



OPS4 Technical Document # 11:

Results of Comparison between Terminal Evaluation Review (TER) and Review of Outcomes toward Impacts (ROtI) Ratings on Outcomes / Process Outcomes and Sustainability / Intermediate States

December 10, 2009

Results of Comparison between Terminal Evaluation Review (TER) and Review of Outcomes toward Impacts (ROtI) Ratings on Outcomes / Process Outcomes and Sustainability / Intermediate States: Note for OPS 4

This note presents results of comparison between the Terminal Evaluation Review (TER) and ROtI Ratings by the Evaluation Office (EO). The key findings are:

- The TER rating on ‘intended outcomes’ and the ROtI ratings on the ‘intended outcomes’ component of ‘process outcomes’ are consistent. While using TER approach ‘intended outcomes’ of 82 percent of the projects were rated as ‘satisfactory,’ using ROtI approach ‘intended outcomes’ of 83 percent of projects were rated as ‘delivered’. For four out of five projects identical ratings were given using these two different approaches.
- There are moderate differences between the TER ratings on ‘intended outcomes’ and the ROtI ratings on ‘process outcomes.’¹ Overall, using the ROtI methodology, the “process outcomes’ of 66 percent of projects were rated to have been ‘delivered.’ For about three out of four projects the ratings were identical on a binary scale even though the phenomena being measured were slightly different.
- There are substantial differences between the ROtI ratings on ‘intermediate states’ for assessing likelihood of achieving global environmental benefits and the TER ratings on ‘outcome sustainability’ to assess risks to sustainability of outcomes achieved.
- Using the ROtI methodology, of the projects reviewed that “delivered process outcomes” and also rated “likely” to achieve global environmental benefits (66/172), 47 percent (31/66) were also assessed to have at least partially achieved their intended impacts by project closing (noted as “+” in the ROtI methodology). In comparison, using the TER methodology of projects rated to have ‘satisfactory’ outcomes and ‘likely’ to sustain them (99/172), 40 percent (40/99) were assessed to have at least partially achieved their intended impacts by project closing.

The conclusions from these findings are:

- The ‘intended outcomes’ aspect of the ratings through the two methodologies give a similar and consistent result.
- Even though the TER approach does not rate ‘process outcomes,’ which is only assessed with ROtI, it seems to capture a significant proportion of the overall process outcome achievements of a project.
- Different phenomena are measured when ‘outcome sustainability’ is rated using the TER methodology, as compared to ‘intermediate states’ rating using the ROtI methodology.
- The predictive value of the TER and ROtI methodologies are comparable in identifying projects that have achieved impacts by project closing.

Methodology

As part of this exercise, three comparisons were made:

- TER 'intended outcome' ratings compared with derived ROTl ratings on the 'intended outcomes' component of 'process outcomes.'
- TER 'intended outcome' ratings compared with ROTl ratings on 'process outcomes.'
- TER 'sustainability of outcomes' ratings compared with ROTl ratings on 'progress to intermediate state.'

Although TER and ROTl methodologies are based on the same documented evidence at project closing (primarily terminal evaluations, final PIRs, agency final evaluations), these comparison pairs use different rating scales and different criteria for assessment.ⁱⁱ The TER outcomes ratings assess the degree and quality of attainment of intended outcomes. The ROTl methodology weighs this assessment along with the relevance of those outcomes achieved in terms of their potential continuation toward global environmental benefits, or their likelihood of progress beyond the outcome stage.

For the purpose of this exercise of examining comparability and complementarities between the two methodologies, their conversion to an identical rating scale was required. To facilitate comparisons, the TER and ROTl ratings, on the given parameters, were converted to a binary scale. Also, while ROTl methodology rated achievement of impacts (or lack thereof), TER methodology does not.ⁱⁱⁱ In the absence of a suitable counterpart, this rating is not directly compared with any TER rating. However, it has been analyzed to assess contingent probabilities.

Only those completed projects from the OPS4 cohort that had both TER and ROTl ratings on all relevant parameters (TER Outcomes; TER Sustainability; ROTl Process Outcomes; ROTl Intermediate State; and, for impact assessments, those that were applicable) were considered for the purposes of this comparison. Of the 210 possible completed projects, 181 met the criteria for all types of ratings, and 172 met the criteria for ratings and are applicable to achieving impact (enabling activities and/or targeted research not included).

