Evidence Without Accountability

A Critique of ODI Global's 2026–2031 Strategy

How the development sector's leading think tank avoids the most important question in development finance

Parminder Brar | mdbreform.com | May 2026

I. The Plan

ODI Global's 2026–2031 strategy diagnoses a world in which multilateralism is imperilled, democratic norms are under strain, and the costs of instability fall on those with the least agency. It promises robust evidence, equitable partnerships with Global South institutions, coalition-building, and narrative-shaping across four priorities: institutions and power, equity and rights, crisis and resilience, and sustainable economic transformation.

ODI will leverage its London and Brussels offices, scale its advisory arm, and use the Fellowship Scheme's 1,300-strong alumni network. Impact will be assessed through "plausible narratives, drawing on metrics and indicators where possible, and inference where necessary."

The flagship contribution of the past strategy period is the MDB client survey — nearly 900 officials across 125 countries — and ODI's role in shaping the G20 MDB reform agenda. ODI now serves as Secretariat of the IDA technical working group ahead of the IDA21 mid-term review. Because ODI occupies such a central convening role in MDB reform debates, its analytical choices — what it measures, what it asks, and what it omits — directly shape the information environment in which replenishment and scaling decisions are made. Its silence on the issues raised below is consequential rather than incidental.

II. The Admission

In a candid 2025 keynote, ODI's Chief Executive Sara Pantuliano described inheriting an institution that faced near-bankruptcy in 2019 — a project-based, donor-driven contract research house, chasing large contracts that offered little space for innovation or independence. Staff fell from 250 to 190. The budget dropped from £40 million to £32 million. Major funding streams were shed.

Charity Commission filings confirm the restructuring. Total income fell from £34.96 million (2020–21) to £27.44 million (2021–22) before recovering to £34.81 million (2024–25). Government contracts collapsed from £4.80 million to under £200,000. ODI Global Advisory revenue dropped from £10.20 million to £3.38 million. Approximately 78 percent of ODI's income comes from charitable research funded by bilateral and multilateral donors — FCDO, EU institutions, Nordic agencies, and the MDBs themselves — with the Fellowship Scheme supported by FCDO, Australia's DFAT, the Global Fund, and WHO.

This means that for much of the previous decade, ODI's enormous output was driven less by a theory of change than by the need to meet contract deliverables. Pantuliano's admission is valuable. The question is whether the reinvention addresses the fundamental problem.

III. The Evidence Gap

The new strategy avoids the most consequential evidence question in development finance. Not because the evidence does not exist, but because the institutional incentives run against confronting it. Throughout this section, "delivery failure" refers to failure to achieve stated development objectives as assessed under the World Bank Group's own independent evaluation methodology.

The survey: what was asked and what was not

ODI's client survey is presented as "the most comprehensive, comparative assessment currently available" of how client countries view MDB performance. It covers 11 MDBs across 125 countries. Its findings have informed G20 Evolution Roadmap discussions, shareholder debates, and replenishment framing.

The survey's questionnaire architecture, however, systematically privileges questions about institutional relevance, demand, and client satisfaction over questions about whether MDB-funded projects actually achieved their development objectives. The survey asks government officials to rate MDBs on: relevance of roles and functions; perceived effectiveness; alignment with country priorities; quality of financing terms; demand for future grants and loans; quality of policy advice; coordination among MDBs; and length of the project cycle. What the survey does not ask — in any form — is whether the projects and programmes financed by those MDBs produced Satisfactory development outcomes as independently evaluated. (See Annex B for the detailed methodological critique.)

The respondent pool itself creates predictable response dynamics. The survey population consists largely of finance ministry, planning ministry, and line ministry officials who manage ongoing MDB financing relationships — actors institutionally embedded within those relationships, who negotiate replenishment access and manage active portfolios. Their incentives to report positively on relevance, alignment, and future demand are structural, not personal. The survey design does not adjust for this.

The consequence is that the survey operationalises "effectiveness" primarily as usability, responsiveness, and financing availability — not as demonstrated development outcomes. It implicitly equates stronger client demand with institutional success, without testing whether increased financing has produced successful results. In a context where IDA has doubled its lending volume while its Satisfactory outcome rate has remained flat at 31 percent for two decades, the distinction between demand and delivery is not academic. It is the central question that any serious assessment of MDB performance should address.

