

Comments from SC Task Force on the PM Test Indicators in the Pilot Year 2005

SUMMARY

The document was prepared by a Science Council (SC) task force (referred to as SC), involving members of SPME and SPIA and the SC Secretariat. The SC assessed the submissions, documentation provided in support of the submissions, and results for the test indicators on Elements 1 (Outputs), 2 (Outcomes), 3 (Impact), 5 (Quality of Research and Managerial Staff) and 6 (Quality and Relevance of Programmes).

The SC concludes that the indicators tested provide a basis for developing a meaningful PM system for programmatic aspect of Centre performance. However, the SC observed problems with indicators in each Element, which have to be revised. In assessment the SC identified weaknesses in the indicator itself, the instructions and definitions that had been provided and the data and additional documents submitted by Centres.

The indicator for outputs, % achievement of output target set in the MTP, can only be fully implemented when they are based on the MTP (MTP 2006-8) designed specifically to cater for PM. A common weakness in the records provided was that terms had not been carefully enough defined and drawing information originally provided using different terms that adopted in the new MTP and PM system led to lack of clarity for the purpose of this indicator.

The outcome indicator testing also revealed some confusion between the concepts output, outcomes and impacts. There was also lack of specificity about the nature of the evidence provided for an outcome. Accurate definitions, clearer instructions and expectation of evidence to support the outcome information will greatly improve the usability of this indicator.

The centre narratives and impact assessment (IA) case studies submitted for PM indicator 3A and 3B were of variable consistency as well as quality and revealed a range of difficulties in attempting to evaluate diverse impact assessment activities without a set of guidelines and criteria on which ratings would be based. Many of the responses for 3a included various general statements that do not clearly relate to specific IA-related activities or outcomes. For effective and consistent rating of responses to be possible, the questions posed need to be specific and query objective information that can be validated and for which ideal expectations are clear. SPIA/SC decided to revise the 3A and 3B impact indicators and establish draft criteria and indicator weightings for rating performance. The draft criteria and indicators were circulated among all Centres for comments and were revised on the basis of preliminary feedback. SPIA/SC will provide the guidelines and frame work as a basis for rating performance on impact in 2006.

The analysis on the “Potential to Perform” indicators of Elements 5 and 6 indicated that at least some of the indicators tested in the pilot year did not provide appropriate measures of the quality of research and managerial staff. Several indicators need a

clearer set of instructions and there is need to consider inclusion of some indicators while adding more appropriate ones.

The SC recommends that for all Elements the instructions need to be improved, the validation process and assessment frameworks need to be better designed and with some the set of indicators need to be revised. The SC will discuss these recommendations in its 4th meeting in September and submit its proposal on these programmatic elements to help implementing a workable PM system starting in 2006.

ELEMENT 1: OUTPUTS

Overview

The SC team examined the pilot outputs and output targets provided by the Centres in 2005 in order to gauge the usefulness of the PM output indicator and to make suggestions for improving the indicator and instructions for providing data in the future. The SC has not made an assessment of Centres based on the pilot output indicator.

In the pilot year, Centres reported output targets based on the MTP 2004-6, which had been prepared following very different guidelines than the current ones are. In MTP 2004-6 Centres variably provided indicators and/or milestones for outputs and activities and these were not linked to any established system of monitoring.

In the pilot year Centres were asked to report for each major program area the 5 most significant outputs and output targets that had been expected to be achieved in 2004. The test indicator was the % achievement of the entire set.

Observations made on 2005 PM report

Centre reports reflected variably the MTP 2004 submissions. In many cases reporting of the precise wording from the MTP made it difficult to assess what kind of a deliverable had been produced. Some Centres described the actual achievements, which helped the assessment. With some Centres, the reported outputs and output targets/milestones did not match with the MTP information. Some Centres reported multiple milestones as a single Output target/milestone. Because Centres were asked to report only a limited number of targets, and because some Centres used large Programs as the unit of reporting while others reported on MTP Projects, there was a large variability in the total number of targets reported. There was also a large difference in the proportion of original milestones/verifiable indicators reported (ranging from 4% to nearly 100%).

These variations in the reporting prevented accurate characterization and assessment of the Output targets as stated. However, the SC made several general conclusions and performed a rough categorization of the output targets.

Not surprisingly, in many cases Centres seem to have reported on those targets only (old milestones or verifiable indicators) that they had achieved or even exceeded in their judgement. Thus eight Centres reported >100% achievement, and the overall mean was 97%.

On the basis of the information provided it appeared that for a considerable number of targets it would be very difficult to obtain evidence on the extent of achievement, particularly regarding whether it had been exceeded. Several of the output targets reported could be better characterized as project milestones, evidence that the project was making progress but not the actual deliverable of the work.

The SC decided to rate the reported output targets to assess whether they represented deliverable products of materials, capacity, policy strategies, practices or other kinds of knowledge. In the process of this rating, some weaknesses were identified. The SC noted that in some cases the Output description was clearer than the Output target description, reflecting different interpretation of the instructions. We took this into consideration in our assessment. The targets were rated in three categories: acceptable, not clear and not a deliverable. The information was always interpreted as acceptable, if it was clear what product was aimed at even if this was not explicitly specified. Differences in the kinds of deliverables Centres would be producing, due to the diverse nature of their research, were carefully taken into consideration. We did not assess the magnitude of the targets or their relevance in terms of Centre mandate or CGIAR goals. Targets that were classified as not deliverables included for instance processes or progress in the research activity (such as funding achieved, project initiated or organisation of research), and outcome.