Conversion of TER Ratings

The EO gives TER 'intended outcome' ratings on a balanced six point scale; wherein a rating of '4', '5' or '6' denotes outcome achievements are in the 'satisfactory' range, and '1', '2' or '3' denotes that achievements are in the 'unsatisfactory' range. On a binary scale, the six point scale rating is converted into 'satisfactory' (= '4', '5' or '6') or 'unsatisfactory' (= '1', '2' or '3').

The EO gives TER 'sustainability' ratings using a balanced 4 point scale; wherein, based on an assessment of risks, a rating of '3' or '4' denotes that achieved outcomes are in the 'sustainable' range, as in, 'likely,' and '1' or '2' denotes that sustainability is in the 'unlikely' range. On a binary scale, the four point scale rating is converted into 'likely' (= '3' or '4') or 'unlikely' (= '1' or '2').

Conversion of ROTl Ratings

The EO gives ROTl 'process outcome' ratings using a four point scale. Within this four point scale the ratings on 'intended outcomes' are embedded.^{iv} ROTl ratings on 'intended outcomes' can be derived on a binary scale by combining 'A', 'B', and 'C' ratings together to denote that intended outcomes were

‘delivered,’ where as ‘D’ denotes that intended outcomes were ‘not delivered’. This derived rating has been used for making the first comparison of ‘intended outcome’ ratings.

For the second comparison, between ROTI ratings on “process outcomes” and TER ratings on “intended outcomes,” the ROTI ratings were combined in a different manner. This comparison examines, according to and within the criteria of ROTI, the broader ‘satisfactory’ cohort (‘A’ or ‘B’) and the broader ‘unsatisfactory’ cohort (‘C’ or ‘D’) for process outcomes, and compares this with the relevant TER cohorts. A rating of ‘A’ or ‘B’ denotes the project’s process outcomes were ‘delivered’ *and* ‘designed’ to address project activities beyond GEF funding, a rating of ‘C’ denotes that the process outcomes were ‘delivered’ but not designed to address post-funding project activities, and a rating of ‘D’ denotes that outcomes were ‘not delivered.’ The terms ‘delivered’ and ‘designed’ to address project activities beyond funding were assumed to be analogous to ‘satisfactory’ and ‘delivered,’ and not ‘designed’ and ‘not delivered’ are analogous to ‘unsatisfactory’. On a binary scale, the four point scale rating is converted into ‘delivered’ (= ‘3’ or ‘4’) or ‘not delivered’ (=‘1’, ‘2’).

The EO gives ROTI ‘intermediate states’ rating using a balanced 4 point scale; wherein a rating of ‘3’ or ‘4’ denotes that impacts leading to global environmental benefits are in the ‘likely’ range, and a rating of ‘1’ or ‘2’ denotes that impacts leading to global environmental benefits are in the ‘unlikely’ range. On a binary scale, the four point scale rating is converted into ‘likely’ (=‘3’ or ‘4’) or ‘unlikely’ (= ‘1’ or ‘2’).

Key Findings

Outcomes

The TER ‘intended outcomes’ ratings and the ROTI ‘intended outcome’ ratings are very consistent.

Table 1 presents a comparison of the TER and ROTI outcome ratings. For 78 percent of the projects the ratings using the two approaches are identical. As per the ROTI ratings, 83 percent of the projects had delivered the ‘intended outcomes.’ In comparison, as per the TER ratings, outcome achievements of a lower proportion of projects (82 percent) were rated in the satisfactory range. In the two thirds where disconnects were found, these were because projects had been rated differently at the margins on the expanded TER rating scale.^v

Table 1: Intended Outcome Rating Comparison

| EO ROTI Intended Outcome Ratings | EO TER Intended Outcome ratings | | |
|----------------------------------|---------------------------------|--------------|------------|
| | Unsatisfactory | Satisfactory | Total |
| Not delivered | 11 (6%) | 19 (10%) | 30 (17%) |
| Delivered | 22 (12%) | 129 (71%) | 151 (83%) |
| Total | 33 (18%) | 148 (82%) | 181 (100%) |

