

# **Meta-evaluations in government and government institutions:**

## **A case study example from the Australian Centre for International Agricultural Research**

Annual Conference  
Australian Agricultural and Resource Economics Society (AARES)  
8-12 February 2010

Madzivhandila T.P. <sup>1,3\*</sup>  
Griffith G.R. <sup>2,3</sup>  
Fleming, E. <sup>3</sup>  
Nesamvuni A.E. <sup>4</sup>

\*The authors are with, respectively, (1) the Agricultural Research Council, Irene, South Africa; (2) the NSW Department of Primary Industries, Armidale, Australia; (3), the University of New England, Armidale, Australia; and (4) the Limpopo Department of Agriculture, Polokwane, South Africa.

AUTHOR'S NOTE:

This paper is part of a PhD study to design a practical program evaluation model for the Limpopo Department of Agriculture, South Africa.

# **Meta-evaluations in government and government institutions: A case study example from the Australian Centre for International Agricultural Research**

Madzivhandila T.P. <sup>1,3\*</sup>, Griffith G.R. <sup>2,3</sup>, Fleming, E. <sup>3</sup> and Nesamvuni A.E. <sup>4</sup>

## **Abstract**

*In this paper we draw on impact assessment work of the Australian Centre for International Agricultural Research (ACIAR) to present an example of meta-evaluation – an evaluation of evaluations – in an agricultural research, development and extension setting. We explore quality issues relating to evaluation studies in the context of government institutions. Program evaluation standards (PES) are divided into categories of utility, feasibility, propriety and accuracy to provide a framework for the meta-evaluation. The PES are presented as a universal measure of evaluation study quality. The intent of using them here is to judge the adequacy of PES as a universal quality measure or meta-evaluation base and to extract useful insights from ACIAR program evaluation activities when developing a meta-evaluation model for the Limpopo Department of Agriculture (LDA). Our meta-evaluation is undertaken of 63 impact assessment reports. First, the literature guiding the conduct of a meta-evaluation is reviewed. Second, an assessment (the meta-evaluation) of the evaluation studies is carried out for 19 sampled reports from a population of relevant reports fitting the dimension of the analysis, and results are presented and discussed. Also, lessons learned are presented, using the framework provided by the meta-evaluation criteria. Third, taking into account the lessons learned, implications are drawn for a proposed systematic meta-evaluation of the LDA. Finally, we conclude that all the PES cannot be equally emphasized in a meta-evaluation model. At ACIAR, 70% of the standards were at least partially addressed. Therefore, we succeeded in using the PES in judging the ACIAR evaluation quality. As such, they can be an important base when developing an evaluation model but should be applied in a contextualized manner.*

## **Key Words**

Meta-evaluation; Evaluation Quality; Program Evaluation Standards; Evaluation Model; Australian Centre for International Agricultural Research; Limpopo Department of Agriculture (South Africa)

---

\* The authors are with, respectively, (1) the Agricultural Research Council, Irene, South Africa; (2) the NSW Department of Primary Industries, Armidale, Australia; (3), the University of New England, Armidale, Australia; and (4) the Limpopo Department of Agriculture, Polokwane, South Africa.

## **AUTHOR'S NOTE:**

This paper is part of a PhD study to design a practical program evaluation model for the Limpopo Department of Agriculture, South Africa.

## 1. Introduction

Every evaluation finding permeates social processes through discussion, dialogue and negotiations, which may be influential to decision making, shaping policies and other kinds of effects (Valovirta 2002). In part, evaluation findings (sometimes called results or evidence) are important for results- or evidence-based social betterment initiatives, performance management and improvement (Blalock 1999; Davies 1999; Wholey 1986, 2001). Findings are more relevant when evaluation is instituted on the premise of organizational learning (Preskill and Torres 1999b; Preskill and Torres 1999a, 2000, 2001; Russ-Eft and Preskill 2001). That is, the way evaluation findings influence social betterment initiatives can be jeopardized when evaluation studies are flawed (Cooksy and Caracelli 2005; Henry 2003). Therefore, in pursuit of rigorous findings, the concern with evaluation quality drives discussions (Cooksy and Caracelli 2009). Despite the importance of getting evaluation right, meta-evaluations – evaluations of an evaluation – reported in the literature are rare (Hanssen et al. 2008). Fitzpatrick et al. (2004) also confirmed the lack of meta-evaluation and recommended its increased use to improve evaluation conduct and practice.

Against this backdrop, we posit that meta-evaluation is the most relevant way to ensure the quality of evaluation findings. It allows one to establish evaluation quality and other details of factors contributing to the influence (or non-influence) of the evaluation's process and findings over time (Oliver 2009) and contributes to continuous improvement of this process. It serves specific functions, such as to provide quality assurance and control, or learning when conceptualized as a part of an evaluation endeavour (Widmer et al. 2007). In this way, meta-evaluation is a powerful tool for increasing the effectiveness and quality of the evaluation work (Henry 2001), and potentially also increasing the use of evaluation findings. According to Fitzpatrick et al. (2004), this is why meta-evaluation is important. Patton (1997) emphasizes the motive for conducting a meta-evaluation to ensure an independent and credible review of an evaluation's strengths and weaknesses (see also Uusikyla and Virtanen 2000). It is a procedure for describing an evaluation activity and judging it against a set of ideas concerning what constitutes good evaluation (Stufflebeam 1974 ). Meta-evaluations are important to control bias (Fitzpatrick et al. 2004; Stufflebeam 1974 2001; Stufflebeam and Shinkfield 2007). They are also conducted to improve subsequent evaluation (Lynch et al. 2003), to address the defensibility of an evaluation's process and results using commonly agreed standards.

Stufflebeam (2001) and Hanssen et al. (2008) argued strongly for increased use of meta-evaluation, either formatively or summatively, to describe and judge information about the strengths and weaknesses of an evaluation. We identified studies using meta-evaluation to: (i) describe aggregated information from several individual evaluations (normally called meta-analysis) (e.g., Ashworth et al. 2004); (ii) systematically control the quality of evaluation studies (e.g., Bustello 2003; Cooksy and Caracelli 2009); or (iii) perform both functions. Studies like Widmer et al. (2007) and Scott-Little et al. (2002) integrate both uses when assessing quality (especially the methodology) of existing evaluation studies while also synthesizing the results of the studies.

Nilsson and Hogben (1983) noted conflict in the literature with regard to what constitutes evaluation quality. Program evaluation standards (PES) (Joint Committee for Programme Evaluation 1994) are central to the discussion of what constitutes quality when conducting meta-evaluations. They are suggested as commonly agreed criteria for evaluation quality in the literature. Examples of those arguing for the use of PES universally include: text books (see Fitzpatrick et al. 2004:444-447; Russ-Eft and Preskill 2001:113-119; Wingate in press) with the exception of Stufflebeam and Shinkfield (2007); and journal papers (Beywl and Speer 2004;

Laubli Loud 2004, for example). In Australasia, literature arguing for extensive use of evaluation standards include Fraser (2001b, 2001a, 2004) and Chatterji (2005), even though they argue for contextualized individual country standards. A manuscript edited by Russon (2000) noted low use PES world-wide, especially in developing countries. Russon's edited work provides the argument for using evaluation standards when designing a meta-evaluation.

## **2. Underpinning Literature**

### **2.1. Meaning of meta-evaluation**

The term meta-evaluation was first introduced by Scriven (1969) but has been confused with meta-analysis, accorded different definitions and used in many different ways (Uusikyla and Virtanen 2000). The definition we use is a systematic review of evaluations to determine the quality of their processes and findings (Bickman 1997; Cooksy and Caracelli 2005, 2009). According to Stufflebeam (2001:183) and Stufflebeam and Shinkfield (2007:651), this determination of quality is the 'process of delineating, obtaining, and applying descriptive information and judgmental information about an evaluation's utility, feasibility, propriety, and accuracy and its systematic nature, competence, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and publicly report its strengths and weaknesses'.

Meta-evaluation differs from meta-analysis, which is applied to synthesize findings from multiple studies (Boruch and Petroino 2004; Rossi et al. 2004). Meta-evaluation evaluates the quality of one or more evaluation studies whereas meta-analysis aggregates and summarizes the findings of several evaluation studies qualitatively or quantitatively. Weiss (1998:48) defined meta analysis as 'the systematic summary of the results from a number of different evaluations of the same kind of programs'. Even though Pawson and Tilley (1997) stress the importance of understanding the role of explanatory mechanisms and contexts (what they call context-mechanism-outcome configuration), the aim of meta-analysis is to provide more estimates of the size of the policy impact (Rossi et al. 2004).

### **2.2. The Types and Purpose of Meta-evaluation**

Meta-evaluation can be used to assess the quality of a single study or a set of studies in different ways. Literature identifies two types of meta-evaluations. First, formative meta-evaluations assist evaluators to plan, conduct, improve, interpret, and report their evaluation studies. Second, summative meta-evaluations – conducted following an evaluation – help audiences see an evaluation's strengths and weaknesses, and judge its value (Cooksy and Caracelli 2005, 2009; Greene 1992; Joint Committee for Programme Evaluation 1994; Stufflebeam 2001). Hanssen et al. (2008) call them proactive meta-evaluation, which is designed to help evaluators before and during conducting an evaluation, and retroactive meta-evaluation, which is designed to help audiences judge completed evaluations (see also Stufflebeam and Shinkfield 2007).

