

Chapter 9 Contents

9. Performance Management	9-1
9.1 Introduction	9-1
9.2 Monitoring and Evaluation General Findings	9-2
9.2.1 Strengths	9-2
9.2.2 Weaknesses	9-3
9.3 Monitoring	9-4
9.3.1 USAID Monitoring	9-4
9.3.2 An Alternative Proposal to Strengthen USAID Monitoring	9-5
9.3.3 Monitoring Innovations by Awardees	9-5
9.4 Evaluation	9-5
9.4.1 Evaluation Design Type	9-7
9.4.2 Program Reviews by USAID	9-7
9.4.3 Findings on Evaluation	9-8
9.4.3.1 Strengths	9-8
9.4.3.2 Weaknesses	9-8
9.5 Reporting	9-11
9.6 Conclusions and Recommendations	9-15
9.6.1 Conclusions	9-15
9.6.2 Recommendations	9-16
Bibliography for Chapter 9	9-22

9. Performance Management

9.1 Introduction

The term *performance management*, as used in this report, encompasses program monitoring, evaluation, communicating results (reporting), and learning from results. See Box 9.1 for USAID/FFP’s definitions of monitoring and evaluation, which are derived from USAID’s ADS 203.3.2. Performance management is arguably where USAID/FFP, Title II Awardees, and FANTA jointly made the greatest progress during the FAFSA-2 time frame. It is thanks to these advances that there was sufficient program performance information available to review in the FAFSA-2.

During the FAFSA-2 time period, USAID/FFP required Awardees to include a results framework in their proposals to depict “the food aid program’s theory of change by laying out the activities and outputs that will lead to short, medium and long-term outcomes and objectives” (FFPIB 09-06, 2009a). USAID/FFP also required an IPTT with project-relevant indicators at the impact, outcome, and output levels. The Awardees established baseline values and targets for these indicators during project start-up and reported progress toward targets annually. It was mandatory to conduct population-based, representative baseline and final evaluation (endline) surveys to measure change in the project’s impact and outcome indicators and to have an endline evaluation done by external evaluators. In addition, USAID/FFP required each project to have a mid-term evaluation. Most of the final evaluations reviewed for the FAFSA-2 included the findings of an external team’s qualitative evaluation substantiated with quantitative data from the surveys. FANTA prepared several Technical Notes on M&E in Title II development programs (Bonnard, 2002; Bergeron et al., 2006a; and Bergeron et al., 2006b) to assist Awardees with the requirements.

Standard indicators. A consensus building process and applied research over several years culminated

Box 9.1. Monitoring and Evaluation

- “Monitoring and evaluation perform two separate but related functions.” (GAO, 2009)
- “**Monitoring** reveals whether desired results are occurring and whether assistance objectives’ outcomes are on track by addressing the ‘what’ of performance and using pre-selected indicators to measure progress toward planned results at every level of the Results Framework.” (FFPIB 09-06, USAID/FFP, 2009a)
- “**Evaluation** answers the ‘why,’ ‘why not’ and the ‘what else’ of performance; it is used on a periodic basis to identify the reasons for success or lack of it, to assess effects and impacts, or to indicate which, among a range of program or project and activity alternatives, is the most efficient and effective. For Title II programs, evaluation is also used to assess the extent to which a program is meeting outcome and impact level objectives.” (FFPIB 09-06, USAID/FFP, 2009a)

in the 2007 release by USAID/FFP of required standard impact and monitoring indicators for development food aid programs (FFPIB 07-02, 2007). Use of standard indicators greatly increased the comparability of results across programs, especially at the impact level. Most notable was the introduction of new indicators to measure household access to food. More recent advances are the

Household Hunger Scale, the *Minimum Acceptable Diet*, and the *Women's Dietary Diversity Score* indicators.²⁷⁰ The standard indicators were updated in 2011 to improve their quality and usefulness, and to demonstrate coordination with the FTF Strategic Results Framework and GHI goals and targets (FFPIB 11-03, USAID/FFP, 2011b). Two 2009 USAID/FFP Information Bulletins describe reporting requirements and M&E responsibilities of Awardees (FFPIB 09-06, 2009a and FFPIB 09-07, 2009b). Prior to the release by USAID/FFP of Information Bulletins from FY 2007 forward, M&E and reporting requirements for development food aid were less formal.

This chapter's findings are based primarily on the strengths and weaknesses in performance management encountered during the FAFSA-2 while reviewing individual project's progress, monitoring, and evaluation reports, data, and information. This is not an in-depth review of performance management by USAID/FFP or Awardees. A useful reference that enriched FAFSA-2 insights was the GAO's 2009 report on its performance audit of M&E in U.S. international food assistance programs and that report's appendix, the December 31, 2008 USAID Report to Congress on "efforts undertaken by the Administrator to conduct oversight of non-emergency food aid programs."²⁷¹

9.2 Monitoring and Evaluation General Findings

The following are some strengths and weaknesses of Title II development programs' overall M&E efforts found during the FAFSA-2 review.

²⁷⁰ The USAID/FFP Strategic Plan for 2006–2010 states that "only 25% of current Title II development programs with access activities include indicators of household food access in their monitoring and evaluation systems" (2005, p. 36)

²⁷¹ This Report to Congress was required under the 2008 FFPA. It is available at <http://www.gao.gov/products/GAO-09-980>.

9.2.1 Strengths

- Harmonizing indicators across similar programs in the same country is very useful for comparing programs and summarizing overall results. Programs in Bolivia, Guatemala, Haiti, Honduras, and Indonesia had harmonized core indicators. While harmonized indicators are advantageous, if they are well chosen and defined, limitations to their usefulness for evaluation are multiplied if harmonized indicators are not selected well, e.g., missing standard indicators, not consistent with the theory of change, or knowledge instead of practice indicators at the highest behavior change outcome level.
- USAID/FFP M&E training workshops conducted by FANTA started in FY 2009 to strengthen Title II development programs' results frameworks, IPTTs (including "F" and USAID/FFP standard indicators), and baseline survey plans drafted by Awardees before or during the workshops.²⁷² Awardees of all new development food aid programs send M&E staff, technical sector leads, and program managers to the workshops.²⁷³ Program chiefs of party and USAID Mission staff also attend. According to FAFSA-2 in-country interviews, the workshops conducted for program start-ups were well received. However, the feedback also indicated that the workshops were too short for the amount of material covered.
- FANTA has conducted periodic training of USAID/FFP staff on various M&E topics.²⁷⁴

²⁷² Workshop duration is five days for Awardees in the field. In July 2011, FANTA-2 organized its first two-day M&E workshop in Washington, DC, for Awardee headquarters staff.

²⁷³ Through December 2011, M&E training workshops had been conducted for programs in Afghanistan, Bangladesh, Burkina Faso, Burundi, Chad, DRC, Ethiopia, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Sierra Leone, South Sudan, and Uganda. Only the programs in Bangladesh, Malawi, Niger, and Uganda were in the FAFSA-2 universe and visited by the FAFSA-2 team.

²⁷⁴ Examples of training topics include: (1) "USAID/FFP Standard Indicators," (2) "How to Review a Results Framework," and (3) "How to Review an IPTT."

- Technical reviews and feedback by FANTA-2 M&E staff have helped improve the quality of Title II program results frameworks, IPTTs, M&E plans, and evaluation survey designs, including sampling and questionnaires.
- In FY 2010, FANTA-2 launched Discussion for Title II M&E (Discussion-TIIME), a listserv where Title II M&E practitioners can learn from each other and access technical advice. The goals are to promote professional development of M&E staff, introduce new staff to Title II M&E, provide updates, and improve M&E quality.
- USAID/FFP hired and placed two new M&E specialists in the West and Southern Africa regional offices in 2011. There now are M&E experts in all three regional Africa offices. Furthermore, all USAID/FFP focus countries now have FFP Officers based in-country, due to increased hiring since 2009.
- The TOPS project FSN Network M&E Task Force has defined core competencies for M&E managers and officers and is working with Awardees to identify gaps and strengthen capacity.
- Title II Awardees have implemented several useful M&E innovations that have strengthened program M&E and that can be adopted by other Awardees. SC/Bangladesh (FY 2005–FY 2010) used geographic information system (GIS) software to map levels of undernutrition; ensure sampling was done correctly for their final evaluation survey; and locate cyclone shelter centers, village water points, and model farms. An M&E system design called “Simple Measurement of Indicators for Learning and Evidence-Based Reporting” (SMILER), developed by CRS, compiles all M&E tools in one system and in one operating manual. This is in use in the CRS/Malawi FY 2009–FY 2014 program. The Malawi program, in collaboration with FANTA-2, also tested cell phone text messaging and Frontline software to collect data on the *Household Hunger Scale* indicator. It was beyond the scope of the FAFSA-2 to assess the quality of these reported innovations.