There are moderate differences between the ROTI ratings on ‘process outcomes’ and the TER ratings on ‘intended outcomes.’ Table 2 presents a comparison of the TER ‘intended outcomes’ and ROTI ‘process outcome’ ratings. For 72 percent of the projects there is no disconnect in the outcome ratings

between these two methodologies. As per the ROTI ratings, 66 percent of the projects had delivered the intended ‘outcomes’ and were ‘designed’ to address post-funding activities. In comparison, as per the TER ratings for achievement of ‘intended outcomes’ of projects, 82 percent were rated in the satisfactory range. There are fewer projects in the ‘unsatisfactory’ range using the TER methodology because the ROTI methodology measures the potential of those outcomes achieved to lead toward intermediate states. That is to say, according to the ROTIs, there were some project that delivered ‘intended outcomes’ but were not designed to ‘feed into a continuing process after GEF funding.’ Overall, even though TER approach does not intend to rate ‘process outcomes,’ which ROTI does, it seems to capture a significant proportion of the overall process outcome achievements of a project.

Table 2: TER Intended Outcome Rating and ROTI Process Outcome Rating Comparison

| EO ROTI Outcome Ratings | EO TER Outcome ratings | | |
|-------------------------|------------------------|--------------|------------|
| | Unsatisfactory | Satisfactory | Total |
| Not delivered | 22 (12%) | 39 (22%) | 61 (34%) |
| Delivered | 11 (6%) | 109 (60%) | 120 (66%) |
| Total | 33 (18%) | 148 (82%) | 181 (100%) |

Sustainability / Intermediate States

There are substantial differences between the ROTI ratings on ‘intermediate states’ for assessing likelihood of achieving global environmental benefits, and the TER ratings on ‘outcome sustainability’ for assessing risks to sustainability of outcomes achieved. Nearly half (45%) of the ratings do not match; which indicates a substantial disconnect between the two approaches. The TER ‘outcome sustainability’ ratings assess likelihood of sustenance of outcome achievements based on an assessment of “risks” to sustainability at the point of project completion. Thus, it relates to the preservation or continuation of gains realized, after the end of the project. It does not attempt to predict whether long term impacts, or global environmental benefits, will be achieved.

In comparison, based on the prevalence of the necessary conditions for replication, scaling-up, and continuation of project activities, the ROTI ratings measure the likelihood that a project will achieve its long term intended impacts. Thus, it attempts to predict the likelihood of attaining more than the gains already realized, and presence of a roadmap to facilitate achievement. The observed inconsistencies between the TER and ROTI ratings illustrate the extent of complementarity of these divergent but related phenomena.

Table 3: Sustainability/ Intermediate States Rating Comparison

| EO ROTI Intermediate States Ratings | EO TER Outcome Sustainability ratings | | |
|-------------------------------------|---------------------------------------|-----------------------|------------|
| | Sustainability Unlikely | Sustainability Likely | Total |
| GEBs Unlikely | 50 (28%) | 57 (31%) | 107 (59%) |
| GEBs Likely | 26 (14%) | 48 (27%) | 74 (41%) |
| Total | 76 (42%) | 105 (58%) | 181 (100%) |

Impact & Likelihood of Attaining Global Environmental Benefits

Using the ROTl methodology, of the projects reviewed that “delivered process outcomes” and also rated “likely” to achieve global environmental benefits (66/172), 47 percent (31/66) were also assessed to have at least partially achieved their intended impacts by project closing (noted as “+” in the ROTl methodology). By comparison, using the TER methodology of projects rated to have ‘satisfactory’ outcomes and ‘likely’ to sustain them (99/172), 40 percent (40/99) were assessed to have at least partially achieved their intended impacts by project closing. Another way of examining this is to assess the counterfactual, or examine the probability of having already achieved impacts by closing, among those projects that did *not* have both the ‘delivered outcomes’ and ‘likely’ rating using the ROTl methodology, and *both* ‘satisfactory’ outcomes and sustainability ‘likely’ rating using the TER methodology. In other words, what are the chances of a project having achieved impact at closing, without having achieved satisfactory ratings across both parameters, according to each methodology? For the ROTl methodology, these errors of omission were 24 percent (25/106), whereas they were 22 percent (16/73) for the TER methodology. Reasons for these errors of omission in the ROTl methodology derive from projects with early impacts which are likely small in scale and not sustainable beyond the outcome stage. Thus, based on the evidence available at the time of these assessments, the predictive value of both approaches seems to be comparable.

Annex 1: Ratings Across All Possible Combinations

If the ratings are compared across all possible combinations, there is perfect consistency in the ratings using the methodologies for 42 percent (76) of projects. There is perfect inconsistency for 12 percent (22) of projects and partial consistency for 46 percent (83).