The main purpose of formative meta-evaluation is to reveal deficiencies in the primary evaluation at a time when they can still be addressed, thus preventing the determination and dissemination of invalid conclusions and increasing the primary evaluation's utility and cost-effectiveness. It takes place while an evaluation is underway in order to provide guidance for improvement (Fitzpatrick et al. 2004; Stufflebeam and Shinkfield 2007). The purpose of summative meta-evaluation is to validate a primary evaluation. It adds credibility to it and enhances users' confidence in the evaluation findings to inform decisions to expand, modify or cancel programs. When a summative meta-evaluation finds serious flaws in a primary evaluation, it can prevent decision makers from taking actions based on faulty information (Wingate in press). It assesses the quality of a completed evaluation, increasing the appropriateness of evaluation processes

and validity of its conclusions (Fitzpatrick et al. 2004; Stufflebeam and Shinkfield 2007). Just as evaluation can improve programs and contribute to knowledge in the field, conducting meta-evaluations can improve quality both formatively and summatively, thereby enhancing the evaluation practice (Hanssen et al. 2008).

### **2.3. Meta-evaluation Conduct Guidelines**

The evaluation quality criteria of a meta-evaluation determine the extent to which an evaluation conforms to commonly held requirements. Many criteria are proposed in the literature. For example:

- According to Stufflebeam (1974), meta-evaluation should: assess the merit; serve the decision making and accountability; assess goals, designs, implementation, and results; provide descriptive and judgmental information and appropriate recommendations; serve all persons who are involved in and affected by the evaluation studies being evaluated; be conducted by both insiders and outsiders; be a process of delineating the questions to be addressed, obtaining the needed information, and using the information in decision making and accountability; and be technically adequate.
- According to Patton (1997) the questions on which to focus a meta-evaluation should include: 'Was the evaluation well done? Is it worth using? Did the evaluation meet professional standards and principles?' Similarly, Scriven (2007) argued that a meta-evaluation can be aided through the use of a key evaluation checklist or standards such as PES (Joint Committee for Programme Evaluation 1994). Patton (1997) further suggested that the use of meta-evaluation should hinge on the politicized context for the evaluation.
- Wholey et al. (2004) noted that an evaluation (or meta-evaluation) should entail a description of resources, activities, process and results caused by the program. In addition, it should identify issues like: relevancy, effectiveness, efficiency and sustainability of the program.
- Bollen et al. (2005) suggested a meta-evaluation review should focus on three general and necessary ingredients to quality evaluations: (a) information on inputs, (b) information on results, and (c) controls for confounding factors. Their argument is that a quality evaluation collects and reports sufficient information on the first two ingredients and addresses the threats to any conclusion of impact presented by the confounding factors. They added confounding factors are an important criterion of quality in any evaluation study.
- Other authors suggest evaluation quality assessment criteria based on the focus of their work. For example: (i) Kirkhart (2000), Christie and Alkin (2003) and Patton (1997) argue that the importance and usefulness of evaluations is when they are utilized; (ii) Cousins and Whitmore (1998) and Weaver and Cousins (2004) emphasize stakeholder participation; (iii) Martens (1999, 2008) argue for consideration of context and the importance of proper communication of evaluation results; (iv) Weiss (1988, 1993, 1999) and Chelimsky (1998) warn about the political nature of evaluation, and the importance of context and actors; (v) Scriven (1995) emphasizes the centrality of systematic judgment in evaluation; and (vi) Chelimsky (1998) highlights the need for evaluation credibility and timeliness of the findings.

### **2.4. Program Evaluation Standards**

There must be a shared understanding and agreement about what constitutes a good evaluation (Fitzpatrick et al. 2004). The PES is considered one of the most importance sources of criteria to achieve this shared understanding and agreement (Fitzpatrick et al. 2004; Lynch et al. 2003; Patton 1997; Russ-Eft and Preskill 2001; Stufflebeam 2001; Stufflebeam and Shinkfield 2007; Wingate in press). They are an explicit set of generally agreed standards of quality evaluation (Joint Committee for Programme Evaluation 1994) that can be useful not only as a framework for the design but also as an assessment of particular evaluation quality (Beywl and Speer 2004). The PES promote consistency in practice and diminish the potential for evaluations to be judged on personal criteria that do little to advance the evaluation profession or educate evaluation

consumers about what constitutes a sound evaluation. This avoids judgments about the quality of evaluations hinging on methodological or theoretical preferences, personal opinion, or arbitrary criteria (Joint Committee for Programme Evaluation 1994). However, there are other opposing views about the applicability of PES as a universal quality measure (Chatterji 2005). The most prominent contest is about the unsuitability of using PES in contexts other than the United States and Canada (Stufflebeam and Shinkfield 2007). But generally, the standards are intended to serve as (i) guides for designing and carrying out sound evaluations and stimulating the use of evaluation findings in appropriate ways; (ii) resources for teaching clients/stakeholders about the purposes for evaluations and what they can expect from evaluative efforts; (iii) a framework for conducting meta-evaluations, or appraisals of the quality of evaluation practices in given projects and programs; (iv) resources in proposal development for developing and evaluating new programs or projects; and (v) guiding criteria for assessments of evaluator knowledge and credibility (Chatterji 2005; Joint Committee for Programme Evaluation 1994).

The PES comprise 30 standards (Joint Committee for Programme Evaluation 1994). Their elements are not discussed individually in detail here, but a summary is presented in Table 1. The general intent of each group of guidelines is described as:

- Utility standards (U1-U7) are intended to ensure that an evaluation will serve the information needs of the intended audience and users. These standards deal with identification of relevant stakeholders, formulation of evaluation questions to address stakeholder information needs, and the usability, clarity, and timeliness of the reports for stakeholders and clients. Evaluation impact is also addressed under the utility standards.
- Feasibility standards (F1-F3) are intended to ensure that an evaluation is designed and conducted in a manner that is prudent, practical, diplomatic, and cost effective. These standards acknowledge the social and political context in which social programs and institutions reside; they stipulate that evaluations be conducted in politically viable ways.
- Propriety standards (P1-P8) are intended to ensure that evaluations are conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by the results. These standards deal with respecting the rights of human subjects, compliance with agreements about the confidentiality of information gathered, the appropriate release of results, and so on.
- Accuracy standards (A1-A12) deal with methodological rigor and the technical adequacy of information on the product, program, institution or service area that is evaluated. They are intended to ensure that quantitative and qualitative procedures are credible and that the information gathered, analyzed and conveyed about various aspects of the program is technically defensible.

Table 1: Joint committee of evaluation standards for ethical consideration (Joint Committee for Programme Evaluation 1994)

<b>Utility</b>	<b>Feasibility</b>	<b>Propriety</b>	<b>Accuracy</b>
<p><b>U1:</b> Stakeholder identification so that their need are addressed</p> <p><b>U2:</b> Evaluator credibility (should be trustworthy and competent)</p> <p><b>U3:</b> Information scope and selection should be broad to address pertinent issues</p> <p><b>U4:</b> Values identification should be based on rational procedures</p> <p><b>U5:</b> Report clarity with essential easy to understand information</p> <p><b>U6:</b> Report timeliness and dissemination to intended users</p> <p><b>U7:</b> Evaluation impact through follow-through by stakeholders</p>	<p><b>F1:</b> Practical procedures to minimise disruption when obtaining information</p> <p><b>F2:</b> Political viability with anticipation of different positions of various interest groups</p> <p><b>F3:</b> Cost effectiveness given its cost and benefits justification</p>	<p><b>P1:</b> Service orientation to assist to address and effectively serve the needs of the full range of participants</p> <p><b>P2:</b> Formal agreements agreed to in writing</p> <p><b>P3:</b> Rights of human subjects designed and conducted to protect their welfare</p> <p><b>P4:</b> Human interactions respecting human dignity so that participants are not threatened or harmed</p> <p><b>P5:</b> Complete and fair assessment in its examination and recording strength and weaknesses</p> <p><b>P6:</b> Full disclosure of findings and limitations should be made accessible</p> <p><b>P7:</b> Conflict of interest should be dealt with openly and honestly</p> <p><b>P8:</b> Fiscal responsibility and accountability when allocating expenditure</p>	<p><b>A1:</b> Clear and accurate program documentation</p> <p><b>A2:</b> Context analyses should be examined in enough detail</p> <p><b>A3:</b> Purpose and procedures should be described in enough detail</p> <p><b>A4:</b> Defensible information sources should be described in enough detail</p> <p><b>A5:</b> Valid information gathering procedures chosen and implemented to reflect true picture</p> <p><b>A6:</b> Reliable information is sufficient for the intended use</p> <p><b>A7:</b> Systematic collected, processed and reported information</p> <p><b>A8:</b> Analysis of quantitative information should be systematically analysed</p> <p><b>A9:</b> Analysis of qualitative information should be systematically analysed</p> <p><b>A10:</b> Justified conclusions should be explicitly justified Impartial</p> <p><b>A11:</b> Reporting should guard against distortion caused by personal feelings and biases</p> <p><b>A12:</b> Meta-evaluation should be formatively and summatively evaluated</p>

## **2.5. Point of departure**

We adopt the view that meta-evaluation is about evaluating evaluation studies based on the profession's standards and principles (Patton (1997), and that meta-evaluation provides a base for assessing the value of an evaluation (Stufflebeam, 2001; Stufflebeam and Shinkfield, 2007). Our hypothesis is that the PES can be appropriate in different contexts and should form the basis of a meta-evaluation. They are commonly agreed to be good evaluation practice (Fitzpatrick et al. 2004) but have their roots in the United States which has a history of evaluation since the early 1960s, mainly in the context of education and health initiatives. Therefore, they are not necessarily suitable for every context outside their origin. Even organizations with a long history of evaluation in developed and developing countries such as ACIAR (chosen for our case study) fall short of using all 30 PES. As a result, it is important for an organization new to evaluation conduct and practice (here LDA) to assess the appropriateness of each of the PES. The checklist in Table 1 is used to judge the adequacy of evaluation designs and reports at ACIAR (see Table 2) to help us benchmark what ACIAR – as an organization in the same field of agriculture (even though in a different context) like the LDA – has been doing to deal with their evaluation quality issues. Based on the empirical evidence, we suggest what LDA can do for their meta-evaluation based on acceptable evaluation profession quality standards.

