
The benefits of both worlds: Towards an integrated mixed-methods approach for evaluating women's empowerment

Melody Mentz

melody@meoldym.co.za

Research and Evaluation Specialist, MelodyM Consulting

Abstract

Increasing demands for credible, rigorous monitoring and evaluation of empowerment in the context of development will challenge theorists and practitioners to find innovative, but credible ways of understanding empowerment. This paper provides a theoretical rationale for, and a practical description of an integrated mixed-methods approach for evaluating the empowerment contributions of the African Women in Agricultural Research and Development program. The paper explores the value of employing mixed-methods for increasing the credibility of results, and for improving understanding of the ways in which the program facilitates change for its participants. This paper assesses the process of data integration and analysis to provide a set of practical recommendations for practitioners seeking to employ similar approaches. The paper finds that managerial commitments, adequate investment in ongoing monitoring and evaluation activities, and openness to methodological innovation are essential for the effective development and use of mixed-method approaches in complex development interventions.

Keywords: Mixed-methods analysis, empowerment, evaluation methodologies, women in agriculture

Introduction

There are increasing demands for credible and rigorous monitoring and evaluation of empowerment at individual, institutional, national, and supranational levels. The global agenda, particularly the Sustainable Development Goals (SDGs), make a clear commitment to evaluating development through processes that are “rigorous and based on evidence, informed by country-led evaluations and data which are high-quality, accessible, timely, reliable, and disaggregated by income, sex, age, race, ethnicity, migration status, disability, and geographic location and other characteristics relevant in national contexts” (Bamberger, Segone and Tateossian, 2016, p.31). Increasing calls for widespread accountability require recipients of development aid to demonstrate the effectiveness and efficiency of their efforts (Strawson, 2015). In addition, there is an increasing acknowledgement of the value of monitoring and evaluation not only for accountability, but also for learning. Again, this is illustrated in the SDG agenda that commits to both “promote accountability to our citizens” and “foster exchanges of best practices and mutual learning” (Bamberger, Segone and Tateossian, 2016, p.31).

One implication of this increased emphasis on monitoring, evaluation, and learning (MEL) is that theorists and practitioners working towards empowerment in the context of development will be increasingly expected to deliver and defend useful, credible, and cost-effective evaluations. A challenge faced by the MEL community is the “super-wicked” nature of the problems that the community seeks to address (term coined by Levin et al., 2012). Wicked problems involve multiple interacting systems; they are characterized by high levels of uncertainty; and there is only imperfect knowledge about both their nature and their solutions (Rittel and Webber, 1973). Furthermore, in the case of super-wicked problems, the time available for solving them is running out; there is no recognized authority on how to solve the problem; and the very persons who seek to solve the problem, also perpetuate it.

In Africa for instance, the agriculture-poverty-gender intersection presents one of the continent's

most pressing and interconnected super wicked problems. Agriculture as a sector has two key roles (and numerous other linked roles) in achieving the SDGs, namely, to provide food and to secure livelihoods. The most recent World Bank Poverty Report, *Poverty in a Rising Africa* (Beegle et al., 2016) predicts that Africa will need to produce 80% more food by 2050 as compared to the global need for 50% more food during the same period. Moreover, more than 700 million people in the world still live in extreme poverty and one quarter of the world's extreme poor live in Africa.

Agriculture is the single largest employer in the world, providing livelihoods for 40% of today's global population. It is the largest source of income and jobs for poor rural households (SDG Fact Sheet, 2015). Africa is also distinguished by a large share of female-headed households, 62% of which have no adult men (15 years or older) (Beegle et al., 2016). As the number of female-headed households continues to increase, and taking into consideration the role of women in agriculture more broadly (regardless of the sex of the household head) the need to identify and implement effective strategies for women's empowerment intensifies. It has been estimated that if female farmers had the same access to resources as men, the number of hungry people in the world could be reduced by up to 150 million (SDG Fact Sheet, 2015). The gender agenda is thus a critical element of the development agenda in Africa.

Despite women's important contributions to the sector, agricultural research and higher education are disproportionately led by men and there is an urgent need for greater representation of women in laboratories, leadership positions and in policy forums where important decisions are made and technological innovations are driven (Beintema and Di Marcantonio, 2010). Part of the solution to overcome this is to provide support to African female agricultural researchers so they can use their unique insights and perspectives to contribute to poverty alleviation and food security at the highest possible level. When implemented, appropriate methodologies are needed for assessing the impact of interventions that address super-wicked problems and facilitate understanding of the nature of the problem and why it persists. These methodologies must allow for a nuanced understanding of what leads to change when it is observed. In other words, MEL must serve both a descriptive and explanatory function.

Mixed-method evaluation approaches have been rising in prominence since the 1980s, and are becoming more popular and sophisticated. Historically, the dominance of experimental and quasi-experimental research approaches has limited innovations in methodology. Despite this, the academic debate is evolving and a growing body of literature describes and verifies alternative approaches, including mixed-methods approaches. It is thus becoming increasingly accepted that the core question is no longer "should methods be mixed?", but rather "how to effectively mix methods?" (Bamberger, 2012; Patton, 2008)

This paper provides a rationale for, and describes the development of, a multi-phase, parallel convergent mixed-methods design applied to evaluating women's empowerment. The design was developed and tested over an eight-year period by the career-development program, African Women in Agricultural Research and Development (AWARD). Since 2008, AWARD, through tailored fellowships, has equipped top women agricultural scientists across sub-Saharan Africa to accelerate agricultural gains by strengthening their science and leadership skills. The details of the program are discussed in Noordeloos (2015). Approximately 70 fellows are accepted on an annual basis.

The fellowship is built on three cornerstones – mentoring, science skills development, and leadership skills development. Fellows benefit from being matched with a mentor (a respected male or female senior scientist in her area of expertise) and are offered a range of courses designed to improve their ability to share their knowledge, through science- and proposal-writing

courses, and to improve their presentation skills. This paper discusses the evolution of AWARDS approach to evaluation, and presents findings from the 2013 and 2014 cohorts to illustrate the approach and its value.

Literature review

The selection of a particular methodology in evaluation does not by default come with guarantee of (or compromise in) quality or rigor as quantitative, qualitative, and mixed-methods studies can be poorly designed, implemented, analyzed, or interpreted. Quality and rigor are ensured through careful design and implementation and the choice of methodology should be based on a combination of factors. These factors include philosophical and practical arguments, as well as the utility of the methods selected for providing the types of information of key interest to the intended users of the evaluation (Patton, 2008).

Paradigmatic considerations – quantitative or qualitative?

Purely quantitative approaches, based on the traditional scientific method, seek to uncover objectively “true” facts. A key assumption with these approaches is that by counting and controlling measurements, an objective reality can be known which is independent and distant from the “knower” (i.e. researcher or evaluator) (Trochim, 2006). Counting in absolute numbers is highly desirable, if not imperative, in many cases. Clearly defined quantitative indicators allow progress tracking over time in a directly comparable fashion for monitoring purposes. For example, 169 targets and 230 indicators have been approved for the SDGs’ follow-up and review processes (Lucks et al., 2016). Without these quantitative forms of data which are collected consistently, systematically, and reliably, it would be impossible to determine the extent to which progress has been made on the targets and indicators. Targets can be monitored at the highest level (as with the SDG example), but also at a programmatic level. In addition, quantitative data can offer clearly defined reference points for longitudinal tracking, group comparisons, and a range of other statistical investigations.

