

SC Assessment and Comments on the PM Results Indicators (2007 data)

15 July, 2008

OUTPUTS

The Output indicator for 2008 is the same as that used for other years. It is a self-reporting assessment of the achievement of output targets¹ in the following categories: materials, policy strategies, practices, capacity, and other kinds of knowledge. The reported achievement of output targets is audited by an external auditor.

In the three previous years, the Science Council has provided a short commentary on what Centers submit as output targets. This year however the SC has not provided such a commentary.

The purpose of the commentary was to assist the Centers by clarifying the definition of output targets (i.e. deliverable results from research and capacity building activities) and the SC recommended leaving out of the final result those targets that did not fulfil the definition (i.e. outcomes, process milestones, on-going activity, vague partnership arrangements, general support/backstopping, technical assistance to partners and routine genebank management).

The results for the 2007 data shown in Table 1 and are based on the audited submissions.

Table 1. Output indicator results and number of output targets

Center	Indicator result (% achievement)	Output targets (#)
Africa Rice	80.4	51
Bioversity	82.2	45
CIAT	98.4	63
CIFOR	87.5	16
CIMMYT	96.3	109
CIP	84.9	146
ICARDA	90.4	73
ICRISAT	98.0	50
IFPRI	84.2	120
IITA	95.0	40
ILRI	78.3	46
IRRI	88.2	68
IWMI	90.7	43
World Agroforestry	92.2	77
WorldFish	87.9	58

¹ The output targets that are defined as: the annual deliverables, defined by quantity and type, expected in a specific year and contributing to achieving the MTP Project Outputs

As in previous years there is a high level of achievement with an average across all Centers of 89.0%. Also as in other years, the range among Center performance in this indicator is small (78.3% – 98.4%). Also, as in previous years, there is a very large difference in the number of output targets that the Center planned and subsequently reported, ranging from 16 to 146.

This year the SC did not provide a commentary, in part because of its recommendation that the output target monitoring should be delinked from the PMS and be replaced by other suitable output indicators. The experience in the use of this indicator over three years suggests that while linked to the PMS it creates perverse incentives. The expectation that Centers receive 100% of their set targets is unrealistic and counter productive in research that ought to involve strategic, innovative and even high risk projects where targets are difficult to achieve but if achieved have potentially very high impact on intended goals. Furthermore, this indicator does not facilitate the research process where a) failure to achieve the expected results may provide important understanding and feedback about the research problem, and b) serendipitous results can have high value, and c) scientists should have the incentive to drop a line of activity that they can tell is going to be unproductive, and shift their attention to something that their research has indicated is more promising. In other words, measuring % achievement of output targets not only provides an incentive to set unambitious targets, it also makes the production of pre-specified output targets more important than the process of discovery inherent to research.

Therefore the SC is proposing to replace the old measure of outputs (using the assessment of output targets achieved) with an assessment of other good proxy indicators of results, such as peer-reviewed publications, capacity strengthening and data. All these are measures that apply equally to any Center and where a suitably designed indicator can pose positive incentives and establish appropriate accountability.

Notwithstanding the above experience with the achievement of output targets as indicator for an annual performance measure, the SC encourages the Centers to continue the internal monitoring of the output targets by its own M&E processes.²

OUTCOMES

As in previous years Centers were instructed to report on the five (5) most significant outcomes in 2007³. The outcomes were to be linked to previous outputs by the Center but

² Internally used it is unlikely that there would be the same disincentives. In addition the SC has provided considerable oversight in strengthening the internal M&E procedures of the Centers through commentary on the relevance of the output targets as outlined in the 3 year rolling Medium Term Plans. And the new e- MTP system (CGMap) allows for the monitoring of achievement of output targets. Thus using SC commentary on MTP and using CGMap to track output achievements, each Center is now able to self monitor achievements in a frame work that would minimise the disincentives outlined above when used for external purposes.

this year there was no requirement to tie the outcome to a specific time frame of when the outputs were derived (This was a departure from previous instructions where the output/output target has to be relatively recent). Each outcome report needed to be supported by evidence of credible documentation from either the Center or target user. The Centers were requested to provide the following information in each outcome description:

- An outcome statement
- What output/output target resulted in the outcome?
- In which MTP was the output/output target identified as planned before its achievement?
- Where was the achievement of the output/output target documented?
- Who used/adopted or was influenced by the output? Identify the kinds of people or entities affected and their location by country.
- How was the output used or adopted? What was the nature of the influence of the output?
- What is the magnitude (in terms of, for example, geographic coverage or reach of intended user groups) of the outcome relative to the intended recommendation domain?
- What is the evidence for the outcome? Specifically, what kind of data were collected or study was conducted? Who collected the data and/or conducted the study? If not included in the outcome evidence, what is the evidence that shows that the outcome is derived from the output/output target?