Our case study context and premise for describing meta-evaluation is of interest for four main reasons:

- Little work has been done on meta-evaluation in the South African public service, particularly LDA. Meta-evaluation is important for a learning organization and it is fundamental for achieving continuous improvements in evaluation quality.
- ACIAR has long been interested in using evaluation to improve their service delivery. This kind of institutional commitment is relatively rare among public institutions. Therefore, important lessons can be learned from their experience.
- The work of both LDA and ACIAR is in agricultural research, development and extension. Both organizations have almost similar programs related to research and development (training included); livestock systems (production, health and fisheries); crop systems (crop production and horticulture), natural resource management (land and water resources, soil management and crop nutrition, and forestry); and economics and agribusiness. Their programs are geared towards sustainability and social betterment in agricultural industries.
- Agriculture is advocated as a pro-poor strategy, playing an important role in economic development and poverty reduction, here called social betterment (Diao et al. 2005, 2007; Meijerink 2005).

Empirical works of this type can provide evaluators, decision makers and policy communities with valuable descriptions of the conduct and practice of meta-evaluation (Bustello 2003; Cook and Gruder 1978; Cooksy and Caracelli 2009; Scott-Little et al. 2002). Such descriptions can provide an important contribution to evaluation practice. Evaluators, decision makers and policy makers can learn what to examine, what methods and instruments to use, whom to talk to and whom to listen to (Fitzpatrick et al. 2004). Following the case study example, we explore some meta-evaluation issues that can be used to enhance evaluations and the meta-evaluation goal of LDA.

## **3. Description of the Evaluand: Evaluation of Impact at ACIAR**

ACIAR is an Australian government statutory authority that operates as part of Australia's aid program within the portfolio of Foreign Affairs and Trade. It contributes to the aid program objectives of advancing Australia's national interest through poverty reduction and sustainable



development (Australian Centre for International Agricultural Research 2009b), what we call social betterment. From its beginning, ACIAR has placed significant emphasis on assessing the impact of the research, development and extension service it funds, particularly focusing on quantifying the returns to research investments (Australian Centre for International Agricultural Research 2009a). It has used these assessments to account to stakeholders and to support improved decision making and management of its funds. ACIAR has for the past 20 years been undertaking formal, independent impact assessment studies (Harding et al. 2008). It therefore has a long history of impact assessment that provides valuable lessons for improving the selection, design and delivery of projects, as well as demonstrating the value of ACIAR as part of Australia's international development assistance program (Harding et al. 2008).

The Impact Assessment Program currently runs two types of finished project evaluations of ACIAR projects (Australian Centre for International Agricultural Research 2009a). The first is a set of 'adoption studies', which is undertaken on all large projects three years after their completion. The second finished project evaluation, which is the focus of our paper, is an impact assessment study. These impact assessment studies are undertaken by external reviewers and involve extensive review of project impact and adoption in the partner country and Australia. The impact assessments provide estimates of the returns to the research investment on a project or suite of projects. Harding et al. (2008) further mentioned that ACIAR has recently developed a database for impact assessment that facilitates systematic analysis of the impact assessment and provides a summary of information, which we have used for our study data.

Within ACIAR, impact assessment aims to identify, provide evidence of and, ultimately, quantify the impacts of its R&D investments. ACIAR's impact assessment activities provide an 'after the event' perspective within the comprehensive monitoring and evaluation (M&E) process it has in place. The impact assessment reports provide project impact information to guide future research activities. The main focus of these commissioned reports is to measure the returns on investment to agricultural research, mainly thorough cost-benefit analysis (CBA) (Davis et al. 2008). However, the emphasis is also given to analyzing the impacts of projects on poverty reduction. When analyzing impact, few studies quantify the types and levels of impact at the beneficiary level (Davis et al. 2008). They provide accountability to stakeholders, as well as a clear measure of the returns to the funds ACIAR invests. Increasingly also, impact assessments provide a basis for improving the research selection process by acting on lessons learned from past projects; and are a valuable learning tool for project participants and project managers.

In their 2005 review process (Raitzer and Lindner 2005), the meta-evaluation or meta-analysis was based on a framework of principles, criteria, and indicators for study credibility, which is derived from a selective review of the methodological literature. Two overarching principles underpinned this review framework—'transparency' and 'analytical rigour'. Transparency embeds three criteria: (i) clearly derived and explained key assumptions; (ii) comprehensive description of data sources; and (iii) full explanation of data treatment. Analytical rigour includes: (i) representative data set utilised; (ii) appropriate data treatment; (iii) plausible counterfactual scenario developed; (iv) adequate consideration of mission-relevance of economic benefits; and (v) plausible institutional attribution (see also Raitzer 2003).

#### 4. Data and Method

This paper reviews the ACIAR impact assessment (evaluation) studies. The database contains impact assessment publications IAS 01 to IAS 63. This impact assessment reports are publicly available at <http://www.aciar.gov.au/publication/term/25>. The first publication was in 1998 and the most recent was in 2009. Before sampling from this database, the first step was to find out whether the dimension of each study can be assessed with the information it contains. After this process, a population of 56 impact assessment study reports remained from which to select our sample. In the second step, these reports were listed in chronological order to form a sampling frame. A random selection procedure was used to ensure that each study had an equal opportunity to be selected. A systematic random selection process was followed (see Babbie 2007:211), and every third report in the population was chosen (a sampling ratio of 35%) to provide a sample of 19 evaluation reports for analysis (see Table 2).

The cornerstone of our evaluation is the choice of PES as assessment criteria, and so we devised, in the third step, a template using the PES criteria to assess the quality of the evaluations undertaken or commissioned by the organization. In the fourth step, using the assessment criteria, a coding sheet was developed to capture information about the nature of the ACIAR programs and basic information about the evaluations. A coding sheet was piloted on two reports in the fifth step to revise the original version. Lastly, the data were summarized across all 19 impact assessment study reports to yield frequency counts for each assessment criteria category. Other studies ( Scott-Little et al. 2002; Widmer et al. 2007) have followed a similar procedure.

Guidance for conducting a meta-evaluation using PES is found throughout the evaluation literature (Hanssen et al. 2008). The purpose is to determine the degree to which each of the evaluation reports addressed the standards established by the Joint Committee on Standards for Program Evaluation. As explained before, the standards are decomposed into four important attributes of an evaluation: utility, feasibility, propriety, and accuracy (see Table 1 for a complete listing). For data analysis, the four-point rating scale as used by the Joint Committee for Program Evaluation (1994) and Scott-Little (2002) consisted of the following:

- a) The standard was addressed.
- b) The standard was partially addressed.
- c) The standard was not addressed.
- d) The standard was not applicable (this category was expanded to include both 'not applicable' and 'unable to judge' when insufficient information was provided in the report).

Table 2: A sample of evaluation reports for meta-evaluation assessment

Number	Title	Report No.	Project No.	Sector	Partner Country	Author(s) (Publication yr)
1	Control of Newcastle disease in village chickens	IAS 01	ACIAR 8334 ACIAR 8717 ACIAR 93/222	Livestock	Malaysia Philippines Sri-Lanka Indonesia Thailand	Centre for International Economics (1998a)
2	Establishment of a protected area in Vanuatu	IAS 03	ACIAR 9020	Natural Resource Management	Vanuatu	Centre for International Economics (1998b)
3	Pigeonpea improvement	IAS 06	ACIAR 8811		Fiji Indonesia India Thailand	Ryan (1998)
4	Australian tree species selection in China	IAS 08	ACIAR 8457 & 8848	Forestry	China	McKenney (1998)
5	Post harvest R&D concerning tropical fruits	IAS 11	ACIAR 8356 & 8844	Horticulture	Thailand	Chudleigh (1998)
6	Improved drying of high moisture seeds	IAS 14	PHT/1983/008 PHT/1986/008 PHT/1990/008	Crops	Thailand	McLeod et al. (1999)
7	Breeding and feeding of pigs in Australia and Vietnam	IAS 17	AS2/1994/023	Livestock	Vietnam	Tisdell and Wilson (2001)
8	Improved methods in diagnosis, epidemiology, and information management of foot-and-mouth diseases in South-East Asia	IAS 21	ASI/1998/067 ASI/1998/035 ASI/1992/004 ASI/1994/038	Livestock	Thailand Laos China	McLeod (2003)
9	Assessment of rodent control projects in Vietnam: Adoption and impact	IAS 24	ACIAR/1998/036 AusAid 2000/024 World vision VN31-174945	Livestock	Vietnam	Palis et al. (2004)
10	Acacia hybrids in Vietnam	IAS 27	FST/1986/030	Forestry	Vietnam	Van Bueren (2004a)
11	Eucalypt tree improvement in China	IAS 30	FST/1990/044 FST/1994/025 FST/1984/057 FST/1987/036 FST/1988/048 FST/1996/125 FST/1977/077	Forestry	China	Van Bueren (2004b)
12	Identifying the sex pheromone of the sugarcane borer moth: Economic impact	IAS 34	CS/1991/680	Horticulture (Pest Management)	Papua New Guinea	Pearce (2005)
13	Management of fruit flies in Pacific	IAS 37	CS2/1989/020 CS2/1994/003 CS2/1994/115 CS2/1996/225	Disease Management (Horticulture)	Pacific Region	McLeod (2005)

14	Water management in public irrigation schemes in Vietnam	IAS 43	LWR2/1994/004 LWR1/1998/034	Natural Resource Management	Vietnam	Harris (2006)
15	Mite pest in honey bees in the Asia Pacific region	IAS 46	AS2/1990/028 AS2/1994/017 AS2/1994/018 AS2/1999/060	Disease Management (Apiary)	Philippines Indonesia Papua New Guinea	Monck and Pearce (2007)
16	Minimising impacts of fungal diseases of eucalypts in South-East Asia	IAS 49	FST/1994/041	Disease Management (Forestry)	Vietnam Thailand	Fisher and Gordon (2007)
17	Impact of increasing efficiency and productivity of ruminants in India by the use of protected-nutrient technology	IAS: 53	AH/1997/115	Livestock	India	Monck and Pearce (2008)
18	A review and impact assessment of ACIAR's fruit-fly research partnerships 1984-2007	IAS 56	CS2/1983/043 CS2/1989/019 CS2/1989/020 CS3/1994/003 CS2/1994/115 CS2/1996/225 CS2/1997/101 CS21998/005 CS2/2003/036 CP/2007/002 CP/2007/187 PHT/1990/051 PHT/1993/083 PHT/1994/133 PHT/2001/027 PHT/2002/086	Disease Management (Horticulture)	Bhutan Cook Islands Fiji Federal States of Micronesia Indonesia Malaysia Papua New Guinea Philippines Samoa Solomon Islands Thailand Tonga Vanuatu Vietnam	Lindner and McLeod (2008)
19	Two stage grain drying in the Philippines	IAS 59	PHT/1983/008 PHT/1986/008 PHT/1990/008	Grain Crops	Philippines	Chupungco et al. (2008)

## 5. Results

The results were analyzed in two ways: first, to determine the degree to which each of the individual standards was addressed across all the reports and, second, to determine the degree to which the individual reports addressed the four categories of standards. The purpose was to look for patterns where standards were addressed in these evaluation reports (see Table 3) and then gauge the degree to which the individual reports addressed the standards (see Figure 1).