9.2.2 Weaknesses

- There are not enough M&E experts in USAID/FFP Washington, in part because the 2008 FFPA does not give USAID/FFP authority to use Title II funds to hire personnel to work in headquarters on non-emergency programs.
- Considerable heterogeneity of indicators, especially in the earlier years of the FAFSA-2 time frame, often precluded being able to assess overall performance across programs or to do meta-analyses.
- In a number of programs reviewed, IPTT indicators were inappropriate for the content of the proposals, the results frameworks, or the state of the art for M&E of a particular intervention. The M&E training workshops, along with the reviews of IPTTs by the AOR and FANTA, should lead to better indicators. However, the FAFSA-2 found that the problems mentioned persist in more recent programs, suggesting a need for more efforts to strengthen IPTTs and more thorough reviews by technical sector experts. Examples of less useful indicators at the highest outcome level include: “knowledge” indicators to evaluate behavior change interventions, when the indicator should measure “practice”; not including **all** the relevant USAID/FFP standard indicators; and not focusing nutrition service delivery indicators on pregnant and lactating women and children under two years in programs targeting these beneficiaries.
- Because they rarely include sample size or CIs, the data in IPTTs are hard to interpret and use.
- Anthropometric data reported were rarely sex-disaggregated.
- Some programs with USAID-funded multi-year extensions did not measure and report results data for the extension period.
- The amount of funds budgeted for M&E by Awardees may be too low. According to several stakeholders interviewed during the FAFSA-2, a lack of funds impairs the quality of Title II M&E.

9.3 Monitoring

9.3.1 USAID Monitoring

Less attention appears to have been paid to strengthening monitoring of Title II development programs than to evaluating them. Monitoring of programs is critical. The FAFSA-2 found that in programs with proactive oversight by USAID, this monitoring was associated with better program results. In contrast, there were other examples of weak technical monitoring by USAID and not holding Awardees to commitments made in their program descriptions and to achieving performance indicator targets. USAID has increased efforts to strengthen its monitoring of Title II programs through training workshops in Africa, called “boot camp” for FFP Officers and staff, and through the development of the CBO Handbook, also known as the Program and Policy Manual. The manual includes information on what to look for during monitoring site visits. At the time of the FAFSA-2, USAID/FFP was preparing a new M&E manual for its staff. Methodologies to systematize USAID field monitoring, with guidance on sampling, indicators, checklists, and analysis, include: (1) the Layers tool developed by FANTA, which is based on Lot Quality Assurance Sampling (LQAS)²⁷⁵; and (2) monitoring plans used by USAID/Ethiopia.

In response to audit findings in Haiti and Madagascar, FANTA developed and introduced Layers to strengthen USAID monitoring of Title II programs in Haiti, Ethiopia, and Madagascar from 2002 to 2004. Field monitors collect, enter, and analyze program information from a small random sample of program sites, using personal data assistant devices (PDAs), to assess the quality of the implementing partner’s operations. From the data collected on inputs, outputs, and processes, Layers generates a report on program performance that USAID sends to the Title II Awardee to use to improve the program. Since 2009, Layers has also been used in Chad, Guatemala, Mali, and Uganda, with TA from FANTA-2 and its subcontractor, TANGO. In these countries, data collection was

outsourced to local research firms contracted by FANTA-2, rather than done directly by USAID field monitors. This is in contrast to Layers implementation in Ethiopia, Haiti, and Madagascar, where USAID field monitors collected the data.²⁷⁶

There is extensive experience with Layers from the initial three countries. The Ethiopia and Haiti Missions continue to use Layers, with their own field monitors collecting the data, but Madagascar has discontinued Layers. It is too early to say if Layers will be continued in the other countries, most of which have just completed one round.

A number of questions raised about Layers in the FAFSA-2 interviews during field visits and in feedback from USAID Missions and embassies (in non-presence countries) to a 2010 survey by USAID/FFP and FANTA-2 can be summarized as follows:

- Is the information generated by Layers worth the time, complexity, and cost, especially when it provides little insight on **why** problems are occurring?
- If Layers is useful but Missions are unwilling to pay for it, how can Layers be financed?
- Do the findings lead to program improvements, especially when improvements require more funds?
- Is outsourcing USAID’s monitoring responsibilities using Layers a good idea?
- Should Layers be used mainly for monitoring commodity management, a function for which it has been especially helpful, versus broader application to technical activities?
- What should be monitored by USAID and what by Awardees?
- Can other limitations of Layers be resolved, for example, not turnkey, not as fast or as efficient as hoped, lack of access to the software in Missions due to USAID restrictions on loading it on its network, too many players, and statistical concerns?

²⁷⁵ See <http://www.fantaproject.org/layers/reference.shtml>.

²⁷⁶ Madagascar outsourced data collection for their last round of Layers.

There are plans by USAID/FFP to reassess the value and role of Layers and to address the issues raised by the field. A key question is, “Does it make sense to do Layers routinely in all Title II programs in all 20 USAID/FFP focus countries”? If not, how can USAID field monitoring be systematized and strengthened and audit concerns addressed?

9.3.2 An Alternative Proposal to Strengthen USAID Monitoring

The USAID/FFP East Africa Regional Office designed a new approach called a “monitoring plan,” which has been implemented by USAID/Ethiopia to monitor its entire large FFP portfolio, including emergency (with WFP) and development programs.²⁷⁷ What is new in this hybrid approach, which strives to instill a greater degree of rigor, is: (1) using Layers only for commodity management monitoring; (2) employing purposive, versus random, sampling of program implementation sites based on issues identified in routine reporting and while observing food distribution, and collecting other information from stakeholder consultations while at these sites; (3) using iterative qualitative inquiry as an essential part of monitoring to determine why program implementation is off track; and (4) holding quarterly meta-analysis meetings of the USAID monitoring team to: (a) identify cross-cutting concerns and trends, (b) formulate policy and operational recommendations that require follow-up, and (c) share lessons learned in monitoring.

The next steps to expand the use of monitoring plans by other USAID Missions are to write normative guidance and for the three Africa-based regional USAID/FFP M&E specialists to work together to design and launch monitoring plans in other focus countries in Africa. USAID/Ethiopia is exceptionally well staffed, with four full-time food aid field monitors. Therefore, a challenge to expanding monitoring plans will be their feasibility in countries with fewer (or no) dedicated full-time FFP monitoring staff. Furthermore, an important question

²⁷⁷ Gregory Collins designed the monitoring plan while assigned to the USAID/FFP East Africa Regional Office in Nairobi and shared it with the FAFSA-2 team on November 28, 2011.

is whether monitoring plans will be as useful as Layers in responding to audit concerns about how well USAID monitors Title II programs, given that Layers is standardized, whereas the qualitative inquiry component of monitoring plans is dependent on the reviewer’s judgment.

9.3.3 Monitoring Innovations by Awardees

There was little information in the Awardees’ reports on their approaches to program monitoring. Several Awardees mentioned using LQAS and PDAs for monitoring. Some said they would benefit from advice on how to determine when LQAS samples are too small for measuring certain common outcome indicators during annual or mid-term monitoring surveys. The managers of the SC/Bangladesh program (FY 2005–FY 2010) wrote a brief case study on integrating program and commodity management using PDAs.²⁷⁸ As part of this integration, SC introduced a health information system called McAid that tracks individual children’s health and nutritional status and use of MCHN services. Based on the information generated by McAid, SC reported the capacity to alert health workers to follow up on malnourished children and to look for positive and negative deviants. Using Pendragon software and PDAs, SC/Bolivia (FY 2002–FY 2009) collected monthly growth monitoring data. Advantages reported included improved data quality, reduced duplication, less time spent transcribing data from written to electronic records, and savings in the M&E budget. It is hoped that these and other innovations can be disseminated to the Title II community through the TOPS project FSN Network M&E Task Force.

9.4 Evaluation

Committed to managing for results, USAID/FFP has required all Title II development programs over the past decade to conduct independent, quantitative evaluations. This focus on evaluation is remarkable given that neither USAID/FFP nor USAID had a dedicated M&E unit during this time period.

²⁷⁸ Attachment H of SC/Bangladesh FY 2009 ARR.

The performance information generated by these evaluations was invaluable to the FAFSA-2 team.

In 2009, USAID reestablished its central evaluation leadership role, staffing a new unit with evaluation experts in the Bureau for Policy, Planning, and Learning, Office of Learning, Evaluation and Research (PPL/LER). In January 2011, USAID released an Evaluation Policy that updates evaluation standards and practices and renews its commitment to learning and accountability through evaluation (see Box 9.2). Evaluations are required for large projects and pilot or innovative development interventions. USAID/FFP is updating its guidance on Title II development program evaluation in accordance with USAID’s new policy. Other USAID programs could learn a great deal from the extensive evaluation experience of USAID/FFP, its Awardees, and TA partners, such as FANTA.

