEG Ratings of the Governance Portfolio in West and Central Africa, 2000–2021*” when he was the Lead Governance Specialist for West Africa. The analysis here is based exclusively on publicly available World Bank data downloaded from official sources. The dataset was compiled and structured for analysis using AI-assisted tools, with all calculations and interpretations reviewed by the author. No confidential or internal World Bank data have been used. The data and analysis contained herein can be circulated freely, either with or without attribution. Comments are welcome. Please respond to fcvstrategy@gmail.com.

PART I

Theme Is Not the Issue

The strategic diagnosis is correct — and has been for a decade

1. The Strategy Is Right

The World Bank Group's Refreshed FCV Strategy makes a sound analytical case. The concentration of extreme poverty in fragile and conflict-affected situations — projected to reach 60% of the world's extreme poor by 2030 — is correctly identified. The three-legged stool of security, justice, and jobs correctly maps the interdependencies that individual sector strategies consistently underperform by ignoring. The four strategic shifts — anticipate fragility earlier, differentiate engagement by FCS type, mobilise the full WBG for jobs, and enhance the operational toolkit — are directionally rational and grounded in the Bank's own operational experience and the external literature.

Private sector development and MSME growth are the correct engines for employment in FCV contexts. Jobs provided by private employers are more durable than those funded by public investment. The Minerals and Metals Compact, AgriConnect, and M300 are plausible instruments. The new FCV classification framework — distinguishing prevention, high-intensity conflict, and turn-around settings — is a genuine advance over blunt FCS designation, and the IDA21 FCV Envelope's three-tier architecture (PRA, RECA, TAA) is structurally aligned with it.

IDA21's thematic architecture — five Focus Areas organised around People, Planet, Prosperity, Infrastructure, and Digital Transformation, with four cross-cutting Lenses including FCV — is appropriately comprehensive. The WBG Scorecard's disaggregation of all 22 outcome indicators by FCS status for the first time is a meaningful transparency commitment. The Grant Facility for Project Preparation and the Facetime Index for staff presence are operationally useful additions.

> The question is not whether the strategy has the right priorities. Jobs and private sector development have been central to WBG FCV strategy for over a decade. The question is whether the Bank's institutional machinery can deliver these priorities at scale in the settings where they are most needed.

2. Why the Delivery Record Demands Priority Attention

The critical point — which this report documents at length in Parts II and III — is that the gap between strategic ambition and institutional delivery is not marginal, recent, or attributable to the difficulty of the operating environment alone. It is structural, persistent across a decade of measurement, and financially very large.

The Wappenhans Report in 1992 identified what it called an 'approval culture' — a preoccupation with new lending at the expense of implementation quality, driven by institutional incentive structures that reward portfolio size over portfolio performance. Thirty-four years and multiple strategy cycles later, the IEG RAP series continues to document the same structural pattern. The scale of the delivery gap documented in this review is not a rounding error. It is the accumulated financial expression of an institution that has not yet resolved the tension between lending ambition and delivery accountability.

The Refreshed FCV Strategy contains sound operational proposals. The evidence demands that it be accompanied by an equally specific institutional reform programme — covering accountability structures, staff incentives, quality control systems, and Board governance — without which the strategic vision will not be realised at the scale the strategy envisions.

PART II

World Bank Results on IDA 2020–2025

The delivery record across instruments, regions, and quality dimensions

3. The Headline Numbers: Scale of Below-Standard Spending

Between FY2015 and FY2025, IDA-eligible countries received **$177.0 billion** in World Bank commitments across the IEG-evaluated portfolio (1,823 projects, IEG Master Database March 2026). Of this, **$117.0 billion — 67%** was committed to projects that did not achieve Satisfactory or better IEG outcome ratings. In FCS countries, **$34.6bn of $47.1bn (74%)** fell below Satisfactory.

These figures use the Satisfactory or above (S+) threshold — not the World Bank's institutional benchmark of Moderately Satisfactory or above (MS+). The distinction matters: MS+ produces headline figures of 79–88%, approximately double the S+ rates shown here, because it treats partial achievement of objectives as an acceptable result. **Annex B** of this report explains in detail why S+ is the methodologically correct standard. Briefly: Satisfactory was the original pass/fail threshold when the rating scale was introduced; the insertion of Moderately Satisfactory in 1994 added granularity, not a new definition of success. A project rated Moderately Satisfactory has, by definition, not been Satisfactory.

3.1 Annual Performance Trend, FY2015–2026

The table below shows S+ rates by evaluation year across all projects, FCS projects, and IDA projects. The trend shows improvement from the nadir years of 2015–2018 (17–31% S+), stabilisation in the mid-2020s, and an apparent improvement in FY2025 — though the FY2025 cohort is the most recently evaluated and the most susceptible to upward revision as additional ICRRs are validated.

Eval. FY	All Projects n	All S+%	FCS n	FCS S+%	IDA n	IDA S+%
FY2015	434	17.5%	76	15.8%	161	14.3%
FY2016	292	31.2%	36	19.4%	113	23.0%
FY2017	361	30.7%	57	8.8%	123	22.8%
FY2018	368	31.5%	57	12.3%	136	24.3%
FY2019	276	37.3%	47	29.8%	109	35.8%
FY2020	295	35.6%	35	37.1%	104	30.8%

Eval. FY	All Projects n	All S+%	FCS n	FCS S+%	IDA n	IDA S+%
FY2021	247	44.5%	61	36.1%	118	39.0%
FY2022	247	41.3%	55	32.7%	96	38.5%
FY2023	286	42.0%	61	31.1%	126	36.5%
FY2024	385	39.2%	101	40.6%	175	34.3%
FY2025	328	49.1%	73	45.2%	142	45.8%
FY2015–25 Average	**3,519**	**38.3%**	**659**	**29.6%**	**1,403**	**31.0%**

Source: IEG ICRR-PPAR database matched with World Bank Projects commitment data. S+ = Satisfactory or Highly Satisfactory only.

The Banks Independent Evaluation Group (IEG) produces on an annual basis a Results and Performance Report of the World Bank Group (RAP). **Annex C** summarizes the key findings of five RAP's 2020 and 2024.

3.2 Performance by Lending Instrument

The World Bank deploys three principal lending instruments, each with materially different outcome profiles. Understanding these differences is essential for instrument selection in FCV contexts.

Instrument	Evaluated	S+%	Below S+%	Total Committed	Below-S+ Committed	Key Characteristic
Investment Project Financing (IPF)	7,141	39.1%	60.9%	$631.9bn	$357.1bn (56.5%)	Finances specific inputs; project team self-evaluation; long implementation cycles
Development Policy Financing (DPF)	1,551	35.6%	64.4%	$342.9bn	$186.4bn (54.4%)	Budget support against policy conditions; rapid disbursement; 10.8% S+ in IDA FCS
Programme-for-Results (PforR)	72	61.1%	38.9%	n/a	n/a	Disburses against verified results; no Unsatisfactory rating in evaluated cohort

IPF is the workhorse, accounting for 81.5% of all evaluated operations. Its 39.1% S+ rate means nearly two-thirds of the portfolio — and $357 billion in commitments — did not reach the Satisfactory threshold. The modal outcome is Moderately Satisfactory, which as noted above represents partial achievement. A note on IPF is at **Annex D.**

DPF performs even worse at 35.6% S+, with performance in IDA FCS at 10.8% — less than one in five operations meeting the threshold. The Sub-Saharan Africa DPF rate is 21.1% S+; Western and

Central Africa records 17.1%. DPF in FCS combines rapid disbursement with weak institutions and generates the institution's worst outcome data. Of $50.3bn committed to African DPF operations, $38.6bn (76.8%) was in below-Satisfactory projects. **Annex E** has a note on DPF's.

PforR is genuinely different. 61.1% S+ across 72 evaluated operations, no Unsatisfactory or Highly Unsatisfactory ratings in the entire evaluated dataset, and an Africa rate of approximately 56% S+ — dramatically better than IPF (30.7%) or DPF (21.1%) in the same region. Three caveats apply: PforR's sample of 72 is small relative to IPF and DPF; PforR requires functioning government systems and is not deployed in the most challenging contexts; and a novelty effect may be inflating early ratings. Nevertheless, the gap is too large to be noise. A note on PforRs is at **Annex F**. **Results-based disbursement against verified outcomes produces materially better development results than input-based or policy-based disbursement. This is among the most important operational findings in this dataset.**

3.3 Jobs-Critical Sectors: Outcome Rates and Dollar Volumes

The table below summarises IEG outcome rates and committed amounts for Global Practices central to the strategy's jobs agenda. **The combined shortfall across these seven GPs — $47.2bn of $65.8bn (71.7%) — is the core financial case for institutional reform.**

Global Practice	IDA S+%	IDA+FCS S+%	IDA Committed	IDA Below-S+	Strategic Role
Finance, Competitiveness & Innovation	23.8%	18.8%	$4.1bn	$3.0bn (73%)	Core MSME / private sector GP
Agriculture and Food	31.2%	28.9%	$8.5bn	$5.8bn (69%)	Rural jobs, food security
Transport	24.7%	19.4%	$11.1bn	$9.9bn (89%)	Market access infrastructure
Energy & Extractives	30.4%	33.3%	$8.8bn	$5.4bn (61%)	Power for productive employment
Governance	21.3%	20.0%	$3.9bn	$3.0bn (77%)	Institutional enabler
Macroeconomics, Trade & Investment	11.4%	9.2%	$17.5bn	$14.3bn (82%)	Largest GP by volume; DPF-heavy. Worst performer
Social Protection & Jobs	54.7%	47.2%	$14.0bn	$7.2bn (51%)	Simpler delivery, direct transfers
TOTAL — jobs-critical GPs	**~32%**	**~26%**	**$65.8bn**	**$47.2bn (71.7%)**	

Scaling an MSME model with this performance profile without first addressing its structural delivery problems will reproduce the same outcomes at greater scale.

Several findings warrant direct attention in the context of the Refreshed Strategy:

- **Finance, Competitiveness and Innovation (FCI)** — the GP responsible for the MSME agenda at the heart of the strategy — shows $3.0bn of $4.1bn committed in IDA (73%) falling below Satisfactory. In IDA+FCS countries, 81% of committed funds ($1.2bn of $1.4bn) are in below-Satisfactory projects. FCI also records the highest variance in annual outcome rates of any GP in the dataset: ranging from 0% Satisfactory (2017, 2019, 2024) to 60% (2025), with no sustained upward trend.

- **Energy & Extractives** — the foundational infrastructure on which every jobs outcome depends — records only 30.4% S+ across IDA and 31.5% in FCS. In IDA+FCS countries, $1.9bn of $2.6bn committed (76%) is in below-Satisfactory projects. This matters directly for the Refreshed Strategy's jobs agenda: without reliable electricity, MSME growth stalls, agro-processing cannot scale, and the formal employment the strategy targets is unreachable. Energy is not a parallel sector objective — it is a precondition for every other jobs outcome. Yet Energy & Extractives in FCV contexts performs at roughly the same level as Agriculture and Transport: a third of projects achieve Satisfactory outcomes, leaving two-thirds of committed capital in portfolios that did not deliver. No jobs strategy for FCV contexts can be credible without a parallel programme to raise Energy performance.

- **Agriculture and Food** — essential for rural employment, food security, and conflict resilience — shows $5.8bn of $8.5bn committed (69%) below Satisfactory in IDA. In IDA+FCS, this deteriorates to $2.0bn of $2.5bn (78%). Although Agriculture shows an improvement trend in recent years (from 4.5% Satisfactory in 2015 to 66.7% in 2024), the base is extremely low and the improvement is recent.

- **Transport** — the infrastructure backbone for market access and economic connectivity — records the starkest finding in FCS: only 16.7% of evaluated projects are Satisfactory or above. Of $6.5bn committed in FCS Transport, $5.6bn (86%) is in below-Satisfactory projects. Transport in FCS is both the lowest performing sector and the sector where poor performance has the greatest structural economic consequence.

- **Macroeconomics, Trade and Investment (MTI)** — which accounts for the largest single GP commitment volume in IDA ($17.5bn) — records the worst outcome rate at 11.4% Satisfactory, with $14.3bn (82%) below Satisfactory. At 82% below standard, MTI represents the largest absolute dollar volume of below-standard spending of any practice group.

Combined jobs-critical GPs (listed in the table above) account for $65.8bn of IDA commitments over this period. Of this, $47.2bn — 71.7% — is in projects that did not achieve Satisfactory outcomes. This is the core financial case for institutional reform alongside strategic ambition.

3.4 FCS Country Performance

The Global FCS Portfolio: Systemic Failure at Scale

Across 51 countries classified as Fragile and Conflict-Affected States in the IEG evaluation database, the World Bank delivered 644 unique IDA projects evaluated between FY2015 and FY2025, representing **$52.7 billion in total commitments.** Applying the Satisfactory or Highly Satisfactory (S+) threshold used consistently throughout this report, **191 of 644 projects — 30% — achieved a passing development outcome rating.** The remaining 453 projects, representing **$36.5 billion, or 70% of committed resources,** went to projects that did not reach the Satisfactory threshold. This is the most complete cross-country picture of World Bank delivery performance in fragile and conflict-affected settings available from the published IEG evaluation record, and it is the baseline against which the Refreshed FCV Strategy's delivery ambitions must be assessed.

> *70% of $52.7bn committed to FCS countries over a decade went to projects that did not reach the Satisfactory threshold. This is not a performance footnote. It is the primary delivery reality of the FCS portfolio.*

IEG Development Outcome Ratings by FCS Country, IDA Projects Evaluated FY2015–2026 (S+ Threshold)

Country	Projects	S+ (S, HS)	Below S+	Total Commit ($m)
Burundi	IDA	20	6 (30%)	$636m
Comoros	IDA	10	5 (50%)	$91m
Congo, Dem. Rep.	IDA	35	5 (14%)	$3,711m
Ethiopia	IDA	14	5 (36%)	$5,372m
Madagascar	IDA	8	0 (0%)	$444m
Malawi	IDA	2	0 (0%)	$100m
Mozambique	IDA	19	4 (21%)	$3,756m
Somalia	IDA	12	8 (67%)	$782m
South Sudan	IDA	15	5 (33%)	$529m
Sudan	IDA	15	5 (33%)	$2,134m
Zimbabwe	Blend	6	4 (67%)	$182m
Burkina Faso	IDA	19	9 (47%)	$1,535m
Cameroon	Blend	11	2 (18%)	$1,086m
Central African Republic	IDA	20	4 (20%)	$621m
Chad	IDA	18	4 (22%)	$559m
Congo, Republic of	Blend	14	3 (21%)	$477m
Côte d'Ivoire	Blend	12	3 (25%)	$690m
Gambia, The	IDA	8	4 (50%)	$165m
Guinea	IDA	3	0 (0%)	$50m
Guinea-Bissau	IDA	11	5 (45%)	$147m
Liberia	IDA	16	4 (25%)	$497m
Mali	IDA	29	7 (24%)	$1,684m
Niger	IDA	22	9 (41%)	$2,079m
Nigeria	Blend	20	12 (60%)	$6,083m
Sierra Leone	IDA	17	0 (0%)	$369m
Togo	IDA	10	4 (40%)	$224m
Afghanistan	IDA	43	5 (12%)	$4,523m
Djibouti	IDA	2	**2 (100%)**	$11m
Yemen	IDA	18	9 (50%)	$2,094m
Kosovo	IDA	12	5 (42%)	$228m
Uzbekistan	Blend	1	0 (0%)	$104m
Kiribati	IDA	10	1 (10%)	$97m
Lao PDR	IDA	5	3 (60%)	$126m
Marshall Islands	IDA	4	0 (0%)	$24m
Micronesia, Fed. States of	IDA	2	1 (50%)	$20m
Myanmar	IDA	11	3 (27%)	$1,890m
Papua New Guinea	Blend	9	2 (22%)	$500m
Solomon Islands	IDA	12	5 (42%)	$143m
Timor-Leste	Blend	5	0 (0%)	$57m
Tuvalu	IDA	6	0 (0%)	$52m
Haiti	IDA	26	2 (8%)	$993m
Nepal	IDA	8	3 (38%)	$498m
TOTAL — All FCS Countries	**644**	**191 (30%)**	**453 (70%)**	**$52,719m**

Source: IEG ICRR/PPAR evaluations, Evaluation FY2015–2026. S+ = Satisfactory or Highly Satisfactory only; Moderately Satisfactory does not qualify. Country FCS classification from IEG ICRR-PPAR database. Where both ICRR and PPAR evaluations exist, PPAR rating is used. Commitment figures from World Bank Projects database (as of 18 February 2026). 643 of 644 projects matched to commitment data.

What the Distribution Reveals

The aggregate 30% S+ rate conceals three structural patterns that are diagnostic rather than incidental.

First, scale compounds failure. The four largest portfolios — Afghanistan (43 projects, $4.5bn), DRC (35 projects, $3.7bn), West Bank and Gaza (35 projects, $917m), and Mali (29 projects,

$1.7bn) — account for 142 projects and $10.8bn in commitments but achieved only **21% S+ collectively. Afghanistan at 12% and Haiti at 8% represent the two worst-performing large portfolios in the dataset.** These are not marginal programmes: they are among the Bank's most consequential FCS engagements, and their performance record carries the greatest institutional weight.

> **Nine countries recorded zero S+ outcomes** across all evaluated projects over a decade: Sierra Leone (17 projects, $369m), Madagascar ($444m), Tuvalu ($52m), Marshall Islands ($24m), Timor-Leste ($57m), Guinea ($50m), Malawi ($100m), Angola ($103m), and Uzbekistan ($104m). Together they represent over $1.4bn in commitments with no project rated Satisfactory. Sierra Leone is the most striking: **17 projects evaluated over a decade with no project rated Satisfactory.**

Second, strong performance is explicable — and the explanations are institutional, not coincidental. Somalia (67%, 8 of 12 projects, $782m) achieved above-average performance through deliberate adaptation of implementation modalities: humanitarian-nexus programming, NGO delivery, and community-driven approaches that bypass failed government systems. Zimbabwe (67%, 4 of 6) reflects a compressed post-reengagement portfolio with careful project selection. West Bank and Gaza (43%, 15 of 35) performs above the FCS average despite active conflict, reflecting three decades of institutional memory and partner network depth. Georgia (60%) and Lao PDR (60%) demonstrate that mid-size portfolios can sustain performance where counterpart capacity is functional.

Third, the Nigeria result requires specific attention. Nigeria records 60% S+ across 20 IEG-evaluated FCS-classified projects ($6.1bn) — the highest S+ rate among the five largest country programmes, substantially above the 30% FCS average. This performance reflects the structural advantage of a Blend country with stronger federal counterpart capacity than most FCS peers. As the Nigeria case study in Section 3 of this annex demonstrates, however, this headline rate conceals a severe concentration of delivery failure in the economic transformation sectors most central to the Refreshed Strategy's agenda.

> *The FCS portfolio failure is not a story of crisis overwhelming delivery capacity. Countries operating in active conflict — Somalia, West Bank and Gaza — achieve above-average outcomes when implementation modalities are deliberately adapted. The countries with the worst outcomes are often those where the Bank deployed standard instruments in structurally weak institutional environments and expected reform commitments to be sustained.*

Case Study: Africa Region — Instrument Selection as the Mechanism of Failure

Sub-Saharan Africa is the principal testing ground for the World Bank's FCV and jobs agenda. It accounts for the largest concentration of IDA commitments, the greatest number of FCS countries, and the highest exposure to the Global Practices most central to the employment and private sector strategy. **If the institutional model is working anywhere, it should be working in Africa.** The evidence suggests that it is not — at least not consistently, and not at the scale implied by financial commitments.

> Across 1,193 IEG-evaluated projects in the Africa Region from FY2010 to FY2025, representing **$119.8 billion in total commitments,** only 25% by number and 24% by value

> reached the Satisfactory threshold. Three in every four dollars committed to Africa across this period went to projects that did not achieve Satisfactory development outcomes.

The headline figures by instrument type are shown below. The instrument breakdown is the analytical centrepiece of the Africa evidence, and its implications are direct.

Africa Region: IEG Development Outcome Ratings by Lending Instrument, FY2010–2025

Instrument	Projects (n)	Commitments (US$bn)	S+ Rate (%)	Below S+ (%)
Investment Project Financing (IPF)	1,169	$100.8bn	~26%	~74%
Development Policy Financing (DPF)	333	$34.7bn	~21%	~79%
Programme-for-Results (PforR)	24	$5.3bn	~58%	~42%
TOTAL — Africa Region (FY2010–2025)	**1,193**	**$119.8bn**	**28.4%**	**71.6%**

Source: IEG evaluation database, FY2010–2025. Africa Region = Eastern and Southern Africa + Western and Central Africa as defined in IEG database. IPF = Investment Project Financing; DPF = Development Policy Financing; PforR = Programme-for-Results. S+ = Satisfactory or Highly Satisfactory only. Commitment figures from World Bank Projects database (as of 18 February 2026).

Three Instruments, Three Stories

Investment Project Financing (IPF) is the dominant instrument, accounting for more than four-fifths of evaluated Africa operations and $100.8bn in commitments. Its S+ rate of approximately 26 percent implies that the majority of Africa's largest and most complex development investments fell short of the Satisfactory threshold. The modal rating is Moderately Satisfactory, which under the S+ benchmark used in this report reflects partial rather than full achievement of development objectives. The performance profile is steady across the period — **it is not improving.**

Development Policy Financing (DPF) performs materially worse: approximately 21 percent S+ across Sub-Saharan Africa, falling below 20 percent in Western and Central Africa. More than three-quarters of committed DPF resources in Africa are associated with below-Satisfactory outcomes. DPF is heavily concentrated in Macroeconomics, Trade and Investment — the largest Global Practice by volume and the weakest by outcome performance. This means that **Africa's macro-reform agenda — the centrepiece of the Refreshed Strategy's economic transformation vision — is being financed through the instrument that demonstrates the lowest probability of achieving Satisfactory outcomes in weak institutional settings.**

> *DPF conditionality assumes that governments have both the political will and institutional capacity to sustain policy reform commitments after disbursement. In Africa's most institutionally constrained settings, this assumption has been wrong three-quarters of the time. The instrument is not being adapted to the evidence.*

Programme-for-Results (PforR) performs materially better, approaching a mid-50 percent S+ rate — significantly above both IPF and DPF. Although PforR operates with a smaller sample size and is not typically deployed in the most extreme fragility contexts, the divergence from DPF is too large to

attribute to selection effects alone. Results-based disbursement against independently verified outputs creates incentive structures that input-based financing and policy conditionality do not. **The Bank has a better-performing instrument. It is not the one dominating the Africa portfolio.**

Country Variance: What the Dispersion Reveals

Eastern and Southern Africa records aggregate S+ rates near 30 percent, but this average conceals sharp dispersion. The Democratic Republic of Congo, with an S+ rate near 14 percent, and Mozambique, near 21 percent, exert a strong downward pull on the regional mean. In Western and Central Africa, performance similarly clusters around 30 percent overall, yet Nigeria records approximately 60 percent S+ while Sierra Leone records zero S+ across seventeen evaluated projects over a decade.

The differentiating variable is not the conflict environment. It is whether the Bank adapted its instruments and delivery architecture to the institutional reality on the ground.

> *The Africa case reinforces the central thesis of this report: theme is not the issue. The Bank's jobs and economic transformation agenda is correctly framed. Delivery architecture, instrument selection, and institutional capacity are the binding constraints — and the instrument portfolio is weighted toward the instruments most likely to fail in the settings where they are most heavily deployed.*

Case Study: Nigeria — The Economic Transformation Sectors Have Never Been Made to Work

Nigeria is the largest IDA borrower in Africa and one of the World Bank's most significant country programmes globally. It is classified as a Blend country — accessing both IBRD and IDA resources — and is designated Non-FCS in the IEG database for the period under review, though the north-east experienced active Boko Haram conflict throughout most of FY2015–2026 and several states are classified as fragile sub-nationally. Nigeria therefore provides a **revealing test case:** if delivery challenges of this severity are observable in a Non-FCS Blend country with relatively strong federal counterpart capacity, they are predictably worse in the Bank's formal FCS portfolio.

Across the 48 IEG-evaluated projects with evaluation years in FY2015–2026, Nigeria received $9.4 billion in total commitments. Only 11 of 51 projects (21.6%) achieved S+. In value terms, **$7.0 billion of the $9.4 billion total — 74.5% — went to projects that did not reach the Satisfactory threshold.** Nigeria's headline rate exceeds the FCS average of 30%, reflecting its Blend classification and stronger federal institutions. What the headline conceals is more important than what it shows.

Nigeria: IEG Development Outcome Ratings by Global Practice, FY2015–2026

Global Practice	Projects	S+ (S or HS)	Below S+	Total Commitment ($m)
Agriculture and Food	5	2 (40%)	3	$707m
Education	4	1 (25%)	3	$525m
Energy & Extractives	1	0 (0%)	1	$200m
Environment, Natural Resources & the Blue Economy	2	0 (0%)	2	$906m
Finance, Competitiveness and Innovation	5	2 (40%)	3	$1,425m
Governance	5	3 (60%)	2	$1,265m
Health, Nutrition & Population	7	3 (43%)	4	$1,341m
Macroeconomics, Trade and Investment	5	0 (0%)	5	$750m
Poverty and Equity	1	0 (0%)	1	$10m

Global Practice	Projects	S+ (S or HS)	Below S+	Total Commitment ($m)
Social Protection & Jobs	3	2 (67%)	1	$1,000m
Transport	5	0 (0%)	5	$963m
Urban, Resilience and Land	2	0 (0%)	2	$400m
Water	3	0 (0%)	3	$570m
TOTAL — All Global Practices	**48**	**13 (27%)**	**35 (73%)**	**$10,062m**

Source: IEG ICRR/PPAR evaluations, Evaluation FY2015–2026. S+ = Satisfactory or Highly Satisfactory only; Moderately Satisfactory does not qualify. Where both ICRR and PPAR evaluations exist, PPAR rating is used. Commitment figures from World Bank Projects database (as of 18 February 2026). Nigeria is classified Non-FCS/Blend for this period. 48 unique projects evaluated; all 48 matched to commitment data. Zero S+ outcomes shown in red.

Seven Global Practices, $4 Billion, Zero Projects Rated Satisfactory

> **Seven of thirteen Global Practices recorded zero S+ outcomes:** Water, Urban Resilience and Land, Energy & Extractives, Environment, Macroeconomics/Trade & Investment, Transport, and Poverty and Equity. Together these seven account for **$4.0 billion in commitments with no project reaching the Satisfactory threshold.** These are not peripheral operations. They include the Bank's largest infrastructure and economic reform investments in its most significant African country programme.

The Water sector record is the most surprising. Three successive National Urban Water Sector Reform Projects — First, Second, Third — were each rated Moderately Unsatisfactory, absorbing $120m, $200m, and $250m across nearly two decades of sequential investment. Each was designed explicitly in response to the failures of the previous one. Each failed for the same reasons: governance constraints in state-level utilities, weak political economy for tariff reform, and counterpart institutional capacity that did not improve at the rate project designs assumed. After $570 million and three attempts, the Bank has **not produced a single project rated Satisfactory in Nigerian urban water.**

Transport is equally surprising in scale. Five projects and $963m with no S+ outcome, including the Federal Roads Development Project ($333m, Moderately Unsatisfactory) and two successive Rural Access and Mobility Projects ($60m and $230m, both Moderately Unsatisfactory). The repetition of the instrument — investment in road infrastructure in a context where maintenance systems and counterpart recurrent budget are structurally insufficient — mirrors the water pattern exactly: **sequential failure without portfolio-level diagnostic response.**

Macroeconomics, Trade and Investment recorded 0% S+ across five Development Policy Operations and $750m in disbursements. Every single one was rated Moderately Satisfactory at best, meaning that $750m in policy-based lending to Lagos and Edo State produced no operation that the Bank itself rated as genuinely Satisfactory. This is the Nigeria DPF record in a context with above-average counterpart capacity and functional state-level government. It is the most direct evidence available of the instrument's limitations in reform-constrained settings.

What Works — and What It Tells Us

The only Global Practices to achieve majority S+ outcomes are Governance (60%, 3 of 5), Social Protection & Jobs (67%, 2 of 3), Health (43%, 3 of 7), Finance/Competitiveness (40%, 2 of 5), and Agriculture (40%, 2 of 5). The pattern is not random.

Governance's relative strength — anchored by the States Fiscal Transparency, Accountability and Sustainability project ($750m, S) and two sub-national governance reform projects — reflects the

comparatively direct link between policy dialogue and measurable institutional change at the state level. Social Protection's two S+ outcomes — the National Social Safety Nets Project ($500m, S) and the Community and Social Development Project ($200m, HS) — reflect a **structural advantage: cash transfer and community-driven delivery is operationally less dependent on the quality of government counterpart institutions** than infrastructure or regulatory reform lending. Health's performance, anchored by the Malaria Control Booster Project, the Polio Eradication Support Project, and the Basic Healthcare Provision Fund, reflects clearly defined disease targets and well-established vertical delivery mechanisms.

> The instruments and sectors that succeed in Nigeria share a common feature: **observable outcomes that can be verified without depending on sustained counterpart institutional reform.** The instruments and sectors that fail share the opposite: they assume that governance, regulatory, or institutional constraints will improve on project timelines. **In Nigeria, they have not — across decade-long series of operations.**

The Nigeria evidence illustrates precisely what the Refreshed FCV Strategy's private sector and infrastructure focus must contend with. The sectors most central to the strategy's jobs and economic transformation agenda — energy, urban infrastructure, financial market development, macroeconomic reform — are precisely the sectors where sustained delivery failure is most entrenched. The Bank has not been unable to design these projects. It has been **unable to resolve the political economy, institutional capacity, and governance constraints that make them fail.**

> That is not a project design problem. It is an upstream diagnostic and counterpart-readiness problem that a GP-level performance record of this consistency — across Water, Transport, Energy, Urban, and Macroeconomics simultaneously — should, by now, have prompted systematic institutional reform.

4. Root Causes: Why the Indicators Are So Poor

IEG rates each project across five quality criteria. Across the 1,403 evaluated IDA projects with data, none of the five criteria is met by a majority of projects in either IDA or FCS contexts. **Quality at Entry is the weakest — met in fewer than 28% of IDA projects — confirming that most delivery failures originate upstream at design, not downstream during implementation.**

Quality Criterion	IDA S+%	FCS S+%	IDA+FCS S+%	Implication
Quality at Entry (QaE)	29.1%	33.3%	31.2%	Projects designed to be approved, not to succeed; upstream origin of most failure
Bank Performance (overall)	27.1%	30.8%	29.0%	Under one-third of projects meet overall Bank performance bar
Quality of Supervision	43.8%	46.2%	44.5%	Better than QaE but still minority; structural underfunding of field supervision
M&E Quality	47.6%	47.1%	44.9%	80 years of M&E requirements; still below 50%; incentive problem not technical
Development Outcome	30.2%	28.0%	26.7%	One in four IDA+FCS projects achieves a satisfactory development outcome

The multi-criteria picture is starker still. Only **17.2%** of IDA projects (242 of 1,403) meet all five criteria simultaneously — the share of projects where the Bank's own quality systems, design, supervision, and M&E all functioned adequately alongside a satisfactory development outcome. At the opposite end, **40.5%** of IDA projects (568 projects, $42.1bn) met **none** of the five criteria. In FCS, 42.8% meet zero of five. In IDA+FCS, 45.3% meet zero of five and $17.2bn was committed to projects at this level.

The IEG evaluation data is consistent with four institutional pathologies identified repeatedly across the RAP series, portfolio assessments, and independent evaluations. These findings recur across a decade of reviews. They are known, well-documented, and the institution has chosen — or been unable — to resolve them.

4.1 Quality at Entry: Designed to Be Approved, Not to Succeed

Average project preparation time is 24 months. The question is not whether the Bank is rushing projects to Board — it is spending two years preparing them. The problem lies elsewhere. Project design in the Bank is fundamentally a negotiation between what TTLs know to be feasible and what operational vice presidencies, sector boards, and country management units need to demonstrate in their lending programmes.

Projects are designed to be approved. The results frameworks that accompany them — theory of change, indicators, implementation arrangements — are assembled to satisfy the review process rather than to reflect a credible operational logic. **In FCS settings, this manifests as projects technically sophisticated at the concept stage but structurally fragile in implementation: they assume government counterpart capacity that does not exist, set disbursement timelines calibrated to stable-country norms, and contain indicators that cannot be observed in active conflict environments.**

The Quality Assurance Group (QAG), which provided independent review of appraisal quality, was disbanded in 2014. Global Practices were expected to absorb its function. **IEG data shows only 27.5% of IDA projects are rated Satisfactory or above on Quality at Entry**, confirming that no effective independent quality check at preparation has replaced QAG. IDA21 introduces the Grant Facility for Project Preparation ($300m over three years) — welcome but marginal against the scale of the QaE problem. The OEE Dashboard simultaneously tracks preparation time with the directional arrow pointing toward shorter times, directly countervailing FCS quality needs.

4.2 Quality of Supervision: A Reporting Budget, Not an Engagement Budget

There is no single fixed "average supervision budget" that applies across all World Bank projects; it varies by size, sector, complexity, and region. However, portfolio data provide useful benchmarks. For IDA-financed projects, the average supervision (implementation support) cost is approximately $277,000 per project (FY25 terms), reflecting staff time, fiduciary and safeguard support across the portfolio. Project planning documents often budget supervision at roughly $180,000 per year in the early years and $120,000 per year in later years of implementation.

In fragile and conflict-affected situations (FCS), supervision costs are materially higher. Average supervision costs increased from approximately $255,000 per project in 2019 to roughly $420,000 in 2024 in FCS contexts, reflecting the need for more intensive field engagement, capacity support, and security arrangements.

Even at these higher levels, supervision typically represents well under 1 percent of total project

commitments. The resource envelope therefore remains modest relative to the operational demands of complex implementation environments.

In FCS contexts, where implementation requires intensive hands-on engagement, frequent field presence, and real-time adaptive management, $400,000 is not a supervision budget. It is a reporting budget. It funds the ISR, not the engagement. The consequence is predictable and documented: ISR ratings diverge systematically from eventual IEG outcome ratings because TTLs are rating the project against what they know from desks and periodic missions, not against implementation reality. IEG has documented this 'optimism gap' in every RAP report since 2015.

