SPIA EXTERNAL REVIEW AND QUALITY RATING SYSTEM FOR *EX-POST* IMPACT ASSESSMENTS (epIAs)

Providing quality assurance for ex post impact assessment studies done by or on behalf of the CGIAR: introducing an external review and quality rating mechanism

The rationale and the plan

The wide variation in the quality of *ex post* impact assessments coming out of the CGIAR Centers has been recognized by both donors and peers alike (Science Council, 2009). Although it exerts no direct control over the quality or publication of those *ex post* IAs, SPIA has been given a mandate to improve the quality and hence credibility of *ex post* IAs within the CGIAR. As such, SPIA, under the Strengthening Impact Assessment in the CGIAR (SIAC) program (2013-2016), is establishing an external review mechanism whereby Centers and CGIAR Research Programs (CRPs) would be encouraged to subject their *ex post* IA reports to external peer review.

The peer review process will led by SPIA Chair (Doug Gollin, Professor, Oxford) as the Chief Editor – much like a journal. In this role, the Chair will be assisted by SPIA members and activity leaders (currently, Erwin Bulte, Bob Herdt, Karen Macours and JV Meenakshi) and the Secretariat staff. The editors will use the review and rating system hosted at Editorial Manager such that a single submission gets externally reviewed and is given a 'quality rating' based on the quality and internal rigor of the study.

Submissions are voluntary, and we would expect authors of good studies – and the CRP Directors and Centers associated with those studies – to be keen to submit those studies for review. Over time, as the number of evaluated IA studies grows, this would create an expectation that all CRP or Center *ex post* IAs to get rated through this system.

SPIA has received encouraging feedback from some donors to the CGIAR (DFID, BMGF), and stakeholders (IEA, Consortium Office). We hope that this process would encourage all other donors and stakeholders to take this exercise seriously and require that all commissioned *ex post* IAs be submitted to this review process, as they clearly have an interest in being able to tell their constituencies that the impact studies are credible. A comparable type of peer review concept has been used by Australia's Council for Rural Research and Development Corporations (RDC) for evaluating outcomes from the R&D projects they fund, and although the context is slightly different (reviews are obligatory)[1], the primary purpose is similar: to provide some assurance of quality, consistency, and comparability in work undertaken to estimate the benefits/impacts from R & D projects, and thereby to increase the value of the impact evaluation work.

Aside from the obvious value to donors in establishing a rigorous evaluation of research impacts, it benefits Centers and CRPs in signalling to researchers and collaborators how research proposals and performance are to be judged by SPIA. It also gives Center- and CRP-based economists the leverage they need to argue more effectively (with colleagues and Directors alike) for the required resources for implementing more impact studies. In addition, to the extent that consistent and comparable estimates are being produced the results from multiple studies can be combined in broader program assessments

Our current thinking is that author(s) will self-classify a manuscript as qualitative, quantitative or mixed methods approach. SPIA will reject those manuscripts which do not constitute an *ex post* IA (defined in the next section). Upon acceptance, the Chief Editors or one of the Associate Editors will identify at least one external reviewer to

[1] Another difference is that the RDC evaluation program samples both successful and randomly selected projects for *ex post* evaluation, and primarily uses cost-benefit analyses to assess returns to rural industries from its R&D. It also commissions *ex ante* project evaluations – primarily to inform decision-making.

provide reviews. While we expect these individuals to hold to rigorous standards, they would not be expected to necessarily follow the same standards as a journal. For example, we would not place such a high premium on novelty of methodology *per se*, whereas other criteria perhaps not assessed so carefully by journals would be given higher weight, e.g., importance to research topic to CGIAR portfolio, clear exposition of key assumptions, evidence of use of reputable data sources, etc. This also implies that some "high quality" journal submissions may receive a "good" but not "excellent rating", and other submissions rejected by journals, for e.g., for not using novel approaches or uniquely contributing to academic literature, may still get a "fair" or "good" (or even "excellent") rating.

Note that this system of review would operate in parallel with any submissions that individuals make to journals or to Center/CRP working paper series. When that is the case, the article submitted through this review process could either be identical to the journal reviewed article or a variant thereof.

Previously, within the context of the CGIAR's annual Performance Measurement System (PMS, 2006-2010), SPIA had been involved in rating Center *ex post* IAs, using external reviewers to evaluate the rigor and quality of Center submitted *ex post* IA (2 per biennium). Some of the criteria used to evaluate those *ex post* IAs is relevant for this exercise as well, but will require some adjustment. A revised set of criteria for evaluation is proposed below.

Studies for review

For the PMS exercise SPIA, in consultation with the Impact Assessment Focal Points (IAFPs), had adopted a fairly strict definition of an *ex post* impact assessment study. We will use a similar (but slightly modified, see below) definition as to what constitutes *ex post* impact assessment and what does not. Thus, the definition used in the Instructions for the Reporting of Performance Indicators for CGIAR Centers (Science Council and CGIAR Secretariat, 2010) for *ex post* IA was:

"… a specialized area of evaluation that is designed to identify and measure consequences resulting from earlier interventions of a program or project. Its timing is ex post IA's defining characteristic: ex post IA takes place after the program's or project's investment has generated the intervention, and sufficient time has elapsed and experience accumulated to assess the intervention's performance in terms of longer term economic, social, and environmental consequences. Ex post IA contributes mainly to accountability and secondarily to learning in the evaluation of agricultural research. Impacts of an intervention may be positive or negative, direct or indirect, and intended or unintended."

