Exemplar

Monitoring and Evaluation of African Women in Agricultural Research and Development (AWARD): An Exemplar of Managing for Impact in Development Evaluation

Paul R. Brandon[1], Nick L. Smith[2], Zenda Ofir[3], and Marco Noordeloos[4]

Keywords

development evaluation, monitoring and evaluation, Africa, women in agriculture, theory of change

Introduction

In this Exemplars case, the fifth and final under the direction of the current coeditors, we present a reflective account of an ongoing, complex, multiyear, multinational monitoring and evaluation (M&E) system conducted for African Women in Agricultural Research and Development (AWARD), an international development program. The program provides African female scientists in agriculture with professional development intended to influence the agriculture sector, and the M&E system supports a *managing for impact* approach to bring about change in individuals and groups in the short term and in institutions and the agricultural sector in the long term.

The preparation and writing of the case was a collaborative effort of the four authors. As in the four most recent Exemplars cases, we begin with a description of the program and evaluation system, followed by an amalgamation in an interview format of the extensive evaluator–editor dialogue that occurred while preparing the case. We conclude with the authors' reflections.

[1] University of Hawaii at Manoa, Honolulu, HI, USA
[2] Syracuse University, Syracuse, NY, USA
[3] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch, South Africa
[4] African Women in Agricultural Research and Development, Nairobi, Kenya

Corresponding Author:
Paul R. Brandon, University of Hawaii at Manoa, 1776 University Avenue, UHS2-214, Honolulu, HI 96822, USA.
Email: brandon@hawaii.edu

Program Description

Rationale and Purpose

Women farmers play an essential role in African agriculture, doing much of the work to produce, process, and market food (Food and Agriculture Organization of the United Nations, 2011). AWARD's benchmarking research across 125 institutions of agricultural research and higher education showed, however, that fewer than one in four professionals are women and that fewer than one in seven of those holding management positions are women (Beintema & Di Marcantonio, 2010). Thus, in recent years, there have been numerous calls for increased leadership roles for women in the African agriculture sector (e.g., Forum for Agricultural Research in Africa, 2006; The World Bank, 2009).

To help deal with this issue, AWARD was established to provide career development to top African women scientists, so that they can contribute to poverty alleviation and food security at all levels of the agriculture sector and to strengthen the voice of African women in agriculture (the *fellows* served by the program) on the farm, in laboratories, in markets, and in policy forums. At the foundation of AWARD is the belief that skilled women leaders are able to offer different and essential insights on the priorities and approaches needed in African agriculture (AWARD, 2009). The program's two primary objectives are to (a) equip women to increase their contributions to African agricultural research and development by making them technically stronger, better networked, and more confident and visible; and (b) close the gaps in information and knowledge about African women in agricultural research and development through research, "vigorous monitoring, evaluation and impact assessment," and training (AWARD, 2010, p. 21). The program is built on the notion that the challenges of small-scale farming require scientific innovation and a new type of leadership at all levels of the sector. It is one of few programs in Africa to emphasize equally the further empowerment of well-educated women and the systematic cultivation of new knowledge for the benefit of the sector and for the development and evaluation communities. The achievements of its fellows can also encourage the agricultural research and development sector to be more responsive to the needs and contributions of women.

AWARD offers 2-year fellowship packages to women with bachelor's, master's, or doctoral degrees. It applies no age limits and tailors the fellowships to the needs of the participants. The demand for AWARD fellowships has been substantially increasing since the program's inception. From 2008 through 2013, AWARD received applications from 3,502 women scientists in some 500 organizations, who competed for a total of 390 available fellowships allocated in cohorts of 60 to 70 per year.

The program was initiated after a 3-year pilot project in Kenya, Uganda, and Tanzania, funded by the Rockefeller Foundation and the U.S. Agency for International Development (USAID). Drawing from lessons learned during the pilot project, an expanded, 5-year program was launched in 9 (later 11) Anglophone sub-Saharan Africa countries in 2008, supported by approximately US$18 million from the Bill and Melinda Gates Foundation and USAID. Supplemental funding was later provided by the Agropolis Foundation and the Alliance for a Green Revolution in Africa. The program has completed its first 5-year phase and was awarded US$22 million for a second 5-year phase starting in 2013. A pilot project offering fellowships to nationals from Francophone Africa has also started during the current year.

Management and Participants

A 13-member steering committee with geographic representation from key partners on the continent, funding agencies, and fellows' institutions oversees AWARD. The steering committee formed an M&E subcommittee to provide periodic advice and guidance in internal M&E discussions. The

program is implemented by a management team of 14 people located in Nairobi, Kenya. The team implements more than 25 training, monitoring, and learning events a year, supported by 22 African trainers who, as part of AWARD's commitment to ensuring sustained benefits on the continent, have taken over all activities in AWARD from the international trainers who were initially engaged.

Since 2008, AWARD has served 320 fellows (with another 70 selected in 2013 as of the writing of this article) from widely differing contexts—42% from universities, 36% from agricultural research institutes, 13% from government, 8% from nonprofit or humanitarian agencies, and 2% from the private sector (adding to 101% due to rounding error). Some of the fellows grew up in remote villages; others are from major cities. They range from recent bachelor's graduates to recognized experts in senior management and leadership positions. The youngest fellow to date has been 22 years old, and the oldest 58. About 65% are mothers at the start of the fellowship, with 31% having at least one child under the age of 5 years. The program provides baby and nanny support during its training and workshop events.