During the FAFSA-2 time period, USAID/FFP required, and continues to require, that Awardees conduct: (1) a population-based household baseline survey in the first year of the program, (2) a mid-term evaluation halfway through the program, and (3) a population-based household endline survey. A qualitative final evaluation, substantiated with quantitative data from the baseline and endline surveys, is encouraged and often done, but not required. Only the endline survey/final evaluation must be done by external consultants to ensure independence. The mid-term evaluation does not require collecting quantitative data, and USAID encourages participatory qualitative assessments (FFPIB 09-06, USAID/FFP, 2009a). Satisfactory review of the baseline and endline survey plans by the USAID/FFP AOR²⁷⁹ is required before data collection begins, and FANTA’s M&E experts also provide technical reviews of these plans, if requested.

²⁷⁹ Usually the CBO.

Box 9.2. USAID’s Evaluation Policy

“The evaluation policy builds on the Agency’s long and innovative history of evaluation, while seeking to redress a decline in the quantity and quality of evaluation practice within the Agency in the recent past. As part of a series of recent reforms known as USAID Forward, the Agency is transforming into a learning organization and a modern development enterprise. The policy is an initial step to strengthen USAID’s evaluation practice as part of the broader reform efforts.”

“USAID Evaluation Practices

- ***Integrate evaluation into design.*** Include evaluation specialists in strategy and project design teams, identify questions, plan for baseline data collection;
- ***Minimize bias.*** Disclosure of conflicts of interest, external evaluation experts as team leads;
- ***Ensure relevance to future decisions.*** Evaluation questions developed with stakeholders and are linked to future decisions;
- ***Use the best methods.*** Qualitative and quantitative methods that generate reproducible and high quality evidence;
- ***Reinforce local capacity.*** Work with local expert evaluation leads, use host country systems and build local capacity;
- ***Be transparent.*** Findings from evaluations are shared publicly and in a timely manner; and
- ***Dedicate sufficient resources.*** Goal of approximately 3 percent of a USAID operating unit’s total program funds to be set aside for external evaluations.”

Source: PPL/LER Briefer, USAID’s Evaluation Policy, 2011.

9.4.1 Evaluation Design Type

Nearly all quantitative evaluation surveys reviewed for the FAFSA-2 followed an adequacy design of simple pre/post comparisons (i.e., changes in outcome and impact indicators between baseline and endline surveys), without comparison groups. The assumption underlying this approach is that effects seen are associated with program activities. This assumption is to be substantiated by sufficient information on program inputs, processes, and outputs. This evaluation design, the least technically challenging and expensive, is a good fit for Awardees with limited technical capacity in survey research, whose primary responsibility is service delivery and not studies. It recognizes that finding a geographically proximate comparison group similar to the program beneficiaries in all or most socioeconomic and other characteristics except program participation is very difficult. However, without a comparison group, it is hard to rule out the possibility that changes may be due to factors unrelated to project activities. Because they lack a credible counterfactual, or comparison group, to control for factors other than the program interventions that might account for the observed changes, current Title II development program evaluation surveys do not meet the USAID Evaluation Policy definition of “impact evaluations.”²⁸⁰ They are instead considered “performance evaluations” for which the policy states that a mix of quantitative and qualitative methods (as used in Title II evaluations) is optimal.

Mid-term evaluations reviewed for the FAFSA-2 were usually qualitative. Since these are internally managed and conducted by Awardees, improving the methodology and maximizing the utility of Title II mid-term evaluations has not been a central focus for USAID/FFP or FANTA. Yet mid-term evaluations are critical for examining whether project implementation is on track and the desired outcomes are being achieved, while there is still time for mid-course corrections. A first step toward defining

the purpose and guidelines for Title II mid-term evaluations was an exchange of information and ideas between Awardees and FANTA-2 about these topics on the Discussion TIIME listserv in 2010.

There were several good examples in the FAFSA-2 of projects being redesigned in response to findings of mid-term evaluations, and ending well with positive impact on nutritional status as a result of fine-tuning interventions, namely, the Bolivia programs, as well as the SC/Bangladesh (FY 2005–FY 2010) and CRS/Haiti (FY 2002–FY 2008) programs. The joint mid-term evaluation of the four Bolivia programs was managed by the USAID Mission with Mission FFP staff participating extensively in the field visits, as they did in the joint final evaluation. The CRS/Haiti program followed a useful two-phase methodology in its mid-term evaluation in January 2006 with assistance from FANTA. In the first phase, quantitative data were collected using LQAS to assess which inputs, outputs, and outcomes were off track compared to targets. The second phase was a participatory qualitative inquiry in the field to probe into constraints or explanations for why certain indicators flagged in Phase 1 were lagging. That information was used to implement corrective actions. In contrast to these promising practices, too few other programs reported using mid-term evaluation findings to adjust project activities and improve implementation, when expected outcomes were not being achieved. Least useful were mid-term evaluations that were just opinion polls of whether project participants were pleased with the project, which were invariably positive, but lacked insight into whether the project was on track to achieve expected results and why or why not.

9.4.2 Program Reviews by USAID

To compensate for the limitations of mid-term evaluations led by Awardees, including the potential for bias in self-evaluation, USAID/FFP is planning to conduct its own “program reviews” to respond to implementation and design problems detected through routine monitoring or other sources, following a methodology designed in the USAID/

²⁸⁰ See http://transition.usaid.gov/policy/evalweb/USAID_Evaluation_Policy_FAQ.pdf.

FFP East Africa Regional Office and implemented in Ethiopia.²⁸¹ Based on FAFSA-2 findings, program reviews would be particularly useful to address design and implementation problems observed in nutrition interventions during field visits. In these cases, it is urgent for USAID/FFP to send teams of technical experts to review program progress and the appropriateness of designs, and to use these findings to work with Awardees on improving performance and redesigning programs (see Section 6.6.2 and Recommendation 37). Therefore, the FAFSA-2 team is in favor of USAID/FFP's plan to do USAID-led program reviews between the baseline and final evaluation surveys, to complement mid-term evaluations done by Awardees. Having team members with technical expertise in the interventions to be reviewed will be essential to the success of designing and conducting effective program reviews.

9.4.3 Findings on Evaluation

The following are examples of some evaluation strengths and weaknesses found by the FAFSA-2 team in Awardees' progress and evaluation reports and outcome and impact data used to assess the overall performance of Title II development programs. These findings were also informed by interviews with USAID, Title II Awardee, and FANTA staff.

9.4.3.1 Strengths

- Based on the example of the *joint* baseline and final evaluation surveys of MCHN programs in Haiti, the FAFSA-2 team thinks that, in a country with multiple Awardees and projects on the same timeline, doing joint quantitative evaluation surveys is a promising practice. The Haiti MCHN final evaluation survey report included overall findings, as well as separate tables with data for each Awardee; both individual and joint findings are critical. Joint surveys standardize evaluation among and across programs, making comparative analysis and cross-program learning possible.

²⁸¹ Gregory Collins designed the program review while in the USAID/FFP East Africa Regional Office in Nairobi and shared it with the FAFSA-2 team on November 28, 2011.

Undertaking a joint survey seems more efficient than each Awardee doing separate surveys, which hampers comparability and likely increases total costs. Programs in Ethiopia and Madagascar conducted joint *baseline* surveys, but due to external factors (e.g., changes in program design, political unrest), joint *final* surveys were done only for some of the Ethiopia programs. The FAFSA-2 team understands that a joint baseline survey will be done for the newest programs in Ethiopia.

- Conducting *joint qualitative final evaluations*, with one team assessing all Title II development programs, in countries with multiple Awardees on similar project timelines, is also a promising practice. Examples are the final evaluations of the Bolivia, Haiti, Honduras, Guatemala (FY 2000–FY 2008), Indonesia, and Nicaragua programs. The FAFSA-2 team learned a lot more about performance, what works, and why from these joint evaluations. Missions in these countries preferred joint evaluations so that they could review the Title II development portfolio as a whole. It is important that joint final evaluation reports include overall findings, as well as separate findings for each Awardee and that the process allows time for individually debriefing each Awardee. But Awardees may not favor the comparative nature of joint evaluations, because they are competitors hoping to win their next award. A drawback is the intense level of coordination involved to plan the field work for a single team, and this may divert Awardees from getting data and monitoring reports organized for the evaluation team.²⁸²

9.4.3.2 Weaknesses

- Until recently, when USAID/FFP issued a new directive, Title II development program Awardees were not required to submit the datasets for their baseline and endline evaluation surveys to

²⁸² This discussion of joint evaluation pros and cons benefited from the insights of the FAFSA-2 team leader, Roberta van Haeften, and Judiann McNulty, an independent HN consultant, both of whom have conducted several joint evaluations of Title II development programs.