The Refreshed Strategy commits to adaptive management in FCS contexts. Adaptive management cannot be operationalised on a supervision budget that, after staff overheads and Washington-based costs, leaves field teams with resources for perhaps one or two missions per year per FCS project. **If the Bank is serious about FCS delivery, supervision resource allocation in FCS must be treated as a first-order budget question.**

4.3 M&E Quality: An Incentive Problem, Not a Technical One

The World Bank has financed development projects for 80 years. M&E frameworks have been mandatory components of project design for decades. IEG M&E ratings have been refined through multiple iterations. And yet approximately half of Bank projects still achieve only Modest or lower M&E ratings — a proportion that has not materially improved over three decades of measurement.

The persistence of this failure is not a technical problem. TTLs who invest in building genuine M&E systems — meaningful baselines, realistic data collection, indicators that detect whether the project's theory of change is working — face two institutional disincentives. First, robust M&E makes project failure visible and attributable: it removes the ambiguity that allows ISR ratings to remain green while implementation deteriorates. Second, building real data systems in FCS contexts takes time, costs money, and requires government counterpart engagement that may not exist.

After 80 years, the diagnosis is settled: the Bank does not underinvest in M&E because it does not know how to do it. It underinvests because no one is penalised for doing it poorly, and good M&E creates accountability risks — of visible failure, attribution, and scrutiny — that inadequate M&E avoids.

5. The Accountability Architecture: A System Designed to Diffuse Responsibility

Across the full IEG evaluation database, 60% of IPF operations — 4,304 of 7,141 evaluated — did not achieve Satisfactory development outcomes. 70% of the $52.7bn committed to FCS countries fell below the S+ threshold. These are not isolated failures. They are the documented output of an institution whose accountability architecture across four interlocking dimensions — outcome attribution, project-level blame, staff incentives, and Board governance — is designed to diffuse responsibility rather than attach it. This section examines all four and identifies what structural reform requires.

> **Core finding:** The Denizer-Kaufmann-Kraay study of over 6,000 World Bank projects establishes that country-level factors explain only 20% of project outcome variation. The remaining 80% is attributable to Bank-level decisions — design quality, supervision investment, TTL performance. IDA21's Sustainable Development Finance Policy holds borrowing governments accountable for the 20%. No mechanism holds Bank teams

accountable for the 80%. This is the central governance issue of the current replenishment architecture.

5.1 The 80/20 Finding: Allocating Responsibility for Outcomes

Denizer, Kaufmann and Kraay (2013) — a World Bank Policy Research Working Paper using over 6,000 projects evaluated between 1983 and 2011 — established the finding precisely: country-level factors explain only 20% of the variation in project outcomes. The remaining 80% varies within countries across projects, attributable primarily to Bank-level factors including the quality of the Task Team Leader. TTL fixed effects are "of comparable importance to country fixed effects" in accounting for outcome variation. The individual TTL's track record across their other projects is a significant predictor of performance on any given operation.

This finding has not altered the Bank's evaluation culture. Published in the Bank's own working paper series and subsequently in the Journal of Development Economics, its core implication — that Bank decisions are as consequential as country context in determining whether projects succeed — has produced no institutional accountability mechanism. ICR narratives continue to foreground country context as the primary explanation for failure. The SDFP holds borrowers accountable for the 20%. No equivalent mechanism exists for the 80%.

5.1.1 The 20% That Is Legitimately the Country's: Project Cycle Evidence

The 20% attributable to country factors is real.

It manifests most directly in the project cycle itself in things like the period from Board Approval to Loan Effectiveness, during which government processes — parliamentary ratification, legal conditions, counterpart designation — control the pace. (**Annex G**) Across 560 IEG-evaluated FCS projects, this averages 152 days. Several FCS contexts extend well beyond 200 days — Congo Republic (324 days), Guinea (294 days), Kosovo (299 days), Nigeria (309 days) — reflecting genuine government process constraints. From Loan Effectiveness to Closing, implementation periods average 5.5 years. These timelines represent the quantifiable expression of counterpart capacity constraints that are legitimately outside the Bank's control. This is the 20%. The appropriate policy response is what the SDFP provides. The table below shows the regional picture.

Region	Countries	Projects	Avg BA→LE (days)	Avg LE→Closing (yrs)	Avg BA→Closing (yrs)
Eastern & Southern Africa	11	156	143	5.5	5.8
Western & Central Africa	14	230	170	5.8	6.3
Middle East, N. Africa & Afghanistan	3	63	94	5.7	6.1
Europe & Central Asia	2	13	279	6.9	9.1
East Asia & Pacific	9	64	124	5.2	5.5
Latin America & Caribbean	1	26	135	5.9	6.3
South Asia	1	8	128	5.7	6.1
OVERALL AVERAGE — All IDA & Blend FCS	—	560	152	5.5	5.9

Source: IEG ICRR–PPAR database matched to World Bank Projects database, 18 Feb 2026 (27,750 projects; 99.97% match rate). BA→LE = Board Approval to Loan Effectiveness. 543/560 projects have valid BA→LE data. Full country-level data in Annex L.

5.2 Nobody Is Responsible When Projects Fail: The Four-Party Blame Loop

When a World Bank lending operation closes with poor outcomes, a four-party attribution sequence plays out in ICRs and ICRRs with documented regularity**. (Annex H)** Each party attributes responsibility to another. None accepts it. The sequence is not accidental — it is the structurally predictable output of an accountability chain designed to distribute responsibility across time, roles, and actors so that no single party bears the consequence.

Party	Attribution in ICR / ICRR	Why This Succeeds Institutionally
Design TTL	Attributes poor outcomes to country context, political instability, or government capacity constraints outside Bank control.	Has typically rotated off the project before closure. No mechanism links design decisions to evaluation outcome. Average implementation period: 8 years. Average TTL tenure: under 2 years.
Closing TTL	Attributes failure to predecessor decisions and a design they inherited mid-implementation.	Has no career stake in what the design TTL decided. In the Chad programme, IEG documented 168 distinct TTLs across 44 projects — individual accountability is architecturally impossible.
Management	Attributes underperformance to borrower government capacity, ownership, or commitment — “factors identified as risks at design.”	The same management that approved the operation formally responds to the evaluation. The Management Action Record allows them to self-certify compliance with IEG recommendations.
Host Government	Primary target of ICR blame: “government ownership was insufficient”; “capacity constraints limited implementation.”	Has no right of reply in the ICRR. Counterpart officials have rotated. Cannot contest the attribution.

5.2.1 The Language of Evasion

ICR language conventions are structurally optimised to diffuse accountability. Three formulations recur across the below-S+ record with diagnostic frequency. “Implementation was affected by the challenging operating environment” applies the legitimate conflict-context exception to predictable challenges that were foreseeable at design and flagged in early ISRs without response — converting a supervision failure into a context narrative. “Government commitment and ownership were insufficient” attributes to the borrower a constraint the Bank’s own guidelines require to be assessed before approval; if it was already visible at design, it represents a preparation failure, not a country failure. “Key lessons for future operations include...” converts accountability for what happened into advice for some future TTL on some future project. IEG’s own 2016 Behind the Mirror evaluation found that ICR lessons “rarely translate into changed practice.” The same five failure modes have recurred in virtually every ICR for two decades. Staff know them. The designs do not change.

5.3 Staff Incentives: Why Over-Ambition Is Rational

Gaël Raballand, Thomas Roundell and Michel Mallberg examined all World Bank public sector and governance investment projects in Sub-Saharan Africa and MENA completed by 2014.

Their finding: "there appears to be little direct correlation between a poorly performing project and task leader career trajectory." (Annex I)

Three IEG evaluations spanning a decade restated the same conclusion. Learning and Results in World Bank Operations (2014): the Bank holds staff accountable for "project design and disbursements more than for fostering knowledge use, engaging clients in learning, and achieving influence." World Bank Group Outcome Orientation (2023): incentives "emerge from what is measured, tracked, and rated in managerial dashboards... and are not always aligned with outcome orientation." The management response to each evaluation rated the recommendation "Moderately Satisfactory" and identified process improvements that addressed nothing structural.

The World Bank's Overall Performance Evaluation demonstrably rewards Board approval of new operations, disbursement and pipeline management, reputational risk avoidance ("staff are aware that even a small error in a transaction can permanently harm a professional career" — CGDev, 2024), and technical credibility through publications and knowledge-sharing independent of project outcomes. It demonstrably does not reward S+ outcome rates as predictors of career advancement, accurate forecasting at appraisal, or learning from failure. 80% of TTLs in IEG's own survey reported that lending pressure crowds out learning.

The consequence is rational institutional behaviour. Projects are consistently designed with timelines (4 years) that are half the actual implementation period (8 years). Objectives are set at levels that exceed country capacity. M&E frameworks measure what was planned rather than what is achievable. Over-ambition signals capability and secures lending approvals. Underperformance at closure attaches to nobody — the responsible party has moved on. "The TTL at appraisal will never close a project" (Raballand et al.) describes the central structural feature of an accountability system without consequences.

> **The optimism bias at World Bank project design is not irrational.** It is the rational response to an incentive structure that rewards ambitious lending approvals and imposes no career consequences for project failure. Addressing it through training or guidance notes — as the Bank has repeatedly attempted, without effect — misidentifies a structural incentive as a behavioural weakness.

5.4 The Circular Evaluation System: Five Structural Failures

The self-evaluation architecture governing project accountability contains five structural failures that collectively ensure independent assessment cannot produce institutional consequences. They are not the result of bad faith. They are the predictable outcome of an evaluation chain that asks people to assess their own work, permits the evaluated to respond to the evaluation, and gives the institution that approved operations the task of receiving findings about those operations.

Problem	Mechanism	Evidence of Failure
1. Self-evaluation	The ICR is prepared by the closing TTL and project team. The team self-rates outcome, Bank performance, and M&E quality.	ICR ratings are higher than IEG ICRR ratings in every RAP cohort FY2020–2024. As evaluation becomes more independent — from ICR to ICRR to field-based PPAR — rated performance consistently declines. This gradient is direct evidence of systematic inflation.
2. Author appointment	The closing TTL controls who contributes to the ICR. No requirement for review by staff unconnected to the project.	Analogous to the Enron auditor-client conflict resolved by Sarbanes-Oxley. No equivalent independence reform has been applied to ICR authorship.

3. IEG rotation	IEG is staffed by World Bank Group employees some of who rotate between IEG and operational units. An evaluator may assess projects they worked on.	ADB's IED requires formal recusal from evaluating projects the evaluator worked on. The World Bank has no equivalent policy, despite ECG Good Practice Standards requiring avoidance of conflict of interest.
4. Management response	IEG sends findings to management before finalisation. Management publishes a formal response alongside the ICRR. The Management Action Record allows management to self-certify compliance with IEG recommendations.	Management consistently rates its own compliance with IEG recommendations more favourably than IEG's independent validation. No comparable institution gives the evaluatee a formal channel to contest the evaluator's findings.
5. Board feedback loop	The Board that approved the operations under review receives IEG's findings through the annual RAP — with no mechanism to hold management accountable without acknowledging co-responsibility for approving the operations.	"There is no annual session in which the Board asks management: you approved 7,141 IPF operations. Sixty per cent failed. Explain what you will do differently." The RAP is received, noted, and the approval cycle continues.

5.4.1 The ACA Gap: Advisory Services Without Accountability

The accountability gap is total for Advisory Services and Analytics (ASA). **(Annex J**) The Bank spends an estimated $1.0–1.5bn annually on ASA — Economic and Sector Work, Technical Assistance, analytical products. Unlike lending, which is formally evaluated through the ICR–ICRR chain, ASA has no equivalent. IEG's RAP 2022 found that Country Learning Reviews cover only one-third of ASA products in terms of documented use or influence; less than half report on higher-level influence; and Activity Completion Summaries are not independently validated. There is no mechanism equivalent to the ICR rating chain that generates a public, evaluable record of whether a given ASA product was used, influenced policy, or was simply produced and filed.

The remedy is the introduction of Activity Completion Assessments (ACAs) — structured self-evaluation documents for ASA products, validated by IEG through random-sample review, with downstream uptake tracking and performance metrics in GP accountability frameworks. IEG's forthcoming RAP 2025 designates ASA as its deep-dive evaluation topic for the first time since 2006, examining over 5,700 products from FY2017–24. Management should commit to the ACA framework before RAP 2025 findings are published — not as a response to them.

5.5 Board Co-optation: How Project Approval Destroys Portfolio Accountability

The World Bank's Board of Executive Directors is a resident body. **(Annex K)** Its 25 Directors maintain offices at Bank headquarters, receive institutional salaries, and meet approximately twice weekly to approve roughly 300 operations per year. The 2009 Zedillo Commission named the problem with precision: "The Board of Executive Directors attempts an impossible trinity of roles: political representation of member state interests, technical banking supervision, and management oversight. No board in the world can perform all three simultaneously." Board approval is not a passive step. It is a formal institutional endorsement. When an approved operation subsequently fails — as 60% do — the Board cannot scrutinise the outcome without simultaneously acknowledging responsibility for the approval. The institutional response is consistent: receive IEG's findings through the annual RAP, note them, and continue.

> The Board approved all 8,764 operations in the IEG evaluation database. *It held management formally accountable for the development results of none of them. The Board's approval function*

> *has successfully substituted for its oversight function, producing an institution in which management is unaccountable for the outcomes of operations it designs, approves, and supervises.*

5.5.1 How Peer Institutions Have Solved This

The World Bank's resident, project-approving Board is now the exception in MDB governance. The Asian Infrastructure Investment Bank (AIIB) adopted a non-resident Board as a foundational principle: project approval delegated to the President; the Board retains strategy, policy, and performance oversight. The New Development Bank (NDB) explicitly adopted the same model to avoid the incentive distortions of resident Board arrangements. The European Investment Bank — the world's largest multilateral lender by volume — operates with delegated lending authority and maintains AAA credit ratings. Ngaire Woods of Oxford identified the problem at the Bank's own 75th anniversary: "The resident Board works with far too much micromanagement of lending. Missing from all these processes is quality oversight of implementation and outcomes." The reform requires delegating project approval to management, establishing an annual management performance review based on portfolio outcomes, and having IEG report directly to a Board Audit Committee before management can respond.

Feature	World Bank	AIIB	NDB	EIB
Board residency	Resident (HQ staff, salaries)	Non-resident	Non-resident	Non-resident
Project approval	Board approves all operations	Delegated to President	Delegated to Pres./VPs	Delegated to management
Board meeting frequency	~75 meetings/year	~6 meetings/year	Quarterly	Periodic
Annual mgmt performance review	Not implemented	Board Performance Ctte	Annual Board review	Board assesses management
Evaluation reporting	Through RAP to full Board	Board Audit Committee	Project Eval. Committee	Board Audit Committee
Annual Board cost	~$50–70m	~$3–5m	~$3–5m	Non-resident equivalent

5.6 The Accountability Asymmetry: What IDA21 Must Still Fix

IDA21's Sustainable Development Finance Policy strengthens the framework under which borrowing governments are held accountable for outcomes. This is appropriate: governments that borrow public money to deliver public services should face performance consequences. The SDFP represents a genuine governance advance for the 20%.

But the framework is analytically indefensible without its symmetric component. Country-level factors explain 20% of project outcome variation. The SDFP holds countries accountable for that 20%. Bank-level factors — design quality, supervision intensity, TTL performance, evaluation honesty — explain 80%. IDA21 introduces no mechanism that connects Bank staff career outcomes to development results. The $100bn replenishment carries an elaborate results matrix and hundreds of performance commitments. It does not include a single instrument that makes Bank teams accountable for the quality of operations they design and supervise.

A symmetric accountability framework requires five structural changes, documented in full in Annexes M and N. (i) Mandatory IEG recusal rules: no evaluator assesses projects or country

strategies they had operational responsibility for, with a minimum three-year cooling-off period — equivalent to the ADB's IED standard. (ii) Attribution-disaggregated ICR reporting: ICRs separately assess Bank performance — design quality, supervision quality, TTL continuity — from country context, with findings tracked against staff records. (iii) Bank Performance Ratings published at country and Practice level with formal consequences: where Practice-level outcome rates fall below agreed thresholds, Management brings a corrective action plan to the Board. (iv) An At-Risk Portfolio Review: a formal quarterly mechanism in which Practice Managers present deteriorating projects to Senior Management with explicit attribution of which design or supervision decisions contributed. (v) Activity Completion Assessments (ACAs) for the ASA portfolio: IEG-validated self-evaluation with performance metrics in GP accountability frameworks, committed before RAP 2025 findings are published.

> *IDA21 introduced borrower accountability without introducing Bank staff accountability. The replenishment commitment to results cannot be credibly delivered by a framework that enforces accountability on one party to the development relationship and systematically ignores it on the other. These are structural failures. They will not be resolved by guidance notes, training programmes, or management commitments to "strengthen learning." They require governance reform: changed incentives, reformed evaluation architecture, and a Board structurally capable of holding management accountable for portfolio results.*

PART III

IFC's PRIVATE SECTOR RECORD

Track record and structural constraints to scaling up

6. IFC's Performance in the Markets That Matter Most

The Refreshed Strategy places significant weight on IFC's role in mobilising private capital for MSMEs, agribusiness, and productive employment in FCV settings. The IEG RAP series from 2020 to 2024 provides a specific and sobering baseline.

6.1 The Pattern: Recovery Followed by Renewed Collapse in FCS

Cohort (CY)	Overall S+%	FCS S+%	IDA/Blend S+%	Work Quality	Additionality
CY2015–17	~30–35% S+ (est.)	22%	n/a	Declining	Declining
CY2017–19	~56% (uptick)	Recovering	n/a	Improving	Improving
CY2019–21	64% S+	56% S+	n/a	60% S+	59%
CY2020–22	60% S+	~6–8% S+ (est.)	36% S+	55% S+	54% overall / 33% in FCS
CY2021–23	Declining	Declining (key drag)	~29% S+ (est.)	Improving slightly	Weakening in FCS/IDA

The 6-8% Satisfactory figure for FCS investment projects in the CY2020–22 cohort (RAP 2023) is the most important single data point in this section. It is **not only a pandemic anomaly**: it follows a period (CY2019–21) when IFC's FCS investment performance recovered to 56% — demonstrating that better outcomes are achievable. The collapse to 8% in the next cohort reflects structural volatility in a segment where IFC's institutional model is fundamentally ill-adapted. The strategy proposes to scale IFC's FCS presence precisely as the outcome data reaches its nadir.

6.2 Additionality: The Core IFC Promise Is Failing in FCS

IFC's distinctive contribution — beyond financing — is additionality: the provision of knowledge, technical standards, ESG improvements, and catalytic mobilisation of other investors that would not occur without IFC involvement. In FCS markets (CY2020–22), IFC realised its anticipated additionality in only **33% of projects**. The gap is widest for non-financial additionality — the dimension most critical to long-term market development:

- Knowledge and innovation additionality: 17 percentage points below anticipated levels
- ESG standards additionality: 22 percentage points below anticipated levels
- Financial mobilisation additionality: modest but more consistent

This finding matters because IFC's argument for PSW concessional capital is premised on additionality: public subsidy is justified by developmental impact beyond the financial return. If IFC is realising additionality in one-third of FCS investment projects, the development case for the PSW subsidy in FCS rests on a minority of deployments. The strategy should require disaggregated additionality reporting per replenishment cycle before scaling FCS commitments.

6.3 Advisory vs Investment: A Structural Divergence

IFC advisory services showed more consistent improvement across the RAP 2020–2024 period. Development effectiveness ratings for advisory projects in FCS improved, attributed partly to shorter project durations, smaller project sizes, and greater co-location of team leaders in the field following a 2017 IEG-IFC work quality study.

This divergence between advisory and investment performance is itself a finding. The mechanisms that improve advisory delivery — shorter cycles, field presence, reduced scope, direct client engagement — are harder to apply to large investment operations in fragile markets**. It suggests that the optimal IFC intervention model in FCS is advisory-first, with investment following only where client quality, counterpart institutions, and market conditions have been established through earlier advisory engagement. The current model does the reverse.**

7. Seven Structural Constraints to IFC Scale-Up in FCS

The 2022 IEG evaluation of IFC and MIGA support for private investment in FCS (FY2010–21) delivers a clear verdict: neither institution was able to scale up business volumes in FCS during the period despite successive strategy cycles and multiple new instruments. IFC's long-term FCS commitments averaged 5.2% of total commitments throughout FY2010–21 with no upward trend.

Seven structural factors explain why. Each is directly relevant to the Refreshed Strategy's private sector ambitions.

Constraint	What It Means	Why PSW/Financing Cannot Fix It
Shortage of bankable projects	Binding constraints in FCS are non-financial — weak governance, underdeveloped regulatory frameworks, and integrity and E&S due diligence concerns	The PSW is principally designed to address financial risk. It cannot create governance, rule of law, or regulatory capacity. Non-financial binding constraints require Bank-IFC coordination on upstream enabling environment, not cheaper capital.
PSW failed to produce volume	Despite $2.5bn in allocated IDA18 funds, the PSW did not generate a net increase in IFC commitments in PSW-eligible countries; non-PSW IFC commitments in PSW-eligible countries actually fell in aggregate	Additionality of the PSW instrument itself is unproven. IDA21 restructures PSW governance but does not resolve the fundamental substitution risk: PSW may replace rather than supplement IFC own-account FCS activity.
Prohibitive cost of doing business	Operating in FCS costs 2.5× more per dollar committed ($48/$1,000 vs $19/$1,000 in non-FCS), compressing margins and deterring deal flow	Cost structure is structural, not cyclical. Concessional capital reduces the cost of money but not the cost of operations. Unless staff incentives, field presence, and transaction cost subsidies are addressed, cheaper capital alone will not change the deal economics.
Negative risk-adjusted returns	IFC's internal review (FY2011–15) found negative risk-adjusted returns across all industry groups in FCS; NPL rates at 4× non-FCS levels	Negative risk-adjusted returns mean IFC is not just failing to make money — it is destroying value in aggregate on its FCS portfolio. No amount of replenishment financing resolves an adverse risk-return profile that is intrinsic to the market segment.
Institutional sustainability	IEG scenario analysis: scaling IFC commitments in IDA LIC and FCS markets from 6% to 15% of total would reduce annual net income by ~$90m	IFC's financial model depends on profits from middle-income markets that cross-subsidise IDA/FCS exposure. Rapid scaling of FCS share would require either (a) a fundamental change in IFC's financial model or (b) recognition that this is a grant-funded, not commercially-driven, activity.
Portfolio concentration	Just six countries (DRC, Lebanon, Nigeria, Mozambique, Myanmar, Congo) accounted for 54% of IFC's FCS business in FY2010–21 — the most commercially accessible FCS markets	The landlocked, low-FDI, lowest-CPIA markets central to the strategy's ambitions are precisely where IFC has been least present. Scaling the six-market concentration is not the same as scaling FCS reach.
Client quality trap	Repeat-client projects performed at 84% S+ in FCS vs 29% for new clients; IFC only 4% of FCS investment projects led by staff in-country at commitment	Institutional bias toward established sponsors limits market development. Field presence stagnated from FY13 to FY19. Decisions are made from Washington about markets that must be understood from the ground.

7.1 What Must Change Before Scaling

The Refreshed Strategy commits to strengthening and scaling IFC's FCS engagement. The evidence above indicates that scaling without prior structural reform will amplify the existing performance gap. Three prerequisites are necessary before FCS scaling targets are set:

- **Disclose IFC development impact baselines per instrument and FCS tier**. IFC's FCS investment outcome rate stood at 11% Satisfactory (CY2020–22). Scaling targets should be conditioned on improvement from this baseline, not set in parallel with it. The strategy should commit to annual reporting of IFC development outcome rates disaggregated by FCS country category and instrument type.
- **Resolve the PSW additionality question**. IDA21 provides $3.2bn to the PSW plus an IFC Concessional Capital Window. The IDA21 governance restructuring reduces Board visibility into individual PSW transactions. Before deploying this capital, a mandatory FCS additionality assessment per replenishment cycle should be required — confirming that PSW is generating net increases in FCS commitments, not substituting for IFC own-account activity.
- **Build the enabling environment before the investment pipeline**. The binding constraints in most FCS markets are non-financial. **The advisory-first model — establish regulatory frameworks, build market systems, develop client capacity, then introduce investment — has a better empirical track record than leading with investment capital.** The Jobs and Economic Transformation Compact proposed in the strategy provides a vehicle for this sequencing if operationalised correctly.

PART IV

IDA21 and the Forward Look

What IDA21 does, what it advances, and where structural gaps remain

8. What IDA21 Is: Architecture and Scale

The 21st Replenishment of IDA, approved by the Board of Governors on 17 March 2025 and effective from 1 July 2025, commits up to **$100 billion** in total commitment authority for FY2026–28 — the largest in IDA's history and an 8% nominal increase over IDA20's $93 billion. The overarching theme is 'Ending Poverty on a Livable Planet.' IDA's hybrid financing model (established IDA18) mobilises approximately $3.50 of commitment authority per $1.00 of Donor grant contribution; more than 70% of IDA20 resources came from non-Donor sources.

8.1 Funding Structure and Allocation Windows

Allocation / Window	Size	Purpose and Key Features
Country Allocations (total)	$67.2bn	Performance-based, country-led; 58% of total replenishment
Performance-Based Allocation (PBA)	$58.1bn	65% of concessional resources; simplified composite terms; 10% grant volume haircut; annual cap reduced to $650m/year

Allocation / Window	Size	Purpose and Key Features
FCV Envelope	$8.8bn	PRA (prevention/resilience), RECA (conflict engagement), TAA (turn-around); enhanced upstream eligibility and grant flexibility for CPIA ≤2.5
GROW Window (new)	$15.9bn	Replaces Regional Window; ring-fenced WHR sub-window ($2.4bn); fully concessional; no PBA contribution required
Scale-Up Window	$10.0bn	$7.0bn non-concessional + $3.0bn concessional SML; expanded eligibility for Regional Organizations
Crisis Response Window	$3.7bn	Last-resort financing; $1.0bn early response facility; grant terms for red-light countries
Private Sector Window	$3.2bn	IFC/MIGA deployment; new public dashboard; + $0.5bn IFC Concessional Capital Window
IDA-GFPP (new)	$0.3bn	Project preparation grants; up to $100m/year; all instruments eligible
TOTAL COMMITMENT AUTHORITY	**$100bn**	8% increase on IDA20; grant element $53bn (54%); IFC Concessional Capital Window adds $0.5bn on IFC balance sheet

IDA21 also reduces country-level policy commitments from over 1,000 in IDA20 to **25 strategic Policy Commitments (PCs)** organised across five Focus Areas (People, Planet, Prosperity, Infrastructure, Digital Transformation) and four Lenses (Gender, Jobs, Private Investment, FCV). The WBG Scorecard introduces 22 outcome indicators disaggregated by FCS status for the first time, and a Facetime Index tracking staff presence in FCS countries is introduced in the OEE Dashboard.

8.2 The FCV Envelope: $8.8bn, Three Tracks

The FCV Envelope retains its three-tier architecture with targeted changes:

- **PRA (Prevention and Resilience Allocation):** Eligibility broadened for earlier upstream engagement before formal conflict thresholds — directly implementing the FCV MTR recommendation on prevention.
- **RECA (Remaining Engaged in Conflict):** Per-cycle cap on top-up raised to $400m. Management discretion to provide grant terms to countries with CPIA ≤2.5 and to channel resources through third-party implementers (UN agencies, NGOs) in zero-capacity settings.
- **TAA (Turn-Around Allocation):** Top-up reduced from 125% to 100% of PBA. Annual transparency reporting introduced. As of Q1 FY25, FCS countries receive 51% of all Country Allocations.

9. What IDA21 Gets Right

IDA21 contains several genuine advances that deserve direct acknowledgement:

- **FCS disaggregation of all 22 Scorecard indicators** is a genuine transparency commitment. For the first time, the Bank will systematically track whether FCS performance is keeping pace with overall portfolio performance across all outcome dimensions.

- **The Facetime Index** operationalises the link between staff field presence and FCS delivery quality — addressing one of the most persistently cited structural weaknesses in the IEG FCS evaluations.
- **The GFPP ($300m)** creates for the first time a dedicated funding stream for project preparation — addressing the upstream design quality problem at source, though at a scale that remains small relative to the quality gap.
- **SimplifIDA (under 500 country-level actions vs 1,000+ in IDA20)** is operationally rational for low-capacity FCS borrowers. Fewer, more strategic policy commitments are more implementable and more measurable than proliferating country-level conditions.
- **IFC Concessional Capital Window ($0.5bn ring-fenced on IFC balance sheet)** creates institutional skin-in-the-game that the PSW previously lacked, providing alignment incentives for IFC on FCS investment quality.
- **RECA grant flexibility and third-party implementation permission** for CPIA ≤2.5 countries is the most important operational improvement in the FCV Envelope, enabling the Somalia model to be deployed systematically rather than case-by-case.

10. The Structural Gaps: What IDA21 Does Not Fix

10.1 No Operational PCs on Quality at Entry, Supervision, or Bank Performance in FCS

IDA21's single FCV-specific Policy Commitment (PC14) requires CPFs and Engagement Notes for FCS countries to include strategic objectives addressing FCV drivers and resilience sources. The baseline is 50% as of FY24. This is a planning-stage commitment about document content, not an operational commitment about delivery quality. **IDA21 contains no specific PC for Quality at Entry in FCS, disbursement rates in FCS, or project restructuring practice in FCS** — the three operational dimensions where the IEG evidence most clearly identifies failure.

This creates a two-to-three year window in which the institution will approve and design FCS projects at a rate consistent with 51%+ of Country Allocations going to FCS, without the operational commitments the evidence already warrants. The IDA21 Mid-Term Review — not scheduled until approximately mid-FY28 — is the first gate at which additional PCs can be introduced. **The gap should be addressed by establishing FCS operational baseline PCs at the outset of the replenishment cycle, not waiting for the MTR.**

10.2 PSW Additionality: Not Resolved by Governance Restructuring

IDA21 restructures PSW governance from a 'retail' model (dual Board approval of individual transactions) to a 'wholesale' model (IDA Deputies approve overall allocation; IFC/MIGA Boards approve individual transactions; IDA Management advisory). This is operationally sensible — it reduces transaction costs and speeds IFC decision-making.

What it does not address is the additionality problem documented by IEG and CGD: **non-PSW IFC own-account commitments in PSW-eligible countries actually fell during IDA18–IDA20**, and private capital mobilisation per dollar of IFC own-account commitment was lower in PSW-supported projects than in non-PSW projects. The wholesale governance model, by reducing IDA Board visibility into individual transactions, may further weaken the external accountability mechanism intended to prevent PSW from substituting for rather than catalysing deeper IFC engagement in FCS.

10.3 RECA Reach: A Mechanism That Has Not Reached Its Intended Beneficiaries

The RECA was designed for states experiencing high-intensity conflict with extremely limited government capacity. In practice, it was applied to only two countries across IDA19 and IDA20: South Sudan and Yemen. The cap increase to $400m per cycle is a calibration, not a redesign.

The structural problem is that RECA eligibility requires states to maintain annual reviews and engage through government channels — a requirement that is, by definition, unworkable in zero-capacity contexts such as Afghanistan after 2021, Sudan after 2023, or Haiti. IDA21's introduction of management discretion to provide grants for third-party implementation in RECA countries with CPIA ≤2.5 is a meaningful improvement. But it remains discretionary. The successor FCV Strategy should define a systematic third-party implementation framework for zero-capacity contexts — with provider accountability standards, minimum service delivery commitments, and an explicit transition pathway — as a standard modality rather than an exception.

10.4 The Scorecard: Visibility Is Not Accountability

IDA21's Scorecard is the most ambitious results measurement architecture in IDA's history, and the FCS disaggregation is genuinely significant. But the IDA20 experience illustrates the gap between visibility and accountability: FCS outcome data was tracked throughout IDA20, the MTR documented persistent quality concerns, and the corrective action was to deepen monitoring rather than restructure incentives.

The IDA21 framework has no pre-agreed corrective mechanism triggered by Scorecard underperformance. **When FCS indicators underperform against targets — as the evidence suggests they will — there is no automatic governance response.** The recommendation is a formal **FCS Results Gap Protocol**: a defined governance process by which Management brings a corrective action plan to the Board when FCS Scorecard outcomes fall below agreed thresholds across two or more consecutive Annual Meeting reporting cycles. This transforms the Scorecard from a transparency instrument into an accountability instrument.

10.5 The Borrower Accountability Asymmetry

IDA21's Sustainable Development Finance Policy (SDFP) extends and strengthens the framework under which borrowing governments are held accountable for development outcomes. This is appropriate. But the SDFP creates an accountability asymmetry that is difficult to justify in light of the Denizer-Kaufmann-Kraay evidence: **country-level factors explain only 20% of project outcome variation**. A framework that holds the country accountable for the 20% while not holding Bank design and supervision teams accountable for the 80% is not a governance framework for development effectiveness. It is a framework that deflects accountability toward the weaker party in the financing relationship.

IDA21 introduced borrower accountability without introducing Bank staff accountability. The replenishment commitment to results cannot be credibly delivered by a framework that enforces accountability on one party to the development relationship and ignores it on the other. The technical reforms required — mandatory recusal rules in evaluation, attribution-disaggregated ICR reporting, Bank performance ratings linked to career consequences, a formal FCS At-Risk Portfolio Review — are set out in Annex N.

11. Consolidated Recommendations

The following eight recommendations are offered in priority order. The first four address the most immediately actionable institutional reforms; the second four address structural architecture changes that require Board-level decisions.