This year the SC applied a new set of assessment criteria that were also published in the Internet⁴. The reason for this change was that having initially focused on improving the quality and clarity of the outcome case submissions, the SC felt that there was need to put more emphasis on i) the clarity of the linkage of the outcome to the Center produced output/output target, ii) the importance and relevance of the outcome, iii) the international public goods nature of the research that had led to the outcome, and iv) the accuracy and direct relevance to the outcome case of the evidence given in support of that case. Also, the SC's assessment in 2007 had brought up several questions regarding border-line cases⁵ and the SC felt that there was need for a more elaborated scoring to better differentiate between excellent, good, intermediate and weak outcome cases.

The criteria included two mandatory conditions: (i) Is the outcome described an actual outcome? and, (ii) Is there a reasonable (plausible) indication that the outcome is derived from a realized Center output or output target? Both of these needed to be fulfilled in order for the outcomes to be eligible for assessment. The SC's modified assessment criteria including the mandatory conditions are shown in Table 2. The eligible outcomes were assessed (scored by each criteria) by a minimum of four reviewers including two consultants. The Center score was calculated on the basis of the scores of remaining 8 criteria (scoring at one decimal) and was adjusted to a score of 0-10. The final Center score was the average of

³ Outcome is the external use, adoption, or influence of a Center output(s) (e.g. by partners, stakeholders, clients)

⁴<http://www.sciencecouncil.cgiar.org/publications/pdf/SC%20PM%20Outcome%20Scoring%20Table.pdf>

⁵<http://www.sciencecouncil.cgiar.org/publications/pdf/SC%20Comments%20on%20PM%20Results%200indicators%202007%20FINAL.pdf>

the 5 individual outcome case scores. Two other criteria were initially included viz. reporting the same output/output target in an earlier year; and the set of cases representing reasonable coverage of the Center’s research agenda, but after the initial scoring these were left out because all Centers were alike regarding these two criteria.

Table 2. – Modified scoring of outcome cases 2008

Criterion	Score
Is the outcome described an actual outcome?	
Is there a reasonable (plausible) indication that the outcome is derived from a realized Center output or output target?	
<i>Basic score if mandatory conditions are reached</i>	1
1. Is the realized output/output target from which the outcome is derived clearly identified in an MTP?	0-1
3. Is the realized output/output target clearly described and verified?	0-0.5
4. Is the outcome description specific on how the output was used to derive the outcome?	0-1
5. Is there clear evidence of the outcome?	0-1
6. Is the linkage from the Center output/output target explicitly documented in the outcome evidence?	0-1
7. Does the outcome show relevant, documented and verifiable adoption /use / influence of outputs from: new knowledge or creative use of knowledge OR innovative research or capacity strengthening OR outcome in more intractable targets	0-2
8. Does the outcome case have an IPG nature?	0-1
Total score for case (provided mandatory conditions are met)	1-8

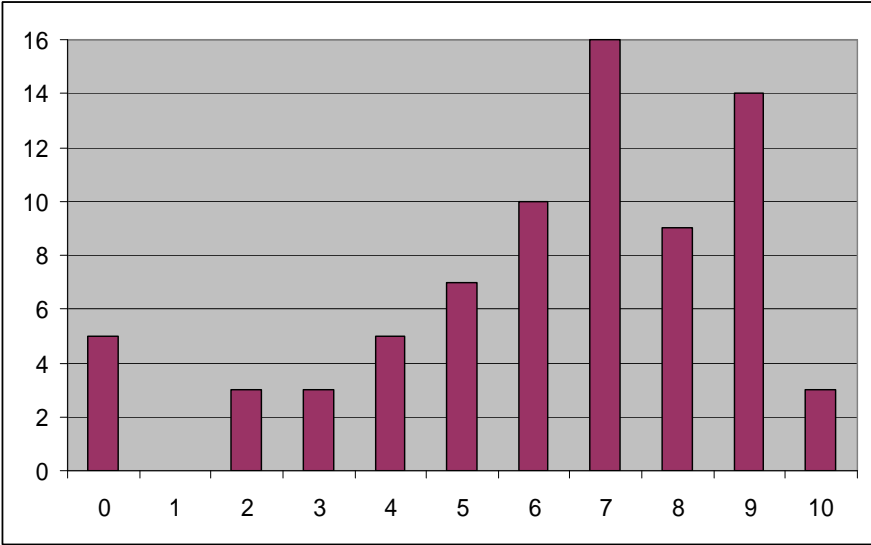
Results

The scores for the PM indicator for outcomes for the Centers are given in Table 3. The mean score for all Centers was 6.2 and median was 6.5. The frequency of the scores for the individual outcomes (a total of 75) is shown in Figure 1.