USAID (FFPIB 11-02, 2011a). Final evaluation reports reviewed for the FAFSA-2 often did not include detailed information on the final survey, i.e., the methodology, sampling, and findings on indicator data compared to baseline values with sample sizes and CIs. Instead, submitted reports tended to be qualitative evaluations that cited some comparative findings from the baseline and endline surveys. In some cases, Awardees submitted additional detailed reports on the endline survey. The lack of information in reports to USAID is understandable because it was not until July 2009, at the end of the FAFSA-2 time period, that USAID/FFP issued FFPIBs 09-06 and 09-07 requiring “confidence intervals along with point estimates, a full description of the survey design type, and sampling methodologies” in baseline and final evaluation study reports from Awardees. But the guidance does not require reporting on actual sample size. Not having a pair of detailed baseline and endline survey reports with methodological information for all of the completed programs reviewed made it difficult for the FAFSA-2 team to ascertain the quality of the data and of the evaluation designs using criteria defined by USAID/FFP and FANTA (Swindale et al., 2004). The FAFSA-2 team had to rely mainly on what the final evaluation team said about the methodology and quality of the surveys (usually conducted by others) in the project’s (qualitative) final evaluation report, if anything, as well as the team’s judgment when obvious flaws were detected. Thus, the number of evaluation surveys found with methodological limitations is probably an underestimate.

- In response to USAID/FFP’s March 2011 special request to Awardees to submit baseline and endline survey datasets for 77 completed Title II development programs for further analysis as part of the FAFSA-2, pairs of baseline and endline datasets were received for only 19 programs. After eliminating surveys with limitations, datasets from only 10 programs could potentially be used, too small a number for meta-analysis.
- The FAFSA-2 team reviewed data quality and design of Title II MCHN evaluations to identify

programs with reliable data to include in the overall analysis of nutritional status impact during the FAFSA-2 time period. (See Section 6.4.1 and Table 6.15 for a discussion of this issue.) Serious limitations were found in the evaluations of 46 percent of the 54 programs that had ended and reported their baseline and endline data. The most common issues (discussed next) were poor-quality anthropometric data, sampling problems, and seasonality differences that made comparison of baseline and final survey data invalid.

- The poor quality of some of the evaluation surveys and of the mid-term and final qualitative evaluations suggests that neither the evaluators nor the Awardees had the necessary technical expertise. Lack of adequate technical expertise on evaluation teams could be due to the selection process by Awardees, but it could also be that there are simply not enough technically strong evaluators available with a basic knowledge of Title II development programs. Only a relatively few names are associated with the better evaluations, for example. Certain expertise also seemed to be missing from many of the evaluation teams, including civil engineers, environmental specialists, and economists/business management specialists.
- The problem of some programs sampling from too large a geographic area at baseline, because the villages that will participate in the project have not yet been selected, may contribute to underestimating the population-based effect of the project when that same large geographic area is sampled and surveyed at endline, including villages that never benefited from project activities.²⁸³
- One reason for conducting the final evaluation survey in a different season is racing to meet both the USAID/FFP deadline for submitting applications for new Title II development programs and the requirement that Title II

²⁸³ This can be controlled during analysis by removing data from the baseline for those villages that never participated in the program, and in the endline sampling only villages that were in the program from the same geographic area as the baseline.

Awardees submit the final evaluation report for their existing program “two months prior to the submission date of the new [Multi-Year Assistance Program] proposal” (FFPIB 09-07, USAID/FFP, 2009b). However, this renders the survey results unreliable.

- Poor-quality anthropometric data are common due to measurement and age estimation errors.
- While conducting their required review of Awardees’ evaluation survey plans, USAID AORs did not always take advantage of FANTA’s survey and evaluation experts, whom USAID/FFP has funded to review the plans for technical soundness. As a result, several bad evaluation designs were approved. The AORs did not have the technical expertise to identify limitations in sampling and evaluation design and, therefore, did not request the Awardee to redesign the survey plan to correct the problems.
- The requirement that external evaluators conduct Title II final evaluations to ensure an independent, unbiased review is compromised at times, because Awardees almost always manage the funding for evaluations, decide who the evaluators will be, and procure their own evaluations and evaluators. Even if an evaluator is independent, it can be difficult to report negative findings to the same Awardee that is paying the consultant’s fee. This approach of Awardee-managed evaluations of Title II programs does not comply with the USAID Evaluation Policy’s standard for minimizing bias, which defines an external evaluator as a third party contractor or grantee managed directly by USAID.
- Several stakeholders interviewed mentioned underfunding as a likely contributor to some of the poorer-quality evaluations of Title II programs. Awardees submit an attachment to the PREP that describes their expenditures on evaluations in the prior year. However, since this budget information is not submitted prior to doing the evaluation, the AORs cannot determine whether Awardees are spending an adequate amount on evaluation as part of their agreement oversight and review of baseline study and final evaluation plans.
- The GAO reported in its 2009 performance audit of M&E in international food assistance that Title II development program evaluations usually do not assess the effectiveness and efficiency of different approaches. The FAFSA-2 review also found that Title II final evaluations do not collect, analyze, and report enough information on service delivery, demand-side behavioral outcomes, and implementation processes. If expected results are not achieved, it is important to determine where in the causal chain the program broke down. The evaluations also do not report adequately on participation in project activities/receipt of project inputs, including food rations, and the duration. Project impacts are rarely disaggregated by socioeconomic status to test if targeting was appropriate. However, this type of in-depth, comparative analysis and disaggregation would require larger sample sizes and evaluation budgets.
- Title II evaluations also did not measure cost and cost-effectiveness of different interventions and approaches.
- In several instances where Awardees hired a research firm to do their evaluation surveys, and separately contracted a qualitative final evaluation team, the endline quantitative data on outcome and impact indicators were not yet available when the qualitative evaluation team did its work in-country. This led to qualitative evaluations that missed the opportunity for rigorous analysis, interpretation, and learning from survey findings.
- A number of final evaluations, especially in the earlier years of the FAFSA-2 time period, were mere opinion polls of whether project participants were pleased with the project, but lacked information on whether the project had achieved its expected results and why. As the USAID Evaluation Policy states, evaluation findings should not be based on anecdotes, hearsay, or the compilation of people’s opinions (USAID, 2011).
- Standard core survey modules were not developed centrally for the different technical sectors during the FAFSA-2 time period. These could have increased efficiency, quality, and comparability

of surveys. However, questionnaires are now available.²⁸⁴

9.5 Reporting

This section discusses FAFSA-2 findings on USAID/FFP and Awardee reporting on Title II development programs. Reports prepared annually by USAID and assessed for the FAFSA-2 include: the International Food Assistance Report (IFAR), written jointly with USDA, and USAID Mission reporting in FACTS, based on information received from Awardees. Reports on program performance prepared by Awardees and assessed in the FAFSA-2 include: ARR with Tracking Tables for Resources and Beneficiaries by technical sectors, IPTTs, SAPQs, and final evaluation reports. While recognizing how invaluable the performance information in these reports is, and congratulating USAID/FFP and its implementing partners for their excellent work in preparing the reports, the FAFSA-2 found that reporting needs to be strengthened to address the following gaps.

- The Foreign Assistance Framework, Standardized Program Structure and Definitions, and indicators developed by USAID and the U.S. Department of State were introduced in 2006, and are referred to in short as the “F” process (USAID/U.S. Department of State, 2010). USAID/FFP shared the results of FY 2010 Mission reporting on Title II programs in FACTS with the FAFSA-2 team. While USAID Missions reported the contribution of Title II development programs to some of the “F” standardized program elements and indicators based on information from Title II Awardees, they did not report on a number of others in which programs supported major activities. This underrepresented the work of Title II development programs in meeting key foreign assistance priorities. Examples

²⁸⁴ On December 20, 2011, USAID/FFP reissued FFPIB 11-03 on “Revision to Food for Peace Standard Indicators Collected in Baseline Surveys and Final Evaluations.” Drafted with TA from FANTA-2, it has guidance on gender-sensitive indicators, Performance Indicator Reference Sheets, and a Standard Indicators Handbook with questionnaires for data collection and tabulation instructions.

of underreporting are discussed further in Section 6.2.2 (MCHN) and Section 8.2.2 (HIV). Improved reporting on the work of Title II programs toward “F” program elements and indicators may also help address the GAO finding that USAID/FFP needs to better link its M&E to key USAID and USAID/FFP goals (2009).