Problems in specifying clear output targets were found across all Centres. Table 1 shows the variability among Centres both in terms of number of targets and appropriateness as a deliverable. In the table the number of targets is shown in three groups. They ranged from 21 to 108. Individual milestones were counted separately when several were reported together as a single target.

Classification of the targets into categories (materials, practices, and so on) occasionally helped to understand what kind of deliverable was intended, but occasionally the classification did not seem appropriate.

In general, the output data reported in 2005 were not suitable to provide a meaningful basis for measuring Centre performance in producing outputs. Based on SC review of the reported output targets, about 40% met the definition of a deliverable, and 20% were considered not deliverables. Even among those that were deliverables, many were not quantified or specified. This is understandable since the process was based on the milestones and indicators developed for MTPs not intended for this purpose.

Table 1. SC assessment of the Pilot PM for outputs

Centre	targets assessed ¹	% okay	% not clear	% not deliverable
1	30 or less	80	16	4
2	30 or less	64	29	7
3	30 or less	57	23	20
4	30 or less	53	39	8
5	30 or less	48	44	8
6	30 or less	44	46	10
7	30 or less	38	48	14
8	30 or less	38	39	23
9	31-60	79	17	4
10	31-60	42	21	37
11	31-60	33	36	31
12	31-60	59	24	17
13	60 or more	44	30	26
14	60 or more	26	45	29
15	60 or more	25	52	23

Conclusions

It will be possible to use the indicator based on annual output targets for PM of outputs PROVIDED that the targets are based on a common understanding about their nature and magnitude and that they are real and identifiable in the MTP of current design.

The PM indicator for output can be fully implemented only when it will be based on a MTP report that has been designed for this purpose, i.e. in 2007 (using targets from the 2006-2008 MTPs).

The SC believes that in the interim, a useful PM system for outputs can be implemented for 2005 results and it will be making recommendations on how to deal with the 2005 target descriptions that are not satisfactory for the PM purpose as they are.

On the basis of the lessons learned from the pilot year exercise, the SC will adjust its assessment of the MTPs 2006-2008, design a mechanism for providing feed-back to Centres and recommend how the instructions for submitting data for the Output indicator can be improved.

¹ When multiple milestones were described as one output target, they were assessed separately.

ELEMENT 2: OUTCOMES

The SC team examined the data provided by the Centres in 2005 in the pilot year for Outcome indicator 2A in order to gauge the usefulness of the approach and to make suggestions for the following year. The SC has not made an assessment of Centres based on the pilot outcomes indicators.

Analysis

In general, the Centre reports on this outcome indicator provided in 2005 were not suitable to provide a meaningful indicator for outcome PM. The SC does however recognize the important steps made by the Centres towards achieving the PM purpose.

General deficiencies in the Pilot year related to Outcome indicator 2A

- There was inconsistency between definitions of “outcome” in the MTP guidelines and PM instructions. This is a major concern that must be corrected in the PM instructions.

In MTP guidelines, the definition is: “the external use, adoption, or influence of a Centre output or outputs (e.g., by partners, stakeholders, clients) that leads to changes in knowledge, attitudes, policies, research capacities, agricultural practices, productivity, sustainability or other factors required in order to achieve the intended impact.”

The PM system has defined outcomes as “changes resulting from uses of Centre outputs by stakeholders and clients, e.g. changes in knowledge, attitudes, policies, research capacities, and agricultural practices.”

The MTP definition focuses on the crucial intermediate step between outputs and impact. The PM definition skips a step and is less clearly distinguished from the definition of impact.²

- There is confusion between what is an output or an activity and what is an outcome. This is a common problem in the outcome reports. In short, activities are the work that generates outputs, outputs are the products of Centre activities, and outcomes are the adoption, use of or influence by those products.

The following are some examples of how activities and outputs were reported as outcomes³:

² The rationale behind the MTP definition is to focus on those steps in the impact pathway that the Centres have the most control over. The Centres themselves value adoption/use/influence as evidence of their performance, as demonstrated by the adoption studies that many do.

³ The examples are not direct citations.

1. The Centre, in collaboration with national partners, follows several approaches to tackle the complex problem of post-harvest toxins, including pre- and post-harvest strategies, food basket survey, bio-ecology of toxin production, biological control through competitive exclusion strategy, and resistance breeding. The Centre has designed public awareness campaigns with governmental organizations, industry champions and others to increase trader and consumer awareness of the deleterious nature of the toxin contamination in specific regions.

Comment: The public awareness campaign designs are the outputs and the food basket survey, bio-ecology of toxin production etc. are research activities. To be an outcome, the PA campaigns would need to have been implemented by the government organizations or other partners, which the statement would need to make explicit with evidence.

2. The Project supports all research related to this crop in hill locations in the target country through five NARS agricultural research stations. It provides resource-conserving practices and more productive crop varieties suited to farmers' circumstances in these fragile hill environments. The project also facilitates, organizes, and provides training to local researchers... Other activities to improve the production of this crop, including the use of soil conserving practices, have been initiated in 6 villages with 84 farmers. The Project ensures that there is sufficient seed of new varieties for farmers to replace old improved or local varieties, which yield very little.

Comment: This excerpt describes both Centre activities (eg. training in soil conservation practices) and outputs (eg. more productive varieties).

3. Work in China and India has demonstrated that investments in agricultural research, rural roads, and education are the most powerful investments for promoting rural economic growth and for reducing poverty, while other investments, particularly irrigation and some welfare programs, have proved much less effective on both scores. Another activity under this program is focused on assessing the impact of different types of infrastructure investments on growth and poverty reduction. Results thus far in China find that low-quality roads, which are mostly rural roads, yielded higher returns (four-times higher) than high-quality roads. Poverty reduction was also greater with low-quality roads in both rural and urban sectors.