What the survey didn't tell respondents

The IEG Master Database — the most comprehensive longitudinal record of development project performance in existence — shows what respondents were never presented with (Brar, 2026). Using the Satisfactory-or-above (S+) threshold — the standard historically treated as the meaningful pass threshold across the Bank Group's evaluation architecture (see Annex A):

Of $177.0 billion committed to IDA-eligible countries between FY2015 and FY2026, $117.0 billion — 68.9 percent — went to projects that did not achieve

Satisfactory development outcomes. The IDA-only Satisfactory rate has been flat at 31 percent for two consecutive decades.

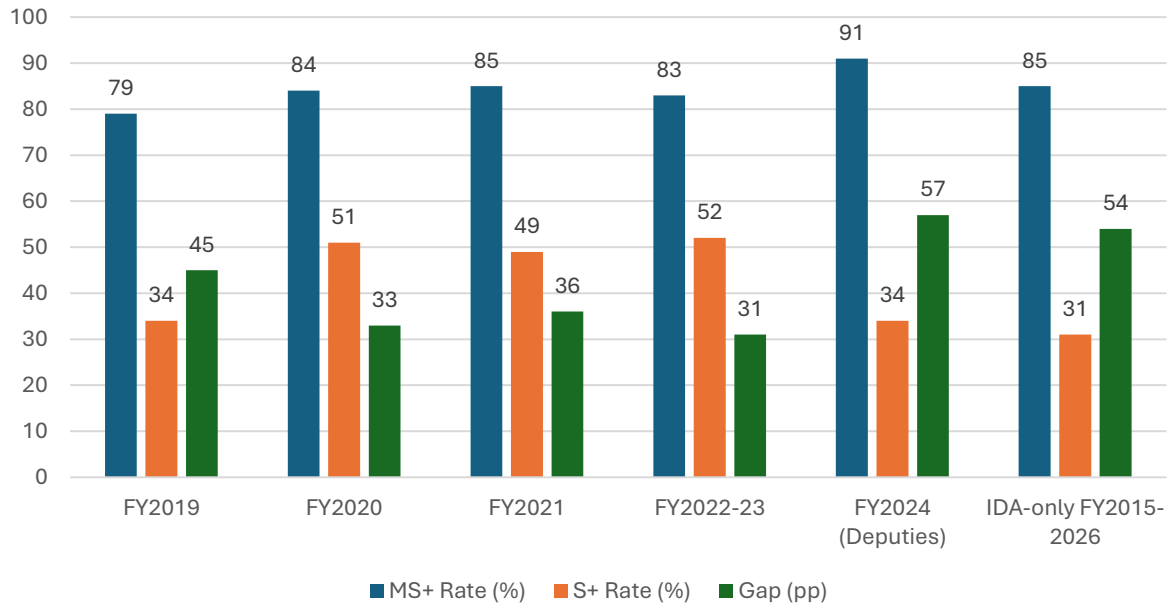
The sectoral data is worse. The Macroeconomics, Trade and Investment Global Practice — the largest IDA portfolio at $26.9 billion — achieves Satisfactory outcomes 11.7 percent of the time. Development Policy Financing in IDA fragile states achieves 10.8 percent. Sub-Saharan Africa, which receives 70 percent of IDA resources, committed $98 billion over the period, of which $71.5 billion — 73 percent — went to below-Satisfactory projects. IFC projects in fragile states collapsed from above 60 percent Satisfactory to 11 percent in CY2020–22.

A finance ministry official in an IDA-eligible country answering the ODI survey had no access to this information. A survey that claims to be the most comprehensive assessment of how client countries view MDB performance, while withholding the most comprehensive assessment of how MDB projects actually performed, is not a diagnostic instrument. It is an exercise in managed perception — and one whose conclusions are readily deployable in support of arguments for expanding MDB lending volumes, precisely the institutional positioning of the organisation that produced it.

The 91 percent problem

The information gap extends into IDA's own governance. The IDA21 Deputies Report — on which 59 donor governments committed $23.7 billion — states that IEG ratings show "satisfactory performance in 91 percent of IDA-financed operations." This uses the Moderately Satisfactory or above (MS+) threshold, which includes every project that partially achieved its objectives. The actual S+ rate is approximately 31 percent. (See Figure 1 and Annex A for the full MS+/S+ methodology.)

Figure 1: IDA Project Outcomes in Percent (IEG)
The Accountability Gap: What Deputies are told vs what the data shows



Deputies from 59 countries committed to IDA 21 on the basis of a reported 91 percent success rate. The actual success rate is less than a third. The institutionalisation of MS+ as the headline metric is not standard practice across the Bank Group — IFC and MIGA evaluation systems

both anchor performance around "Successful" or "Satisfactory" thresholds rather than intermediate categories (see Annex A). The softening of the benchmark is primarily a feature of World Bank lending-side corporate reporting.

ODI, which serves as Secretariat of the IDA technical working group and positions itself as the leading independent voice on MDB reform, has never flagged this gap. For an institution whose strategy is built around "robust and compelling evidence to inform our ideas and recommendations," the omission is difficult to reconcile with the claim.

Why this has persisted

The persistence of the MS+ benchmark and the survey's design choices are not accidental. They are sustained by a political economy in which no major actor has strong incentives to redefine success downward. Shareholders prefer expandable institutions with positive headline metrics; replenishment politics reward optimism; management's credibility rests on demonstrated results; and think tanks that depend on the same shareholders for funding face significant institutional risk in challenging the reporting framework.

ODI's position illustrates the dynamic precisely. Its financial structure — 78 percent of income from donors who are simultaneously IDA shareholders — creates strong incentives against adversarial scrutiny. The survey itself was funded by the Open Societies Foundation and the Gates Foundation, reviewed in draft by MDB staff, and disseminated through channels that amplify its reform-friendly conclusions. The new strategy, for all its eloquence about independence, does not resolve this conflict. It points toward more narrative building, more coalition convening, more advisory services — activities that keep ODI useful to the institutions whose performance it might otherwise examine.

What this costs

When reform debates centre on capital adequacy rather than delivery capability, the consequences are operational. Instruments that consistently fail — DPF in fragile states at 10.8 percent — continue to be deployed at scale because the failure rate is invisible in headline figures. Proven alternatives — Programme-for-Results at approximately 61 percent — remain a small fraction of the portfolio. Implementing partners with strong delivery records are excluded from IDA's highly centralised delivery architecture.

At 31 percent Satisfactory, with the Bank's largest Global Practice succeeding one time in ten and DPF in fragile states failing nine times in ten, "the question that every implementing partner, every concessional financier must ask is: why should 100 percent of IDA resources flow exclusively through a delivery platform that fails at this rate?" (Brar, 2026).

IV. The Larger Problem

This note uses ODI as the case study, but the problem is systemic. MDB reform discourse increasingly privileges politically functional metrics over outcome-accountable evaluation, allowing expansion and replenishment to proceed without confronting underlying delivery performance. Once demand, satisfaction, coordination, and institutional relevance become proxies for success, actual delivery performance disappears from the conversation entirely.

ODI is a capable institution with talented researchers. The Fellowship Scheme has built genuine relationships across governments in the poorest countries. But capability without accountability produces sophisticated analysis of everything except the question that matters most: after six decades and trillions of dollars, is the delivery system working?

Annex A: The S+ / MS+ Benchmark

This analysis uses the Satisfactory or above (S+) threshold from IEG's six-point outcome scale (1 = Highly Unsatisfactory through 6 = Highly Satisfactory). Moderately Satisfactory sits at 4 — below the Satisfactory threshold of 5. A project rated MS has partially or mostly achieved its objectives but has not, by the Bank's own definition, satisfied them.

The use of Satisfactory/Successful as the meaningful benchmark is not an idiosyncratic reinterpretation. It reflects the standard that operates across the World Bank Group's own evaluation architecture — everywhere except the lending-side corporate reporting that governs IDA and IBRD headline metrics.

IFC investment evaluations use a four-point scale (Highly Successful, Successful, Mostly Unsuccessful, Unsuccessful) in which Successful is explicitly the pass/fail point. There is no "Moderately Successful" buffer. IEG's RAP reports present IFC development outcome results against the "Mostly Successful or above" threshold for the expanded six-point scale, but the original four-point architecture established Successful as the operative standard.

MIGA guarantee evaluations historically used a four-point scale (Excellent, Satisfactory, Partly Unsatisfactory, Unsatisfactory) through FY2019, with Satisfactory as the explicit pass/fail threshold — no intermediate category existed. After FY2019, MIGA shifted to a six-point scale aligned with IFC (Highly Successful through Highly Unsuccessful), which now includes a "Mostly Successful" category. But crucially, IEG's RAP reporting continues to centre on "Successful or better" as the meaningful success threshold for MIGA. RAP 2024 reports that "72 percent of guarantee projects [were] rated satisfactory or better" — using Successful, not Mostly Successful, as the positive benchmark.

The institutional inconsistency is stark. IFC anchors performance around Successful. MIGA anchors performance around Successful. But IDA/IBRD corporate reporting normalised Moderately Satisfactory or above as its headline metric — a category that, by the Bank's own rating definitions, encompasses projects that have not achieved their objectives satisfactorily. The softening of the benchmark is not uniform across the Bank Group. It is primarily a feature of lending-side corporate reporting — the reporting that governs the largest concessional fund in the world and that directly informs the replenishment negotiations through which 59 donor governments committed $100 billion.

Where IDA's S+ rate is 31 percent and its MS+ rate is 91 percent, the 60-percentage-point gap is the space in which delivery failure is absorbed, normalised, and rendered invisible to shareholders. The argument is not that MS-rated projects are failures in an absolute sense. Many produce meaningful benefits despite implementation shortcomings. The concern is institutional: once MS+ becomes the headline success metric, partial achievement and full achievement become analytically indistinguishable in corporate reporting and replenishment negotiations.

IEG's own analyses note that project ratings "increasingly cluster in the moderately satisfactory or satisfactory points of the scale" — a structural artifact of the MS+ corporate target. The MS+ benchmark has a ceiling effect: at 83–88 percent MS+, further improvement is mathematically limited, making the metric uninformative. S+ captures whether teams are delivering against stated objectives, not merely avoiding outright failure.

The distortion is most severe in FCS contexts. FCS projects recorded 74–77 percent MS+ in FY2021 versus 91 percent for non-FCS — a 15-point gap. Under S+, FCS projects achieve roughly 35–40 percent against approximately 60 percent for non-FCS — a 20–25 point gap. MS+ compresses the FCS performance gap by lumping partially-achieving FCS projects into the same "acceptable" category as fully-achieving non-FCS projects. For an institution deploying $100 billion, the difference between a 15-point and a 25-point performance gap determines whether the FCV Strategy's objectives are being met or whether the metric has been chosen to obscure that they are not.

For the complete methodological case, see Brar (2026), Annex B.

Annex B: The ODI MDB Client Survey — A Methodological Critique

The ODI Global MDB client survey (Prizzon and Zeka, 2022; Prizzon et al., 2026) is the most widely cited assessment of how borrowing country governments view MDB performance. Its findings have directly informed G20 Evolution Roadmap discussions, shareholder debates, and IDA replenishment framing. ODI itself states that the survey is designed to ensure that "the views and demands of client countries should inform and shape [MDB] strategies and financing instruments." The survey's influence is not in question. Its analytical architecture is.

Five structural problems compromise the survey's value as an independent assessment of MDB effectiveness.

1. The survey measures perceptions without grounding them in independently verified outcomes

The questionnaire asks respondents to rate MDBs on: relevance of roles and functions (financing, policy advice, research, convening); perceived effectiveness of individual MDBs across these functions; alignment with country priorities; quality of financing terms and modalities; demand for future grants and loans; quality of policy advice and technical assistance; coordination among MDBs at the country level; and speed of the project cycle.

What the questionnaire does not ask — in any form — is whether MDB-funded projects in the respondent's country achieved their stated development objectives. Respondents are never presented with country-level IEG performance data, sector outcome records, IDA success rates, instrument-level failure rates, or the MS+/S+ discrepancy documented in this note and in Brar (2026). The survey effectively asks "How satisfied are you with the institution?" without asking "Did the institution actually achieve its stated development objectives?" In evaluation methodology, the distinction between perceived effectiveness and demonstrated effectiveness is fundamental. The survey collapses it entirely.

2. The respondent pool creates structural response bias

The survey population consists of finance ministry officials, planning ministry officials, line ministry officials, and MDB country office staff — actors institutionally embedded within ongoing MDB financing relationships. These respondents negotiate replenishment access, manage active portfolios, and in many cases seek larger allocations. The survey's own methodology notes acknowledge some limitations of online questionnaires but do not address the structural incentive toward positive evaluation that arises from surveying actors who depend on the institution being evaluated.

This creates what survey methodology literature would call institutional dependency bias: the respondent's professional interest is aligned with the institution's continued relevance and expansion. The survey design does not adjust for this — through control questions, outcome-data contextualisation, or respondent stratification that might isolate the dependency effect.

3. The questionnaire architecture privileges expansion logic

The survey's question categories — relevance, alignment, future demand, coordination, project cycle speed — are framed around the institutional logic of scaling and responsiveness. They measure whether MDBs are relevant, aligned, and in demand. They do not measure whether

MDB operations have produced successful development outcomes, whether specific instruments have failed systematically, or whether lending volume has outpaced delivery capability.

This framing implicitly equates stronger client demand with institutional effectiveness. The finding that "57% expect demand for MDB financing to grow" is presented as evidence of institutional value. But demand for concessional finance from institutions offering below-market terms is structurally guaranteed — it tells you nothing about whether the financing produces results. In a context where IDA has doubled its lending volume while its Satisfactory outcome rate has remained flat at 31 percent, the distinction between demand and delivery is the most important question the survey could ask. It does not ask it.

4. Contradictory evidence was not disclosed to respondents

The IEG Master Database is publicly available. The data showing that 68.9 percent of IDA commitments between FY2015 and FY2026 went to below-Satisfactory projects is drawn from the World Bank's own evaluation system. ODI, as an institution with deep expertise in MDB operations and a formal role in IDA's technical working group, possessed access to — and analytical familiarity with — this evidence.

When respondents are asked "How effective is IDA?" without being told "IDA's S+ outcome rate is 31 percent in countries like yours," they are evaluating the institution under materially incomplete informational conditions. Informed survey design, when evaluating institutional performance, requires contextual disclosure of independently verified outcome data — particularly when that data contradicts the perception environment the survey is designed to capture. This is not a minor methodological omission. It is the central design choice that determines whether the survey functions as a diagnostic instrument or as a mechanism for validating institutional legitimacy.

5. The survey entered replenishment politics while presenting itself as technocratic research

ODI itself states that the survey has informed Evolution Roadmap discussions, G20 processes, donor debates, and replenishment framing. The 2026 report was reviewed in draft by staff from eight MDBs, including the World Bank. Its dissemination coincides with the IDA21 mid-term review process for which ODI serves as Secretariat.

The survey therefore functioned politically — shaping the informational environment in which replenishment and scaling decisions are made — while presenting itself as neutral evidence-generation. This is the operational expression of the broader thesis of this note: MDB reform discourse increasingly substitutes stakeholder perception and financing demand for independently evaluated development outcomes. The ODI survey is the most prominent instrument of that substitution.

None of this implies intentional manipulation. The survey architecture itself — what it asks, who it asks, and what it withholds — produced institutionally useful outcomes through structural design choices rather than deliberate deception. That is a more consequential critique, because structural biases are harder to correct and easier to reproduce than individual acts of suppression.

References

Brar, P. (2026). *IDA: Why Is the World's Largest Concessional Fund Delivering Satisfactory Outcomes on Only 31% of Its Portfolio?* April 2026. mdbreform.com/ida-performance-record/

Brar, P. (2026). *Revised World Bank FCV Strategy: Suggestions for Consideration. Annex B: The Case for S+ as the Correct Performance Benchmark*. February 2026. mdbreform.com

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

IEG (2024). *Results and Performance of the World Bank Group 2024: Appendix A — Methodology*. Washington, DC: World Bank.

IEG (2015). *World Bank Project Performance Ratings Codebook*. Washington, DC: World Bank.

ODI Global (2026). *ODI Global Strategy 2026–2031*. London: ODI Global.

Pantuliano, S. (2025). Keynote, On Think Tanks Conference 2025. "Beyond Budgets: Shaping the Future."

Prizzon, A. et al. (2022). *Country perspectives on multilateral development banks: a survey analysis*. London: ODI.

Prizzon, A. et al. (2026). *Reforming multilateral development banks: perspectives from client countries*. London: ODI Global.

UK Charity Commission (2025). *ODI Global, Charity 228248: Financial History FY2021–FY2025*.

World Bank Group (2025). *IDA21 Deputies Report*. March 2025.

Parminder Brar is the founder of mdbreform.com and a former World Bank operational manager with postings in FCV countries in Africa between 2003 and 2023.