

CONSULTATIVE GROUP ON INTERNATIONAL AGRICULTURAL RESEARCH  
SCIENCE COUNCIL AND CGIAR SECRETARIAT

**SUGGESTIONS TO IMPLEMENT THE CGIAR PERFORMANCE  
MEASUREMENT SYSTEM**

SCIENCE COUNCIL SECRETARIAT

OCTOBER 2005

## TABLE OF CONTENTS

	Page
Summary	3
Element 1 – Outputs	6
Conclusions from Pilot Year	6
Definition of Output and Output target	6
Submission of Output Reports	6
Assessment and Verification of Output Reports	7
Outstanding Issues	7
Element 1 – Outcomes	8
Conclusions from Pilot Year	8
Definition of Outcome	8
Indicators for Outcomes	9
Instructions to Centers	9
Assessing Outcome Reports	10
Element 1 – Impact	10
Conclusions from Pilot Year	10
Revised IA Indicators in 2006	11
Elements 5 and 6 – Quality of Research and Managerial Staff and Quality and Relevance of Programs	13
Conclusions from Pilot Year	13
Indicators for Staff Quality	13
Indicators for Program Quality	14
Conclusions	15
Annex I	17
A. SC Assessment of Tentative Outcome Indicators not Found Suitable	
B. SC Assessment of Tentative Program Quality Indicators not Found Suitable	17
Annex II	18
Examples of Outcome Statements	
Annex III	20
Checklist for SPIA Rating of a Center's <i>Ex-post</i> Impact Assessment Study Rigor	

## CGIAR SCIENCE COUNCIL

### SUGGESTIONS TO IMPLEMENT THE CGIAR PERFORMANCE MEASUREMENT SYSTEM

#### SUMMARY

The Science Council (SC) is fully committed to the implementation of a successful Performance Measurement (PM) system. It will be an essential component of the new CGIAR Monitoring and Evaluation (M&E) system. This document builds on earlier SC comments. It presents recommendations for improving indicators for research performance and comments on the overall PM system.

The SC has been involved in developing indicators for five PM Elements: Outputs, Outcomes, Impact (as part of the "Results" elements), Quality of Research Staff and Quality and Relevance of Programs (as part of the "Potential to perform" elements). In the pilot year, 2005, several indicators in these elements were tested. Observations and lessons learnt were reported in a paper "Comments from SC Task Force on the PM Test Indicators in the Pilot Year 2005", which was submitted to the CGIAR Director in June.

Based on the lessons learned in the 2005 implementation of the PM system, the SC recommends using a limited set of indicators in 2006 and beyond. The SC recommends that the "Results" category will include three elements: Outputs (one indicator), Outcomes (one indicator), and Impact (one annual indicator and one triennial indicator). The SC recommends that two elements currently included in category "Potential to perform", namely Quality of Research and Managerial Staff, and Quality and Relevance of Programs be merged into a single element, Quality and Relevance of Current Research (four indicators). The SC has not reviewed the indicators for Finance and Institutional Health, which were not part of its responsibility. It considers both elements suitable for the PM system.

The SC's recommendations are summarized in Table 1. The indicators recommended for three "Results" elements and one "Potential to perform" element are judged to have the potential to accurately and credibly reflect real performance. In addition, they are a parsimonious and appropriately balanced set that do not place an undue burden on Centers or overemphasize potential to perform over actual performance. The SC recommends that:

- Additional indicators are not added to the set unless the analysis over 2-3 years reveals shortcomings in the system, at which stage the shortcoming should be carefully assessed;
- An appropriate balance of all elements and indicators is maintained and it is monitored how they as a set reflect real performance;
- The Challenge Programs (CP) are included in the CGIAR PM system for all parts that are applicable. The "Results" elements are equally suitable for CPs as for Centers taking into account the lag time in reporting on outcomes and impact;
- Effective, transparent and timely verification mechanism needs to be developed for all indicators of the PM system, particularly as the results are used for resource allocation, among other purposes.