Comment: The results described in this statement are research outputs (i.e., the findings on the relative effectiveness of different interventions) which are yet to be translated into outcomes.

Examples of the type of statements that describe outcomes are:

1. Improving access of the poor to global markets: The Program is an outcome of a broad consultative process and the core alliance formed between the Centre, donor agencies and the worldwide commodity industry and trade... The programs have been implemented in four countries by local and international non-governmental organizations, farmer organizations, local governments and civil society organizations.

Comment: While the Centre describes the program as an “outcome,” the program design is an output produced by the consultative process. The adoption of the program in the 5 nations is the outcome. This statement would be strengthened by information about when the program was adopted by the 5 nations.

2. Improved crop materials with two or more important traits were grown by at least 10% of farmers in eastern and southern Africa. As a result of efforts by the regional project, this material that yields 30% more than other popular varieties under dry conditions was grown on 250,000 hectares in 2004, and enough seed to sow 1 million hectares was produced for use in 2005.

Comment: The outputs are the improved maize materials. The outcome is the adoption of the materials on a specific number of hectares in a specific region and in a specific and recent year.

3. The Centre has posted growing numbers of fully-documented and rehabilitated datasets on its web-site. Since October 2004, these datasets are now fully downloadable from the Centre’s web site, before then datasets were provided on CD-ROM. In 2004, the number of sets requested or downloaded was 4,467, up from 3,004 in 2003. Of those, fifty-five percent were requested from developing countries. In 2004, the Centre commissioned a survey of the use of the datasets by student requestors. Of the responders (42 – or about 10 percent), 2 used datasets for undergraduate theses, 10 for Master’s theses, and 22 for Ph.D. theses. The remaining students used the sets for research papers or class projects. Eighty-one of the responders said the datasets were “very important” for their research.

Comment: The outputs are the improved datasets and their availability on the web. The outcome is not only documented with the proxy measure of number of downloads, but has been more directly assessed through the survey which discovered the nature and extent of actual use (albeit only for a small percent of a subset of the requestors). Another strength of this statement is that it describes use in the year targeted by the PM system (2004).

- There is lack of specificity about the nature of the evidence provided for an outcome. A description of the evidence supporting an outcome statement gives credibility to the statement.

Examples of clear and sufficiently detailed descriptions of the evidence follow:

1. The Water Challenge paper was summarized in a full-page article in The Economist (Economics Forum, 13 May 2004). The Copenhagen Consensus results were widely discussed in the world's top media (The Economist, The New York Times, Science, etc.). The Copenhagen Consensus book has been nominated for "best book of 2004" by The Economist.
2. A global readership survey and impact assessment of POLEX [Forest Policy Experts listserv] was completed in 2004. Response rates were 9% (English speakers), 17% (French speakers), 25% (Spanish speakers), and 56% (Japanese speakers).
3. Study surveyed 210 farm households in 8 villages in 2002 and again in 2004.

Conclusions

The SC observed several weaknesses in the information provided. Outputs were often described as outcomes, and the evidence available for the outcomes was not always clear. A problem related to the instructions was that outcomes were described more as impacts thereby differing from the MTP definition of adoption, use and influence of outputs.

The SC makes the following recommendation on the basis of these observations:

- The PM System definition of "outcome" needs to be revised so that it is consistent with the MTP definition. With inconsistent definitions, the PM system does not hold Centres accountable for achieving the expectations they have identified in the MTPs.
- There is a need to ensure that Centres clearly understand the definitions in PM reporting.

The SC will make recommendations for reporting and scoring the 2005 outcome data for indicator 2A and consider the complete set of outcome indicators making recommendations for revising and implementing the set in a workable PM system.

ELEMENT 3: IMPACTS

3A Developing and Implementing an Impact Assessment Process and Culture

The centre narratives for PM indicator 3A were of variable consistency. Some were good on describing their impact assessment (IA) strategy and on the processes they have set in train to institutionalise IA. Many described completed IA studies but others did not. Frequent mention was made of future activities (intentions) rather than focusing on IA activities during the last 3 years. Due to the lack of specificity in the time frame in the pilot question posed to the centres, some described the complete history of IA while others only described the recent past. This has been corrected in the revised guidelines for

next year to ensure that specific time frames are used to monitor progress (see attached appendix).

As an aside, SPIA found the histories of impact assessment in the centers (as produced by a couple of centers) to be useful background information and of possible broader interest to the CG investors as well as other centers. Thus, SPIA believes that it would be valuable to ask each centre to prepare a brief history of its involvement in IA up to the present so that this information is available as a baseline from which future progress can be gauged (one-page maximum). This summary of each center's IA history could be published by the Science Council or Alliance Office.

Some centres mentioned their focus on the R→D continuum in order to enhance their impact and the need for baselines to measure impact. Joint impact studies with NARS were a feature of some centre programs. An issue is the desirable balance between project IAs and program or institutional IAs. Few centres have explicit linkages between such IAs and staff performance monitoring, which arguably is a desirable objective in the IA institutionalization process.