In evaluation, quantitative approaches are typified by the ideals of hard, reliable, context-free data which serves as proof of the observed phenomenon (Trochim, 2006). Within the practice of evaluation, the most outstanding example of the purely quantitative approach is Randomized Control Trials (RCTs) “which estimate the mean net impact of an intervention by comparing results between a randomly assigned control group and experimental group or groups” (Better Evaluation, 2016). The appeal of approaches of this nature is the perception of tangibility and proof, and there has been idealization of these approaches as the “gold standard” (Trochim, 2006). However, there are practical limits to the validity of the approach (see Shadish, Cook and Campbell, 2001) and there are situations in which its application is not appropriate from a philosophical, ethical, or practical perspective (Olofsgård, 2014).

Essentially, quantitative methods are suited to measuring levels and changes in impacts and to drawing inferences from observed statistical relations. They are less effective in understanding process, which are the mechanisms by which an intervention results in a series of events that leads to a desired (or unanticipated) impact (Bourguignon and Pereira da Silva, 2003). They provide limited contextual information and reduce highly complex situations or narratives into reductionist numbers (Bamberger, 2012).

The second broad approach is experience-based and qualitative which involves observing, focusing on the experience of individual “knowers”, and enabling them to describe their experience which is multilayered and complex (Cohen, Manion and Morrison, 2011). Therefore,

rather than seeking to uncover “objective facts”, qualitative approaches provide evaluators with insights into the lived experiences and interpretations of the participants and allows for a greater depth of understanding into the complexity of a given situation, experience, or context. Rather than seeking objectivity, qualitative approaches acknowledge the inherent influence that the evaluator brings to the interpretation of textual, observational, or visual data (Miles, Huberman and Saldana, 2013). Qualitative approaches are also particularly useful for theory building, through grounded theory approaches (Glaser, 1998), and for the verification of theories of change (Vogel, 2012).

Despite the opportunity to understand specific contexts, phenomena, or experiences in a great level of detail, qualitative approaches are often limited in their generalizability and are frequently criticized for subjectivity and lack of credibility (Bamberger, 2012). Thus, in the evaluation context, irrespective of their individual benefits, there are limitations to the mono-method approaches described above. Therefore, mixed-approaches has become a growing field of literature, with academics rising to the challenge of describing in greater detail the philosophical and practical implications of mixing in evaluation.

Mixed-method approaches have several advantages, including that they allow for answering descriptive and explanatory questions, allow for the assessment of change, and help to explore the nature and pathways of change (Pereznieto and Taylor, 2014). In the case of AWARD, the MEL approach needed to serve both an explanatory and a descriptive purpose. In other words, data was needed in order to answer the question of how change is facilitated, not only what (if any) change has occurred. Purely quantitative approaches were deemed unlikely to uncover these nuances.

Transformative approaches in evaluation

Transformative paradigms in evaluation acknowledge that realities are constructed and shaped by social, political, cultural, economic, and racial/ethnic values. This necessitates that power and privilege must be considered in the evaluation process. Evaluators embracing a transformative paradigm explicitly bear social justice issues in mind so that their work becomes intertwined with a political agenda (broadly defined) and are action-oriented towards generating increased equity within society (Mertens, 2007).

According to Mertens (2007), methodological inferences based on the underlying assumptions of the transformative paradigm reveal the strength of mixed-methods. A qualitative dimension is needed to gather participant perspectives on their experience of the process, while a quantitative dimension provides the opportunity to demonstrate outcomes that have credibility. Thus, mixed methodologies provide a mechanism for addressing the complexities of evaluation in culturally complex settings that can provide a basis for social change.

Transformative evaluation approaches are acknowledged for being particularly important in the context of empowerment. In 2014, the *Gender and Development Journal* published its first ever special issue exclusively on monitoring, learning, and evaluation. In this issue, Pereznieto and Taylor (2014), reviewed monitoring, evaluation, and learning methods and approaches used in 70 cases of women’s empowerment. Their study concluded that mixed-methods are not only beneficial, but required if empowerment is to be comprehensively assessed. In this same issue, (Bowman and Sweetman, 2014 pp 206) speak of the importance of listening to nuanced and individual assessments of the changes that development makes, specifically to the lives of women:

“Good MEL arrives at a conclusion drawing on multiple perspectives, systematically and transparently. It should aim to enable previously unvoiced perspectives to come to the fore

and challenge the dominance of professional or expert researchers mining the experience of grassroots women and girls whose own identity and location means their views are seldom, if ever, sought.”

Evaluators seeking to understand empowerment need to understand the quantitative changes which have taken place, but also the experiences and processes underlying these changes which can only be fully articulated through qualitative data. From its inception, the AWARD program intended for its MEL to serve an empowering role. In other words, MEL was never intended to only understand empowerment, but to also enable it (Noordeloos, 2015). This intention is aligned clearly with a transformative approach, and thus the benefits of rigorous mixed-methods were adopted by the program from the outset.

Practical considerations for method selection

Pragmatic issues related to "how" data and information are accessed and collected, as well as for what purposes it is analyzed, reported, and used should also be considered. A pragmatic approach is concerned with what works; it is problem-centered, pluralistic, and orientated towards real-life practice. In a pragmatic approach, any method may be used to understand and address the problem – making the purpose of the investigation the central concern (Creswell, 2013).

In the AWARD context, several competing priorities needed to be balanced and weighed against each other. For instance, the requirements by funders to demonstrate impact versus the voices of the participants and target audience; and the need to provide evidence of outcomes and impacts versus the importance of using MEL insights in adaptive management. In addition, time and budget had to be weighed up against ideal MEL plans (Bamberger, Rugh and Mabry, 2012).

Furthermore, mixed-methods approaches are appropriate when there are widely differing contexts for the evaluator and the evaluated, when both tangible and intangible outcomes and impacts are of interest, and when language may present a significant barrier to reliable data collection. The cross-cultural context in which the program is implemented (eleven countries across Africa) increases the probability that differences in the use of jargon and English language mastery are likely to influence responses to surveys. Therefore, balancing these priorities and managing the practicalities of a cross-cultural study necessitated a mixed-methods approach.

Developing the components of an integrated mixed-method approach

The decision to use a mixed-methods approach must be followed by a careful consideration of the evaluation design. Key decisions in the design relate to how the data will be employed to address the stated evaluation questions, and the extent to which quantitative and qualitative data will be integrated.

Mixed-methods evaluation approaches may be employed for several purposes (Greene, Caracelli and Graham, 1989). These include triangulation (to increase the validity of data and minimize bias), complementarity (to enhance the strengths and minimize the weaknesses of individual methods), development (to help use the results of one method to enhance another), initiation (to catalyze new or unexpected insights from the juxtaposition of data and methods representing different perspectives), and expansion (to increase the overall scope of research).