Table 3. Outcome scores

Center	1	2	3	4	5	mean
A.Rice	6.5	6.1	4.1	7.6	0	4.9
Bioversity	2.7	0	5.9	6.5	1.8	3.4
CIAT	8.8	5.3	7.6	4.7	4.7	6.2
CIFOR	5.4	6.2	5.5	7.4	8.9	6.7
CIMMYT	6.8	7.3	6.9	7.2	4.1	6.5
CIP	7.6	6.5	8.8	7.1	8.8	7.8
ICARDA	7.5	4.1	8.6	4.9	7.1	6.4
ICRISAT	6.2	9.8	9.4	7.3	6.5	7.8
IFPRI	9.4	0	7.6	8.2	9.4	6.9
IITA	5.6	1.8	2.5	5.9	4.4	4.0
ILRI	9.8	9.2	3.5	5.4	5.9	6.8
IRRI	8.8	9.4	9.4	7.1	7.1	8.4
IWMI	5.9	8.7	10	8.1	7.2	8.0
W.Agroforestry	8	7.6	4.7	9.2	2.4	6.4
WorldFish	0	0	5.9	2.9	6.5	3.1

Figure 1. Frequency of individual outcomes case scores at scale 0-10*



*Frequency of cases that scored: 0, >0≤1, >1≤2; ... >9≤10

These results in 2008 are not directly comparable with the earlier years’ results where 67 % of the outcomes received the highest possible scores. In the past, and particularly in 2007 which used a scale 0-2, it was not possible to distinguish among outcomes in terms of clarity of attribution to the results that were claimed to have led to the outcome, the importance and relevance of the outcome, etc. In 2008 these have been added as subjective criteria. Thus whereas in earlier years several Centers reached the highest possible score, this year the Center scores ranged from 3.1 to 8.4. The SC believes that the earlier scoring system did not provide the incentives for the Centers to improve their outcome cases. Further the SC felt

that while many cases formally fulfilled the criteria set at the time, there were very large differences in the relevance and the linkage to the Center's research among them. With the more elaborate scoring with a higher number criteria taking the quality, relevance and attribution aspects into account, the ceiling to be reached requires more work from all Centers to monitor and document their outcomes. It also requires more care in selecting cases where sufficient time has passed from the delivery of the research results or capacity to allow outcomes to accrue.

Despite the differences in scoring in the three years, some comparisons across years can be made. There was a lower proportion of cases that were discarded in 2008 (5 outcome cases compared to 7 in 2006 and 10 in 2007) and many more Centers scored above the average (10 out of 15 in 2008 compared to 7 out of 15 in the previous two years). This, in the context of a new scoring scale, suggests that there has been improvement in the performance indicator reporting.

In summary the SC believes that the outcome indicator can be a good proxy for real outcomes, and thus real performance. It needs, however, to be adjusted to the volume of research and expectation of outcome specific to each Center. i.e the number of cases requested for evaluation from each Center needs to be made relative to the Center size. The SC is also confident that the indicator sets appropriate incentives because it encourages planning towards outcomes, monitoring progress along the impact pathway and paying attention to appropriate baseline and success indicators so that the documentation of the outcome and its attribution to the research results can be credibly done.

There is however scope for improvement:

Timing

This year Centers were allowed to report outcomes that had derived from research in the past without limiting the time when the results (outputs) had been achieved. The outcomes however had to have been documented in the year of reporting i.e 2007.⁶ Rather than reporting outcomes from outputs where sufficient time had passed for the outcome to happen, in many cases outcomes were reported prematurely. The problem with choosing an outcome to be reported while the research is still going on, or a very early pilot outcome, is that the partners in the research may be subsidised to uptake the research, or there may be strong project-dependent incentives to use and adopt the emerging results. Thus there is no good measure of the real outcome, i.e. the use or adoption of the results for the beneficiaries' needs independently of the project arrangements.

Another example of a too early stage to report outcome is when the research results are used by the research commissioner but the actual outcome from the results is pending on further action (and often follow up project finding).