- Until FY 2011, the Tracking Tables for Beneficiaries and Resources that USAID/FFP required Awardees to submit with ARRs continued to use eight technical sectors unique to USAID/FFP, and not the “F” program elements used by the rest of USAID to describe similar activities.²⁸⁵ This gap was partially closed with the new ARR, PREP, and AER guidance issued by USAID/FFP for FY 2011/2012 to better align Title II reporting with the foreign assistance standardized program structure and definitions. The new instructions require Awardees to classify resources and beneficiaries in Tracking Tables by 14 “F” program elements versus the 8 technical sectors used before.²⁸⁶
- There are several reasons for saying the “F” reporting gap was only “partially closed” by USAID/FFP’s revised guidance. The 14 program elements selected do not include any for reporting on public infrastructure constructed, repaired, or maintained using Title II resources. This could be remedied by reporting on standard Program Element 4.4.3 “Transport Services” for roads and by reporting on Program Element 5.2.2 “Mitigation” for protective infrastructure to enhance emergency preparedness and disaster management. However, the “F” reporting structure only captures other infrastructure supported by Title II, e.g., water management

²⁸⁵ The USAID/FFP technical sectors and management information system predated the “F” process and FACTS. The eight technical sectors are listed in Section 1.2.1.4.

²⁸⁶ The 14 “F” program elements now used in Title II reporting are: civic participation; HIV/AIDS; maternal and child health; family planning and reproductive health; water supply and sanitation; nutrition; basic education; social assistance; agricultural sector capacity; strengthen microenterprise productivity; natural resources and biodiversity; protection and solutions; assistance and recovery; and capacity building, preparedness, and planning.

infrastructure for irrigation, flood control, drought prevention, and watershed protection in Sub-Element 4.5.2.2 “Land and Water Management” under Program Element 4.5.2 “Agricultural Sector Capacity,” and FFW in Sub-Element 3.3.3.4 “Self-Help Programs” under Program Element 3.3.3 “Social Assistance.” But Title II Awardees will not report at the sub-element level in their annual Tracking Tables sent to USAID/FFP, and USAID Missions are unlikely to report on Title II activities in these sub-elements in FACTS. In addition to the infrastructure work of Title II going unreported, there are a number of “F” program elements and many sub-elements supported by Title II for which the program will continue to get no recognition because they are not reported by Awardees or USAID Missions.

- **The IFAR** is a critical annual report to Congress on highlights of USAID and USDA international food assistance. However, it does not report on the impressive quantitative results of Title II development programs. Thus, it does not effectively tell the story of the importance of Title II development programs in improving nutrition and food security among some of the world’s most vulnerable populations. Yet such impact and outcome data are available in the Awardees’ evaluations and IPTTs. The IFAR reports stand in sharp contrast to USAID’s Reports to Congress on its Global Health and Child Survival (GHCS) Program, which are full of data on dramatic progress in saving children’s lives, increasing coverage of essential health services, and changing health and nutrition behavior for the better (USAID, 2009 and 2010).
- An important look-up table in the IFAR is the appendix on USAID Title II development activities by Awardee and country, which reports “recipients” of direct food aid. However, beneficiaries of Title II-supported activities that do not receive food rations are not reported in that table or anywhere in the IFAR. This is another example of undercounting Title II, and the many lives it touches, that is especially striking in countries with small direct food distribution components. Nearly twice as many beneficiaries

were reported in the FY 2009 Tracking Tables as the number of food “recipients” in the FY 2009 IFAR.

- There may be errors in filling the Total Food Aid Commodity (MT) column of the Resources Tracking Tables due to leaving out from this figure the food that was monetized, and only reporting food for direct distribution, e.g., CRS/Malawi FY 2009. The form has a separate column for Monetization Budget (\$), but the Total Food Aid Commodity (MT) column is supposed to include both food tonnage for monetization and direct distribution.
- Most Tracking Tables lacked essential identifiers (i.e., country, name of Awardee, submission date, fiscal year, and agreement number).
- The number of HN beneficiaries reported by Awardees in the annual Tracking Tables is misleading in cases where household members receiving food rations in MCHN or HIV programs are counted as HN beneficiaries, because these family members do not receive any HN interventions. Including such household members inflates the number of beneficiaries listed as receiving HN programming, making it difficult to discern how many people in the mother-child or PLHIV target groups ultimately receive HN interventions. Household members that are only food recipients need to be counted somewhere, but it would be better to have a separate category for this.
- Existing reporting makes gauging the coverage and scale of individual Title II projects difficult. Projects with low coverage of the expected number of beneficiaries with service delivery and project activities, or that reach few people in the geographic area of influence, are unlikely to achieve positive population-based changes in outcome and impact indicators.
- **The SAPQ** is an Excel-based questionnaire designed by FANTA that Awardees complete annually to assist USAID/FFP in collecting standard data across countries and programs on indicators designed to measure Strategic

Plan implementation progress. Such reporting is essential to aggregating data across all Title II development programs and reporting on standard indicators' overall impact. However, the SAPQ process has not been useful for overall performance reporting to date. Awardees submitted SAPQs each year since FY 2008, with FANTA tabulating these data annually. However, the FAFSA-2 team found no examples of USAID/FFP using SAPQ data. The FAFSA-2 team could not use the tabulated SAPQ data for its review due to concerns about Awardee reporting errors and the aggregation of Awardee data for analysis, e.g., the weighted average prevalence of underweight from all baselines done in the reporting year compared to the weighted average prevalence of underweight in all endlines done in the same reporting year. A better way to measure nutritional impact across all programs would be to calculate the annual percentage point change in underweight and stunting achieved by each completed program between its baseline and final surveys and then to calculate the average annual percentage point change in underweight and stunting across all programs that ended that year.²⁸⁷ The FAFSA-2 team questions whether a separate SAPQ form is needed when data on changes in indicators between baselines and endlines could be obtained from the IPTTs Awardees submit with their ARR, especially if a standard IPTT template/online form were designed to facilitate data processing and analysis and to fulfill SAPQ-like functions. The main value of the SAPQ to the FAFSA-2 was the summary contact and program information in the first block (Awardee name, country and program location in-country, start and end dates, program name, and award number), as this information was not readily available elsewhere.

- There is no standard format or template for **the IPTT**, although USAID/FFP provides an illustrative example with its RFA. Allowing every Awardee to design its own IPTT has led to varying quality in the reporting and further

complicated the process of aggregating overall data. Common problems found in IPTTs in the FAFSA-2 were: (1) format, such as too small a font; (2) completeness, such as no identification of country, Awardee, date, or award number; no sample sizes; no CIs; no labeling of the month and year of the baseline (no year is specified in the USAID/FFP sample form either) and endline surveys or the fiscal or calendar years when data were collected (e.g., FY 1 taken literally from the USAID/FFP example instead of FY 2011); (3) mixing monitoring and survey data for the same indicator on the same line, leading to faulty comparisons; and (4) not distinguishing final year results from cumulative life-of-agreement results. Several of these problems come from copying the USAID/FFP sample IPTT, which needs improvement.

- **The IPTT** is not a required part of program final evaluation reports, yet it should be because it summarizes the impact and outcome of the project on key indicators. Including this document in the final evaluation report would ensure that evaluators review these results and that readers get a fuller picture of the project's achievements.
- **No final performance (or end of program) report** is required from the Awardee, as USAID/FFP considers the final evaluation report written by the external evaluators to serve as the final program report. Consequently, most Awardees do not submit final reports. Yet it is the Awardee, not the evaluator, who is best able to explain and document the logical structure of the project and what was accomplished, including: the interventions and implementation approaches used, the theory of change or development hypotheses tested, the program model employed, and the project's inputs, outputs, number of people and communities benefited by each component (and by multiple components/integration) and for how long, actual coverage compared to planned coverage, and cost.
- **Final evaluation reports** the FAFSA-2 team reviewed seldom clearly described the development hypotheses and program models

²⁸⁷ This is the method used in the FAFSA-2 and described in Section 6.4.

the projects employed. Basic information, such as what interventions programs delivered to how many people for how long, was often lacking. The requirement to include this information is not specified in USAID/FFP's guidance (FFPIB 09-07, 2009b). Furthermore, evaluators usually visit programs for a month or less and are not as familiar with the specific interventions and approaches used as are the implementers. As such, many evaluations present mainly the quantitative survey data on outcomes and impacts without much explanation as to why programs did or did not achieve targets. The lack of this information in the final evaluation report and the lack of a requirement for a final performance report by the Awardee make it difficult to learn which approaches did or did not work and why, and inhibit replication of successful program models.