No centre explicitly relates *ex post* IA to *ex ante* IA or priority setting. Indeed only one or two centres reported conducting *ex ante* IA. Many of the responses strayed well beyond the boundaries of impact assessment. For example, some centres are formally embracing the ILAC paradigm, but it is not clear how this relates to their program on IA. Impact pathway analysis (IPA) was claimed to be used by the centres to guide their decision-making, help in monitoring and evaluation, and position them better for later *ex post* IA. Some centres, particularly non-commodity centres, state that impact pathway analysis is the best they can offer in the short-term to be more accountable to donors in an *ex ante* context. The proof of the pudding will be the extent to which the IPAs are actually used in later *ex post* IAs, and not as a substitute for them. Some centres report training programs in IA for non-economists as part of the institutionalization process. Many centres have or are in the process of recruiting specialists dedicated to overseeing the IA program.

Some examples of useful (informative and substantive) IA-related activities/results reported:

- List of IA studies completed during last three years (but full citation still req'd)
- Centre-wide working group on IA formed in 2004 and objectives/rationale outlined
- One full-time IA position recently established with core funds
- Aggregate spreadsheet maintained (currently based on 15 case studies) to tally net benefits of success stories compared against total investments of the Centre
- Several regional workshops and joint impact studies implemented with NARS [but more details are needed]
- Independent external consultant serves as impact coordinator for the Centre [and] staff from throughout the centre engaged in IA through an interdivisional WG ...
- Workshop focused on lessons learned from IA studies
- Two-week training course on IA methods for its NARS collaborators

- CCER of the Centres IA work completed
- Centre has developed an evaluation and IA strategy document in which objectives of IA at the Centre are spelled out clearly
- Building an IA culture and capacity within the Centre through training, technical assistance and support in both planning and evaluation [but more specifics req'd]
- All research project proposals submitted now include a statement of how the impacts of the project will be assessed

Some examples of less-informative IA activities/results reported:

- Establishing better linkages between research output and development projects [how?]
- Centre is in the process of incorporating systems, livelihood, and poverty orientation emphasized in the new strategic plan, into its IA research.
- Centre will be holding an Impact Workshop in May 2005 [report following year]
- Participating in the development of impact assessment tools for INRM research [report when activity completed/manual produced]
- Centre continues to gauge its impact across research areas and the satisfaction of its clients through regular ex-post impact studies and CCERs [but no evidence]
- General statements such as:
 - “there is high demand by the biophysical scientists for socio-economists’ involvement in the planning and implementation of research projects [for] characterization of rural livelihoods and gender, ...”
 - “M&E provides an integrated system of reflection and communication that helps strengthen project design and implementation and stimulates partnership with project stakeholders.
 - “it is recognized that an effective M&E system must be planned, managed, and sustained with adequate resources.. and should be linked with results, outputs, outcomes, and impact summarized in the project logframes”

Given the wide variability in the types of information reported by the centres, e.g., various mixes of IA studies, future meetings, monitoring processes, staffing trends, and the adoption of specific research orientations, there is clearly a need for greater specificity in the questions posed. In addition, many of the responses included various general statements that do not clearly relate to specific activities or outcomes. For effective and consistent rating of responses to be possible, the questions posed will need to be more specific and will need to query objective information that can be validated and for which ideal expectations are clear. Accordingly, the 3a impact indicator statement and criteria to be used for the 2006 submissions has been revised in the light of these observations and considerations.

3B Impact Studies

PM indicator 3B involves the submission of two centre impact studies undertaken during the past three years, i.e., 2002-2004. This first round of case study submissions (see table Indicator 3b) was the exception in the sense that the three year limit was not followed (given that it is the trial year), i.e., several are 2005 publications..

The cases submitted were highly variable in quality and rigour and showed the whole range of problems encountered when attempting to do impact assessments for research other than technology development and crop germplasm improvement. A number of the studies were refereed publications in international journals and books. Some of the submissions are adoption studies rather than traditional impact assessments; and some of the studies are qualitative and descriptive with little analytical work involved. Many of the cases are ones done for SPIA earlier in its work on the impacts of CGI research (Evenson and Gollin eds., 2003), the impacts of the CGIAR on poverty alleviation, and the impacts of CGIAR NRM research (ongoing activity). While all the studies were looked at, SPIA did not attempt to rate them. Given that the centres did not have prior knowledge of SPIA's evaluation criteria, rating cases on the basis of those criteria would not have produced valid and useable results.

The purpose of the review of cases this year was to use them, in addition to the theory and the generally accepted best practices for impact assessment, to derive a clear, practical set of criteria for developing ratings. These will be given to the centres to use in preparing their submissions in coming years. These criteria and their suggested weightings in the overall rating are provided in the attached appendix.

SPIA discussed at length the best timing for submissions for 3B and SPIA is recommending that each center submit its two cases every three years, with approximately one-third of the centers submitting next year, the next third the following year, and the final third submitting three years from now⁴, in order to ensure each of the case studies receives an adequate review. For 2006, existing or revised versions of the 2005 submissions would be accepted, but with the centre's understanding that they would be judged on the basis of the attached set of criteria.

⁴ For example, IRRI, CIMMYT, WARDA, CIP and ILRI (primarily commodity-focused centres) would submit in 2006; ICRISAT, ICARDA, IITA, CIAT, WorldFish and ICRAF (primarily commodity + NRM oriented centres) would submit in 2007; and IFPRI, IPGRI, CIFOR and IWMI (primarily policy-oriented centres) would submit in 2008.

Table 2. Indicator 3b. Impact studies submitted by the Centres based on work done in the period 2002-2004 to document the difference that the center has made in the lives of poor people.