The AWARD developers designed the program to help ensure that, in addition to the fellows, there are others who can gain from the program, with the intent to create a multiplier effect. Each fellow is matched with a carefully chosen mentor—a senior scientist or leader in a university or research organization—who also participates in AWARD's capacity-strengthening activities. Furthermore, AWARD fellows "share forward" during their second year by mentoring junior women scientists, called fellows' mentees, and conducting role-modeling events. When considering all fellows, mentors, and fellows' mentees to date, AWARD has directly engaged with 715 scientists in nearly 20 countries. Through its implementation, the program is linked to approximately 200 organizations in Africa, including leading agricultural research and development institutes, universities, science laboratories, and companies. In addition, more than 26,000 people, primarily female high school students, have been reached through role-modeling events by AWARD fellows. The program has also received considerable attention in the news media, as shown by a lengthy list of news items on its website (AWARD, 2013).

Program Components

Since the beginning, the program has included three key components or *cornerstones*. The first, Fostering Mentoring Partnerships, provides (a) a 5-day mentoring workshop for the fellow–mentor pairs, tailored for an African context while drawing from international experience; (b) monthly mentoring meetings for a year; and (c) an opportunity to mentor a junior woman scientist of the fellow's choice during the second year of the fellowship. The mentors are senior male and female scientists or leaders in the sector. Each fellow is required to prepare with her mentor a *purpose roadmap* describing her vision for her life and career future, thus helping to direct the mentors' guidance. The second component, Developing Leadership Capacity, is a leadership course, also tested internationally and tailored to the African context, that provides intensive training to the post-bachelor fellows for 5 days and to the post-master's and postdoctoral fellows for 7 days in interpersonal skills (focusing on personalities, styles, gender, and other diversities), communication skills, conflict management, and strategies to influence and build alliances. In addition, AWARD provides support to enable fellows to hold "role-modeling events" in schools, organizations, or communities. Some are also given an opportunity to represent AWARD at national, regional, or international events. The third component, Building Science Skills, provides a menu of options depending on need, including a laptop computer and Internet service, membership in a professional association, attendance at a science conference, a course in science writing for publication or fund-raising, a short course in gender-responsive agricultural research, and a highly competitive opportunity to undergo advanced science training over several months in an advanced research organization on the continent and around the world. Some of these elements were adjusted for the second phase, based on lessons learned during the first.

Monitoring and evaluation. Beginning with its second phase, a fourth component, Monitoring and Evaluation, has had equal status with the other three. The component reflects the program's focus on *managing* for impact (Guijt & Woodhill, 2002), rather than only *measuring* impact. It therefore stresses *adaptive management*—an approach emphasizing timely adjustments as contexts change, new information emerges, and weaknesses or gaps in program design or implementation are encountered—as well as accruing knowledge useful for the long term. The M&E system shifts attention from a simple primary preoccupation with achieving immediate or long-term impacts to promoting approaches, processes, and systems that are likely to achieve, enhance, and sustain positive impacts, both during program execution and after it has been terminated. It also requires that primary stakeholders have a feeling of ownership of the M&E and that program participants understand its value. The M&E system thus addresses and challenges (a) a reductionist approach to development evaluation and obsession with measuring impact; (b) the widespread overemphasis on accountability, reporting, and control; and (c) the notion that outcomes and impacts are largely linear and predictable. (We use the word *outcomes* for short-term and intermediate effects and *impacts* for long-term sector or societal effects.) It also addresses the need for gathering knowledge that can help strengthen management and scaling towards sustained positive impacts.

In its pilot phase, AWARD had experienced the consequences of the inadequate data about women in agricultural research and development that had been collected for many years before the program began. The management team had a strong sense from the pilot experience about the value of collecting useful data. This experience was part of the foundation for the program's core objective to close the gaps in information and knowledge about African women in agricultural research and development, as well as part of the reason why funding agencies supported M&E generously with about 11% of the AWARD budget—a percentage at the high end of the range of allocations for monitoring and evaluation of development evaluations.

Development of the M&E system. At its inaugural Phase I meeting in 2008, the AWARD steering committee opted for an evaluation approach that was rigorous, use-focused, and guided by a set of principles that stated, among others, that the M&E activities were to be useful multidirectionally, not only to address the accountability requirements of funders but also for continuous program improvement and knowledge generation. Empowering the African stakeholders was its first priority. Stakeholders were considered to be in three groups: *primary* stakeholders, including the AWARD team and its implementing partners (e.g., trainers), the program steering committee, program funders, and fellows; *secondary* stakeholders, including mentors, fellows' mentees, and the participating institutions; and *tertiary* stakeholders who, although not program participants, are the funders, designers, and implementers of other fellowship, mentoring, science, and leadership policies and strategies, as well as communities of practice (e.g., M&E or gender in agriculture) with such priorities. The guiding principles specified, among others, that the M&E system would

1. apply appropriate and innovative methods rigorously for timely and credible results. Stakeholder perceptions and insights would be respected, and stakeholders would be engaged in setting up the system, with systematic triangulation helping to ensure credible data and information.
2. be cognizant of the complexity of the program, with an open-systems perspective to allow insights into interactions, both predictable and unpredictable, among the interventions and the different contexts.
3. build useful knowledge by trying to understand why, how, for whom, and under what circumstances the program operates effectively. What makes for success or failure would, within the limitations of this type of program, be as important as determining outcomes and impacts.

4. be efficient. Data and information would be available when needed, efficiently delivered in formats that speak to different audiences, and used for ongoing incremental and strategic adjustments to the program design and implementation, as well as for scaling up.