#	Recommendation	What It Requires	When
1	**Establish a delivery baseline with financial accountability**	Formally acknowledge that 71.7% of committed funds in jobs-critical IDA GPs did not achieve Satisfactory outcomes (2015–2025). Set time-bound S+ improvement targets tracked through the IDA21 MTR, framed in both project-count and committed-dollar terms.	IDA21 MTR (mid-FY28)
2	**Reform MSME/FCI model before scaling**	FCI at 23.8% S+ in IDA shows no sustained improvement trend. Scaling targets should be conditional on explicit QaE improvement benchmarks and accountability mechanisms, not set parallel to them. QAG equivalent for FCS QaE review should be established.	FCV Strategy launch
3	**Disclose IFC FCS baselines before setting scaling targets**	IFC's FCS investment outcome rate stood at 11% (CY2020–22). Publish annual IFC development outcome rates disaggregated by FCS country tier and instrument type. Make further FCS scaling conditional on measurable improvement from this baseline.	Immediate / Annual
4	**Require mandatory FCS additionality assessment for PSW**	Confirm that each $bn deployed through PSW is generating net increases in IFC FCS commitments, not substituting own-account activity. IDA21's reduced Board visibility through wholesale governance makes an independent additionality verification mechanism more, not less, important.	IDA21 Year 1
5	**Introduce FCS operational Policy Commitments for IDA21**	Add PCs on Quality at Entry (minimum S+ rate), disbursement rates, and project restructuring practice for FCS operations, without waiting for the MTR. The evidence base already warrants these commitments.	Board: FY26
6	**Establish an FCS Results Gap Protocol**	Define a formal governance process triggered when FCS Scorecard outcomes fall below agreed thresholds across two consecutive Annual Meeting cycles. Management brings a corrective action plan to the Board. Converts the Scorecard from visibility to accountability.	IDA21 MTR
7	**Systematise third-party implementation for RECA-eligible countries**	Convert IDA21's discretionary permission for third-party implementation in CPIA ≤2.5 RECA countries into a standard modality with provider accountability standards and transition pathways. Somalia's model (67% S+ in active conflict) is the template.	FCV Strategy
8	**Align Board governance with portfolio oversight**	Individual project approval by the Board co-opts governance into the approval culture. Move the Board toward portfolio-level oversight — sector performance reviews, GP outcome targets, management accountability for aggregate outcome rates. Reform the Board Governance structure as described in Annex N.	Medium-term

References and Data Sources

Primary Data

- Own analysis. IEG ICRR-PPAR Evaluation Database (accessed February 2026) matched to World Bank Projects Database (27,750 projects, data as of 18 February 2026) by Project ID (P-code). Match rate: 7,772 of 7,774 IEG-evaluated projects (99.97%). S+ = Satisfactory or Highly Satisfactory only.

IEG RAP Series

- IEG (2020–2024). Results and Performance of the World Bank Group, RAP 2020, 2021, 2022, 2023, 2024. Washington DC: World Bank Group IEG.

IEG Thematic Evaluations

- IEG (2022). IFC and MIGA Support for Private Investment in Fragile and Conflict-Affected Situations, FY2010–21. Washington DC: World Bank Group IEG.
- IEG (2016). Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group. Washington DC: World Bank.
- IEG (2015). External Review of the Independent Evaluation Group. Washington DC: World Bank.
- IEG (2021). Chad Country Programme Evaluation. Washington DC: World Bank.

IDA Replenishment

- IDA (2025). Securing Our Shared Future Together: Report from the Executive Directors on the Additions to IDA Resources — Twenty-First Replenishment (IDA21). Approved 17 March 2025, effective 1 July 2025. Washington DC: World Bank.

Academic and Policy Literature

- Denizer, C., Kaufmann, D., and Kraay, A. (2013). Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance. Journal of Development Economics, 105: 288–302.
- Gurara, D., Le Borgne, E., Morris, R. and Wane, W. (2023). Is the IDA Private Sector Window Crowding In Private Investment? CGD Policy Paper 281. Washington DC: Center for Global Development.
- Raballand, G., Roundell, K., and Mallberg, F. (2015). Bank Staff Incentives and Behaviors: Lessons from Public Sector Governance Projects. World Bank presentation.
- Wappenhans, W.A. et al. (1992). Effective Implementation: Key to Development Impact. Report of the Portfolio Management Task Force. Washington DC: World Bank.
- Woods, N. (2019). Remarks at World Bank 75th Anniversary Celebrations. Cited in IEG governance assessment literature.
- Zedillo Commission (2009). Modernizing the World Bank Group: Report of the High-Level Commission on Governance. Washington DC: World Bank.

Technical Annexes in This Report

The following technical annexes are attached to this report and contain the detailed analysis, data tables, and instrument-level findings that underpin the main body:

- Annex A — Jobs, Private Sector and the Analytical Framework
- Annex B — S+ Benchmark Justification: Why Moderately Satisfactory Is Not an Acceptable Threshold
- Annex C — IEG RAP 2020–2024 Compiled Performance Data
- Annex D — Investment Project Financing: Outcome Performance and Lessons Learned (n=7,141)
- Annex E — Development Policy Financing: Outcome Performance and Lessons Learned (n=1,551)
- Annex F —Programme-for-Results: Outcome Performance, Design Logic, and Lessons Learned (n=72)
- Annex G — Project Cycle Timeline Data: Board Approval to Closing by Country
- Annex H — Project-Level Accountability: The Circular Evaluation Problem and the Blame Loop
- Annex I — Staff Incentives and Institutional Behaviour: Raballand et al. Analysis
- Annex J — Advisory Services and Analytics 2020 to 2025 – The Case for Periodic Independent Review
- Annex K — Board Governance and Co-optation: The Accountability Architecture Flaw

ANNEX A — Jobs, Private Sector, and the Analytical Framework

Jobs, Private Sector and the Analytical Framework

There is consensus that the priorities for FCV countries should be on jobs and the private sector. There is a fifteen year history on this. Identifying the priority has not been the issue. Delivering on it has been.

A brief summary on the key findings of Bank reports is attached below.

1. WDR 2011: Conflict, Security, and Development

The 2011 World Development Report treated jobs as one of three foundational pillars of stability — alongside security and justice — arguing that when markets fail to provide employment opportunities, the likelihood of violence rises significantly.
Jobs as a peacebuilding tool. Unemployment — especially among youth — was identified as a major driver of grievance and recruitment into armed groups or criminal networks. Job creation was not just a development goal but a conflict prevention mechanism.
Early wins through employment. The report placed particular emphasis on delivering rapid, tangible results through early job creation to build trust in new institutions and governments. Quick employment schemes were seen as confidence-building measures, with a specific call for job creation through the private sector at an early stage of post-conflict transitions.
Practical instruments recommended. These included large-scale public works and community-based employment schemes, provided they did not crowd out private sector activity. Alongside these, the report recommended expanding access to finance and broadening access to assets, skills, and work experience — all intended to catalyze private-sector-led growth over time.
Private sector as contingent partner. The 2011 report was cautious about the private sector's immediate role. It recognized that private investment is crucial for sustainable job creation but acknowledged that conflict environments are deeply unattractive to investors. The framing was primarily that public schemes should not displace eventual private sector activity, rather than placing the private sector at the center of the response immediately.
Institutional pre-conditions. The report argued that without legitimate institutions capable of delivering security and justice, the private sector would not invest and labor markets would not function. Jobs and private sector development were therefore dependent on getting governance right first.

---

2. Pathways for Peace (2018, UN–World Bank)

The 2018 joint UN–World Bank study shifted focus from post-conflict recovery to conflict prevention, giving the private sector a more prominent and proactive role.
Exclusion from economic opportunity as a root cause. Pathways for Peace placed exclusion — from power, resources, opportunity, and security — at the center of its analysis. Lack of economic inclusion, particularly for youth and marginalized groups, creates grievances that fuel conflict. This made job creation and economic access not just a recovery tool but a preventive one.
Private sector as a key partner in prevention. Unlike the 2011 report, Pathways for Peace explicitly identified the private sector as a necessary coalition partner for conflict prevention — alongside civil society, women's groups, and faith communities. The most successful countries, it found, mobilize all of these actors to provide incentives for peace and manage social tensions.
Development policy must be conflict-sensitive. The report argued that macroeconomic policies, investment climates, and institutional reforms must be adapted to address inequalities and unmet aspirations rather than simply pursuing growth. Development programs that ignore distributional concerns can themselves fuel conflict.

Prevention is cost-effective for the private sector too. The report made a strong "business case" for prevention, estimating that conflict costs the global economy up to $13.6 trillion per year, and that prevention could save between $5 billion and $70 billion annually — an argument partly intended to engage the private sector by demonstrating the economic returns to investing in stable societies.

---

3. FCV Strategy 2020–2025

The FCV Strategy represented the World Bank Group's first dedicated, institution-wide strategy on fragility, conflict, and violence. On jobs and the private sector, it marked a significant operational step forward — moving from analytical framing to institutional commitment and concrete targets.
Jobs explicitly listed as a high priority. The strategy placed special emphasis on six high-priority issues in FCV settings, with creating jobs and economic opportunities listed as the third priority , alongside investing in human capital, macroeconomic stability, community resilience, justice, and the security sector. This was a notable elevation — jobs were no longer just a by-product of peacebuilding but a named strategic objective.
Private sector placed at the center of the sustainable development model. The strategy recognized that the private sector lies at the center of a sustainable development model in FCV settings, and that IFC and MIGA are significantly scaling up their efforts. This was a sharper and more unambiguous statement than anything in the 2011 WDR or Pathways for Peace, formally embedding the private sector as central to — not merely supportive of — development in fragile states.
Catalyzing private investment as a development approach. The strategy stated that FCV-affected countries need a "development approach that catalyzes private sector development to complement public efforts," and committed to engaging a wide array of clients including local MSMEs and regional and multinational firms. This included leveraging financing to incentivize investments and influence critical policy reforms that address the root causes of fragility.
Helping countries transition out of fragility through local private sector growth. One of the strategy's four pillars was helping countries transition out of fragility by promoting approaches that can renew the social contract between citizens and the state, foster a healthy local private sector, and strengthen the legitimacy and capacity of core institutions. This was significant because it linked private sector development explicitly to state legitimacy — something more implicit in the earlier reports.
IFC and MIGA set ambitious targets. IFC committed to delivering 40 percent of its overall long-term finance in IDA and FCS countries and 15–20 percent in low-income IDA and IDA FCS countries by 2030. MIGA similarly targeted a greater share of guarantees in IDA and FCS countries. These were not aspirational statements but quantified, corporate commitments — a major step from the earlier reports' more qualitative framing.
Challenges in implementation were significant. By the time of the strategy's mid-term review, important gaps had emerged. The Bank Group was found to lack an effective, holistic approach to private sector development, with efforts not achieving intended impact. IFC and MIGA faced barriers to scaling up in FCS due to poor business environments, high risks, a lack of bankable projects, and the need to adapt their own business models to FCS contexts and realities.
The Mid-Term Review acknowledged further work is needed. The review found that while the private sector is critical for building resilience and supporting transitions out of FCV, challenging investment conditions often limit the Bank's capacity to support private investment and growth, highlighting the importance of developing more integrated approaches. It recommended that the Bank advance its private sector engagement as one of five key priorities for adaptation.

3.1 Comparing All Three Across Key Themes

Theme	WDR 2011	Pathways for Peace 2018	FCV Strategy 2020–2025
Primary focus	Post-conflict recovery	Conflict prevention	Operational strategy across the FCV spectrum
Role of jobs	One of three pillars for stability	Core tool for addressing exclusion	Named high-priority objective
Role of private sector	Important but contingent on institutions first	Active partner in prevention coalitions	Central to sustainable development model; formal targets set
Instruments	Public works, access to finance, skills	Inclusive development, redistributive policy	IFC/MIGA investment, blended finance, MSME support, guarantees
Framing	Jobs follow security and governance	Jobs and inclusion *are* governance and security	Private sector development as a transition strategy
Key challenge identified	Unattractive investment climate	Distributional inequalities in economic opportunity	Lack of bankable projects; business model adaptation required

3.2 An Analytical Arc Across the Three Documents

Reading the three documents together traces a clear — if still incomplete — evolution in World Bank thinking. The 2011 WDR established that jobs matter for peace but was cautious about the private sector's immediate role. Pathways for Peace shifted the frame from recovery to prevention and elevated the private sector to a coalition partner, while making the economic case for engagement. The FCV Strategy 2020–2025 attempted to operationalize that ambition at an institutional scale, setting targets, mobilizing IFC and MIGA, and naming job creation as a strategic priority.
What is also clear from the mid-term review and independent evaluations, however, is that aspiration has outpaced delivery. The structural barriers to private investment in FCV contexts — risk, informality, poor infrastructure, limited rule of law, lack of bankable projects — remain deeply entrenched. The Bank has not yet solved what all three documents implicitly struggle with: how to make private sector engagement commercially viable and development-effective in the world's most difficult environments at the same time. The next FCV strategy, now in preparation, will need to grapple with that gap far more concretely than its predecessors.

4. IEG Independent Evaluation of the 2020–2025 FCV Strategy

The Independent Evaluation Group published its full evaluation of the 2020–25 FCV Strategy in November 2025 — making it the definitive external assessment of whether the strategy delivered on its ambitions. The headline verdict is one of qualified credit for ambition, but significant concern about implementation gaps, missing infrastructure, and unmet promises across nearly every dimension.

Overall Assessment: Ambitious but Structurally Flawed from the Start

The strategy was the first to cover the entire Bank Group and influenced FCV approaches in other multilateral and bilateral agencies. It was timely, ambitious, and comprehensive. However, the strategy lacks a theory of change that links improvements in its operational framework to achieving desired FCV outcomes at the global and country levels. Without that connective tissue, it was never clear *why* specific actions were expected to lead to specific outcomes — and this made the strategy difficult to implement, monitor, or learn from.

Compounding this, the strategy is not supported by an implementation plan detailing actions to be taken and responsibilities for achieving intended outcomes across the Bank Group — meaning that even where the direction was right, there was no roadmap for who should do what, by when.

On Private Sector and Jobs: The Goals Were Not Achieved

This is where the evaluation is most critical, and most directly relevant. The Bank Group lacks an effective, holistic approach to private sector development — a "One World Bank" priority — with efforts not achieving the intended impact so far. IFC and MIGA face barriers to scaling up engagement in FCS due to poor business environments, high risks, a lack of bankable projects, and the need to adapt their own business models to FCS contexts and realities.

While both IFC and MIGA have gradually deployed new approaches and instruments in FCS, they have not exhibited an upward trend sufficient to achieve the targets. The ambitious quantitative commitments the Bank Group made — including IFC's pledge to deliver 40 percent of its overall long-term finance in IDA and FCS countries — have not been met.

The root problem, IEG found, goes beyond financing. A shortage of bankable projects that meet IFC and MIGA standards and criteria, more so than the availability of finance, is the key constraint to scaling up business in FCS. Non-financial risks include those arising from weak governance, uncertainty, underdeveloped regulatory regimes, poorly functioning institutions, and market characteristics of most FCS countries. The blended finance instrument created specifically to unlock private investment — the IDA Private Sector Window — also fell short. An early IEG assessment shows that while the PSW allowed IFC and MIGA to support some projects in new markets and sectors, its usage was below expectations, as the financial risk mitigation offered by the PSW is only one of the factors deterring private investments in high-risk countries. The private sector in low-income and fragile countries needs more than credit.

IEG also flagged a performance deterioration: the performance of World Bank projects in FCS was stable between FY2015–19 and FY2020–25, while ratings of IFC's investments in FCS have declined significantly. This is a notable finding — at the very moment the Bank was committing to scaling up IFC in FCS, IFC's development performance in those settings was going backwards.

On Financing More Broadly: Growth Was Real but Misleading

The World Bank Group's financing support to FCS countries increased during the evaluation period. However, this increase was driven by the new classification of large borrowers such as Ethiopia, Nigeria, and Ukraine as FCS since FY20, as well as in part by one-time commitments related to the COVID-19 pandemic response. In other words, the headline numbers flattered performance — much of the apparent growth reflected classification changes and emergency spending rather than a genuine strategic deepening.

Meanwhile, the growth in FCS financing has not been met with proportionate increases in administrative budgets, which have declined in real terms. Resources dedicated to project

preparation, advisory services and analytics decreased, which could adversely affect the quality of the operations.

4.1 On Programming and Conflict-Sensitivity

Country Program Frameworks in FCS have increasingly integrated FCV concerns but need to strengthen links to operational programming through better use of diagnostics, conflict and fragility sensitive portfolios, and FCV indicators. The Bank's Risk and Resilience Assessments improved in quality, but IEG found persistent gaps in linking those diagnostics to actual operational decisions — meaning that even where the analysis was good, it was not driving programming choices.
Country-level evaluations identified five recurring problems: a lack of adaptation of operations to FCV circumstances; successes in using phased, adaptive approaches that gradually scale up support; calibration of lending volumes with absorptive capacity; high risks to sustainability in institution building and service delivery; and weak accountability and learning due to a lack of FCV-relevant results frameworks and indicators.

4.2 On Staffing: Commitments Were Not Kept

While the total number of staff deployed in FCS increased largely due to the redesignation of countries as FCS, there has not been a significant increase in the number of staff willing to deploy to FCS. Staff and management also strongly perceive that the FCV Strategy's commitments to enhance staff incentives and career development have not been met. The strategy promised a different talent model for difficult environments; in practice, few of those promises were honoured.

4.3 What IEG Said the New Strategy Must Fix

IEG's evaluation was explicitly designed to feed into the next FCV strategy. Its key lessons for the new strategy include:
On private sector and jobs specifically: Donor strategies should identify leading constraints to, and opportunities for, private initiative and support mechanisms tailored to the specific country context to overcome the barriers. This task requires both public and private interventions. It has been challenging for donors to scale up support for private sector investment in FCV, and effective models and approaches of engagement have yet to emerge.
Addressing the constraint of bankable projects warrants further shifts towards efforts to develop projects and changes in the institutions' business models. IFC cannot simply bring its standard model into FCS environments and expect it to work — the business model itself needs to change.
On theory of change and implementation planning: The next strategy must have a clear causal logic and a specific, monitorable implementation plan — the absence of both was the most fundamental structural failure of the 2020–25 strategy.
On country differentiation: FCV country strategies and programs must consider — and become more responsive to — the specific drivers of fragility and conflict in each country. These factors may include security risks and violence, elite capture of resources, weak institutional capacity, and social exclusion. Country teams could take a leading role in diagnosing the drivers of fragility, and this diagnostic could be better tailored to provide operational guidance.

4.4 Where Does This Leave the Overall Picture?

Dimension	What the Strategy Promised	What IEG Found
Private sector centrality	Central to sustainable development model	No holistic approach; intended impact not achieved
IFC/MIGA volume targets	40% long-term finance in IDA/FCS by 2030	Insufficient upward trend; targets not on track
IFC performance in FCS	Scaling up with impact	Ratings have *declined* significantly
Blended finance (PSW)	Unlock private investment	Usage below expectations; money alone not enough
Staffing incentives	Enhanced career development in FCS	Commitments not met; staff perceive the gap clearly
Financing growth	Genuine increase in FCS support	Growth largely driven by reclassification and COVID
Conflict sensitivity in operations	Portfolios shaped by FCV diagnostics	Diagnostics improving but not yet driving operations
Theory of change	Implicit throughout	Absent; a fundamental design flaw

The overall verdict of the IEG evaluation is that the 2020–25 FCV Strategy was the right response to the right problem, but it was never operationalized with the tools, plans, incentives, or institutional models that would have been needed to deliver. On the specific question of jobs and the private sector — the most ambitious part of the strategy's development model — the gaps are widest, and the distance between aspiration and achievement is greatest. The next FCV strategy will need to either lower its ambitions or dramatically rethink how the Bank Group engages business in the world's most difficult environments.

ANNEX B — S+ Benchmark

Methodological Note: Rationale for Adopting Satisfactory or Above (S+) as the Performance Benchmark

1. Overview and Purpose

This annex explains and justifies the decision to use Satisfactory or above (S+) — rather than the World Bank Group's current institutional benchmark of Moderately Satisfactory or above (MS+) — as the primary performance threshold in this analysis. The case rests on three pillars: the historical origins of the rating scale, the analytical and accountability properties of each threshold, and the documented empirical behavior of the MS rating category over the IEG RAP 2020–2024 period.

The choice is not cosmetic. The gap between MS+ and S+ is large — typically 30 to 45 percentage points across the portfolio. Projects rated Moderately Satisfactory (MS) represent a structurally ambiguous category that has, over time, accumulated a disproportionate share of the portfolio and diluted the discriminatory power of aggregate performance reporting. Using MS+ as an acceptable threshold understates performance problems, particularly in Fragile and Conflict-affected Situations (FCS) contexts where even marginal compliance with objectives represents a significant accountability gap.

2. Historical Origin of the Six-Point Scale

The World Bank's project evaluation system was built on a four-point rating scale — Highly Satisfactory (HS), Satisfactory (S), Unsatisfactory (U), and Highly Unsatisfactory (HU) — in which Satisfactory was the explicit pass/fail threshold. A project was acceptable if it achieved a Satisfactory rating; anything below was a failure.

In 1993, the scale was extended with new anchors at the extremes to allow finer gradation of outstanding and deeply problematic projects, while leaving the central pass/fail logic intact. The critical change came in 1994, when the scale was expanded again, this time *"to provide more nuanced distinctions near the center"* (IEG Codebook, 2015). Two new midpoint categories were inserted: Moderately Satisfactory (MS) and Moderately Unsatisfactory (MU). The stated purpose was to allow evaluators to distinguish between projects that clearly met objectives and those that partially did so — not to redefine adequacy.

The critical implication is that MS was never designed to be the institutional benchmark for acceptability. It was designed to give evaluators more precision in the middle of the distribution. Satisfactory remained the natural threshold for genuine project success. The subsequent institutional decision to adopt MS+ as the corporate scorecard target — driven by pressure to show broad portfolio improvement — effectively lowered the bar without changing the methodology.

Table B.1. Evolution of the World Bank Project Rating Scale

Period	Scale Type	Rating Points	Pass Threshold
Pre-1993	4-point (original)	HS, S, U, HU	S (Satisfactory)
1993	4-point (expanded extremes)	HS, S, U, HU retained; HS/HU added as new anchors	S (Satisfactory)

Period	Scale Type	Rating Points	Pass Threshold
1994–present	6-point (current)	HS, S, MS, MU, U, HU	MS+ adopted as institutional benchmark

Source: IEG World Bank Project Performance Ratings Codebook (2015); IEG RAP series. Authors' synthesis.

---

3. The Problem with Moderately Satisfactory as an Acceptable Threshold

3.1 Semantic and Conceptual Ambiguity

The Moderately Satisfactory rating is, by definition, not Satisfactory. The modifier 'moderately' indicates a departure from the full meaning of the category it qualifies. A project rated MS has *partially* or *mostly* achieved its objectives, but it has not satisfied them. To treat MS as an acceptable result conflates a partial with a full achievement — a conflation that would not be accepted in any other evaluation context.

In numerical conversion used by IEG itself (where 1 = HU, 2 = U, 3 = MU, 4 = MS, 5 = S, 6 = HS), MS is assigned the value of 4 — the first point above the midpoint on a scale where Satisfactory scores 5. This scoring confirms that MS represents a below-Satisfactory outcome on an ascending scale. Calling a score of 4 out of 6 'acceptable' when the scale was originally anchored at 5 is a methodological retreat.

3.2 Rating Clustering and Portfolio Distortion

IEG's own RAP analyses have noted, repeatedly, that project ratings *"increasingly cluster in the moderately satisfactory or satisfactory points of the scale"* (RAP 2020). This clustering is a structural artifact of the institutional incentive created by the MS+ corporate target. When teams know that MS is the minimum acceptable rating, they calibrate self-assessments and project designs to reach that floor — not to exceed it. The result is a distribution with a large, uninformative mass of projects rated MS, which IEG validates without significant downgrade.

This is not a neutral statistical phenomenon. It represents a systematic incentive for convergence toward marginal acceptability rather than genuine achievement. Using MS+ as the benchmark therefore rewards this behavior analytically, treating a project that partially achieved its objectives in the same category as one that fully achieved them.

3.3 The Empirical Gap Between MS+ and S+

The magnitude of the MS+/S+ gap reveals how much analytical work the MS tier is being asked to do. As the table below illustrates, consistently 30 to 45 percentage points of the portfolio sit in the MS tier alone — rated above MU but below S. These are projects the institution classifies as 'acceptable' but that, by the original standard, would not have been.

Table B.2. MS+ versus S+ Portfolio Gap, World Bank Lending Projects (FY2019–FY2023)

Fiscal Year	% Projects MS+	% Projects S+	Gap (MS+ minus S+)	Implication
FY2019	79%	34%	45 percentage points	Nearly half of 'acceptable' projects marginal

Fiscal Year	% Projects MS+	% Projects S+	Gap (MS+ minus S+)	Implication
FY2020	88% (→ revised 84%)	51%	~33–37 percentage points	Improvement concentrated in MS tier
FY2021	85%	58% (of MS+)	~36 percentage points	42% of MS+ projects only marginally acceptable
FY2022–23	83–84%	~50–54%	~30–34 percentage points	Plateau; gap remains structural

Source: IEG RAP 2021, 2022, 2023, 2024. Note: S+ figures for FY21 based on IEG statement that 58% of MS+ projects were rated S or HS. FY22–23 S+ figures estimated from available distribution data.

These are not small rounding differences. In FY2019, for every ten projects the Bank counted as 'performing,' only three to four would have passed the original S threshold. Even at the high point of FY2020 — when the Bank reported an 88% MS+ rate — fewer than half of all projects met the Satisfactory standard. The institutional headline performance figure is thus substantially more optimistic than the underlying evidence supports.

---

4. Particular Relevance for FCS Contexts

The case for S+ as the meaningful benchmark is strongest in FCS settings, where the analytical stakes are highest and the consequences of misdiagnosis most severe. FCS operations consistently underperform the broader portfolio under both MS+ and S+ measures. However, the performance gap is far more visible under S+, where FCS projects are estimated to fall 20–30 percentage points below the non-FCS performance level, compared with a 10–17 percentage point gap under MS+.

The FY2021 data are illustrative. FCS projects recorded a 74–77% MS+ rating, compared with 91% for non-FCS — a gap of roughly 15 percentage points that, at first glance, appears manageable. Under the S+ standard, however, FCS projects are estimated to achieve roughly 35–40% S+, against approximately 60% for non-FCS — a gap of 20–25 percentage points. The MS+ measure is therefore compressing and concealing the FCS performance gap by lumping partially-achieving FCS projects into the same 'acceptable' category as fully-achieving non-FCS projects.

Table B.3. MS+ versus S+ Performance Gap in FCS Contexts

Context	MS+ Rating (%)	S+ Rating (%)	Diagnostic Value
Overall portfolio (FY21)	85%	~50%	MS+ masks wide variance
Non-FCS projects (FY21)	91%	~60%	Solid performance confirmed by S+
FCS projects (FY21)	74–77%	~35–40% (estimated)	MS+ overstates FCS performance considerably
Country programs – FCS (FY13–23)	55%	~25–30% (estimated)	FCS gap only visible under S+ standard

Source: IEG RAP 2022, 2024. FCS S+ figures estimated from available distribution data; authors' calculations.

This matters operationally. The FCV Strategy 2020–2025 set ambitious objectives around improved development outcomes in FCS settings. Measuring progress against an MS+ benchmark allows modest, partial achievement to be counted as progress. Measuring against S+ makes clear that full objective achievement in FCS settings remains the exception rather than the rule, which is the honest assessment that strategy review and resource allocation decisions require.

---

6. Accountability and Diagnostic Properties of S+

An evaluation benchmark should maximize the diagnostic value of the resulting data — that is, it should sharply distinguish performing from underperforming operations, and it should be sensitive to genuine trends in portfolio quality. The MS+ benchmark fails on both counts.

First, MS+ is a low-resolution signal. Because the MS tier absorbs a very large share of the portfolio, shifts in MS+ rates tend to reflect movement in and out of MU and MS — i.e., movement among marginal projects — rather than genuine improvement in quality. The observed jump in MS+ from 79% in FY2019 to 88% in FY2020 was partly driven by restructuring and adaptive management that prevented some borderline MU projects from slipping below the MS floor. This is useful to know, but it should not be confused with a 9-percentage-point improvement in the quality of the portfolio.

Second, the corporate MS+ target creates a ceiling effect at the reporting level. Once the portfolio reaches 83–88% MS+ — as it has in recent years — further improvement on that metric is mathematically and practically limited, making MS+ an uninformative indicator for tracking performance over time. The S+ metric, by contrast, shows continued room for improvement and more meaningfully captures whether project teams are delivering results against stated objectives, not merely avoiding outright failure.

Third, IFC's own rating methodology for investment projects does not use a six-point scale — it uses a four-point scale (Highly Successful, Successful, Mostly Unsuccessful, Unsuccessful), in which Successful (S) is explicitly the pass/fail point. By using S+ as the World Bank lending benchmark, this analysis aligns the cross-institutional comparisons to a common standard of genuine achievement, rather than tolerating methodologically inconsistent thresholds across arms of the Bank Group.

---

7. Conclusion

The adoption of S+ as the primary benchmark in this analysis reflects a deliberate methodological choice grounded in: (i) the historical architecture of the World Bank rating system, in which Satisfactory was always the meaningful pass threshold; (ii) the recognized limitations of the MS category as a dumping ground for marginal projects that technically avoid failure without genuinely achieving their objectives; (iii) the large and well-documented empirical gap between MS+ and S+ rates across the portfolio; and (iv) the particular need for honest, high-resolution measurement in FCS contexts, where the consequences of misdiagnosis are most severe.

This analysis does not reject the MS+ metric as meaningless — it remains useful for tracking broad portfolio trends, particularly at the lower end of the distribution. But treating MS+ as the benchmark for 'acceptable performance' treats a project that partially achieved its objectives as equivalent to one that fully achieved them. It rewards marginal success and obscures the FCS performance gap that the FCV Strategy 2020–2025 was explicitly designed to close.

S+ is not a harder standard for the sake of being harder. It is the correct standard because it is the one the Bank's own original methodology intended — and because it is the only standard that honestly answers the fundamental question: did this project deliver what it promised?

References

IEG (Independent Evaluation Group). 2015. IEG World Bank Project Performance Ratings Codebook. Washington, DC: World Bank.

IEG. 2020. Results and Performance of the World Bank Group 2020 (RAP 2020). Washington, DC: World Bank.

IEG. 2021. Results and Performance of the World Bank Group 2021 (RAP 2021). Washington, DC: World Bank.

IEG. 2022. Results and Performance of the World Bank Group 2022 (RAP 2022). Washington, DC: World Bank.

IEG. 2023. Results and Performance of the World Bank Group 2023 (RAP 2023). Washington, DC: World Bank.

IEG. 2024. Results and Performance of the World Bank Group 2024 (RAP 2024). Washington, DC: World Bank.

IEG. 2025. World Bank Group FCV Strategy 2020–2025: An Independent Evaluation. Washington, DC: World Bank.

IEG Glossary of World Bank Group Ratings. Available at: ieg.worldbankgroup.org.

ANNEX C — IEG RAP Data

IEG Results and Performance Reports 2020–2024

Key Trends, Performance Data, and Implications for FCV Work

1. Purpose and Scope

This note summarises key findings and performance data from the five annual Results and Performance of the World Bank Group (RAP) reports published by the Independent Evaluation Group (IEG) between 2020 and 2024. The RAP series is IEG's primary vehicle for tracking and communicating the development effectiveness of the World Bank, the International Finance Corporation (IFC), and the Multilateral Investment Guarantee Agency (MIGA) across their project and country programme portfolios.

Each RAP draws on IEG's independent validation of self-evaluations submitted by World Bank teams (Implementation Completion and Results Reports), IFC (Expanded Project Supervision Reports), and MIGA (Project Evaluation Reports). The period 2020–2024 covers the full life of the FCV Strategy 2020–2025, the COVID-19 pandemic and its aftermath, and the post-pandemic global shocks that created a persistently challenging operating environment.

The annex to this note presents the underlying data in tabular form. The note itself draws out the main trends, their implications for performance in fragile and conflict-affected situations (FCS), and cross-cutting themes.

2. World Bank Lending Performance: Overall Trends

The most prominent story in the World Bank's project outcome ratings over this five-year period is the dramatic spike in FY2020, followed by a plateau at historically high levels.

2.1 The FY2020 Spike

Under the S+ threshold, FY2019 project outcome ratings stood at 34 percent, rising to an estimated 51 percent in FY2020 — the largest single-year improvement in the period. RAP 2021 reported this as a 9 percentage-point jump in MS+ terms (79 to 88 percent); under S+, the improvement reflects a mix of genuine performance gains and restructuring activity. On the numerical rating scale (1–6), the average outcome rating rose from 4.1 to 4.4. Subsequent re-evaluation revised the MS+ figure to 84 percent (S+ equivalent: approximately 42 percent), highlighting the lag effect in ICR validation data.

2.2 Stabilisation at Elevated Levels

Following the FY2020 spike, outcome ratings stabilised. Under S+, FY2021 is estimated at approximately 49 percent (RAP 2022 reported 85 percent MS+) and FY2022 at approximately 52 percent (RAP 2023: 83 percent MS+) — maintaining the average outcome rating at 4.3 on the six-point scale, the highest sustained level since FY2012. RAP 2024, covering the period FY2020–2023, confirmed that ratings remained broadly stable across consecutive years, with no statistically significant deterioration.

2.3 COVID-19: Less Impact Than Expected

Despite fears that pandemic-related implementation disruptions would damage portfolio performance, RAP 2023 — the first RAP with a substantial cohort of COVID-19-affected projects — found no statistically significant correlation between COVID-19 exposure and project outcome ratings. Teams responded by restructuring projects more frequently (from an average of 1.9 restructurings per project pre-pandemic to 2.6 in the RAP 2023 cohort) and by making earlier course corrections to results frameworks. RAP 2024 corroborated this finding, attributing resilience to adaptive management and the Bank's use of repurposed existing projects.

2.4 Monitoring and Evaluation Quality

M&E quality ratings improved consistently across the period, rising from 51 percent rated High or Substantial in FY2019 (RAP 2020) to 57 percent in FY2020, 64 percent in FY2021, and 63 percent in FY2022. Improvement in M&E was identified as a partial driver of improved outcome ratings — not because projects were performing better in absolute terms, but because they were better documenting and measuring their results.