- Central to USAID's Evaluation Policy is *transparency*, achieved by publicly sharing evaluation findings in a timely manner as widely as possible. Transparency is a guiding principle that USAID/FFP has followed for many years by requiring Awardees to post their ARRs, baseline surveys, and mid-term and final evaluation reports to the DEC. Yet, when the FAFSA-2 team looked for Title II final evaluations on the DEC, it could find only 23 of the 67 documents, i.e., only one-third of those that should have been there. (See Table 9.1 at the end of this chapter.) The GAO also found only 16 percent of Title II program reports on the DEC that should have been there and found that USAID/FFP had not ensured that Awardees routinely comply with this requirement (2009). According to USAID's 2011 Evaluation Policy, within three months of completing an evaluation, the report should be posted on the DEC. Some Awardees told the FAFSA-2 team that they are reluctant to post their reports on the DEC and share their tools because other PVOs may use these to compete against them during the RFA process for new Title II awards.
- Once reports are submitted to the DEC, the FAFSA-2 team learned that there can be lengthy delays in actually posting them. Moreover, searching the DEC for Title II program

evaluations by logical keywords does not readily produce reports that are on the DEC.

- Another impediment to transparency and to learning from results is that there is no requirement for public posting of the evaluation data from Title II development programs. Evaluators determine what results data to include in final evaluation reports and which are to be posted on the DEC. Grantees of USAID's CSHGP are required to post their evaluation data on the CSHGP website, where other grantees can see and learn from them or use the data for secondary analysis. This could be a good model to follow.
- **Database on interventions and approaches.** There is no reporting on or a USAID/FFP database with standardized information on common interventions and approaches used in each Title II program. Therefore, USAID/FFP has no readily available data source to describe what interventions and approaches it is supporting and where. The FAFSA-2 team had to hand count this information by reading and tallying from the program documentation, using an Excel spreadsheet designed for the purpose (HN only), a tedious, time-consuming, and massive process that should be automated going forward. This is another example of underrepresentation of the Title II program in terms of the many technical areas it works in. Not documenting this also makes it difficult to detect interventions and approaches that are rarely done that should be promoted more. A lack of descriptive information on what Title II programs are doing in different countries also impedes the ability to coordinate and co-program with other USAID programs and partners. There are many other needs and requests for this information that cannot be met.
- **Data for decision making.** Although this is an important step in USAID's performance management process, the extent to which USAID/FFP and Awardees are using Title II M&E data to improve program design, implementation, and management is unclear from the program documentation and interviews undertaken during the FAFSA-2. The FAFSA-2 found few examples where mid-term evaluation findings drove mid-

course corrections and little technical guidance by USAID/FFP based on evidence from the many Title II evaluations conducted. The FAFSA-2 team did not encounter any USAID/FFP-commissioned cross-cutting studies or in-depth analyses of Title II evaluation results to advance organizational learning other than the FAFSA and the FAFSA-2. Much greater use could be made of the evaluation data and findings for systematic reviews, meta-analyses, secondary analyses, learning, adapting, and decision making.

9.6 Conclusions and Recommendations

9.6.1 Conclusions

M&E General

- There were major advances in Title II M&E during the FAFSA-2 time period, e.g., standardized indicators; new indicators; normative guidance from USAID/FFP in Information Bulletins; the requirement for independent, population-based, quantitative evaluation surveys; and M&E training workshops.
- The shortage of M&E expertise in both USAID/FFP and Awardee organizations is a constraint. There are not enough qualified evaluators.

Monitoring

- Valuable efforts to strengthen monitoring of Title II programs by USAID/FFP included Layers and monitoring plans, but more attention to improving monitoring is needed.

Evaluation

- USAID/FFP's evaluation requirements for Title II development programs were more extensive and ambitious with respect to quantitative performance data than those of other USAID programs in the same time frame. As a result, there is a wealth of outcome- and impact-level performance data available to assess the Title II development program overall.

- USAID/FFP's expectation that all its Awardees have the capability to design and conduct, or to oversee others to implement, quality quantitative evaluation surveys is unrealistic. Specialized sampling and survey research skills are essential to ensure that data collected are reliable, valid, and generalizable. These skills are scarce among the Awardees. Thus, the laudable requirement to do quantitative evaluations with the goal of generating solid data on program effectiveness was frustrated by the poor quality of many (46 percent of MCHN evaluations) of the baseline and final surveys. Problematic baseline and endline surveys with questionable data represent a great deal of wasted effort and resources. Such evaluation surveys reduce the amount of reliable evidence about program outcomes and impact. Furthermore, the considerable effort Awardees put into these surveys likely detracted from the more important task of implementing their programs well. This experience argues strongly for USAID/FFP centralizing, professionalizing, standardizing, and making independent the conduct of future Title II development program evaluations, including baseline and final evaluation surveys. The 2002 FAFSA also recommended reducing Awardees' responsibility for conducting complex and burdensome evaluation activities, such as baseline and final evaluation surveys, and instead suggested using external TA to improve evaluation quality and timeliness.
- The contribution to organizational learning of many Title II final evaluation reports is reduced by the absence of: an adequate description of the development hypothesis; the interventions and implementation approaches used; the project's inputs, outputs, and processes; and the number of people and communities benefited (compared to the expected number) by each component and by multiple components (integration), and for how long and at what cost. The reader is often unable to discern the logical structure of the project and what it was supposed to accomplish and is not given sufficient detail to be able to replicate a successful model. Most reports focus mainly on the outcome and impact indicators and do not describe the program model or analyze

why results were or were not achieved so the development community can learn from the experience.

Reporting

- The Title II development program is seriously underrepresented, undercounted, and undervalued in terms of the many technical areas it supports and results it achieves. There is no central database that tracks interventions and approaches ongoing in field programs. Not knowing what is being supported where has many drawbacks. It impedes being able to readily quantitatively describe/aggregate ongoing programs, as well as coordinating and co-programming with other USAID programs. Recent steps by USAID/FFP to better align Title II reporting and “F” reporting by using the same program elements and indicators are positive developments, but gaps remain. For example, there is nowhere to report infrastructure and FFW supported by Title II.
- The IPTT is an invaluable tool for Awardees and USAID to monitor, record, and report the progress and final results of each Title II program using agreed-on targets and indicators. The usefulness of a number of IPTTs was reduced by the lack of a standard template, format issues, and incomplete information. Furthermore, unless made publicly available on the DEC as part of the final evaluation or final report on the program, the data reported in IPTTs are not available for broader learning and accountability.

9.6.2 Recommendations

All recommendations listed below are for actions by USAID/FFP unless otherwise noted.

General M&E

- Work together with Awardees to harmonize indicators in countries with multiple Title II development programs and ensure that data are collected and analyzed in a standardized way. The USAID/FFP standard indicators should be universally used to aid comparing performance across projects worldwide. In-country harmonization would further facilitate

performance management at that level. Individual programs could have additional indicators, as needed.

Monitoring

- Strengthen monitoring of Title II development programs by USAID/FFP and Awardees. **(Recommendation 8)**²⁸⁸

Evaluation

- Have Title II program baseline and final evaluation surveys done independently and standardized by contracting a professional, central survey research organization. The centrally funded survey research firm would select and contract local data collection firms and supervise all of the field work. The central survey research firm would also do the data analysis and report preparation and dissemination.²⁸⁹ Another essential task for this central contractor would be creating a worldwide database with the Title II evaluation data, similar to the *StatCompiler* used for accessing DHS data.²⁹⁰ The central evaluation survey research contractor should make data publically available electronically; strive for as much standardization and cross-program, cross-country comparability as feasible; and undertake in-depth analyses based on adequate sample size. For example, in-depth analyses could include disaggregating evaluation results to compare effectiveness of different intervention packages, impact by length of exposure to the intervention, or participation and results by socioeconomic status or other sub-groups. **(Recommendation 11)**

²⁸⁸ The numbers after certain recommendations are the same as those assigned to the major recommendations in the FAFSA-2 summary report.

²⁸⁹ If monetization funds in the Awardees’ agreements are the only way USAID/FFP has to fund local survey costs, then Awardees could use monetization proceeds to hire the local survey data collection agency, with the central survey research organization overseeing the selection process and supervising the performance of the local firm. Ideally, however, the evaluations should be entirely external and directly managed by the central survey research contractor, with no direct involvement of Awardees.

²⁹⁰ <http://www.statcompiler.com>.