**Table 1 - SC Recommendations for Implementing the CGIAR PM System in 2006 and Beyond**

Pilot elements and indicators	PM Elements in 2006
<b>A. RESULTS</b>	
<b>OUTPUT</b>	
<b>Tested:</b> % MTP output targets achieved in 2004	<b>Suitable.</b> The SC will work with Centers to guide the planning of output targets so that they are measurable and clear, and significant enough to result in a difference towards achieving CGIAR goals. The Centers will report on achievement of targets.
<b>OUTCOME</b>	
<b>Tested:</b> Centers report the five most significant outcomes in 2004 in relation to the Center's MTPs submitted in 2001, 2002, and 2003	<b>Suitable.</b> The SC will assess and rate 5 documented outcome cases which represent adoption, use or influence of Center outputs.
<b>IMPACT</b>	
<b>Tested:</b> Steps taken to establish a comprehensive set of impact assessment results across research areas, i.e., performance in terms of developing and implementing an impact assessment <i>process</i> and <i>culture</i> .	<b>Suitable.</b> Centers specify and document the ex-post impact (epIA) assessment activities, outputs and outcomes over the past year, particularly related to: (a) epIA studies; (b) innovation and advancement of epIA; (c) communication/dissemination and capacity enhancement; and (d) impact culture.
<b>Tested:</b> Two best impact studies done in the period 2002-2004 that document the difference that the Center has made in the lives of poor people	<b>Suitable (every three years).</b> Centers submit two best impact studies completed in the three year period that document the change towards achieving CGIAR goals. SC/SPIA will rate the studies for rigor.
<b>STAKEHOLDER PERCEPTIONS</b>	
<b>Tested:</b> Donor ratings of Center performance	<b>Not suitable.</b> The PM system is intended to inform the stakeholders. The PM system should provide reliable and verifiable data about performance as input to stakeholders for formulating their perceptions. As source of information about performance, stakeholder perceptions are vulnerable to low response rate (observed in the test year and likely to get worse if donors are not convinced that this indicator is meaningful), variable levels of familiarity, and subsequent biases. The SC recommends not using this element in the PM system, while recognizing the value of stakeholder perceptions in priority setting processes and in Center specific external reviews.
<b>B. POTENTIAL TO PERFORM</b>	
<b>QUALITY OF RESEARCH STAFF*</b>	<b>QUALITY AND RELEVANCE OF CURRENT RESEARCH</b>
<b>Tested:</b> % of top choices who accept offer for research positions	<b>Not suitable.</b> Dependent on Center specific recruitment patterns (10 one year, 1 in another year), and may be difficult to verify due to confidentiality of recruitment information. In test year did not allow

Pilot elements and indicators	PM Elements in 2006
	differentiation. The SC recommends not using this indicator.
<b>Tested:</b> Peer reviewed publications per researcher	<b>Suitable.</b> Peer reviewed journal articles per researcher multiplied by impact factor.
<b>Tested:</b> # of national, regional, international, and CGIAR awards received per researcher	<b>Not suitable.</b> The indicator is too influenced by different “awards culture” in different countries, and differences between scientific fields. The SC recommends not using this indicator.
<b>QUALITY AND RELEVANCE OF PROGRAMS*</b>	<b>QUALITY AND RELEVANCE OF CURRENT RESEARCH</b>
<b>Tested:</b> Scientific papers published with developing country partners	<b>Suitable.</b> Scientific papers published with developing country partners <b>Potential (not yet tested).</b> EPMR rating of quality of research programs <b>Potential (not yet tested).</b> Citation indices
<b>INSTITUTIONAL HEALTH</b>	<b>Suitable.</b> The SC has no recommendations for this element.
<b>FINANCIAL HEALTH</b>	<b>Suitable.</b> The SC has no recommendations for this element.

\* The SC recommends that these elements are merged into a new element, Quality and Relevance of Current Research, which includes indicators on input quality (staff quality), output quality (EPMR rating) and relevance (co-publishing and citations).

## **ELEMENT 1 – OUTPUTS**

### **Conclusions from Pilot Year**

In the PM pilot year 2005 one indicator was tested: for each major program outlined in the Medium-Term Plans (MTP) 2004-2006, Centers were asked to identify the 5 most significant outputs that had been expected to be produced in 2004 and report whether the outputs were achieved, partially achieved, not achieved or exceeded. The outcome indicator for PM was the percent of output targets achieved.

The SC concluded that 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.

### **Definition of Output and Output Target**

It is important that all Centers report on the same unit of analysis, namely the output targets. The annual targets should be set in the MTP clearly so that their measurement and monitoring of achievement can be accurately done, proven and verified. The SC will check the compatibility of the MTP guidelines and the PM instructions for 2006.