In the first approach, as shown in Table 3, many of the standards were addressed in the ACIAR impact assessment reports.

Table 3: Frequency of ratings by standard

Standard	Addressed	Partially addressed	Not addressed and Unable to judge
▪ U1: Stakeholder identification*			19
▪ U2: Evaluator credibility	19		
▪ U3: Information scope and selection*	19		
▪ U4: Values identification*	19		
▪ U5: Report clarity*	19		
▪ U6: Report timeliness and dissemination*		19	
▪ U7: Evaluation impact*		19	
▪ F1: Practical procedures	19		
▪ F2: Political viability		4	15
▪ F3: Cost effectiveness			19
▪ P1: Service orientation*	16	3	
▪ P2: Formal agreements			19
▪ P3: Human Interactions*		2	17
▪ P4: Rights of human subjects			19
▪ P5: Complete and fair assessment*	19		
▪ P6: Full disclosure of findings*	19		
▪ P7: Conflict of interest	19		
▪ P8: Fiscal responsibility			19
▪ A1: Program documentation*	19		
▪ A2: Context analyses*	12	7	
▪ A3: Described purpose and procedures*	16	3	
▪ A4: Defensible information sources*	19		
▪ A5: Valid information gathering procedures	18	1	
▪ A6: Reliable information	18	1	
▪ A7: Systematic information	19		
▪ A8: Analysis of quantitative	19		
▪ A9: Analysis of qualitative	1	1	17
▪ A10: Information justified*	19		
▪ A11: Impartial reporting*	19		
▪ A12: Meta-evaluation*			19

\* Standards that are relevant to reporting (see Joint Committee for Programme Evaluation 1994; Scott-Little et al. 2002)

The standards that were almost fully addressed are: evaluator credibility (U2), information scope and selection (U3), values identification (U4), report clarity (U5), Practical procedures (F1), service orientation (P1), complete and fair assessment (P5), full disclosure of findings (P6), conflict of interest (P7), program documentation (A1), prescribed purpose and procedures (A3), defensible information sources (A4), valid information gathering procedures (A5), reliable

information (A6), systematic information (A7), analysis of quantitative (A8), information justified (A10), and impartial reporting (A11)

As shown by the asterisks in Table 3, the Joint Committee identified 17 of the 30 PES that were particularly relevant to judging the technical quality of evaluation reports (Joint Committee for Programme Evaluation 1994; Scott-Little et al. 2002). We categorized only three of these standards (stakeholder identification (U1), human interaction (P3) and meta-evaluation (A12)) as very poor compliance (not addressed or unable to judge). We also observed low compliance with report timeliness and dissemination (U6) and evaluation impact (U7) with no report fully addressing the standard. Also, the standard of context analysis (A2) was only partly addressed by 7 out of 19 reports (37%), with the other 12 reports fully addressing this standard. Otherwise, the standards particularly relevant to judging the quality of evaluation reports were met in a large majority of the reports.

In the second approach, we judged the quality of the evaluation studies by defining a satisfactory score. Judging if the four attributes of utility, feasibility, propriety and accuracy were addressed in the reports, we assigned the satisfactory level using the aggregated PES category rather than by individual standard.

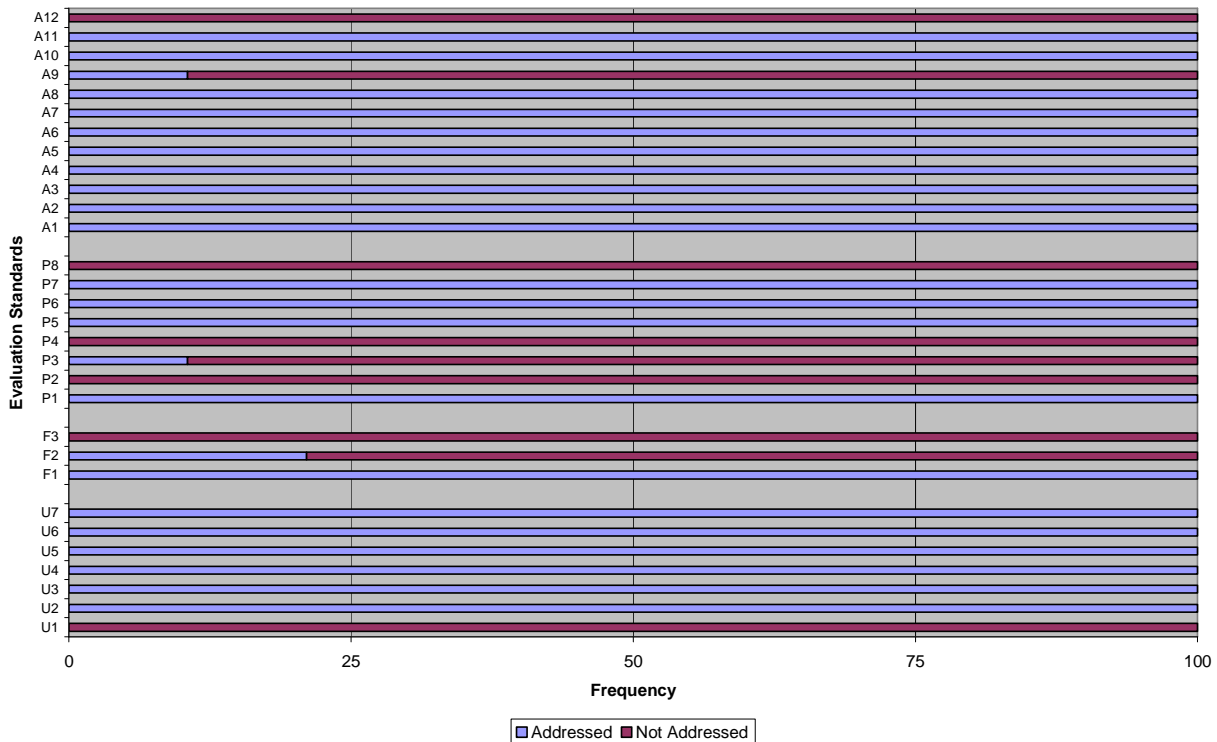


Figure 1: A diagram representing frequency of standards identified in each report

From Figure 1, the following is the presentation of satisfactory level rating:

- Utility: 86% of the standards within the category were at least partially addressed.
- Feasibility: 33% of the standards within the category were at least partially addressed.
- Propriety: 50% of the standards within the category were at least partially addressed.
- Accuracy: 83% of the standards within the category were at least partially addressed.
- Total PES: 70% of the standards were at least partially addressed.

The ACIAR impact assessment studies satisfied standards in two categories, utility and accuracy. The design of their evaluation is intended to ensure that: (i) an evaluation will serve the information needs of the intended users (utility); and (ii) an evaluation will reveal and convey technically adequate information about the features that determine the worth of the program. The ACIAR evaluation was satisfactory on the propriety standard, which is intended to ensure that an evaluation will be conducted legally, ethically and with regard for the welfare of those involved in the evaluation, as well as those affected by results. It performed unsatisfactorily with the feasibility standard, which is intended to ensure that an evaluation will be realistic, prudent, diplomatic and frugal.

## **6. Discussion**

### **6.1. The overall evaluation design**

The ACIAR impact assessment studies are standardized with the guidelines providing the basis for ensuring consistency. The overall evaluation design and reporting is guided by the guidelines for assessing the impacts of ACIAR's research activities (Davis et al. 2008). In addition, the ACIAR database for impact assessments provides an outline of the database structure and a guide to its impact evaluation operation (Centre for International Economics 2009). The effect of a highly standardized format is evident from the results of this study. We observed that standards are most likely to be fully or partially addressed in all the impact assessment reports, but there are a few exceptions. When designing a meta-evaluation model, standards guidelines for database structure and operations, and impact assessment can be valuable for LDA.

### **6.2. Did program evaluation standards apply?**

The focus of our discussion is aligned mainly towards those standards that are partially addressed, not addressed or in the unable to judge category. Our reasoning is that learning and improvement are an important part of meta-evaluation principles (Preskill and Torres 1999b; Preskill and Torres 1999a, 2000, 2001; Russ-Eft and Preskill 2001).