- Along with the survey report, require the survey organization to submit a checklist of whether quality survey performance standards were met. (**Recommendation 12**)
- Centralize and professionalize independent *qualitative* evaluations of Title II development programs through several contracts with qualified firms. These contractors would not necessarily be the same as the survey contractor mentioned previously. (**Recommendation 11**)
- Consider the *total* dollar value of *all* Title II development projects in a country as one block to determine whether together they are large enough to meet the criteria for a performance evaluation of all of them to be conducted according to the USAID Evaluation Policy. Continue to encourage external evaluations of smaller Title II projects for which performance evaluation is not required. (**Recommendation 14**)
- Have the central evaluation survey contractor do *joint* baseline and final evaluation surveys for all Title II programs in countries with multiple Awardees and projects with similar timelines. USAID/FFP should also make joint final *qualitative* evaluations the standard in countries with multiple Awardees. Surveys should measure common indicators across programs, as well as indicators that each Awardee wants to evaluate. Evaluation reports should contain joint findings as well as individual findings for each Awardee. (**Recommendation 15**)
- Ensure that there are enough professionals available with the relevant technical expertise and knowledge of Title II development programs to meet the demand for independent, high-caliber evaluators. This could be accomplished by: (1) contracting qualified firms to do the evaluations; (2) having a core professional staff dedicated to Title II program evaluation at such firms, as well as short-term consultants hired to do the evaluations; and (3) inviting Title II development program evaluators to attend technical capacity building and knowledge-sharing events organized by USAID/FFP, its TA partners, or Awardees. (**Recommendation 13**)
- When programs are extended by a year or more, require the Awardee to commit to and measure new results targets through the new end date.
- **Mid-term evaluations.** USAID/FFP and FANTA should promote further experimentation by other Title II projects with the two-phase approach of collecting quantitative data on project indicators using LQAS, and then doing qualitative inquiry to identify reasons for lagging indicators and implementing an action plan to improve the program (the CRS/Haiti example). The utility of this methodology for mid-term evaluations could be determined through testing in other programs. FANTA and CRS could prepare a Technical Note on the mid-term evaluation methodology used in Haiti to aid others to replicate and experiment with it. Similarly, FANTA should disseminate information on promising approaches used by other Awardees and provide TA to develop better models for Title II mid-term evaluations. FANTA could then document more effective methodologies in a Technical Note. (**Recommendation 9**)
- Conduct USAID-led *program reviews* to complement mid-term evaluations done by Awardees. Ensure that the reviews are done by qualified teams. (**Recommendation 10**)

Reporting

- Require all Title II Awardee reporting documents, including the IPTT, to be in at least size 10 point font. All documents should be labeled with the date, country, name of Awardee, years of the program, and agreement number, including the IPTT and Resource and Beneficiary Tracking Tables, which are often submitted as separate documents from the main ARR.
- Add results data to the **IFAR** and make it look more like USAID's GHCS Report to Congress. Similarly, ensure that the GHCS report includes more Title II results. Another column should be added in the IFAR appendix table on Title II development programs to report direct beneficiary numbers for Title II project activities, not just food recipients. (**Recommendation 18**)

- Standardize the **IPTT** form with a required template, as the current IPTT example included in the RFA is contributing to errors. Have Awardees complete the IPTT that they submit to USAID as an online form that won't allow omissions. An IPTT submitted as an online form could serve as the data source for standard indicators for overall performance reporting, in lieu of the separate SAPQ form. The IPTT template should include the date of the baseline and final evaluation surveys (calendar month and year), sample size for each indicator, name of country, Awardee organization, agreement number, separate rows for monitoring data versus survey data for the same indicator, and CI.
- Ensure that all relevant USAID/FFP **standard indicators** are included in the IPTT, that indicators are logically appropriate for the program model, that behavior change indicators at the highest outcome level measure practices not knowledge, and that the groups measured for people-level indicators (the denominator) are consistent with the project's target groups. Use FANTA TA, as needed.
- Post final IPTT data online for access by other Awardees. If such data were readily available, they could also be used in the IFAR to tell the story of the impact of the program and for secondary analysis by Awardees or USAID. (**Recommendation 17**)
- In the **Tracking Table for Beneficiaries**, make another category for household members receiving food in MCHN and HIV activities and instruct Awardees not to include these people under the Nutrition, MCH, or HIV program elements. It would be useful to redesign the AER to have food recipient categories for MCH-mother, MCH-child, and MCH-household member, plus FFW, PLHIV, and PLHIV-household member.
- Design a standard table for inclusion in ARRs that would better report on how many direct beneficiaries received each of a project's key interventions, how many received multiple interventions, and the percent coverage of the total eligible population in the project area or of the expected number to be covered per project targets.
- Add program elements for infrastructure to the AER and to the Resources and Beneficiaries Tracking Tables in the PREP and ARR (use the "F" Program Element 4.4.3, "Transport Services," for work on roads and Program Element 5.2.2, "Mitigation," for protective infrastructure to enhance emergency preparedness and disaster management). Establish a method for capturing information on other kinds of Title II-assisted infrastructure and FFW, which fits the definition of "F" Sub-Elements 3.3.3.4, "Self-Help Programs," and 4.5.2.2, "Land and Water Management." (**Recommendation 19**)
- Develop a database to better describe the interventions and approaches used in Title II development programs. The data should come from a new, standard, electronic, online reporting form that Awardees would be required to submit annually to USAID/FFP for each Title II development program. In the form, Awardees would check off the standard types of interventions and approaches they are implementing in each program from a menu with definitions, agreed upon by USAID, Awardees, and other technical experts and stakeholders.²⁹¹ (**Recommendation 20**)
- Require final performance reports by Awardees, have more detailed guidelines for required sections in final evaluation and final performance reports, and include the completed IPTT in both. Follow the USAID Evaluation Policy Checklist for Assessing USAID Evaluation Reports.²⁹² Require proof in the final performance report that Awardees have posted the final evaluation to the DEC (e.g., assigned DEC numbers or other proof of transmittal). (**Recommendations 16 and 21**)

²⁹¹ The Excel spreadsheet developed for the review of HN in the FAFSA-2 could be used as a starting point.

²⁹² http://transition.usaid.gov/policy/evalweb/evaluation_resources.html.

- To maximize the utility of *final evaluation reports* and *final performance reports* for learning and accountability, these reports need to clearly describe the program model and its implementation in enough detail to allow for replication. This should include information on the: (1) development hypothesis; (2) interventions and approaches; (3) inputs, outputs, and processes; (4) final IPTT with results, sample size, and CIs for all indicators; (5) number of people and communities benefited, by each separate component and by multiple components, compared to targets, and for how long; and (6) cost. **(Recommendation 21)**
- Review whether the **SAPQ** form is the most efficient method for collecting data on standard indicators from Awardees or whether a required online IPTT template could serve the same purpose. USAID/FFP should use the results data reported in the SAPQs. If reports produced from the SAPQ information are not essential to USAID/FFP for performance management, then consider eliminating the requirement for Awardees to submit the SAPQ. If the SAPQ continues to be required, FANTA and USAID/FFP agree that it should be redesigned for Awardees to complete as an online form with data cleaning filters to reject errors. This would also speed up delivery and analysis of the forms.
- Post summaries of Title II programs on the USAID/FFP and/or TOPS websites, following the example of the CSHGP one-page reports on specific child survival projects and their results featured on the CSHGP home page.²⁹³
- Enforce the requirement that Awardees post project reports and evaluations to the DEC. **(Recommendation 16)**
- Request for all Awardees with final evaluation reports for completed Title II development programs not found in the DEC by the FAFSA-2 (see Table 9.1) to immediately post these missing

reports to the DEC and to provide USAID/FFP with the assigned DEC number for the document. **(Recommendation 16)**

- **DEC.** Work with personnel responsible for the DEC to improve the keyword search so that it readily locates Title II reports, and require Awardees to submit reports identified by these keywords.

Learning from Results

- Use the Title II development results data and evaluation findings for decision making and learning.
- Have a technical contractor analyze Title II development program-wide (worldwide) results data and evaluation findings annually and organize knowledge-sharing events to foster learning from the evaluations. The focus could be to review one technical sector or approach in-depth each year. In addition, this contractor should present the program-wide results in a format that USAID/FFP can readily use for high-level reporting and wide dissemination to communicate what Title II development programs accomplish. **(Recommendation 2)**

²⁹³ See <http://www.mchipngo.net/controllers/link.cfc?method=home>.