Centre	Title	Authors	Type of Publication / Journal
CIAT	Analysis of a Cassava Integrated R & D Approach: Has It Really Contributed to Poverty Alleviation?	M. Gottret & M. Raymond	Ch 12 <i>in</i> Agric. Research and Poverty Reduction (CIAT Impact Series 2) (2003)
CIAT	Impacts of the Use of Bean Technology on Poverty Reduction: Evidence from Uganda	S. David, R. Kirkby & S. Kasozi	Ch 13 <i>in</i> Agric. Research and Poverty Reduction (CIAT Impact Series 2) (2003)
CIFOR	C&I Indicators: NRM case study	M. Spillsbury	<i>SPIA commissioned study (2005)</i>
CIFOR	Influencing the global forest policy agenda—evaluation of CIFOR reseach	M. Spillsbury & P. Bose	In-house publication (2004?)
CIMMYT	Economic Impact of Leaf Rust resistance	C. Marasas, M. Smale & R. Singh	Economics Program Paper 04-01 (2004)
CIMMYT	Geospatial Dimensions of Poverty & Food Security—Case study of Mexico	M. Bellon et al.	CIMMYT publication (DRAFT) (2004)
CIP	Potato Genetic Improvement in Developing Countries-CIP's Role	T. Walker et al.	Ch 15 <i>in</i> Crop Variety Improvement... by Evenson and Gollin (2003)
CIP	Impact of Farmer Field Schools in Knowledge and Productivity: Potato Farmers in the Peruvian Andes	E. Godtland et al.	Published <i>in</i> Economic Development and Cultural Change (2004)
ICARDA	Impact of International and National Investment in Barley GI	A. Aw-Hassan & K. Shiheed	Ch 11 <i>in</i> Crop Variety Improvement... by Evenson and Gollin (2003)
ICARDA	Influence of the Spillovers to Australia on impacts of ICARDA	J Brennan, A Aw-Hassan & T Nordbloom	Published in Food Policy (2003)
ICRAF	Impact of Agroforestry-based Soil Fertility Replenishment Practices on the Poor in W. Kenya	F. Place et al.	<i>SPIA/IFPRI Poverty Impact study (2005)</i>
ICRAF	Impact of NRM Technologies on Small-scale farmers: Fertilizer Tree Fallows in Zambia	O. Ajayi et al.	<i>SPIA commissioned study (2005)</i>
ICRISAT	Wealth Generation through Chickpea Revolution: On-farm IPM of Chickpea in Nepal	S. Pande, V. Bourai & R. Neupane	ICRISAT Info. Bulletin No. 66 (2002)
ICRISAT	Impact of Vertisol Technology in India	P. Joshi et al.	ICRISAT Impact Series No. 10 (2002)
IFPRI	Science and Poverty: An Interdisciplinary Assessment of the Impact of Ag Research	Meinzen-Dick et al.	IFPRI Food Policy Report (2004)
IFPRI	Impact Assessment of IFPRI's Research & Related Activities Based on Economywide Modelling	Kym Anderson	IFPRI Impact Assessment Discussion Paper No. 21
IITA	An Impact Assessment of IITA's Benchmark Area Approach	B. Douthwaite et al.	In-house publication (shortened version of a paper soon to appear in a journal special issues) (2004?)

IITA	Harnessing Nature in Africa: Biological Pest Control Can Benefit ...	P. Neuenschwander	Commentary <i>in Nature</i> Vol. 432 (2004)
ILRI	Assessing the Impact of the SASA/CASREN Technology Interventions in the Sweet Potato-Pig Production Systems in Zitong County (China)	D. Pezo	ILRI Report (2004)
ILRI	Genetically Improved Dual-Purpose Cowpea: Adoption and Impact in West Africa	P. Kristjanson et al.	ILRI Impact Assessment Series No. 9 (2002)
IPGRI	An Analysis of IPGRI's Influence on the Int'l Treaty on Plant Genetic Resources ...	R. Sauve & J. Watts	Published <i>in Agricultural Systems</i> (2003)
IPGRI	Midterm evaluation of Participatory Mgmt of date palm genetic resources in the Maghreb project	H. Stewart	Internal project report (2004)
IRRI	Rice Research, Technological Progress, and Impacts on the Poor: Bangladesh Case Study	M. Hossain et al.	IFPRI EPTD Discussion Paper No 110 (2003) [IFPRI/SPIA Poverty study]
IRRI	International Research and Genetic Improvement in Rice: Evidence from Asia and Latin America	M. Hossain et al.	Ch 11 <i>in Crop Variety Improvement...</i> by Evenson and Gollin (2003)
IWMI	Impact Assessment of IWMI Research on Accounting for Water Use and Productivity	M. Giordano & R. Namara	Internal Document (prepared for Board of Trustees Meeting) (2004)
IWMI	Impact Assessment: Methodology & Application to the Mgmt of Soil Erosion Consortium Project	A. Maglinao et al.	IWMI Report ???
WARDA	Diffusion and Adoption of NERICA Rice in Cote d'Ivoire	A. Diagne	Working paper presented at pre-conf IAAE Workshop in Durban (2003)
WARDA	The Demand for Rice Varietal Characteristics Technologies	A. Diagne et al.	Paper presented at Symposium in Tokyo (2004)
WFC	Aquaculture Extension Impacts in Bangladesh	P. Thompson et al.	World Fish Centre publication (2005)
WFC	Development and Dissemination of Integrated Aquaculture – Agriculture Technologies in Malawi	M. Dey et al.	<i>SPIA commissioned study (2005)</i>

ELEMENT 5: QUALITY OF RESEARCH AND MANAGERIAL STAFF

The SC has considered the information from the Pilot PM to measure Quality of Research and Managerial Staff (Element 5) including some comparison with indicator 6B in of Quality and Relevance of Programs (Element 6) and has the following observations and suggestion for the implementation of a PM for these measures.