#### *Did utility standards apply?*

The result showing that 86% of the standards within the utility category were at least partially addressed proves that the utility standard can apply when designing a meta-evaluation model. Stakeholder identification (U1) was the only standard not addressed, even though responsive evaluation recently has been identified as a major alternative to more traditional approaches to program evaluation (Maxwell 1984). Other than responding to ACIAR's evaluation needs, as the commissioner of the impact assessment studies and intended user of the results, no other stakeholders' needs were indicated in all the reports. The studies are used by ACIAR mainly to guide their investment decisions even though reports are widely distributed. The lead researcher(s) in most reports acknowledged the contribution of partner organizations in Australia and in partner countries.

#### *Did feasibility standards apply?*

ACIAR performed poorly with regard to this standard with only one of the three standards at least partially addressed. This category of standards did not apply for ACIAR and we suggest their revision for the LDA meta-evaluation. Scott-Little et al. ((Scott-Little et al. 2002) also observed difficulty in the application of these standards. First, the political viability (F2) which observes the different positions of interest groups did not apply. The reason we attribute to this result is that their impact evaluation is more aligned towards CBA. Their CBA is rooted in the limited use of primary data from the partner country where the work was done. Second, the cost effectiveness of the evaluation studies was not presented in all the reports. In most cases, the

cost of undertaking an evaluation is a contractual issue which is hardly disclosed and the benefit of the impact evaluation study difficult to quantify.

*Did propriety standards apply?*

Using our 75% satisfactory level, propriety standards did not apply for the ACIAR given that only one-half of the standards within this category were at least partially addressed. Beywl (2000) and Scott-Little et al (2002) observed a similar result, positing the reasons that propriety standards deal mainly with formal agreements (P2), consent and respect for human subjects (P3 & P4), and fiscal responsibility and accountability when allocation impact assessment budget (P8). In most cases, this type of information will not be addressed explicitly in the reports (Beywl 2000; Scott-Little et al. 2002). Therefore, the violations, if any, were marginal. The assumption is that evaluators will make reasonable efforts to meet the P3 and P4 standards by engaging in ethical practices.

*Did accuracy standards apply?*

Our results showed that 10 of the 12 standards within this category were at least partially addressed. The exceptions were systematic analysis of qualitative information (A9) and meta-evaluation (A12). Davis et al. (2008) noted the high cost of collecting primary qualitative impact data and the frequent use of meta-evaluation to improve conduct and practice of organization evaluation. Even though meta-evaluation was not addressed in individual study reports, it is used at ACIAR (see Raitzer and Lindner 2005). The presentation of accuracy standards acknowledges that evaluators could potentially gather both quantitative and qualitative data (Joint Committee for Programme Evaluation 1994). ACIAR also acknowledges that situation in their impact assessment guide (Davis et al. 2008) and database structure and operations framework (Centre for International Economics 2009). But the use of quantifiable qualitative data was not evident. The accuracy standards did apply and can be used when designing a meta-evaluation with ease.

*Did the overall standards framework apply?*

Partial address of 70% of the overall standards falls just short of the satisfactory level of 75%, but some standards do not fit the ACIAR context of their reporting. They are cost effectiveness (F3), formal agreements, fiscal responsibility (P8) and meta-evaluation (A12), which can be argued to be in place. With this consideration, 83% of the overall standards can be argued to apply. While all of the standards are important, they may not be equally important to all audiences. Results of this analysis suggest that the standards likely to be of most interest to stakeholders outside the program – such as funders – were addressed in the evaluation reports, while standards of most interest to ACIAR staff – such as some of the feasibility and propriety standards – were less likely to be addressed in the reports.

### **6.3. Lessons learned**

Our study noted the following positive attributes that can be taken on board when designing a meta-evaluation model. They are:

- Organizations commissioning evaluations often place high priority on using findings for program improvement and to inform decision making. Most ACIAR impact assessment reports contain a foreword by the CEO, noting that the ‘study provides some very useful lessons for guiding future investments in research, in particular the need to take local industry and policy conditions into account when developing research activities’. However, other than report clarity (A5) and report timeliness and dissemination (U6), organizations should have formal mechanisms for assessing whether their evaluations are used effectively or not (Oliver 2009).



- At the start of each report, the impact assessment studies describe and document clearly the program being evaluated (A1), the context in which the program exists (A2), and the purpose and procedures that the study followed (A3).
- ACIAR commissions independent assessments for: defensible (A4), valid (A5) and reliable (A6) information; evaluator's credibility (U2); and minimization of bias (F1) when monitoring and evaluating the effects of their projects. They include contracted consultancies to AgTrans Research, Centre for International Economics, eSYS Development, Temtac in addition to consultancies undertaken by world renowned professionals and academics.
- Because of the clear documentation of methods and design protocols in all the reports examined, information scope and selection (U3) and programs values identification (U4), the analysis of quantitative (A8) and qualitative (A9), impartial justification of conclusions (A10) and undistorted reporting (A11) standards could be applied quite easily with the ACIAR evaluations.
- The ACIAR studies revealed attention to systematic research and methodological issues that meet academic standards. Their reports document in detail the national context of their work, design phases, and methodological refinements. Results are reported mainly in the form of descriptive statistics and graphs. In terms of reporting, numbers in this series are distributed internationally to selected individuals and scientific institutions, and are also available from ACIAR's website.
- The end-users of the evaluation report may have very different perspectives and information needs from the funders of the evaluation (Scott-Little et al. 2002). Standards dealing with the evaluation contract and expenditures, for example, may be of more interest to the program staff who have paid for the evaluation, while standards dealing with the reliability and validity of the data may be of more interest to funders who want to know if the intervention they funded 'worked'.

Literature tends to position meta-evaluation quality as based on the PES. We have learned from reviewing evaluation reports at ACIAR that it is difficult to include all 30 PES that ultimately affect evaluation quality. However, unless analysts applying the meta-evaluation model carefully consider and use these standards, the organization will be unable to describe and explain its evaluation quality. All that said, we posit that the design of a meta-evaluation should be grounded in a set of contextualized standards.

#### **6.4. Limitations, Implications and Future Research**

ACIAR and LDA have similarities and differences. In terms of similarities, they both operate in the agricultural sector and have a social betterment focus. But ACIAR has a much longer history of evaluation, operates in a developed country even though its work is in developing countries, and is a statutory authority. LDA, in contrast, has little experience in evaluation, operates in a developing country and is a government department. These differences have a confounding effect on the arguments we advance in the previous section. They provide limitations for the study and have implications when we assimilate lessons learned in to design of an LDA meta-evaluation.

When analyzing the PES, Fraser (2004) noted that their use depends on a number of preconditions that sometimes do not hold true in other environments (see also Stufflebeam and Shinkfield 2007). According to Beywl (2000), the usefulness of the standards to evaluations in developing countries depends on the following three assumptions, which are important for our ACIAR case study. First, the PES make certain prescriptions about the type of evaluation (defined as an evaluation model in this paper), the evaluation purposes, the training of evaluators, and the relationships between evaluators, sponsors, and recipients of programs and

services. When the conditions are met, even in developing countries such as South Africa, the standards tend to apply quite well. Second, most research and evaluation methodologists trained in Western countries tend to comply with the standards, or make reasonable adaptations to the standards even when they are not aware that they exist. Third, most evaluations are expected to be conducted by external evaluators, whose roles are distinct from the program sponsors, participants and stakeholders – an assumption that was met with the ACIAR evaluations.

The PES can contribute to assuring quality and credible evaluations, but to provide a sound basis for evaluation quality and suitability they need to be contextualized. The only area in which revisions seem necessary pertain to the feasibility and propriety standards. In terms of possible revisions to the PES, the language of the standards needs to be broadened to make them more adaptable to evaluation situations in a developing country such as South Africa which only recently adopted a government-wide monitoring and evaluation system (The Presidency 2005, 2007). Putting aside the limitations of a single case approach in making broader generalizations, the overall PES framework applied in the ACIAR case study appears to be quite applicable to the LDA case.

In summary, the limitations, implications and proposed future research include the following:

- The PES framework and meta-evaluation model has not yet been proved to be a useful guide to implementing evaluation quality measures in South Africa and is presented here with the potential to stimulate scholarly debate among those reading or studying it.
- For evaluation practitioners, the paper provides insights into how PES make sense of the evaluation quality discourse, which can affect how to introduce future quality measures.
- The paper focuses on a government department and statutory authority; future research should explore the influence of the PES in other settings with regard to quality.
- To provide empirical evidence, the paper used a single case study approach. Whilst the literature covered is extensive and the sample is reasonable, it is not comprehensive. The intention is to draw attention to the kinds of variables that need to be conceptualised, observed and included when a meta-evaluation is studied or implemented.
- The proposals for the LDA meta-evaluation are based on the theoretical literature and the ACIAR case study; therefore, future research needs to include empirical evidence that encompasses the South African context.

## **7. Proposed Systematic Meta-evaluation Framework for LDA**

A meta-evaluation design is similar to any other program evaluation design in that it involves balancing the probable cost of answering evaluation questions with the likely credibility and usefulness of the meta-evaluation results (Wholey et al. (2004). Therefore, an evaluation design is the set of decisions required to carry out the needed evaluation (Stufflebeam and Shinkfield 2007), in a systematic manner (Fitzpatrick et al. 2004; Weiss 1998) to establish the value of an evaluand (Scriven 1995). The value is measured in the strength of the evidence produced; credibility to policy makers, managers, and other intended users of the results; and the use of information in influencing policies, decisions and activities (Wholey et al. 2004). Therefore, like any other evaluation, a meta-evaluation should identify the questions that will be answered, what will be measured, and what sets of analyses will be applied to the measures to answer the questions. A comprehensive and systematic meta-evaluation design would include some representation of the steps shown in Figure 2. The steps we are suggesting are not exhaustive, but the purpose is to provide a framework of the most commonly used steps in a meta-evaluation design.