Use of retrospective baselines

The retrospective baselines design has gained prominence as a convenient, valid method for measuring self-reported change (Klatt and Taylor-Powell, 2005). Retrospective baselines are

administered later than a traditional pre-test, most frequently at the same time as the post-test. In these types of pre-tests, the respondents are asked to answer questions about their level of understanding or a skill *after* an intervention by reflecting back on their understanding prior to the intervention.

Retrospective baselines present several advantages. In the first instance, retrospective baselines allow for a single point of administration. The approach allows fellows to reflect on their experiences at a single point in time (at the end of the fellowship) and not at multiple points in time. The approach has been proven to be effective in avoiding the response-shift effect, which occurs when a respondent's frame of reference or evaluation standard changes significantly during the course of the intervention. If participants misunderstand basic terms or concepts associated with a specific construct, then results from traditional pre-test questions may be misleading. This response-shift effect can under- or overestimate the actual program effects (Lam and Bengo, 2003). The method thus avoids the response-shift effect by clearing up constructs before participants are asked to make assessments and avoids introducing constructs to participants before they are ready to be introduced to them.

A study conducted by Howard et al., (1981) showed that retrospective baselines are no more susceptible to social desirability demands than the traditional pre-test. The ability of participants to recall accurate information after an intervention, especially after a longer period, may pose threats to validity. However, research by Lamb and Tschillard (2005) found that there were few differences in the results between using retrospective baselines and the conventional pre-test/post-test approach. This finding suggests that the retrospective baselines are a good method to use if it is difficult or impossible to use traditional pre-tests.

Retrospective baselines result in less missing data than the traditional pre-post design. Unless respondents fail to complete the questionnaire appropriately, there will always be pre-test and post-test data for each participant. The only missing data will be from people who did not complete the questionnaire or from people who skipped items (Raidl et al., 2004; Klatt and Taylor-Powell, 2005).

Overcoming the limitations of self-report data through triangulation

There has been much debate about the reliability of self-reported data, however, various studies show that valid responses can be obtained through self-reported measures when certain conditions are present (Baird, 1976; Pace, 1985; Pike, 1995) and when items do not threaten respondents by asking questions about highly sensitive topics (Bradburn and Sudman, 1988). Kuh (2005) summarized the conditions conducive to accurate and valid self-report data, asserting that the validity is high if the information requested is known to the respondent, questions are phrased clearly and unambiguously, questions refer to recent activities, the respondent thinks the question merits a thoughtful response, and answers do not threaten privacy.

Alternatively, since the early 1980s, researchers have been investigating the intersection between cognitive psychology and survey methodology (Jobe and Mingay, 1991), providing evidence from the field of human cognition which suggests that the inaccuracies of self-report surveys extend beyond topic sensitivity and social desirability. Critics of self-report measures contend that this method of data collection is susceptible to a wide range of inaccuracies due to the nature of the response process and how humans store and recall memories (Porter, 2009). The widely-referenced response model proposed by Tourangeau (1984) highlights the substantial cognitive effort required of participants in responding to self-report surveys. The nature of this process, as well as the energy required to optimize the quality of responses, makes participants susceptible to providing inaccurate answers. Tourangeau's model (1984) states that when participants respond to

questions, they must work through four processes, namely, comprehension and understanding of the question posed; retrieval of information related to the question; judgment and/or estimation of the response; and, reporting the actual response. Inaccuracies in responses can occur during any one of the four phases.

Various problems related to the first process (comprehension and understanding of the question posed) have been identified, including misinterpretation, the use of jargon, vague and/or ambiguous wording and response categories. If respondents do not understand (i.e. interpret) the intended meaning of the words in the question in the same way that the researcher does, then inaccurate (and by implication, invalid) data is potentially being collected. The lack of detail provided in many surveys, often in an attempt to minimize survey length, may contribute to a plurality of interpretations (Porter, 2009). In addition, misinterpretation of survey questions may be particularly pertinent when respondents are not completing the survey in their mother tongue, as is the case for the majority of the program's participants. Furthermore, survey questions may be difficult to comprehend if the questions employ discipline-specific jargon. These terms are used in a very specific manner by the evaluator, but may have little (if any) meaning to respondents. All of these challenges highlight ways in which inaccurate data may be collected through quantitative self-report perceptual surveys, even if it was not the intention of the participant to misrepresent themselves.

Despite the concerns raised, self-report assessment remains one of the most widely implemented and useful tools for collecting large amounts of data (Borden and Zak-Owens, 2001). It is, however, important for researchers and evaluators to recognize the limitations of these measures and to work towards constantly improving the reliability and validity of the data collected from their surveys in order to appropriately use the data in the intended contexts to achieve specified outcomes.

The remainder of this paper discusses the evolution and application of the mixed-methods approach employed by the program in its monitoring and evaluation efforts.

Methods

A convergent parallel, multiphase design was selected based on the specific purposes articulated by the program for the MEL component of the program's work. This is illustrated in Figure 1.

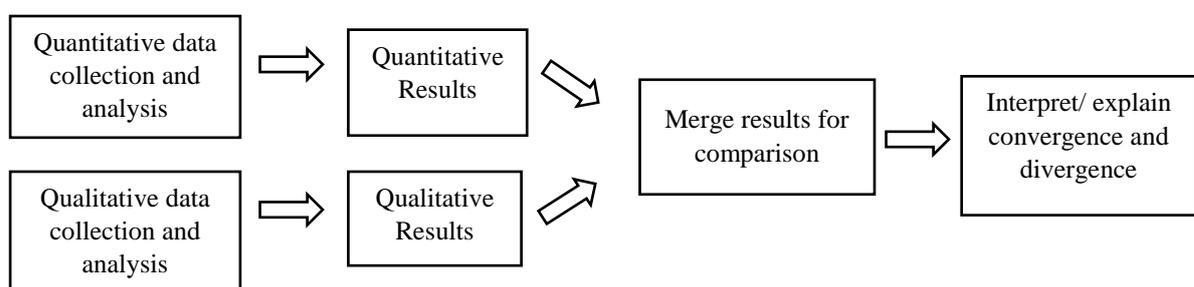


Figure 1: A convergent parallel design

The multi-phase component of the design was added through annual data collection from participants from the 2013 and 2014 cohorts through a similar format data collection tool. This longitudinal design allows for comparison between cohorts to identify stable trends and patterns, but also to increase overall sample size to allow for analysis by other variables of interest, for

example age, country, employment organizational type, etc.

Evaluation data collection tool

The author of this paper worked with the program over a period of five years to design and implement MEL related to the empowerment of program participants. During this period, the data collection tool (final evaluation form) underwent significant refinement in order to ensure ever increasing levels of credibility and rigor, and to ensure that the results of the evaluation are useful for both reporting and program implementation.

Mixed-data from fellows' final evaluation forms were used for three specific purposes, namely, triangulation, complementarity and development. Thus, the data collection tool (final fellow evaluation form) was specifically designed with these purposes in mind.