Table 9.1. Availability of Title II Final Evaluations in USAID's DEC as of June 2011

Country	Awardee	Dates of Program	Grant Number	Final in DEC?		DEC Number
				Yes	No	
Bangladesh	CARE	FY05–10	FFP-A-00-04-00079	1		PD-ACP-746
	SC	FY05–10	FFP-A-00-04-00080	1		PD-ACP-903
Bolivia	ADRA	FY03–09	FFP-A-00-02-00052	1		PD-ACO-837
	CARE	FY02–09	FFP-A-00-02-00028			
	FH	FY02–09	FFP-A-00-02-00057			
	SC	FY02–09	FFP-A-00-02-00056			
Burkina Faso	CRS	FY04–10	FFP-A-00-03-00076		1	
	AFRICARE	FY04–10	FFP-A-00-04-00078		1	
Cape Verde	ACDI/VOCA	FY03–08	FFP-A-00-02-00022		1	
Chad/Mali	AFRICARE	FY03–08	FFP-A-00-02-00094		1	
Ethiopia	CARE	FY02–05	FFP-A-00-03-00002		1	
	CRS	FY03–07	FFP-A-00-02-00095		1	
	REST	FY03–07	FFP-A-00-03-00004		1	
	SC US	FY03–07	NA		1	
	WV	FY03–08	FFP-A-00-03-00003		1	
	CARE	FY05–08	FFP-A-00-05-00032		1	
	CRS	FY05–08	FFP-A-00-05-00027			
	FH	FY05–08	FFP-A-00-05-00030			
	REST	FY05–08	FFP-A-00-05-00028			
	SC US	FY05–08	FFP-A-00-05-00029			
	WV	FY05–09	FFP-A-00-05-00031			
WV	FY05–09	FFP-A-00-05-00031				
Ghana	ADRA	FY02–08	FFP-A-00-02-00015	1		PD-ACH-634
	CRS	FY03–08	FFP-A-00-03-00081		1	
	OICI	FY04–09	FFP-A-00-04-00085		1	
	TNS	FY06–10	NA		1	
Guatemala	CARE	FY01–08	FFP-A-00-01-00012	1		PD-ACT-247
	CRS	FY01–07	FFP-A-00-02-00007			
	SC	FY00–07	FFP-A-00-02-00027			
	SHARE	FY03–07	FFP-A-00-03-00015			
Guinea	ADRA	FY00–09	FFP-A-00-00-00085	1		PD-ACD-461
	AFRICARE	FY01–08	FFP-A-00-01-00034		1	
	OICI	FY05–09	FFP-A-00-04-00074		1	
Haiti	CARE	FY02–08	FFP-A-00-02-00055	2		PD-ACJ-025 PD-ACN-886
	CRS	FY02–08	FFP-A-00-02-00053			
	SC	FY02–08	FFP-A-00-02-00045			
	WV	FY02–08	FFP-A-00-02-00054			
Honduras	CARE	FY01–08	FFP-A-00-01-00002	1		PD-ACP-775
				1		PD-ACP-998
	ADRA	FY05–09	FFP-A-00-05-00001	1		PD-ACP-894
	SC	FY05–09	FFP-A-00-04-00071			
	WV/ACDI/VOCA	FY05–09	FFP-A-00-05-00002			
India	CRS	FY02–06	FFP-A-00-02-00017	1		PD-ACH-642
	CRS	FY07–10	FFP-A-00-07-00025		1	
	CARE	FY02–06	FFP-A-00-02-00011	1		PD-ACI-026
	CARE	FY07–10	FFP-A-00-07-00024		1	

Country	Awardee	Dates of Program	Grant Number	Final in DEC?		DEC Number
				Yes	No	
Indonesia	CARE	FY05–08	FFP-A-00-04-00070		1	
	CRS	FY05–08	FFP-A-00-04-00069		1	
	MC	FY05–08	FFP-A-00-04-00072		1	
	SC	FY05–09	FFP-A-00-04-00076		1	
	WV	FY05–08	FFP-A-00-04-00068		1	
Kenya	ADRA	FY03–08	FFP-A-00-03-00082		1	
	CARE	FY04–09	FFP-A-00-03-00085	1		PD-ACQ-780
	CRS	FY01–06	FFP-A-00-01-00003		1	
	FH	FY04–08	FFP-A-00-03-00090		1	
	WV	FY03–07	FFP-A-00-02-00001		1	
Liberia	LIAP Consortium (CRS)	FY07–10	FFP-A-00-07-00008		1	
Madagascar	ADRA	FY04–09	FFP-A-00-03-00096		1	
	CARE	FY03–09	FFP-A-00-03-00078		1	
	CRS	FY03–08	FFP-A-00-03-00089		1	
Malawi	I-LIFE Consortium (CRS)	FY05–09	FFP-A-00-04-00066		1	
Mauritania	WV	FY01–07	FFP-A-00-01-00044		1	
Mozambique	ADRA	FY02–08	FFP-A-00-02-00008		1	
	AFRICARE	FY02–08	FFP-A-00-00-00013		1	
	CARE	FY02–08	FFP-A-00-02-00014		1	
	FH	FY02–08	FFP-A-00-02-00010		1	
	SC	FY02–08	FFP-A-00-02-00009		1	
	WV	FY02–08	FFP-A-00-02-00012		1	
Nicaragua	ADRA	FY02–09	FFP-A-00-02-00002	1		PD-ACO-867
	CRS	FY02–09	FFP-A-00-02-00004			
	PCI	FY02–09	FFP-A-00-02-00005			
	SC	FY02–09	FFP-A-00-02-00003			
Niger	AFRICARE	FY00–07	FFP-A-00-00-00087		1	
Rwanda	ACDI/VOCA/AFRICARE	FY00–05	FFP-A-00-00-00037	1		PD-ACD-646
	ACDI/VOCA/AFRICARE	FY05–10	FFP-A-00-04-00073		1	
	CRS	FY00–09	FFP-A-00-00-00086	1		PD-ACD-647
	WV	FY04–09	FFP-A-00-04-00084		1	
Senegal/Gambia	CRS	FY02–07	FFP-A-00-02-00029		1	
Sierra Leone	CARE (AFRICARE, CRS, WV)	FY04–07	FFP-A-00-04-00020	1		PD-ACN-868
	CARE (AFRICARE, CRS, WV)	FY07–10	FFP-A-00-07-00007	1		PD-ACR-358
Uganda	ACDI/VOCA	FY02–06	FFP-A-00-02-00016	1		PD-ACI-108
	AFRICARE	FY02–06	FFP-A-00-02-00073	1		PD-ACG-740
	CRS	FY02–06	FFP-A-00-02-00006	1		PD-ACG-739
	SC	FY03–09	FFP-A-00-03-00086		1	
	WV	FY03–09	FFP-A-00-03-00079		1	
Zambia	LOL	FY04–09	FFP-A-00-04-00001	1		PD-ACM-615

Bibliography for Chapter 9

Bergeron, Gilles; Deitchler, Megan; Bilinsky, Paula; and Swindale, Anne. 2006a. "Monitoring and Evaluation Framework for Title II Development-Oriented Projects." Technical Note No. 10. http://www.fantaproject.org/downloads/pdfs/TN10_MEFramework.pdf.

Bergeron, Gilles; Swindale, Anne; Deitchler, Megan; and Bilinsky, Paula. 2006b. "Evaluating Title II Development-Oriented Multi-Year Assistance Projects." Technical Note No. 11. http://www.fantaproject.org/downloads/pdfs/TN11_MYAP.pdf.

Bonnard, Patricia. 2002. "Title II Evaluation Scopes of Work." Technical Note No. 2. <http://www.fantaproject.org/downloads/pdfs/tn2SOW.pdf>.

GAO. 2009. "USAID Is Taking Actions to Improve Monitoring and Evaluation of Nonemergency Food Aid, but Weaknesses in Planning Could Impede Efforts." <http://www.gao.gov/products/GAO-09-980>.

Swindale, Anne; Deitchler, Megan; Cogill, Bruce; and Marchione, Thomas. 2004. "The Impact of Title II Maternal and Child Health and Nutrition Programs on the Nutritional Status of Children." USAID/FFP Occasional Paper No. 4. <http://www.fantaproject.org/downloads/pdfs/ffpOP4.pdf>.

USAID. 2009. *Two Decades of Progress: USAID's Child Survival and Maternal Health Programs*. Washington, DC: USAID.

USAID. 2010. *Striving for a Healthy Future*. Global Health and Child Survival Report to Congress, Fiscal Year 2009. Washington, DC: USAID.

USAID. 2011. *Evaluation Policy*. Washington, DC: Bureau for Policy, Planning and Learning, Office of Learning, Evaluation and Research, USAID.

USAID/FFP. 2005. *Strategic Plan for 2006-2010*. Washington, DC: USAID.

USAID/FFP. 2007. "New Reporting Requirements for Food for Peace." FFPIB 07-02.

USAID/FFP. 2009a. "Monitoring and Evaluation Responsibilities of Food for Peace Multi-Year Assistance Programs Awardees." FFPIB 09-06.

USAID/FFP. 2009b. "Description of Food for Peace Awardee Reporting Requirements." FFPIB 09-07.

USAID/FFP. 2011a. "Submission of Baseline/Final Evaluation Datasets by Food for Peace Awardees." FFPIB 11-02.

USAID/FFP. 2011b. "Revision to Food for Peace Standard Indicators Collected in Baseline Surveys and Final Evaluations (Reissuance)." FFPIB 11-03 (Reissued).

USAID/U.S. Department of State. 2010. "Foreign Assistance Standardized Program Structure and Definitions." <http://www.state.gov/documents/organization/141836.pdf>.

USAID and USDA. 2011. *U.S. International Food Assistance Report 2010*. Washington, DC: USAID.