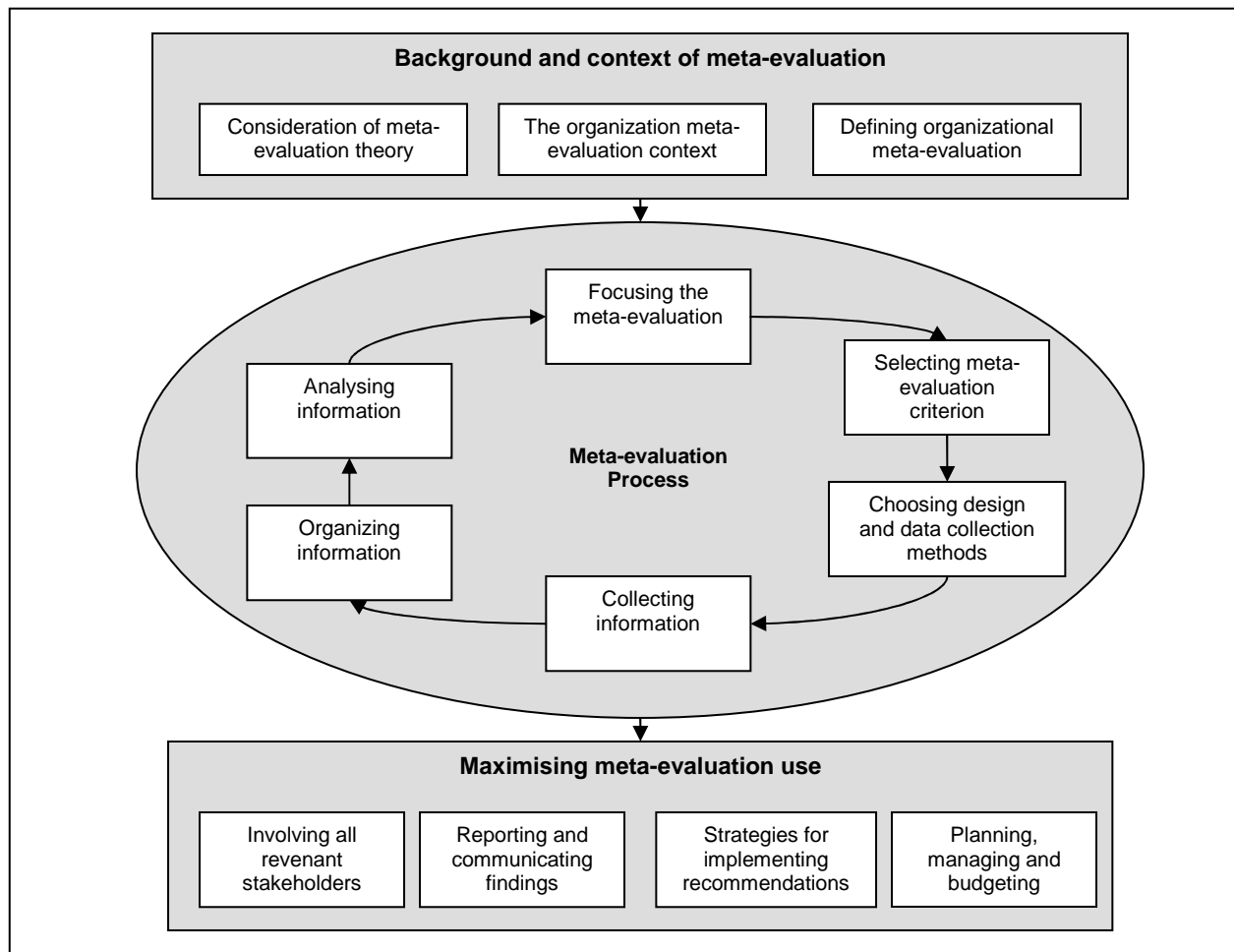


Figure 2: A comprehensive and systematic meta-evaluation model (adapted from Russ-Eft and Preskill (2001) and Stufflebeam and Shinkfield (2007:492-504))

### Step 1: Background and context of meta-evaluation

- Consideration of meta-evaluation theory: Properly practised meta-evaluation leads to direct and incontestable improvements in organizational systems, programs and practices (Fitzpatrick et al. 2004). Meta-evaluation theory can suggest what constitutes a bad or a good meta-evaluation (Shadish et al. 1991). Evaluation underpinned by theory will avoid what Scriven (1996) calls 'sin by omission'.
- The organization meta-evaluation context: We believe context is critical to understand meta-evaluation current practice and its value for quality and learning in the organization (Rossi et al. 2004; Russ-Eft and Preskill 2001). Therefore, the preceding discussion of the meta-evaluation should consider the constraints in evaluation work (Stufflebeam and Shinkfield 2007).
- Defining organizational meta-evaluation (and its quality): Evaluation quality and meta-evaluation may have different meanings to people in the organization (Nilsson and Hogben 1983; Uusikylä and Virtanen 2000; Widmer et al. 2007). Therefore, it is important to clarify

what meta-evaluation is and how it can be conducted in ways that provide useful information within the organizational context and learning nexus (Russ-Eft and Preskill 2001). For too long, organizations have neglected integrating PES into their meta-evaluation (Läubli Loud 2004). Therefore, it is due for organizations to integrate PES as a measure of evaluation quality criteria.

## **Step 2: The process**

- **Focusing the meta-evaluation:** This should reflect the rationale and purpose of the meta-evaluation, how the findings will be used, who the potential stakeholders and audiences might be, and what key questions the meta-evaluation should answer (Owen 2006; Rossi et al. 2004; Russ-Eft and Preskill 2001). Stufflebeam and Shinkfield (2007:493) exhibit an evaluation design checklist that details micro-steps that can be followed to focus a meta-evaluation.
- **Selecting meta-evaluation criteria:** There are different criteria for use in judging the quality of evaluation studies and reports which are disconcerting to evaluators and consumers alike (Fitzpatrick et al. 2004). There is no consensus on what constitutes evaluation quality. However, PES can provide a guideline.
- **Choosing the design and data collection methods:** A compromised meta-evaluation might yield invalid conclusions and mislead audience (Stufflebeam and Shinkfield 2007). Literature notes quantitative and qualitative paradigms dichotomy (Chelimsky 1997; House 2001). This dichotomy influences the meta-evaluation key questions, design and the choice of data collection methods. To make sure both perspectives are embraced, it is a good idea to conduct meta-evaluation using a team (Russ-Eft and Preskill 2001) or using mixed methods (Creswell and Plano Clark 2007; Datta 1997; Greene and Caracelli 1997; Greene 2007). Our related study (Madzivhandila et al. 2009) on knowledge construction elaborate on evaluation methodological issues.
- **Collecting information:** Different methods to collect information have different strengths and weaknesses (Creswell and Plano Clark 2007; Greene and Caracelli 1997). Therefore, the appropriateness of the chosen method(s) is very important. One consideration for choosing data collection methods is the extent to which data already exist within the organization that can allow fair judgment of evaluation quality (Russ-Eft and Preskill 2001). Due to space constraints, we will not elaborate on the menu of data collection methods. Work by Babbie (2007), Babbie and Mouton (2001), Creswell (2009), Madzivhandila et al. (2009), and Russ-Eft and Preskill (2001) provides a detailed menu of evaluation methods. The evaluation design checklist presented by Stufflebeam and Shinkfield (2007:493-294) details 10 micro-steps that can be followed to collect information for a meta-evaluation.
- **Organizing information:** Each evaluation requires an effective approach to information management (Stufflebeam and Shinkfield 2007). A functional system to file, control and retrieve data needs to be established as some data must be coded for later analysis and summary. Stufflebeam and Shinkfield (2007:294) detail three micro-steps required to organize information for a meta-evaluation.
- **Analysing information:** Analysis of data should be keyed to answering the meta-evaluation questions on quality (Stufflebeam and Shinkfield 2007). Both quantitative and qualitative data analysis should support judgments of the quality of an evaluation study or report against the quality criteria, mostly the PES (Joint Committee for Programme Evaluation 1994).

Stufflebeam and Shinkfield (2007:294) detail 11 micro-steps that can be followed to analyse information for a meta-evaluation.

### **Step 3: Maximizing use of meta-evaluation findings**

- Involving all relevant stakeholders: Meta-evaluation can serve the interest and information needs of several stakeholders (Greene 1988; Lincoln 1990; Reineke 1991). They can be program designers, developers, customers, future and former participant, community members, members of professional evaluation community, members of the organization, advisory boards, and etc. (Russ-Eft and Preskill 2001). Each of these stakeholders may have interest in the outcomes of the meta-evaluation.
- Reporting and communicating findings: Organizational learning is one of the important objectives of a meta-evaluation to secure appropriate use (Stufflebeam and Shinkfield 2007). Therefore, reporting and communicating the findings is a critical aspect. According to Russ-Eft and Preskill (2001), reporting and communicating serve two purposes: reporting the meta-evaluation findings and communicating about the meta-evaluation itself. Communication takes place long before findings are available due to interactions with stakeholders. Highly interactive ways (e.g. working sessions and meetings and verbal presentations) and less interactive ways (e.g. written reports, posters, newsletters, memos and executive summaries, internet communications) can be used to report or communicate findings (Russ-Eft and Preskill 2001). Stufflebeam and Shinkfield (2007:295) detail nine micro-steps that can be followed to report information after a meta-evaluation.
- Strategies for implementing recommendations: According to Owen (2006) recommendations are suggested courses of action, advice to policy makers, programme managers or providers about what to do in the light of the evidence and conclusions. Therefore, strategies are required to incorporate suggested recommendations in to the very fabric of the organization. Russ-Eft and Preskill suggest the use of participatory and collaborative approaches to ease implementation of meta-evaluation recommendations.
- Planning, managing and budgeting a meta-evaluation: Recognizing that meta-evaluation should be on-going and integrated with how programs are evaluated (Russ-Eft and Preskill 2001), there are better times than others to conduct a meta-evaluation that produce useful and valid information. Also, there is a need to plan for how, when and by whom the meta-evaluation will be conducted (Fitzpatrick et al. 2004). Developing a management plan can help ensure that the meta-evaluation is executed as planned. For an organization to allocate a budget for a meta-evaluation is an important requirement, whether the meta-evaluation is done internally or it is outsourced (Russ-Eft and Preskill 2001).

## **8. Conclusions**

We have stated that the 30 PES are divided into four categories. First, the utility standards primarily focus on the need for evaluations to be responsive to the needs of stakeholders. Second, the feasibility standards require the evaluator to pay attention to the cost effectiveness and real world constraints of conducting realistic, prudent, diplomatic, politically sensitive and frugal evaluations. Third, the propriety standards protect the rights of all people involved or affected by the program. These require evaluations to be conducted legally, ethically, and with due regard for their welfare. Lastly, the accuracy standards are concerned with the degree to which the evaluation is providing valid, high quality information.

In the ACIAR case study, we sought to determine the extent to which the ACIAR evaluation studies conform to the 30 PES. We found that, by design or not, there is non-use or low use of some standards. The following are the nine violations (not addressed or unable to judge) reported from their sample of 19 evaluation studies: evaluation stakeholders identification; practical procedures; political viability; formal agreements; rights of human subjects; human interactions; fiscal responsibility; analysis of qualitative information; and the use of meta-evaluation. The findings from the case study suggest that even for an organization with a long history of evaluation, it is difficult to consider equally and emphasize all the PES. How organizations rate the value of each standard rests on the context (who they are) and the purpose of each evaluation they undertake.

For LDA, the value and quality of its meta-evaluation augmented by the PES will be measured by the strength of evidence it produces, the credibility of its meta-evaluation to policy makers, managers and other intended users of the results, and the use of the meta-evaluation information in influencing organizational policies, decision making and activities.

We conclude that, the PES were developed to guide the myriad decisions and choices that an evaluator must consider to ensure quality. Although the standards do not dictate what to do in different contexts, they highlight the necessary expectations (standards) and pitfalls of evaluation practice in the modern world. They confirm and validate what constitutes a good evaluation practice (Russ-Eft and Preskill 2001). If evaluation within an organization is an old activity (like at ACIAR) or new activity (like at LDA), using contextualized standards to guide an evaluation study adds extra legitimacy to evaluation quality.

## 9. References

- Ashworth, K, et al. (2004), 'Meta-evaluations: Discovering what works better in welfare provisions', *Evaluation*, 10 (2), 193-216.
- Australian Centre for International Agricultural Research 'Impact Assessment Programme', *measuring impact* <[http://www.aciar.gov.au/measuring\\_impact](http://www.aciar.gov.au/measuring_impact)>, accessed 26 October 2009.
- 'Research that works for developing countries and Australia', *Home* <<http://www.aciar.gov.au/home>>, accessed 26 October 2009.
- Babbie, E (2007), *The practice of social research* (Belmont: Wadsworth).
- Babbie, E and Mouton, J (2001), *The practice of social research* (Cape Town: Oxford University Press).
- Bacmann, N, et al. (2000), 'The program evaluation standards in international settings', in C Russon (ed.), (The Evaluation Center Occasional Papers Series).
- Beywl, W (2000), 'Standards for evaluation: On the way to guiding principles in German evaluation', in C Russon (ed.), *The Program Evaluation Standards in international settings* (Kalamazoo: The Evaluation Center, Western Michigan University), 60-65.
- Beywl, W and Speer, S (2004), 'Data- and literature-based reflections on Western European evaluation standards and practices', *New Directions for Evaluations*, 104 (Winter), 43-54.
- Bickman, L (1997), 'Evaluating evaluation: Where do we go from here?', *Evaluation Practice*, 18 (1), 1-16.
- Blalock, A. B (1999), 'Evaluation research and the performance management movement: From estrangement to useful integration?', *Evaluation*, 5 (2), 117-49.
- Bollen, K, Paxton, P, and Morishima, R (2005), 'Assessing international evaluations: An example from USAID's democracy and Governance program', *American Journal of Evaluation*, 26 (2), 189-203.
- Boruch, R. F and Petroino, A (2004), 'Meta-analysis, sytematic reviews, and research syntheses', in J.S Wholey, H. P Hatry, and K. E Newcomer (eds.), *Handbook of practical programme evaluation* (San Francisco: Jossey-Bass).
- Bustello, M (2003), 'Evaluation of gender mainstreaming: Ideas from a meta-evaluation study', *Evaluation*, 9 (4), 383-403.
- Centre for International Economics (1998a), 'Controll of new castle disease in village chickens', *Impact assessment study report: IAS 01* (Canberra: Australian Centre for International Agricultural Research).
- (1998b), 'Establishment of a protected area in Vanuatu', *Impact assesment study report: IAS 03* (Canberra: Australian Centre for International Agricultural Research).
- (2009), 'ACIAR Database for Impact Assessments (ADIA): an outline of the database structure and a guide to its operation', *ACIAR impact assesment series: IAS 60* (Canberra: Australian Centre for International Reserach).
- Chatterji, M (2005), 'Applying the Joint Committee's 1994 Standards in international contexts: A case study of education evaluations in Bangladesh', *Teachers College Record*, 10 (October), 2372-400.
- Chelimsky, E (1997), 'Thoughts for a new evaluation society', *Evaluation*, 3 (1), 97-118.
- (1998), 'Theories of evaluation practice', *American Journal of Evaluation*, 19 (1), 35-55.
- Christie, C. A and Alkin, M. C (2003), 'The user-oriented evaluator's role in formulating a program theory: Using a theory-driven Approach', *American Journal of Evaluation*, 24 (3), 373-85.
- Chudleigh, P (1998), 'Post harverst R&D concerning tropical fruits', *Impact assesment study report: IAS 11* (Canberra: Australian Centre for International Agricultural Research).
- Chupungco, A, Dumayas, E, and Mullen, J (2008), 'Two stage grain drying in the Phillipienes', *Impact Assessment study report: IAS 59* (Canberra: Australian Centre for International Agricultural Research).
- Cook, T. D and Gruder, C. L (1978), 'Meta-evaluation research', *Evaluation Quarterly*, 2 (1), 5-51.
- Cooksy, L. J and Caracelli, V. J (2005), 'Quality, context, and use: Issues in achieving the goals of metaevaluation', *American Journal of Evaluation*, 26 (1), 31-42.
- (2009), 'Meta-evaluation in practice: Selection and application of criteria', *Journal of MultiDisciplinary Evaluation*, 6 (9), 1-15.
- Cousins, B. J and Whitmore, E (1998), 'Framing participatory evaluation', *New Directions in Evaluation*, 8, 5-23.
- Creswell, J. W (2009), *Research design: Qualitative, quantitative, and mixed methods approaches* (Thousand Oaks: SAGE publications, Inc).

- Creswell, J. W and Plano Clark, V. L (2007), *Designing and conducting mixed methods research* (Thousand Oaks: SAGE Publications, Inc).
- Datta, L (1997), 'pragmatic basis for mixed methods designs', *New Directions for Programme Evaluation*, 74 (Summer), 33-46.
- Davies, I. C (1999), 'Evaluation and Performance Management in Government', *Evaluation*, 5 (2), 150-59.
- Davis, J, et al. (2008), 'Guidelines for assessing the impacts of ACIAR's research activities', *ACIAR impact assesment series: IAS 58* (Canberra: Australian Centre for International Reserach).
- Diao, X, et al. (2005), 'The Role of Agriculture in Pro-Poor Growth in Sub-Saharan Africa', (Washington, D.C.: IFPRI).
- (2007), 'The role of agriculture in development: Implications for Sub-Saharan Africa', *Research report 153* (Washington, DC: IFPRI).
- Fischer, H and Gordon, J (2007), 'Minimising impacts of fungal disease of eucalypts in South East Asia', *Impact assesment study report: IAS 49* (Canberra: Australian Centre for International Agricultural Research).
- Fitzpatrick, J. L, Sanders, J. R, and Worten, B. R (2004), *Program evaluation: Altrenative approaches and practical guidelines* (Boston: Pearson).
- Fraser, D (2001a), 'Development of AES standards', *Evaluation Journal of Australasia*, 1 (1), 59.
- (2001b), 'Beyond Ethics: Why We Need Evaluation Standards', *Evaluation Journal of Australasia*, 1 (1), 53-58.
- (2004), 'National Evaluation Standards for Australia and New Zealand: Many questions but few Answers', *New Directions for Evaluations*, 104 (Winter), 67-78.
- Greene, J. C (1988), 'Stakeholder participation and utilization in programme evaluation', *Evaluation Review*, 12 (2), 91-116.
- (1992), 'A case study of evaluation auditing as metaevaluation', *Evaluation and Program Planning*, 15 (1), 71-74.
- (2007), *Mixed methods in social inquiry* (San Francisco: Jossey-Bass).
- Greene, J. C and Caracelli, C. J (1997), 'Defining and describing the paradigm issues in mixed-methods evaluation', *New Directions for Programme Evaluation*, 74 (Summer), 33-46.
- Hanssen, C. E, Lawrenz, F, and Dunet, D. O (2008), 'Concurrent Meta-Evaluation: A Critique ', *American Journal of Evaluation*, 29 (4), 572-82.
- Harding, M, Jiang, T, and Pearce, D (2008), 'Analysis of ACIAR's returns on investment: Appropriateness, efficiency and effectiveness', *ACIAR impact assesment series: IAS 63* (Canberra: Australian Centre for International Agricultural Research).
- Harris, D. N (2006), 'Water management in public irrigation schemes in Vietnam', *Impacy assesment studies report: IAS 43* (Canberra: Australian Centre for International Agricultural Research).
- Henry, G. T (2001), 'How modern democracies are shaping evaluation and the emerging challenges for evaluation', *American Journal of Evaluation*, 22 (3), 419-29.
- (2003), 'Inflential evaluations', *American Journal of Evaluation*, 24 (4), 515-24.
- House, E. R (2001), 'Unfinished business: Causes and values', *American Journal of Evaluation*, 22 (3), 309-15.
- Joint Commitee for Programme Evaluation (1994), *Programme evaluation standards* (Thousand Oaks: SAGE publications, Inc).
- Kirkhart, K. E (2000), 'Reconceptualizing evaluation use: An intergrated theory of influence', *New directions for evaluation*, 88 (Winter), 5-24.
- Laubli Loud, M. M (2004), 'Setting standards and providing guidelines', *Evaluation*, 10 (2), 237-45.
- Läubli Loud, M. M (2004), 'Setting standards and providing guidelines: The means toward what End?', *Evaluation*, 10 (2), 237-45.
- Lincoln, Y. S (1990), 'Multiple stakeholders, several policy interests: To whom are we ultimately ethically responsible?', *Evaluation Practice*, 11 (3), 261-62.
- Lindner, R and Mc Leod, R (2008), 'A review and impact assessment of ACIAR's fruit fly research partnerships 10984-2007', *Impact assesment study report: IAS 56* (Canberra: Australian Centre for International Agricultural Research).
- Lynch, D. C, et al. (2003), 'Descriptive Metaevaluation: Case Study of an Interdisciplinary Curriculum', *Evaluation and the Health Professions*, 26 (4), 447-61.
- Madzivhandila, T. P, et al. (2009), 'Knowledge construction component of practical programme evaluation: The move towards improving evaluation influence for social betterment', *Designing a practical*



- evaluation model for the agricultural Research, Development and Extension (RD&E) initiatives* (Armidale: University of New England).
- Martens, D. M (1999), 'Inclusive evaluation: Implications of transformative theory for evaluation', *American Journal of Evaluation*, 20 (1), 1-14.
- (2008), *Transformative research and evaluation* (New York: The Guilford Press).
- Maxwell, G. S (1984), 'A rating scale for assessing the quality of responsive/illuminative evaluations', *Educational Evaluation and Policy Analysis*, 6 (2), 131-38.
- Mc Kenney, D. W (1998), 'Australiuan tree species selection in China', *Impact assessment study report: IAS 08* (Canberra: Australian Centre for International Agricultural Research).
- Mc Leod, R (2003), 'Improved methods in diagnosis, epidemiology, and information management of foot-and-mouth deseases in South-East Asia', *Impact assessment study report: IAS 21* (Canberra: Australian Centre for International Agricultural Research).
- (2005), 'Management of fruit flies in Pacific', *Impact assessment study report: IAS 37* (Canberra: Australian Centre for International Agricultural Research).
- Mc Leod, R, Isvilanonda, S, and Wattanuchariya, S (1999), 'Improved drying of high moisture seeds', *Impact assessment study report: IAS 14* (Canberra: Australian Centre for International Agricultural Research).
- Meijerink, G (2005), 'The role of agriculture in development', *International policy brief* (Wageningen: BO Cluster International).
- Monck, M and Pearce, D (2007), 'Mite pest of honey bees in the Asia Pacific region', *Impact assessment study report: IAS 46* (Canberra: Australian Centre for International Agricultural Research).
- (2008), 'Impact of increasing efficiency and productivity of ruminants in India by the use of protected-nutrient technology', *Impact assessment study report: IAS 53* (Canberra: Australian Centre for International Agricultural Research).
- Nilsson, N and Hogben, D (1983), 'Meta-evaluation', in E. R House (ed.), *Philosophy of evaluation* (San Francisco: New Directions for Programme Evaluation).
- Oliver, M. L (2009), 'Meta-evaluation as a means of examining evaluation influence', *Journal of MultiDisciplinary Evaluation*, 6 (11), 32-37.
- Owen, J. M (2006), *Programme evaluation: Forms and approaches* (New York: The Guilford Press).
- Palis, F. G, Sumalde, Z. M, and Hossain, M (2004), 'Assessment of rodent control projects in Vietnam: Adoption and impact', *Impact assessment study report: IAS 24* (Canberra: Australian Centre for International Agricultural Research).
- Patton, M. Q (1997), *Utilization-focused evaluation* (London: Thousand Oaks).
- Pawson, R and Tilley, N (1997), *Realistic evaluation* (Los Angeles: SAGE Publications, Inc.).
- Pearce, D (2005), 'Identifying the sex pheromone of the sugarcane boremoth: Economic impact', *Impact assessment study report: IAS 34* (Canberra: Australian Centre for International Agricultural Research).
- Preskill, H. S and Torres, R. T (1999a), *Evaluative inquiry for learning organization* (Thousand Oaks: SAGE Publishers).
- Preskill, H. S and Torres, R. T (1999b), 'Building capacity for organizational learning through evaluative inquiry', *Evaluation*, 5 (1), 42-60.
- Preskill, H. S and Torres, R. T (2000), 'Learning dimensions of evaluation use', *New directions for evaluation*, 88 (Winter), 25-38.
- (2001), 'The learning dimension of evaluation use', *New directions for evaluation*, 88.
- Raitzer, D. A (2003), 'Benefit-cost meta analysis of investment in the agricultural research Centres of the CGIAR', (Rome: Food and Agricultural Organization).
- Raitzer, D. A and Lindner, R (2005), 'Review of the returns to ACIAR's bilateral investments', *Impact Assessment Series no. 35* (Canberra: Australian Centre for International Agricultural Research).
- Reineke, R. A (1991), 'Stakeholder Involvement in Evaluation: Suggestions for Practice', *Evaluation Practice*, 12 (1), 39-44.
- Rossi, P. H, Lipsey, M. W, and Freeman, H. E (2004), *Evaluation: A systematic approach* (London: Sage Publications, Inc).
- Russ-Eft, D and Preskill, H. S (2001), *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change* (New York: Basic Books).
- Ryan, J. G (1998), 'Pigeonpea improvement', *Impact assessment study report: IAS 06* (Canberra: Australian Centre for International Agricultural Research).

- Scott-Little, C, Hamann, M, and Jurs, S. G (2002), 'Evaluations of after-school programs: A meta-evaluation of methodologies and narrative synthesis of findings', *American Journal of Evaluation*, 23 (4), 387-419.
- Scriven, M (1995), 'The logic of evaluation and evaluation practice', *New Directions for Evaluation*, 68 (Winter), 49-70.
- (1996), 'The Theory behind Practical Evaluation', *Evaluation*, 2 (4), 393-404.
- Shadish, W. R, Jr., Cook, T. D, and Leviton, L. C (1991), *Foundations of program evaluation: Theories of practice* (California: Sage Publications, Inc).
- Stufflebeam, D. L (1974 ), 'Meta-evaluation', *Occasional paper series #3*.
- (2001), 'The metaevaluation imperative', *American Journal of Evaluation*, 22 (2), 183-209.
- Stufflebeam, D. L and Shinkfield, A. J (2007), *Evaluation theory, models, and applications* (San Francisco: John Wiley & Sons, Inc).
- The Presidency (2005), 'Proposal and implementation plan for the Government-Wide Monitoring and Evaluation System', (Pretoria: Government of the Republic of South Africa, The Presidency).
- (2007), 'Policy framework for the Government- Wide Monitoring and Evaluation System', in The Presidency (ed.), (Pretoria: Government of Republic of South Africa, The Presidency).
- Tissdell, C and Wilson, C (2001), 'Breeding and feeding of pigs in Australia and Vietnam', *Impact assessment study report: IAS 17* (Canberra: Australian Centre for International Agricultural Research).
- Uusikyla, P and Virtanen, P (2000), 'Meta-evaluation as a tool for learning:A case study of the European Structural Fund, evaluations in Finland', *Evaluation*, 6 (1), 50-65.
- Uusikylä, P and Virtanen, P (2000), 'Meta-evaluation as a tool for learning: A case study of the European structural fund evaluations in Finland', *Evaluation*, 6 (1), 50-65.
- Valovirta, V (2002), 'Evaluation utilization as argumentation', *Evaluation*, 8 (1), 60-80.
- Van Bueren, M (2004a), 'Acacia hybrids in Vietnam', *Impact assessment study report: IAS 27* (Canberra: Australian Centre for International Agricultural Research).
- (2004b), 'Eucalypt tree improvement in China', *Impact assessment study report: IAS 30* (Canberra: Australian Centre for International Agricultural Research).
- Weaver, L. and Cousins, B. J (2004), 'Unpacking the participatory process', *Journal of MultiDisciplinary Evaluation*, 1, 19-40.
- Weiss, C. H (1988), 'Evaluation for decisions: Is anybody there? Does anybody care?', *Evaluation Practice*, 9 (1), 5-19.
- (1993), 'Politics and evaluation: A reprise with mellower over tones', *Evaluation Practice*, 14 (1), 107-09.
- (1998), *Evaluation* (New Jersey: Prentice Hall).
- (1999), 'The interface between evaluation and public policy', *Evaluation*, 5 (4), 468-86.
- Wholey, J. S (1986), 'Using evaluation to improve government performance', *Evaluation Practice*, 7 (2), 5-13.
- (2001), 'Managing for results: Roles for evaluators in a new management era', *American Journal of Evaluation*, 22 (3), 343-47.
- Wholey, J. S, Hatry, H. P, and Newcomer, K. E (eds.) (2004), *Handbook of practical programme evaluation* (San Franscisco: Jossey-Bass).
- Widmer, T, Blaser, C, and Falk, C (2007), 'Evaluating Measures Taken Against Right-Wing Extremism', *Evaluation*, 13 (2), 221-39.
- Wingate, L. A (in press), 'Meta-evaluation: Purpose, prescriptions, and practice', in B McGaw, P. L Peterson, and E Baker (eds.), *International encyclopedia of education* (San Diego: Elsevier).