⁶ This year many of the documents lacked dates and in several cases it was not clear whether the outcome had been documented or observed in 2007 making it difficult to know why the case was presented in 2007 when it seemed that outcomes had been accumulated over previous years. The instructions will need to be improved in this area.

Linkage to outputs/output targets planned in the MTP

Because there was no restriction on when the outputs were achieved there often was no clear linkage of the outcome to the output targets. This was particularly so for outputs in MTPs preceding 2006-2008 when outputs were often quite generically described (output targets were recorded for the first time in MTP 2006-2008 when the SC began its more detailed commentary on the MTPs). In such cases the case description needed to be clear enough to show the linkage from the outcome to specific Center research. Several cases described outcomes from pipe-line research where outcomes are continuously accumulated.

Attribution

Attribution was a problem in some cases. Particularly when the Center has a facilitator's role it appeared that the research itself had been done by others and the Center's added value to help achieve the outcome was not clear. In other cases the description of the outcome was not clear on what the research output leading to the outcome was. Cases presenting a package of activities over time by different actors posed the difficulty to understand what the Center's contribution to the specific outcome was.

Evidence

Among the evidence there still were very few studies explicitly designed to document the outcome. In many cases the Centers placed the burden on the SC reviewers to seek for the evidence among multiple sources or very large or very generic documents not directly relevant to the case. It was not possible to verify the legitimacy of the evidence that was in local languages, such as Indonesian. There is need to restrict the volume of evidence material, to request it to be provided principally in English (Spanish and French being acceptable when they are the main global languages used by the Center), and to emphasise the importance of Center's own efforts to document outcomes.

Characteristics of outcome cases receiving different scores

The cases that received a top score (9-10)

The case description was clear about the research or capacity outputs that led to the outcome, evidence was mostly direct and appropriate, outcomes had had sufficient time to accrue from the outputs having been delivered, the outcome case was important and relevance and the research and outcome had clear IPG attributes. Slight reduction in scoring reflected some weakness in any of these aspects.

The cases that received a medium-high score (7-8)

The outcomes were mostly very relevant and often with IPG attributes but there was lack of clarity on the role that Center research had in the outcome to accrue, on output-outcome linkage or on the evidence. In some cases the outcomes were so early that their importance and relevance could not be judged yet.

The cases that received a medium score (5-6)

The linkage from Center research to the outcome was often by association and not described nor shown by evidence. Several cases lacked relevance or IPG attributes. In many cases the Center's role was rather minor in the research and other steps leading to the outcome.

The cases that received a medium low score (3-4)

Attribution was difficult due to weak description and absence of evidence. These cases received the scores they did because it was very plausible that there was a Center involvement in research leading to the outcome and the outcome has sufficient merits to yield a score albeit rather low.

The cases that received a low score (1-2)

In these cases the attribution was very problematic, but also the outcome cases were weak. In some cases the outcomes represented a very minor result compared what could be expected from the activity or cases represented simply an uptake of technical assistance with almost no research and outcome merits.

The cases that scored 0

These cases did not fulfil the mandatory conditions:

- not outcomes
- outcome only anticipated
- internal use of research results
- clear lack of attribution to the Center

IMPACTS

For evaluating Centers' performance in 2007 (2008 PMS exercise), the SPIA/SC continued to use the checklist format introduced last year where information requested under three major criteria was used for scoring the 3a impact indicator (weights indicated)⁷:

- (1) Ex-post impact assessment (epIA) studies / advancement of epIA methods (70%),
- (2) Building an impact assessment culture at the Center (20%),
- (3) Communication / dissemination and capacity enhancement (10%).

An overall modifier score was given for other impact related activity reported.

Only relatively minor changes were made to specific subcomponents this year. This reflected a conscious decision on the part of SPIA to strive to maintain continuity and thereby allow for consistent measurement and comparison of performance over time. Hence, only three modifications were made:

- (i) 'Unpublished reports' was dropped from the publication venue list (Criterion I.B.1) to encourage Centers to submit publications in a more advanced stage (as revised versions of previously submitted IA publications are not accepted). MS and PhD theses are now counted under Criterion III.B.4.
- (ii) 'Annual budget/expenditures' devoted to epIA work in the Center need not be reported for Criterion I.C. Centers need to only report senior full time equivalent (FTE) staff allocated for epIA. SPIA also agreed that the 1% 'yardstick value' for Criterion I.C was too low and, based on experience in other agencies (ACIAR, Embrapa), raised

⁷ The complete checklist with weights assigned to the components of each major criterion is appended (Annex I).

the yardstick value for the percentage of budget allocated to ex post impact assessment to 1.25%.