As the SPIA strategy (SIAC program) now encompasses a broader swathe of activities aimed at enhancing our understanding of the R-to-D impact pathway, studies that examine carefully the causal chain of events along this pathway, e.g., especially micro-level studies using experimental or quasi-experimental methods that provide evidence on the impact of CGIAR derived technologies to adoption households or other relevant populations, will now be included in the set of studies subject to review along with the more traditional '*ex post* IAs' .

However, to avoid submissions that fall outside the purview of SPIA's mandate (*ex post* IA + micro-level impact pathway assessments), we offer the following clarification with respect to what does (and what does not) constitute an *ex post* IA (in the broader sense of the term)[2]:

- An *ex post* IA refers to a journal article, conference paper, book chapter, report or any other publication that was first drafted in 2014 or 2015 (i.e. two immediate years prior to the launch year 2016) or an early version

[2] This clarification is absolutely necessary as Centers under the PMS exercise submitted numerous other types of studies peripheral to SPIA's ex post IA mandate.

of a manuscript drafted in the current calendar year (2016) that documents empirically the impact of the CRP's or Center's research or research-related output.

- The impacts measured could be either short-term, medium-term or long-term but must be linked to a clearly discernible intervention (technology, management practice, policy) derived from CGIAR research.
- The *ex post* IA must include some measurement of adoption and some measure of *ex post* impact as a result of that adoption.
- Studies focused primarily on adoption constraints analyses, pilot technology evaluations (except those examining technology adoption to impact hypotheses), farmer preference and demand type studies and *ex-ante* impact assessments are not regarded as an *ex post* IA.
- While there may be an element of *ex-ante* in many *ex post* IAs, there must be some measurement of adoption and *ex post* impact to qualify as a bona-fide study.

Proposed criteria for evaluation of the studies

A major objective of the quality review of the IA study is to assess how well the particular claims made by the study stand up against the evidence provided, based on the narrative description, the key assumptions made, the methods, data and analyses used and the results obtained, and whether appropriate limitations of the study have been identified. In other words, the evaluation exercise specifically focuses on the rigor of the methods used in the study. The intent to evaluate the evidence in support of the claim rather than the significance and relevance of the claim (e.g. type or scale of impact).

In order to ensure consistency in the application of the evaluation methodology to the ex post IAs to be reviewed, SPIA wishes to make sure that all peer reviews undertaken adhere to a similar set of criteria in the evaluation. What follows is a set of proposed criteria that could be used by editors/reviewers in evaluating the quality of *ex post* impact assessment studies submitted to SPIA by CRPs and Centers[3]

1) Clear presentation of the assessed research and resulting innovation

The study, either internally or with supplemental information, must adequately describe how the Center's or CRP's activities have contributed to specific improvements in the relevant innovation / policy.

2) Transparent and reasonable assumptions

Are the major assumptions regarding the research assessment methodology (in all components of the analysis) transparent?

3) Plausible impact pathway

Are the intermediate steps between research output and impact carefully described? Does it seem plausible /reasonable i.e., adequately justified? Are confounding factors carefully considered?

4) Clarity of overall impact claim

Is there a good (qualitative or quantitative where possible) description of the relevant direct and indirect outcomes claimed from Center/CRP research?

5) Appropriate counterfactual

a) 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?

[3] As mentioned, this draws on a partial list of criteria SPIA used in the annual PMS, which also used an independent external peer reviewer's assessment. Some of these experts included John Antle, Tom Walker, Mitch Renkow, Daniel Karanja, Bruce Gardner, John Brennan, Suresh Pal. It also draws on the CRRDC paper on Procedures for Evaluation of R & D (January 2013).

b) For micro-level studies of technology adoption, how is self-selection into the "treatment" handled? For macro studies, are appropriate models/arguments presented that justify the counterfactual?

6) Reliable and representative data on adoption

a) Are the methods used to estimate adoption clearly described? Is the sample frame clear?
b) Are various sources of measurement error recognized and duly described?

7) Reliable and representative data on yields, incomes, other outcomes and benefit-cost analyses

a) Are the methods used to estimate productivity gains, unit cost reductions and other outcomes described clearly?
b) Is the choice of indicator for particular outcomes appropriate? Does it align well with description of the impact pathway?
c) Is the description of the methods for assessing outcomes sufficiently detailed to allow someone to replicate the study? Are there alternative sources for such estimates that have not been considered but that would represent a more rigorous and/or cost-effective approach?

8) Quality of sensitivity analysis

Has there been suitable sensitivity analysis to assess the robustness of the conclusions to changes in the underlying assumptions/parameters? Are lower-bound estimates provided (conservative scenario)?

***9) Sound attribution of benefits to research* and, if relevant, attribution to specific institutions**

a) If the study attempts to attribute the benefits of adoption of the innovation to research, is it clearly described and justified, i.e., are potential mitigating factors sufficiently addressed?
b) If the study attempts to attribute the benefits of adoption of the innovation to the particular Center or CRP, is the method for doing so clearly described and adequately justified, i.e., are potential mitigating factors sufficiently addressed?

10) Credibility of extrapolation of results

Does the study make reasonable or plausible extrapolations or generalizations to a wider target group outside the sample frame?

11) Results framed with appropriate qualifiers

a) Are the obvious limitations of the results of the study clearly explained and/or contextualized?
b) For quasi-experimental and experimental studies, have alternate treatment effects been considered and results discussed in annexures? Have authors provided a clear rationale for choice of methods?