In the MTP, the definitions are as follows:

- **Outputs** are the products of research with a defined time line, contributing to reaching the Center goals by offering solutions to problems identified during the planning process.
- **Output Targets** are the annual deliverables, defined by quantity and type, expected in a specific year and contributing to achieving the MTP Project Outputs. Output targets are deliverables in knowledge, technology, policies, or services with an expected date of delivery. For the purposes of the MTP logframe, only output targets most likely to significantly contribute to the overall objective of the research should be listed.

### **Submission of Output Reports**

In the PM reporting in 2006, Centers will be requested to report on achievement of the annual output targets set in MTP 2005-2007 for 2005.

As the MTPs for 2005-2007 were done following old MTP guidelines that were not designed for a PM system, it is likely that the output indicators and milestones planned by Centers in their MTP 2005-2007 are not appropriate for PM reporting. In the SC's assessment of the test year results reflecting MTP 2004-2006, only 43% of targets reported in the Pilot year for 2004 were clear and accurate. Over 20 % of targets were not considered outputs at all.

Of issue then is the quality and clarity of the output targets (defined as output indicators and milestones in the MTP 2005-2007). The SC has commented on the Center MTPs for 2006-2008, which will be the first MTPs specifically prepared for output monitoring in the PM system. The SC still observed lack of clarity and specificity in the output targets. Furthermore, several MTP Projects listed a very large number of output targets, many of which were either activities or of minor importance relative to the goal of the project. The SC recognizes that

Centers are in a transition from an old type of MTP to one fully suitable for output monitoring. To facilitate the transition, it will meet with the Center Deputy Directors Committee in late 2005 to discuss the alignment of MTP with the PM system.

The SC recommends that following the MTP feedback and discussions, the Centers and CPs revise their 2005 output targets internally and report on an improved set in 2006, covering however, all projects and projects areas as planned in the MTP 2005-2007. Similarly the Centers should revise their 2006 output targets for reporting in 2007.

The SC recommends that Centers reduce the number of output targets by eliminating any that do not represent a major deliverable contributing to the overall output. Centers should aim roughly at one output target for an annual investment of about US\$ 500,000. Naturally, some variation in the investment required to achieve a specific target is expected, given the variety of intended outputs.

### **Assessment and Verification of Output Reports**

The Centers should report only fully achieved output targets. The output targets should be so clear that their achievement can be unambiguously monitored.

Centers should also be prepared to provide evidence of achievement for verification purposes. Published paper, other document, record, web site reference and researcher report are suitable forms of evidence<sup>1</sup>. Other forms of verification may be acceptable depending on the specific target.

The SC does not intend to provide further feedback on either the 2005 or the 2006 output targets to be reported on in 2006 and 2007, respectively. The SC will assess the output submissions in 2006 and provide a commentary recommending approval of the Centers own indicator or otherwise. For accommodating for a transition year, the expectation for clear output targets could be set at 70% (2/3) of the total output targets reported in 2006. This limit could go up as experience is gained.

The SC will not verify the results but it recommends that the System Office will develop a suitable verification mechanism for this PM element.

### **Outstanding Issues**

The SC acknowledges the moral hazard with the output indicator based on self-set targets. Centers may be tempted to set very low targets so that they could receive a high indicator value. In particular, using the PM results for funding decisions could result in such a trend, which the SC find highly undesirable. The SC strongly encourages the Centers to set demanding research targets, to include new and even risky areas of research in the MPTs and to aim at challenging outputs. In its MTP commentary the SC will highlight these aspects and also assess the output targets for relevance and significance in their contribution

---

<sup>1</sup> Publications per se are not outputs, although they are often a means of dissemination and a proof of achievement.

to solving pertinent problems and in producing international public goods. In addition, so that Centers are not unduly penalized for setting ambitious targets in their MTPs, the PM system will allow Centers to provide a separate note of clarification to give reasons for a) not having achieved certain targets, b) having achieved other deliverable results that were not planned for the given year; or c) having significantly exceeded expected results. In the last case Centers should elaborate the difference between the planned output target and the achieved one. The output indicator, nevertheless, should be % of targets fully achieved.

As 2006 is a transition year in output reporting, the SC expects that there are further lessons to be learnt in reporting, assessment and verification. After the next PM reports have been submitted there may be need to revise the way the reports are assessed.