(iii) 'Website hits' (Criterion III.A.3) was removed as a component of Criterion III because of the wide disparity of statistics reported by Centers and lack of clarity as to what it represents.

All other components and subcomponents and their weightings remain unchanged.

To improve the quality of submissions this year, SPIA provided individual feedback to Center impact assessment focal points (IAFPs) on last year's results. IAFPs had already had access to the overall results and methodology used by SPIA in last year's evaluation⁸, but it was felt that individual feedback was also needed, particularly for those Centers who had submitted studies under Criterion I.A and I.B (65% of the final score) that were not considered legitimate epIAs and hence not accepted. These included adoption constraints analyses, farmer preference and demand type studies, pilot technology evaluations, and *ex-ante* impact assessments. While these studies are useful in their own right, and may well be counted as outputs or outcomes, none of these qualify as *ex post* impact assessments (epIAs) for purposes of this exercise⁹. SPIA also recognizes that there may be an element of *ex-ante* in many epIAs. However, while projections in *ex-post* studies are not uncommon, there must be some measurement of adoption and *ex-post* impact to qualify. EpIAs that were revisions of previously submitted studies or were very similar to current or previously submitted studies were also not counted. In preparation for this year's exercise, IAFPs were encouraged to seek clarification about a given study if they were in doubt, and several did.

Like last year, SPIA was fairly strict about counting only those studies that document adoption and impact (*ex-post*) of Center research or research related activities. Members carefully evaluated the characteristics of each of the studies submitted based on the summary description of the studies provided under Criterion I.A. In spite of these efforts to encourage Centers to submit only epIAs under Criterion I.A., the number of submitted epIA studies accepted as legitimate this year actually declined, although total submissions was about the same, reflecting a significantly lower percentage of acceptable studies this year. Of the 93 studies submitted last year, 74 (80%) were accepted as epIAs, whereas only 47 (52%) of the 90 studies submitted this year qualified as legitimate epIAs. This, as will be shown below, largely explains the lower overall performance achieved this year (mean score of 5.8 vs. 6.4 last year). There was considerable variability across Centers in this respect, however (see Annex Table 2).

Apart from assessing whether a study submitted by a Center actually qualifies as an epIA (as discussed above), SPIA does not perform a 'verification' role with respect to the data submitted by the Centers under Criterion I.B. For studies which qualify as epIAs, all information about those studies submitted by the Center is accepted at face value.

⁸<http://www.sciencecouncil.cgiar.org/publications/pdf/SC%20Comments%20on%20PM%20Results%200indicators%202007%20FINAL.pdf>

⁹ An epIA study refers to a published journal article, conference paper, book chapter (but not entire edited book), report or other publication that has entered the public domain, which is not a revised version of an earlier submission.

Verification was done this year on selected Centers/epIA submissions by the CGIAR Secretariat for quality control.

SPIA used the same method of weighting and scoring for each Criterion/component as described in last year's report (see footnote 2) with the few changes as noted above. Thus, adequate account is taken of differences in size of Centers and each Center is considered in relation to benchmark levels of achievement rather than relative to other Centers' performance. This allows Centers to monitor their own progress from year to year and de-emphasizes performance only in relation to other Centers' results.

Scores from Criteria I, II, III, and overall modifier points for each Center were calculated on an EXCEL spreadsheet. Scores for individual criterion and aggregate final scores (0-10 scale) are reported for each Center in Table 4.

Table 4. Center Scores for Impact Indicator 3A in 2008

Center	Criteria*				Total Points	Final adjusted score (0-10 scale)
	I (max =70)	II (max=20)	III (max =8)	Modifier to overall score**		
Africa Rice	50.6	12.8	8.0	4.9	76.3	7.6
Biodiversity	3.0	14.7	4.4	4.1	26.2	2.6
CIAT	39.1	13.5	6.5	3.1	62.2	6.2
CIFOR	40.5	14.0	2.7	4.7	61.9	6.2
CIMMYT	34.7	16.1	5.3	4.1	60.2	6.0
CIP	38.0	16.2	2.0	2.9	59.1	5.9
ICARDA	51.2	13.7	4.5	4.9	74.3	7.4
ICRISAT	45.8	16.2	7.5	3.3	72.7	7.3
IFPRI	36.3	15.0	2.8	3.3	57.5	5.8
IITA	47.8	12.5	7.4	2.7	70.3	7.0
ILRI	24.4	8.6	1.8	3.4	38.2	3.8
IRRI	28.4	9.1	2.1	4.6	44.2	4.4
IWMI	21.6	9.2	0.9	5.2	36.8	3.7
W. Agroforestry	40.2	10.7	0.0	1.3	52.2	5.2
WorldFish	54.1	12.0	6.3	2.3	74.7	7.5
Average	37.0	13.0	4.1	3.7	57.8	5.8