An initial effort in 2008 to have a consulting firm use outcome mapping to help the first round of fellows identify expected changes was unsuccessful. With the appointment in 2009 of an M&E coordinator in the program team (replaced in 2011 by the fourth author), assisted by a part-time external evaluation advisor acting as an internal evaluator (the third author), a renewed effort was made to establish a working M&E system. The program design, the evaluation of the pilot project, and the evaluations of similar programs together provided enough substance for a credible theory of change, but this top-down logic had to be complemented by program participants' bottom-up insights of expected changes. A consultative process with fellows, mentors, and the program team led to a comprehensive results framework that spelled out short-term, intermediate, and long-term impacts at the individual, organizational, and sector levels, complemented by a set of detailed assumptions. The results were not seen as appearing in a linear fashion but organized in a *sphere of control*, *sphere of influence*, and *sphere of interest*—outcome mapping concepts that gave a sense of progression while providing for feedback loops and complicated connections among change pathways. All impact pathways were not predicted, as it was seen as important to allow for and track emergent developments. The M&E system was to help confirm those pathways most critical for success, clarify the interconnectedness of the different program interventions where such pathways were difficult to predict, try to identify tipping points, and highlight the role of external contributions to change. It also studies whether *transformative change* is being achieved. An empowerment model in the literature (Rowlands, 1997; Solava & Alkire, 2007; VeneKlasen & Miller, 2007) was later adapted for AWARD's focus on leadership in science. It matched almost perfectly the changes fellows were predicted to experience, thus further strengthening the change hypothesis.

The innovative combination of M&E system elements is unusual in development evaluation in three primary ways. First, it uses a credible, though partial, theory of change founded on stakeholder insights, several available evaluations of leadership programs (Center for Creative Leadership, 2007), an empowerment model from the literature, and a program design based on the study of previous programs of a similar nature. It can be used to guide monitoring and for deductive analysis and learning, yet also allows inductive analysis and for identifying unpredictable patterns, using practices similar to developmental evaluation. Second, the system provides for extensive monitoring, self-evaluation, and internal evaluation of the design, implementation, and emerging program effects, drawing from successes and mistakes, while also focusing on frequently neglected aspects such as transformation and sustainability. Longitudinal tracking is to be established in Phase II. It also provides for external, possibly independent, evaluation. Special evaluative studies on topics such as social return on investment, institutional change, and sustainability are to be conducted. An external summative evaluation will be conducted at the end of Phase II. Third, the system uses a variety of mixed methods to collect rich data and a variety of methods for systematically analyzing the data. The evaluators are using aspects of realistic evaluation (Pawson & Tilley, 1997) and contribution analysis (Mayne, 2011; Stern et al., 2012; White & Phillips, 2012) for tracing causal pathways from the individual to the sector level. The data are examined for unintended—especially negative—consequences, as part of a focus on how best to enable and sustain program ideas and impacts. Furthermore, the analyses will focus on connections between change and context, using up to 13 variables for fellows, as well as studies at organizational and national levels. To ensure systematic reporting of findings, the M&E team synthesizes in a set of brief summaries all analyzed data and information as they emerge, organized by topic for planning, improving, reporting, or sharing further. The findings are intended to assist in both operational and strategic decision making on an ongoing basis, such as when they were used to inform the design of AWARD Phase II.

The M&E team. The AWARD leadership made a major effort to ensure that the responsibility for monitoring and evaluation was not only carried by the M&E Coordinator, but by the program team as a whole. However, as a result of team members' heavy workload and capacity constraints in the initial phase of the program, this did not always work well, and a significant amount of data analysis had to be outsourced during the first phase of the program. In November 2008, an independent evaluation specialist, the third author—who had done the evaluation of a pilot program in 2007 that led to AWARD—was officially appointed as an M&E advisor to the AWARD steering committee. Since October 2009, she has frequently played the role of a part-time internal evaluator, initially guiding the theory of change and M&E system development and later engaging in some of the more substantive data analyses. Staffing changes eventually led to a stronger internal M&E team, with a more senior M&E Coordinator (the fourth author) supported by an M&E Officer and two program assistants. Since 2012, they have taken over much of the internal evaluator's role.

Instruments and data collection. The M&E team developed the data collection instruments with input from the program team and participants. Data collection focuses on the quality and quantity of implementation progress and on expected and unexpected outcomes and impacts. The frequency of data collection is determined by the nature and use of the data. Implementation data are collected immediately after activities or events, and descriptive data, progress data, and achievement data are collected at the beginning, after 1 year, and at the end of the 2-year fellowship. Special studies are conducted when needed. Long-term longitudinal data tracking of the first and subsequent rounds of fellows began in 2013; its implementation was slower than expected. Findings are documented in a series of brief reports and updated as new data become available. They are intended to serve as a basis for communication in different formats for different purposes and audiences; their formats are currently being improved, with better use of interactive online techniques.

The large number of data sources available from fellows includes their curriculum vitae and application forms, baseline and progress journals, course and event evaluations, contributions to progress monitoring meetings, e-mail messages, and impact stories prepared by the fellows. These are complemented by mentors' and fellow mentees' reports about the changes they have experienced, as well as those they have observed in fellows; component leader and team reflection session records; dashboard tracking of milestone achievement; and rare external stakeholder surveys.

The M&E team triangulates reports among accounts of the fellows, the mentors, and the fellows' mentees and also during case studies with others who are at arm's length from the program.