The SC discussion is ongoing and it plans to provide more specific recommendations after the SC in September after discussions with the WG.

The approach taken by the SC has been to:

- a) Consider the tested indicators 5A, 5B and 5C and 6B together on the grounds that both elements are considered evidence of “potential to perform” and thus there should be some consistency of results across the indicators.
- b) Consider if there are plausible alternative explanations – other than quality of staff or quality of programs – for the differences between Centers.
- c) See if there is sufficient variation in the results to distinguish high performing from low performing Centres. (Note that one year of pilot data might not give us enough information to answer this question.).
- d) Provide some subjective analysis, based on a) a review of EPMR analysis and b) on current and last year’s SC commentary on the MTP of the Centres, of how the ranking in the PM indicators (for the one year only) matches on information (subjective) from these other sources. The SC suggests that for the indicators to have credibility to those being assessed, they must have ‘face validity’ as well of course as the “evidence based” validity which is the objective of the PM indicators.⁵

SC has made some preliminary recommendations for change and identified areas for further discussion.

Consistency

We compared the three indicators (5A peer-reviewed publications), 5B (awards, and 5C (top-choice appointments) with each other and with the indicator 6B (papers co-authored with developing country partners). As 5A is at least partly included in 6B significant correlation can be assumed, and it was observed ($r = 0.65$). We assumed that high publishing records would be correlated with high number of awards. No correlation was observed with the full set of data ($r = -0.02$) but excluding three Centres with the three Centres with publishing records above the mean (>3 peer-reviewed publications per scientist) changed the analysis results notably ($r = 0.12$). Significant correlation was observed between 5B and 6B when one Centre with highest publication record was removed ($r = 0.52$). This suggests that publications and awards do reflect scientific quality, but the input data may need to be more accurate. The nearly complete lack of variability in 5C, discussed later, made correlation analysis meaningless.

⁵ “Face validity” means common sense validity based on multiple sources of information and experience, as opposed to validity that has been supported by some kind of evidence collected for this purpose. Face validity is important for political credibility.

Plausible alternative explanations

There may be other reasonable explanations for the patterns that are observed in the data. For example, the results for any of the indicators could be affected by unclear instructions. For example in indicator 5A (peer-reviewed publications), the data may be skewed because:

- ❖ The highest performer might have included publications that shouldn't have been included and the lowest performer might have done the opposite.
- ❖ The number of venues for publication or the ease of the peer-review process probably varies across the Centres and the scientific fields represented by the Centres. There is a need for a clear set of guidelines on the acceptance of journals, conference papers etc.
- ❖ Centres with more scientists who are not fluent speakers of languages commonly used in relevant scientific journals may have lower rates of publication. There is a need to ensure that all externally peer-reviewed publications in non English language that meet research standards are accepted.
- ❖ Other mission-focused demands, such as project and Centre administration and participation in Centre training programs, may lead to diminished time or opportunities for publishing, without in any way reflecting the quality of Centre staff as scientists.
- ❖ The results could reflect the attention given by each Centre to professional development and publication. This explanation is based on an assumption that participating in and presenting at professional conferences facilitates publication. In the pilot data, there is no correlation between the number of publications and the number of training days, which may reflect the lack of variability in the latter or the inclusion of other kinds of training than conference participation in indicator 7B6 on staff development. These indicators, particularly 5A are likely to reflect both the ethos of the institute in promoting scientific visibility of staff and the quality of staff to produce publications.

Indicator 5B (awards) may be affected by:

- ❖ The ratio of senior to junior scientists, with Centres with more senior scientists having a higher average # of awards than those with more junior scientists.
- ❖ No clear definition of the importance of the award.

Indicator 5C (% top-choice appointments) is probably affected by:

- ❖ The number of positions there were to be filled.

The SC is examining some of the alternative explanations and is exploring how to make corrections to the Pilot PM and plans to provide a new set of parameters for consideration by the WG in September 2005.

Variation across the Centers

To be a meaningful performance measure, an indicator has to vary sufficiently across Centres to distinguish between high and low performers. Looking at the preliminary results for indicators 5A, 5B, 5C, and 6B, it looks like there is meaningful variation across 5A, 5B, and 6B, but not for 5C. Eight Centres reported 100% top-choice appointments (5C) and the lowest score was 88%. The rate of 9 out of 10 top choice candidates accepting a position is not meaningfully different from 10 out of 10.

The lack of variation in 5C could also be affected by the definition of the indicator. Would the results have more variation if they were weighted for # of positions that were open during

the year or # of top candidates that were from outside the Centre? (Note that these are alternative explanations for the results).

The lack of variation could also reflect an unusual year. One of the problems with assessing the variation of the indicators on a single pilot year is that some indicators might detect meaningful change over time without being that useful in any given year. Two possible patterns could emerge over time. First, there could be a drop in the average for all Centres. If next year, none of the Centres had a 100% acceptance rate, it would be worth asking if the entire system had lost stature or was not compensating scientists at comparable rates as comparable institutions. Second, while this year all the Centres clustered at the top, next year there might be more variation.

So, for 5C, the Science Council should consider whether (1) the measure has face validity, (2) it needs some kind of adjustment to account for other factors, and (3) it is likely to be useful over time even though it doesn't seem particularly useful this year.