The SC emphasizes that this indicator is intended to cover all the different kinds of activities Centers engage in, including external human capacity development, and it is not sensitive to the orientation of the Center; e.g. commodity versus policy focus. The indicator is purposefully simple and yet it accommodates the differences between the Centers. The assessment of the relevance of the research agenda, balance between different activities and suitability of the output targets for monitoring achievement is done by the SC in its review of the MTPs.

## **ELEMENT 2 – OUTCOMES**

### **Conclusions from Pilot Year**

In the PM Pilot year, only one indicator was tested; Centers were requested to report the 5 most significant outcomes that appeared in 2004 from outputs expected under their MTPs for 2001, 2002, and 2003. The Center reports on the outcome indicators were not suitable to provide a meaningful indicator for outcome performance. Centers, however, made important steps towards achieving the PM purpose for this element.

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 outcomes as the adoption, use and influence of outputs.

### **Definition of Outcome**

An unambiguous and mutually agreed definition of outcomes is crucial for compiling evidence of outcome and for accurate interpretation of the information.

Outcome as intended in the PM system is “the **external use, adoption, or influence** of a Center output or outputs (e.g. by partners, stakeholders, clients)”.

These outcomes in turn may lead to changes in knowledge, attitudes, policies, research capacities, agricultural practices, productivity, sustainability or other factors required in order to achieve the intended impact.

This definition, which is given in the MTP guidelines, focuses on the crucial intermediate step between outputs and impact, which needs to be clearly distinguished from impact (intermediate or ultimate). Along the impact pathway, Centers have most control over adoption/use/influence (outcome) and regular self-monitoring (e.g. through adoption studies) gives them feedback on their own performance. The change/impact occurring as a result Centers have little or no influence over.

Centers need to more carefully distinguish between outputs and outcomes. Activities are the work that generates outputs, outputs are the products of Center activities, and outcomes are the adoption, use of or influence of those products.

### **Indicators for Outcomes**

In 2004, the Performance Measurement Working Group (PMWG) discussed a plan to implement 5 indicators for outcomes. The SC considered the currently proposed measures of achievements of outcomes and found them in general not suitable for measuring results. The SC's comments on the tentative outcome indicators are provided in Annex 1, A.

The SC recommends that only one indicator is used, based on Centers' outcome reports, which is parallel to the Impact indicator 3B; Centers would be asked to report a certain number of annual outcomes deriving from research outputs specified in their MTP, and the reports would be made subject to the SC's assessment and rating. The indicator would be the SC's rating following the assessment of Center reports according to a detailed procedure (see below 4.5).

### **Instructions to Centers**

In the MTP logframe, Centers are instructed to state the expected outcome from their outputs. These are perceived outcomes occurring for the intended users that Centers also are requested to specify. It is expected that Centers study, document and collect evidence for outcomes occurring, and the PM outcome report is based on this evidence. Although Centers cannot be made accountable for achieving outcomes, over which they have much less control than for outputs, Centers can be expected to monitor the outcomes of their entire research agenda.

In 2006, Centers will be asked to report on the 5 most significant outcomes in 2005 resulting from outputs that the Center made available in 2002-2004. The outcomes can be specific to a particular annual output achievement (in the current MTP termed output target), or to an overall output from a Project as specified in the MTP. Each outcome report must be supported by evidence of credible documentation done either by the Center or others.

To make the information in the outcomes provided more specific, the SC recommends that the Centers will respond to the following prompts (in the on-line system) for each outcome:

- What output(s) resulted in the outcome? In which MTP was the output first identified? [maximum 50 words]

- Who used/adopted or was influenced by the output? Identify the kind of people or entities affected and their location by nation. [maximum 50 words]
- How was the output used or adopted? [maximum 150 words]
- What is the evidence for this outcome? Specifically, what kind of study was conducted? Who conducted it? [maximum 100 words]
- Please provide a reference or source (if not in a formal document) where the evidence can be reviewed.

These prompts would allow assessment of the reported outcomes. Verification would be based on the availability of the outcome document.

### **Assessing Outcome Reports**

The SC proposes a score based on the following criteria:

- Criterion 1: Of the five cases, how many are truly outcomes? Each case is scored 1 if it is an outcome and 0 if not. Total possible score for this item = 5.
- Criterion 2: Of the cases that scored “1” in the first item, how many have been clearly and specifically articulated (see tables in Annex 2)? Each case that scored “1” in the first item is scored “1” on this item if it meets the criterion and “0” if it does not. Total possible score for this item = 5.
- Criterion 3: Of the cases that scored “1” in the 2<sup>nd</sup> item, is there a verifiable source for the outcomes described? Total possible score for this item = 5.

The quality score would be derived by summing the scores for each of the criteria. So, with three criteria, the highest possible score would be 15 and the lowest possible would be 0.

This quality score will be applied in 2006 (on 2005 data) and will need to be reassessed after its first implementation. Eventually a more refined rating scale could be developed, once there is a common understanding of the expectations for outcome studies among the Centers.

## **ELEMENT 3 – IMPACT**

### **Conclusions from Pilot Year**

In the PM pilot year two Impact indicators were tested; the Centers were asked to provide a brief narrative on the steps taken towards developing and implementing an impact assessment (IA) process and culture.

The Center narratives and 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 elicit 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 Centers for comments and were revised on the basis of preliminary feedback. The revised guidelines and framework as a basis for rating performance on impact in 2006 is elaborated below.

### **Revised IA Indicators in 2006**

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 Center'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.

Sub-indicator 3A focuses on the overall performance of the impact assessment program, and should be reported annually. Sub-indicator 3B focuses on the rigor of specific impact studies and is assessed on a triennial basis.

**Historical narrative (one-time submission, for year 2006 only)** - please submit a historical narrative (less than 500 words) that describes previous impact assessment activities at the Center prior to the year 2005. The narrative should explain how impact assessment has been institutionalized in the Center, as well as major ex post impact assessment products produced prior to the period for the current PM 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 during 2005 at your Center, 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><b>epIA studies:</b> Please provide a complete list of ex-post impact studies of technologies/outputs developed at least in part by the Center, which have been produced over the past year (indicate these in scientific citation format and italicize authors who are Center 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 Center should make meaningful attempts to assess major impacts that can be reliably attributed to its work. This implies that those Centers 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. Centers 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><b>Innovation and advancement of epIA methods and processes:</b> Please list all IA methods papers or computer models produced over the past year by your Center. 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>Centers 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><b>Communication/dissemination and capacity enhancement:</b> 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 Center 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><b>Impact culture (internal feedback and learning for user relevance):</b> a) Please provide a list of internal workshops convened by the Center's impact assessment unit/specialists within the past year to help orient and improve the relevance of planned and ongoing research of the Center (for each, describe the theme and number of Center participants); b) Please provide 2 examples of systematic evaluation of user relevance of Center 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 Center'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 Center'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 2003-05 that document the difference that the center has made in achieving CGIAR goals. SPIA will rate these for rigor (*submissions from each Center once every three years*). SPIA's checklist for rating *ex-post* impact assessment study rigor is given in Annex 3.

## **ELEMENTS 5 AND 6 – QUALITY OF RESEARCH AND MANAGERIAL STAFF AND QUALITY AND RELEVANCE OF PROGRAMS**

### **Conclusions from Pilot Year**

In the PM pilot year 2005, three indicators were tested for Quality of Research and Managerial Staff. Centers were requested to report: a) the number of all papers published in peer-reviewed journals, books, monographs as well as book chapters per scientist (as defined); b) the number of national, regional, international, and CGIAR scientific and development awards per scientist; and c) the percentage of Center's top choice candidates for research and managerial positions.

For Quality and Relevance of Programs, one indicator was tested; Centers were requested to report the number of scientific papers published with developing country partners in refereed journals, conference and workshop proceedings.

The SC analyzed the tested indicators for staff quality by looking at consistency among indicators, also including other elements, where consistency would be expected, plausible alternative explanations for the results (other than potential to perform), variation across Centers, and variation between the pilot PM data and other sources of evidence. This analysis indicated that at least some of the indicators tested in the pilot year were not appropriate measures of the quality of research and management staff. A further observation was that the Pilot indicators did not make meaningful distinctions between the Centers' potential to perform. The SC's analysis is reflected in the recommendations below for implementing this element in 2006, including modification of the set of indicators under a new element, improvement of the instructions for reporting data, and instructions for verification of the 2005 data for 2006 PM.

The SC also considered the feasibility of assessing the quality and relevance of programs on the basis of 2003 and 2004 External Program and Management Reviews (EPMR) and Center Commissioned External Reviews (CCER), as suggested in the Pilot year, and 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.

The SC recommends that the element Quality of Research and Managerial Staff be merged with the element Quality and Relevance of Programs into a new element Quality and Relevance of Current Research, and that peer-reviewed publications be included as one indicator in the new element.

### **Indicators for Staff Quality**

The indicator "percentage of Center's top choice candidates for research and managerial positions" does not seem to provide sufficient variation to be meaningful in any particular year; it is too dependent on Center specific recruitment patterns (10 one year, 1 in another year); and it may be difficult to verify due to confidentiality of recruitment information. This kind of indicator would be better used by the EPMR which covers several years. The SC

recommends that this indicator not be included in the PM system. The awards indicator the SC considers as too much influenced by different “awards culture” in different countries, and differences between scientific fields. The SC recommends not using this indicator.

The SC feels that the definition of peer-reviewed publication must be more clearly stated. This indicator should include peer reviewed journal articles, books, book chapters and conference papers peer-reviewed prior to publishing. The SC recommends that databases used by other scientific institutions would be used also by the CGIAR to guide the inclusion of publications as peer-reviewed journal articles. The Thomson Scientific Master Journal List by Thomson ISI is a suitable international database. Furthermore, for sake of transparency and easy verification, Centers should provide a list of the publications that they include in the count.

The publications/scientist indicator should also reflect the Impact Factor of the journals in which the research is published. This requires a more complex algorithm for calculating the indicator because Impact Factors are not available for all the peer-reviewed venues that would be included in the count of publications. Not only are there no impact factor calculations for books, impact factors are not calculated for some high quality regional publications. Thus, the question of how to weight publications by impact needs to be resolved for this indicator to be an accurate representation of research quality.

The SC will prepare instructions for Centers on reporting peer-reviewed publications when the decision on the nature of this indicator has been reached.

### **Indicators for Program Quality**

The SC considers that the indicator tested in the pilot year, “Scientific papers published with developing country partners” sufficiently reflects the relevance of programs and should be kept. However, rather than reporting the actual numbers, percentage of all publications would be more appropriate. This indicator includes both peer-reviewed and non-peer-reviewed publications, and therefore inclusion/exclusion of publications is likely to be less of a problem than in the indicator on peer-reviewed publications.

The SC considers that EPMR rating of the *quality* of research programs, endorsed by the SC, is a feasible indicator. The SC is developing a set of criteria that all future EPMRs would draw from to assess quality of research in a systematic fashion and to provide a rating. These criteria will cover the quality of staff, processes and outputs. For example, awards, which the SC considers weak as an independent indicator to be included in the PM system, should be an important criterion in the EPMR rating of science quality where the EPMR panel can assess the nature and significance of the awards. The SC will need to decide on a rating scale. The SC will also need to advise when this indicator can be first used. The SC feels that while a high rating would remain over the period between EPMRs, Centers must be given an opportunity to improve a low rating through a mechanism which needs to be defined.

The SC does not recommend that CCERs be used in the PM as suggested in the pilot year. CCERs are an increasingly important input into the EPMRs, which should adequately reflect the credible judgment of the CCERs in overall EPMR assessment. The CCERs themselves are

program specific and are variable in purpose and quality and therefore not suitable for Center-level PM rating.

The SC recommends that the EPMR Panel assessment of the *relevance* of Center programs not be transformed into a PM indicator. The assessment of program relevance should take place first and foremost in the SC's assessment of Center MTPs. In the context of the PM system, relevance is assessed by the ability of the Centers to demonstrate outcomes and impacts. In general, the EPMR's in-depth assessment of all aspects of Center performance, including quality and relevance should be the main mechanism for communicating about Center performance to the CGIAR community, including influencing donors' resource allocation decisions. Because the EPMR draws on a variety of data sources, applies diverse criteria to reach conclusions, and verifies the source information, they are a more valid source of information about program quality and relevance. Thus, while EPMRs can contribute to the PM by providing a rating on research quality, their value extends well beyond the PM system, particularly in the general communication about Center performance. If the annual indicators are inconsistent with the judgments of the EPMR Panels, the validity of the indicators will be called into question.

The SC recommends that Citations per researcher, not tested but suggested to be considered as a quality and relevance of programs indicator, be included as the fourth indicator in a new element on quality and relevance of research. It feels that citation indices sufficiently reflect the Center relevance of research. This indicator is suggested for papers published in the last 6 years, relative to average by journal. The SC endorses this measure as a useful indicator as it has an advantage of being relatively easy to capture and having credibility as a measure of performance in the research community. As an outcome indicator, as initially suggested, it reflects "outcome" as the use of an output and the International Public Good nature of the results. As a program quality indicator, it reflects the recognition and visibility of Center research as used by peers. Besides the average for the journal, SC suggests as comparators norms of other Centers and alternative suppliers/peers. The mechanism and process for collecting the citation data need to be decided.

In 2004, the PMWG discussed a plan to implement further two indicators for program quality. The SC's comments on these tentative indicators are provided in Annex 1, B.

## CONCLUSIONS

The preliminary PM system includes indicators of two types: "Results" and "Potential to Perform." The SC emphasizes that only the "Results" indicators demonstrate directly how well the Centers are fulfilling their missions and producing solutions to problems towards achieving CGIAR goals. While measures of institutional and financial health and research quality and relevance may be valuable as a measure of changes in potential to perform over time, they say little about actual performance in any given year. As a result, and especially in comparison with the rich detail provided in an EPMR or the PM data on actual performance, these measures are relatively weak indicators of a Center's potential to perform, which should be taken into account in the way they are interpreted and used.

Year 2005 was a pilot year. More testing and refinement is needed before the PM system can fulfill its main purposes of facilitating Centers' performance management and enhancing accountability. As the PM system moves into its second year, the SC recommends that only a limited set of very robust indicators be included, with additional indicators considered only if those recommended above are found to be problematic.

The SC considers that an effective, transparent and timely verification mechanism needs to be developed for all indicators of the PM system, particularly as the results are used for resource allocation, among other purposes. The SC notes that invariably Centers have both well performing and less well performing sectors. Therefore the SC emphasizes that the PM is only one although an important component of the overall M&E system, where the other components, CCERs and EPMR provide more detailed information on Center program and management performance. The SC encourages using all the components of the M&E in Centers striving to improve their own performance and donors considering which Centers and components of research to support.

## Annex I

### A. SC Assessment of Tentative Outcome Indicators not Found Suitable

*% achievement towards MTP Outcome Targets for use, adoption, or influence of center outputs.* The indicator relies on Centers being held fully accountable for achieving outcomes. This is not realistic because outcomes cannot be fully predicted as in general they are into the realm where Centers have less control (relative to outputs). However, Centers should monitor outcomes across their entire research portfolio, and the cases documented by Center or someone else are reported for the suggested outcome indicator. The suggested indicator, which is based on the MTPs, allows the Center to choose outcome cases which show positive results, but over time outcome evaluation should cover all past project components.

*% of NARS/ ARI/ private sector partners reporting use of CGIAR material (collected in stakeholder survey every two years).* This is an unrealistic indicator, which would be time consuming and vulnerable to low response rate or biases. It is expected that Centers' own adoption studies and other evidence they provide for their claims of outcome should capture the adoption/use/influence by those stakeholder groups intended as users of the outputs.

*% of CGIAR trainees reporting use of skills learned in training (from trainee survey done at system level every 2-3 years).* As the previous, this is a very weak and unrealistic indicator. Trainee surveys have proven to yield poor and biased results and are costly to organize. Focus on one target group separately biases the results as this target group is not equally significant for all Centers. Furthermore, only perceptions of formal training would be obtained. The other problems with this indicator are that it assumes that (1) skills are learned in training, and (2) trainees come from and return to positions in which there are opportunities to use those skills.

*Technologies or varieties developed with CGIAR materials and released by NARS/partners.* This measure is redundant with the suggested indicator of reporting outcome. Its weakness is focus only on material outputs, when the Centers also produce capacity, practices, policy strategies and other information. If the CGIAR wants this kind of information, it should be extracted from the outcome statements, or collected separately, not as part of the PM.

### B. SC Assessment of Tentative Program Quality Indicators not Found Suitable

In 2004, the PMWG discussed a plan to implement 2 additional indicators for Quality and Relevance of Research Programs, not discussed above. The SC sees several weaknesses with these tentative indicators: Rating of relevance to partner goals, and % of budget in line with System Priorities would not be considered further.

*Rating of relevance to partner goals.* This indicator, like the stakeholder perceptions indicator discussed earlier, is vulnerable to biases due to poor response to surveys and variable definition and understanding of relevance. Stakeholders' views are better gauged through Center specific surveys as an input to EPMP and during Center's strategic planning and priority setting exercises.

*% of program budget in line with System Priorities.* The SC does not consider this a realistic indicator.

**Annex II**

**Examples of Outcome Statements**

**Table A - Example of unclear outcome statements**

<b>Criterion</b>	<b>Example</b>	<b>Explanation</b>
What output resulted in the outcome?	“[The Center] embarked on research several years ago, aimed at gauging the impact of associations of farmer participatory groups.”	This focuses on the research process or activity. A successful statement would identify the output (product) of the research.
Who used/adopted the output?	Research was conducted in 5 countries.	The statement needs to identify the countries and institutions that used the research results and in what way. They are not necessarily the same as those that participated in the research process, although there is often overlap. Indeed spillovers of outcomes beyond immediate collaborators is to be encouraged
How was the output used?	“The associations are evidently exerting a significant influence on local decisions about rural development.”	The nature of the influence needs to be specified with appropriate references/citations/testimonials
What is the evidence for this outcome?	Groups participating in the research were surveyed.	This sounds like the process of the research rather than a study of how the research was used.
What is the source of this evidence (provide a reference)?	[reference to document prepared for the PM system]	To be verifiable, the evidence for the outcome has to be detailed in some kind of formal statement like appropriate references/citations/testimonials. Ideally, this is a formal report or journal article. Statements that are prepared as communications/media briefs are not considered as evidence.

**Table B - Examples of appropriate outcome statements**

Criterion	Good Example
What output resulted in the outcome?	The new sweet potato varieties selected from CIP material specifically for dual food-feed use (for both humans and pigs); they exhibit high yield in both roots and vines, both of which are used extensively for pig feed in Vietnam. <sup>2*</sup>
Who used/adopted the output?	Farmers in the northern region of Vietnam.
How did they use the output?	The targeted farmers planted the new varieties. The use of the new varieties spread over time.
What is the evidence for this outcome?	Study surveyed 210 farm households in 8 villages in 2002 and again in 2004. High-yielding modern varieties derived from the Center had been adopted by more than 80% of all households.
What is the source of this evidence (provide a reference)?	Peters, D., T.T. Nguyen, T.H. Mai, Ty.Y. Nguyen, N.T. Pham, and K. Fuglie. 2005. Rural income generation through improving crop-based pig production systems in Vietnam: Diagnostics, interventions and dissemination. <i>Agriculture and Human Values</i> 22 (1, March): 73-85.
Criterion	Good Example
What output resulted in the outcome?	The "Proposed Framework for Transition to Integrated and Participatory Water Resources Management in the Ferghana Valley" prepared by IWMI and partners
Who used/adopted the output?	Adopted on May 16, 2003 by the ministries in charge of water resources management in the three participating countries (Uzbekistan, Kyrgyzstan, and Tajikistan).
How did they use the output?	An immediate result was the establishment of three new Canal Management Organizations (CMOs) for the Khujabaqirgan Canal in Tajikistan, the Aravan-Akbura Canal in Kyrgyzstan, and the South Fergana Canal in Uzbekistan. Creation of these CMOs, based on a partial restructuring of concerned district water management organizations, is a new management concept for these countries, initiated by the Project by recognizing the management of water resources along hydrographic (basin) boundaries instead of administrative boundaries.
What is the evidence for this outcome?	External review*
What is the source of this evidence (provide a reference)?	Integrated Water Resources Management Project in the Ferghana Valley: External Review of Phase II and Recommendations for Phase III. Prepared by PA Government Services, Washington DC, January 2005.

<sup>2</sup> Ideally, the statement would provide more detail about numbers and years of releases and the extent of use of Center materials in developing the cultivars.

**Annex III**

**Checklist for SPIA Rating of a Center's  
Ex-post Impact Assessment Study Rigor**

**1) Clear presentation of assessed research (A Necessary Condition)**

The study must adequately describe how the Center's activities have contributed to specific improvements in the relevant technologies/policy recommendations and what the costs were for the Center, and its partners-- if applicable.

Comments: \_\_\_\_\_  
\_\_\_\_\_

**2) 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 (0)

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 (0)

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 (0)

Comments: \_\_\_\_\_  
\_\_\_\_\_

**4) Realistic counterfactual (15%)**

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

- Exceptional (4)
- Good (3)
- Satisfactory (2)
- Not satisfactory (0)

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 (0)

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 (0)

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 (0)

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 (0)

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 (0)

Comments: \_\_\_\_\_

---

**7) External input (15%)**

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 Center 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 Center staff; the study is not subject to external peer review (0)

Comments: \_\_\_\_\_

---