3. Performance in Fragile and Conflict-Affected Situations

FCS performance is the most critical dimension for assessing the World Bank's progress under the FCV Strategy. The picture is mixed: ratings in FCS improved in the earlier years of the period and held relatively stable, but they consistently lag the overall portfolio, and the growing share of FCS operations in the portfolio creates structural downward pressure.

FCS ratings improved but remain below average: Under S+, FCS project ratings are estimated at approximately 30 percent in FY2012–14, rising to approximately 35 percent in FY2017–19 (RAP 2020 reported MS+ figures of 69 and 77 percent respectively). The overall portfolio S+ rate was 34 percent in FY2019. RAP 2021 highlighted continued FCS expansion, with FCS countries' share of the closing portfolio growing from 15 percent to 23 percent between FY2019 and FY2020.

Bank performance rating in FCS trails portfolio average: Under S+, the Bank performance rating in FCS is estimated at approximately 44 percent in FY2021, compared to approximately 59 percent for the overall portfolio — a gap of roughly 15 percentage points (RAP 2022 reported 81 and 91 percent under MS+). The quality of supervision in FCS was rated lower, and the Bank performance rating in FCS declined marginally between FY2020 and FY2021 even as overall performance rose. S+ estimates derived from IEG rating distribution data; see Annex B.

Growing FCS share creates structural challenge: RAP 2024 reported that the proportion of closed operations with full or partial exposure to FCS increased from 31 percent in FY2020 to 37 percent in FY2023. Because FCS operations historically carry lower outcome ratings, this compositional shift creates a structural drag on the Bank's overall portfolio performance, even if individual FCS project quality is improving.

Simplified design and access-focused objectives improve FCS outcomes: RAP 2024 found that FCS operations with objectives focused on expanding access to services received significantly higher average outcome ratings. Institutional capacity challenges were more prevalent in FCS contexts, and their negative effect on ratings became more pronounced with each additional challenge reported.

4. IFC Investment Projects: Reversal, Recovery, and Renewed Decline

IFC's investment project performance tells a more volatile story than the World Bank's. Having experienced a multi-year decline that reached its nadir in CY2017 (41 percent of projects rated Mostly Successful or better), IFC saw a meaningful recovery through 2019–2021, followed by a renewed deterioration in the most recent cohort — with the decline particularly severe in FCS, Africa, and IDA/blend countries.

4.1 Recovery: RAP 2021 and RAP 2022

RAP 2021 reported that IFC's investment outcome ratings reversed a declining trend for the first time in a decade, with 56 percent of projects in the FCS evaluated cohort (CY2019–21) rated Mostly Successful or better — up steeply from just 22 percent in CY2015–17. Overall IFC development outcome ratings improved to their highest level in years. RAP 2022 confirmed this recovery was continuing, supported by improved IFC work quality ratings and stronger additionality performance.

4.2 Renewed Decline: RAP 2023 and RAP 2024

RAP 2023 marked the reversal of this recovery. For the CY2020–22 cohort of 170 investment projects, IFC's overall development outcome ratings stood at 60 percent Satisfactory or better — a decline from 64 percent in CY2019–21. More concerning were the disaggregated figures: only 11 percent of FCS investment projects were rated Mostly Successful or better, and Africa projects stood at 27 percent, IDA/blend at 36 percent.

IFC work quality ratings also declined — from 60 percent success in CY2019–21 to 55 percent in CY2020–22 — as did additionality success rates (59 percent down to 54 percent). In FCS markets specifically, IFC only realised its anticipated additionality in 33 percent of projects. The gap was widest for nonfinancial additionality — the provision of knowledge, standards, and technical assistance beyond financing — which IEG repeatedly identified as the most challenging dimension for IFC in difficult markets.

RAP 2024 confirmed that this declining trend in FCS/IDA contexts was continuing into the CY2021–23 cohort. IDA/blend countries overall showed only 46 percent of projects rated Mostly Successful or better (CY2021–23), with the FCS subgroup within IDA/blend identified as a key driver of that weakness. The share of IDA/blend projects in IFC's evaluated portfolio declined from 35 percent (CY2013–15) to 27 percent (CY2021–23), partly due to India's IDA graduation — masking what is actually a worsening performance trajectory in the remaining IDA/FCS countries.

4.3 IFC Advisory Services

IFC advisory services projects showed more consistent improvement across the period. RAP 2022 found that development effectiveness ratings for advisory projects in FCS had risen significantly, attributing improvements to IFC's response to a 2017 joint IEG-IFC study on work quality — particularly the shortening of advisory project durations and sizes, and the co-location of team leaders in the field. RAP 2023 reported overall advisory services work quality at 59 percent satisfactory, though preparation and design work quality ratings were satisfactory in fewer than half of projects.

6. MIGA Guarantee Projects: Sustained but Fragile Stability

MIGA's performance over the period showed sustained high ratings on a six-year rolling basis, reflecting a decade-long improvement trend. However, signs of fragility emerged in the most recent data, partly attributable to pandemic-related market challenges and persistent weaknesses in MIGA's M&E and self-evaluation systems.

RAP 2020 reported MIGA ratings at 69 percent Satisfactory or better (S+) in FY2013–18. RAP 2022 noted a further increase to 70 percent in FY2016–21. RAP 2023 reported that MIGA guarantee project development outcome ratings remained stable over six years but were slightly lower in the most recent three years. A significant M&E concern was identified: 69 percent of MIGA project outcomes were not tracked by MIGA's Development Effectiveness Indicator System, with 10 percent of project-level outcomes and 13 percent of foreign investment-level outcomes unverifiable. RAP 2024 flagged that 19 MIGA projects (45 percent of planned self-evaluations for FY2021–23) remained pending, creating uncertainty about whether the stable ratings trend would be maintained.

7. Country Programme Performance

RAP 2022 introduced systematic analysis of country programme performance, finding that Completion and Learning Reviews covering FY2019–21 rated World Bank Group performance as Good or higher in 17 of 21 reviewed country programmes. IFC and MIGA contributed to 23 percent of country

programme objectives in CPFs reviewed. The FCS trend could not be tracked due to insufficient sample sizes.

RAP 2024 extended this analysis, confirming that while project-level ratings had stabilised, country-level results frameworks remained weak — 83 percent of Completion and Learning Review validations reported major inadequacies in results frameworks, undermining the ability to demonstrate contribution to higher-level development outcomes.

8. Cross-Cutting Themes Across the RAP Series

8.1 Work Quality Drives Performance Across All Institutions

Across all five RAPs and all three institutions, one finding was consistent: the quality of front-end work — project preparation, design, and appraisal — was the strongest predictor of development outcome ratings. For the World Bank, preparation time and quality at entry were significantly correlated with outcome ratings. For IFC, work quality ratings (especially at project preparation) were strongly and positively associated with development outcome ratings across every RAP. For MIGA, preparation and design weaknesses directly undermined outcome achievement.

8.2 M&E Quality Is Improving but Remains Uneven

M&E quality improved across the period for the World Bank. However, IFC and MIGA both showed persistent weaknesses in results measurement — particularly for market-level outcomes (IFC) and foreign investment-level outcomes (MIGA), which are harder to measure and attribute. RAP 2023 found that 8 percent of IFC intended outcomes could not be verified at all due to lack of evidence. The overall conclusion across five RAPs is that improvements in documentation and measurement contributed to improved ratings — but results measurement gaps continued to undermine accountability.

8.3 The FCS Performance Gap Has Not Closed

Despite the FCV Strategy's stated ambition to improve Bank Group performance in FCS, the RAP series shows no convergence between FCS and non-FCS performance. The gap in World Bank Bank performance ratings (10 percentage points in FY2021), IFC's extremely low FCS investment outcome ratings (11 percent Mostly Successful in CY2020–22), and the growing share of FCS projects in the portfolio all point to a structural and persistent challenge. The RAP series consistently identified institutional capacity constraints, high business risks, lack of bankable projects, and implementation disruptions from conflict as the key barriers — none of which have been resolved during the strategy period.

8.4 Adaptive Management Matters, Especially in Crisis Contexts

RAP 2023 and RAP 2024 both highlighted that teams that restructured projects earlier and more frequently during implementation achieved better outcome ratings — and that this was especially valuable during COVID-19 and in FCS contexts. RAP 2024 found that crisis-related operations in FCS performed better when they were focused, simple, and realistic — reinforcing the Mid-Term Review's finding that ambition must be calibrated to context.

9. Conclusions and Implications

The five RAPs from 2020 to 2024 tell a story of genuine improvement at the portfolio level — particularly for the World Bank — alongside persistent and in some cases worsening challenges in the contexts that matter most: FCS, IDA, and Africa. The headline ratings improvement is real but partly explained

by better measurement, a growing Human Development portfolio (consistently the highest-rated practice group), and pandemic-era adaptive management. It does not reflect a breakthrough in the most difficult environments.

For IFC, the period began with a promising reversal of its long decline, particularly in FCS advisory services, but ended with renewed deterioration in investment project outcomes — especially in FCS/IDA contexts — just as the Bank Group was scaling up its commitments to these markets under the FCV Strategy. The gap between IFC's strategic ambitions and its development effectiveness performance in fragile settings is the most urgent performance challenge visible in the RAP data.

The data in the annex below provides the underlying evidence base for these conclusions and should be read alongside the IEG's full evaluation of the FCV Strategy 2020–2025, published in November 2025.

IEG RAP 2020–2024: Compiled Performance Data

Table A1. World Bank Lending – Project Outcome Ratings (S+), FY2019–FY2023

Fiscal Year	% S+ (All Projects)	Avg. Rating (1–6)	% S+ (FCS)	% S+ (IDA)	Source RAP
FY2019	34%	4.1	~25–30% (est.)	~28–32% (est.)	RAP 2020
FY2020	51%→42%* (est.)	4.4	Improving (est.); FCS share ↑ 15%→23%	~38–42% (est.)	RAP 2021
FY2021	~49% (est.)	4.3	~35–40% (est.)	n/a	RAP 2022
FY2022	~52% (est.)	4.3	Stable; FCS share ↑	n/a	RAP 2023
FY2020–23	Stable ~4.3	4.3	FCS exposure: 31%→37% of portfolio	n/a	RAP 2024

**FY2020 S+ initially estimated at 51% (based on 88% MS+, RAP 2021); revised to ~42% based on 84% MS+ (RAP 2022). S+ figures derived from IEG distribution data; see Annex B.*

Table A2. World Bank – Bank Performance and M&E Quality Ratings

Fiscal Year	Bank Perf. S+ (All, est.)	Bank Perf. S+ (FCS, est.)	M&E High/Subst. (All)	M&E High/Subst. (FCS)	Source RAP
FY2018–19	~50–55% (est.)	n/a	51%	n/a	RAP 2020
FY2020	~55% (est.)	n/a	57%	52%	RAP 2021
FY2021	~59% (est.)	~44% (est.)	64%	57%	RAP 2022
FY2022	~56% (est.)	n/a	63%	60%	RAP 2023
FY2023	Stable	Stable; gap persists	Improving	Improving	RAP 2024

Table A3. IFC Investment Projects – Development Outcome Ratings

Cohort (CY)	Overall S+	FCS Projects	Africa Projects	IDA/Blend Projects	Work Quality	Source RAP
CY2015–17	~30–35% S+ (est.)	~14% S+ (est.)	n/a	n/a	Declining	RAP 2020
CY2017–19	~56% (uptick)	Recovering	n/a	n/a	Improving	RAP 2021
CY2019–21	64% S+	~35% S+ (est.)	n/a	n/a	60% S+	RAP 2022
CY2020–22	60% S+	~6–8% S+ (est.)	~18% S+ (est.)	~23% S+ (est.)	55% S+	RAP 2023
CY2021–23	Declining	Declining (key drag)	Long-term ↓	~29% S+ (est.)	Improving slightly in FCS	RAP 2024

Note: S+ = Satisfactory or better (IFC terminology). MS+ (Mostly Successful or better) figures converted to S+ equivalents using IEG distribution data; estimates marked (est.). IFC uses different rating terminology from the World Bank. FCS share of IFC active portfolio grew from 7% to 10% (2019–2022); IDA/blend share grew from 27% to 32%.

Table A4. IFC Investment Projects – Additionality Success Rates

Cohort (CY)	Overall Additionality S+	FCS Countries	Africa Countries	IDA/Blend Countries	Source RAP
CY2019–21	59%	n/a (improving)	n/a	n/a	RAP 2022
CY2020–22	54%	33% (anticipated → realised)	37%	47%	RAP 2023
CY2021–23	Declining in FCS/IDA contexts	Key driver of IDA/blend decline	Share ↑ >5pp in active portfolio	Weakening	RAP 2024

Note: Nonfinancial additionality gap in FCS: knowledge/innovation 17 percentage points below anticipated; ESG standards 22 percentage points below (RAP 2023).

Table A5. MIGA Guarantee Projects – Development Outcome Ratings (6-Year Rolling Average)

Period	% S+ (by number)	Project-Level Outcomes (full achievement)	Foreign Investment-Level Outcomes (full)	Source RAP
FY2013–18	69%	n/a	n/a	RAP 2020
FY2014–19	68%	n/a	n/a	RAP 2021
FY2015–20	66%	n/a	n/a	RAP 2022
FY2016–21	70% (highest)	55%	31%	RAP 2022
FY2017–22	Stable/slightly lower	Stable	More difficult; 44% unachieved	RAP 2023
FY2021–23	Uncertain (19 projects pending eval.)	n/a	n/a	RAP 2024

Note: 69% of MIGA project outcomes were not tracked by the Development Effectiveness Indicator System (RAP 2023). 45% of FY2021–23 planned self-evaluations remain pending (RAP 2024).

Table A6. RAP Series Summary: Annual Reports, Coverage, and Thematic Focus

RAP	Published	WB Data Coverage	IFC/MIGA Coverage	Key Theme	FCS Finding
2020	Nov 2020	FY2012–19	CY2015–19	Outcome levels framework; COVID-19 emerges	FCS S+: est. ~30%→35% (FY12–17); MS+ was 69%→77%
2021	2021	FY2012–20	CY2016–20	Steep jump in WB ratings; IFC reverses decline	FCS share ↑ 15%→23%; broad improvement
2022	2022	FY2013–22	CY2017–21	Country programme analysis added; IFC additionality intro'd	Bank perf. gap in FCS: 81% vs 91% (all countries)
2023	2023	FY2012–22	CY2018–22	First major COVID cohort; IFC outcome types; MIGA M&E gaps	IFC FCS: ~6–8% S+ (est.); additionality 33%

RAP	Published	WB Data Coverage	IFC/MIGA Coverage	Key Theme	FCS Finding
2024	2024	FY2013–23	CY2019–23	Better Bank evolution; institutional capacity; access objectives	FCS share 31%→37%; IDA/blend FCS key drag on IFC

Table A7. Key Performance Metrics at a Glance: Start vs End of Period

Metric	~Start of Period (2019–2020)	~End of Period (2022–2023)	Trend
World Bank Lending			
Outcome rating S+ (all projects)	34% (FY2019)	~52% (FY2022)	▲
Average outcome rating (1–6 scale)	4.1 (FY2019)	4.3 (FY2022)	▲
Bank performance S+ (FCS)	~50% (FY2020) (est.)	~44% (FY2021) (est.)	→
M&E quality High/Substantial	51% (FY2019)	63% (FY2022)	▲
FCS share of closing portfolio	15% (FY2019)	37% (FY2023)	▲▲
IFC Investment Projects			
Development outcome S+ (all)	~56% improving (CY2017–19)	60% (CY2020–22) → declining	▲ then ▼
Development outcome S+ (FCS)	~14% (CY2015–17) (est.)	11% (CY2020–22)	▲ then ▼▼
Work quality S+ (investment)	60% (CY2019–21)	55% (CY2020–22)	▼
Additionality success (all)	59% (CY2019–21)	54% (CY2020–22)	▼
Additionality success (FCS)	n/a	33% (CY2020–22)	Low
MIGA Guarantee Projects			
Development outcome S+ (6yr rolling)	69% (FY2013–18)	70% peak (FY2016–21) → slightly lower	▲ then →
Project-level outcomes (fully achieved)	n/a	55% (FY2016–21 cohort)	Moderate
Foreign investment-level outcomes (fully achieved)	n/a	31% (FY2016–21 cohort)	Low

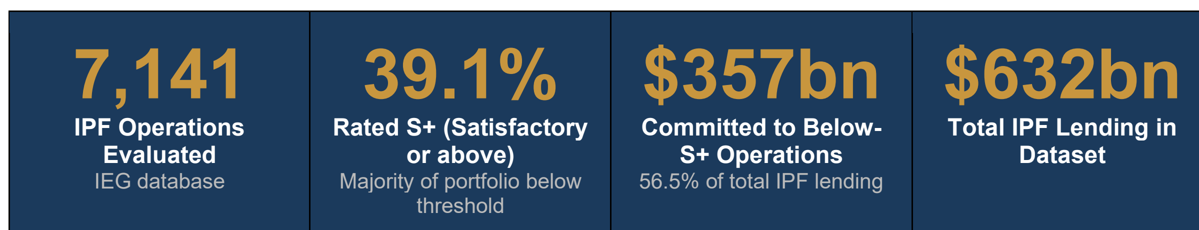
Key: ▲ = improvement; ▼ = deterioration; → = stable; ▲▲ = significant increase; ▼▼ = significant deterioration.

Sources: IEG Results and Performance of the World Bank Group 2020, 2021, 2022, 2023, and 2024 (Independent Evaluation Group, World Bank Group). All data extracted from published IEG reports and associated chapter-level analyses available at ieg.worldbankgroup.org.

ANNEX D — IPF

Investment Project Financing: Outcome Performance and Lessons Learned

Source: IEG evaluation database (n = 7,141 IPF operations). All figures derived from IEG outcome ratings, quality ratings, and lessons text. S+ denotes Satisfactory or Highly Satisfactory. Below S+ denotes Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory, or Highly Unsatisfactory.



1. Overview: What Is Investment Project Financing?

Investment Project Financing (IPF) is the World Bank's principal lending instrument, providing funding for specific, time-bound development projects with defined inputs, activities, outputs, and outcomes. Unlike Development Policy Financing (DPF), which provides general budget support against policy conditions, IPF finances discrete investments — infrastructure, social services, institutional strengthening, technical assistance — and disburses against expenditures incurred in project implementation.

IPF is the workhorse of the World Bank portfolio, accounting for 7,141 of the 8,767 evaluated operations in the IEG dataset — 81.5 per cent of all evaluated projects. Its applications span transport, energy, health, education, water, agriculture, governance, financial sector development, social protection, and environment. Operations range from targeted technical assistance credits of a few million dollars to major infrastructure programmes exceeding $1 billion.

IEG evaluates IPF operations on the same six-point outcome scale applied across the portfolio. S+ (Satisfactory or above) is the standard threshold for a positive outcome assessment. This annex analyses performance across the full IPF dataset, with particular attention to regional disaggregation, FCS performance, time trends, the relationship between preparation quality and outcomes, and the lessons recorded in IEG evaluation reports.

2. Outcome Performance: The Headline Numbers

2.1 Portfolio-Wide Performance

Across the 7,063 IPF operations with IEG outcome ratings, 39.1 per cent were rated S+. This means that 60.9 per cent of evaluated IPF operations — nearly two-thirds — did not reach the Satisfactory threshold. Of $631.9 billion in total IPF lending in the dataset, $357.1 billion — 56.5 per cent — was committed to operations subsequently rated below Satisfactory.

The outcome distribution is as follows:

Outcome Rating	Operations	Share of Rated Portfolio
Highly Satisfactory	297	**4.2%**
Satisfactory	2,462	**34.9%**
Moderately Satisfactory	2,492	**35.3%**
Moderately Unsatisfactory	991	**14.0%**
Unsatisfactory	698	**9.9%**
Highly Unsatisfactory	123	**1.7%**

The modal outcome rating is Moderately Satisfactory — a category that represents partial achievement of objectives and meaningful shortfalls in development outcomes. The combined weight of the three below-S+ categories (35.3% + 14.0% + 9.9% + 1.7% = 60.9%) indicates a portfolio in which underperformance is the norm, not the exception.

> **CORE FINDING**
> *60.9% of evaluated IPF operations — nearly two-thirds — fall below the Satisfactory threshold. The modal outcome rating is Moderately Satisfactory, indicating widespread partial achievement rather than outright failure. Of $632bn in total IPF lending, $357bn was committed to below-S+ operations.*

2.2 Regional Disaggregation

Regional performance varies substantially. East Asia and Pacific (EAP) and Europe and Central Asia (ECA) significantly outperform the portfolio average, while Sub-Saharan Africa and MENA record the weakest results. The gap between the best and worst performing regions exceeds 20 percentage points.

Region	Operations	S+ Rate	Total Committed	Below-S+ Committed
East Asia and Pacific (EAP)	1,191	**49.1%**	**$116.7bn**	**$49.0bn**
Europe and Central Asia (ECA)	1,205	**45.4%**	**$77.9bn**	**$38.7bn**
South Asia (SAR)	624	**41.3%**	**$102.9bn**	**$57.8bn**
Latin America and Caribbean (LAC)	1,277	**40.4%**	**$117.6bn**	**$63.9bn**
Western and Central Africa (WCA)	1,047	**31.2%**	**$55.1bn**	**$36.7bn**
MENA	744	**33.2%**	**$65.9bn**	**$38.8bn**
Eastern and Southern Africa (ESA)	997	**30.2%**	**$93.2bn**	**$69.8bn**

Sub-Saharan Africa (combining ESA and WCA) records a 30.7 per cent S+ rate across 2,044 operations — the largest regional grouping in the dataset. Of $148.2 billion committed to African IPF operations, $106.5 billion (71.8 per cent) was in operations subsequently rated below Satisfactory. EAP achieves S+ on 49.1 per cent of operations — more than 18 percentage points above the African rate.

ESA's performance is particularly concerning given the concentration of IDA lending and FCV Envelope resources in that sub-region. At 30.2 per cent S+, ESA records the lowest regional rate in the dataset — below even WCA (31.2 per cent). Of $93.2 billion committed to ESA IPF operations, $69.8 billion (74.9 per cent) was in below-S+ operations.

AFRICA IPF PERFORMANCE GAP
Africa (ESA + WCA combined): 2,044 operations, 30.7% S+, $148.2bn committed. Of that, $106.5bn — 71.8% — was in below-S+ operations. The region receiving the largest share of IDA IPF resources consistently produces the weakest outcomes.

2.3 FCS Performance

IPF performance in Fragile and Conflict-affected Situations (FCS) is materially worse than in non-FCS contexts, though the gap is less extreme than for DPF. This reflects the somewhat greater adaptability of investment lending — which can target specific, bounded interventions — relative to the broader policy and institutional ambitions of budget support.

Classification	Operations	S+ Rate	Total Committed	Below-S+ Committed
FCS	813	**26.6%**	**$50.0bn**	**$34.6bn**
Non-FCS	3,562	**36.1%**	**$366.4bn**	**$215.8bn**

At 26.6 per cent S+, the FCS IPF outcome rate is nearly 10 percentage points below the non-FCS rate — and well below half. Of $50.0 billion committed to FCS IPF operations, $36.3 billion (72.5 per cent) was in below-S+ operations. The lesson record for FCS contexts (Section 3.8) reveals a consistent pattern: inadequate adaptation of project design to fragile environments, unrealistic timelines, institutional capacity overestimation, and insufficient supervision intensity.

2.4 Trend Over Time

The IPF outcome trajectory shows a striking decline followed by a partial recovery. Early-period performance (FY2000–2009) of approximately 47 per cent S+ deteriorated sharply in FY2010–2014 to 23.6 per cent — the worst performance in the dataset — before recovering to 41.9 per cent in FY2020–2024.

Evaluation Period	Operations	S+ Rate	Total Committed
FY2000–2004	1,490	**49.0%**	**$120.4bn**

Evaluation Period	Operations	S+ Rate	Total Committed
FY2005–2009	1,308	**44.7%**	**$101.1bn**
FY2010–2014	1,199	**23.6%**	**$90.0bn**
FY2015–2019	1,406	**29.4%**	**$124.0bn**
FY2020–2024	1,130	**41.9%**	**$140.2bn**

The FY2010–2014 collapse — from 44.7 per cent to 23.6 per cent S+ — is the most significant discontinuity in the IPF dataset. Its causes likely include: the geographic expansion of the portfolio into more fragile contexts; the surge in post-crisis lending following the 2008 global financial crisis; increased complexity of operations; and possible IEG rating standards adjustments. The recovery in FY2020–2024 to 41.9 per cent S+ is notable but does not restore early-period performance levels, and may partly reflect the shorter time available for post-project evaluation of more recent operations.

2.5 Performance by Practice Group and Global Practice

IPF performance varies by practice area. Infrastructure operations — transport, energy, urban — perform above the portfolio average (42.3% S+), as does Human Development (40.0%). Equitable Growth, Finance and Institutions (EFI), which includes governance, financial sector, and public resource management, records the lowest practice group rate at 32.7%.

Practice Group	Operations	S+ Rate	Committed
Infrastructure	2,445	**42.3%**	**$293.9bn**
Human Development	1,699	**40.0%**	**$160.7bn**
Sustainable Development	1,782	**36.6%**	**$119.3bn**
Equitable Growth, Finance & Institutions (EFI)	1,027	**32.7%**	**$56.0bn**

At the Global Practice level, the highest-volume practices reveal meaningful performance dispersion:

Global Practice	Operations	S+ Rate	Committed
Social Protection & Jobs	380	**46.2%**	**$46.9bn**
Transport	799	**45.6%**	**$121.0bn**
Education	725	**41.6%**	**$63.0bn**
Energy & Extractives	650	**41.5%**	**$101.3bn**
Environment & Blue Economy	583	**37.4%**	**$21.2bn**
Agriculture and Food	538	**36.5%**	**$35.7bn**
Finance, Competit. & Innov.	498	**34.5%**	**$35.9bn**
Health, Nutrition & Pop.	594	**34.1%**	**$50.9bn**

Global Practice	Operations	S+ Rate	Committed
Water	629	**36.3%**	**$58.7bn**
Governance	417	**30.0%**	**$16.8bn**
Urban, Resilience & Land	996	**40.1%**	**$71.6bn**

Governance operations record the lowest Global Practice S+ rate at 30.0 per cent across 417 operations. Health, Nutrition and Population (34.1%), Finance and Competitiveness (34.5%), and Water (36.3%) all underperform the portfolio average. Transport (45.6%), Social Protection (46.2%), and Energy (41.5%) outperform. The concentration of underperformance in governance and social sector lending reflects the same institutional-capacity and ownership dynamics identified in the lesson record.

3. The Quality at Entry Effect: The Single Most Important Finding

The most striking quantitative finding in the IPF dataset is the near-perfect relationship between Quality at Entry (QaE) ratings and development outcomes. IEG's QaE assessment — the quality of project design at the preparation and appraisal stage — is an almost deterministic predictor of whether a project will achieve S+ outcomes.

Quality at Entry Rating	Operations	S+ Outcome Rate
Highly Satisfactory	233	**93.5%**
Satisfactory	2,869	**68.6%**
Moderately Satisfactory	1,695	**25.9%**
Moderately Unsatisfactory	1,122	**4.5%**
Unsatisfactory	1,045	**8.0%**
Highly Unsatisfactory	71	**0.0%**

The gradient is extraordinary. Projects with Highly Satisfactory QaE achieve S+ outcomes 93.5 per cent of the time. Projects with Satisfactory QaE achieve S+ outcomes 68.6 per cent of the time. Projects with Moderately Satisfactory QaE — the rating for operations that are adequate but not excellent at preparation — achieve S+ outcomes only 25.9 per cent of the time. Projects with Moderately Unsatisfactory QaE achieve S+ outcomes 4.5 per cent of the time. Projects with Highly Unsatisfactory QaE achieve S+ outcomes 0.0 per cent of the time.

There is no comparable variable in the dataset with this predictive power. The implication is not merely that preparation quality correlates with outcomes — it is that poor preparation is close to a sufficient condition for a below-S+ outcome, and high-quality preparation is close to a sufficient condition for a satisfactory one.

THE SINGLE MOST IMPORTANT FINDING IN THE IPF DATASET

> Quality at Entry is nearly deterministic of development outcomes. A project rated Moderately Unsatisfactory at QaE has a 4.5% chance of achieving a Satisfactory or better outcome. A project rated Highly Satisfactory at QaE has a 93.5% chance. Improving preparation quality is the highest-return intervention available to improve IPF performance — and the data indicate it is currently failing in a large share of the portfolio.

Of the 7,035 IPF operations with both QaE and outcome ratings:

- 233 are rated Highly Satisfactory at QaE — just 3.3% of the portfolio. This is vanishingly rare.
- 2,869 are Satisfactory at QaE — 40.8%.
- 1,695 are Moderately Satisfactory — 24.1%.
- 2,238 are below Moderately Satisfactory (MU, U, or HU) — 31.8% of the portfolio enters implementation already assessed as inadequately prepared.

Nearly a third of IPF operations proceed to implementation despite a preparation quality assessment below Moderately Satisfactory. Given that below-MS QaE is associated with S+ outcome rates of 4–8 per cent, this represents a predictable source of portfolio underperformance — one that is visible at the approval stage and not systematically addressed.

4. Supervision Quality and M&E

4.1 Quality of Supervision

Quality of Supervision (QoS) ratings show a broadly better distribution than QaE, but significant scope for improvement remains. The modal rating is Satisfactory (3,647 operations, 51.6%), followed by Moderately Satisfactory (1,663, 23.5%). A combined 1,325 operations (18.8%) receive below-MS supervision ratings.

Supervision quality matters particularly in IPF because the instrument's multi-year implementation period creates opportunities for course correction that DPF does not offer. The lesson record consistently identifies inadequate supervision — insufficient mission frequency, poor continuity of task team leadership, failure to identify and escalate problems early — as a compounding factor in poor outcomes. 'Consistency and stability is needed in supervision teams, both on the Borrower's and the Bank's side' is among the most frequently recurring formulations across evaluations.

4.2 M&E Quality

M&E quality across IPF operations is poor. The distribution:

M&E Quality Rating	Operations	Share
High	120	**2.8%**
Substantial	1,680	**38.6%**
Modest	2,138	**49.1%**
Negligible	473	**10.9%**

The dominant M&E rating is Modest — 49.1 per cent of all rated IPF operations. Only 2.8 per cent achieve a High rating. Combined, 60.0 per cent of IPF operations receive M&E ratings of Modest or Negligible. The consequences of this are identified consistently in the lesson record: without robust baseline data and outcome indicators, projects cannot demonstrate development impact even where implementation is successful. They also cannot identify implementation problems in time to correct them.

6. Lessons Learned: Synthesis from IEG Evaluations

The lessons text appended to IEG evaluation reports for the 4,302 below-S+ IPF operations with substantive lesson records reveals a consistent set of recurring themes. Frequency analysis identifies the themes that appear most persistently across regions, sectors, and time periods. These are not novel insights — they have been documented for decades. Their persistence in the lesson record is itself a finding about the Bank's institutional capacity to act on evaluation evidence.

Lesson Theme	Frequency (% of below-S+ lessons)
Institutional / capacity weakness	**85.2%**
Programme design / complexity	**75.7%**
M&E / results framework	**55.0%**
Stakeholder / community engagement	**41.9%**
Technical quality / analytical base	**34.4%**
Ownership / government commitment	**31.1%**
Project management / coordination	**25.2%**
Procurement / financial management	**24.6%**
Safeguards / social / environment	**23.9%**
Supervision / Bank performance	**21.5%**
Crisis / external shock	**18.7%**
Sequencing / phasing	**16.7%**

6.1 Institutional and Implementation Capacity (85.2%)

Institutional capacity constraints are the dominant lesson theme — appearing in 85.2 per cent of all below-S+ lesson records. This is higher than the equivalent frequency for DPF (77.3%), reflecting the greater operational complexity of IPF: investment projects require governments not merely to adopt policies but to procure goods and services, manage construction, operate facilities, maintain financial management systems, and report results — all activities that depend critically on institutional capacity.

The lesson record identifies several specific capacity failures:

- Project Implementation Units (PIUs) and Project Management Units (PMUs) that are insufficiently staffed, inadequately resourced, or disconnected from the line ministries responsible for sustained service delivery after project closure.
- Technical ministries without the engineering, procurement, or financial management capacity to implement project components without continuous Bank hand-holding.
- Governments unfamiliar with Bank procurement rules and fiduciary requirements — 'when a new client is unfamiliar with Bank policies and procedures, additional resources should be allocated to supervision and technical assistance to familiarize the client with Bank procedures.'
- Institutional reforms embedded in project design that require more time and capacity to implement than the project's legal disbursement period allows.

The lesson drawn repeatedly is that project complexity must be calibrated to demonstrated implementation capacity, not to the theoretical scope of what reforms are needed. 'Projects in countries with weak institutional capacity should be less complex, and provide more extensive institutional support and technical assistance.'

> **RECURRING LESSON (85.2% FREQUENCY)**
> *Institutional capacity constraints are the dominant explanation for IPF underperformance. Operations are routinely designed with complexity and ambition that exceed the borrowing government's demonstrated capacity to implement. Adaptation of project design to capacity constraints — not simplification as a concession but as a design principle — is the most consistently identified corrective measure.*

6.2 Programme Design and Complexity (75.7%)

Design failures appear in 75.7 per cent of below-S+ lesson records. Unlike DPF — where design failure primarily means over-ambitious conditionality — IPF design failure takes multiple forms:

- Excessive scope: projects attempt to address too many components across too many sub-sectors, diffusing implementation effort and management attention. 'A well-focused project design than a complex one' is among the most persistent recommendations.
- Unrealistic timelines: project design assumes implementation periods that are insufficient for the activities planned, particularly for infrastructure projects in low-capacity environments where procurement, construction, and commissioning typically take longer than projected.
- Component mismatch: individual project components are not designed with the same rigour as the primary activity, and underperforming components drag aggregate outcomes below the S+ threshold.
- Hybrid project risks: complex hybrid operations — combining investment, institutional reform, and policy change — require management capacity and government bandwidth that few low-income borrowers possess simultaneously. 'Complex hybrid projects supporting both investment and ambitious sectoral reforms should only be approved when there is strong ownership and strong capacity to implement.'
- Reform components in infrastructure projects: adding governance or institutional reform components to infrastructure investments frequently results in neither being done well.