Variation between PM system data and other sources of evidence

The SC conducted an informal comparison of the PM data on research staff quality results with the results of EPMRs where possible. The purpose of the comparison was to check the "face validity" of the ranking of the Centres based on the PM data. The Centre ranking derived from the PM system data is not consistent with information on the quality of research staff drawn from other sources. For example, the reports of publishing rates provided in the PM system were notably different from those reported in recent EPMRs (the latter were mostly lower and in some cases considerably lower), reflecting variations in the definition of "peer-reviewed" and in the ways in which the rates were calculated. Also, looking at the overall results in the PM test year, the aggregated results from the indicators available so far do not correspond with the EPMR conclusions.

Conclusions

The analysis discussed above indicates that at least some of the indicators tested in the pilot year were not appropriate measures of the quality of research and management staff. The SC also questions whether the Pilot indicators make meaningful distinctions between Centres potential to perform. The SC therefore recommends modification of the set of indicators, improvement of the instructions for reporting data, and instructions for verification of the data for 2005 PM. These recommendations will be discussed in detail during the SC September meeting.

ELEMENT 6: QUALITY AND RELEVANCE OF PROGRAMS

The SC has considered the feasibility of assessing the quality and relevance of programs on the basis of 2003 and 2004 EPMRs and CCERs, as suggested in the Pilot year. The SC also examined the effectiveness of the indicator "Number of scientific papers published with developing country partners in 2004 in refereed journals, conference and workshop proceedings" as a measure of relevance.

Quality and relevance based on EPMRs and CCERs.

The SC does not believe that the current CCERs can give information useful for the PM purpose. Current CCER are often program specific and are variable in purpose and quality. There is an approach to upgrade the CCERS and to have the Centres Boards responsible for their implementation. Even so, it is unlikely that the CCERs can provide the Centre wide commentary that is needed for the PM process. Thus the SC does not believe that the CCERs can play this additional role for the PM.

The SC did look at the most recent EPMRs to consider if that process can also be used for the PM purpose. Again, there is a great diversity in the approach and type of information provided by the EPMR in terms of the quality and relevance of the programs. Thus the SC does not believe that it should rate Centres on the basis of the recent past EPMRs.

The SC is exploring an approach where future EPMR panels are asked to give a simple rating (say in scale of three) of both quality and relevance of programs, which the SC may endorse if it finds the report assessment credible on basis of clear justification and evidence. The SC will consider developing a set of criteria for use by EPMRs for this assessment, so that all EPMRs would apply similar approach when considering the rating.

The SC notes that the EPMR rating would be on a 5-year basis and this is not entirely satisfactory for PM on an annual basis. However, EPMRs should provide strong evidence of Centre performance that would be used for the same purposes as the PM system, namely for improving performance, providing accountability and guiding donors in their fund allocation if they wish.

In addition to the EPMR assessment, the SC is exploring if, in its annual assessment of MTPs, it can provide an overall assessment of relevance and some aspects of quality of the research programs that can be combined with that of the EPMR. The SC recommends that it explore how the assessment of MTPs can be used for indicator information in this Element.

Scientific papers published with developing country partners

This indicator has been discussed in the previous section as correlated with indicator 5A, which is at least partly included in 6A, and as potentially correlated with other Element 5 indicators tested reflecting publication productivity of research staff.

This indicator has similar weaknesses as were observed for indicator 5A. However, as the data do not include only peer-reviewed publications, inclusion/exclusion of publications is likely to be less of a problem. Yet, verification of data is important, and analysis of the raw data would likely allow recommendations for improving the instructions for this indicator.

APPENDIX

Impacts

Impacts are the longer range social, environmental and economic benefits that are consistent with the center's mission and objectives (e.g. increased agricultural productivity, improved food distribution, etc.).

The performance indicator for impacts focuses on the centre's capacity to measure the impacts of its previous research through ex post impact assessment. Ex post impact assessment is here defined as empirical analysis that improves understanding of the distribution and/or magnitude of effects attributable to CGIAR research outputs or capacity enhancement activities in the context of the CGIAR's goals of poverty alleviation, food security and environmental protection. Studies that identify the scale of adoption to date and likely pathways to impact, as well as small or large scale studies of actual and expected economic, social or environmental impact are included.

Subindicator 3A focuses on the overall performance of the impact assessment programme, and should be reported annually. Subindicator 3B focuses on the rigour of a specific impact study and is assessed on a triennial basis.

Historical narrative (for year 2006 only) - please submit a historical narrative (less than 500 words) that describes impact assessment activities at the centre through year 2004. The narrative should explain how impact assessment has been institutionalized in the centre, as well as major ex post impact assessment products produced prior to the period for the current performance measurement exercise. This description will not be rated.

3A – Overall Impact Assessment Performance. Specify and document (up to 1000 words) the ex-post impact assessment activities, outputs and outcomes over the past year at your centre, particularly related to: (a) ePIA studies; (b) innovation and advancement of ePIA; (c) communication/dissemination and capacity enhancement; and (d) impact culture, as detailed below.

Criteria/Questions	Indicators of performance	Expectation	Weight
<p>ePIA studies: Please provide a complete list of ex-post impact studies of technologies/outputs developed at least in part by the Centre, which have been produced over the past year (indicate these in scientific citation format and italicize authors who are Centre staff). Please provide a two to three sentence description of each, which specifies the research assessed, the geographic or thematic/programmatic coverage of the study, the impact indicators analyzed, and the approximate cost of the study.</p>	resources invested in ePIA studies (e.g. # number of studies, expenditure)	<p>The Centre should make meaningful attempts to assess major impacts that can be reliably attributed to its work. This implies that those Centres with a portfolio that includes more readily assessable categories of research (i.e. more crop genetic improvement) should be expected to have a greater number of large-scale quantitative studies. Centres with more difficult to assess portfolios will be expected to produce evidence that their research has been influential in fostering desirable shifts in policy or resource management practices, and should eventually have at least a limited set of quantitative examples of attributable impact.</p>	40%
	ePIA coverage of research areas with few prior studies (e.g. policy and NRM research)		
	studies with wide geographic scope		
	peer-reviewed publications		
	studies that trace research outputs to CGIAR goals		
<p>Innovation and advancement of ePIA methods and processes: Please list all IA methods papers or computer models produced over the past year by your Centre. In addition, please describe specific examples of how the studies listed above have applied innovative methods to overcome attribution problems, and explain any methodological workshops convened. .</p>	innovative methods applied in ePIA studies	<p>Centres should be advancing the relevance, robustness, and comprehensiveness of ex-post impact assessment coverage across their research portfolios. Methods will need to be continually improved for this to be possible.</p>	20%
	methodological papers produced		
	impact assessment models developed		
	methods workshops convened		
<p>Communication/dissemination and capacity enhancement: Please specify how the findings of ePIAs have been disseminated. In particular, please list the numbers of external workshops held, as well as the numbers of different media (such as briefs, newspaper articles, etc.) produced to communicate impact findings. Capacity strengthening activities and materials for partner IA specialists should be detailed, as well. Please also describe whether your Centre has its own website for impact assessment (in addition to participating in the Systemwide website).</p>	publication of ePIA briefs	<p>Ex-post impact assessment results should be communicated widely to intended audiences through appropriate media, including a portfolio of conferences, journal articles (for legitimacy), grey literature, briefs, a website, and potentially popular media. It is also valuable if methods and techniques are communicated to partners, such as NARS through capacity enhancement activities and materials.</p>	20%
	dissemination of ePIA findings in popular media		
	conferences/workshops for external audiences		
	impact assessment website and numbers of hits /downloads		
	seminars, workshops or training materials for NARS in the area of impact assessment		
<p>Impact culture (internal feedback and learning for user relevance): a) Please provide a list of internal workshops convened by the centre's impact assessment unit/specialists within the past year to help orient and improve the relevance of planned and ongoing research of the centre (for each, describe the theme and number of centre participants); b) Please provide 2 examples of systematic evaluation of user relevance of centre research outputs produced within the past year (two sentences to describe data collection and analysis for each, and major findings); c) Please provide specific examples of how empirical ex-post impact assessment findings have been applied as a basis for quantitative ex-ante impact projections that underpin the centre's priority-setting procedures, or have been used to validate earlier ex-ante work.</p>	internal workshops conducted by impact assessors to orient and plan research	<p>An important element of a Centre's impact culture is that adequate efforts are made to ensure that research outputs are relevant to the conditions and preferences of intended adopters and hence will generate outcomes and ultimate impact. Impact assessors should conduct workshops that help to plan and improve the relevance of future research, while systematic evaluations of the appropriateness of outputs to client needs should be undertaken regularly. Internal feedback loops from evaluation to the refinement of <i>ex ante</i> impact assessments and research outputs comprise a critical means for fostering relevance.</p>	20%
	clear examples of systematic empirical evaluation of relevance of research outputs		
	clear examples of use of specific ex-post impact assessment results in ex-ante analysis		

3B. Submit the two best impact studies your center has done in the period 2003-2006 that document the difference that the center has made in the lives of poor *and food/nutrition insecure people and the environment*. SPIA will rate these for rigor (*submissions from each Centre once every three years*)

**Checklist for SPIA rating of a
Centre's ex-post impact assessment study rigour**

1) Reasonable and transparent assumptions (15%)

a) Are the major assumptions regarding the assessment methodology (in all components of the analysis) reasonable and/or adequately justified (transparent)?

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

2) Clear presentation of assessed research (10%)

a) Does the study adequately describe how the Centre's activities have contributed to specific improvements in antecedent technologies/policy recommendations and what the costs were for the centre and its partners, if applicable?

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

3) Reliable and representative data (20%)

a) Are the sources or methods of derivation of data on adoption, productivity gains/unit cost reductions and other benefits clear? (5%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

b) Do adoption, productivity gains/unit cost reductions and other benefit data appear to be based on scientifically sound survey and other instruments? (15%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

4) Realistic counterfactual (15%)

a) Is the counterfactual scenario clear? (5%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

b) Does the counterfactual appear to represent a plausible scenario (including other potential sources of technical and policy change) in the absence of the assessed research outputs? (10%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

5) Sound attribution of benefits to research (20%)

a) Is the method of attribution of benefits to the research output clearly described? (5%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

b) Is the method of attribution of benefits to the research output appropriate (i.e. are potential mitigating factors sufficiently addressed)? (10%)

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

c) Has there been a suitable amount of sensitivity analysis to assess the robustness of the conclusions? (5%)

- Exceptional (4)
- Good (3)

- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

6) Distance down the impact pathway (15%)

a) Does the analysis make a reasonable effort to trace impact pathways to CGIAR goals (economic, environmental, food security and distributional), given the attribution difficulties of the assessed research output?

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (1)

Comments: _____

7) External input (5%)

Who has authored and reviewed the study?:

- Author(s) include external experts; the study is subject to external peer review (4)
- Author(s) only include centre staff; the study is subject to external peer review (3)
- Author(s) include external experts; the study is not subject to external peer review (2)
- Author(s) only include centre staff; the study is not subject to external peer review (1)

Comments: _____