* Criterion I = epIA studies/ advancement of methods; Criterion II = building an IA culture; Criterion III = communication/dissemination & capacity building; Criterion IV = other impact related activity

** Modifier to overall score is combined score for of Criteria IV, IIIA.4 and IIIB.4 (max = 7)

Performance this year compared to last year is generally poorer for 9 of the 15 Centers. This is also reflected at the System level as the mean score in 2007 was 5.8 compared to 6.4 for 2006. A comparison of individual criterion scores for 2007 and 2006 (Annex Table 3) shows

that the overall lower performance can be attributed to the average lower score observed for Criterion I (37 vs 44 points for 2007 and 2006, respectively). The mean scores for criteria II and III and the modifier did not change dramatically. The significantly lower number of qualified epIAs submitted in 2007 (47) compared to 2006 (74) largely accounts for the low scores in Criterion 1.

Three-year results: 2005, 2006, 2007

SPIA/SC recognizes that modest or even major fluctuations in annual scores for any indicator are to be expected -- for a variety of reasons -- and hence advises against placing too much weight on an individual year's score, as this may not be the best measure of 'performance'. SPIA/SC encourages donors and other stakeholders to consider assessing performance of any given Center considering several years of scores together. For this reason, three-year moving averages are being recommended for use by the SC when assessing performance for any indicator. Multiple (continuous) years of results are also useful for Center management in tracking their own performance over time

Table 5 provides the results of the 3A and 3B impact indicator scores for every Center for the three-year period 2005 to 2007. A number of important observations can be made. First, ICARDA stands out as a consistently high performing Center for the 3A impact indicator every year, with final scores well above the CGIAR mean each year. It has the highest three-year average score at 7.7. Three other Centers (Africa Rice, ICRISAT and WorldFish) have demonstrated consistent improvement in their 3A scores over the three-year period and consequently, have also established relatively high means over the period, 6.9, 6.8, and 6.7, respectively. A large group of other Centers are close to or above the three-year CGIAR mean of 6.0, with CIP considerably higher (6.6). Four Centers – IRRI, IWMI, Bioversity and ILRI – are lagging behind the others with mean 3A impact scores of 5.2, 4.5, 4.5, and 4.4, respectively. The single most important factor influencing these scores is the number of legitimate epIAs submitted (Criterion I.A) each year, which was 2, 1, 0 and 1 for these four Centers in 2007. Without a sufficient number of qualifying epIAs to document a Center's impact, it is not possible to achieve a high 3A impact indicator score.

Two other points are worth noting. There is general consistency (positive correlation) between 3A and 3B scores for these Centers, although there are some exceptions. Thus, the three lowest performing Centers with respect to the 3A impact indicator are also the Centers with the lowest 3B scores. Two of the Centers with the highest 3A mean scores also had 3B scores well above the mean (ICARDA and ICRISAT). The other two highest 3A performing Centers for 3A had 3B scores slightly above the mean level. The exceptions are CIMMYT and IRRI whose 3A scores – at or below the mean – pale in comparison with their top performance in 3B which measures the rigor and quality of selected epIAs. These are two Centers with considerable research success that has translated into impact on the ground and that have deep experience and breadth in conducting epIAs. Overall, the consistency in Centers' relative performance between 3A and 3B is reassuring.

Although not shown in the table, there is also a positive correlation between scores of 3A and the component I.C – the latter being level of investment in impact assessment activity. Centers who have allocated relatively more resources to impact assessment activities

(measured in full time equivalent senior staff time against Center's total budget) typically have higher 3A impact scores. This result is consistent with expectations and argues for greater investment in resources committed to documenting impact, particularly for lowest performing Centers.