Analysis and use of evaluation results. The steering committee used the findings at the end of the first phase of the program to make some fairly important modifications in the proposal that they submitted for Phase II. In addition, they used the results as part of the adaptive management of the program. The M&E findings have also fostered reflection and discussion within the management team—not only when producing the year-end program review but also in near-real time. One example is that the program revised its fellowship package for post-bachelor's fellows, as it learned through M&E data that its assumptions about the practical context and needs were not accurate (e.g., a number of post-bachelors fellows did not match the "junior researcher" profile but were actually heading a lab or research department). A second example is a renewed and enriched understanding of the mentoring process. Through comprehensive analysis of mentoring-related M&E data, the program found, for example, that the male and female mentors were equally effective (for the female fellows) and that the second year of mentoring was not nearly as effective as the first. As a result of these and related findings, in Phase II, AWARD dropped the second year of the post-bachelors fellows' mentoring and provided additional incentives for male mentors to serve in the program. Third, fellowship application forms have provided not only a means of selecting participants but also have become a valuable piece in the M&E system. To date, the program has more

than 3,000 vitae on file that comprise a data set offering many different research and evaluation angles. Furthermore, the program receives frequent requests from agricultural organizations, donor agencies, and their partners about snippets of the evaluation results about who has done certain things and who has shown progress. The first round of analysis of M&E data provided the foundation with a refined understanding of the AWARD fellows and their specific contexts. Since 2012, AWARD has used a much more refined matrix of an in-country talent pool to guide its selection process. Finally, knowledge has been developed about the factors affecting success, the change logic, and the synergistic effects of the program. With more work in Phase II on the interaction with context, the management team expects that transfer to other contexts will be easier.

The M&E team does not see monitoring and evaluation as one-way traffic to stakeholders. Instead, they promote an interactive process of engagement that, when feasible, uses the mentors, fellows, and fellows' mentees to help interpret unclear findings. This has been a challenge. Initial difficulties with ineffective data collection systems, data gaps, and insufficient capacities and time to conduct such processes effectively stymied such efforts, yet also provided significant lessons on how to do this better. Consequently, not all needs of all stakeholder groups could be addressed during the first phase. The M&E team is structuring the annual monitoring meetings to be much more useful to the participants and hopes to motivate the newly formed AWARD alumnae network to engage with analysis and interpretation. During Phase II there will also be a much stronger effort to disseminate new insights in innovative ways.

Interview

Editors: To what extent is using a detailed theory of change unusual for evaluations in developing countries?

Evaluators: When we started in 2009, it was still uncommon, certainly in terms of how we have been using it. Around that time, the Network of Networks on Impact Evaluation (NONIE) and the International Initiative for Impact Evaluation, commonly known as 3ie, started to promote theory-based impact evaluation. This gave it a higher profile. Unfortunately, theories of change are still mostly being used for control in development initiatives. They tend to be unsophisticated and have become prescriptive rather than evolving with implementation. In recent years, there have been serious critiques of the frequent misuse of the log frame approach (LFA) in development planning and evaluation. Hummelbrunner (2010) quite rightly refers to it as "logic-less frame," "lack-frame," and "lock-frame." He notes that the LFA often fails to reflect development's "messy realities," something with which we heartily concur. We find that logframes are primarily done for donors during the proposal phase and then put on a shelf until reporting time, when everyone runs around trying to count many things that are supposed to be entered into a few columns in a table.

Editors: Has your use of a theory of change met your expectations?

Evaluators: It has exceeded our expectations. M&E should enlighten stakeholders and generate useful knowledge about development. A theory of change approach is one way to ensure this while also giving direction to the program implementers. We have tried to balance a deductive and inductive approach, because we wanted to be rigorous and systematic by connecting upfront planning and clarity of direction, yet allow for emergence.

AWARD is an integrated, complicated program, with different interventions expected to contribute to a large number of intermediate outcomes at individual and institutional levels, as well as at the sector level. With the theory of change, we could show that AWARD is in part so successful because each component leads to multiple changes and also enhances other components, similar to the synergistic effect of different chemicals in a mixture. We believe we can use this information to show that AWARD is more than the sum of its parts and thus a good value proposition. It is also a relatively expensive program, so we need to determine whether all its components are necessary to achieve the

desired impact. Working with a nonlinear, quite detailed theory of change has proved to be essential for this purpose.

The graphical display of the theory of change has been particularly helpful. There was a real "aha moment" when everyone in AWARD suddenly saw the logic of the program design clearly spelled out through their own efforts and how the theory of change related to the outcomes and assumptions. Another aha moment came when we found that the theory of change resonated very well with an empowerment framework in the literature. There was almost complete overlap between this framework and the changes expected at the individual level. This confirmed to us that the designers were very experienced, very well informed by the literature, or both. In our experience, theories of change will be much better if we draw on what is known from well-documented frameworks and experiences. So this very nice blending of bottom-up and top-down approaches to theory of change development greatly strengthened its credibility.

Editors: How useful are you finding your theory of change in explaining why AWARD is and is not working?

Evaluators: We have found it to be extremely useful. We could have followed an outcome mapping or developmental evaluation approach without thinking in advance about the causal model. But making it explicit has helped to guide the whole M&E system and to test it in quite a rigorous way.

One of the challenges with this type of program is that fellows enter at different stages of their professional life and respond differently to program interventions. In spite of this, we have been able to identify some fairly generic impact pathways and some important success factors. We can link a specific activity to a specific set of intermediate outcomes or an outcome to several activities. We discovered some reinforcing feedback loops and several patterns per grouping—for example, at the post-bachelor's, post-master's, and postdoctoral levels. We are still working on disaggregating the data.

Editors: You said that you found an empowerment model that was an excellent match with your theory of change. Tell us more about the purpose of this aspect of AWARD.

Evaluators: Some months after we developed the theory of change, we came across a framework in the literature that resonated very well with the outcomes in the theory of change, that related to change at the individual level, and that could be readily adjusted for AWARD's focus on leadership in science. The framework treats empowerment as an expansion of "agency," or what people are free and able to do and achieve in pursuit of their goals or values. It postulates that there are four possible displays of agency that can lead to empowerment as a leader in science: "Power from within" involves a fellow's growth in inner strength—in her willingness, confidence, and motivation to induce change in line with her own vision and values. "Power to do" refers to her increasing access to resources and capacities to progress in her professional life. "Power over" involves a fellow's growing ability to exercise control over professional and personal decisions, over resources, and being better able to deal with professional or social power relations and hierarchies. "Power with" involves a fellow purposefully focusing on advocating for, and enabling change collectively with others. When we analyzed the individual empowerment components of our theory of change, we could fit them completely into this framework, thus strengthening their credibility. Theories of change use too few actual theories derived from practical experience!