6.3 M&E and Results Framework Design (55.0%)

M&E deficiencies appear in 55.0 per cent of below-S+ lesson records, consistent with the rating data showing the majority of IPF operations receive Modest or Negligible M&E ratings. The specific failures identified include:

- Results frameworks that measure outputs (facilities constructed, trainings conducted, beneficiaries reached) rather than outcomes (utilisation rates, service quality improvements, behavioural change).
- Baselines that are not established at project preparation, making it impossible to assess change attributable to the project.
- Outcome indicators that are 'inconsistent between objectives, results framework and financial tables' — internal incoherence in project design documents that is carried through to implementation.
- M&E systems that are not 'realistic and implementable' in the project context — designed for high-capacity environments and imposed on low-capacity ones.
- Absence of any functioning data system in FCS contexts, making project-level M&E impossible without prior investment in data infrastructure.

6.4 Stakeholder and Community Engagement (41.9%)

Stakeholder and beneficiary engagement appears with substantially greater frequency in IPF lessons (41.9%) than in DPF lessons — reflecting the direct interface between investment projects and communities, households, and local institutions. The lesson record identifies:

- Projects designed without adequate consultation with intended beneficiaries, resulting in low utilisation of completed facilities or services that do not match local needs.
- Community participation mechanisms that are formulaic rather than substantive — 'participation' recorded in project documents without genuine influence on design decisions.
- Resettlement and compensation processes that generate conflict with affected communities, delaying implementation and creating legacy disputes that persist after project closure.
- Failure to engage with local political economy — the interests, incentives, and resistance of local stakeholders — in project design and implementation planning.

6.5 Technical Quality and Analytical Underpinning (34.4%)

Technical and analytical quality concerns appear in 34.4 per cent of below-S+ lesson records. For IPF, this includes:

- Engineering assessments at appraisal that underestimate construction costs, site conditions, or technical complexity — producing cost overruns and completion delays.
- Market and demand analyses that are overly optimistic — projects designed around assumed beneficiary demand that does not materialise at the intensity projected.
- Institutional assessments that do not adequately evaluate the borrower's technical capacity to implement specific project components.
- Sector diagnostics that are outdated or insufficiently specific to the sub-national context of implementation.

6.6 Government Ownership (31.1%)

Government ownership appears in 31.1 per cent of IPF lesson records — somewhat lower than for DPF (37.8%), because the bounded nature of investment projects creates less dependence on

sustained political commitment than broad policy reform programmes. Nevertheless, ownership failures are significant in IPF:

- Counterpart funding commitments that are not honoured, stalling implementation when government budget allocations are diverted.
- Projects initiated under one political administration that lose priority under a successor — particularly significant for multi-year institutional reform operations.
- Line ministries that are not genuine owners of project design, creating implementation resistance that task teams struggle to overcome.
- 'To ensure Government's ownership and commitment the Bank should have worked closely with Government officials in the design of this project' — a lesson that recurs across decades and suggests that formal ownership processes do not reliably produce substantive buy-in.

6.7 Procurement and Financial Management (24.6%)

Procurement and financial management failures appear in 24.6 per cent of below-S+ lesson records — a lower frequency than might be expected given the operational centrality of these functions to IPF implementation. The lesson record identifies:

- Procurement delays as the most common implementation bottleneck in infrastructure projects — particularly in countries where procurement regulations differ significantly from Bank requirements.
- Financial management systems that are inadequate for the reporting requirements of Bank-financed projects, creating audit findings and disbursement holds.
- Counterpart funding arrears — governments failing to contribute their agreed share of project costs, creating financing gaps and delaying contractor payments.
- PIU staffing with procurement expertise that is not sustained through project implementation, creating institutional fragility when key staff depart.

6.8 FCS-Specific Lessons

The FCS lesson record within IPF contains consistent themes that amplify the portfolio-wide lessons in several specific respects:

- Post-conflict and fragile settings require simpler, more targeted project designs with longer implementation periods. 'For countries emerging from post-conflict situations, it is necessary to start with simple rehabilitation projects. Reform components should be modest in design and allow for longer implementation periods.'
- Project restructuring in FCS contexts should be proactively pursued at the first sign of implementation difficulty, not treated as a last resort. The lesson record suggests restructuring is initiated too late in a significant number of cases.
- Political risk assessment at appraisal is consistently inadequate in FCS operations — 'the SAR is deficient by not having raised the issue of political risk.'
- In FCS, institutional intermediaries outside government (NGOs, community organisations) frequently outperform government agencies as implementing entities and should be considered as alternatives to government PIUs where capacity is severely constrained.
- Multi-year approvals and phased operations are preferable to single, large operations in volatile environments: 'for projects in countries with fragile political conditions and instability, a new project is preferable over additional financing for the scale-up of activities.'

- Supervision intensity must be higher in FCS — more frequent missions, greater flexibility to adapt, and more proactive escalation of implementation problems — but is frequently not resourced accordingly.

> **FCS IPF PATTERN**
> FCS operations record 26.6% S+. The lesson record shows this is not primarily caused by conflict disruption — it reflects systematic design failures: over-complex projects, inadequate capacity assessment, unrealistic timelines, and insufficient supervision intensity. These are correctable through design and resourcing decisions made before approval.

7. Why Are Outcome Ratings So Poor? A Synthesis

7.1 Poor Preparation Is the Proximate Cause

The QaE data make the proximate cause of IPF underperformance clear: projects that are inadequately prepared at the design stage fail. Projects with Moderately Unsatisfactory or lower QaE ratings achieve S+ outcomes less than 8 per cent of the time. With 2,238 operations (31.8% of the portfolio) entering implementation with below-Moderately Satisfactory QaE ratings, the scale of foreseeable underperformance is substantial.

The primary levers are at the preparation stage: more rigorous institutional assessment, more realistic scope and timeline, better calibrated results frameworks, and genuine (not formulaic) engagement with government ownership and stakeholder needs. The data suggest these disciplines are not being consistently applied across a significant share of the portfolio.

7.2 Complexity Is Systematically Overestimated

The design and complexity lesson (75.7% frequency) points to a structural bias toward over-ambitious project design. The incentive structure that produces this bias is identifiable: task teams need to justify lending volumes; country programs need to demonstrate comprehensive development engagement; stakeholder expectations — both in the borrowing country and at headquarters — favour broad scope. The result is persistent project complexity that exceeds the implementation capacity of the contexts into which projects are deployed.

7.3 The Africa Gap Is Structural, Not Incidental

The 30.7 per cent S+ rate for African IPF operations is not explained by a few outlier failures. It reflects the systematic interaction of the portfolio's three major performance determinants in a context where all three are most challenging: institutional capacity is weakest (85.2% of lessons), preparation quality is lower (the QaE distribution for African projects skews below the portfolio average), and the proportion of FCS-classified operations is highest. The instrument has not been adapted to these structural characteristics; it has been applied at scale without adequate accommodation of them.

7.4 M&E Weakness Makes the Outcome Problem Invisible at the Time

With 60 per cent of IPF operations rated Modest or Negligible on M&E, a large share of the portfolio cannot demonstrate — or detect — its own performance problems during implementation. This creates a structural lag: problems that should be visible during supervision become apparent only at project closure evaluation. By that point, disbursement is complete and restructuring options are

exhausted. The M&E problem is therefore not merely a reporting failure — it is an implementation management failure with direct consequences for outcome quality.

7.5 The FY2010–2014 Collapse Has Not Been Fully Explained

The collapse of the S+ rate from approximately 47 per cent (FY2000–2009) to 23.6 per cent (FY2010–2014) is the most significant unexplained feature of the IPF data. Possible factors — portfolio expansion into harder contexts, crisis lending surge, IEG methodology changes — are not mutually exclusive. The partial recovery to 41.9 per cent in FY2020–2024 is encouraging but incomplete, and may partly reflect the shorter post-project time available for the most recent evaluations.

8. Implications for the IDA21 Period

IDA21's $100 billion replenishment, with its specific commitments on FCS, climate, human capital, and governance, will be substantially delivered through IPF. The analysis in this annex identifies several implications:

- The FCV Envelope architecture concentrates resources in the contexts where IPF performs worst — FCS countries, predominantly in Africa. Without deliberate adaptation of preparation quality standards, project complexity ceilings, and supervision intensity for FCV Envelope operations, the historical outcome rates will be reproduced at scale.
- IDA21 human capital targets — in health, education, and social protection — are primarily delivered through IPF. Human Development IPF (40.0% S+) performs near the portfolio average, but the lesson record identifies M&E weakness and government ownership as specific challenges in social sector lending that are not yet adequately addressed.
- The QaE finding is the single most actionable implication of this analysis. Raising preparation quality from Moderately Satisfactory to Satisfactory is associated with a near-tripling of S+ outcome rates (25.9% → 68.6%). This is a design, staffing, and incentive issue — not a resource availability one — and it is within the Bank's control.
- Governance operations (30.0% S+, 417 operations) represent the weakest Global Practice performance in the IPF data. Governance is central to IDA21's cross-cutting priorities. The combination of weak governance IPF outcomes and the instrument's central role in IDA21 delivery creates a specific accountability risk that warrants targeted attention.
- The M&E gap means that a significant share of IDA21 operations will be unable to demonstrate results against the replenishment's results matrix commitments — not because results were not achieved, but because the measurement infrastructure was not built into project design. Mandatory minimum M&E standards, enforced at quality enhancement review, represent the most direct corrective available.

> **SYNTHESIS**
> *The IPF outcome problem is large in scale ($357bn in below-S+ lending), consistent in geography (Africa, FCS), identifiable at the preparation stage (QaE effect), and well-documented in the lesson record (85% of lessons cite institutional capacity, 76% cite design complexity). The data do not indicate an instrument that is fundamentally unsuited to its contexts. They indicate an instrument that is systematically under-prepared, over-designed, and under-supervised relative to what its deployment contexts require.*

Data source: IEG evaluation database. All calculations based on IEG outcome, quality, and M&E ratings. Lessons synthesis derived from qualitative analysis of lesson text appended to IEG evaluations of below-S+ IPF operations (n = 4,302 with substantive lesson records). Committed amounts in current USD.

ANNEX E — DPF

Development Policy Operations: Outcome Performance and Lessons Learned

Source: IEG evaluation database (n = 1,551 DPF operations). All figures derived from IEG outcome ratings, quality ratings, and lessons text. S+ denotes Satisfactory or Highly Satisfactory. Below S+ denotes Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory, or Highly Unsatisfactory.

DPF Operations Evaluated IEG database	Rated S+ (Satisfactory or above) vs ~45% portfolio average	Committed to Below-S+ DPF Operations 54.4% of total DPF lending	Total DPF Lending in Dataset

1. Overview: What Are Development Policy Operations?

Development Policy Financing (DPF) — formerly known as Development Policy Loans (DPLs) and Structural Adjustment Credits (SACs) — provides general budget support to governments conditional on the adoption of agreed policy and institutional reforms. Unlike investment lending, which finances specific inputs (infrastructure, equipment, services), DPF finances government expenditure broadly, disbursing rapidly against a programme of prior actions. The instrument accounts for a substantial share of World Bank and IDA lending, particularly to countries experiencing balance-of-payments pressures or seeking support for structural reform programmes.

DPF operations are evaluated by the Independent Evaluation Group (IEG) on the same six-point outcome scale applied across the portfolio: Highly Satisfactory, Satisfactory, Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory, and Highly Unsatisfactory. S+ (Satisfactory or above) is the standard threshold for a positive outcome assessment.

This annex analyses outcome performance across the full IEG DPF dataset (n = 1,551), disaggregated by region, FCS status, practice group, and time period, and synthesises the lessons recorded in IEG evaluation reports.

2. Outcome Performance: The Headline Numbers

2.1 Portfolio-Wide Performance

Across the 1,551 DPF operations with IEG outcome ratings in the dataset, only 35.6 per cent were rated S+. This compares unfavourably with the broader World Bank portfolio average and represents a persistent structural underperformance of the instrument. The corollary is that 64.4 per cent of evaluated DPF operations — the majority — failed to reach the S+ threshold.

The financial scale of this underperformance is significant. Of $342.9 billion in total DPF lending captured in the dataset, $186.4 billion — 54.4 per cent — was committed to operations subsequently rated below Satisfactory.

CORE FINDING
The majority of evaluated DPF operations — 64.4% — are rated below Satisfactory. This is not a marginal performance shortfall. It represents a systemic outcome problem with the instrument, concentrated in Africa and FCS contexts and worsening over time.

The outcome distribution is as follows:

Outcome Rating	Number of Operations	Share of Rated Portfolio
Highly Satisfactory	39	**2.5%**
Satisfactory	512	**33.1%**
Moderately Satisfactory	586	**37.9%**
Moderately Unsatisfactory	278	**18.0%**
Unsatisfactory	119	**7.7%**
Highly Unsatisfactory	12	**0.8%**

2.2 Regional Disaggregation

The regional breakdown reveals that the DPF outcome problem is heavily concentrated in Sub-Saharan Africa and MENA, while Europe and Central Asia (ECA) and Latin America and the Caribbean (LAC) substantially outperform the portfolio average. This pattern has persisted across evaluation periods.

Region	Operations	S+ Rate	Total Committed
Europe and Central Asia (ECA)	304	**53.0%**	**$74.8bn**
Latin America and Caribbean (LAC)	337	**47.3%**	**$113.8bn**
South Asia (SAR)	66	**47.0%**	**$12.5bn**
East Asia and Pacific (EAP)	179	**27.4%**	**$87.5bn**
MENA	161	**29.2%**	**$41.9bn**
Africa (ESA + WCA combined)	502	**21.1%**	**$50.3bn**

Africa accounts for 502 DPF operations in the dataset — the largest single regional grouping — yet records the lowest S+ rate at 21.1 per cent. Of $50.3 billion committed to African DPF operations, $38.6 billion (76.8 per cent) was committed to operations rated below Satisfactory. ECA achieves S+ on 53.0 per cent of operations, more than two and a half times the African rate.

The Western and Central Africa (WCA) sub-region records a 19.3 per cent S+ rate across 285 operations, closely matched by Eastern and Southern Africa (ESA) at 23.5 per cent across 217 operations. Both are among the lowest-performing sub-regions in the entire portfolio.

AFRICA DPF PERFORMANCE GAP
Africa accounts for the largest share of DPF operations in the dataset yet records the lowest S+ rate (21.1%). Of $50.3bn committed, $38.6bn — 76.8% — was in operations rated below Satisfactory. The instrument is being most heavily deployed in the region where it most consistently fails.

2.3 FCS Performance

DPF performance in Fragile and Conflict-affected Situations is significantly worse than in non-FCS contexts. Across 171 FCS-classified DPF operations in the dataset, the S+ rate was 19.3 per cent — less than half the already-low portfolio average.

Country Classification	Operations	S+ Rate	Total Committed
FCS	171	**19.3%**	**$15.6bn**
Non-FCS	921	**34.5%**	**$222.0bn**

The FCS context creates compounding challenges for DPF: weak institutional capacity limits the government's ability to implement prior actions meaningfully; macroeconomic instability disrupts programme assumptions; political instability undermines ownership continuity; and the rapid-disbursement logic of DPF creates incentives to proceed even where reform conditions are unlikely to be sustained. The lesson record (Section 3) confirms all four dynamics.

2.4 Trend Over Time: A Deteriorating Trajectory

The S+ rate for DPF operations has deteriorated significantly since the early 2000s. The early-period performance (FY2000–2009) of approximately 48 per cent S+ has not been recovered in any subsequent period:

Evaluation Period	Operations	S+ Rate	Total Committed
FY2000–2004	257	**47.5%**	**$70.1bn**
FY2005–2009	254	**48.6%**	**$43.9bn**
FY2010–2014	330	**29.5%**	**$66.9bn**
FY2015–2019	324	**25.1%**	**$67.5bn**
FY2020–2024	282	**30.9%**	**$66.4bn**

The FY2010–2014 period marks a sharp discontinuity: the S+ rate fell from approximately 48 per cent to 29.5 per cent, a decline of nearly 20 percentage points. It has not recovered. The FY2015–2019 period recorded the worst performance in the dataset at 25.1 per cent S+. This deterioration coincides with the geographic expansion of DPF into more fragile and lower-capacity contexts, increased use of the instrument for crisis response, and the continued application of ambitious conditionality in environments where implementation capacity is demonstrably weak.

2.5 Performance by Practice Group

DPF performance varies substantially by practice area. The Equitable Growth, Finance and Institutions (EFI) practice group — which accounts for the large majority of DPF operations (1,194 of 1,551) — records a 32.0 per cent S+ rate. Human Development DPF operations substantially outperform at 56.9 per cent S+, while Infrastructure (44.9 per cent) and Sustainable Development (40.7 per cent) sit between these extremes.

Practice Group	Operations	S+ Rate	Committed
Human Development	131	**56.9%**	**$34.1bn**
Infrastructure	136	**44.9%**	**$37.3bn**
Sustainable Development	81	**40.7%**	**$16.3bn**
Equitable Growth, Finance & Institutions (EFI)	1,194	**32.0%**	**$254.6bn**

The concentration of poor performance in EFI-managed operations reflects the nature of those operations: fiscal adjustment, financial sector reform, governance, private sector development, and public resource management — all of which depend heavily on political ownership, institutional capacity, and sustained implementation over time horizons that DPF instruments are not well-designed to support.

2.6 Bank Performance Ratings

IEG's quality ratings for Bank performance illuminate the supply-side contribution to poor outcomes. Across the DPF portfolio:

- Quality at Entry: 584 operations rated Satisfactory; 345 Moderately Satisfactory; 248 below Moderately Satisfactory (Moderately Unsatisfactory, Unsatisfactory, or Highly Unsatisfactory).
- Quality of Supervision: 702 operations rated Satisfactory; 258 Moderately Satisfactory; 165 below.
- M&E Quality: only 17 operations rated High; 303 Substantial; 361 Modest; 78 Negligible. The dominant M&E rating is Modest — barely adequate.

Weak Quality at Entry is particularly consequential for DPF because the instrument's rapid disbursement profile means that design flaws are rarely corrected before funds have been committed. Poor M&E quality means that programmes frequently cannot demonstrate results even where reforms are implemented, and the lesson record is correspondingly impoverished.

3. Lessons Learned: Synthesis from IEG Evaluations

The lessons text appended to IEG evaluation reports for the 995 below-S+ DPF operations with substantive lesson records reveals a consistent set of recurring themes. These are not marginal observations: they appear with high frequency across decades, regions, and instrument types, and they have not been resolved. What follows is a synthesis of the dominant lesson themes, ordered by frequency of appearance in the data.

3.1 Institutional and Implementation Capacity (77.3% of below-S+ lessons)

The most frequently cited lesson — appearing in more than three-quarters of all below-S+ DPF evaluation reports — is that the borrowing government's institutional and implementation capacity was insufficient for the reform programme being supported. This takes several forms in the lesson text:

- Governments lack the technical capacity within line ministries to draft, pass, or operationalise required legislation.
- Prior actions are met on paper without being translated into substantive implementation — 'action plans alone should be avoided as main loan conditions since they do not guarantee that the authorities will carry out the reforms.'
- Operations designed with limited absorptive capacity assumptions are then executed in environments where capacity is even weaker than assessed.
- Parallel technical assistance is frequently identified as essential but absent: 'good ESW and TA are absolutely essential for reforms to succeed in a country with capacity constraints.'

The lesson drawn repeatedly is that DPF should not be approved where institutional capacity is insufficient for meaningful implementation, and that where it is approved, parallel TA operations are not optional — they are preconditions for any reasonable prospect of success.

> **RECURRING LESSON (77.3% FREQUENCY)**
> *Institutional capacity constraints are the dominant explanation for DPF underperformance. The instrument is routinely deployed in environments where the government cannot implement the reforms it has nominally agreed to. Parallel technical assistance is consistently identified as essential but inconsistently provided.*

3.2 Programme Design and Scope (75.9% of below-S+ lessons)

The second most frequently cited theme concerns the design of the DPF operation itself. Three specific design failures recur across the lesson record:

- Over-ambition: operations attempt to reform too many sectors simultaneously, diffusing both Bank and government effort and exceeding borrower management capacity. 'Reform programs need to be well-focused so as not to diffuse Bank and counterpart efforts too broadly.'
- Unrealistic objectives: programmes are designed with objectives that are not achievable within the operation's timeframe given existing capacity constraints.
- Insufficient focus: operations with many objectives in different sectors produce shallow engagement across all of them rather than deep reform in any. 'A more narrowly-targeted operation is more likely to be successful' is one of the most commonly cited lessons in the dataset.

These design failures are not a recent phenomenon — they appear with equal frequency in evaluations from the early 2000s as in more recent ones. The structural incentive to lend at volume, and to satisfy multiple stakeholder interests in programme design, persistently overrides the lesson that focus is the primary predictor of success.

3.3 Conditionality Design (48.1% of below-S+ lessons)

Conditionality design failures appear in nearly half of all below-S+ lesson records. The most important distinctions identified by IEG are:

- Process conditions vs. results conditions: conditionality linked to the preparation of plans, strategies, or draft legislation — rather than to actual implementation or measurable results — 'does not guarantee that the authorities will carry out the reforms required.' IEG consistently favours conditionality 'linked to concrete actions, actual implementation, and results in the main reform areas.'
- Tranche structure: multi-tranche operations with floating tranches can provide useful flexibility when reform timelines are uncertain; one-tranche operations based solely on prior actions 'do not allow monitoring of implementation.' The appropriate structure depends on the reform context, but this choice is frequently made without adequate consideration.
- Over-counting prior actions: reforms that are included as conditions to satisfy donor expectations rather than because they are critical to programme objectives produce compliance without substance.
- Legal and regulatory sequencing: operations requiring legal reforms frequently underestimate the time required for parliamentary processes or the issuance of implementing regulations. 'The conditionality should extend to the issuance of application decrees, and not only to the revision of laws.'

> **RECURRING LESSON (48.1% FREQUENCY)**
> *Prior actions that require the preparation of plans or draft legislation rather than concrete implementation systematically overstate reform progress. The Bank has known this for twenty-five years and continues to accept process conditionality as evidence of reform commitment.*

3.4 Monitoring and Evaluation Quality (59.2% of below-S+ lessons)

M&E deficiencies appear in nearly 60 per cent of below-S+ lesson records — a finding consistent with the rating data, which shows the dominant M&E quality rating for DPF is Modest. The consequences of weak M&E are identified as:

- Inability to demonstrate development outcomes even where reforms were implemented — results frameworks measure inputs and processes, not changes in economic or social conditions.
- Absence of baseline data making it impossible to assess whether change occurred.
- 'The lack of such a system for this operation seems to have been costly' — a formulation that appears, with variation, across decades of evaluations.
- In some FCS contexts, no functioning data system exists from which baseline data could be drawn: 'the efficacy of Bank and other donor interventions cannot be judged adequately in the absence of a minimum of basic data.'

The specific lesson drawn is that DPF results frameworks should be designed to measure outcomes — changes in economic behaviour, institutional performance, or service delivery — rather than the adoption of policies whose ultimate effect is unverified.

3.5 Ownership and Political Commitment (37.8% of below-S+ lessons)

Government ownership of the reform programme — as distinct from acceptance of donor conditions — appears as a decisive variable in 37.8 per cent of below-S+ lesson records. The distinction the lesson text consistently draws is between:

- Formal commitment: government signs off on the programme, accepts the prior actions, and receives disbursement.
- Substantive ownership: the relevant ministries, civil servants, and political leadership are genuinely invested in the reform agenda and will sustain it beyond the operation's disbursement period.

Where formal commitment exists without substantive ownership, reforms are implemented minimally, reversed after disbursement, or simply not operationalised beyond the signing of decrees. 'Conditionality cannot substitute for strong government commitment in a politically sensitive program' and 'the components of the loans that were owned by the country were generally successfully implemented' are among the most frequently recurring formulations.

In FCS contexts, ownership is additionally complicated by political instability, government turnover, and split authority between executive and legislative branches. Operations that do not adequately assess — and honestly reflect — the degree of genuine ownership are systematically set up to fail.

3.6 Macroeconomic Instability (32.0% of below-S+ lessons)

Approximately a third of below-S+ lesson records identify macroeconomic instability as a contributing factor in poor outcomes. DPF is particularly sensitive to the macroeconomic environment because:

- Structural reform programmes assume a degree of economic stability that enables long-term planning; inflation, exchange rate volatility, and fiscal crises disrupt the conditions under which reforms can be sustained.
- 'Financial sector loans with financial deepening objectives should be implemented in periods of macroeconomic stability' — a lesson drawn from multiple African countries and repeated across decades.
- Operations designed at appraisal with one macroeconomic assumption are then implemented under materially different conditions, rendering the programme design obsolete.

The structural tension here is that DPF is frequently used precisely in response to macroeconomic distress — as a crisis instrument. Using a budget support instrument requiring sustained reform implementation in conditions of economic instability creates a fundamental contradiction between the instrument's deployment rationale and its implementation requirements.

3.7 IMF Programme Alignment and Donor Coordination (38.4% of below-S+ lessons)

IMF programme alignment and donor coordination appear in 38.4 per cent of below-S+ lesson records. The specific concerns are:

- DPF operations that assume IMF programme support is in place — and that disbursement of IMF resources validates the macroeconomic framework — are exposed when Fund programmes lapse or are not renewed.
- 'The existence of a Fund program should not be a sufficient condition' for proceeding with DPF — a formulation that appears across multiple African evaluations where IMF programme compliance was formal rather than substantive.
- Donor coordination failures lead to contradictory conditionality, competing reform priorities, and government time spent managing donor relationships rather than implementing reforms.
- 'Partnership between Government, other Donors, the IMF and the Bank is crucial for the success of adjustment operations.'

3.8 Sequencing and Reform Overload (29.1% of below-S+ lessons)

Sequencing failures appear in 29.1 per cent of lesson records — the complement of the design-focused lessons in 3.2. Sequencing concerns include:

- Financial infrastructure reform preceding financial sector policy reform — 'financial sector institution building projects should precede financial sector policy loans.'
- Civil service reform and expenditure management systems being superimposed on existing inadequate systems rather than built sequentially.
- 'Attempting to introduce a new system of expenditure prioritization by superimposing it on an existing inadequate budgeting system is not likely to be successful.'
- Privatisation conditionality designed around numerical targets rather than process milestones — 'it is inadvisable to rely on numerical targets to measure success of privatization.'

4. Why Are Outcome Ratings So Poor? A Synthesis

The question of why DPF outcome ratings are so poor admits a layered answer. The data and lesson record together point to five structural explanations, each reinforcing the others.

4.1 The Instrument–Context Mismatch

DPF is a fast-disbursing budget support instrument premised on the existence of a credible reform programme and a government with the capacity and commitment to implement it. The data show that the instrument has been most heavily deployed in precisely the contexts where these preconditions are least likely to hold: Africa (502 operations, 21.1% S+), FCS countries (171 operations, 19.3% S+), and low-capacity states where institutional constraints are severe.

The result is predictable: a fast-disbursing instrument reaches its disbursement milestones — prior actions are met, tranches are released — but the underlying reforms are not implemented with the depth or sustainability that would produce positive development outcomes. The instrument measures compliance with conditions; IEG measures outcomes. The gap between these two measures is the outcome gap.

4.2 Structural Incentives Toward Volume

The lesson record contains explicit references to Bank internal incentives that prioritise lending volume over outcome realism. One Malawi evaluation states directly: 'incentives within the Bank can motivate unwise lending — the desire to transfer resources and establish HIPC eligibility were important motivations to the initiation of the operation, despite the poor macroeconomic climate and lack of reform progress under the previous credits.'

These incentives — to lend, to maintain country programmes, to meet replenishment commitments, to establish HIPC eligibility — persistently override the analytical finding that the reform context is insufficient for a successful DPF. The lesson has been identified. It has not been addressed.

4.3 Process Conditionality That Does Not Guarantee Reform

The shift from input and output conditionality toward results-based conditionality has been advocated by IEG for decades. The data suggest it has not been achieved. Operations that accept the preparation of action plans, passage of legislation, or issuance of decrees as evidence of reform — rather than demonstrated implementation — systematically overstate progress. Governments rationally focus on satisfying disbursement conditions; if conditions do not require actual reform implementation, actual reform implementation will frequently not occur.

4.4 Analytical Underpinning Gaps

Economic and Sector Work (ESW) is the Bank's primary diagnostic instrument for understanding the reform environment prior to DPF design. The lesson record identifies inadequate ESW as a recurring factor in poor outcomes — 'a well-conceived and carefully executed program of economic work is essential to underpin policy lending operations.' Where ESW is absent, outdated, or insufficiently specific to the sectors being reformed, DPF design is built on incomplete foundations. The macroeconomic assessments appended to DPF documents are criticised in several evaluations as 'unbalanced and lacking realism.'

4.5 The Declining S+ Trend Is Not Random

The trajectory from approximately 48 per cent S+ in FY2000–2009 to 25–31 per cent S+ in FY2010–2024 reflects a deliberate policy choice: expanding DPF into more fragile, lower-capacity, and macroeconomically stressed environments in response to the IDA FCS allocation architecture, crisis lending demands, and replenishment commitments. As the deployment context became harder, the instrument design did not adapt commensurately. The result is a portfolio where the most challenging contexts receive the most conditionality-heavy instrument with the weakest outcome track record.

> **SYSTEMIC FINDING**
> The DPF outcome problem is not a collection of individual project failures. It is a systemic misalignment between instrument design, deployment context, conditionality structure, and outcome expectations — reproduced across twenty-five years of evaluation evidence without structural correction.

6. Implications for the IDA21 Period

The IDA21 replenishment commits $100 billion to IDA-eligible countries, the majority of which are in Sub-Saharan Africa — the region with the worst DPF outcome track record in the dataset. Several implications follow directly from the analysis:

- The FCV Envelope and related IDA21 architecture will channel significant resources to FCS contexts through DPF. The 19.3% S+ rate for FCS DPF operations suggests that this disbursement will not, on current form, translate into satisfactory development outcomes at the programme level.
- IDA21 results commitments include significant governance and institutional reform targets. DPF is a primary vehicle for these commitments. The lesson record is unambiguous: these targets will not be achieved through process conditionality in low-capacity environments without sustained, adequately-resourced parallel technical assistance.
- The Scorecard's disaggregation of FCS results creates measurement visibility without creating the accountability mechanisms or instrument redesign required to address the underlying performance gap.
- Bank performance ratings — Quality at Entry, Quality of Supervision — show meaningful scope for improvement. Strengthening these ratings through more realistic programme design, genuine prior assessment of ownership, and robust M&E frameworks represents the most tractable near-term lever available to the Bank.

The case for DPF instrument reform — not abolition, but structural redesign — is supported by the full weight of the evaluation evidence reviewed in this annex. The starting point for that redesign is

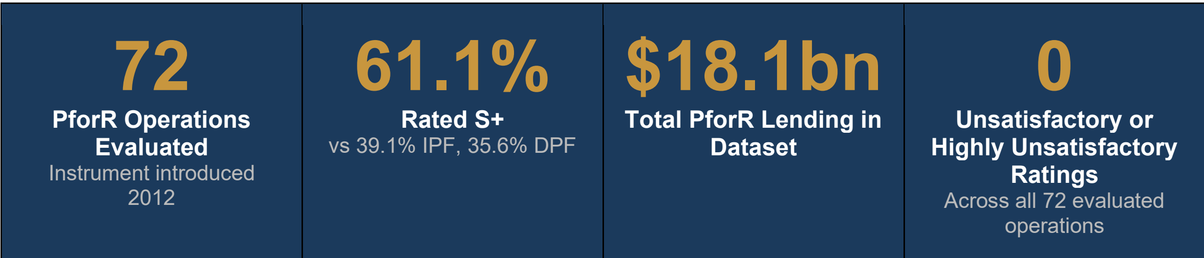
the lesson the data most consistently identify: match the instrument to the context, not the context to the instrument.

Data source: IEG evaluation database. All calculations based on IEG outcome, quality, and M&E ratings as recorded in the dataset. Lessons synthesis derived from qualitative analysis of lesson text appended to IEG evaluations of below-S+ DPF operations (n = 995 with substantive lesson records). Committed amounts in current USD.

ANNEX F — PforR & Staff Incentives

Program-for-Results: Outcome Performance, Design Logic, and Lessons Learned

Source: IEG evaluation database (n = 72 PforR operations with outcome ratings). All figures derived from IEG outcome, quality, and M&E ratings. S+ denotes Satisfactory or Highly Satisfactory. PforR was introduced in 2012 and the evaluated cohort covers the period FY2016–2025. Comparisons to IPF and DPF figures draw on the same dataset.



1. What Is Program-for-Results?

Program-for-Results (PforR) is the World Bank's newest lending instrument, introduced in 2012. It disburses funds against the achievement of pre-defined Disbursement-Linked Indicators (DLIs) — verified results — rather than against expenditure inputs (as IPF does) or policy conditions (as DPF does). This is a fundamental structural difference. A government receives PforR funds only after an independent verification agent confirms that agreed results have been achieved.

PforR supports existing government programmes rather than creating parallel implementation structures. It uses the government's own financial management, procurement, and accountability systems. The Bank conducts a Systems Assessment at design to evaluate whether those systems are adequate; where they are not, the Systems Improvement Plan embedded in the PforR design specifies what improvements the government must make.

The instrument is explicitly results-based: DLIs are typically a mix of output indicators (facilities built, systems established) and outcome indicators (utilisation rates, coverage levels, institutional performance improvements). Disbursement of tranches is conditional on verified achievement — not on government attestation, and not on compliance with procedural conditions.

> **THE CORE DESIGN LOGIC**
> *PforR inverts the disbursement logic of other instruments. DPF disburses against policy conditions that may not produce sustained reform. IPF disburses against expenditure on inputs that may not produce results. PforR disburses only when independent verification confirms results have been achieved. This is why it performs better.*

2. Outcome Performance: A Genuinely Different Picture

2.1 Portfolio-Wide Performance

The PforR outcome distribution is qualitatively different from both IPF and DPF. Across all 72 evaluated operations with outcome ratings, 61.1 per cent achieve S+. No operation in the evaluated dataset is rated Unsatisfactory or Highly Unsatisfactory. The tail of severe failure that characterises the IPF and DPF portfolios is absent.

Outcome Rating	Operations	Share
Highly Satisfactory	12	**16.7%**
Satisfactory	32	**44.4%**
Moderately Satisfactory	22	**30.6%**
Moderately Unsatisfactory	6	**8.3%**
Unsatisfactory	0	**0.0%**
Highly Unsatisfactory	0	**0.0%**

The comparison with IPF and DPF is stark:

Instrument	Operations Evaluated	S+ Rate	Unsatisfactory or below
PforR	72	**61.1%**	**0 (0.0%)**
IPF	7,141	**39.1%**	**821 (11.5%)**
DPF	1,551	**35.6%**	**131 (8.5%)**

PforR outperforms IPF by 22 percentage points and DPF by 25.5 percentage points on S+ rates. The absence of Unsatisfactory ratings across 72 operations — compared to 11.5 per cent U/HU in IPF and 8.5 per cent in DPF — is the most striking single feature of the data.

2.2 The Reliability Question: Why Is the PforR Data Unusual?

Before attributing PforR's outperformance entirely to instrument design, three interpretive caveats are necessary:

Small sample: 72 evaluated operations is a fraction of the IPF (7,141) and DPF (1,551) cohorts. PforR was introduced in 2012 and the evaluated cohort is heavily weighted toward FY2020–2025. Statistical conclusions from 72 operations carry wide uncertainty bands.

Selection bias: PforR requires a functioning government programme, adequate country systems, and a credible DLI verification framework. It is not deployed in the most challenging implementation environments. The instrument is being used in contexts where success is more achievable — the comparison is not apples-to-apples.

Institutional incentives around a new instrument: Both Bank management and borrowing governments have incentives to demonstrate the success of a new instrument that the Bank has publicly championed. This does not mean ratings are inflated, but IEG has noted a 'novelty effect' in previous instrument evaluations.

CAVEAT ON THE DATA
The 61.1% S+ rate is encouraging and likely reflects genuine design advantages of the PforR instrument. But 72 operations is too small a sample to draw definitive conclusions, selection bias is real (PforR goes to better contexts), and the complete absence of Unsatisfactory ratings — across any region, any sector, any context — is statistically unusual and warrants scrutiny as the cohort grows.

2.3 Regional Performance

Regional variation in PforR is substantial and illuminating. East Asia and Pacific records a 92.3 per cent S+ rate — but with only 13 operations. South Asia (40.0% S+, 15 operations) is the weakest regional performer. Africa combined records approximately 56 per cent S+ — significantly better than Africa's IPF rate (30.7%) or DPF rate (21.1%).

Region	Operations	S+ Rate	Committed
East Asia and Pacific	13	**92.3%**	**$4.4bn**
MENA	9	**77.8%**	**$1.7bn**
Eastern and Southern Africa	21	**61.9%**	**$4.2bn**
Europe and Central Asia	6	**50.0%**	**$0.5bn**
Western and Central Africa	4	**50.0%**	**$1.6bn**
South Asia	15	**40.0%**	**$4.5bn**
Latin America and Caribbean	4	**25.0%**	**$1.1bn**

The Africa data is particularly notable. ESA records 61.9 per cent S+ on PforR — compared to 30.2 per cent S+ on IPF and 23.5 per cent S+ on DPF in the same sub-region. If this performance is sustained as the PforR cohort in Africa grows, it would represent a significant argument for expanding PforR deployment in contexts where IPF and DPF have failed.

2.4 FCS Performance

In a counterintuitive finding, PforR performs marginally better in FCS contexts (66.7% S+, 12 operations) than in non-FCS contexts (60.0% S+, 60 operations). This is the opposite of the pattern observed in IPF and DPF, where FCS consistently records the worst outcomes.

The explanation is likely selection: PforR in FCS requires an existing government programme with adequate systems — a high bar in fragile contexts. The 12 FCS PforR operations in the dataset are therefore not representative of the FCS portfolio broadly; they are the subset of FCS operations where government capacity was sufficient to support PforR deployment in the first place.

2.5 Quality Ratings

Quality ratings for PforR are substantially better than for IPF or DPF across all dimensions:

Quality Dimension	PforR	IPF	DPF
M&E: High or Substantial	**85.0% (61/72)**	**40.2%**	**20.8%**
M&E: Modest or Negligible	**15.0% (11/72)**	**59.8%**	**73.5%**
QaE: Satisfactory or above	**66.7% (48/72)**	**44.1%**	**45.1%**
QoS: Satisfactory or above	**77.8% (56/72)**	**56.3%**	**56.9%**

The M&E quality comparison is the most significant. Across PforR operations, 85 per cent receive Substantial or High M&E ratings — compared to 40 per cent for IPF and 21 per cent for DPF. This is not coincidental: the PforR instrument mandates results verification as a condition of disbursement. M&E is not optional or peripheral in PforR — it is structurally embedded in the financing mechanism. The instrument forces M&E quality in a way that neither IPF nor DPF does.

THE M&E ADVANTAGE
PforR achieves 85% Substantial or High M&E ratings compared to 40% for IPF. This is because PforR disburses against verified results — the instrument cannot function without functioning M&E. This is the clearest argument for PforR's structural superiority: it makes good measurement a financial precondition, not a design aspiration.

3. Is PforR Set Up to Move Money Out the Door?

This is the central accountability question about the PforR instrument, and the data allow a nuanced answer.

3.1 The Concern

Critics of PforR have argued that Disbursement-Linked Indicators can be designed to be easily achievable — to legitimate disbursement while providing the appearance of results-based accountability. If DLIs are set at levels below what would have happened without Bank financing (low additionality), or if 'results' are defined as process steps rather than outcomes, PforR disburses readily against non-additional or non-meaningful change. This would make PforR a sophisticated version of the same volume-over-outcomes incentive structure that drives DPF and IPF underperformance.

3.2 What the Lesson Record Shows

The lesson text for PforR operations — drawn from both S+ and below-S+ evaluations — identifies DLI design quality as the most consequential variable in PforR performance. Specific failure modes identified in the evaluated cohort:

- DLIs set at levels that were achievable without Bank financing — 'the DLIs were appropriate for the areas associated with family assistance but were of limited use for the area associated with improving skills development; the level of ambition was limited.'
- DLI scope too diffuse — 'spreading the financing so thin limited the financial incentives to reach ambitious targets' (Croatia health programme, where PforR funding was less than 10 per cent of total programme cost).

- Process DLIs rather than outcome DLIs — 'action plans agreed upon with government can identify actions that are deemed necessary for achieving outcomes, but should not be a substitute for project readiness.'
- DLIs that require readiness preconditions not met at approval — 'action plan items were in fact measures necessary for successful project launch and initial implementation; they were not fulfilled until two years into the programme.'

These are recognisable variants of the conditionality design failures documented in DPF. The structural difference is that in PforR, they tend to produce Moderately Satisfactory rather than Unsatisfactory outcomes — the verification mechanism prevents complete disbursement failure even when DLI design is poor.

3.3 What the Data Suggest

The honest answer is: the PforR instrument has better safeguards against money-out-the-door dynamics than DPF or IPF, but is not immune to them. The safeguards are structural — independent verification, results-linked disbursement, mandatory M&E. The vulnerabilities are in DLI design — low ambition, process framing, inadequate additionality assessment.

The cleaner lesson from the successful PforR operations is different from the IPF/DPF lesson record. Vietnam's Rural Water Supply PforR (Highly Satisfactory) specifically notes: 'the verification-based disbursement process was an important learning-by-doing tool that incentivised results — the verification protocols included the engagement of a credible independent agent with clear responsibilities and capacity.' Uganda's Municipal Infrastructure PforR notes that local development forums 'fostered transparency and accountability and improved relationships between local governments and their constituents.'

This suggests PforR, when well designed, generates accountability dynamics that extend beyond financial disbursement to actual behavioural change in implementing institutions. This is qualitatively different from anything identified in the IPF or DPF lesson records.

> **ON THE MONEY-OUT-THE-DOOR QUESTION**
> PforR is structurally less vulnerable to volume-over-outcomes dynamics than IPF or DPF. Independent verification of results is a genuine safeguard. But DLI design quality determines whether that verification mechanism produces meaningful accountability or merely legitimises pre-determined disbursement. The lesson record identifies low-ambition and process-framed DLIs as the primary failure mode — the same conditionality design problem that characterises DPF, in a different structural form.

4. Lessons Learned: What PforR Evaluations Show

4.1 Lessons from Successful Operations (S+ cases)

The lesson record from S+ PforR operations identifies a distinct set of success factors not commonly found in IPF/DPF lessons:

- Government alignment: PforR works best when it supports a programme the government already owns — 'the PforR instrument was effective in supporting the bridge programme, which had already been part of the Government's vision and plan.' The absence of this alignment is consistently associated with poorer outcomes.

- Long-term series engagement: multiple PforR operations in the same sector allow design improvement across generations — 'this programme benefited from a series of urban operations; the government used lessons from the preceding operations.' Vietnam ran four follow-on PforR operations after its first; Ethiopia's urban programme drew explicitly on lessons from ULGDP I.
- Verification system quality: successful operations invest in credible independent verification agents — 'a transparent M&E system at national and city levels, linking it with the MIS and independent verification, was an important learning-by-doing tool.'
- Sector-level engagement: PforR creates institutional space for dialogue that IPF does not — 'this approach committed not only the Ministry of Transport but also the Ministry of Economy and Finance — the main financier of the programme.'
- Institutional incentive alignment: Kenya's National Safety Net PforR: 'institutional strengthening is facilitated when a confirmed source of financing is readily available — time-bound institutional reforms triggered by policy reforms benefited from timely availability of technical assistance financing.'

4.2 Lessons from Below-S+ Operations

Below-S+ PforR lessons converge on three failure modes:

- Instrument-context mismatch: 'the PforR design of sequential DLI targets left little room for slippage; the implementation period left little room for gradual capacity building' (Mozambique Disaster Risk). PforR's results-linked disbursement requires a minimum level of government systems functionality — when that threshold is not met, performance suffers.
- DLI ambition and framing: 'DLIs were appropriate for some areas but of limited use in others; the level of ambition regarding how the PforR would drive reform in all sectors was limited' (Brazil Ceará). Process DLIs without outcome verification produce disbursement without transformation.
- Political instability: 'government policy nationalised what had been a decentralised system, disrupting the programme's financial flows to local governments' (Tanzania Urban Local). PforR shares with DPF vulnerability to political ownership reversal — the DLI structure does not fully insulate operations against government policy changes.

6. Why PforR Outperforms: A Structural Explanation

The PforR performance advantage is real but requires careful explanation. It is not simply that PforR is deployed in better contexts — selection bias explains some but not all of the gap. The structural explanations that the data support are:

- Disbursement is conditional on verified results, not on inputs expended or conditions adopted. This eliminates the failure mode — common in IPF and DPF — in which money is disbursed against nominal compliance while development outcomes are not achieved.
- M&E is a financial precondition, not a design aspiration. The instrument cannot disburse without functioning measurement and verification. This eliminates the 60% Modest/Negligible M&E that characterises IPF and produces the 85% Substantial/High M&E rate in PforR.
- PforR uses government systems, which forces a realistic assessment of those systems at preparation and creates incentives for their improvement. IPF's parallel PIU model can sidestep weak government systems without fixing them; PforR cannot.

- The DLI structure makes government ownership legible — governments negotiate DLIs and accept performance accountability for them in a way that prior actions in DPF or expenditure signing in IPF does not require.

The limits of the structural argument are equally important: with 72 evaluated operations, no severe FCS deployments, and a portfolio concentrated in middle-income and stronger IDA contexts, the PforR performance record cannot yet support confident generalisations about the instrument's performance at scale in the most challenging environments.

SUMMARY

PforR genuinely outperforms. The instrument's structural design — disbursement against verified results, mandatory M&E, use of government systems — addresses the core failure modes of IPF and DPF. With 72 evaluated operations, the record is promising but not yet conclusive. The instrument's extension into harder FCS and low-capacity contexts — which is the direction of IDA21 — will be the real test of whether the design advantages survive deployment in the most challenging environments.

Data source: IEG evaluation database. All calculations based on IEG outcome, quality, and M&E ratings. Committed amounts in current USD. PforR introduced 2012; evaluated cohort FY2016–2025.

ANNEX G — Project Cycle Timeline Data

FCS Countries (IDA & Blend): Average Project Cycle Timelines — Board Approval to Loan Effectiveness, and Loan Effectiveness to Closing

This table presents, for each IDA and Blend FCS country, the average number of days between: (i) Board Approval and Loan Effectiveness (BA→LE), a measure of the time required for a project to become legally operative after Board approval; and (ii) Loan Effectiveness and Project Closing (LE→Closing), the active implementation period. Averages are calculated across all IEG-evaluated projects with evaluation years in FY2015–2026, matched by project code (P-code) to Board Approval Date, Loan Effective Date, and Project Closing Date from the World Bank Projects database. Data is available for 543 of 560 projects for the BA→LE interval and 559 of 560 for the LE→Closing interval.

The overall average across the FCS IDA/Blend portfolio is 152 days from Board Approval to Loan Effectiveness and 2,019 days (5.5 years) from Loan Effectiveness to Closing. These figures carry two distinct implications for FCV strategy design. The BA→LE interval — averaging 5 months — reflects legal, fiduciary, and counterpart-readiness processes that must be completed before funds can flow. In FCS contexts this interval is particularly sensitive: government counterparts change, security conditions shift, and the assumptions embedded in the project design at Board approval may no longer hold by the time the project becomes effective. Prolonged BA→LE periods in the most fragile contexts — Congo Rep. (324 days), Nigeria (309 days), Kosovo (299 days), Guinea (294 days) — effectively compress the available implementation window. The LE→Closing interval is not in itself a performance metric, but it calibrates the realistic timeframe within which a Bank project must demonstrate results in FCS. At 5.5 years on average, this is already a shorter horizon than most institutional-capacity-building or governance-reform objectives require to produce durable change — which helps explain the outcome pattern documented in Annex C and Annex D.

The fastest BA→LE performers — Somalia (70 days), Afghanistan (70 days), Gambia (72 days), Tuvalu (92 days), Solomon Islands (91 days) — are not necessarily the best-resourced counterpart environments; many reflect the use of pre-agreed legal frameworks, emergency lending instruments, or simplified effectiveness conditions in contexts where the alternative is no engagement at all. The slowest — Congo Rep. (324 days), Nigeria (309 days), Kosovo (299 days), Guinea (294 days), Cameroon (240 days) — include several Blend countries where more complex financing structures and IBRD-standard legal documentation requirements add time. The data supports a practical recommendation: in FCS contexts, effectiveness conditions should be the minimum necessary for legal operability, with remaining safeguard and capacity requirements addressed through implementation support, not as pre-conditions for fund flow.

Country	IDA / Blend	Projects	Avg BA → LE (days)	Avg LE → Closing (days)	Avg LE → Closing (years)	Avg BA → Closing (years)
Eastern & Southern Africa | 156 projects BA→LE: 149d LE→Close: 2045d (5.6yr) BA→Close: 2025d (5.5yr)						
Burundi	IDA	20	172	1834	5.0	5.4
Comoros	IDA	10	105	1264	3.5	3.7
Congo, Dem. Rep.	IDA	35	232	2461	6.7	7.3
Ethiopia	IDA	14	106	2640	7.2	7.5
Madagascar	IDA	8	111	2460	6.7	7.0
Malawi	IDA	2	145	1064	2.9	3.3
Mozambique	IDA	19	166	2351	6.4	6.8
Somalia	IDA	12	70	1727	4.7	4.9
South Sudan	IDA	15	132	1696	4.6	5.0
Sudan	IDA	15	98	1480	4.1	4.3
Zimbabwe	Blend	6	100	1961	5.4	5.6
Western & Central Africa | 230 projects BA→LE: 172d LE→Close: 1984d (5.4yr) BA→Close: 2160d (5.9yr)						
Burkina Faso	IDA	19	152	2092	5.7	6.1
Cameroon	Blend	11	240	2169	5.9	6.6

Country	IDA / Blend	Projects	Avg BA → LE (days)	Avg LE → Closing (days)	Avg LE → Closing (years)	Avg BA → Closing (years)
Central African Rep.	IDA	20	115	1690	4.6	4.9
Chad	IDA	18	167	1707	4.7	5.1
Congo, Republic of	Blend	14	324	2560	7.0	7.9
Côte d'Ivoire	Blend	12	127	1515	4.1	4.5
Gambia, The	IDA	8	72	1645	4.5	4.7
Guinea	IDA	3	294	2837	7.8	8.6
Guinea-Bissau	IDA	11	148	1963	5.4	5.8
Liberia	IDA	16	119	1659	4.5	4.7
Mali	IDA	29	150	2107	5.8	6.2
Niger	IDA	22	138	2212	6.1	6.4
Nigeria	Blend	20	309	2486	6.8	7.7
Sierra Leone	IDA	17	160	1736	4.8	5.1
Togo	IDA	10	126	1508	4.1	4.4
Middle East, North Africa & Afghanistan | 63 projects BA→LE: 94d LE→Close: 2031d (5.6yr) BA→Close: 2227d (6.1yr)						
Afghanistan	IDA	43	70	1993	5.5	5.6
Djibouti	IDA	2	214	2208	6.0	6.6
Yemen	IDA	18	137	2101	5.8	6.1
Europe & Central Asia | 13 projects BA→LE: 279d LE→Close: 2501d (6.8yr) BA→Close: 3324d (9.1yr)						
Kosovo	IDA	12	299	2383	6.5	7.3
Uzbekistan	Blend	1	42	3924	10.7	10.9
East Asia & Pacific | 64 projects BA→LE: 124d LE→Close: 1897d (5.2yr) BA→Close: 2022d (5.5yr)						
Kiribati	IDA	10	123	1209	3.3	3.6
Lao PDR	IDA	5	125	2205	6.0	6.4
Marshall Islands	IDA	4	135	1323	3.6	4.0
Micronesia, Fed. States of	IDA	2	138	2074	5.7	6.1
Myanmar	IDA	11	152	2426	6.6	7.1
Papua New Guinea	Blend	9	126	2301	6.3	6.5
Solomon Islands	IDA	12	91	1718	4.7	4.9
Timor-Leste	Blend	5	158	2535	6.9	7.4
Tuvalu	IDA	6	92	1366	3.7	4.0
Latin America & Caribbean | 26 projects BA→LE: 135d LE→Close: 2163d (5.9yr) BA→Close: 2292d (6.3yr)						
Haiti	IDA	26	135	2163	5.9	6.3
South Asia | 8 projects BA→LE: 128d LE→Close: 2098d (5.7yr) BA→Close: 2226d (6.1yr)						
Nepal	IDA	8	128	2098	5.7	6.1
OVERALL AVERAGE — All IDA & Blend FCS		**560**	**152**	**2,019**	**5.5**	**5.9**

Notes: IDA and Blend lending group countries only. Averages calculated across IEG-evaluated projects with Evaluation FY 2015–2025, matched by P-code to World Bank Projects database (as of 18 February 2026). BA→LE = Board Approval Date to Loan Effective Date. LE→Closing = Loan Effective Date to Project Closing Date. Observations excluded where interval is zero or negative (data entry anomalies) or exceeds 3,000 days for BA→LE. 543/560 projects have valid BA→LE data; 559/560 have valid LE→Closing data. BA→LE colour scale: green ≤90d; amber 91–180d; orange 181–300d; red >300d. LE→Closing and BA→Closing colour scales: light to dark blue by duration (longer = darker).

ANNEX H — Project Level Accountability – The Circular Evaluation Problem

Project-Level Accountability — The Circular Evaluation Problem

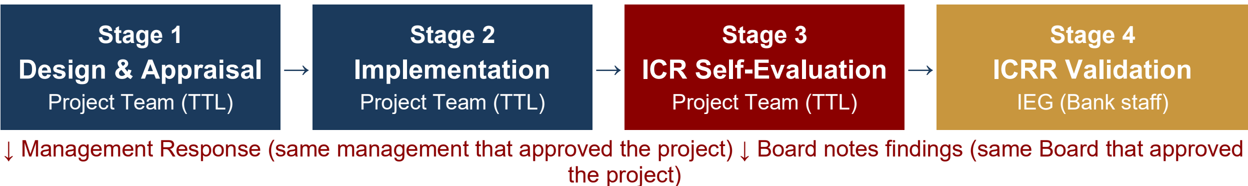
Part I

1. The World Bank's Project Completion Accountability Chain

At the core of the World Bank's project accountability system is a structural paradox: the same team that designs a project, implements it, and then writes the primary evaluation of whether it succeeded. This annex documents the circular logic embedded in the World Bank's implementation completion and results reporting (ICR) process, quantifies the rating inflation it produces, compares the World Bank's approach against other major multilateral development banks and global best practice, and proposes the structural reforms necessary to make project-level evaluation credible.

1.1 How the ICR Process Works: The Accountability Chain

When a World Bank lending operation closes, the task team leader (TTL) and project team are required to prepare an Implementation Completion and Results Report (ICR). The ICR is the primary formal record of whether the project achieved its development objectives. It contains outcome ratings, assessments of Bank performance, and lessons for future operations. Understanding who prepares each element of this chain — and who reviews it — is essential to evaluating whether the system produces credible accountability.



Stage 1 — Project Design and Appraisal: The project team, led by the TTL, prepares the Project Appraisal Document (PAD) or Programme Paper. The TTL selects peer reviewers. The Quality Assurance Group (QAG), which provided independent review of appraisal quality, was disbanded in 2014. Global Practices were expected to absorb its function. IEG data shows only 44% of IPF projects are rated Satisfactory or above on Quality at Entry, confirming that no effective independent quality check at preparation has replaced QAG.

Stage 2 — Implementation: The project team implements the operation and prepares Implementation Status and Results Reports (ISRs) throughout the project lifecycle. ISRs are self-assessments. The TTL responsible at implementation is typically a different person from the TTL at appraisal — average project implementation takes 76 months against a planned 48 months, with multiple TTL changes across the lifecycle. This discontinuity means the TTL preparing the ICR typically did not design the project.

Stage 3 — ICR Self-Evaluation: The ICR is prepared by the closing TTL and project team. The team self-rates outcome, Bank performance (quality at entry and quality of supervision), and M&E quality. The closing TTL selects who contributes to the document. The team's own ratings determine the first official record of the project's success or failure. The World Bank's ICR guidelines require candour,

but the institutional incentive structure — described in the Raballand analysis in this report — produces systematic upward bias.

Stage 4 — IEG Validation (ICRR): The ICR is submitted to the Independent Evaluation Group (IEG), which reviews it within 180 days. The IEG ICR reviewer interviews the last TTL and produces an Implementation Completion and Results Report Review (ICRR). The ICRR is a desk-based validation: it checks the evidence and reasoning in the ICR and arrives at its own ratings. IEG may downgrade (or upgrade) the team's self-ratings. Coverage is 100% of all completed lending operations. However, this is a validation of a self-evaluation, not an independent evaluation from primary evidence.

Management Response: IEG sends findings to management before finalisation. The same management that approved and supervised the project can comment on IEG's assessment. Comments are published alongside the ICRR. Management cannot alter IEG ratings, but the response creates a parallel record contesting or qualifying findings.

Board: The Board receives aggregated IEG findings through the annual Results and Performance (RAP) report. The same Board that approved the individual projects reviewed in any given annual cohort hears the aggregate outcome data. There is no mechanism by which the Board holds management accountable for the patterns of underperformance the RAP documents year after year.

2. The Five Structural Problems

2.1 The Self-Evaluation Problem: The Team Rates Itself

The ICR is, formally, a self-evaluation. The team that implemented the project writes the official assessment of whether the project worked. This is the most fundamental accountability problem in the chain.

IEG's own 2016 evaluation of the Bank Group's self-evaluation systems — published as Behind the Mirror — found that incentives in the ICR system are 'not conducive to conducting high-quality self-evaluation.' The report noted that 'most staff do not view the self-evaluation systems as a source of timely, credible, and comprehensive information.' Writing a rigorous, honest ICR that documents design failures does not benefit the team that designed the project. The institutional incentive is to document lessons in general terms while protecting the team's ratings.

> "Individual authors of self-evaluation reports often learn something from visiting the project and writing up their analysis — it would be strange otherwise — but little knowledge flows beyond the authors themselves. It is rare that business units analyse completed self-evaluations or mine their lessons. Lessons rarely turn into revised policies, guidelines, or procedures. IEG's evaluations point out the same weak spots and missed opportunities, year after year." — IEG, Behind the Mirror (2016)

The consequence is systematic upward bias: ICR outcome ratings are higher than IEG ICRR validation ratings, in every annual cohort across the RAP 2020–2024 series. The team rates its own project more generously than independent review confirms. This is not a coincidence or a marginal rounding difference. It is a persistent structural pattern produced by a system that asks people to assess their own work.

2.2 The Author Appointment Problem: The Team Selects the Evaluator

In the World Bank's ICR system, the closing TTL and project team do not merely write the primary self-evaluation — they also effectively control who contributes to it. There is no mechanism requiring that ICR peer review be conducted by staff with no connection to the project. The team can invite colleagues, former teammates, or managers who supervised the project to review the draft ICR.

This mirrors a well-documented problem in financial auditing, where auditors were selected and paid by the companies they audited — a conflict of interest so fundamental that it contributed to the collapse of Enron and the subsequent Sarbanes-Oxley reforms requiring auditor independence. The World Bank's ICR system does not require equivalent independence. The 2014 disbandment of QAG removed the main external check on the self-evaluation quality of individual operations without replacing it with an equivalent mechanism.

2.3 The IEG Rotation Problem: The Evaluators Are Bank Staff

IEG is formally independent: its Director-General reports to the Board rather than to Bank management, and it has its own budget and recruitment procedures. These are genuine and important protections. However, IEG is staffed by World Bank Group employees. In practice, staff move between IEG and operational Vice Presidencies in both directions.

A staff member who spent five years as a TTL in the Education Global Practice, designing and implementing operations in East Africa, may subsequently join IEG as an evaluator of education projects in East Africa. The formal independence of IEG's reporting line does not prevent the evaluator from having personal knowledge of, relationships with, or institutional loyalty to the teams and approaches they are now asked to evaluate. Conversely, staff who have worked in IEG return to operational roles, carrying evaluation culture into the institution but also illustrating that IEG is not a distinct career path but part of the general Bank employment pool.

The External Review of IEG commissioned by the Board (2015) noted that 'the possibility of term extensions can create tension with IEG's key attribute of institutional and operational independence' even at the Director General level. The ECG Good Practice Standards on independence of evaluation functions identify staff movement between evaluation and operational units as a structural independence risk. The ADB's IED explicitly addresses this through a recusal rule: IED evaluators must exclude themselves from any evaluation involving a project, programme, or strategy on which they worked. The World Bank has no equivalent formal requirement.

2.4 The Management Response Problem: The Evaluatee Responds to the Evaluation

The ICR-ICRR process concludes with management providing a formal response to IEG's findings. This response is published alongside the ICRR and becomes part of the permanent project record. The management providing this response is the same management that approved the project, supervised it, and oversaw the preparation of the ICR that IEG is validating.

At the portfolio level, the Management Action Record (MAR) system requires management to self-assess its own compliance with IEG recommendations. IEG independently validates these self-assessments. The pattern is consistent: management consistently rates its own compliance with IEG recommendations more generously than IEG's independent validation confirms. The MAR system replicates at the portfolio level exactly the same structural problem that exists at the project level: the evaluatee rates itself.

> The management response system gives the same institution that approved and implemented the operations a formal, published channel to contest the findings of the evaluation of those operations. It is the institutional equivalent of allowing a company's management to append its own assessment

of the external audit, disputing its findings, and having that assessment published alongside the audit report as equally weighted evidence.

2.5 The Board Feedback Loop Problem: The Approver Reviews the Outcomes

At the governance level, the Board receives IEG's aggregated findings through the annual RAP report. The Board that approves this year's RAP findings is the same Board — or a Board constituted on the same model — that approved the operations being evaluated. The co-optation problem described in the Board Governance annex to this report is not only a structural issue for management accountability: it means the body that should be holding management to account for project outcomes is itself a co-author of those outcomes.

There is no annual session in which the Board asks management: 'You approved 7,141 IPF operations in the last two decades. Sixty per cent failed to achieve satisfactory development outcomes. Explain what you will do differently.' The RAP is received, noted, and the approval cycle continues.

3. Quantifying the Problem: ICR vs ICRR Rating Divergence

The ICR-ICRR rating divergence is the clearest quantitative evidence that the self-evaluation system produces inflated assessments. The following patterns are documented across the IEG RAP 2020–2024 series and the IEG evaluation database.

Pattern	Frequency / Direction	Implication
ICR ratings higher than ICRR validation	Consistent pattern across IEG RAP reports 2020–2024. RAP 2024 notes ICR-ICRR divergence remains a standing feature of the evaluation system	Task teams systematically rate their own projects more favourably than independent validators. Self-evaluation overstates performance.
PPARs (field-based) diverge further from ICR than ICRRs (desk-based)	PPARs — which involve country visits and stakeholder interviews — tend to produce lower ratings than desk-based ICRRs. Coverage: approximately 20% of portfolio	The further evaluation moves from the task team's own documentation and into independent field evidence, the more performance is downgraded. The primary accountability mechanism (ICR) is the one furthest from independent evidence.
Management response disputes IEG findings	Management response system (MAR) allows management to rate its own compliance with IEG recommendations. Management consistently rates its own compliance higher than IEG's assessment	Same pattern at portfolio level as at project level: the evaluatee rates its own performance more favourably than the evaluator. MAR is not an independent verification; it is management self-certification.

The significance of the PPAR finding is particularly important. PPARs are field-based evaluations conducted by IEG staff who visit project sites, interview beneficiaries and implementing agencies, and review evidence independently of the project team's documentation. They represent the closest approximation to genuinely independent project evaluation that the Bank Group undertakes. Their

coverage — approximately 20% of the portfolio — means that for 80% of completed operations, the primary evidence is either the task team's ICR (self-evaluation) or IEG's desk-based ICRR (validation of that self-evaluation). Neither involves independent field evidence.

The consistent finding that PPARs produce lower ratings than ICRRs, and ICRRs produce lower ratings than ICRs, is a gradient of independence: as evaluation becomes more independent of the task team's documentation and more reliant on primary field evidence, performance appears worse. This gradient directly implies that the headline S+ rates reported in the IEG database — already poor — are likely overstating actual development performance.

4. Comparative Analysis: How Other MDBs Manage Project Completion Accountability

The table below summarises the project completion evaluation architecture of the World Bank and five comparator institutions, along with the ECG/OECD-DAC good practice standard.

Institution	Who Prepares Completion Report	Evaluator Independence	Staff Rotation Firewall	Management Response
World Bank	Project team (TTL) — same staff who designed and implemented the project	IEG is within the Bank Group; DG reports to Board but staff are Bank employees; IEG staffers regularly rotate back to operations	None formal. Staff move freely between IEG and operational VPs. No restriction on evaluating sectors where they previously worked	Same management that approved and supervised the project comments on its own performance record; comments are published alongside IEG ratings
Asian Development Bank (ADB)	Project/operations team prepares Project Completion Report (PCR) as self-evaluation	IED is organisationally independent; reports to Board via Development Effectiveness Committee. **IED evaluators must recuse themselves from any evaluation involving a project they worked on.**	Explicit recusal rule: IED staff who worked on a project cannot evaluate it. Formal conflict-of-interest policy	Management reviews draft; IED Director General finalises findings independently. Management response published separately; does not alter IED ratings
African Development Bank (AfDB)	Operations department prepares Project Completion Report (PCR) immediately after project closure	IDEV is independent; Evaluator General reports directly to Board via CODE. Staff from operations cannot move directly into evaluation roles without clearance	Formal separation between operational divisions and IDEV. IDEV has own recruitment and budget controlled by Board, not management	Management prepares formal Action Plan in response to IDEV findings. Board (CODE) reviews both IDEV findings and management response. IDEV tracks

Institution	Who Prepares Completion Report	Evaluator Independence	Staff Rotation Firewall	Management Response
				implementation of recommendations
European Bank for Reconstruction and Development (EBRD)	Operations staff prepare Operation Performance Assessment (OPA) for every project	EvD is independent; Head reports to Board Audit Committee. Board-commissioned 2019 Kirk Report found management was not meeting self-evaluation obligations. EvD designs OPA template itself; operations staff fill it in	EvD uses random selection for independent audits — operations staff do not control which projects receive full independent evaluation. EvD Director has fixed term, non-renewable	Management cannot alter EvD findings. Kirk Report (2019) specifically flagged management non-compliance with self-evaluation. EvD has authority to report non-compliance directly to Board Audit Committee
Inter-American Development Bank (IDB)	Office of Evaluation and Oversight (OVE) is fully external to operations. Project teams prepare progress reports; OVE does independent evaluation — not a validation of a self-evaluation	OVE is structurally separate: no joint recruitment pipeline with operations. OVE Director reports to Board of Governors. OVE produces its own evaluations from scratch, not validating task team self-assessments	OVE staff cannot be seconded from or to operational VPs. Evaluators are recruited independently; career path does not include operational roles	OVE findings go directly to Board. Management prepares formal response to Board, not to OVE. Board adjudicates any disputes between OVE and management independent of management hierarchy
ECG / OECD-DAC Best Practice Standard	Self-evaluation (by operational teams) is explicitly supplemented or replaced by independent evaluation. Self-evaluation alone is not accepted as the primary accountability mechanism	Independence defined across 4 dimensions: organisational, behavioural, conflict-of-interest avoidance, and protection from outside interference. ECG GPS (2010): evaluators must have no operational stake in what they evaluate	Evaluators who worked on an operation cannot evaluate it. Cooling-off period before former operational staff may join evaluation function. Rotation from evaluation back to operations is a structural independence risk	Management response is published but advisory only. Findings and ratings are set by the independent evaluation unit. Board, not management, is the primary audience for evaluation findings

4.1 Key Distinctions from Best Practice

Several features of the World Bank's approach are structurally weaker than comparators and international standards.

- **No formal recusal rule for IEG evaluators.** ADB's IED requires evaluators to exclude themselves from evaluating projects they worked on. This is the single most direct structural protection against the rotation problem, and the World Bank does not have it.

- **Validation of self-evaluation rather than independent evaluation.** The IDB's OVE produces independent evaluations of operations — it does not validate a self-evaluation prepared by the operations team. The World Bank's ICRR is a validation of the ICR, not an independent evaluation. The primary source document — the one with the institutional authority of the task team — remains the ICR.
- **No Board Audit Committee for evaluation findings.** The EBRD's EvD reports findings to the Board Audit Committee, which has an independent mandate to assure evaluation quality and management compliance. The World Bank's IEG findings reach the Board through the RAP, which goes to the full Board — but via management, not directly to an oversight committee.
- **Management Action Record self-certification.** The MAR system allows management to certify its own compliance with IEG recommendations. The AfDB's IDEV tracks recommendation implementation independently, reporting to the Board's CODE committee. The Bank's system conflates management self-assessment with accountability.
- **PPAR coverage too low.** Field-based PPARs — the only evaluation modality that goes beyond documentation to primary evidence — cover approximately 20% of the Bank's portfolio. For 80% of closed operations, no field-based independent evidence is ever gathered. The IDB and ADB have higher proportions of field evaluations in their evaluation portfolios.

6. ECG Good Practice Standards and OECD-DAC Principles

The Evaluation Cooperation Group (ECG) was established in 1996 specifically to harmonise MDB evaluation practice. Its Good Practice Standards (GPS) on independence — approved in 2010 with the World Bank Group as lead drafter — define independence across four dimensions that are directly relevant to the structural problems documented in this annex.

Organisational Independence

The evaluation function must have independent budget authority, its own staffing and recruitment processes, and formal reporting lines to the Board rather than to management. IEG meets these standards formally. The structural weakness is that formal organisational independence is undermined by the staff rotation practices described above — independence in the organisational chart coexists with deep personal and institutional connections between IEG and operational units.

Behavioural Independence

Evaluators must be free from undue influence and must not be subject to management pressure on findings. The management response mechanism — while it cannot alter IEG ratings — creates a formal institutional channel through which management can publicly contest, qualify, and minimise IEG findings. This constitutes a form of pressure on the evaluation function that ECG GPS identifies as a behavioural independence risk.

Avoidance of Conflict of Interest

This is the dimension where the World Bank's IEG most clearly falls below the ECG standard, and where the ADB's IED is most clearly superior. ECG GPS requires evaluators to have no operational stake in what they evaluate. Staff rotation between IEG and operational VPs creates direct conflicts of interest that the formal independence of IEG's reporting line does not address. The absence of a formal recusal rule is a gap against ECG's own standards — which the World Bank helped draft.

Protection from Outside Interference

IEG is reasonably well-protected against external interference. The DG's reporting line to the Board, non-renewable term, and budget independence provide meaningful structural protections. This is the dimension where IEG most clearly meets the ECG standard.

7. What Reform Would Look Like: Structural Changes Required

The problems identified in this annex are structural, not individual. They are not the result of bad-faith behaviour by task teams or IEG staff. They are the predictable consequence of evaluation architecture that asks people to assess their own work, does not formally separate evaluators from the work they evaluate, and provides no enforcement mechanism when evaluation findings document failure. Structural problems require structural solutions.

7.1 Mandatory Recusal Rule for IEG Evaluators

IEG should adopt a formal policy, equivalent to ADB's IED requirement, that evaluators must recuse themselves from any project, programme, country strategy, or sector evaluation involving work they were personally responsible for during their time in operational units. A cooling-off period — a minimum of three years — should apply before former operational staff may evaluate work from units in which they previously worked. This is the single most direct reform available to close the conflict-of-interest gap against ECG GPS.

7.2 ICR Authorship Reform: Independent Co-Authorship or External Validation

The ICR should be prepared either by an independent reviewer appointed by IEG (not by the task team), or co-authored by an IEG-appointed evaluator alongside the task team. At minimum, ICR peer review should be conducted by staff who have had no prior involvement with the project and have no reporting relationship to the project manager. The practice of allowing the task team to select its own ICR peer reviewers should be replaced by IEG-assigned review.

The IDB model — in which OVE produces independent evaluations rather than validating self-evaluations — represents the gold standard. A version of this for the World Bank would involve IEG selecting a sample of completed operations for fully independent evaluation from primary evidence, with no reliance on the ICR as the primary document. PPAR coverage should be expanded from 20% to a minimum of 40% of the closed portfolio, with random selection to prevent operational teams from being able to predict or influence which projects receive full field evaluation.

7.3 Eliminate the Management Response to Individual Project Evaluations

The management response to individual ICRRs should be discontinued. The institutional interest in contesting evaluation findings should be channelled to the portfolio level, where management should be required to explain patterns of underperformance in response to the annual RAP — to the Board, in a formal session, not through a published parallel record alongside individual project ICRRs.

The MAR system should be reformed so that management's compliance with IEG recommendations is independently assessed by IEG and reported to the Board, without management self-certification being accorded equal weight to IEG's independent validation.

7.4 Direct IEG Reporting to Board Oversight Committee

IEG's annual RAP findings should be presented directly to a Board oversight committee — modelled on the EBRD's Board Audit Committee arrangement — before management receives and responds to them. The current sequence, in which management shapes its response before the Board receives the findings, limits the Board's ability to assess findings without prior management framing. IEG should brief the Board's oversight committee on RAP findings first; management's response should come second, as a reaction to Board questions rather than as the framing through which the Board first encounters the data.

7.5 Performance Accountability Linked to Evaluation Findings

The structural disconnection between project outcome ratings and staff career trajectories — documented in the Raballand analysis in this report — means that even a reformed evaluation system will not produce accountability unless evaluation findings are linked to individual and institutional performance consequences. Three minimum requirements: project outcome ratings at closure should be visible in TTL performance assessments, with explicit recognition that designing poor projects has professional consequences; aggregate outcome ratings for Global Practices should be reported to the Board and linked to resource allocation decisions; and Practice Managers whose portfolios consistently underperform the institutional S+ rate should be required to present improvement plans to senior management and the Board.

7.6 The Underlying Governance Connection

Project-level evaluation reform cannot be separated from Board governance reform. The reason the circular accountability problem described in this annex persists — despite being identified clearly in IEG's own evaluations, ECG peer reviews, and multiple external assessments — is that no governance body exists with both the information and the authority to enforce change.

IEG has the information but no enforcement authority. The Board has the authority but is structurally co-opted through project approval. Management has neither the incentive nor the mandate to reform a system that serves its institutional interests by diffusing accountability.

A non-resident Board, delegating project approval to management, and spending its meeting time on portfolio performance review rather than individual project sign-off, would be able to receive IEG's annual findings without the conflict of interest that currently prevents the Board from holding management accountable for patterns of performance the Board itself helped create. Board governance reform and evaluation reform are therefore complementary requirements, not alternatives. Neither alone is sufficient.

Part II

The Accountability Void: When Projects Fail, Nobody Is Responsible

Across the IEG evaluation database, 60% of IPF operations — 4,304 of 7,141 evaluated — did not achieve satisfactory development outcomes. The question this figure raises, and that the World Bank's accountability architecture systematically avoids answering, is: if a project fails, who is responsible, and what happens to them?

The answer, in the current institutional architecture, is: nobody is responsible, and nothing happens to them. This is not a coincidental outcome — it is the predictable result of an accountability chain designed to diffuse responsibility across time, roles, and institutional actors. This section documents how the diffusion mechanism operates, examines the empirical evidence on what actually drives project failure, and proposes specific structural reforms to create genuine individual accountability for poor performance.

9. The Four-Party Blame Loop

When a World Bank lending operation closes with poor outcomes, the following attribution sequence typically plays out in the ICR and ICRR:

- The design Task Team Leader attributes poor outcomes to the country context: political instability, government capacity constraints, or implementation environment factors outside the Bank's control.
- The closing Task Team Leader, who may have inherited the project mid-implementation, attributes failure to predecessor decisions and a design it did not create.
- Management attributes underperformance to borrower government capacity, ownership, or commitment — institutional factors the Bank identified as risks at design but did not price into the probability of success.
- The host government has right of reply in the ICRR, and no mechanism to contest the attribution.

The host government — the primary institutional target of blame in this sequence — cannot contest the Bank's evaluation. It is not a party to the evaluation process. Its counterpart officials may have rotated out of their positions over the project's 6–8 year lifecycle. The government's formal accountability mechanism, the Borrower Performance rating (discontinued from ICRs in 2017), was itself inevitably prepared by the Bank team that had been working with that government, producing a further layer of self-interested assessment.

> The blame loop has no closing mechanism. The design TTL blames the country context and is no longer accountable. The closing TTL blames predecessors and inherits a narrative they didn't create. Management blames implementation. The host government is blamed as an institution but has no voice. IEG validates the outcome but cannot name individuals. The Board notes the pattern and approves the next tranche of lending to the same country. The cycle repeats.

9.1 The TTL Rotation Problem in Numbers

The rotation frequency that enables this diffusion of responsibility is not marginal. IEG country programme evaluations and research on the Bank's staff database document a pattern of routine, systemic TTL turnover that makes individual accountability structurally impossible under the current model.

Context	Projects Examined	Average TTLs per Project	Key Finding
Chad portfolio (IEG CPE, 2021)	44 investment projects	4 TTLs per project (2× Africa average); 11 projects averaged 6 TTLs each	168 TTLs across 44 projects; high rotation cited as 'a major factor complicating delivery'
World Bank IPF portfolio — planned vs actual duration (Raballand et al.)	All SSA/MENA IPF with ICRs by 2014	TTL at appraisal rarely same as at closure	Average implementation 76 months vs 48-month planned duration; TTL who designed project 'will never close it'
Sample of 200 projects — appraisal vs implementation TTL (Springer 2022)	200 projects across regions	Majority change from appraisal to implementation	Average project length ~500 days shorter when same TTL from appraisal to implementation; continuity matters for performance
DPF operations — Chad 2016–2018	Fiscal Consolidation DPO (one operation)	5 TTLs in 2 years; 2 served less than 12 months	'Lack of consistent leadership not conducive to maintaining strong policy dialogue'

The significance of the Chad data is not that Chad is unusually difficult — though it is. It is that 168 TTLs managed 44 investment projects across the evaluation period. When a project in Chad closes with below-satisfactory outcomes, the ICR is prepared by a closing TTL who likely supervised the project for 1–3 years of its 5–8 year lifecycle. The original design TTL may have left the Bank. There may have been four intermediate TTLs between design and closure. The accountability trail is cold before the ICR is drafted.

The Raballand analysis documented in the Staff Incentives section of this report makes the structural implication explicit: "The TTL at appraisal will never close a project — different incentives at entry and at closing." This is not an incidental observation. It describes the central feature of the Bank's project accountability architecture. The person with the most information about why a project was designed as it was — and therefore the person best positioned to assess whether poor outcomes reflect design failure — is the person with the least accountability at closure.

10. The Empirical Evidence: Country Context Is Not the Primary Explanation

The institutional narrative of host-government blame has an empirical problem: the data contradicts it. The most comprehensive study of World Bank project outcomes — Denizer, Kaufmann, and Kraay (2013), using over 6,000 World Bank projects evaluated between 1983 and 2011, conducted by World Bank economists — produced a finding that directly challenges the default attribution.

> "Country-level macro measures of the quality of policies and institutions are strongly correlated with project outcomes [...] However, a striking feature of the data is that the success of individual development projects varies much more within countries than it does between countries [...] country-level factors account for only 20 percent of the total variation in project outcomes. The remaining 80 percent of the variation occurs within countries across projects." — Denizer, Kaufmann and Kraay (2013), World Bank Policy Research Working Paper 5646.

The implications of this finding are profound and directly contradict the institutional narrative deployed in ICRs. If country-level factors — government quality, institutional capacity, political stability — explained the majority of project failure, we would expect outcomes to be relatively consistent within countries and to vary primarily between countries. The data shows the opposite. Within any given country, some Bank projects succeed and others fail. The variation is primarily across projects within countries, which means it is primarily explained by project-level factors, including design quality, supervision intensity, and the quality of the task team leader.

Denizer et al. document this directly: "task team leader fixed effects are of comparable importance to country fixed effects in accounting for the variation in project outcomes." The quality of the individual TTL — measured by their average outcome rating across all their other projects — is a highly significant predictor of any given project's outcome. This means that the assignment of TTL quality to a project is, on average, as important a determinant of outcome as the governance and institutional quality of the host country.

This finding has not altered the World Bank's ICR attribution culture. Despite being published in a World Bank working paper series and subsequently in a major peer-reviewed journal, its implication — that the Bank's own people are as important as the country context in determining outcomes — has not been operationalised in evaluation practice. ICRs continue to foreground country context as the primary explanatory factor for failure. The Denizer-Kaufmann-Kraay baseline remains an academic finding rather than an institutional standard.

10.1 The 20/80 Finding and the Blame Asymmetry

The 20/80 finding creates a direct accountability asymmetry. The World Bank attributes the majority of poor project outcomes to the 20% of variation explained by country factors, while systematically understating the 80% explained by within-country project variation — which includes the quality of Bank design, supervision, and TTL performance. This inversion has predictable institutional consequences:

- TTLs learn to foreground country context in ICR narratives as the primary explanatory variable for failure, regardless of whether the evidence supports this attribution.
- Career incentives are structured to reward narrative construction that diffuses blame, rather than candid self-assessment of design and supervision quality.
- The Bank systematically underinvests in the accountability infrastructure — design quality review, supervision intensity, TTL performance tracking — that would make individual-level accountability possible, because the institutional incentive is to externalise failure rather than internalise it.

The result is that the institution most responsible for the 80% — the Bank — holds accountable the institution responsible for the 20% — the borrower government. This is not simply an analytical error. It is a governance arrangement that serves the Bank's institutional interests and the career interests

of its staff, at the cost of honest assessment of where the primary accountability for poor outcomes lies.

11. The Language of Evasion: How ICRs Structure Blame

ICR language conventions are not random — they reflect the institutional incentives shaping who writes the document and what they need to say. Several recurring ICR formulations systematically displace accountability from the Bank and from individual staff members to contextual and institutional factors outside anyone's specific control.

"Implementation was affected by the challenging operating environment"

This formulation is institutionally appropriate when genuine disruption occurs — war, pandemic, natural disaster. It becomes evasive when applied to predictable challenges that were either foreseeable at design, flagged in early ISRs, and not addressed, or are simply a restatement of the country's pre-existing structural conditions that were present when the project was approved.

"Government commitment and ownership were insufficient"

Government ownership is a legitimate constraint. It is also one that the Bank's own operational guidelines identify as a primary assessment criterion at design. If government ownership was insufficient to deliver project objectives, the question the ICR should answer — but routinely does not — is why the project was designed with objectives that required ownership the government did not have, and why this was not identified at Quality at Entry.

"The project would have benefited from a longer implementation period"

This formulation appears in ICRs where the project simply failed to deliver in the time allotted. It implicitly attributes failure to a planning decision — the project duration was too short — rather than to the quality of implementation or the achievability of objectives. It obscures the question of whether the project was over-ambitious relative to the context, which is a design failure.

"Key lessons for future operations include..."

The lessons section is the ICR's accountability release valve. By framing failures as lessons for future operations, the ICR converts accountability for what happened into advice about what should be done differently — by someone else, in some future project, to be managed by some future TTL. The lesson format is explicitly forward-looking, which means it is explicitly non-attributive. Nobody is responsible for the failure; someone should learn from it.

> IEG's own 2016 Behind the Mirror evaluation found that lessons in ICRs "rarely turn into revised policies, guidelines, or procedures" and that "individual authors of self-evaluation reports often learn something from visiting the project and writing up their analysis — but little knowledge flows beyond the authors themselves." The lesson format serves accountability purposes (it demonstrates reflection) while producing virtually no institutional learning (the lessons are not acted upon). It is the perfect instrument for an institution that needs to appear accountable without being so.

11. Structural Reforms to Create Genuine Individual Accountability

The five reforms below address the specific accountability failure mechanisms documented in this section and in the preceding sections of this annex. They are designed to be mutually reinforcing: each reform closes a gap that the others would leave open if implemented in isolation.

Reform	Mechanism	What It Fixes
Project accountability passport	A structured record attached to every project — updated at each TTL transition — documenting the design decisions taken, by whom, at what stage, with what rationale. The passport follows the project throughout its lifecycle. At ICR, the closing TTL and IEG reviewer can trace specific outcome failures to specific decision points and the individuals responsible.	Eliminates the ability of subsequent TTLs to simply inherit the project narrative without attributing specific failures to specific decisions and decision-makers. Makes the design-outcome link visible and traceable. Creates a factual record that cannot be managed by ICR framing.
Formal handover protocol with accountability transfer	When a TTL transition occurs, the incoming TTL must formally document their assessment of: project design quality, whether objectives remain achievable, adequacy of M&E framework, and quality of supervision to date. This assessment is shared with the Practice Manager and filed in the project record. If the incoming TTL believes the project is not on track to achieve its objectives, they must flag this at transition, not only at closure.	Breaks the silent inheritance pattern. Creates a contemporaneous record of each TTL's assessment at the moment they take over. Makes it impossible for a closing TTL to claim, at ICR, that problems only became apparent late in implementation, when the handover record shows they were flagged earlier.
Attribution-disaggregated ICR reporting	ICR outcome assessment should formally disaggregate causes of underperformance into: (a) design quality at entry — attributable to appraisal TTL and team; (b) supervision quality during implementation — attributable to each TTL by phase; (c) borrower and government performance — attributable to the client; (d) external factors — political, macroeconomic, force majeure. Each factor rated separately. IEG ICRR validates the attribution, not just the outcome rating.	Forces explicit apportionment of responsibility. Makes it impossible to simply attribute failure to government performance without separately accounting for Bank performance at design and during supervision. The 20/80 finding (Denizer et al.: country factors explain 20% of variation, project factors 80%) becomes a reference point — ICRs that attribute 80%+ of failure to country factors would require specific justification against the empirical baseline.
Mandatory Bank performance rating in every ICRR	IEG currently rates Bank performance (Quality at Entry and Quality of Supervision) in ICRRs for 100% of operations. These ratings should be formally incorporated into Practice Manager performance reviews and Global Practice portfolio assessments. A Practice where Bank performance ratings are consistently Moderately Unsatisfactory or below should be required to present an improvement plan to Senior Management. Individual TTLs with multiple Bank performance ratings below Moderately Satisfactory should have this reflected in their annual OPE.	Closes the loop between evaluation findings and career consequences. Currently Bank performance ratings are produced but have no institutional consequences. Linking them to individual and practice-level performance assessments makes the evaluation machinery consequential, not advisory.

Reform	Mechanism	What It Fixes
Country attribution standard: require reference to Denizer-Kaufmann-Kraay baseline	ICRs and ICRRs that attribute failure primarily to country-level factors should be required to reference the empirical baseline: that country-level factors explain approximately 20% of the variation in World Bank project outcomes, while 80% varies across projects within countries. Any ICR attributing more than 40% of outcome failure to country factors should require senior management endorsement and a specific explanation of why the Bank's own design and supervision are not implicated.	Directly counteracts the 'host government blame' default. Forces authors to engage with the empirical evidence that project outcomes are primarily determined by project-level and Bank-level factors, not country context alone. Replaces qualitative narrative with reference to quantitative evidence.

11.1 The Deeper Requirement: Continuous Accountability Across the Project Lifecycle

All five reforms above address accountability for what happened after the fact — at handover, at ICR, at closure. They are necessary but not sufficient. The deeper problem is that no accountability mechanism exists during the project lifecycle that links real-time performance to individual consequences.

A project that is deteriorating — flagged as 'moderately unsatisfactory' in ISRs, missing DLIs, receiving poor supervision scores — continues under the same Practice Manager who approved the design, with the same TTL rotation system that makes individual accountability impossible, and with the same Board approval that locked in the design and disbursement schedule at appraisal. Nothing stops the deterioration generating consequences until the ICR is written, at which point it is too late and the responsible parties have moved on.

What is required is an At-Risk Portfolio Review — a formal quarterly mechanism in which Practice Managers present deteriorating projects to Senior Management, with explicit attribution: why is this project failing, what design or supervision decisions contributed, and what is the accountability consequence. This converts the ISR system from a passive monitoring record into an active accountability trigger. It does not wait for the project to close. It surfaces accountability while the project can still be salvaged, and before the blame loop has time to form.

11.2 The Borrower Accountability Asymmetry Must Be Addressed

IDA21's Sustainable Development Finance Policy extends and strengthens the framework under which borrower governments are held accountable for development outcomes. This is appropriate. Governments that borrow public money to deliver public services should be accountable for results.

But the SDFP framework creates an accountability asymmetry that is difficult to justify in light of the Denizer-Kaufmann-Kraay evidence. If country-level factors explain only 20% of project outcome variation, a framework that holds the country's government accountable for outcomes while not holding the Bank's design and supervision teams accountable for the 80% of variation they influence is not a governance framework for development effectiveness. It is a framework that deflects accountability toward the weaker party in the financing relationship.

A symmetric accountability framework would include: SDFP-style performance monitoring for borrowers (which exists), alongside Bank Performance Ratings published at country and Practice level (which exist but have no consequences), linked to a formal Bank Performance Review in which

the Bank's own management is required to account to the Board for patterns of poor Bank performance, with improvement plans and consequences for Practice Managers whose portfolios consistently underperform (which does not exist).

IDA21 introduced borrower accountability without introducing Bank staff accountability. The replenishment commitment to results cannot be credibly delivered by a framework that enforces accountability on one party to the development relationship and ignores it on the other.

References and Sources

IEG (2016). Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group. Washington DC: World Bank. https://ieg.worldbankgroup.org/evaluations/roses

IEG (2015). External Review of the Independent Evaluation Group of the World Bank Group. Washington DC: World Bank.

IEG (2020, 2021, 2022, 2023, 2024). Results and Performance of the World Bank Group (RAP). Washington DC: World Bank.

Evaluation Cooperation Group (2010). Good Practice Standards on Independence of Central Evaluation Departments of International Finance Institutions. ECG. https://www.ecgnet.org/content/independence-ceds

Raballand, G., Roundell, K., and Mallberg, F. (2015). Bank Staff Incentives and Behaviors: Lessons from Public Sector Governance Projects. World Bank presentation, 2015.

Asian Development Bank (2021). ADB Evaluation Principles. Manila: ADB. https://www.adb.org/sites/default/files/page/612241/adb-evaluation-principles-2021.pdf

African Development Bank / IDEV. Project Completion Report Evaluation Notes and Annual PCREN Synthesis Reports. Abidjan: AfDB. https://idev.afdb.org

European Bank for Reconstruction and Development / EvD (2020). Project Self-Evaluation in EBRD: Discussion Paper. London: EBRD.

Kirk Report (2019). Independent External Evaluation of EBRD's Evaluation System. London: EBRD Board of Directors.

Inter-American Development Bank / OVE. Evaluation methodology and independence framework. Washington DC: IDB. https://publications.iadb.org

OECD-DAC (2019). Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use. Paris: OECD.

IEG ICRR Guidance Manual (2024). Guidance Manual for Independent Evaluation Group Validators: ICR Reviews for Investment Project Financing. Washington DC: World Bank.

Denizer, C., Kaufmann, D., and Kraay, A. (2013). Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance. Journal of Development Economics, 105: 288–302. [World Bank Policy Research Working Paper 5646, 2011.]

Bulman, D., Kolkma, W., and Kraay, A. (2015). Good Countries or Good Projects? Comparing Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance. World Bank Policy Research Working Paper 7245.

IEG (2021). Chad Country Programme Evaluation. Washington DC: World Bank. https://ieg.worldbankgroup.org/evaluations/chad-country-program-evaluation

Raballand, G., Roundell, K., and Mallberg, F. (2015). Bank Staff Incentives and Behaviors: Lessons from Public Sector Governance Projects. World Bank presentation.

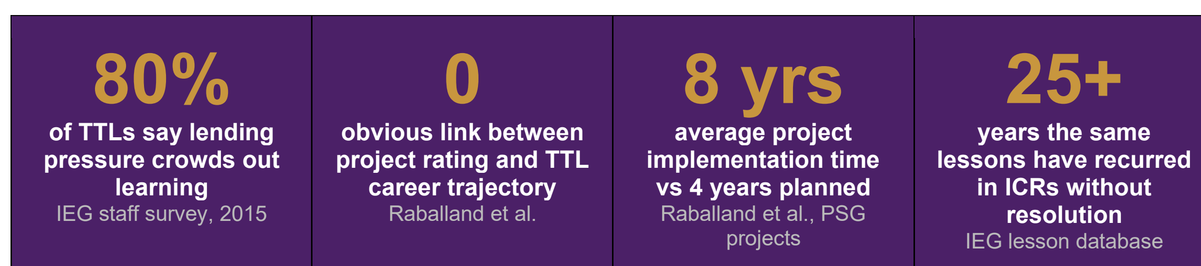
Springer/Review of International Organizations (2022). Managing Performance and Winning Trust: How World Bank Staff Shape Recipient Performance.

IEG (2016). Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group. Washington DC: World Bank.

ANNEX I — Staff Incentives and Institutional Behaviour: Raballand et al.

The Accountability Gap: Staff Incentives, Career Trajectories, and the Failure to Learn from Poor Outcomes

This annex draws on: Raballand, Roundell, and Mallberg, 'Bank Staff Incentives and Public Sector Reform Projects' (presentation, World Bank, Washington DC); IEG evaluations of staff incentives and learning (2015, 2023, 2024); Center for Global Development analysis; and peer-reviewed research on World Bank staff behaviour. The central question is why the same lessons appear in evaluation reports across twenty-five years without being acted upon.



1. The Raballand Analysis: Key Findings

Gaël Raballand, Thomas Roundell, and Michel Mallberg examined all World Bank public sector and governance investment lending projects undertaken since 2000 in Sub-Saharan Africa and MENA for which ICRs were completed by 2014. Their analysis focused on a question that institutional evaluations consistently avoid: not what went wrong in projects, but what incentives explain why the same things keep going wrong.

Their methodology used ICR documents — the completion reports prepared by task teams and reviewed by IEG — as a dataset for coding patterns in project design, failure modes, and the treatment of staff incentive questions. They supplemented this with analysis of Task Team Leader (TTL) career trajectories following project closure.

The core findings are worth stating directly:

> *"There appears to be little direct correlation between a poorly performing project and task leader career trajectory."*
> — Raballand, Roundell, and Mallberg

> *"Over-ambition and neglect of incentives and behavioral biases in projects may be possible as long as task leaders are not accountable for project designs and outcomes. The TTL at appraisal will never close a project — different incentives at entry and at closing."*
> — Raballand, Roundell, and Mallberg

> *"Strong incentives for behaviors which reinforce problems of the fallacy of planning and the optimism bias, creating limited learning and improvements to avoid the mistakes of the past."*
> — Raballand, Roundell, and Mallberg

These findings are presented in a 2015 working paper — itself prepared, with apparent irony, for the World Development Report on behavioural biases. They describe an institution whose staff know the lessons (the same five problems recur in virtually every ICR: lack of political will, lack of flexibility, lack of local knowledge, over-ambition, lack of data) but whose career incentive structure provides no mechanism for translating that knowledge into changed behaviour.

2. The Structural Accountability Gap

2.1 The TTL Discontinuity

The most important structural finding from the Raballand analysis is the TTL discontinuity: the staff member responsible for project design at appraisal is almost never the staff member responsible at project closure. The average implementation period in the PSG dataset is 76 months — over six years — against a four-year planned duration. Staff rotate, move to other projects, change regions.

This creates a structural disconnection between the incentives that shape design decisions and the incentives that would operate if accountability for outcomes existed. The TTL who designs an over-ambitious project with an unrealistic timeline, weak conditionality, and inadequate M&E does not bear the reputational cost of those design decisions at closure. A different person — or no specific person — absorbs the accountability.

> **THE TTL DISCONTINUITY**
> The person who designs a project and the person who closes it are almost never the same. In a system where career consequences attach to project outcomes, this would create accountability gaps. In the World Bank system, where career consequences do not attach to project outcomes, the TTL discontinuity is simply irrelevant — there are no consequences at either end.

2.2 What Careers Actually Reward

The IEG Learning and Results evaluation (2015) found that the World Bank's Overall Performance Evaluation (OPE) — the annual staff review that determines salary increases — does not reinforce learning from project outcomes. IEG's own survey found that for 97 per cent of Bank staff, the Board is not among the three biggest problems for learning and knowledge.

What the performance evaluation system does appear to reward:

- Lending approvals: the Board approval of a new operation is the clearest signal of productivity. Management attention concentrates on reviews prior to approval — the moment at which 'actual results can be evaluated' receives no equivalent management or Board space (Center for Global Development, 2024).

- Disbursement and portfolio management: keeping projects moving through the pipeline, managing restructuring without escalating to Board, maintaining country programme relationships.
- Avoiding reputational risk: 'staff members and managers are aware that even a small error in a transaction that results in a reputational problem can permanently harm a professional career, reinforcing a compliance mindset' (CGDev, 2024).
- Technical credibility within Global Practice: publication records, knowledge sharing, participation in learning weeks. These are individually valuable activities that exist independently of project outcome quality.

What the system does not appear to reward:

- Achieving S+ project outcome ratings as a predictor of career advancement.
- Accurate forecasting at appraisal — projects that are designed conservatively and deliver against conservative targets are not rewarded above projects that are designed ambitiously and fail.
- Learning from failures: 'the annual OPE does not reinforce learning' (IEG, 2015). 80 per cent of TTLs in IEG's survey reported that lending pressure crowds out learning.

2.3 The Optimism Bias and Why It Persists

Raballand identifies the optimism bias and planning fallacy as central behavioural mechanisms. Projects are consistently designed with timelines (4 years) that are half the actual implementation period (8 years). Objectives are set at a level that exceeds what country capacity can achieve. M&E frameworks are designed to measure what was planned rather than what is achievable.

The behavioural literature on why these biases persist in institutional settings is well developed. In an institution where:

- Ambitious project design signals engagement and seriousness at approval
- Project failure at closure does not attach to the TTL who designed the project
- Conservative design that succeeds is not rewarded above ambitious design that fails
- The institutional culture around project preparation concentrates attention on getting approval, not on designing for implementation

— the optimism bias is not a psychological glitch to be corrected by training. It is a rational response to the incentive environment. Staff behave in ways that are rewarded. Over-ambition at design is rewarded (it signals capability, secures lending). Under-performance at closure is not penalised (the responsible party has moved on).

> **THE STRUCTURAL LOGIC**
> The optimism bias at World Bank project design is not irrational. It is the rational response to an incentive structure that rewards ambitious lending approvals and imposes no career consequences for project failure. Addressing it through training or guidelines — as the Bank has repeatedly attempted — treats a structural problem as a behavioural one.

3. Corroborating Evidence: What Other Analyses Find

3.1 IEG's Own Evaluations

IEG's evaluations of the Bank's learning systems are remarkably consistent in their findings, and remarkably consistent in not being acted upon:

- Learning and Results in World Bank Operations (2014): 'The Bank holds staff accountable for project design and disbursements more than for fostering knowledge use, engaging clients in learning, and achieving influence.' The OPE 'has in the past given insufficient weight to learning and knowledge sharing.'
- World Bank Group Outcome Orientation (2023): 'Incentives emerge from what is measured, tracked, and rated in managerial dashboards, regional vice presidents' and country directors' performance agreements, and Corporate Scorecards. These signals and incentives are not always aligned with outcome orientation.'
- Learning in World Bank Lending (2024): 'The World Bank holds staff accountable for project design and disbursements more than for fostering knowledge use, engaging clients in learning, and achieving influence.'

The same finding, restated across three separate evaluations spanning a decade. The Bank's evaluation function documents the problem repeatedly. The management response to each evaluation rates the recommendation 'Moderately Satisfactory' and identifies actions that address process (new guidance notes, training programmes, knowledge sharing incentives) without addressing structure (career consequences for project outcomes).

3.2 Center for Global Development (2024)

The CGDev analysis of the Bank's incentive structure under President Banga is particularly sharp on the governance dimension: 'Key governance decision-making procedures are centred on lending approvals. The Board must approve all lending operations and devote countless time to review their merits. At the implementation or completion stages, where actual results can be evaluated, no equivalent management or Board space is provided. A Board playing a more strategic role and devoting more attention to outcome accountability will send a completely different message.'

The analysis identifies a specific mechanism: 'Management attention concentrates on multiple and exhausting reviews prior to approval to minimise criticism.' The institutional machinery for preparing a project for Board approval is elaborate, competitive, and high-stakes. The institutional machinery for evaluating what that project actually produced is the ICR — a document prepared by the task team (with obvious incentive to minimise apparent failure) and reviewed by IEG, which has no mechanism for attaching consequences to findings.

3.3 Academic Research on Staff Behaviour

Peer-reviewed research (Bayerlein et al., The Review of International Organizations, 2022) finds that World Bank staff do influence recipient performance — but primarily through supervision quality and network activation. The implication is that supervision investment is a significant lever for project outcomes — which is consistent with the IEG QoS data showing that supervision quality is a strong predictor of project outcomes across IPF.

The Gino-Staats (2015) findings cited in the Raballand presentation — bias toward success (fear of failure and overreliance on past performance) and bias toward fitting in (believing conformity is necessary) — are institutional dynamics that the academic literature consistently associates with organisations that do not reward failure disclosure. The Bank's 80 per cent TTL figure on lending pressure crowding out learning is consistent with an organisation where the dominant signal is to prioritise approvals.

> *"If President Banga is serious about outcomes, he will need to commit to a specific number of outcomes for which the institution will be held accountable over the course of the next five years. The World Bank's scorecard tracks high-level development outcomes and outputs from operations; no commitments are made to outcome targets."*
> — Center for Global Development, 2024

4. The IDA21 Connection

The staff incentives problem is directly relevant to IDA21 for three reasons:

- IDA21's $100 billion replenishment is the largest in IDA's history. It carries an elaborate results matrix with hundreds of commitments. The institutional machinery for approving and committing those $100 billion is in place — it is what the Bank is optimised to do. The institutional machinery for ensuring those commitments translate into development outcomes is the same machinery that Raballand, IEG, and CGDev identify as inadequate.
- The FCV Envelope deploys resources to the contexts where institutional accountability is weakest — FCS countries with high TTL turnover, difficult supervision environments, and the longest distance between appraisal and closure. These are the contexts where the TTL discontinuity is most severe and where the absence of outcome accountability is most consequential.
- IDA21 introduces new performance incentives for borrowers (the Sustainable Development Finance Policy). It introduces no equivalent performance accountability for Bank staff. The asymmetry — holding recipient governments accountable for outcomes while not holding Bank staff accountable for the quality of the operations through which those outcomes are to be achieved — is a structural feature of the IDA architecture that IDA21 does not address.

> **THE ACCOUNTABILITY ASYMMETRY IN IDA21**
> IDA21 strengthens performance accountability for borrowing countries through the SDFP and other mechanisms. It does not strengthen performance accountability for Bank staff. The Raballand analysis, confirmed by three IEG evaluations and independent research, indicates that Bank staff incentives are a significant independent variable in project outcomes. Strengthening borrower accountability without addressing Bank incentive structures is a partial solution to a structural problem.

6. What Would Address the Problem

The Raballand analysis and the corroborating literature converge on what structural changes would be required — as distinct from the process changes (new guidance notes, training programmes) that have been repeatedly implemented without effect:

- Outcome accountability in performance evaluation: project outcome ratings at closure — particularly patterns of below-S+ ratings on operations designed by a specific TTL — should be legible in performance evaluation. This does not require mechanical punishment; it requires making the connection visible.

- TTL continuity incentives: where a TTL is responsible for project design, mechanisms that maintain some form of ongoing accountability through the implementation period would reduce the design-closure discontinuity. This is a staffing and workload management question, not a purely policy one.
- Board-level outcome review: rebalancing Board time from approval review to outcome accountability — scheduled reviews of project performance that carry the same institutional weight as approval meetings — would send the institutional signal that the CGDev analysis identifies as absent.
- Honest ICR incentives: the current ICR is prepared by the task team that implemented the project. IEG's review is the check on that self-assessment. Strengthening IEG's independence and the consequences of divergence between ICR and ICRR ratings would improve the quality of the institutional learning record — which is currently a document that reflects task team incentives to minimise apparent failure.
- DLI ambition standards in PforR: the one instrument where outcome accountability is structurally embedded is PforR. Expanding PforR deployment, raising DLI ambition standards, and mandating independent verification across other instrument types would extend PforR's structural advantages more broadly.

Sources: Raballand, Roundell, and Mallberg, 'Bank Staff Incentives and Public Sector Reform Projects' (World Bank, 2015); IEG, 'Learning and Results in World Bank Operations' (2014, 2015); IEG, 'World Bank Group Outcome Orientation' (2023); IEG, 'Learning in World Bank Lending' (2024); Center for Global Development, 'Changing the World Bank's Incentives from Inputs to Outcomes' (2024); Bayerlein et al., 'Managing Performance and Winning Trust', Review of International Organizations (2022); Gino and Staats, 'Why Organizations Don't Learn', Harvard Business Review (2015).

ANNEX J — Advisory Services and Analytics (2020 to 2025) – The Case for Periodic Independent Review

Advisory Services and Analytics (ASA), 2020–2025: Scale, Costs, Oversight, and the Case for Periodic Independent Review

ASA delivery between FY2020 and FY2025 has consistently run into the high hundreds and, in several years, exceeded one thousand tasks annually when Economic and Sector Work (ESW) and Technical Assistance (TA) are included. Drawing on the Results and Performance of the World Bank Group (RAP) series published by the Independent Evaluation Group (IEG), ASA represents one of the largest non-lending budget envelopes in the institution, with annual expenditures in the range of approximately US$0.9–1.1 billion. Over FY2020–FY2025, cumulative ASA spending is therefore on the order of US$5.5–6.0 billion. A significant share of this total is externally financed through trust funds, meaning a portion of the portfolio is oriented toward donor rather than Bank-determined strategic priorities — a structural feature that compounds the alignment and oversight challenges documented below.[1]

The table below consolidates reported ASA volumes and expenditure envelopes over the period. Exact counts vary modestly depending on internal classification, but the order of magnitude is stable and consistent with RAP reporting and budget documents.

Fiscal Year	Estimated ASA Tasks Delivered	Estimated ASA Expenditure (US$ billion)
FY2020	~1,050–1,150	~0.95–1.0
FY2021	~1,100–1,200	~1.0
FY2022	~1,150–1,250	~1.0–1.05
FY2023	~1,150–1,300	~1.05
FY2024	~1,200–1,300	~1.05–1.1
FY2025 (est.)	~1,200+	~1.0–1.1

What IEG Has Found: Use, Influence, and the Monitoring Gap

IEG's most comprehensive recent assessment of ASA effectiveness is the RAP 2022, which conducted the first systematic examination of ASA use and influence at the country programme level, analysing Completion and Learning Review (CLR) Reviews for a random sample of 50 countries over the FY2013–22 period. Its core findings are clear:[2]

- ▶ Topic relevance is not the problem. More than 80 percent of ASA products matched government policy priorities and Country Partnership Framework objectives. The Bank is producing relevant analysis.

- ▶ Influence is the gap. Less than half of CLR Reviews reported on higher-level ASA influence — defined as documented impact on policy dialogue, uptake in government programmes, or integration into CPF objectives and Bank operations.

> ▶ Coverage is the structural failure. CLR Reviews reported on average on only one-third of the ASA programme in terms of documented use or influence. CLRs do not appear to report on ASA that is not used.

The monitoring architecture is the root cause. The World Bank systematically self-evaluates every lending project through Implementation Completion and Results Reports, which IEG then independently validates. ASA has no equivalent: there is **no systematic self-evaluation, no IEG validation of outputs, and no portfolio-level quality review.** Activity Completion Summaries exist but are neither independently validated nor designed to capture use and influence. Download metrics from the Open Knowledge Repository are available at the item level but are not integrated into Global Practice performance dashboards, not linked to cost metrics, and not used to inform resource allocation decisions. There is currently no institutional mechanism equivalent to a lending portfolio review that periodically assesses ASA outputs as a consolidated body of work. IEG concluded directly that, without a deliberate focus on use and influence, the World Bank risks becoming a "Report Bank" rather than a "Knowledge Bank."[2]

An important qualification applies. Where CLR Reviews did report on higher-level influence, in the majority of cases ASA was found to have been genuinely influential — Kenya, Rwanda, Argentina, and Poland all showed strong ASA integration with lending and policy reform. The problem is systematic under-documentation and the absence of institutional infrastructure to capture evidence of use, not evidence of universal failure.

Historical Precedent and the Persistence of the Pattern

A dedicated IEG evaluation of World Bank research covering 1998–2005 reviewed nearly 4,000 research outputs and found that too large a share was "undistinguished," lacking either strong policy relevance or academic distinction. Quality was uneven, monitoring was infrequent, and the absence of systematic portfolio-wide review allowed weaker products to persist. The evaluation recommended sharper selectivity, stronger quality control, and more regular independent assessment. Managerial overstretch was also identified: as flagship studies proliferated, quality control mechanisms did not scale proportionately, and incentives favoured production volume over demonstrated use.[3]

> *The scale of ASA now substantially exceeds, in financial terms, the research portfolio examined in 2006. The RAP 2022 findings confirm that the structural dynamics identified then — volume growth without matched oversight, diffuse accountability, incentives oriented toward production rather than use — have persisted as an institutional pattern.*

The FY26 Reorganisation: Institutional Acknowledgement and Its Limits

In late 2025, the World Bank Group announced a major reorganisation structured explicitly around a "Knowledge Bank" model, establishing a Knowledge and Learning Department and an Outcomes Department. The stated objective is to strengthen the operational use of knowledge across the institution. IEG's assessment is cautionary. This is the Bank's third major knowledge reorganisation in a decade: the 2014 restructuring created joint departments that were subsequently disbanded; the 2018 restructuring partly reversed that design. IEG's evaluation of the 2014 reforms found that they failed because culture and incentives were left unaddressed.[4]

> *The FY26 reorganisation strengthens rather than resolves the case for periodic independent review. A new departmental architecture that aspires to supercharge knowledge production requires external accountability commensurate with its ambition — and with the resources committed.*

The Case for Periodic Independent Review

Given cumulative expenditures exceeding US$5 billion over five years, the institutional case for systematic review is compelling. ASA should be subject to periodic, independent portfolio assessment every five years, analogous in rigour to lending portfolio reviews. Such a review should examine quality distribution, evidence of policy uptake, alignment with corporate priorities, timeliness, and cost-effectiveness. It should draw not only on self-reported Completion and Learning Reviews, but also on independently verifiable metrics, including usage analytics, citation patterns, trust fund alignment, and evidence of incorporation into lending operations and policy reform processes.

In addition, ASA would benefit from the introduction of a structured ratings architecture analogous to that applied to investment projects. IEG currently assigns outcome ratings, risk-to-development-outcome ratings, and quality-at-entry assessments to lending operations. A comparable framework could be developed for ASA, with ratings for relevance, technical quality, timeliness, and demonstrated uptake, assigned by IEG on a sample basis annually, with periodic deep-dive portfolio reviews conducted by external panels of independent experts.

Regular external supervision would serve three purposes. First, it would strengthen incentives for selectivity and quality at entry. Second, it would introduce a transparent feedback loop linking knowledge production to measurable engagement and operational use. Third, it would enhance Board oversight by providing consolidated evidence on whether large non-lending budget allocations are generating proportional development value.

IEG has designated World Bank ASA as the deep-dive topic for RAP 2025, which will examine over 5,700 country-specific ASA products from FY2017–24 — the first comprehensive independent evaluation at this scale since 2006. RAP 2025 is being conducted in consultation with the newly established Knowledge and Learning Department and Outcomes Department, and its commissioning validates the proposition that the current oversight architecture is insufficient. The objective is not to constrain knowledge production, but to align scale with accountability. When annual expenditures approach US$1 billion, when cumulative five-year spending exceeds US$5 billion, and when IEG's own evaluations confirm that oversight is self-reported and fragmented, reliance on decentralised monitoring is institutionally insufficient. Periodic independent review — supported by a ratings architecture, external expert panels, and independently verifiable use metrics — would bring ASA oversight into alignment with the scale of resources committed and the ambition of the institution's knowledge mandate.[5]

Sources

1. Results and Performance of the World Bank Group (RAP) series, Independent Evaluation Group, World Bank, FY2020–FY2025 editions; World Bank administrative budget documents, FY2020–FY2025. Trust fund composition: World Bank, Realizing the World Bank Group's Knowledge Potential for Effective Development Solutions: A Strategic Framework (2021).
2. Independent Evaluation Group. Results and Performance of the World Bank Group 2022 (RAP 2022). Washington, DC: World Bank, 2022. Chapter 6: Advisory Services and Analytics in Country Programs. See also IEG Blog, "World Bank — a 'Knowledge Bank' or a 'Report Bank'?" 2022.
3. Independent Evaluation Group. An Evaluation of World Bank Research, 1998–2005 (Report No. 86003). Washington, DC: World Bank, 2006.
4. Independent Evaluation Group. "The Knowledge Bank: Five Recommendations for Getting the World Bank Group Reorganization Right." IEG Blog (Rasmus Heltberg), December 10, 2025; IEG, Knowledge Flow and Collaboration Under the World Bank's New Operating Model (2023).
5. Independent Evaluation Group. Concept Note: Results and Performance of the World Bank Group 2025 (RAP 2025). Washington, DC: World Bank, September 2024. 5,720 country-specific ASA products, FY2017–24 closed as of June 2025.

ANNEX N — Board Governance

The Board Co-optation Problem: How Project-Level Approval Destroys Management Accountability

This annex examines the governance mechanism by which World Bank management has structurally insulated itself from accountability for poor development outcomes. It draws on the analysis of Ngaire Woods (University of Oxford), the 2009 Zedillo Commission Report, IEG evaluations of the Bank's results orientation, comparative governance analysis of the Asian Infrastructure Investment Bank (AIIB) and New Development Bank (NDB), and corporate governance literature on the board–management accountability relationship.

2×/wk	~300	0	60%
World Bank Board meets Approving individual projects	**Projects approved p.a.** Each requires Board sign-off	**Annual outcome reviews** Of management portfolio performance	**of projects rated below S+** For which no Board is held accountable

1. The Accountability Architecture and Its Central Flaw

The World Bank's Board of Executive Directors is a resident body. Its 25 Directors and Alternate Directors maintain offices at Bank headquarters in Washington DC, receive salaries and benefits from the institution, and participate in its day-to-day operations. The Board meets approximately twice a week. The dominant function of those meetings is the approval of individual lending operations — reviewing project documents, clearing questions with management, and formally authorising each loan.

This arrangement is a legacy of the Bank's founding architecture, designed for an era when the Bank made a small number of large loans and Board members could plausibly evaluate individual projects. Today, the Bank approves roughly 300 operations per year totalling $30–40 billion in annual commitments. The Board reviews each one. Management prepares each one. The same management whose operations the Board is supposed to oversee is also the management whose operations the Board has just approved.

THE CO-OPTATION MECHANISM

When a Board approves an individual project, it becomes a co-author of that project. It cannot subsequently hold management accountable for a project's failure without simultaneously acknowledging its own role in approving that project. Project-level approval does not strengthen oversight — it structurally prevents it. This is not a side effect of the current governance model. It is, from management's perspective, its most valuable feature.

> *“The resident board works with far too much micromanagement of lending. Missing from the World Bank is a forum for longer-term agenda-setting and genuine multilateral cooperation on development assistance.”*
>
> — Ngaire Woods, ‘From Intervention to Cooperation: Reforming the IMF and World Bank’ (2008)

Woods’s observation, repeated at the Bank’s 75th anniversary celebrations in 2019, identifies the central paradox: a Board occupied with approving hundreds of individual operations each year has no time, no institutional mandate, and no structural independence to exercise genuine strategic oversight of management performance. The two functions — approving operations and holding management accountable for results — are mutually exclusive when performed by the same body on the same operations.

2. The Zedillo Commission: ‘An Impossible Trinity’

In 2009, former Mexican President Ernesto Zedillo chaired a High-Level Commission on Modernisation of World Bank Group Governance appointed by President Zoellick. The Commission’s diagnosis of the Board is among the sharpest ever produced by an authorised external review:

> *“The Board of Executive Directors attempts an impossible trinity of roles: political representation of member countries, managerial responsibility for Bank policy, and oversight of those same policies. The result is a fundamental conflict of interest — the Board cannot pass judgment on management’s performance without also passing judgment on its own performance.”*
>
> — Zedillo Commission, ‘Repowering the World Bank for the 21st Century’ (2009)

The Commission’s specific recommendations included:

- The Board should meet ‘a few times a year, not twice weekly’ — shifting its function from operational co-management to strategic governance.
- A framework for annual performance review of the President, with clear criteria, should be established — the President has never been subject to such a review.
- The Board should be rebalanced toward developing countries (a 50/50 split was proposed) to ensure the institution is accountable to those it serves, not only to its largest donors.
- Separate boards for IBRD and IDA were proposed, reflecting the distinct interests of middle-income and low-income country shareholders.

The response from the existing Board was, predictably, to discuss the recommendations extensively and implement none of the structural ones. As one participant at the CGDev launch of the Zedillo Report observed, there had been ‘similar reports calling for governance reforms with strikingly similar recommendations that have been ignored’ stretching back to 2005. The Board does not reform itself, because the existing arrangement serves the interests of the largest shareholders — and of management.

WHY REFORM HAS NOT HAPPENED

The resident Board model serves the interests of the three parties who would need to agree to change it: (1) Management benefits from Board co-optation, which prevents accountability for results; (2) Major shareholders (US, European countries) benefit from resident representation, which gives them disproportionate influence over individual lending decisions; (3) Directors benefit from the status, salary, and influence that resident Board membership provides. Reform requires all three to agree to reduce their own power. It has not happened in fifteen years of advocacy.

3. The Mechanics of Co-optation: How It Works in Practice

The co-optation of the Board through project approval operates through several specific mechanisms that are worth making explicit:

3.1 Information Asymmetry at Approval

The Board approves projects on the basis of documentation prepared by management. The Project Appraisal Document, the Programme Paper, and supporting technical documents are management's account of why the operation is well-designed and likely to succeed. The Board has no independent analytical capacity to contest those assessments. It relies on management's self-evaluation at precisely the moment when management's institutional incentives run most strongly toward optimism.

The Quality Assurance Group (QAG), which did provide some independent quality review at preparation stage, was disbanded. The Global Practices that replaced it have not delivered equivalent quality oversight — as the IEG data on Quality at Entry (only 44% rated Satisfactory or above across the IPF portfolio) makes clear. The Board approves projects whose quality it cannot independently assess, based on documentation from the party it is supposed to oversee.

3.2 The Approval as Formal Endorsement

Board approval is not a passive administrative step. It is a formal decision by the institution's governing body that the operation is sound, that the design is appropriate, and that the commitment of funds is justified. When that operation subsequently fails to achieve its development objectives — as 60 per cent of IPF and 65 per cent of DPF operations do — the Board's formal endorsement is on the record.

This creates a structural disincentive for the Board to scrutinise IEG outcome ratings seriously. A Board that takes poor outcome ratings seriously must simultaneously acknowledge that it approved those operations. The institutional response — consistently documented in IEG's management responses — is to rate recommendations 'Moderately Satisfactory,' commit to process improvements, and avoid engaging with the structural question of why 60 per cent of the portfolio consistently fails.

3.3 The Annual Cycle Crowding Out Strategic Oversight

The Board's twice-weekly approval schedule means that its institutional attention, staff time, and relationship with management are dominated by the pre-approval pipeline. The CGDev analysis (2024) identifies this precisely: 'management attention concentrates on multiple and exhausting reviews prior to approval to minimise criticism. At the implementation or completion stages, where actual results can be evaluated, no equivalent management or Board space is provided.'

The result is an institution whose internal machinery — from preparation to Board approval — is designed to move money and secure sign-off. The machinery for evaluating what that money produced is IEG, which has no enforcement powers, no ability to attach consequences to its findings, and whose recommendations are responded to by the management it evaluates.

THE ACCOUNTABILITY VOID

The World Bank has no mechanism by which poor development outcomes produce consequences for management. IEG identifies failures. Management responds to IEG. The Board approved the operations being criticised. No party in this loop has both the information and the structural independence to hold management to account for results. This is the accountability void at the centre of the institution's governance.

4. Global Best Practice: How Modern Development Institutions Govern

The World Bank’s resident, project-approving Board is now the exception rather than the model in modern MDB governance. Three institutions — the AIIB, the NDB, and the European Investment Bank — have adopted governance architectures that separate strategic oversight from operational management in ways the World Bank has not.

4.1 The Asian Infrastructure Investment Bank (AIIB)

The AIIB, founded in 2016 with 111 member countries and $100 billion capitalisation, adopted a non-resident Board of Directors as a foundational governance principle. The distinction from the World Bank model is structural and deliberate:

World Bank Board (Resident Model)	AIIB Board (Non-Resident Model)
25 resident Directors, based in Washington DC	12 part-time, non-resident Directors
Meets 2× per week; agenda dominated by project approvals	Meets quarterly; agenda focused on strategy, policy, oversight
Approves every individual lending operation	Has delegated project approval authority to the President (75% majority required for delegation)
Board co-opted into management through operational approval	Board maintains structural separation; can hold management accountable for portfolio results
No formal annual management performance review	Audit and Risk Committee; Policy and Strategy Committee; oversight mechanism mandated in Articles of Agreement
IEG recommendations non-binding; management self-responds	Chief Evaluation, Integrity and Oversight Unit reports directly to Board, not to management

The AIIB’s founding documents explicitly required the Board to ‘establish an oversight mechanism’ with principles of transparency, openness, independence and accountability — a requirement absent

from the World Bank's original Articles of Agreement. The Board's three committees (Policy and Strategy; Budget and Human Resources; Audit and Risk) structure its engagement around governance functions rather than individual transaction review.

The delegation of project approval to the President is not a weakening of governance — it is its precondition. As Norton Rose Fulbright observed: 'the Board is able to delegate its authority in respect of the operations of the bank to the president, allowing the board to focus on strategy, policy and the parameters within which the bank must operate.' This is modern corporate governance: a board that sets strategy and holds management accountable for executing it, rather than co-managing operations with management.

> *"The non-resident Board of Directors has clearly provided powers to approve policy and operations, delegate authority and supervise management; its oversight powers are highlighted. AIIB's President has traditional powers with the express possibility of delegated authority."*
>
> — Natalie Lichtenstein, Inaugural General Counsel, AIIB; Oxford Law Blogs (2018)

4.2 The New Development Bank (NDB)

The NDB, established by the BRICS countries in 2014, also adopted a non-resident Board model with explicit attention to the incentive structure that resident Board models create. The NDB's General Strategy document (2017) states directly: 'A lean governance structure helps streamline the decision-making process. The non-resident Board of Directors reduces administrative costs and — more importantly — helps the Board focus on high-level policy issues and particularly complex projects rather than routine day-to-day operations.'

The NDB's Credit and Investment Committee, composed of the President and Vice Presidents, holds project approval authority up to limits established by the Board. The Board reviews complex or high-value operations directly; routine transactions proceed through management-level authorisation. The governance principle is explicit: the Board's time is too valuable to spend on routine approvals — it should be reserved for strategic oversight and accountability.

4.3 The European Investment Bank (EIB)

The EIB, the world's largest multilateral lender by volume, operates with a non-resident Board of Directors that does not approve individual operations. Lending decisions below certain thresholds are made by management under delegated authority. The Board sets strategy, approves the annual operational programme, monitors portfolio performance, and holds management accountable for aggregate results. It is not in the business of approving individual motorways or renewable energy projects.

The EIB model is significant because it demonstrates that delegated lending authority is compatible with the highest credit ratings (AAA), the strongest governance standards, and substantial lending volumes. The argument that project-level Board approval is necessary for fiduciary soundness or credit quality is not supported by the EIB's record.

4.4 Corporate Governance Best Practice

The principle that operational co-management compromises oversight accountability is not specific to development banks. It is a foundational principle of modern corporate governance, codified in the OECD Principles of Corporate Governance (2023) and in the UK Corporate Governance Code:

Governance Principle	Implication for MDB Board Design
Separation of governance and management	Boards set strategy and hold management accountable; they do not co-manage operations. When a board approves individual transactions, the line collapses.
Independence of oversight from operations	The oversight body must be structurally independent of what it oversees. A Board that co-approves operations cannot independently assess their outcomes.
Management accountability for results	CEOs and senior management are held accountable by boards for portfolio outcomes, not for individual transaction quality. Accountability at the aggregate level creates incentives across the whole portfolio.
Delegation with accountability	Boards delegate operational authority to management and hold management accountable for how that authority is exercised. Delegation is not abdication — it is the precondition for genuine accountability.
Time allocation to oversight	Effective boards spend their time on strategy, risk, performance review and succession — not on approving individual contracts, investments or loans.

The World Bank's governance model fails every one of these principles. Its Board is resident, operationally co-managerial, and structurally unable to hold management accountable for the results of operations it has approved. This is not a peripheral design detail — it is the central governance failure of the institution.

6. The Numbers: What Co-optation Looks Like in Practice

The consequence of the co-optation structure is visible in the IEG data. Across the period FY2000–2024, the World Bank's Board has approved every one of the 8,764 operations in the IEG evaluation database. Of those:

Instrument	Total Approved	Below S+ at Evaluation	Committed to Below-S+
IPF	7,141	4,304 (60.3%)	$357.1bn
DPF	1,551	1,000 (64.5%)	$186.4bn
PforR	72	28 (38.9%)	$6.2bn
ALL	8,764	5,332 (60.8%)	$549.7bn

The Board approved all 8,764. It held management accountable for the results of none. There is no annual review in which management is required to explain why 60 per cent of approved operations fail to achieve satisfactory development outcomes, what design changes would address the pattern, or what consequences attach to a portfolio in which the majority of lending consistently underperforms.

This is not an accountability system with modest effectiveness. It is an accountability system that does not exist. The Board's approval function has successfully substituted for its oversight function, producing an institution in which management is effectively unaccountable for the development results of its operations.

THE MANAGEMENT BENEFIT

It would be naive to treat this as an accidental governance failure. The structure benefits management directly and persistently. A management team that must secure Board approval for every operation has an incentive to invest in Board relations at the preparation and approval stage — where career rewards concentrate — rather than in implementation quality and development outcomes, where no equivalent accountability exists. The co-optation of the Board through project approval is the institutional mechanism that makes the staff incentive problem described in the Raballand analysis possible.

6.2 Cost Comparison

The structure of MDB Boards is not merely a governance choice; it has direct fiscal and operational consequences. Evidence compiled by Prizzon et al. (2022) shows that resident Boards impose significant fixed costs on institutions. In legacy MDBs, annual Board of Directors costs range from approximately US$17 million to US$84 million. At the World Bank, Board costs alone amount to US$84.4 million per year — the single largest governance-related budget line.

These are direct expenditures. They exclude indirect and opportunity costs: management time devoted to preparing Board papers; staff time responding to detailed operational queries; extended review cycles; and the cumulative institutional burden associated with frequent Board engagement in transaction-level decisions. When these indirect costs are factored in, the full economic footprint of a resident Board is materially higher than the headline budget numbers suggest.

Table 1: Costs of the Board of Governors and Board of Directors (US$ million)

Institution	Residency of BoD	BoG Cost	BoD Cost
World Bank	Yes	6.30	84.40
IADB	Yes	4.00	21.90
AfDB	Yes	9.41	21.36
AsDB	Yes	2.36	17.10
EBRD	Yes	1.56	17.91
CDB	Yes	Travel & subsistence only	25.00
CABEI	Yes	n/a	n/a
EIB	No	Not applicable	Travel & subsistence only
CAF	No	n/a	n/a
IsDB	No	n/a	6.39
AIIB	No	Travel & subsistence only	Travel & subsistence only
NDB	No	n/a	n/a

IFAD	No	Not applicable	5.04
Gavi	No	Not applicable	Travel & subsistence only
GFATM	No	Not applicable	1.60
GCF	No	Not applicable	0.33
GPE	No	Not applicable	1.95
TDB	No	Not applicable	0.81

Note: n/a = not available. Data as of June 2021.

Cost differences are reinforced by stark variation in meeting intensity and agenda volume. Resident Boards meet far more frequently and process substantially larger numbers of items, reflecting deeper engagement in operational approvals rather than high-level strategic oversight.

Table 2: Board Meetings – Number of Meetings and Items Considered (2016)

MDB	AfDB	World Bank	IADB	AsDB	EBRD	CAF	GF	AIIB
Number of meetings	41	75	37	43	25	3	2	6
Total hours met	123	137	n/r	~100	n/r	18	32–40	55
Items considered/approved	82	184	n/r	72	n/r	30	31	59
Informal meetings	8	37	n/r	48	n/r	3	2	n/a

Note: n/r = not reported; n/a = not available. Data on Board meeting frequency and workload reflects the most recent publicly available cross-MDB comparative analysis (ODI 2022, drawing on IDEV 2018), and may not capture changes in meeting practices in individual institutions since that period.

The contrast is structural rather than incidental. In 2016, the World Bank Board met 75 times and considered 184 items. By comparison, CAF met only three times and AIIB six times. High meeting frequency and large approval volumes are consistent with Boards functioning as operational gatekeepers. Lower meeting frequency in non-resident models is consistent with a governance design that emphasises delegation, risk tolerance frameworks, and strategic supervision.

Taken together, the evidence suggests that resident Boards are associated with:

- Higher fixed governance costs;
- Greater operational involvement;
- Increased management transaction burden; and
- Reduced institutional distance between Board and day-to-day execution.

Non-resident models, by contrast, demonstrate that MDBs can operate with lower governance overhead and fewer meetings while maintaining oversight — provided clear delegation frameworks and accountability mechanisms are in place.

7. What Reform Would Look Like

The reform implied by the analysis is not radical in terms of precedent — AIIB, NDB, and EIB have all implemented versions of it. It is radical in terms of the consequences it would have for the existing distribution of power and influence at the World Bank. The specific elements are:

7.1 Delegated Lending Authority

The Board should delegate project approval authority to the President for operations within policy-compliant parameters. This would follow AIIB and NDB precedent. The Board would retain approval authority for operations above specified financial thresholds, operations in new or high-risk categories, and operations that raise novel policy questions. The default for routine, policy-compliant lending would be management authorisation.

This is not a reduction in governance — it is a reallocation of the Board's finite attention from operational co-management to strategic oversight. A Board that does not spend its time approving hundreds of individual operations per year has time to evaluate management's performance on the portfolio those operations constitute.

7.2 Annual Management Performance Review

The Zedillo Commission recommended a formal annual performance review of the President, with clear criteria and Board ownership of the process. This has never been implemented. Such a review would, at minimum, include assessment of portfolio outcome performance (IEG S+ rates by region, instrument, and practice), quality-at-entry trends, and management's response to IEG findings. The review would be public.

This would create the first formal mechanism in the Bank's history by which management is held accountable for development outcomes. It would also change the signal received by management throughout the institution: portfolio results, not lending approvals, would become the primary accountability metric.

7.3 IEG Reporting Directly to the Board

Currently, IEG reports to the Board through a committee, but its recommendations are responded to by management. Management's responses to IEG recommendations are themselves rated by IEG — an arrangement that creates an appearance of accountability while generating none. A reformed structure would have IEG findings presented directly to the Board in a dedicated annual session, without management intermediation, at which the Board holds management to account for patterns of poor outcomes documented by IEG.

7.4 Moving to a Non-Resident Board

The deeper reform — consistent with Woods, Zedillo, AIIB, NDB, and EIB precedent — is moving the Board to a non-resident structure. Non-resident Directors maintain their national roles and independence from the institution's day-to-day operations. They are not embedded in the culture, relationships, and incentive structures of the Washington DC headquarters. They bring the external perspective that genuine oversight requires.

The main objection — that non-resident Directors cannot exercise adequate oversight without day-to-day presence — is answered by the performance of the AIIB and NDB, both of which have maintained AAA ratings and strong governance assessments with non-resident boards. The objection is not empirical; it is an interest-based defence of the existing arrangement by those who benefit from it.

THE REFORM LOGIC

Delegating project approval to management is not a weakening of accountability — it is its precondition. A Board that approves everything can be held accountable for nothing. A Board that delegates operations and holds management accountable for portfolio results creates, for the first time in the World Bank's history, an accountability mechanism with actual consequences. This is what AIIB, NDB, and EIB have built. It is what the World Bank needs.

8. The IDA21 Connection

The governance failure described in this annex is directly consequential for IDA21. The $100 billion replenishment commits IDA to results across hundreds of operations in the world's most challenging development environments. Those commitments are made by management to shareholders at replenishment meetings. They are executed through the lending pipeline managed by management. They are evaluated by IEG. And they are subject to oversight by a Board that co-approved every operation through which those results were supposed to be achieved.

IDA21's results matrix is, in this governance context, a monitoring framework without an accountability mechanism. Shareholders can observe whether targets are met. They cannot hold management accountable for missing them, because the Board they use to exercise that accountability has already approved the operations through which the targets were to be achieved. The same structural loop — management designs, management implements, Board approves, IEG evaluates, management responds, no consequences attach — governs IDA21's $100 billion as it governed every previous IDA replenishment.

THE IDA21 GOVERNANCE GAP

IDA21 has introduced strengthened borrower accountability (SDFP, performance triggers). It has not introduced strengthened Board oversight of management. Shareholders at the IDA21 replenishment made commitments to their publics about development results. Those commitments cannot be enforced through the existing governance structure. The co-optation of the Board through project approval means that the accountability chain from taxpayer contribution to development outcome has a structural break at the institutional level. No amount of results monitoring closes that gap without governance reform.

9. Summary

The World Bank's Board of Executive Directors approves every individual lending operation. In doing so, it becomes co-author of every operation and is structurally prevented from holding management accountable for the development results of those operations. This is the central governance mechanism that makes management unaccountable for development outcomes across a portfolio in which 60 per cent of operations consistently fail to achieve satisfactory results.

This is not an unseen problem. Ngaire Woods identified it publicly at the Bank's own 75th anniversary. The Zedillo Commission documented it formally in 2009. The AIIB, NDB, and EIB have all adopted governance models that avoid it. The World Bank has not reformed, because reform

would require the three parties who benefit from the current arrangement — management, major shareholders, and resident Directors — to agree to reduce their own operational influence.

The reform is straightforward in principle: delegate project approval to management, require an annual management performance review based on portfolio outcomes, and have IEG report directly to a Board that holds management accountable for its findings. This is modern corporate governance. It is what development institutions established in the 21st century have adopted as their baseline. It is what the World Bank should have adopted in 2009 when the Zedillo Commission recommended it, and should adopt now.

Sources: Ngaire Woods, 'From Intervention to Cooperation: Reforming the IMF and World Bank' (Policy Network, 2008); Woods, 'Unelected Government: Making the IMF and the World Bank More Accountable' (Brookings Review, 2003); Zedillo Commission, 'Repowering the World Bank for the 21st Century' (2009); Center for Global Development, 'Changing the World Bank's Incentives from Inputs to Outcomes' (2024); AIIB Articles of Agreement and Oversight Mechanism Paper (2019); NDB General Strategy 2017–2021; Natalie Lichtenstein, 'Setting Up the AIIB', Oxford Law Blogs (2018); Norton Rose Fulbright, 'AIIB's Role in Asian Infrastructure Development' (2019); Congressional Research Service, 'Asian Infrastructure Investment Bank' (2023); The Diplomat, 'A Decade of Questionable Governance at the AIIB' (2026); IEG evaluation database (8,764 operations, $974bn committed). Prizzon, A., Bains, M., Chakrabarti, S., Pudussery, J. (2022). *Governance of Multilateral Development Banks: Options for Reform.* ODI Report, September 2022.