In 2013, the program began using a retrospective baseline approach to collect quantitative data after investigating the feasibility in literature (see earlier discussion). The limited amount of missing data when using retrospective baselines was considered a significant benefit in the program context where large amounts of missing data from traditional pre- and post-test designs restricted the analysis of change for earlier cohorts. The ability to minimize response shift bias was also an important factor for the program when dealing with concepts such as gender-responsiveness, role modelling, and mentoring.

To overcome some of the challenges of using self-report data, the program needed to apply stringent criteria for the triangulation of quantitative and qualitative data. Careful attention to qualitative question design focused the collection of verifiable data consisting not only of perceptual descriptions, but also eliciting detailed information about specific examples of change to serve as complementary information to the quantitative responses. This purposeful question design resulted in qualitative data that served to enhance the credibility and verifiable nature of survey responses, rather than merely providing detailed self-perceptions. A full description of the integrated approach to triangulation is described in a later section of this paper.

The program's African Women in Science Empowerment (AWSEM) framework, which is directly linked to its theory of change, identified five types of empowerment needed for female agricultural researchers. The five types of empowerment are referred to as powers, each of which is illustrated in Table 1.

To evaluate the extent to which fellows have been empowered during the fellowship, all five of these dimensions are measured both by quantitative and qualitative questions in the final program evaluation form. In addition, the programs contributions to the change in each dimension, was investigated by quantitative and qualitative questions related to both the change and the contribution. The specific format and design of the questions is discussed in the sections below. For the purposes of this paper, illustrative examples are drawn from the findings related to the category, Power from Within – a fellow's inner power (See Table 1).

Table 1: Description of questions in final evaluation form related to each empowerment dimension

Power subdomains	Description of questions in final evaluation form related to each Power
<p><i>Power from Within</i></p> <ul style="list-style-type: none"> ▪ Self-confidence ▪ Self-knowledge ▪ Motivation ▪ Vision and direction 	<p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power from Within.</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power from Within.</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p>
<p><i>Power to Do</i></p> <p>Access to</p> <ul style="list-style-type: none"> ▪ Knowledge and information ▪ Opportunities ▪ Contacts and networks <p>Leadership capacities</p> <ul style="list-style-type: none"> ▪ Present oneself professionally ▪ Manage diversity ▪ Leverage team talents ▪ Formally mentor others ▪ Negotiate ▪ Network <p>Scientific capacities</p> <ul style="list-style-type: none"> ▪ To conduct research ▪ To conduct gender-responsive research ▪ To fundraise ▪ To present (work or research) 	<p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power to Do (access domains).</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power to Do (access domain).</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p> <p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power to Do (leadership domains).</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power to Do (leadership domain).</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p> <p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power to Do (scientific domains).</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power to Do (scientific domain).</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p>

Power subdomains	Description of questions in final evaluation form related to each Power
<p>Power Over</p> <ul style="list-style-type: none"> ▪ Career Progress ▪ Visibility ▪ Professional Recognition, 	<p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power Over</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power Over</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p>
<p>Power With</p> <ul style="list-style-type: none"> ▪ Participating in collaborative activities ▪ Leading collaborative activities 	<p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power With</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>One qualitative question asked fellows to reflect on the most significant changes in their Power With</p> <p>One qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p>
<p>Power to Empower</p> <ul style="list-style-type: none"> ▪ Efforts to raise awareness of gender responsive research & development ▪ Efforts to strengthen capacities around gender responsive research & development ▪ Efforts to influence on norms, policies and strategies for gender responsive research & development 	<p>QUANT</p> <p>Quantitative questions asked fellows to reflect on their change with regards to each aspect of the Power to Empower</p> <p>One quantitative question asked fellows to reflect on the strength of AWARDS contributions to the change they experienced.</p> <p>QUAL</p> <p>Three qualitative question asked fellows to reflect on the most significant changes in their Power to Empower</p> <p>Three qualitative question asked fellows to describe the ways in which AWARD contributed to the changes they experienced.</p>

Quantitative question design and content

Quantitative data were collected to provide systematic comparisons over time and for reporting purposes. This included monitoring type data as well as data that could be used to answer evaluation questions. The quantitative data collected spanned both the factual and the perceptual.

Quantitative questions asked fellows to rate themselves on each aspect (outcome) of the five powers at the start and the end of the fellowship. This retrospective baseline approach allowed for individual level change to be calculated for each fellow on each identified outcome from the quantitative data (See Table 2).

Table 2: Questions for retrospective baselines (final evaluation questionnaire)

	Very Low	Low	Moderate	High	Very High	Not applicable
START of the Fellowship						
END of the Fellowship						

Where *quantitative factual data* were required, attempts were made to collect data in a manner that ensures the verifiability of the data. For example, fellows were asked to indicate the number of publications they had contributed to during the fellowship. To verify the answers provided, full publication data in a CV-type format was requested. This approach limits the extent to which self-report bias or social desirability could impact the way the questions were answered.

The perceived strength of the program's contribution was also examined for each power through a quantitative question; the typical format of which is illustrated in Box 1.

Box 1: Question on AWARD's contribution (Final Evaluation Questionnaire)

<p>To what extent was the Fellowship a factor in bringing about any changes to your inner power?</p> <p><input type="checkbox"/> I did not experience any changes to my 'inner power'</p> <p><input type="checkbox"/> Negative factor</p> <p><input type="checkbox"/> No factor</p> <p><input type="checkbox"/> Minor positive factor</p> <p><input type="checkbox"/> Moderate positive factor</p> <p><input type="checkbox"/> Major positive factor</p>
--

Qualitative question design and content

Whilst quantitative questions ask fellows to rate themselves on each outcome related of the type of empowerment in question, qualitative questions related to a particular power were presented to fellows following the series of quantitative questions. The qualitative questions did not prompt fellows to describe changes in all aspects of the power. Rather, fellows were asked to report on the most important changes they have experienced related to a power and to provide concrete examples illustrating these changes. The stories reported by fellows thus illustrate aspects of the changes that are most salient and meaningful to them, rather than to provide a comprehensive description of changes on all aspects of a particular type of empowerment. The typical format of these questions was "*Please tell us about the changes you have experienced, and give concrete examples of how you have changed since joining AWARD in terms of your inner power.*"

The way in which the program contributed towards changes in a particular area of empowerment was assessed by a single qualitative question, typically phrased as follows: *We are especially interested in the role that the Fellowship has played in bringing about these changes. Please explain how your Fellowship experience influenced the changes in your inner power.*

Analysis and management of data

The analysis of the data from the final evaluation forms was done in an iterative process. The first

step included a descriptive analysis of the quantitative data and a deductive thematic coding of qualitative responses. After this initial analysis, quantitative and qualitative data were integrated for triangulation purposes, and further descriptive analyses were drawn from the integrated data set.

A final step of the investigation was to revisit the qualitative data to identify actual narratives from the fellows that provide in-depth and contextual insights to supplement the quantitative and triangulated findings. The data analysis procedures are discussed in greater detail in the sections below.

Quantitative data analysis

Quantitative data was analyzed using the Statistical Package for Social Sciences (SPSS). As a first step in the analysis process, all quantitative data was analyzed, merging variables where necessary, and performing transformations. Important data transformations include, for example, comparing fellows' *start* and *end* of fellowship assessments of their empowerment to determine change in a particular outcome. Based on these data transformations, the number of fellows who experienced change in a particular outcome could be determined and the scale of change could be identified.

Table 3 and Figure 2 illustrates a set of typical descriptive quantitative findings from the analysis of change in fellows' level of confidence (one of the Power from Within outcomes). This initial phase of quantitative analysis focuses on purely descriptive goals through the production of frequency tables and appropriate data visualizations.

Table 3: Percentage of fellows with quantitative evidence for change in self-confidence

	Percentage with change
All Fellows	98%
2013 cohort	100%
2014 cohort	96%
Post-bachelors	94%
Post-master's	93%
Post-doctoral	93%

To identify any possible trends/differences between groups, quantitative data was disaggregated by cohort of Fellowship and Level of Qualification (post-bachelors level, post-master's level and post-doctoral level). Due to practical considerations including time, human resources, and budget allocations, these are the only two variables by which disaggregation was done for all variables. However, the potential for disaggregation on several other criteria exists, for example, country of residence or organizational type (see figure 2).

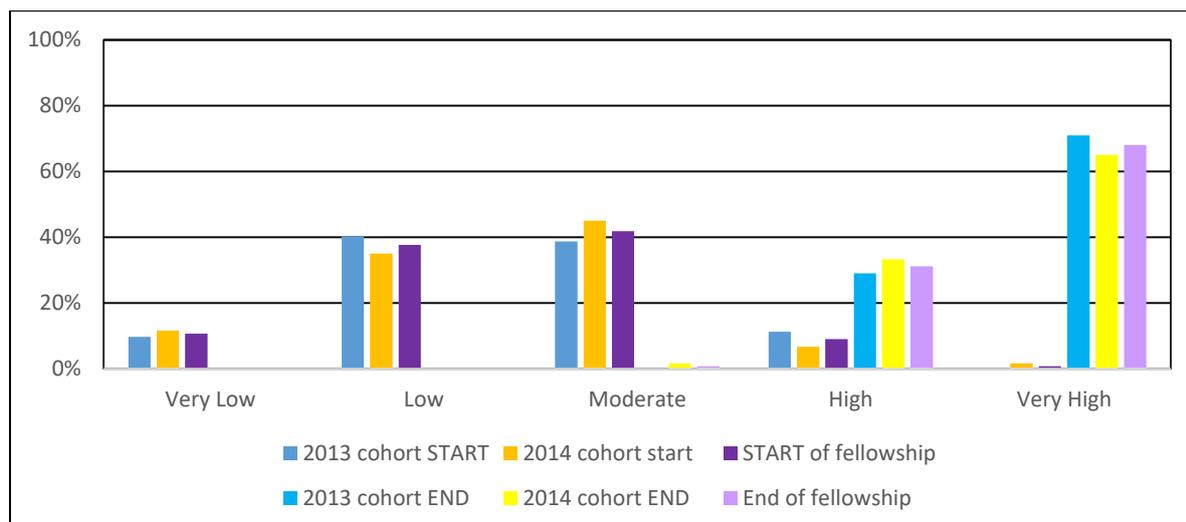


Figure 2: Rating of fellows' self-confidence before and after the program

Source: AWARD 2016 evaluation report

From this example, it is noted that 98% of fellows reported quantitative changes in their self-confidence during the fellowship period, and the distribution of confidence ratings shifted from predominantly low at the start of the fellowship to predominantly high at the end. This analysis was repeated for each outcome within each one of the powers.

Qualitative analysis

Deductive qualitative coding was conducted using Dedoose (www.dedoose.com), an online mixed-methods data analysis tool. The program conducted a detailed review of available Qualitative Data Analysis (QDA) tools during in 2014, and selected Dedoose for several reasons. These include the highly intuitive nature of the interface, the comparatively low costs of using the platform, the potential for simultaneous collaboration on a single project by multiple coders, and most importantly, the mixed-methods analytic capabilities of the tool.

Detailed code descriptions were agreed upon to improve consistency in the coding process. Data were coded by a small team of researchers, all of whom have been engaged with the program for at least three to five years. Resource constraints limited the extent to which all coded excerpts could be verified by a second researcher. However, for each power, a percentage of excerpts (usually 10-15%) were selected for verification. In cases where there were two coders who differed in more than 10% of cases, the full set of codes for the particular power were revisited.

Qualitative stories of change were deductively coded according to the empowerment framework and rated in terms of their credibility as either lackluster, convincing, or compelling. Box 2 summarizes the broad criteria for coding the qualitative stories.

The combination of convincing and compelling stories was counted as credible evidence. Fellows who have credible evidence in their descriptive stories were considered to have qualitative evidence of change in that particular power and were included in the integrated mixed-methods analysis. Fellows with lackluster evidence of change were considered to have no qualitative evidence of change.

Box 2: Criteria for coding the qualitative stories

- NO CHANGE
- LACKLUSTRE: Mention change but no examples or description of change provided
- CONVINCING: Changes on one or more sub-domain, with at least one example or description provided. Examples or descriptions either lacking in detail or not verifiable, in other words not a 'wow' story
- COMPELLING: two or three examples across multiple subdomains of the power, has to reflect behavior change or directly verifiable evidence

Qualitative questions also asked fellows to reflect on the role that the program played in bringing about the changes they experienced. Fellows were asked to describe specific program activities and to explain how these activities enabled or contributed towards the changes they experienced. These qualitative responses were coded to identify which program activities played the most prominent role in the development of each of the five powers. Using the Dedoose mixed-methods analysis tool, code application data at the individual fellow level was exported from Dedoose into Excel for integration with the quantitative data.

Integration of qualitative and quantitative data

For the purpose of triangulation, quantitative and qualitative data were merged in SPSS at the individual fellow level, and this matched data was thus used for the integrated analysis.

Integrated data were used for three primary purposes, (i) triangulation, (ii) complementarity, and (iii) development. The process of triangulation allows for higher levels of confidence in the findings reported, as well as a more in-depth understanding of how the program contributed towards the changes that fellows reported. In addition to the purposeful design of questions to address specific needs, mixed-methods data also allowed the program to understand context-specific factors. These included identifying the contributors to differences in findings for fellows who come from different countries, organizational contexts, as well as differences in levels of qualification. From this perspective, the qualitative and quantitative data were used in a complimentary manner. By matching a fellow's quantitative and qualitative responses, it was possible to identify questions which were not uniformly understood between participants, or which were frequently interpreted in ways other than intended by the evaluation question design. Over the eight years of implementation, responses from fellows which indicated that questions were being misinterpreted were used to reformulate questions for more uniform understanding. This application of mixed-methods analysis illustrates the developmental purposes of the approach.

Using mixed-methods data for triangulation

Triangulated change for a power was determined by matching a fellow's quantitative change to their qualitative stories. Change was considered triangulated if both quantitative change and a credible story were found for each fellow. These findings are referred to by using the phrase "*triangulated evidence of change in Power X*". This level of triangulation increases confidence in the credibility of the quantitative responses by requiring a credible narrative to support the quantitative data and decreases the effect of respondents who may not have responded thoughtfully to the Likert-type questions in the final evaluation form.

A similar approach was followed when triangulating the program's contribution to change. Quantitative responses were matched to qualitative responses. If a fellow indicated that the program had played a role in their quantitative response and had written a credible narrative in their qualitative response, the program's contribution to change was considered verified. These findings are referred to by using the phrase "*triangulated evidence of AWARD's contribution to change.*" The combination of quantitative and qualitative data allows for the number of fellows who provided feasible evidence to be combined with the rating of the strength of the program's contribution to identify the proportion of fellows for whom the program played a particularly strong role in facilitating the changes experienced. These cases were classified as having "*strong verified evidence*", which denotes the number of fellows who said the program played a major positive role in the change, and provided a narrative to support this assertion. By requiring an explanatory narrative of how the program contributed, fellows could not simply provide overly positive quantitative responses to "please" the evaluator.

The triangulation analyses were taken a step further to identify the number of fellows who had triangulated evidence of change as well as triangulated evidence of the program's contribution to the change. This is the highest level of triangulation, and represents the set of fellows for whom triangulated data is available for both the change they have experienced and the program's contribution towards that change. This "double" triangulation brings a high level of rigor into the findings and could be considered the best reflection of the program's contribution to facilitating change from the perspective of the fellows. These findings are referred to by using the phrase "*triangulated evidence of change in Power X as a result of AWARD*". Figure 3 illustrates the typical findings from this approach. The figure illustrates the percentage of fellows (for each cohort and by qualification level) who reported changes as a result of program activities (labelled as AWARD role), as well as the percentage of fellows whose data supported the assertion that the program played a strong verified role in these changes.

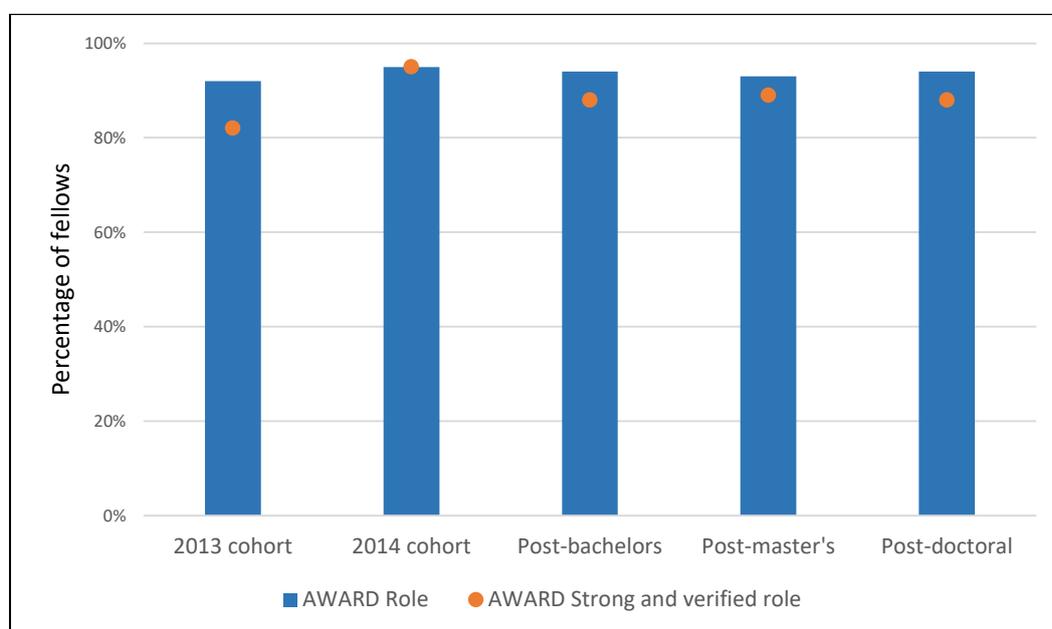


Figure 3: Sample findings on AWARD's contribution to change

Source: AWARD 2016 Evaluation Report

The combination of triangulated evidence of change and the triangulated evidence of the program's contribution is a measure to offset the fact that many fellows may have experienced changes in a specific power, but that these changes may be influenced by a multiplicity of factors (not only the program). As illustrated by the findings from the Power from Within, 93% of fellows *provided triangulated evidence of change in their Power from Within as a result of AWARD*. The program played a strong verified role in facilitating Power from Within changes for 89% of fellows. The slightly lower percentage of fellows (89%), illustrates the value of investigating change and the programs contribution towards change separately for a more realistic assessment of the program's role in facilitating change.

Using mixed-methods data for complementarity

As noted in the literature review, mixed-methods research can be used for the purpose of complementarity. This involves enriching the understanding of findings in ways that would not have been possible with mono-method approaches.

Triangulated results give the overall picture of change for the group in its totality, whilst qualitative stories provide context on the significance of the change for the individual. For each of the sub-domains within a power, the integrated findings were supplemented by revisiting the narratives that fellows provided so as to understand the significance of the change and the contextual nature of the change from the fellows' perspectives and using their voices.

Examples of fellows' narratives are presented below. The first excerpt is an example of a credible story that relates to the shifts a fellow experienced in her level of self-confidence. The concrete and verifiable actions taken because of the change qualify this narrative as a credible story. The text in [brackets] illustrate how the thematic coding would have been applied to an excerpt.

The AWARD fellowship has contributed a lot on positive changes in my life. Before AWARD I was timid and could not trust my own inner power so most of my decision depended on other people's decisions. But now I am self-confident and assertive person. [evidence of change in confidence] I went ahead and applied for Borlaug Laube Women in Triticum AWARD, and I am currently one of the awardees. [concrete action taken because of the AWARD fellowship which can be verified] Before AWARD I could have listened to other people's opinion, but now I believe in myself. (AWARD final evaluation form, 2013).

In the second excerpt the same fellow describes the way in which the program contributed to the changes in Power from Within (inner power), and notes specific programmatic activities which contributed to this change.

The AWARD fellowship experience played a concrete and vital role in influencing the changes I experienced in my inner power through the series of training I undergone during the fellowship. The most important and most interesting of all the training is the women leadership course. This serve as an eye opener for me and it really helped me in self-awareness and increased my confidence. (AWARD final evaluation form, 2012).

The coded narratives of how the program contributed towards change, enabled the evaluation team to identify the activities in the fellowship that made the most important contribution to change within a power (without prompting), and to translate these findings into quantitative summaries.

Based on these quantitative summaries of the coded qualitative data, the five most important activities contributing to a power were identified. Table 4 (Mentz, 2016) illustrates the three most valuable activities for the 2013 and 2014 cohorts for developing the Power from Within. In the case of Power from Within, the leadership training stood out as the most important activity regardless of phases, and the mentoring relationship and mentoring orientation workshop were

also identified as important.

Table 4: Most important fellowship activities for the development of fellows' inner power

	Cohort	Post-bachelors		Post-master's		Post-doctoral		All Fellows	
		n of fellows	% of fellows	n of fellows	% of fellows	n of fellows	% of fellows	n of fellows	% of fellows
Leadership training	2013	10	56%	12	52%	10	67%	32	57%
	2014	7	50%	13	54%	8	57%	28	53%
Mentorship activities	2013	9	50%	7	29%	4	27%	20	35%
	2014	3	20%	10	35%	5	33%	18	31%
Mentoring Orientation Workshop	2013	10	56%	4	17%	5	33%	19	33%
	2014	3	20%	10	35%	5	33%	18	31%

Divergent findings between cohorts could be further investigated by re-analyzing qualitative responses for the various groups, specifically with the intent to identify factors that may contribute to these differences. Additional sources of information, for example, course evaluations, could also be consulted to identify any challenges experienced in a particular cohort that may be influencing the results.

Discussion, recommendations and conclusion

The AWARD program has worked towards developing a rigorous and credible systematic approach to monitoring and evaluating the empowerment of its fellows. At a programmatic level, the result of this process is a more nuanced understanding of how the fellowships contribute to changes at the individual level and what changes can be anticipated during the fellowship period.

Although the program has made significant strides in refining the evaluation of fellows' empowerment, there are limitations to the approach, areas in need of improvement, and scope for additional research. A primary limitation of the approach is that both the quantitative and qualitative information is collected from the same source (i.e. a single questionnaire completed only by the fellow herself). Various measures have been taken to collect verifiable information, to triangulate and to rate qualitative contributions. However, the integration of information from additional sources (e.g. supervisor or mentor reports on a fellow's work) could strengthen the reliability of the findings even further.

There is also tremendous scope for approaching the wealth of qualitative information from an inductive perspective, rather than a deductive one. The primarily deductive approach taken by the program has strengthened the triangulation, but has limited other interpretations of the data which are beyond the scope of the deductive framework. Although we returned to the qualitative stories when assessing the complementarity of the quantitative and qualitative data results, this is not a substitute for a full, thoughtful, inductive analyses of the qualitative data themselves. This alternative has not been fully explored due to budget and time constraints.

At a broader level, the approach serves as an exemplar of how the application of mixed-method evaluation approaches could be applied in the context of women's empowerment in a

development context, in a manner that addresses accountability requirements (from donors or governments), while still providing voice and context-sensitive knowledge. There are numerous lessons that can be extracted from the author's five-year experience of working with the program on the design, refinement, and implementation of the approach to evaluate the empowerment of the female scientists.

Ensure a clear logic and purpose for selecting integrated mixed-methods. The logic and purpose of the selection should carefully consider the nature of the program and the context within which it is implemented, along with pragmatic elements such as resource availability and time. Well-implemented, rigorous, and integrated mixed-methods approaches are time and resource intensive. Using mixed-methods for the sake of mixed-methods is not adequate reasoning.

Invest time early on in conceptualizing data collection processes and tools. Although a mixed-method approach was embedded in the MEL from the very onset of the program, the development of a fully-integrated approach was the result of iterative processes of conceptualization and implementation. Data collection processes and tools should be given careful consideration during the program design phase and should be consistently refined – but ideally not overhauled entirely in the course of the project implementation. The more consistent data collection tools are over time, the greater the potential for comparability of results.

Consider analytic approaches during the planning phase. Using multiple tools for data analysis, such as Dedoose and SPSS, requires detailed analytic planning prior to analysis. Attention to extreme detail is needed to ensure that variables are coded or captured in a useful manner. Time is needed to conduct pilot analyses on small samples or dummy data to test the proposed approaches to integration prior to embarking on a large-scale analysis. Significant amounts of resources can be saved in this manner.

Select the tool most appropriate for analysis. Numerous tools are available (varying in cost and skill level required for use) for mixed-methods studies. In some cases, the differences between the tools are minimal, but in other cases, the analytic capacities of a tool can severely limit the evaluation. Careful consideration should be given to the (i) costs of use, (ii) the availability and affordability of human resource capacity to use the selected tool and, (iii) the future uses of the data. Furthermore, engage experts for advice and conduct desktop reviews when selecting the most appropriate tools in cases where new technologies are being considered.

Dedicate adequate time for analysis, reporting, and sense-making. The detailed level of integration illustrated in the approach necessitates allocating adequate amounts of time for data analysis, reporting, and sense-making activities. In this case, a typical time frame for processing data from a cohort of fellows is approximately six months, including time for quantitative analysis, qualitative analysis (and verification), data integration, and additional analysis. The detailed and technical nature of the analysis process requires careful reporting and communication of MEL results, an activity which should be factored into the project planning. In the program, results were reported in two steps. First, reports were presented to the program implementation team for discussion and sense-making. Secondly, analyses and report findings were finalized for broader audiences, including presentation to the program's steering committee. The process of presenting and refining MEL findings through discussions with program staff were highly beneficial to both staff and the MEL team.

Dedicate adequate amounts of resources (human and financial). The time and the high-level skills required to successfully implement an integrated mixed-methods analysis calls for the investment of adequate resources in the process, both human and financial. MEL project staff may not have all the requisite skills to implement studies of this nature and thus, outside expertise may need to be drawn upon during the design and implementation phases. Programs without dedicated MEL

staff may need to commission external resources which may be costly. The benefits and costs of employing an integrated mixed-methods approach should therefore be assessed prior to beginning the exercise. If deemed valuable and feasible, the exercise should be appropriately invested.

However, new approaches to MEL also present an opportunity for program staff to acquire new skills. In the current example, a proportion of the budget for the evaluation was allocated to internal staff development to capacitate staff internally to engage meaningfully with the evaluation findings and to conduct similar studies in future.

Document approaches and ensure processes that lead to credible and verifiable outcomes. One of the criticisms that has been levelled at mixed-methods approaches is the lack of clarity in describing and documenting processes and methodologies. Well-documented mixed-methods studies could contribute to the body of knowledge in the evaluation field and could also improve the implementation of the approach in the program. A well-documented process can increase the consistency of evaluations carried out by different evaluators and the fidelity of implementation of the approach in different settings.

Taking all of the above into consideration, the experience of the program confirms the assertion in literature that mixed-methods approaches are well suited to MEL that focus on a transformational agenda (Pereznieta and Taylor, 2014). Although pragmatic in approach, they do not represent a compromise in quality, rigor, or credibility if approached and implemented systematically and with adequate time and resource investment.

References

- Baird, L.L. (1976) *Using self-reports to predict student performance*. New York: The College Board.
- Bamberger, M. (2012) 'Introduction to Mixed Methods in Impact Evaluation', *InterAction*. Available at: [https://www.interaction.org/sites/default/files/Mixed%20Methods%20in%20Impact%20Evaluation%20\(English\).pdf](https://www.interaction.org/sites/default/files/Mixed%20Methods%20in%20Impact%20Evaluation%20(English).pdf)
- Bamberger, M., Segone, M. and Tateossian, F. (2016) 'Evaluating the Sustainable Development Goals: With a "No One Left Behind" Lens Through Equity-Focused and Gender-Responsive Evaluations', *EvalPartners*. Available at: <http://www.evalpartners.org/sites/default/files/documents/evalgender/Eval-SDGs-WEB.pdf>
- Bamberger, M., Rugh, J. and Mabry, L. (2012) *Real world evaluation: working under budget, time, data and political constraints*. California: SAGE.
- Beegle, K., Christiaensen, L., Dabalen, A. and Gaddis, I. (2016) *Poverty in a rising Africa*. Washington DC: World Bank.
- Beintema, N. and Di Marcantonio, F. (2010) *Female Participation in African Agricultural Research and Higher Education: New Insights. IFPRI Discussion Paper 00957*. Washington DC: IFPRI and Award. Available at: <https://www.asti.cgiar.org/pdf/ifpridp00957.pdf>
- Better Evaluation. (2016) *Better Evaluation: Sharing Information to Improve Evaluation*. Available at: <http://www.betterevaluation.org/en> (Accessed 15 March 2017).
- Better Evaluation. (2016) *Approaches*. Available at: <http://betterevaluation.org/en/approaches>
- Bradburn, N.M. and Sudman, S. (1988) *Polls and surveys: understanding what they tell us*. San Francisco: Jossey-Bass.

- Borden, V. and Zak-Owens, J.L. (2001) *Measuring quality: choosing among surveys and other assessment of college quality*. Washington DC: American Council on Education and Florida State University.
- Bourguignon, F. and Pereira da Silva, L.A. (eds.) (2003) *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*. Washington DC: World Bank and Oxford University Press.
- Bowman K. and Sweetman C. 2014. Introduction to gender, monitoring, evaluation, and Learning. *Gender and Development*. 22 (2) pp 201-212
- Cohen, L., Manion, L. and Morrison, K. (2011) *Research methods in education*, 7th edition. London: Routledge.
- Creswell, J.W. (2013). *Research design: qualitative, quantitative, and mixed methods approaches*. 4th edition. California: SAGE.
- Glaser, B.G. (1998) *Doing grounded theory: issues and discussions*. California: Sociology Press.
- Greene, J.C., Caracelli, V.J. and Graham, W.F. (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational Evaluation and Policy Analysis*, 11(3), pp. 255–274.
- Howard, G.S., Millham, J., Slaten, S. and O'Donnell, L. (1981) 'Influences of Subject Response Style Effects on Retrospective Measures', *Applied Psychological Measurement*, 5(1), pp. 89 – 100.
- Jobe, J.B. and Mingay, D.J. (1991) 'Cognition and Survey Measurement: History and Overview', *Applied Cognitive Psychology*, 5, pp. 175-192.
- Klatt, J., and Taylor-Powell. E. (2005). 'Using the retrospective post-then-pre design. Quick tips #27.' Madison: University of Wisconsin-Extension. Retrieved from <http://www.uwex.edu/ces/pdande/resources/pdf/Tipsheet27.pdf>
- Kuh, G.D. (2005) 'Putting Student Engagement Results to Use: Lessons from the Field', *Assessment Update*, 17(1), pp. 12–13.
- Lamb, T.A. and Tschillard, R. (2005) 'Evaluating Learning in Professional Development Workshops: Using the Retrospective Pretest', *Journal of Research in Professional Learning*.
- Lam, T.C. and Bengo, P. (2003) 'A Comparison of Three Retrospective Self-Reporting Methods of Measuring Change in Instruction Practice', *American Journal of Evaluation*, 24(1), pp. 65-80.
- Levin, K., Cashore, B., Bernstein, S. and Auld, G. (2012) 'Overcoming the Tragedy of Super Wicked Problems: Constraining Our Future Selves to Ameliorate Global Climate Change', *Policy Sciences*, 45(2), pp. 123–152.
- Lucks, D., Schwndt, T., Ofir, Z., El-Saddik, K. and D'Errico, S. (2016). Counting critically: SDG follow-up and review needs interlinked indicators, monitoring and evaluation. (IIED Briefing papers). Available at: <http://pubs.iied.org/pdfs/17363IIED.pdf> (Accessed 10 January 2017).
- Mentz, M. (2016). AWARD Phase II Monitoring and Evaluation Report.
- Mertens, D.M. (2007) 'Transformative Paradigm: Mixed Methods and Social Justice', *Journal of Mixed Methods Research*, 1(3), pp. 212–225.
- Miles, M.B., Huberman, A.M. and Saldana, J. (2013) *Qualitative data analysis: a methods sourcebook*. 3rd edition. California: SAGE.

- Noordeloos, M. (Ed). (2015) *Empowering African women scientists through career-development fellowships*. Nairobi, Kenya: African Women in Agricultural Research and Development. Available at: <http://awardfellowships.org/joomla-website/images/downloads/publications/AWARD%20Empowering%20African%20women%20scientists%20through%20career-development%20fellowships-reduced.pdf>
- Olofsgård, A. (2014) *Randomized controlled trials: strengths, weaknesses and policy relevance*. Stockholm: Fritzes Customer Service.
- Pace, C.R. (1985) *The credibility of student self-reports*. University of California: Center for the study of evaluation.
- Patton, M.Q. (2008) *Utilization-focused evaluation*. 4th edition. California: SAGE.
- Pereznieto, P. and Taylor, G. (2014) 'A Review of Approaches and Methods to Measure Economic Empowerment of Women and Girls', *Gender & Development*, 22(2), pp. 233–251.
- Pike, G. (1995) 'The Relationship between Self-Reports of College Experiences and Achievement Test Scores', *Research in Higher Education*, 36(1), pp. 1-21.
- Porter, S. R. (2009, November). *Do college student surveys have any validity?* Paper presented at the Association for the Study of Higher Education, Vancouver.
- Raidl, M., Johnson, S., Gardiner, K., Denham, M., Spain, K., Lanting, R., Jayo, C., Liddil, A. and Barron, K. (2004) 'Use Retrospective Surveys to Obtain Complete Data Sets and Measure Impact in Extension Programs', *Journal of Extension*, 42(2). Available at: <https://www.joe.org/joe/2004april/rb2.php> (Accessed 20 March 2017).
- Rittel, H.W.J. and Webber, M.M. (1973) 'Dilemmas in a General Theory of Planning', *Policy Sciences*, 4(2), pp. 155–169.
- SDG Fact Sheet. (2015) *Sustainable Development Goals: Fact Sheet*. Available at: http://www.un.org/sustainabledevelopment/wp-content/uploads/2015/08/Factsheet_Summit.pdf
- Shadish, W.R., Cook, T.D. and Campbell, D.T. (2001) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Strawson, T. (2015) *Improving ODA allocation for a post-2015 world*. United Kingdom: UNDESA. Available at: http://devinit.org/wp-content/uploads/2015/01/DI_UN_Improving-ODA-allocation-for-a-post-2015-world_21-January-2015.pdf (Accessed 20 March 2017).
- Tourangeau, R. (1984) 'Cognitive science and survey methods: a cognitive perspective', in Jabine, T. et al. (eds.) *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*. Washington DC: National Academy Press, pp. 73-100.
- Trochim, W.M.K. (2006) *Research methods knowledge base*. Available at: <https://www.socialresearchmethods.net/kb/> (Accessed 20 March 2017).
- Vogel, I (2012) *Review of the use of 'Theory of Change' in international development: Review report*. UK: Department for International Development. Available at: http://www.theoryofchange.org/pdf/DFID_ToC_Review_VogelV7.pdf