The theory of change provides the foundation for monitoring, reflection, and evaluation, but we make sure it is not a "recipe" or "template" for data collection. So we ask a lot of open questions in order to get a nuanced understanding of the program participants' experiences, reactions, and relationships. We deductively and inductively analyze patterns in the rich qualitative and quantitative information. Importantly, the theory of change helps AWARD to focus its knowledge generation; it helps us to know what we know, what we don't know, and what we can still know.

Editors: To what extent is the program actually using the information being provided by the M&E system?

Evaluators: This is very important. Some examples are given in the Description section. However, with all the challenges the M&E system faced, it has been difficult to serve stakeholders beyond the program team and steering committee. The team has little time for anything beyond immediate implementation of about 25 training events and meetings throughout Africa every year. The program also has to achieve approximately 25 additional annual milestones, so the individual participants' involvement in working with the data has been less than desired. Still, they appear to have benefited significantly from their engagement with the theory of change and M&E design process. They continue to use the data for both strategic and operational purposes, and they complain bitterly when information for their immediate use is not available. We are still struggling to provide information in a timely manner and to match it with the significant data needs of the team and steering committee. We intend to get this right during the second phase.

Implementation of efforts to make M&E useful for the fellows was initially poor. Only in the last year have the fellows started to buy in and understand its utility. M&E is not a common practice in scientific research environments! We have developed a session with fellows and their mentors to inspire them about the utility of credible M&E. We know it is now effective (after several less successful efforts) because of much higher ratings for the quality of this component in the first week-long exposure of fellows to AWARD, as well as better responses to our data collection requests. We are also going to engage fellows and mentors through the new alumnae network and in the annual monitoring meetings to help us understand some unclear findings. Of course, our assumption is that when we build monitoring and self-reflection capacities, we are also strengthening the program team and participating fellows and mentors in these much-needed skills. We are aware that this type of approach may influence the M&E data, so we limit these engagements. There is a trade-off, but we believe the benefits outweigh the risk.

It is worth noting that the program team is extremely committed to their work. Their engagement and experience provide them with intuition and observation that complement the evidence. Decisions are frequently not only based on external evidence, but reflect their first-hand experience as well.

Editors: Since you have adopted an unconventional and ambitious approach to conducting development evaluation, no doubt you have learned a lot that might help others who adopt it. Can you tell us any more about the difficulties and benefits of your approach?

Evaluators: Our main challenges and risks are intertwined and relate primarily to M&E capacity and resources. A comprehensive M&E system that is realistic and locally owned, while remaining credible and useful enough for the desired purposes, needs people with diverse capabilities. Triangulation needs to be rigorous, which is difficult to achieve with limited resources. We need to measure change in a convincing manner among different cohorts of very diverse fellows from diverse contexts. Their change trajectories can vary significantly. We still tend to skim the surface. We need larger numbers of fellows for disaggregated data, and we need more comparative institutional case studies to understand some of the patterns and nuances that make for success or failure in specific contexts. Measuring the many interrelated intangible variables such as confidence, motivation, leadership, and influence requires significant amounts of verifiable qualitative information. Many of the tangible changes at the individual level, and especially at the sector level, will emerge long after AWARD has ended. And of course, such lengthy causal pathways will be increasingly difficult to trace through process tracing or contribution analysis, but perhaps not impossible.

We now know enough about the theory of change to focus more on the institutional and sector level impacts. This will help limit the scope of our data collection, but also shift focus to the longitudinal data collection—with all the challenges this will entail!

Understandably, donors have a main interest in accountability which they assess through reports, but we see that as only one of many benefits of evaluation. Fortunately, our funders agreed that we needed to experiment with an M&E system that can serve different purposes. They placed its

ownership completely in our hands but also ensure that the work we do is credible. At the same time, we are completely transparent about our own successes and failures. Most program teams and evaluators do not have this luxury. Many are currently under pressure to implement impact evaluations using conventional experimental designs that they believe are not the most appropriate or will not yield sufficient benefits. Many of these experiences are not made public so that others can learn from them. This means that a lot of evaluation resources are wasted.

Editors: Tell us more about these challenges—the problems that you have dealt with in the program and what difficulties you expect to encounter as the work progresses.

Evaluators: We can talk for days about this, with many layers to unpack, but let's summarize those that come to mind immediately. First, there has been insufficient stable M&E capacity in the team from the beginning. The approach is intensive and requires quite a wide set of evaluation expertise. It is not ideal to depend upon consultants who do not "live with implementation." The most critical consequence has been that there was insufficient attention early on to the basic systems needed to manage all the data and information AWARD produces. If an implementation team has never been exposed to good knowledge management systems, it's hard for them to understand their benefits. They need to be convinced; we are starting to do this now. We are working to embed tailor-made systems in the day-to-day work so that we increase efficiency and effectiveness in data collection, analysis, and communication.

Second, our data management, including analysis, is still largely outsourced. We are slowly changing to more in-house work where this makes sense. We believe that even in programs of this relatively modest size, the implementation team should have sufficient M&E expertise in-house to do the majority of the work and control any engagement with external evaluation expertise. External and independent evaluators and consultants should only be used for highly specialized areas of work that offer specific challenges or for greater independence.