Table 5. 3A and 3B Impact Indicator Scores (2005-2007)

Center	3A Scores			3A 3-year Average	3B Scores* (2005)
	2007	2006	2005		
Africa Rice	7.6	7.4	5.6	6.9	6.7
Bioversity	2.6	4.7	6.2	4.5	4.2
CIAT	6.2	6.6	6.5	6.4	5.3
CIFOR	6.2	6.0	6.4	6.2	8.1
CIMMYT	6.0	7.0	7.2	6.7	8.8
CIP	5.9	7.0	6.9	6.6	7.7
ICARDA	7.4	7.9	7.7	7.7	7.7
ICRISAT	7.3	7.3	5.9	6.8	8.5
IFPRI	5.8	5.5	6.9	6.0	8.1
IITA	7.0	7.7	3.8	6.2	6.1
ILRI	3.8	4.9	4.6	4.4	3.1
IRRI	4.4	7.0	4.2	5.2	8.5
IWMI	3.7	3.8	6.0	4.5	1.7
World Agroforestry	5.2	5.9	5.9	5.7	6.8
WorldFish	7.5	7.0	5.5	6.7	6.7
Average	5.8	6.4	6.0	6.0	6.5

* Information for this indicator is collected and evaluated by SC/SPIA once every three years. As this was last done in 2006 (for 2005), the indicator rating carries over for two years, this being the final carry over year. For the 2009 PMS, Centers will be submitting two new ePIAs for evaluation by SC/SPIA

It is important to emphasize that the 3A impact indicator and, likewise, the 3B impact indicator measure a Center's efforts to rigorously document impact from its past research (as part of the accountability imperative towards CGIAR stakeholders & partners) and to institutionalize impact culture among its own researchers and partners. It is not a measure of the magnitude of impact.

SPIA/SC is aware that some donors and stakeholders would welcome an indicator which measures actual impact and, preferably, in a way that allows comparisons across Centers. While recognizing the value of such an indicator and potential for using these results for strategic allocation purposes, one should keep in mind the immense challenges and the associated dangers in trying to measure and compare actual impact in CGIAR goal terms across Centers. It must be emphasized that the 15 Centers research outputs and impact

pathways are highly diverse and target different direct and indirect channels of impact and types of impact, e.g., economic, social and environmental, not all of which lend themselves to straightforward measurement. SPIA is working with Centers to develop appropriate economic, environmental and social indicators to measure impacts from more difficult to assess research areas such as policy, biodiversity and training, but more progress must be made in this direction. Also the age of Centers has a substantial influence on the size and extent of realized impacts and as the Center age distribution is wide, it is felt that such comparisons are fraught. For these reasons, until now, SC/SPIA has focused on commitment to documenting impact and building an impact culture based on the premise that over the long run a culture of commitment to document impact is positively correlated with the impact of the Center's research.

However, to begin to approach the issue of measuring actual impacts, consideration is being given to requesting all Centers to conduct periodic (perhaps to coincide with EPMRs) meta-analyses of their cumulative success stories in an attempt to compile and quantify the size and nature of the economic and non-economic impacts over time. Methods to value and compare these impacts among Centers – in CGIAR goal terms – would be the next step.

ANNEX

Annex Table 1. Distribution of weights for 3a scoring exercise

	Component Weight	Indicator weights
CRITERION I. epIA studies/advancing methods	70	
I.B. EpIA studies		
1. Publication venue	10	
Refereed journal		10
Book chapter		7.5
Conference paper (includes proceedings)		5
In-house publication (reviewed externally)		5
In-house publication (not reviewed externally)		2.5
2. (Co-) authorship	5	
With other CG Center scientists		1.67
With NARS scientists		1.67
With ARI scientists		1.67
Center only scientists		0.83
3. EpIA coverage	10	
Commodity improvement		2.5
NRM related		5
Policy related		5
Biodiversity related		5
Training/capacity building related		5
Other (specify)		5
4. Distance down the impact pathway covered by the study	15	
Uptake/adoption (field surveys)		3.75
Influence (bibliometric/citation analysis, key informant surveys, etc.)		3.75
Intermediate impacts (improved yield/quality, lower risk, higher income, conserve resources, increase market access/efficiency, develop human capacity)		7.5
Ultimate impact (poverty, food security, environment)		15
5. Geographical breadth of impacts assessed by the study	15	
Single location within single country assessment		1.9
Multi-locations (regions) within single country assessment		3.8
Multiple countries (~ 2-5) assessment		7.5
Global assessment (i.e., spread over several continents)		15.0
6. Advances in new methods/models for epIA embodied in the study	10	
Addresses non-economic impacts		2
Addresses differential effects (different target groups)		2