Comparison of TER and ROTI Ratings (n=181)

| ROTI Ratings | TER Ratings | | | | Total |
|------------------------|---------------------------|-------------------------|-------------------------|-----------------------|------------|
| | Unsatisfactory / Unlikely | Unsatisfactory / Likely | Satisfactory / Unlikely | Satisfactory / Likely | |
| Not Delivered/Unlikely | 16 (9%) | 3 (2%) | 15 (8%) | 19 (10%) | 53 (29%) |
| Not Delivered/Likely | 3 (2%) | 0 (0%) | 4 (2%) | 1 (1%) | 8 (4%) |
| Delivered/Unlikely | 5 (3%) | 2 (1%) | 14 (8%) | 33 (18%) | 54 (30%) |
| Delivered/Likely | 3 (2%) | 1 (1%) | 16 (9%) | 46 (25%) | 66 (36%) |
| Total | 27 (15%) | 6 (3%) | 49 (27%) | 99 (55%) | 181 (100%) |

Annex 2: Ratings Terms and Description for ROTI and TERs

Desk Review of Outcomes to Impacts (ROTI)

Rating Descriptions

| Process Outcome Rating | Rating on Progress Toward Intermediate States |
|---|---|
| A The project's intended outcomes were delivered, and were designed to feed into a continuing process, with specific allocation of responsibilities after GEF funding. | A The conditions necessary to achieve intermediate states are in place and have produced secondary outcomes or impacts, with high likelihood that they will progress toward the intended Global Environment Benefit. |
| B The project's intended outcomes were delivered, and were designed to feed into a continuing process, but with no prior allocation of responsibilities after GEF funding. | B The conditions necessary to achieve intermediate states are in place and have produced secondary outcomes or impacts, with moderate likelihood that they will progress toward the intended Global Environment Benefit. |
| C The project's intended outcomes were delivered, but were not designed to feed into a continuing process after GEF funding. | C The conditions necessary to achieve intermediate states are in place, but are not likely to lead to impact. |
| D The project's intended outcomes were not delivered. | D The conditions necessary to achieve intermediate states are unlikely to be met. |
| U/A Unable to assess. | U/A Unable to assess. |

Impact Rating

“+” Measurable impacts achieved and documented within the project life-span.

Blank No impacts achieved.

Terminal Evaluation Review (TER)**Rating Descriptions**

| Outcome Rating | | Likelihood of Sustainability | |
|-----------------------|---------------------------|-------------------------------------|---------------------|
| 6 | Highly Satisfactory | 4 | Likely |
| 5 | Satisfactory | 3 | Moderately Likely |
| 4 | Moderately Satisfactory | 2 | Moderately Unlikely |
| 3 | Moderately Unsatisfactory | 1 | Unlikely |
| 2 | Unsatisfactory | Blank | Unable to assess |
| 1 | Highly Unsatisfactory | | |
| Blank | Unable to assess | | |

ⁱ The delivery of ‘intended outcomes’ in the ROTI rating scale is only one component of the ‘process outcome’ rating, whereas the TER methodology assesses only ‘intended outcomes.’ See Annex 2 for full ratings descriptions.

ⁱⁱ The TER uses a 6-point scale for “outcome” rating and a 4-point scale for “likelihood of sustainability;” the ROTI methodology uses a 4-point scale each for “process outcomes” and for “intermediate states.” See Annex 2 for full ratings descriptions.

ⁱⁱⁱ Although TER methodology addresses project impacts, it does not rate it. ROTI approach rates it as: ‘achieved or partially achieved,’ and ‘not achieved or not yet achieved’ at the time of project closing. See Annex 2 for full ratings descriptions.

^{iv} ‘Delivery of outcomes’ is one component of the ROTI rating; the second component assesses the quality of those outcomes in terms of addressing next steps toward intermediate states. See Annex 2 for full ratings descriptions.

^v Since the ROTI scale does not measure intended outcomes on a gradation of marginal achievement (as the TER “marginally satisfactory” and “marginally unsatisfactory” ratings), and instead only indicates intended outcomes that were “delivered” (‘A,’ ‘B,’ ‘C’) and “not delivered” (‘D’), the marginal assessment of ‘intended outcomes’ is only possible on the TER scale.