Third, the management team has been keen to learn, but they have been under huge time pressure, and they haven't been able to engage as they ideally should. This year we want to build in more regular and active opportunities for the different primary stakeholders to engage with the M&E data. The scientists who enter AWARD generally do not appreciate the process and value of M&E, so we are developing innovative ways to engage them in feedback and in certain analyses. Our intent is that they see the real value of frank self-assessment, coupled with more external and independent evaluation. We intend using annual progress monitoring meetings for this purpose, where fellows and mentors from several countries come together for 2 days. We also will use the alumnae network that will soon be established.

Fourth, we are collecting a significant amount of qualitative information and analyzing it systematically, adding a lot to our work. It has been a challenge to prioritize collecting and analyzing manageable amounts of data for a leadership and program team who want evidence for everything, including why, how, for whom, under what conditions, and at what cost, the program works—and what might work when scaled. So we have to try to connect the success and failure factors with context. That is a challenge. Initially, we watered down our instruments to focus on selected quantitative information, so that participants were not burdened with too much work. Later, we included more qualitative information. So we have a very significant focus on how to make our data collection and analysis systems more efficient.

Finally, the problem we find hardest to resolve is the integrity and utility of the data from the first few rounds, when we were still struggling and experimenting with the best approach. Most data collection systems in development evaluation struggle with credibility because people do not feel their ownership and utility and therefore do not care about data quality. In our case, this led to gaps in data. A still greater challenge was the fact that the key concepts were not clear and our vignettes or rubrics were not sufficiently developed to cultivate common understanding when rating. Definitions for terms such as "gender responsiveness," "innovation," and "influential" needed

to be clarified and shared early on. Fortunately, we have qualitative information that helps us assess how some of these concepts were initially understood, but it still makes comparison with later rounds difficult. We moved away from an initial focus on overly complex progress markers (a useful outcome mapping concept). This further complicated comparison between rounds. We also needed to provide structure to impact stories, which meant we ran the risk of leading participants somewhat in their responses.

Given all these early mistakes, we might appear not to provide much of an exemplar! However, these are challenges many programs face in the field. One could say that we have been exemplary in acknowledging and facing them head-on, without waiting for others to point them out. In the process, capacities have been cultivated and much has been learned. We can now share these experiences with others first-hand in a systematic and detailed manner. In the end, this creates a ripple effect, adding value to AWARD's achievements. We are quite confident that these issues will be resolved for the second phase, although we expect that the longitudinal tracking, which is to start in 2013, will bring some new challenges. Our experience shows how long it can take to build and embed the systems and capacities to serve a program well in countries where a critical mass of M&E expertise still has to be nurtured. For us, it adds to the ongoing argument that short-term, time-bound funding of fragmented interventions is detrimental to sustained development. We are grateful that AWARD's funders are supporting a decade-long journey!

Editors: You work in multiple countries and with numerous cultural groups. Have you encountered any particular ethical or cultural issues that have required special attention in the M&E work?

Evaluators: A practical issue stems from our use of extensive forms to capture participants' impact stories and examples of progress. The emerging stories on the forms don't always match the rich picture one hears during interviews. This is probably a universal challenge, but perhaps even more so in a particularly oral culture such as in sub-Saharan Africa. The positive side of this is that our results likely underplay AWARD's effects. We are going to try to address this issue as we move forward in the second phase.

Another issue is that we cannot always assume that all those to whom the fellows report want the best for the fellows. Power dynamics are at play, especially since most of these supervisors are men. There are complicated and sometimes very sensitive issues in our work context. There's a potential for jealousy and backlash in an environment where significant growth opportunities like those provided through an AWARD fellowship are scarce. We need to take this into account in our data collection strategy, particularly when considering a context in our interpretation of M&E data. Of course, we are very careful to keep sensitive information connected to specific individuals confidential.

We have become better at considering the specific context of each unique individual in the AWARD program when analyzing M&E results. The starting points, cultural contexts, and personal and professional environments are very different for different individuals. Simply aggregating participants by year of fellowship or education level often leads to either diluted or simply misguided conclusions. Now that we have data on a larger group of fellows, the second phase will also allow us to do more refined analyses of the influence of both institutional and country contexts. This should inform AWARD's upcoming effort to adjust the program for implementation in Francophone Africa.

Editors: Finally, to round out the picture, can you say a little about how your M&E system is different from other M&E systems in development evaluations?

Evaluators: Black-box, accountability-driven M&E remains pervasive in development evaluation. In those kinds of studies, evaluators set up simplistic, rigidly held logical frameworks, or "log frames." They tend to adopt inappropriate baselines, ignore development trajectories, and report on a few short-term quantitative monitoring indicators with a rhetorical nod to learning. They often examine what is easy to measure rather than what is significant for development in the long term.

Sometimes they "parachute evaluation teams into the field" without ensuring systematic work. The current obsession with measuring impact means that development evaluators pay scant attention to interventions as open systems. They tend to ignore negative consequences that may outweigh positive results and do not work in a manner that enhances the long-term sustainability of ideas, institutional systems, and impacts.

We wanted to test the limits of our type of approach in practice and learn how some limitations could be overcome. We were well aware of many of the difficulties it would bring. But if evaluation cannot cope with complexity, we should say so and start every assignment by stating that as a caveat. Our profession should focus on innovations that can help us tackle this issue head-on. Those trying to deal in a pragmatic way with complexity are gaining ground but tend to be less influential. This might be because of the inevitably scary terminology but certainly because simple, yet impressive-sounding numbers and solutions resonate so much better with overwhelmed politicians, policy makers, and even technocrats.