Addresses positive and negative effects		2
Addresses multiplier effects (other sectors)		2
Employs novel methods (combines quantitative & qualitative, participatory approaches, etc.)		1
Other methodological advances (specify :)		1
I.C. Annual Expenditure on epIA activity	5	
Staff resources dedicated to epIA:		
CRITERION II. Building an IA Culture	20	
II. A. List of Internal Workshops	5	
II.B. Systematic Evaluation of User Relevance	5	
II.C. Using epIA in Planning/Priority Setting	5	
II.D. Baseline Studies	5	
CRITERION III. Communication/dissemination & Capacity Building	8	
III.A.1 epIA briefs	2	
III.A.2 popular media	1.5	
III.A.4 Other dissemination	modifier	
III.B.1 Conferences/workshops	2	
III.B.2 Training materials	1.5	
III.B.3 IA visiting specialists	1	
III.B.4 Other activity	modifier	
CRITERION IV. Other Impact Related	modifier (+)	
OVERALL TOTAL	98	

Annex Table 2. Number of Submitted and Accepted epIA studies in 2006 and 2007 for Criterion I.A, 3A Impact Indicator.

Center	2007			2006		
	No. of Studies Submitted	No. of Studies Accepted as epIA	Percent accepted	No. of Studies Submitted	No. of Studies Accepted as epIA	Percent accepted
Africa Rice	5	3	60.0	11	9	81.8
Bioversity	4	0	0.0	6	1	16.7
CIAT	9	4	44.4	12	11	91.7
CIFOR	4	2	50.0	3	1	33.3
CIMMYT	6	3	50.0	7	5	71.4
CIP	3	2	66.7	4	3	75.0
ICARDA	5	5	100.0	4	4	100.0
ICRISAT	6	4	66.7	5	4	80.0
IFPRI	4	4	100.0	5	5	100.0
IITA	20	7	35.0	12	12	100.0

ILRI	2	1	50.0	3	3	100.0
IRRI	5	2	40.0	8	8	100.0
IWMI	2	1	50.0	2	1	50.0
W. Agroforestry	8	4	50.0	7	4	57.1
WorldFish	7	5	71.4	4	3	75.0
Total	90	47	52.2	93	74	79.6

Annex Table 3: Comparison of 2006 and 2007 3A Performance Monitoring Scores

Center	Criteria I (Max =70)		Criteria II (Max =20)		Criteria III (Max =8)		Modifier to overall score		Total Score (Out of 105)		Final adjusted score (0- 10)	
	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006	2007	2006
Africa Rice	50.6	52.6	12.8	6.5	8.0	7.8	4.9	6.8	76.3	73.6	7.6	7.4
Bioversity	3.0	21.9	14.7	15.5	4.4	6.3	4.1	3.8	26.2	47.4	2.6	4.7
CIAT	39.1	50.6	13.5	6.3	6.5	4.9	3.1	4.3	57.4	66.1	6.2	6.6
CIFOR	40.5	38.1	14.0	14.3	2.7	5.3	4.7	2.5	61.9	60.2	6.2	6.0
CIMMYT	34.7	47.5	16.1	10.9	5.3	4.5	4.1	7.3	60.2	70.1	6.0	7.0
CIP	38.0	50.1	16.2	14.8	2.0	3.6	2.9	1.3	59.1	69.8	5.9	7.0
ICARDA	51.2	48.5	13.7	16.9	4.5	7.1	4.9	6.3	67.3	78.7	7.4	7.9
ICRISAT	45.8	45.5	16.2	16.9	7.5	7.0	3.3	3.8	72.7	73.2	7.3	7.3
IFPRI	36.3	45.3	15.0	8.8	2.8	0.0	3.3	1.3	57.5	55.3	5.8	5.5
IITA	47.8	56.1	12.5	14.9	7.4	6.2	2.7	0.0	70.3	77.1	7.0	7.7
ILRI	24.4	34.2	8.6	10.7	1.8	2.5	3.4	1.3	38.2	48.6	3.8	4.9
IRRI	28.4	54.5	9.1	8.5	2.1	4.5	4.6	2.5	44.2	70.0	4.4	7.0
IWMI	21.6	21.8	9.2	11.7	0.9	2.8	5.2	1.3	36.8	37.5	3.7	3.8
W. Agrofor	40.2	41.2	10.7	13.1	0.0	4.2	1.3	0.5	52.2	59.0	5.2	5.9
WorldFish	54.1	52.7	12.0	10.4	6.3	6.7	2.3	0.5	73.6	70.3	7.5	7.0
Average	37.0	44.0	13.0	12.0	4.1	4.9	3.7	2.9	57.8	63.8	5.8	6.4