There is also a need for more examples that work from realistic evaluation or developmental evaluation perspectives, and we wanted to contribute to this. In developing countries, capacities and institutions are generally much weaker and less productive than in developed countries. If such countries are to grow, development strategies need to be comprehensive and integrated, with well sequenced interventions, and positive results that have to be sustained over long periods. We have to grapple with how to evaluate under these circumstances. It is imperative that we understand the interplay between interventions and their changing, often highly unpredictable environments, and deal with issues such as "transformation," "sustainability," and "resilience."

Reflections

M&E for Development Programs

AWARD's M&E evaluation system is meant to be empowering for the implementers, participants, decision makers, and funders. It reflects the program's notion that empowerment for leadership in science requires comprehensive strengthening of power in four aspects ("power from within," "power to do," "power over," and "power with"), plus seeking transformative change at the organizational and sector levels. That is not to say that it is an empowerment evaluation or participatory evaluation per se—it is empowering in the sense of cultivating ownership and understanding and being mindful and respectful of local contexts, as well as generating knowledge for similar programs and for development more generally.

To achieve this ownership and generate credible, useful knowledge, AWARD's M&E system had to develop practices addressing issues that others in the development evaluation community are likely to encounter if implementing a similar M&E system. These practices include (a) building the system on a set of explicit values for driving and justifying M&E designs and practices; (b) ensuring that the utility and value of evaluation is experienced by multiple African stakeholders, so as to counter the notion that evaluation is an externally imposed practice of little local utility; (c) building in a direct link between planning, M&E, and the adjustment of strategy and operations; (d) presenting and conducting monitoring as something beyond mechanically reporting on a set of high level indicators; (e) ensuring rigor, including thorough systematic triangulation, whenever possible; and (f) balancing the measurement of impact with ensuring and assessing the capacities, knowledge, and systems to manage for impact. Furthermore, with the program's emphasis on adaptive management, the M&E system is designed to go beyond process evaluation and to focus in an ongoing manner on collecting and using information addressing the intended program impacts, with implications for bringing about change across the agriculture sector and the region.

AWARD's M&E system is intended to help bring systems and complexity thinking into evaluation in practical, even if still limited, ways. The system is designed to help evaluators work with

predictable and unpredictable aspects of programs and deal with interventions as open systems. It is necessary to connect outcomes to multiple influences and to examine the extent to which the program components strengthen one another's impact, thus justifying the presence of each component.

These are not trivial matters. Development practices very frequently do not yield positive results within what are usually artificially established timeframes. Evaluators are increasingly being challenged to ensure that the profession can deliver on what is required and, in turn, significantly improve poor aid and development practices. The AWARD M&E system does not provide an ideal solution; its practices still need improvement and will be most relevant for selected "small *n*" development interventions. All of this draws attention to the need for much more systematic and documented work on innovations as open, complex adaptive systems. Development efforts will increasingly need such evaluation solutions.

AWARD's M&E Approach

The features of the AWARD evaluation system, with its use of realistic evaluation principles and methods such as contribution analysis, stand in stark contrast to the increasing advocacy of experiments and quasi-experiments to study development programs in Africa and elsewhere. Experiments are powerful tools, but their methodological requirements (particularly the aspects necessary to ensure internal validity) make their use limited to studying development programs with well-defined and stable interventions, circumscribed program contexts, and a limited breadth of outcomes that can meet psychometric requirements. They do not address issues such as program sustainability and resilience over the long term. Furthermore, program personnel are more likely to provide credible data if they believe in the eventual utility of the evaluation findings. The AWARD M&E experience shows the value of developing evaluation designs that are sensitive to the locally relevant values, cultures, and contexts.

The extensive use of a detailed nonlinear theory of change, allowing for deductive and inductive analyses, is an exemplary aspect of the M&E system. AWARD built its theory of change through close, iterative interaction with stakeholders and drew from existing knowledge and frameworks to strengthen it. The theory of change has been critical in shaping the M&E system and in thinking through findings. It has been extremely useful for adaptive management and has been at the core of AWARD's managing for impact approach. It has also satisfied the requirements of funding agencies and government funders.

The AWARD M&E is built on the notion that it is short-sighted to waste resources on development interventions without cultivating a nuanced understanding *and management* of the interventions and their dynamic cultural and organizational contexts. The theory of change helps focus attention on collecting evaluation data addressing causes, contexts, and mediators. The program steering committee and the AWARD management team can use these data to make modifications to the program and to understand more fully how to bring about the program's intended effects. The theory of change also is intended as a primary tool for producing knowledge about what contributes to AWARD outcomes and impacts. The evaluation system thus values both instrumental and conceptual uses of evaluation findings. It is unclear at this relatively early point that the approach will provide knowledge applicable to a wider scale program and the elements are necessary to facilitate this, but the M&E team will be working to clarify this during Phase II.

The M&E team admits that the long-term success of addressing complexity remains to be shown, but to date, it appears to have been successful. The evaluators are confident that the evaluation has shown that the program components have contributed in an integrated manner to empowerment at the individual level and that implementation has occurred largely as intended, including with the desired quality. They intend to focus the M&E efforts during the second phase on tracking institutional and sector change and on refining their understanding of issues such as better measuring some

of the more intangible outcomes: transformative change, gender-responsiveness, innovation, and the sustainability of AWARD's model and ideals.

Successes and Challenges

The AWARD evaluation is well resourced (funded at about 11% of the program's budget) and strongly supported by the steering committee and the funders, all of which make for conditions that allow the M&E system to have a strong and effective role in development. As a novel and ambitious endeavor, however, it has encountered obstacles along the way, as any seasoned evaluator would predict is likely to occur. The third and fourth authors of this Exemplars case have expanded fully on these obstacles, resulting in a full account that we intend to be of use to other development evaluators adopting the approach.

The AWARD evaluation shows that M&E systems need significant evaluation expertise embedded in an organization or program. The expertise should not come solely from external sources. The M&E team needs to understand a spectrum of possible designs and act in internal evaluator roles, supporting their work through occasional special studies and external or independent evaluations. AWARD reinforces the argument that evaluation is not a simple task that ill-prepared professionals can do.

In setting up the system, the M&E team found out who wanted to know what among the primary stakeholders, as well as for what purpose, so that priorities could be established using a systematic process drawing from the theory of change. The AWARD leadership was prepared to reflect on evidence of both successes and mistakes, couple it to intuition and their own observations, and adjust quickly and effectively—that is, to use adaptive management well. The M&E team has found that the program steering committee and management team have demonstrated the value of adaptive management and that a culture of openness about performance is imperative for success.

The M&E system is cultivating an evaluative culture among primary stakeholders by demonstrating the value of M&E as quickly as possible and has done as much as possible, within reason and resource availability, to get priority information as rigorously as circumstances allow. It is building M&E capacities among primary stakeholders, and most important for delivery, developing efficient data collection and analysis systems to alleviate the burden of work. The M&E team admittedly failed initially in every one of these efforts, but purposefully learned through each and adjusted wherever possible.

Many of the challenges encountered by the AWARD M&E are common in complex M&E systems serving multiple stakeholder groups. The M&E team had to deal with barriers in developing and maintaining a wide-ranging M&E system that, to date, had limited success in serving all the identified stakeholder groups. The M&E team also found challenges in supporting the management team's efforts to engage deeply all stakeholder groups in the use of evaluation findings, and it has grappled with the value added by using consultants versus the dependency that occurs in having to rely on external expertise.

The Future of AWARD M&E

There may be an implicit presupposition in the work reported here that at some point the AWARD program will be reasonably stable and less dynamic and the M&E system will be static, fully integrated, and sufficiently resourced. The remaining evaluation tasks then would be the continued implementation of the M&E work. There seems a real likelihood, however, that the program may continue to be modified and evolve and that the M&E system will also need to continue to be modified. The M&E system may have to be as dynamic as the program itself, and the evaluation may need to adopt an emergent, adaptive evaluation strategy—perhaps an M&E *process* rather than

an M&E *system*. Such a process may itself need to be guided by its own theory of change—a theory of evaluation change that operates alongside the theory of program change.

Perhaps most importantly, the AWARD team seeks to know what works, why, and under what conditions AWARD achieves its intended impacts—a tall order, and one that requires the use and refinement of approaches such as realistic evaluation. The program theory of change is complex, with interrelated variables, some of which are difficult to operationalize, and some of the definitions of progress markers and the evaluation instruments need revision. These challenges are not unique to this evaluation or program setting, however—they are familiar to any evaluator seeking to collect substantial amounts of data and provide results in a timely manner. It will be of considerable interest to the evaluation community to learn the extent to which and how the AWARD M&E system successfully addresses these challenges in the years to come.

Declaration of Conflicting Interests

Funding

References

African Women in Agricultural Research and Development. (2009). *Sowing success: Results report 2008/09*. Retrieved from http://www.awardfellowships.org/about-us/results-report-200809.html

African Women in Agricultural Research and Development. (2010). *The theory of change of AWARD*. Nairobi, Kenya: Author.

African Women in Agricultural Research and Development. (2012). *Summary of AWARD's M&E system*. Nairobi, Kenya: Author.

African Women in Agricultural Research and Development. (2013). *AWARD in the news*. Retrieved from http://awardfellowships.org/media/award-in-the-news.html

Beintema, N. M., & Di Marcantonio, F. (2010). *Female participation in African agricultural research and higher education: New insights* (International Food Policy Research Institute Discussion Paper 00957). Washington, DC: International Food Policy Research Institute/African Women in Agricultural Research and Development.

Center for Creative Leadership. (2007). *The handbook of leadership development evaluation*. San Francisco, CA: Jossey-Bass.

Food and Agriculture Organization of the United Nations. (2011). *The state of food and agriculture women in agriculture: Closing the gender gap for development*. Retrieved form http://www.fao.org/docrep/013/i2050e/i2050e00.htm

Forum for Agricultural Research in Africa. (2006). *Framework for African agricultural productivity/Cadre pour la productivité agricole en Afrique*. Accra, Ghana: Author.

Guijt, I., & Woodhill, J. (2002). *Managing for impact in rural development: A guide for project M & E*. Rome, Italy: International Fund for Agricultural Development.

Hummelbrunner, R. (2010). Beyond logframe: Critique, variations and alternatives. In N. Fujita (Ed.), *Beyond logframe; Using systems concepts in evaluation* (pp. 1–34). Tokyo, Japan: Foundation for Advanced Studies on International Development.

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In R. Schwartz, K. Forss, & M. Marra (Eds.), *Evaluating the complex* (pp. 53–96). New Brunswick, NJ: Transaction.

Pawson, R., & Tilley, N., (1997). *Realistic evaluation*. London, England: Sage.

Rowlands, J. (1997). *Questioning empowerment: Working with women in Honduras*. Oxford, England: Oxfam.

Solava, I., & Alkire, S. (2007). *Agency and empowerment: A proposal for internationally comparable indicators*. OPHI Working Paper. Oxford, England: Oxford Poverty and Human Development Initiative.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations* (Department for International Development Working Paper 38). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

The World Bank. (2009). *Gender in agriculture sourcebook*. Washington, DC: Author.

VeneKlasen, L., & Miller, V. (2007). *A new weave of power, people & politics: The action guide for advocacy and citizen participation*. Sterling, VA: Stylus.

White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework* (Working Paper 15). New Delhi, India: Global Development Network, International Initiative for Impact Evaluation. Retrieved from http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf