

PURCHASE FOR PROGRESS MONITORING AND EVALUATION SYSTEM DESIGN CONSIDERATIONS: FINAL REPORT



**World Food
Programme**

MSI MANAGEMENT
SYSTEMS
INTERNATIONAL
A Subsidiary of Coffey International, Ltd.

Prepared by World Food Programme and
Management Systems International

coffey  international development
SPECIALISTS IN DEVELOPING COMMUNITIES

Version 2, December 7, 2010

Table of Contents

ACKNOWLEDGMENTS.....	IV
NEW IN THIS VERSION	V
INTRODUCTION.....	1
WFP LOCAL PROCUREMENT PRACTICES	3
The Potential Benefits and Risks of Local Procurement	3
Local and Regional Procurement.....	4
Purchase for Progress	5
MONITORING AND EVALUATION SYSTEM DESIGN METHODOLOGY	9
Refining the Logical Framework.....	9
Developing Data Collection Instruments.....	9
Validating the Monitoring and Evaluation Approach	10
MONITORING AND EVALUATION OBJECTIVES & DESIGN.....	12
Monitoring and Evaluation Objectives	12
Evaluation Questions.....	13
MONITORING AND EVALUATION SYSTEM OVERVIEW.....	17
Monitoring and Evaluation System Functions.....	17
Logical Framework	18
Data Collection.....	27
Data Types, Sources, and Collection Procedures	27
Data Collection Schedule.....	30
Data Analysis.....	33
Calculating Indicator Values.....	33
Incorporating Qualitative Data.....	34
Disaggregating by Sex	35
DESIGN CONSIDERATIONS	36
Global Versus Country-Level Monitoring and Evaluation.....	36
Impact Assessment Approach & Sampling.....	37
Alternative Approaches to Impact Assessment	37
The P4P Impact Assessment Approach.....	38
Implementation Examples	42
Defining P4P Beneficiaries	44
How Does P4P Define Beneficiaries?	45
Counting Beneficiaries	46
When is a Participant not a Beneficiary?	47
Frequency of Beneficiary Estimates.....	48
Measuring Household Income	49
Evaluating Standard Local and Regional Procurement	50
Seasonality	52
Monitoring.....	53
Maintaining Data Quality	54
Country Office Capacity.....	55

Evaluations	56
Case Studies.....	56
ANNEX A.....	58
P4P Monitoring and Evaluation System Components	58
ANNEX B.....	59
Design Team Field Visits	59
Pretesting Data Collection Instruments and Procedures.....	60
ANNEX C.....	61
Purchase for Progress Logframe Matrix	61
REFERENCES.....	73

Table of Figures

Figure 1. P4P Monitoring and Evaluation Questions	14
Figure 2. Assessment Criteria.....	15
Figure 3. Illustrative Detailed Evaluation Questions.....	16
Figure 4. P4P Outcomes and Indicators	20
Figure 5. Elements of Farmer Data Collection Instrument	27
Figure 6. Elements of Farmers' Organization Data Collection Instrument	28
Figure 7. Elements of Trader Data Collection Instrument	29
Figure 8. Data Collection Schedule.....	32

ACKNOWLEDGMENTS

The P4P monitoring and evaluation design team owes an enormous debt to the many people who collaborated in the design effort. Arlene Mitchell, Alesha Black and others at the Bill and Melinda Gates Foundation offered steady guidance, constructive feedback and encouragement throughout the design process. Colleagues in WFP's Food Security Analysis Unit: Joyce Luma, Jan Delbaere, Chiara Brunelli, Cinzia Papavero, Claudia Ah Poe, Issa Sanogo and George Muammar also deserve a special thank you for the many hours of fruitful discussion and critique of the approach and for their contribution to the questionnaires. Thanks are also due to Stephen Were Omamo, Deputy Director, Policy, Planning and Strategy Division of WFP for his unfailing support and insightful guidance throughout the design process.

The candid feedback and practical perspective of country office staff in the 21 P4P pilot countries contributed substantially to the end product. The four WFP Country Offices that hosted the design team's pretesting mission (Burkina Faso, Malawi, Uganda and Zambia) deserve a special commendation. The exceptional coordination of the pretests reflected the time and resources the country offices committed to the effort and assured a productive field experience. The design team is indebted to all those who participated in the pretests for the intellectually stimulating discussions and the camaraderie that made a lengthy mission informative as well as enjoyable and is deeply grateful to the smallholder farmers, farmers' organization staff, traders and others who gave generously of their time to meet with the team during the pretest process.

Finally, a Technical Review Panel of development experts met with the design team in May 2008 to provide guidance on system design. These individuals (Erin Coniker Lentz, Care/Cornell; Lawrence Haddad, Institute for Development Studies; Dave Tschirley, Michigan State University; Penny Ward, Mutengo Consulting; Maximo Torero, IFPRI; Rashid Shahidur, IFPRI; Rick Hodges, Natural Resources Institute; Marta Valdés García, Oxfam Intermon and Alexander Woollcombe, Oxfam International) made time in their busy schedules to travel to Rome to share their extensive knowledge and experience with the design team. The entire P4P Coordination Unit thanks these individuals for their unselfish contributions of time and intellectual energy and for the thoughtful feedback and guidance.

This second version of the report also owes much to the Technical Review Panel members, WFP country offices, and WFP headquarters units personnel who have contributed to the evolving monitoring and evaluation system.

NEW IN THIS VERSION

The second version of the report documents a number of changes in the monitoring and evaluation system that the Coordination Unit has implemented in response to input from advisors and lessons from a year of implementation. Key changes relative to the first version of the report include:

- The task of collecting and analyzing baseline data emphasized the intense effort required to collect primary survey data annually. To minimize this burden and create space for reflection and learning, the Coordination Unit, on the recommendation of the Technical Review Panel, has changed the requirement for surveys of smallholder farmers from annual surveys to surveys in years one, three, and five.
- The requirements of rigorous impact assessment (i.e., identifying and collecting data from non-participating farmers' organizations and farmers) also represented a substantial burden. Furthermore, for a number of reasons not all countries provide an environment suitable to impact assessment. On the recommendation of the Technical Review Panel, the Coordination Unit will now pursue impact assessment only in selected countries that offer good prospects for meaningful impact assessment. All other countries are not required to collect data from non-participating farmers' organizations and farmers.
- To increase the usefulness of data, all countries will follow panels of farmers. This means that countries will collect data from the same farmers in each year that they conduct surveys of farmers.
- This version of the report develops a precise definition of beneficiaries to aid in reporting.
- This version of the report describes the external evaluations and how they fit into the overall monitoring and evaluation system.
- This version of the report also provides additional information and guidance on conducting case studies.

INTRODUCTION

As the world's largest humanitarian agency, and one focused on food assistance, the World Food Programme (WFP) purchases large quantities of food for distribution in emergency and other situations. Over the past two decades, WFP has purchased an increasing quantity of this food from developing countries – often from surplus-producing regions of countries in which it distributes food. In 2008, the organization bought more than 2.8 million metric tonnes of food worldwide with a value of more than \$1.4 billion. Purchases from developing countries accounted for 76 percent of the quantity and 78 percent of the value of all of WFP's food purchases (World Food Programme 2009).

Food aid and development professionals generally agree that procuring food locally “provides much greater net benefits for rural and urban populations than equivalent expenditure on tied food aid ...”(Coulter; Walker, and Hodges 2007). Therefore, the rising trajectory of local food procurement is consistent with WFP's 2008-2013 Strategic Plan which establishes a goal of “using [WFP's] purchasing power to support the sustainable development of food and nutrition security systems, and transform food and nutrition assistance into a productive investment in local communities.”

However, little is known about how local procurement generates benefits and how the way in which buyers structure local procurement affects the types of benefits, how those benefits are distributed, and whether it produces sustainable change in the food system. To maximize the development impacts of its local procurement, WFP must learn 1) how to leverage its presence in markets to promote market development, market access, and increased incomes for smallholder farmers and 2) how, when, and under what conditions to buy locally to maximize development impacts without unduly compromising food assistance objectives.

Purchase for Progress (P4P) is a pilot programme designed to help WFP answer these questions. Over a five-year period in 21 countries¹, WFP will test alternative means of procurement and examine the impact of these purchases on market development, access to markets, and the welfare of smallholder farmers. WFP will use what it learns to more effectively advocate for smallholder-friendly policy, influence the actions of other agricultural market development stakeholders, and ultimately transform the way it buys in developing countries.

Purchase for Progress focuses on learning how best to structure WFP's procurement to benefit smallholder farmers. The emphasis on learning brings monitoring and evaluation to the forefront of programme design. Deliberative, comprehensive, and rigorous monitoring and evaluation will be crucial to assessing the effectiveness of interventions under various operating environments, identifying best practices, and learning how and under what conditions alternative procurement mechanisms are most appropriate. The remainder of this report describes the monitoring and evaluation system developed for P4P.

The report contains five main chapters covering different aspects of the monitoring and evaluation system.

¹ The 21 P4P pilot countries are Afghanistan, Burkina Faso, Democratic Republic of Congo, El Salvador, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Laos, Liberia, Malawi, Mali, Mozambique, Nicaragua, Rwanda, Sierra Leone, Sudan, Tanzania, Uganda, and Zambia.

- The chapter on **WFP Local Procurement Practices** reviews the evolution of WFP’s local procurement and provides the context necessary to understand the design of P4P and the questions the monitoring and evaluation system is designed to address.
- The **Monitoring and Evaluation System Design Methodology** chapter describes the approach that WFP employed to design the monitoring and evaluation system.
- The **Monitoring and Evaluation Objectives & Design** chapter outlines the objectives of the monitoring and evaluation system.
- The **Monitoring and Evaluation System Overview** chapter provides a summary overview of the functions and components of the monitoring and evaluation system.
- The **Design Considerations** chapter presents a detailed accounting of design issues and the solutions adopted to address the issues. The discussion provides additional insight into the design of the system and documents the justification for critical decisions that the design team made.

This report is one component of a larger body of reports and manuals that document the overall monitoring and evaluation system for P4P.² The report describes the monitoring and evaluation system (i.e. its objectives, components, and implementation) and perhaps more importantly, provides the rationale for the design that WFP has adopted to monitor and evaluate the achievements of P4P.

² Annex A lists the components of the P4P monitoring and evaluation system.

WFP LOCAL PROCUREMENT PRACTICES

WFP has been procuring locally since at least 1985.³ In all but one year between 1986 and 2008, the quantity purchased from developing countries exceeded that from developed countries. With a fivefold increase in total procurement between 1986 and 2008, WFP is spending an ever increasing amount of money to buy food in developing countries. In 2008, it bought over 2.77 million metric tonnes of food from developing countries with a value of over \$1 billion US.

This chapter traces the evolution of WFP's local procurement practices and thus provides the context for P4P and a foundation for understanding the monitoring and evaluation system on which this report focuses. The chapter briefly catalogs the potential benefits and risks associated with local procurement and describes how WFP has structured its local procurement in response. It frames P4P as WFP's attempt to learn how to maximize development benefits while managing risk. The chapter concludes with a description of the salient features of P4P.

The Potential Benefits and Risks of Local Procurement

Procuring locally is often preferable to importing food for several reasons.

- Lower transportation costs and (sometimes) lower commodity prices mean that with a given budget, WFP can often buy more locally produced food than imported food therefore assisting more people (Sserunkuuma & Associates Consult 2005;Tschirley and del Castillo 2007).
- Also, because of the proximity of production to need, WFP can often obtain and deliver locally produced food more quickly than imported food.
- Finally, locally produced food is more likely than imported food to suit local preferences.

Local procurement also unquestionably produces development impacts. Coulter, et al. (Coulter and others 2007) conclude that: "It (local procurement) provides much greater net benefits for rural and urban populations than equivalent expenditure on tied food aid ...". The direct impact of local procurement is obvious – it injects large amounts of money into the local economy. In 2008, the World Food Programme alone spent about \$1.1 billion (78 percent of what it spent on food) to buy food in 73 developing countries.⁴ These expenditures almost certainly had some development impact by increasing incomes and employment. Local procurement may also produce less obvious benefits such as higher farm gate prices, strengthened formal markets, and increased investment by traders and processors (Sserunkuuma & Associates Consult 2005).

Local procurement, however, also entails unique risks for WFP. Potential risks include:

- Procuring locally, particularly when procured quantities represent a large share of marketed surpluses, can drive up commodity prices and threaten the food security of households that must purchase food (Tschirley and del Castillo 2007).

³ WFP procurement records.

⁴ WFP procurement records.

- Erratic or procyclical (with prices) procurement can accentuate seasonal fluctuations in commodity prices which also potentially threaten the food security of vulnerable segments of the population (Wandschneider and Hodges 2005).
- A higher likelihood of default on local procurement contracts creates potential pipeline risks for WFP that threaten its ability to respond to emergencies in an appropriate and timely fashion.
- Less stringent local quality controls may mean that available food does not meet WFP's quality standards which are designed to safeguard the health of recipients.⁵
- To the extent that local procurement (and particularly procurement directly from farmers' organizations) deals in smaller quantities, it may also increase the administrative cost of procurement (Lynch 2006).
- Finally, if WFP purchases non-surplus food at harvest time prices from households who can't afford to replenish food stocks at higher prices later on, then P4P may actually reduce food security.⁶

In designing a local procurement strategy, WFP must weigh the potential benefits and costs associated with alternative means of procurement.

Local and Regional Procurement

To manage these risks, and control procurement costs, WFP has traditionally conducted most of its local procurement through competitive tenders with larger traders that have the capacity to deliver the required quantities and qualities in a timely manner. WFP pre-qualifies suppliers and requires that they post substantial bid and performance bonds. Contract specifications also include delivery points (often the main WFP warehouse in the country), stringent quality standards, and bagging and marking requirements. This is the standard Local and Regional Procurement (LRP) model.

As it began to procure increasing amounts of food locally, WFP became more interested in structuring its procurement specifically to generate development impacts – especially for smallholder farmers who often produce a majority of a country's marketed surplus. To the extent that WFP can direct its procurement to smallholder farmers (many of whom live in poverty), and to the extent that buying from smallholders affects household income, its local procurement can serve the dual purpose of providing food assistance while also reducing poverty and addressing the root causes of hunger.

The Agricultural and Marketing Support Project represents WFP's most prominent attempt to enhance the development impacts of its local procurement. The project bought maize and beans from smallholder farmers in Uganda starting in 2000. However, because it was prohibitively expensive to buy small quantities from many individual farmers, WFP implemented its local

⁵ WFP's quality standards prevent dangerous food from entering the pipeline. However, poor quality may lead WFP to reject delivered food, causing pipeline breaks which put at risk WFP's ability to deliver food in a timely manner.

⁶ This situation does not necessarily increase food insecurity. In the context of high post-harvest losses, it may be perfectly rational for a household to sell non-surplus food and buy it back later at higher prices.

purchases largely through farmers' organizations that could aggregate commodities from many farmer members, thus reducing transaction costs. Because many farmers' organizations did not have the resources or capacity to participate in WFP's normal tendering process (e.g., could not provide large bid and performance bonds, could not aggregate large quantities, lacked the resources to transport to WFP warehouses, etc.) WFP relaxed some of these requirements for farmers' organizations. In particular, under the Agricultural and Marketing Support Project, WFP reduced the minimum tonnage requirement, eliminated the bid bond requirement, and arranged and paid for transportation (Sserunkuuma & Associates Consult 2005).

While the Agricultural and Marketing Support Project tested alternative local procurement models that purchased closer to smallholder farmers, it failed to meet its target of providing ten percent of locally procured maize and beans. In fact, a review of the project concluded that it never exceeded 4.7 percent (Sserunkuuma & Associates Consult 2005). The review also found that many farmers' organizations lacked the capacity and resources to sell to WFP even under the relaxed procedures. The organizations had difficulty raising cash for performance bonds, often could not aggregate sufficient quantities of commodities from their members, did not have access to adequate storage, or lacked the management capacity to determine appropriate bid prices when responding to a tender.

Purchase for Progress

Purchase for Progress represents WFP's attempt to learn how to leverage its evolving role in food markets to enhance the impact of its procurement on smallholder farmers. This section reviews the salient features of P4P that are relevant to understanding the rationale for the design of the monitoring and evaluation system. WFP's proposal to the Bill and Melinda Gates Foundation provides a much more comprehensive description of P4P for readers who desire a greater level of detail.

Purchase for Progress is a five-year pilot project emphasizing learning – learning how best to leverage WFP's presence in developing countries' food markets to increase the incomes of smallholder farmers thereby reducing poverty, chronic hunger, and food insecurity. Specific programme objectives are:

- To identify and share best practices for WFP, NGOs, governments, and agricultural market stakeholders to increase profitable smallholder/low income farmer engagement in markets.
- To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.
- To identify and implement best practices for increasing sales to WFP and others with a particular focus on smallholder/low income farmers.
- To transform WFP food purchase programmes so they better support sustainable small-scale production and address the root causes of hunger.

To achieve these objectives, P4P specifically addresses the primary constraints to profitable smallholder participation in markets. WFP's experience with the Agricultural and Marketing Support Project and a reading of a large agricultural development literature suggest that the chief constraints include:

- **Low agricultural productivity** – Smallholder farmers often face depleted soils and limited access to productivity-enhancing technologies (e.g., fertilizers, improved seeds, mechanization) and practices (e.g., knowledge of the appropriate use of fertilizers and pesticides), capital, labor, and sometimes land. These factors combine to keep many smallholder farmers producing at, or only slightly above, subsistence levels.
- **Limited access to markets** – Their (often) remote location coupled with poor transportation infrastructure and low production volumes limit smallholder farmers' access to markets by increasing marketing cost. Therefore, many smallholder farmers have few market outlets other than itinerant traders who buy from the farm gate. These traders generally face little competition and have greater knowledge of market prices than farmers – factors that can contribute to low, and sometimes non-remunerative, farm gate prices. Low prices exacerbate the low productivity problem because they do not provide market-based incentives to invest in increased production.

Many smallholder farmers have joined farmers' organizations in order to improve access to markets through group marketing. However, many farmers' organizations lack the capacity or resources to aggregate sufficient quantities or qualities to satisfy large buyers such as WFP. They may also lack access to adequate storage facilities or financing to time sales to higher priced markets. These constraints make it difficult for many farmers' organizations to participate in WFP tenders, even under relaxed tendering requirements.

The generally poor quality of smallholder crops also limits access to markets and reduces prices. Quality has several dimensions, with moisture content being the most important because grains with high moisture content will spoil. Other aspects of quality include percentage of foreign matter, small grains, broken grains, and discolored grains. In countries with dry climates during harvest, smallholder farmers can usually reduce moisture content to acceptable levels without special machinery. In other countries they generally do not have access to the mechanical dryers necessary to attain acceptable moisture levels. Smallholders can generally manually sort out foreign matter and small, broken, and discolored grain if they have the price incentive to do so. However, few markets reward smallholders for higher quality. In part this is due to a lack of quality control along the market chain. A smallholder farmer gains little from improving quality if traders mix his or her crops with others on the way to market. Similarly, intermediate traders have little incentive to pay for quality if subsequent commodity handlers mix commodities of varying quality together before sale to the ultimate buyer.

- **Limited access to credit** – Many smallholder farmers have little or no access to credit. Limited cash reserves constrain their ability to buy inputs and make other investments in agriculture. The need for cash also often forces farmers to sell commodities soon after harvest when prices are typically at their lowest.

Purchase for Progress seeks to address these constraints by increasing smallholder farmers' capacity to produce surpluses and enhancing farmers' organizations' capacity to aggregate surpluses and sell

them at remunerative prices.⁷ However, WFP specializes in procuring and distributing food and does not have specific expertise in enhancing agricultural production or market access. Purchase for Progress therefore seeks to learn how WFP can best use its procurement to catalyze partnerships to address smallholder farmers' constraints to sustainable access to remunerative markets. This statement encompasses the three pillars of P4P – procurement, partnerships, and learning and sharing. The following brief review of each pillar illustrates WFP's vision of how P4P works.

1. **Procurement** - Procurement is the main tool that WFP has at its direct disposal to influence the way food systems work. WFP will use its procurement as a tool for building the capacity of farmers' organizations. Purchase for Progress specifies four "procurement modalities" aimed at addressing different constraints that farmers' organizations may face selling to WFP.⁸ The modalities are designed to build the capacity of farmers' organizations so that they can eventually participate in competitive tenders solicited by WFP and other buyers. These modalities are competitive tendering, direct contracting, forward contracting, and transformation (processing). Beyond direct engagement with farmers' organizations, WFP expects to learn how it can best use its status as a large and respected buyer of commodities to affect change at many levels of the food system including policy, capacity building, and market development and, ultimately, to improve the welfare of smallholder farmers. By using what it learns, WFP intends to transform its local procurement activities to support sustainable market access and improved welfare for smallholder farmers.
2. **Partnerships** – Because WFP lacks specific expertise in agricultural production and market development – key constraints to smallholder engagement in markets – forming productive partnerships with organizations that do have this expertise is central to P4P's success. WFP will use its reputation and its market presence to catalyze supply-side partner support to help smallholder farmers and farmers' organizations increase their capacities for agricultural production and market access.
3. **Learning and Sharing** – As a pilot programme, P4P focuses primarily on learning. In particular, it seeks to identify and share best practices for
 - 1) using local procurement (of WFP and others) to provide incentives for smallholder investment in agriculture, encourage and support market development, and enhance sustainable and profitable access to markets for smallholder farmers and
 - 2) structuring its own local procurement to most effectively address the dual objectives of increasing development impacts while still providing food assistance efficiently, safely, and in a timely manner.

Purchase for Progress also emphasizes sharing lessons with other market development stakeholders to promote widespread adoption of pro-smallholder market development and procurement practices.

⁷ Farmers' organizations may gain higher prices by aggregating sufficient quantities to enable them to approach larger buyers directly rather than through a middleman, by increasing commodity quality thereby earning a premium price, or by adding value by transforming commodities into processed products.

⁸ P4P may use these modalities in conjunction with various market platforms including commodity exchanges and warehouse receipt systems.

Each of the 21 P4P pilot countries faces a different set of challenges and opportunities. Therefore, while the three pillars described above provide a broad general framework for implementing P4P, individual countries have considerable latitude to tailor P4P to suit their own particular environment, needs, and opportunities.

Zambia provides a prime example of adapting P4P to a country-specific environment. Capitalizing on the opportunity presented by the nascent Zambia Agricultural Commodity Exchange (ZAMACE), WFP in Zambia has determined that it can best support market development and market access for smallholder farmers by channeling all of its local procurement (P4P and LRP) through the commodity exchange. On the supply-side, it is supporting the development of rural ZAMACE-certified warehouses to facilitate smallholder access to the exchange.

The Zambia model stands in stark contrast to the *typical* P4P model of buying directly from farmers' organizations or small and medium traders. Consequently, Zambia will require a somewhat different approach to monitoring and evaluation. The "Design Considerations" chapter describes the monitoring and evaluation implications of the Zambia P4P programme.

MONITORING AND EVALUATION SYSTEM DESIGN METHODOLOGY

The monitoring and evaluation system for P4P represents a collaboration between WFP and Management Systems International (MSI), a Washington, D.C. based consulting firm. Prior to engaging MSI, the P4P Coordination Unit in Rome had begun to flesh out the requirements of the monitoring and evaluation system contained in the P4P proposal submitted to the Bill and Melinda Gates Foundation. In doing so, it had elicited feedback from other WFP operational units, country offices, and a technical review panel on specific aspects of the monitoring and evaluation system – particularly the approaches to sampling and impact assessment. The monitoring and evaluation design team, consisting of P4P Coordination Unit and MSI personnel, worked together to further refine and finalize the monitoring and evaluation system.

During the first week of MSI's engagement, team members met with P4P and other WFP operational units and personnel. These meetings provided the MSI team members with a grounding in P4P, the perspectives of key WFP operational units and personnel on the important aspects of monitoring and evaluation for P4P, and the potential for the activities of other WFP operational units to contribute to the P4P monitoring and evaluation process.

Refining the Logical Framework

The logical framework matrix (logframe) summarizes the development hypothesis of P4P and forms the foundation of the overall monitoring and evaluation system. The draft logframe included in the proposal approved by the Bill and Melinda Gates Foundation reflects the outcome of consultation with WFP staff. The design team worked to refine the draft logical framework and indicators by broadening consultation within WFP to include country office staff and also incorporated the views of development and food assistance experts, government agencies, and other stakeholders. As it refined the logframe, the design team also sought additional input from the Bill and Melinda Gates Foundation.

The logframe specifies project objectives, the outcomes expected to contribute to achieving the objectives, the outputs that lead to outcomes, and some of the key activities that produce outputs.⁹ For each outcome and output, the logframe defines at least one quantitative indicator designed to measure progress towards achieving the outcome or output, specifies the sources of data for the indicator, and summarizes assumptions and risks that may affect the contribution of the output or outcome to the next higher level in the logframe hierarchy. Assumptions and risks represent the external factors (outside of the direct control of P4P) that must occur for the development hypothesis to function as intended and for the project to yield the expected results.

Developing Data Collection Instruments

The monitoring and evaluation system relies primarily on four sources of quantitative data – a survey of farmers' organizations, a survey of smallholder farmers, a survey of traders, and data

⁹ The logframe does not specify a complete list of project activities since these may change as the project evolves and because many of the activities will be specific to different countries' implementation plans.

transcribed from farmers' organization records. After refining the logframe, the design team turned its attention to developing the four baseline data collection instruments.

While developing drafts of the data collection instruments, the team reviewed, and borrowed freely from, existing instruments – particularly for some modules of the instrument aimed at smallholder farmers. In particular, modules on agricultural production, marketing, and income draw from data collection instruments developed by the International Food Policy Research Institute (IFPRI) and Michigan State University's Food Security Group (FSG). Survey instruments developed by WFP's Vulnerability Assessment and Mapping (VAM) unit contributed to modules on food consumption, livelihood sources, expenditure, household assets, and housing and amenities. Drawing from existing instruments helps to incorporate other organizations' extensive field experience in developing questions that effectively collect complex data. In the case of WFP instruments, it also ensures that the data are directly comparable to other WFP datasets for purposes of validating survey data or sampling methods.

WFP is implementing the P4P pilot in 21 countries on three continents. The programme therefore operates under a wide variety of agricultural environments and practices, policy settings, governments, economic influences, cultures, and languages. This, coupled with the fact that each pilot country has some latitude in how it implements P4P, makes it impossible to develop a common set of data collection instruments that will satisfy every need of all pilot countries. The data collection instruments therefore focus on the data necessary to calculate indicator values for the overall (global) monitoring and evaluation effort. Individual country offices may add to the instruments as necessary to accommodate their own data needs so long as they do not compromise the core questions.

The baseline instruments formed the basis for developing instruments for follow-up surveys and interview guides for case studies. Case studies are in-depth qualitative studies of individual farmers, farmers' organizations, and traders, focused on gaining an understanding of how and why P4P performs in different contexts. The accompanying *Purchase for Progress Monitoring Manual* contains data collection instruments for baseline and follow-up surveys and case study interview guides.

The quality of data collected through surveys depends critically on well developed and tested data collection instruments, competent and well trained enumerators and field staff, and systematic field procedures. The design team traveled to four P4P pilot countries¹⁰ to pretest data collection instruments and procedures, test enumerator training procedures, and familiarize country office staff with the monitoring and evaluation system and data collection procedures. Annex B contains additional detail about how the design team conducted the field visit activities.

Validating the Monitoring and Evaluation Approach

The design team also used the field visits as an opportunity to validate the monitoring and evaluation approach in a variety of settings. In particular, the team met with WFP country office staff to review the country's P4P implementation plan, explore the logic of the proposed intervention (e.g., how it would contribute to P4P objectives), and discuss country-specific implementation issues (e.g. how best to implement the sampling strategy). The team also met with current and potential partners and other stakeholders to gain a better understanding of the

¹⁰ Burkina Faso ,Malawi, Uganda and Zambia, .

perspectives of project partners and stakeholders, the context in which P4P is operating and the supply-side support that WFP would draw on when implementing P4P. Input from partners and stakeholders contributed to discussions with country office staff to validate the logic of P4P in general and of the country-specific implementation plans in particular.

The validation exercises changed the way the design team viewed some aspects of P4P and were crucial to adapting the monitoring and evaluation approach to be more relevant to how P4P worked in practice and to accommodate unique features of some countries' implementation plans.

The market for quality provides one example of validating the P4P approach. Prior to the field visits, the design team understood that one of the main ways that WFP expected P4P to increase smallholder farmers' incomes was by enhancing access to markets for higher quality commodities, thus allowing farmers to capture the price premium associated with quality. However, the field visits suggested that the market for higher quality commodities is often very thin – sometimes consisting only of WFP and a few other food assistance buyers. Thus, while quality improvement may provide access to additional markets for some farmers, at present the larger share of the market may not be willing to pay for quality or may not trust the marketing chain to maintain quality between the farm gate and the buyer.¹¹ This suggests that in the short-term it may make little sense to focus extensively on smallholders' capacity to produce quality commodities. Based on this experience (in multiple pretest countries), the design team shifted the focus of the monitoring and evaluation system to place more emphasis on quantity (i.e., increased production), and less on quality, as the more likely source of increased smallholder incomes.

The impact assessment implications of Zambia's implementation of P4P provide another example of information gathered during the field visits influencing the monitoring and evaluation approach. Zambia does not intend to procure directly from a set of pre-defined "participating" farmers' organizations. Instead, it will channel all of its local procurement through the nascent commodity exchange (ZAMACE). As the final chapter of this report will make clear, this implies that Zambia must employ a completely different approach to impact assessment and sampling than more "typical" P4P countries.

¹¹ The design team heard on a number of occasions that traders blended commodities during transport thus eliminating incentives for farmers to produce quality.

MONITORING AND EVALUATION OBJECTIVES & DESIGN

A monitoring and evaluation system provides a framework for assessing programme performance. Monitoring and evaluation systems typically address two aspects of performance, 1) monitoring progress towards accomplishing specified activities and producing expected outputs and 2) evaluating programme performance relative to achieving intended outcomes and impacts. To perform these functions, a monitoring and evaluation system must have clearly defined objectives and evaluation questions that address the information needs of specific audiences.

The P4P monitoring and evaluation system is a global system in that it establishes common procedures and data collection tools across the 21 P4P pilot countries. Each country may add to the system (e.g., add questions to data collection instruments) to accommodate monitoring and evaluation needs specific to their programmes but may not remove or change elements of the global system.

This chapter reviews the objectives of the P4P monitoring and evaluation system and outlines the key questions that determined the system design.

Monitoring and Evaluation Objectives

The P4P pilot focuses first and foremost on learning how WFP can best leverage its considerable presence and procurement in local food markets to support sustainable and profitable smallholder engagement in markets and ultimately to improve their welfare. To establish the conditions for learning, P4P will test a number of interventions (e.g., specific ways of buying from farmers' organizations and small to medium traders, building capacity, developing partnerships, etc.) and observe their context-specific performance relative to defined criteria. The monitoring and evaluation system is the observation tool. It defines the measurement criteria, establishes methods to collect data for measurement, and describes how to analyze the data to assess performance.

Identifying best practices (i.e., learning) depends on understanding the outcomes and impacts of P4P. Therefore, the global monitoring and evaluation system emphasizes performance evaluation over monitoring operational progress. In particular, it focuses on establishing the data and processes necessary to understand how, why, and under what conditions P4P interventions are working, or not, so that WFP can determine which approaches work best in different situations.

The scope of the monitoring and evaluation system encompasses both P4P and LRP. WFP is seeking ways in which its LRP footprint can more explicitly support agricultural and market development while delivering tangible livelihood benefits to smallholder farmers. Ultimately, WFP needs to know how it might restructure its approach to local procurement to be as smallholder friendly as possible while not unduly putting at risk WFP's core objectives. The best approach may involve a combination of current LRP practices with identified best practices from the P4P pilot.

The monitoring and evaluation system has several audiences with distinct information needs. These include:

1. **WFP** – WFP's interest in the monitoring and evaluation system is to generate the necessary evidence to identify best practices in market development and procurement, share these practices with other market development stakeholders, and employ practices to enhance its

own contribution to agricultural and market development and smallholder farmer welfare. To the extent that WFP headquarters is interested in the programme management functions of the monitoring and evaluation system, it will be from the perspective of headquarters-directed activities and outputs and will place less emphasis on country-specific activities.

2. **WFP country offices** – Country offices will place a greater emphasis on the programme management function of the monitoring and evaluation system but will also contribute to learning at the global level. To the extent that activities are country-specific, they will appear in the country-level monitoring and evaluation systems rather than in the global system. Although country offices, as a group, will contribute to the global learning objectives, their specific needs in terms of the learning aspect of the monitoring and evaluation system focus on those aspects that are relevant to their individual programme contexts.
3. **Stakeholders** – Other agricultural development stakeholders (e.g., partners, governments, etc.) are an equally important audience for monitoring and evaluation system results. Sharing results is a key objective of P4P and this means using the monitoring and evaluation system to develop compelling evidence of context-specific best practices that stakeholders will be eager to incorporate into their own operations. Engaging other stakeholders is vital to most effectively achieving long-term objectives.

Evaluation Questions

The rationale for P4P rests on the premise that: *in the long run, to efficiently buy the quantities it needs locally, WFP will buy primarily through competitive processes (e.g., competitive tenders, commodity exchanges, etc.).* However, the discussion of the second chapter suggests that smallholder farmers face significant constraints to profitably accessing competitive markets. Chief among these constraints are limited production capacity and limited market access.

Purchase for Progress also accepts the (testable) hypothesis that: *smallholder farmers will likely fare better when marketing as a group (i.e., through farmers' organizations) than when marketing individually (e.g., to traders, at the mill gate, etc.).* However, most farmers' organizations lack the capacity to effectively aggregate and market members' commodities – especially through competitive processes to demanding and quality conscious buyers such as WFP.

This reasoning sets up the first of two overarching evaluation question for P4P:

- What procurement modalities/platforms and practices best support capacity building (of both smallholder farmers and farmers' organizations) and create an enabling environment for sustainable and profitable smallholder engagement in staple food markets?

The previous discussion suggests that the relevant criteria for assessing performance relative to this question include smallholder farmers' production capacity, farmers' organizations' marketing capacity, and measures of smallholder farmers' welfare.

Once WFP identifies best practices for developing markets and building capacity to access markets, it must then determine how best to structure its own future procurement activities. This raises the second primary evaluation question:

- How can WFP optimize its local procurement to achieve the dual objectives of maximizing benefits to smallholder farmers while providing safe food in a timely and efficient manner?

Procurement options include variations on P4P modalities as well as the LRP model through which WFP currently procures. To select the optimal approach, WFP needs to consider the impact of its procurement on smallholder farmers’ production, livelihoods, and marketing capacities under both P4P and LRP and also take into account the timeliness of deliveries, the ability of smallholder farmers to meet the quality demands of WFP, and the overall cost efficiency of pro-smallholder purchases.

Figure 1 summarizes the two key evaluation questions mapped onto the general (outcome level) criteria used to assess the questions.

Figure 1. P4P Monitoring and Evaluation Questions

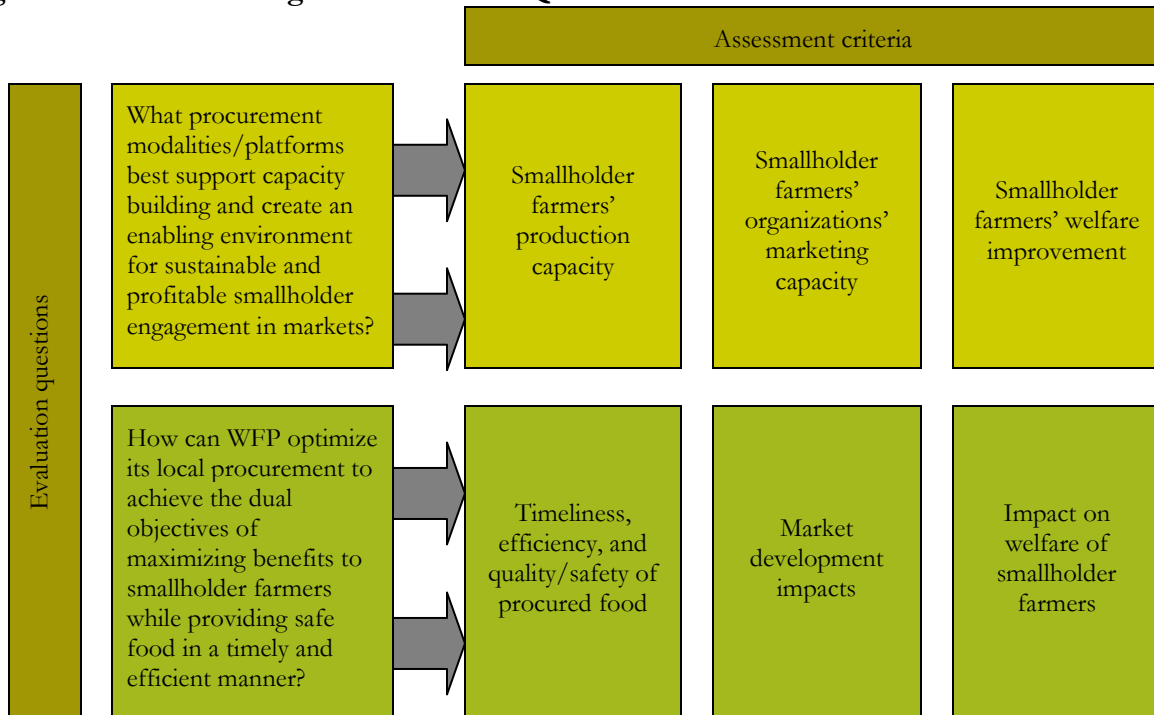


Figure 1 provides a high-level perspective of the key evaluation questions of P4P. The assessment criteria are broad statements that are not operationally useful (i.e., are not measurable). Figure 2 defines the measurable concepts associated with each of the assessment criteria. These are still not operationally precise indicators but they do illustrate the types of data and analyses anticipated to address the key evaluation questions. The next chapter develops specific quantitative indicators for measuring performance relative to these criteria.

Figure 2 implies that the quantitative assessment criteria alone are sufficient to address the key evaluation questions. However, a much richer and more nuanced set of questions is necessary to support the key context-specific learning objectives of P4P. To a very large extent, augmenting the quantitative data with qualitative data will be vital to answering these questions and fully informing the learning process. Figure 3 contains some illustrative detailed evaluation questions that the monitoring and evaluation system will address.

Figure 2. Assessment Criteria

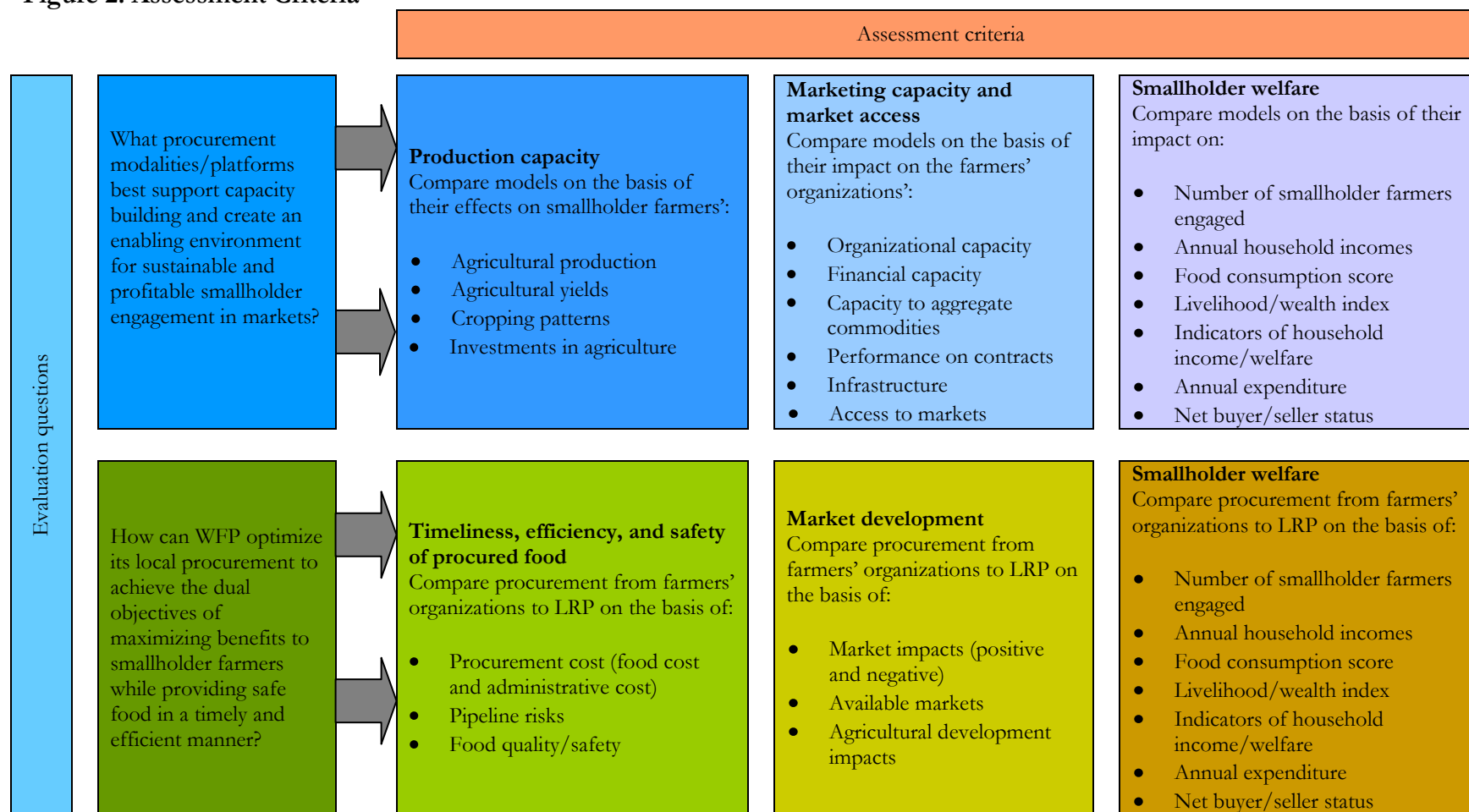


Figure 3. Illustrative Detailed Evaluation Questions

- 1. Has P4P affected agricultural practices and production levels of participating smallholder/low income farmers (male and female)?**
 - 1.1. If yes, how did it do so? Which interventions were most effective and under what conditions?
 - 1.2. If no, why not? What constraints still exist and how can WFP address these constraints?
- 2. Has P4P affected access to markets for participating farmers' organizations (male and female) and increased their capacity to profitably engage in these markets?**
 - 2.1. If so, how? Which interventions were most effective and under what conditions?
 - 2.2. If no, why not? What constraints still exist and how can WFP address these constraints?
- 3. Has P4P affected the lives of participating smallholder farmers and traders (male and female)?**
 - 3.1. If yes, how? How did farmers respond to the incentives provided by P4P? Are impacts sustainable?
 - 3.2. If no, why not? Was the development hypothesis incorrect? What additional constraints exist and how can WFP address these constraints?
- 4. How well has P4P engaged smallholder farmers (male and female) in WFP procurement? Which methods of procurement best engage smallholder farmers and under what conditions?**
- 5. Has P4P had adverse effects on markets (prices and price stability)?**
 - 5.1. If yes, why and under what conditions? How can WFP structure its local procurement to minimize or prevent adverse market impacts?
 - 5.2. If not, what procurement management tools mitigated potential adverse market impacts?
- 6. Has WFP employed appropriate mechanisms to identify best practices and shared those practices effectively with others?**
 - 6.1. If yes, what methods were particularly effective and under what conditions?
 - 6.2. If not, why not? How could WFP have enhanced learning and sharing?
- 7. How do the full costs of procuring locally (e.g., transaction costs, risks, etc.) compare to the full benefits (e.g., development impacts, etc.)? Under what conditions is it best to procure locally and what methods of local procurement work best in different situations?**

MONITORING AND EVALUATION SYSTEM OVERVIEW

The previous chapters have developed the context for P4P, defined the objectives of P4P and of the P4P monitoring and evaluation system, and posed the key evaluation questions for P4P in very broad terms. This chapter provides a basic overview of the P4P monitoring and evaluation system. It focuses on broad themes, basic procedures, and how the different components of the system contribute to monitoring and evaluation objectives.

The chapter first describes the basic monitoring and evaluation functions that guided the system design. It then describes the programme's development hypothesis (presented in a logical framework) with specific measurable indicators of performance. The final two sections review data collection procedures and outline a plan for data analysis. The plan for data analysis specifically discusses how the monitoring and evaluation system will combine qualitative and quantitative data to more completely assess the performance of P4P and to facilitate learning. The following chapter provides much of the detailed rationale for the design of the system. Separate baseline and monitoring manuals¹² provide the detailed guidance country offices require to implement the system at the country level.

Monitoring and Evaluation System Functions

A monitoring and evaluation system is a framework for monitoring programme processes and evaluating performance. The P4P monitoring and evaluation system serves three primary functions – monitoring the operational progress of the programme, testing/validating models of pro-smallholder agriculture and market development, and evaluating programme performance.

- Monitoring is the ongoing process of assessing whether the programme is performing specified activities and producing expected outputs and outcomes. Monitoring contributes primarily to the day-to-day management of the programme. The monitoring function, at least as it relates to objective 2, is somewhat more relevant to the individual P4P pilot countries where actual implementation occurs than it is to headquarters. Monitoring of objectives 1, 3, and 4 will be more relevant to headquarters than to the pilot countries.
- Evaluation establishes whether the programme is contributing to anticipated outcomes and impacts (e.g., increasing agricultural production, access to markets, smallholder incomes, etc.). Both individual pilot countries and headquarters will benefit from evaluation but it will contribute most to the global (i.e., headquarters) monitoring and evaluation objectives.

¹² The baseline manuals consist of the *Purchase for Progress Monitoring & Evaluation Baseline Sampling Strategy*, the *Purchase for Progress Data Collection Manual*, and the *Purchase for Progress Enumerator Training Manual*. The monitoring manual is the *Purchase for Progress Monitoring Manual*.

Logical Framework

The logical framework approach is a detailed method for assessing programme performance. The logical framework matrix, or logframe matrix, concisely articulates the development hypothesis of P4P as a hierarchy of impacts, outcomes, outputs, and activities where activities (actions) produce outputs (direct results of the actions) which influence outcomes (conditions affected by the outputs) which, in turn, contribute to impacts (ultimate development objectives to which P4P contributes). The logframe matrix also specifies measurable indicators to determine if outputs are accomplished and the extent to which P4P contributes to impacts and outcomes; sources of data for the indicators; and risks and assumptions that may affect programme performance.¹³

Indicator Characteristics

Specific – A specific indicator is precisely stated and unidimensional so there is no ambiguity about what it measures.

Measurable – Indicators should be measurable (preferably in quantitative terms) whenever possible to facilitate interpretation and enhance precision.

Achievable – The indicator should be realistically achievable within the stated timeframe.

Relevant – A performance indicator should measure as closely as possible the result it is intended to measure.

Time-bound – The indicator should refer to a specific period of time in which the result will be achieved.

Ultimately, the performance of P4P depends on the extent to which it accomplishes the four programme objectives. However the P4P objectives statements are multifaceted and too broadly stated to serve as measurable gauges of performance. The monitoring and evaluation design team, with input from personnel in other WFP operational units and other internal and external stakeholders, dissected the objectives into more specific impact and outcome statements. The team then developed outputs and associated activities related to each outcome – again drawing on WFP personnel familiar with the development of P4P, P4P documentation, and input from stakeholders.

With a succinct statement of the development logic of P4P, the team then developed measurable, quantitative indicators for each impact, outcome, and output; defined the sources of information for each indicator; and summarized the risks and assumptions associated with outcomes and outputs. The resulting global logframe matrix (contained in Annex C) serves as the foundation for the P4P monitoring and evaluation system.

Because individual country offices will implement P4P in slightly different ways, the P4P Coordination Unit encourage each country to develop a country-specific logframe by drawing from and/or refining the global logframe. However, all country logframes will retain the core set of indicators necessary to support monitoring and evaluation at the programme level and at minimum will administer the standardized data collection instruments.

Figure 4 summarizes the outcomes extracted from the full P4P logframe matrix along with their associated indicators and the rationale for the indicators. Objective 1 emphasizes the core rationale for P4P – learning how best to structure local procurement and sharing lessons with other

¹³ Many manuals describe the logical framework in greater detail. Two of particular relevance are (World Food Programme: Office of Evaluation and Monitoring) and (Swedish International Development Cooperation Agency (Sida) 2004).

agricultural development stakeholders. Outcomes 1.1 and 1.2 describe the conditions WFP expects to prevail if P4P is successful in achieving objective 1. Objective 2 encompasses most of the actual implementation of P4P at the country office level. It addresses increasing the production capacity of smallholder farmers (Outcome 2.1), enhancing the marketing capacity of farmers' organizations (Outcome 2.2), market access (Outcome 2.3), and smallholder farmer welfare (Outcome 2.4). The large number of indicators for Outcome 2.4 captures different aspects of welfare and also provides some insurance if measures of income turn out to be unreliable. Objective 3 focuses on WFP's local procurement. WFP's procurement database, PASTFood, and evaluation of country office assistance programmes will provide data on procurement statistics and the timeliness, efficiency, and safety of procured food. Objective 4 captures the long-term transformation of WFP's local procurement.

The assessment criteria of Figure 2 specify comparisons. The first question tests alternative procurement modalities in different contexts to draw out lessons and identify best practices. The second question compares the relative performance of P4P and LRP in different contexts in order to craft a smallholder friendly procurement strategy. In terms of the quantitative data, these will generally entail comparison of indicator values. For many of the impact and outcome indicators, the comparison is a measure of programme impact. The "Data Analysis" section describes how to calculate impacts.

The logical framework largely reflects global objectives, outcomes, outputs, and activities. Because P4P is the sum of the experience of each pilot country, each country must report against the indicators in the logframe. The global logframe provides all countries with a common monitoring and evaluation framework and ensures consistent measurement of global indicators across countries. However, since each country will implement P4P in a slightly different way contingent on country-specific constraints and opportunities, they will each augment the logframe to incorporate their own specific activities, outputs, outcomes, and objectives.

Figure 4. P4P Outcomes and Indicators

Objective 1: To identify and share best practices for WFP, NGOs, governments, and agricultural market stakeholders to increase profitable smallholder/low income farmer engagement in markets.		
Outcomes	Indicators	Indicator rationale
Outcome 1.1: WFP and other agricultural development stakeholders collaborate to identify procurement and market development best practices from P4P experience	Number of completed compilations (by WFP) of best practice programming and policy recommendations from P4P	Periodic reports summarizing the key best practice programming and policy recommendations are a fundamental part of documenting the lessons derived from P4P and sharing those lessons widely.
Outcome 1.2: By the end of the project, agricultural development stakeholders (e.g., governments, NGOs, private sector, donors, etc.) have integrated smallholder/low income farmer-focused market development and procurement best practices into their operations, procedures, or policies	Number of participating stakeholders that have incorporated best procurement and market development recommendations into their operations, procedures, or policy documents	The wider (beyond WFP) success of P4P depends on other agricultural development stakeholders adopting the resulting best procurement and market development recommendations and incorporating them into their operations, procedures, and policies.

Figure 4 (continued)

Objective 2. To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.		
Outcomes	Indicators	Indicator rationale
Outcome 2.1: By the end of the project, participating smallholder/low income farmers have increased their marketable surpluses of staple commodities.	Average per farm marketable surplus of staple commodities produced by smallholder members of participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity and sex of farmer)	Increased production of surpluses (i.e., above household need) staple commodities is one of the key pathways to increased incomes for smallholder farmers under P4P. It is critical that the M&E system determine whether P4P is affecting production or not.
	Average per farm quantity of staple commodities sold by participating smallholder/low income farmers (relative to baseline and comparison group, disaggregated by sex of household head)	To increase their incomes, smallholder farmer not only have to produce greater surpluses, they also have to sell them. This indicator measures smallholder farmers' total sales (to all sources) of staple commodities. It does not attempt to determine whether farmers are selling from what they need for household consumption.
	Average (per smallholder farm) post-harvest losses of staple commodities as a percentage of annual production (relative to baseline and comparison group, disaggregated by commodity and sex of household head)	Post-harvest losses are high in many countries in which WFP buys. High post-harvest losses threaten households' food security and also limit the quantity available for sale, thereby reducing potential household income.

Figure 4 (continued)

Objective 2. To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.		
Outcomes	Indicators	Indicator rationale
Outcome 2.2: By the end of the project, participating smallholder/low income farmers' organizations have increased their capacity to aggregate and market their smallholder members' marketable surpluses of staple commodities	Average proportion of smallholder members' staple commodities sold through participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity and sex of registered farmers' organization member)	P4P aims to improve the capacity of farmers' organizations to market their members' staple commodities. Improved marketing capacity should translate into higher prices for many organizations members relative to what they could earn selling independently to traders. If P4P is successful, member farmers may choose to increase the proportion of their commodities they sell through their farmers' organization relative to the proportion they sell through other channels.
	Average (per registered member) quantity of staple commodities sold through participating farmers' organizations (relative to baseline and comparison group, disaggregated by sex of registered farmers' organization member)	One indicator of a farmers' organization's capacity to market its members' commodities is the quantity of staple commodities it markets. An organization that is increasing its capacity to market should be able to collect and sell more. Because this indicator measures the marketing capacity of the farmers' organization it focuses on all farmers' organization members, not only smallholders.
	Average (over participating smallholder farmers' organizations) price received for commodities as a percentage of the highest price in that locality during the marketing season	One desired outcome of increasing the capacity of farmers' organizations to market their members' commodities is that the organization receives a better price. This may occur because the organization develops more markets, improves the quality of the commodities it sells, or has the flexibility to hold commodities for sale at higher priced seasons (i.e., the lean season). Comparing the average price (weighted by quantity) received by the farmers' organization to the highest price recorded in the locality during the marketing season captures the farmers' organization's ability to capture a higher price for any of these reasons.

Figure 4 (continued)

Objective 2. To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.		
Outcomes	Indicators	Indicator rationale
Outcome 2.3: By the end of the project, participating smallholder/low income farmers' organizations have increased access to markets for staple commodities	Average quantity of staple commodities sold by participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity)	The quantity of staple commodities sold by a farmers' organization is a key indicator of its marketing capacity. Growth in the indicator for participating farmers' organizations (relative to the baseline and comparison group) indicates an increase in marketing capacity.
	Average size of sale of staple commodities by participating smallholder/low income farmers' organization (relative to baseline and comparison group, disaggregated by commodity)	As farmers' organizations increase their capacity to aggregate and market their members' staple commodities and as members increase their production, the farmers' organizations should be able to provide larger quantities and participate in larger tenders. This indicator measures one aspect of farmers' organizations' organizational and marketing capacity.
	Average number of different geographic markets sold into by participating smallholder/low income farmers' organizations (relative to baseline and comparison group)	The geographic marketing reach of a farmers' organization is a measure of its capacity to reach more markets for its members' commodities.

Figure 4 (continued)

Objective 2. To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.		
Outcomes	Indicators	Indicator rationale
Outcome 2.4: By the end of the project, the sale of staple commodities is contributing to improved welfare for households of participating smallholder/low income farmers	Average percentage contribution of sale of staple commodities to household incomes of participating smallholder/low income farmers (relative to baseline and comparison groups, disaggregated by sex of household head)	One way that P4P expects to increase household income is through increased production and sale of staple commodities. This indicator measures the extent to which staple commodities contribute to net household income.
	Average food consumption score of participating smallholder/low income farmer households (relative to baseline and comparison groups, disaggregated by sex of farmers' organization member)	P4P aims to improve the well-being of smallholder farmer households. The food consumption score is one measure of well-being. Given the potential problems with measuring income change, this indicator also serves as a proxy for income because the quantity and variety of foods a household consumers should increase with increasing incomes. Other WFP surveys collect the FCS. This indicator may therefore provide an opportunity to validate results against data from other surveys.
	Average household asset score (HAS) of participating smallholder/low income farmer households (relative to baseline and comparison groups, disaggregated by country and sex of farmers' organization member)	The household asset score is a measure of the household's ownership of assets and serves as a measure of wealth.
	Average annual household expenditure (food and non-food) by smallholder farmer households (relative to baseline and comparison groups, disaggregated by sex of household head)	P4P aims to increase the incomes of smallholder farmer households. Income is difficult to measure accurately for smallholder farmers. Annual expenditure is a proxy for income because expenditure should rise with increasing incomes.
	Percentage of participating smallholder/low income farmers who are net sellers of staple commodities (i.e., produce more than they consume) (relative to baseline and comparison group, disaggregated by commodity)	This indicator provides another perspective on production of marketable surpluses. Instead of measuring the average size of marketable surpluses (as does another indicator) it measures the percentage of smallholder households who produce more than they consume.

Figure 4 (continued)

Objective 3. To identify and implement best practices for increasing sales to WFP and others with a particular focus on smallholder/low income farmers.		
Outcomes	Indicators	Indicator rationale
Outcome 3.1: The quantity of WFP's purchases from smallholder/low income farmer associations increases by 30% annually throughout the five-year P4P pilot phase	Quantity of food purchased annually by WFP from smallholder/low income farmers' organizations (disaggregated by commodity, procurement modality, and country)	This indicator will facilitate tracking of WFP local procurement through P4P and LRP. It will contribute to tracking the percentage of procurement by modality within P4P, the growth of P4P procurement (as specified in the Bill and Melinda Gates Foundation proposal), and the proportion of local procurement through P4P.

Figure 4 (continued)

Objective 4. To transform WFP food purchase programmes so they better support sustainable small-scale production and address the root causes of hunger.		
Outcomes	Indicators	Indicator rationale
Outcome 4.1: By 2013, WFP has transformed its programming, policies, rules, and regulations to incorporate a strategic focus on local procurement with a focus on smallholder/low income farmers	Financial regulations and procedures revised to incorporate pro-smallholder procurement	WFP's current payment mechanisms are not particularly suited to accommodating the unique needs of farmers' organizations. Revision of financial regulations and procedures to better accommodate the constraints of farmers' organizations will enhance the ability of P4P to meet its objectives.
	Job descriptions reflect needs/skills required to effectively manage pro-smallholder local procurement.	WFP job descriptions need to clearly spell out the required competencies, skills and tasks of staff engaged in pro-smallholder local procurement.
	Program guidance manual revised to reflect a strategic approach to Local Procurement.	Pro-smallholder local procurement may have advantages (e.g., cost, timeliness of delivery, suitable to local tastes) in meeting WFP's needs but it also has potential disadvantages (e.g., procurement cost). Country Offices are more likely to use local procurement and target it effectively to smallholder farmers if they recognize that it has development objectives and explicitly consider these objectives when deciding how and when to use the approach.
	Number of P4P pilot countries in which risk management strategies explicitly acknowledge risks associated with pro-smallholder procurement	Pro-smallholder local procurement potentially introduces unique risks into WFP's procurement process (e.g., market distortion, pipeline risks, quality risks, etc.). Building pro-smallholder local procurement into country-level risk management strategies is an indicator of explicit consideration of these risks and thus of institutionalizing pro-smallholder local procurement models within WFP. The indicator assumes that P4P modalities are an effective way to procure locally.
	Percentage of PRROs, EMOPs, and country programmes that incorporate pro-smallholder local procurement as a programme component (disaggregated by country)	Explicitly including P4P modalities among its strategies to procure food is an indicator that WFP is incorporating P4P into its operations. The indicator assumes that pro-smallholder local procurement modalities are an effective way to procure locally.

Data Collection

The P4P monitoring and evaluation system relies on a number of sources of quantitative and qualitative data. This section describes the primary sources of data, collection procedures for each type of data, how each type of data contributes to the monitoring and evaluation system, and a schedule for data collection. The separate *Purchase for Progress Data Collection Manual* provides the detailed instruction necessary to prepare for and conduct data collection activities in the field.

Data Types, Sources, and Collection Procedures

The monitoring and evaluation system collects qualitative and quantitative data from a variety of primary and secondary sources. The main primary data sources include:

- **Surveys of smallholder farmers** –A household panel data set collected through surveys of smallholder farmers in years one, three, and five supplies most of the quantitative data on smallholder farmers’ agricultural practices, production, and marketing; food security; housing and household characteristics; livelihood sources; and welfare. Countries selected for impact assessment will administer the survey to random samples of participating farmers and similar non-participating farmers to facilitate impact assessment. All other pilot countries will collect data only from participating farmers. The series of support manuals developed for P4P fully describe data collection procedures.¹⁴ Figure 5 illustrates the general content of the *Farmer Livelihood and Agricultural Production* data collection instrument.

Figure 5. Elements of Farmer Data Collection Instrument

Module A – Farmers’ organizations Membership and role in organizations, services provided
Module B – Agricultural production/practices Crops planted/harvested/sold, value of production, production expenses, agricultural practices
Module C – Food security Household food consumption/diversity, net buyer/seller status
Module D – Livestock Household income and expenses related to livestock
Module E – Livelihood activities/other income Other sources of household income

¹⁴ Relevant manuals include the *Purchase for Progress Monitoring & Evaluation Baseline Sampling Strategy*, the *Purchase for Progress Data Collection Manual*, the *Purchase for Progress Enumerator Training Manual*, and the *Purchase for Progress Monitoring Manual*.

Module F – Borrowing and access to credit Borrowing/credit history
Module G – Expenditure Annual food and non-food expenditure
Module H – Household assets Assets ownership
Module I – Housing and amenities Housing materials, facilities
Module J – Household information Characteristics of household members

- Surveys of farmers’ organizations** – An annual panel data set collected through surveys of farmers’ organizations provides the primary quantitative data on farmers’ organizations’ structure and membership; facilities and services provided to members; and marketing activities. Countries selected for impact assessment will administer the survey to all farmers’ organizations that are participating in P4P and to a sample of similar organizations that are not participating in P4P. Non-impact assessment countries will collect data only from participating organizations. Figure 6 illustrates the general content of the farmers’ organization data collection instrument.

Figure 6. Elements of Farmers’ Organization Data Collection Instrument

Module A – Organization structure & membership Membership, organizational structure & capacity, staffing
Module B – Facilities & services Organization facilities, services
Module C – Marketing Aggregation and sale details (how, what, to whom, quality, challenges, contracting/tendering, etc.

- Surveys of traders** – Surveys of commodity traders in years one, three, and five represent the main sources of quantitative information about traders’ activities; capacities; purchases and sales; suppliers; buyers; and marketing. It is not possible to construct a comprehensive sampling frame for traders given the wide range of both transient and permanent traders that operate in any given country context. Therefore, country offices will interview a representative, rather than a random, sample of traders. However, a random sample is not

necessary for monitoring and evaluation purposes because there is no need to extrapolate from the sample to the population of traders. Figure 7 illustrates the general content of the trader data collection instrument.

Figure 7. Elements of Trader Data Collection Instrument

<p>Module A – Business information Owner and business characteristics</p>
<p>Module B – Employees Questions for employees</p>
<p>Module C – Trader capacity/profile Assets, business practices, facilities, capacity</p>
<p>Module D – Purchases and sales Quantities bought and sold</p>
<p>Module E – Suppliers Sources of commodities</p>
<p>Module F – Buyers Buyers of commodities</p>
<p>Module G – Marketing Marketing behavior and challenges</p>

- **Farmers’ organization records** – Review of farmers’ organization records provides additional quantitative data on farmers’ organizations’ membership and sales.¹⁵ Country offices will collect data from all participating farmers’ organization records on a regular basis to document organizations’ receipts and sales of commodities. Some farmers’ organizations do not as yet keep the records required for purposes of monitoring and evaluation. Country offices will endeavor to build this capacity in farmers’ organizations as P4P progresses. The *Purchase for Progress Monitoring Manual* contains complete details on collecting data from farmers’ organization records.
- **Case studies of smallholder farmers, farmers’ organizations, and traders** – In-depth case studies will follow a relatively small number of participating smallholder farmers and farmers’ organizations and a selection of traders throughout the P4P pilot. The case studies

¹⁵ The rationale for collecting these data from records is to save time during the farmers’ organization interviews and also to improve the accuracy of the data.

will provide qualitative evidence about how P4P affects these actors. The qualitative data will augment the quantitative data to provide insight into how and why P4P is working, or not working. The *Purchase for Progress Monitoring Manual* contains comprehensive guidance on how to conduct case studies.

- **Monitoring procurement** – Ongoing monitoring of WFP’s local procurement will produce detailed data about all aspects of local procurement (i.e., tender participants, winners, and performance; purchases from farmers’ organizations; procurement modalities; quantities; prices; quality; etc.). These data will contribute to tracking smallholder participation, the proportion of local procurement conducted through P4P, farmers’ organizations’ performance on contracts and other relevant aspects of procurement. The Food Procurement and P4P units in Rome are developing queries to extract procurement data from the PASTFood database. The *Purchase for Progress Monitoring Manual* provides details on procurement monitoring.
- **Feedback/input from country offices, stakeholders, and other experts** – Those with hands-on experience implementing, or engaging with, P4P are in the best position to fully understand how P4P is, or is not, working and why. Purchase for Progress incorporates many opportunities for country office staff, stakeholders, partners, WFP headquarters staff, and other experts to provide input on P4P’s performance. Feedback from these sources is a crucial element in the learning component of P4P and it is critical that WFP capture and compile these data. The P4P Coordination Unit will regularly compile qualitative feedback from country office quarterly reports, country-level stakeholder meetings, headquarters-level steering committee meetings, annual meetings of a technical review panel, and an annual review workshop that will bring together WFP and partner staff from around the globe.

Purchase for Progress will also draw from secondary data sources whenever possible to support the monitoring and evaluation system. Secondary data sources will vary by country but the primary sources WFP expects to use include:

- **Market data** – Understanding how P4P is affecting markets and how the price that smallholder farmers receive for their commodities under P4P compares to prevailing market prices requires accurate data on local prices. When possible, WFP will rely on secondary sources for these data. But, when a country does not have reliable price information, WFP may have to collect these data itself.
- **Agricultural statistics** – National and regional agricultural statistics will contribute tremendously to understanding the context within which P4P operates and to designing, managing, and monitoring P4P. WFP will compile relevant agricultural data from existing sources when it is available.

Data Collection Schedule

Figure 8 illustrates the schedule for collecting monitoring and evaluation data over the course of the five-year P4P pilot. The figure’s rows represent data collection activities rather than types of data. There are eight different data collection activities: 1) baselines, 2) P4P monitoring, 3) LRP monitoring, 4) procurement monitoring, 5) market monitoring, 6) case studies, 7) qualitative data, and 8) evaluations.

The primary data collection effort begins in year one with baseline surveys of smallholder farmers, farmers' organizations, and traders. The baselines establish the starting point, i.e., conditions prior to the start of P4P, against which to assess change. Country offices will also begin procurement monitoring, market monitoring, case studies, and qualitative data collection in the first year.

In years three and five, WFP will follow up on the baselines by re-surveying the smallholder farmers, farmers' organizations, and traders who responded to the baseline. These follow-up surveys will provide the data necessary to monitor changes in conditions and indicators that have occurred since the baseline. The resulting panel datasets will provide rich information about the dynamic responses of individuals, organizations, and businesses to P4P. The data will contribute to monitoring and to impact assessment.

Also during years two through five, WFP will continue monitoring procurement and markets, conducting case studies, and collecting qualitative data. In years three and five, WFP will commission external mid-term and final evaluations. The data collected during normal monitoring and evaluation should be sufficient to support the evaluations. However, an independent evaluator may decide to collect additional data.

Answering the second of the two key evaluation questions requires comparing P4P with LRP on the basis of procurement cost, pipeline risks, food quality, market impacts, market development impacts, agricultural development impacts, the number of smallholder farmers engaged, and smallholder household welfare measures. To support this comparison, WFP must monitor LRP procurement as well as P4P. Because the traders from whom WFP buys through LRP do not typically keep records of the smallholder farmers from whom they buy, monitoring the impacts of LRP on smallholder farmers is difficult. WFP, with input from the Technical Review Panel, is in the process of finalizing the approach to monitoring LRP's impact on smallholder farmers.

Figure 8. Data Collection Schedule

	Year 1	Year 2	Year 3	Year 4	Year 5
Baseline (survey data from farmers, farmers' organizations, and traders)					
Follow-up surveys (survey data from farmers, farmers' organizations, and traders)					
LRP monitoring (survey data from farmers, farmers' organizations, and traders)					
Procurement monitoring (data from WFP procurement database – PASTFood and from quarterly reports)					
Market monitoring (data from secondary sources or from WFP country offices)					
Case Studies (qualitative data from farmers, farmers' organizations, and traders)					
Qualitative Data (quarterly reports; stakeholders, steering committee, and experts)					
Evaluations (survey data from farmers, farmers' organizations, and traders)			Mid-term		Final

Data Analysis

The monitoring and evaluation system collects a tremendous amount of quantitative and qualitative data from surveys, procurement records, quarterly reports, market monitoring, and other sources. These data have little utility without a clear plan for how they will be used to answer the questions of interest to WFP and other audiences. The quantitative and qualitative data serves several purposes. At the most basic level, much of the quantitative data feeds directly into calculating values for logframe indicators. It also contributes to developing profiles of programme participants and non-participants. The profiles will help analysts compare respondents who are participating in P4P to those who are not participating and thus assess the similarity of the comparison groups required for impact assessment. They will also contribute to comparing survey respondents to other populations to determine how well respondents represent the intended populations (e.g., all farmers' organizations, smallholder farmers, etc.). They will also contribute to analyses of how P4P is working.

The qualitative data will help enhance understanding of the how and why of P4P performance, or lack thereof. This section outlines a plan for analyzing data that includes discussions of calculating and using indicator values and organizing and incorporating qualitative data.

Calculating Indicator Values

Calculating values of the logframe indicators represents the most basic level of data analysis. Changes in indicator values over time represent the fundamental quantitative measures of programme performance. Output level indicators monitor the degree to which activities are executed. Outcome and impact level indicators measure changes in the conditions that WFP expects P4P to influence. In general, outcome level indicators measure gradual change in conditions measured by trends in indicator values over time. Output indicators, on the other hand, are more likely discrete measures of whether an activity has produced the desired output or not.

The baseline data collection effort provides the data to establish the initial (i.e., pre-P4P) values for logframe indicators. For impact assessment countries and for the indicators that rely on data from smallholder farmers and farmers' organizations, the baseline will establish initial indicator values for both participating and non-participating farmers and organizations. The biennial (every two years) surveys provide the data to update the indicator values as the programme progresses.

The logframe matrix specifies many of the outcome and impact indicators as impacts. Impacts are the change in the indicator that is *attributable to P4P*. Establishing attribution requires a comparison of current indicator values to baseline values for both participating and non-participating groups. Specifically, the impact of P4P on indicator i is:

$$impact = (v_{pc} - v_{ipb}) - (v_{inc} - v_{inb})$$

where v_i represents the value of indicator i , the subscript p denotes participating groups, the subscript n denotes non-participating groups, the subscript c denotes current values, and the subscript b denotes baseline values.

The "Indicator Reference" defines each indicator precisely, provides formulas for calculating indicator values, describes how to calculate impacts for those indicators that require it, and explains how to discount monetary indicator values when necessary.

The indicators contribute directly to understanding the evaluation questions in Figure 1.

Incorporating Qualitative Data

Qualitative data from case studies, country office quarterly reports, and notes or minutes of stakeholder and steering committee meetings will contribute immensely to learning. The quantitative indicators illustrate *what* is happening but the qualitative data is necessary to interpret the quantitative indicators and develop an understanding of *why* and *how* things happen. The qualitative data will be crucial to understanding why a specific action produces a particular result in a given situation. The questions of why and how expected and unexpected results occur is crucial to course correction and learning.

Much of the qualitative data is narrative. For instance, transcribed comments from case studies, written material from quarterly reports, and meeting minutes are all narrative. The narrative format of the qualitative data makes it much more difficult than quantitative data to compile, analyze, and interpret. Analysts must first determine how to efficiently transfer qualitative narrative from reports, notes, meeting minutes, and other documents and media into a common format and location for access and analysis. Once the data reside in one place in a common format, analysts must summarize it and organize it in a manner that facilitates drawing lessons from the amassed body of qualitative evidence. Both of these pose significant challenges.

The P4P Coordination Unit in Rome is exploring alternative procedures for efficiently transferring data from reports and other narrative sources. One approach under consideration is to develop a web-based system for country office reporting. The system would automatically organize responses into a common database. Another possible approach is to develop reporting forms (e.g., in Adobe Acrobat or Microsoft Word) that would automatically link responses in form fields to a common database. Either of these approaches would address the problem of physically transferring narrative data into a database. WFP is still weighing the relative merits of the two approaches.

Summarizing the narrative data and compiling and organizing it in a format that facilitates learning and sharing will also be a challenge. A web-based system may require those who generate the data (i.e., narratives) to categorize it under pre-defined key words or topics. But, however the data come into the system, an analyst (probably the P4P reports officer) will have to edit, synthesize, and organize the data around key words and concepts.

The most fertile ground for learning in P4P will likely be forums where WFP and other stakeholders and experts meet and discuss their experiences implementing, observing, and interacting with P4P. Purchase for Progress provides for several such forums including bimonthly (i.e., every other month) country-level stakeholder meetings and monthly headquarters-level steering committee and stakeholder group meetings. The country offices and the P4P Coordination Unit will enter the notes and minutes from these meetings into the database described above.

The geographic extent of P4P across 21 countries and three continents limits the opportunities for face-to-face learning and sharing across country P4P pilots. However, this level of sharing is crucial to learning about the context-specific performance of P4P and to validating best practices. Purchase for Progress will seek to build on WFP's in-house communication platform, the Practical Advice Sharing System, or PASS-it-on, to facilitate cross-country sharing of lessons learned, best practices, and experiences. PASS-it-on is a collection of knowledge sharing tools that can be disseminated internally via the PASS-it-on website. The tools include:

- **After Action Reviews (AARs)** – After Action Reviews are short one-day workshops that focus on lessons learned. Country offices may organize AARs within the P4P Coordination Unit, within the WFP country office, or among stakeholders. The AARs provide a forum for sharing experiences and drawing lessons from those experiences. Country offices will summarize lessons in short reports and post the reports on the PASS-it-on website.
- **“Let’s Talk” Sheets** – Let’s Talk Sheets are an opportunity for each country office to describe its experiences with different elements of P4P. Periodically, the P4P Coordination Unit will ask country offices to describe their experience with a particular aspect of P4P. Responses will contribute to understanding alternative approaches for dealing with the context-specific challenges each country faces implementing P4P. The P4P Coordination Unit may incorporate the Let’s Talk Sheet into the regular quarterly reporting framework by asking country offices to respond to a special question in each report.
- **Cross-site visits** – Cross-site visits provide an opportunity for stakeholders from one country to observe how another country has addressed a common challenge. Cross-site visits can help educate and motivate existing stakeholders. They may also serve to engage new stakeholders by demonstrating the successes of P4P.
- **“How We Did it” case studies** – Case studies are in-depth descriptions of how a country office addressed a particular problem. The narrative reports should provide sufficient detail that another country office, stakeholder, or government facing the same challenge can determine what elements of the problem match their situation and adapt the solution to their context.

The P4P Coordination Unit in Rome will draw the narrative reports generated by these tools into the database described earlier.

Disaggregating by Sex

Although the discussion of general data analysis procedures does not specifically mention sex, the differential impacts of P4P by sex are a critical element of P4P and, consequently, of the monitoring and evaluation system. Whenever appropriate and meaningful, the data analysis will disaggregate impacts and programme processes by sex. This will help WFP understand the sex dimension of P4P; assess whether P4P is impacting smallholder farmers, farmers’ organizations, and traders equitably with regard to sex; allow the programme to address inequities; and facilitate a sex sensitive and sex balanced approach to pro-smallholder local procurement.

DESIGN CONSIDERATIONS

This chapter catalogues a number of issues that arose during the course of designing and validating the P4P monitoring and evaluation system. The chapter describes each issue, the challenges it posed, and the approach the design team adopted to address the issue. The recitation of issues and solutions documents the design team's thinking and provides insight into the implementation of the monitoring and evaluation system. It should also serve as a guide to implementers as issues arise in the future.

Global Versus Country-Level Monitoring and Evaluation

Purchase for Progress is a global programme with global objectives. The P4P Coordination Unit housed in WFP's Rome headquarters manages the overall programme with the aim of facilitating context-specific learning about market development and procurement best practices. The global perspective, and global management of the monitoring and evaluation system, is crucial to integrating the experiences of each of the 21 pilot countries into context-specific learning.

Even though P4P is a global programme, 21 individual countries share responsibility for field-level implementation (i.e., defining country-level objectives, designing country-specific interventions, forming partnerships, procuring, working with stakeholders, policy advocacy, etc.). It is crucial that each country contribute to the common global monitoring and evaluation system because the global indicators, especially for objective 2, represent the sum of experience over all pilot countries. The common approach to monitoring and evaluation embodied in the global logframe is essential to effective learning from the P4P pilot.

The global monitoring and evaluation system thus represents a joint effort of the P4P Coordination Unit in Rome and the individual country pilots. The P4P Coordination Unit bears primary responsibility for overall programme implementation, providing guidance, programme coordination, and the programme components that are internal to headquarters (i.e., developing and implementing the monitoring and evaluation system, coordinating learning and sharing across the pilots, and facilitating transformation of WFP's procurement activities) while country programmes focus on field-level implementation. In general, logframe activities under objectives 1, 3, and 4 are largely headquarters-level activities while activities under objective 2 relate primarily to country-level implementation.

Because each of the 21 pilot countries presents a different environment, they have significant latitude to tailor their specific implementation of P4P to suit their own unique constraints and opportunities. To manage their country pilots, individual countries are likely to have unique monitoring and evaluation needs that the global system does not address. Therefore, individual countries have the flexibility to augment the global logframe and monitoring and evaluation system to incorporate their own unique activities, outputs, outcomes, and indicators.

The P4P Coordination Unit in Rome bears responsibility for guiding and coordinating monitoring and evaluation activities at the global level. The individual pilot countries, however, will implement monitoring and evaluation of their country-specific programme elements.

Impact Assessment Approach & Sampling

The four P4P objectives mention a number of expected impacts. In particular, they speak of 1) increasing profitable smallholder farmer engagement in markets, 2) increasing smallholder farmers' production, organization, and marketing capacities, 3) raising smallholder farmers' income, and 4) increasing sales by low-income farmers to WFP.

Because of the emphasis on impacts, the monitoring and evaluation system developed a specific approach to impact assessment. Impact measures focus on farmers' organizations and their smallholder members. Therefore, the impact assessment component of the P4P monitoring and evaluation system deals only with assessing the impact of P4P on various behaviors of and outcomes for smallholder farmers and farmers' organizations.

Alternative Approaches to Impact Assessment

Approaches to impact assessment for P4P must address two issues – they must accurately measure the change, if any, in the variable of interest (e.g., household income, agricultural production, etc.) and they must plausibly attribute that change to P4P. For example, in the context of assessing the impact of P4P on agricultural production, an impact assessment must determine first whether agricultural production has, in fact, changed and then establish whether any measured change is a result of P4P.

Attributing a measured change to a particular intervention means ruling out (or controlling for) other factors that could have accounted for the change. The three common approaches to impact assessment are 1) comparison groups, 2) before and after comparisons, and 3) statistical approaches. The remainder of this section briefly reviews each of these approaches and the strengths and weakness of each in the context of P4P.

The *comparison group* approach controls for external factors by comparing those who are potentially affected by P4P to those who are not. For example, if P4P is expected to increase agricultural production, then the approach would compare changes in agricultural production of a group of farmers who were potentially affected by P4P to changes in production of a similar group who were not. If the two groups are identical in every respect except for participation in P4P, then any observed differences between the two groups over time are likely attributable to P4P. The approach controls for external factors because both groups, if they are truly identical, are exposed to the same factors and respond in the same way, in aggregate, to those factors.

The comparison group approach produces the best measure of impact *if the two groups are identical in every important respect except for participation in P4P*. This is a high hurdle to clear however and in practice it is nearly impossible to ensure that the groups are identical. If the groups are dissimilar, then observed differences between the groups over time may arise from the dissimilarities rather than from P4P and the approach will over or understate the impact of P4P. Comparison group approaches often use statistical methods to control for observed difference between comparison groups.

The *before and after* approach compares the situation before an intervention to the situation after the intervention. For example, with P4P, the approach would compare agricultural production before P4P to the level of production after some period of experience with P4P. The approach is capable of identifying correlation between agricultural production and participation in P4P but it does not explicitly control for external factors. The before and after approach may work as an impact

assessment approach if few other important external factors exist but this is certainly not the case with P4P.

Finally, *statistical approaches* use econometric techniques to control for external factors. For example, the approach might regress agricultural production on a set of variables expected to affect production. The variables would include a measure of participation in P4P as well as data on external factors. Given adequate data, the analysis would estimate the separate effects of P4P and other factors on agricultural production.

The availability of adequate data is the single greatest constraint to employing statistical approaches. Obtaining reliable measures of impact from a statistical approach would require complete and accurate data on the entire range of factors likely to affect production. This is difficult in any country, and particularly in developing countries where WFP is implementing P4P. Inaccurate or incomplete data will lead to unpredictable bias in impact estimates. Furthermore, even if data are available, it is difficult to identify every external factor that might affect production, much less find reliable measures of those factors.

The P4P Impact Assessment Approach

The P4P monitoring and evaluation system employs a comparison group approach to impact assessment and this affected the design of the baseline. Identifying comparable groups is challenging but the other available approaches are not well suited to impact assessment. The before and after approach is inadequate because it does not control for the myriad external factors that may affect the impacts of P4P. The statistical approach is impractical because it entails prohibitive data requirements.

The M&E design team initially anticipated conducting a rigorous impact assessment in each of the 21 pilot countries. This would require that each country establish baselines for groups of participating and non-participating farmers' organizations and member smallholder farmers. In preparation, most of the 21 countries collected baseline data in 2009 or early 2010. The initial impact assessment design also specified that each country would conduct follow-up surveys of participating and non-participating farmers' organizations, farmers, and traders each year.

Managing such intensive data collection in 21 countries strained the capacities of many country offices to collect, manage, and analyze the data. It also tested the capacity of the P4P Coordination Unit to provide technical guidance to country offices and to aggregate the data to facilitate monitoring and evaluation and learning. In fact, by late 2010, only about half of the countries had submitted baseline data to the Coordination Unit and few, if any, of the datasets were adequately cleaned and organized for comprehensive analysis.¹⁶

The realities of implementation in some countries also limit prospects for meaningful impact assessment and contributed to the decision to limit impact assessment to selected countries that represent the range of strategies and provide good prospects for impact assessment. For example, in several countries farmers' organizations, and thus their farmer members, are migrating from non-participating to participating status thus eroding the control group necessary for impact assessment.

¹⁶ The P4P Coordination Unit is in the process of creating an independent "data hub" to provide technical assistance to country offices in data collection and analysis and to manage the monitoring and evaluation data.

In some cases this is a consequence of implementation decisions. In others it is the result of outside influences (e.g., the Government of Rwanda’s decision to replicate P4P-like interventions nationwide).

Mindful of the intense effort required to collect large quantities of survey data, several members of the P4P Technical Review Panel suggested limiting impact assessment to a few countries with good prospects for impact assessment. Based on this input, and its experience with the baseline, the P4P Coordination Unit revised its impact assessment approach to one that is better aligned with the WFP’s capacities to manage the process. The revised impact assessment approach will conduct rigorous impact assessment in a group of “impact assessment” countries selected to represent the range of P4P approaches and on the basis of their potential for meaningful impact assessment.¹⁷ Furthermore, it will collect survey data for monitoring and impact assessment only in years three (mid-term) and five (final). WFP will estimate impacts in the non-impact assessment countries by extrapolating from impact assessment countries with similar programmes and by drawing on data collected from participating farmers’ organizations and farmers in the non-impact assessment countries. Table 1 summarizes key differences between the original and revised impact assessment approaches.

Table 1. Comparison of Original and Revised Impact Assessment Approaches

	Original approach	Revised approach
Frequency of surveys	Annually	Years 1, 3, and 5
Collect data from samples of participating farmers, farmers’ organizations, and traders	All countries	All countries
Collect data from samples of non-participating farmers, farmers’ organizations, and traders	Impact assessment countries only	All countries
Collect data from the same samples of respondents (panels) in each data collection cycle	No	Yes

The revised approach does not collect as much data as the original approach – household survey data will only be available in years one, three, and five. However, countries will continue to collect data from other sources (i.e., farmers’ organization surveys, case studies, market monitoring, LRP monitoring, procurement monitoring) continuously. The advantages of the revised approach outweigh this disadvantage. The advantages are:

- The less intensive primary data collection effort is much more aligned with the capacities of country offices and the P4P Coordination Unit in Rome to collect and analyze data.
- The pace of annual surveys left little time for reflection and learning. While limiting surveys to years one, three, and five will generate less data from households, it will create space for fully exploring the data, reflecting on what it reveals, and learning. It will also give country offices more time to prepare for follow-up surveys.

¹⁷ The potential for impact assessment depends on the focus of the country’s strategy, how well the country is implementing the strategy, and the quality of its baseline in terms of sampling and data quality. The P4P Coordination Unit is reviewing characteristics of all 21 pilot countries to select the countries that will pursue rigorous impact assessment.

- Focusing on a limited number of countries for impact assessment will improve the quality of impact assessment overall. Not all countries have equal prospects for meaningful impact assessment. Focusing impact assessment efforts on countries with the greatest prospects for learning makes better use of resources and will produce more reliable and defensible estimates of impact.
- What the revised approach sacrifices in terms of data quantity, it makes up for in quality. The panel dataset collected under the revised approach will provide a much richer source of information about the dynamic impacts of P4P than the original approach of independent samples. Furthermore, fewer surveys will permit countries to focus more resources on the remaining surveys thus improving the quality of data. The panel dataset also allows WFP to count individual beneficiaries.¹⁸

Switching to a panel data approach also poses risks. The original samples of households selected by country offices may not be large enough to absorb the attrition in a panel over a five-year period. In some countries it may be necessary to augment the panel if attrition rates become too high. This will compromise the panel to some extent but will still provide better information than the original approach.

Implementation Issues

1. Ensuring the similarity of comparison groups is the primary challenge when implementing the comparison group approach. The impact assessment component of the monitoring and evaluation system focuses on the impacts of P4P on farmers' organizations and their smallholder members. Participating farmers are members of participating farmers' organizations. Therefore, the issue of selecting comparable groups comes down to selecting comparable farmers' organizations with one criteria of comparability being the similarity of their smallholder members. Comparison groups of smallholder farmers are then random samples from the membership of participating and non-participating farmers' organizations.
2. Random assignment is the best way to ensure comparable comparison groups. In the context of P4P, this would involve identifying a set of farmers' organizations that were potential P4P participants and then randomly assigning organizations into participating and non-participating groups. However, random assignment was not feasible in most pilot countries. In many cases, country programmes identified interested, relevant, and capable supply-side partners first and then worked with all the organizations that those partners supported. This strategy "used up" (in some cases all of) the farmers' organizations that would have been the best comparison group in a given location (i.e., other organizations with which the partner worked). In some cases (e.g., Tanzania) so few organizations met the basic criteria for participation in P4P that few qualified organizations remained to form a comparison group. A few countries were far enough advanced in their implementation that random assignment was no longer feasible.

¹⁸ The original approach, based on independent samples of farmers in each year, allowed analysts to determine only whether all participants had, on average, attained the \$50 annual income increase necessary to qualify as a beneficiary or not. Therefore, either all farmers in a country were beneficiaries (if they had, on average, increased incomes by at least \$50) or all failed to benefit. The panel dataset will allow WFP to estimate the number of individual beneficiaries in each country.

3. Also, non-P4P farmers' organizations may disintegrate over time as members migrate to participating organizations. It is difficult to determine how much risk this poses to the sampling strategy and impact assessment approach.

The monitoring and evaluation system specifies the following sampling approaches for the three target groups (i.e., farmers' organizations, smallholder farmers, and traders).

- **Farmers' organizations** – WFP will work with a limited number of farmers' organizations in each country. Therefore, country offices will collect monitoring and evaluation data from *all* participating farmers' organizations and sampling is not necessary. The number of non-participating farmers' organizations, however, is potentially large. The monitoring and evaluation system instructs country offices to identify a set of farmers' organizations that are similar to participating organizations except that they are not participating in P4P and then to select a random sample of organizations from this set to represent non-participating organizations.¹⁹ General criteria for identifying similar organizations include:
 - Weather, soils, and other location-specific conditions that affect agricultural production;
 - Agricultural practices (i.e., degree of mechanization, access to and use of inputs, cultivation and cropping patterns, etc.);
 - Access to markets for staple commodities and cash crops. Consider not only physical access (e.g., distance to markets, road conditions, etc.), but also factors such as access to organizations, facilities, or services that make it easier to market commodities (e.g., access to a warehouse, access to credit, the marketing assistance of a farmers' organization, or other assistance providers, etc.);
 - Similar socioeconomic characteristics; and
 - Similar number and sex and farm size composition of members.

The monitoring and evaluation system also suggests that geographic proximity may be a good proxy for many of these criteria and recommends that country offices select non-participating organizations that are geographically close to participating organizations. However, if the groups are too close, then members of participating groups may migrate to participating groups.

It is also possible in some countries that, as the P4P programme matures and expands, it will incorporate some of the non-participating groups. The monitoring and evaluation system suggests that country offices select a large enough sampling frame of non-participating groups so that an adequate number remain in the sampling frame at the end of the five-year pilot.

In some countries, country offices have had to expend considerable time and effort to generate sufficient information about non-participating farmers' organizations to assess their comparability with participating organizations.

¹⁹ The instructions suggested the country offices select a random sample of non-participating farmers' organizations of roughly the same size as the number of participating organizations.

- **Farmers** – The sampling frames for smallholder farmers are the smallholder members of participating and (the sample of) non-participating farmers’ organizations. Country offices will select a random sample of smallholder farmers from each group. When farmers’ organizations are many and widely dispersed geographically, country offices may cluster samples to control data collection effort and cost.
- **Traders** – In most, if not all, P4P pilot countries it will not be possible to develop comprehensive lists of traders from which to select representative samples. Fortunately, the P4P monitoring and evaluation system does not require representative samples of traders, nor does it specify rigorous impact assessment for traders. Instead the monitoring and evaluation system will focus on describing the role of traders in marketing smallholders’ commodities and how that role changes as a result of P4P. This analysis does not require representative samples or comparison groups.

When sampling is required, the sampling strategy specifies sample sizes sufficient to produce a five percent margin of error at a 95 percent confidence level. Limited resources for data collection restrict this level of confidence only to describing the characteristics of participating and non-participating groups and to comparisons between the two groups. Estimates for sub-populations (e.g., sex, region, crops, etc.) will have larger margins of error.

Implementation Examples

The following case studies illustrate the challenges P4P pilot countries are facing in implementing the impact assessment approach and how they are addressing these challenges. The P4P Coordination Unit has not determined whether the countries in the following examples will be impact assessment countries but the examples illustrate the challenges of impact assessment.

Zambia

Zambia’s P4P programme capitalizes on the opportunity presented by the nascent Zambia Agricultural Commodity Exchange (ZAMACE). WFP in Zambia has determined that it can best support market development and market access for smallholder farmers by channeling all of its local procurement (P4P and LRP) through the exchange. On the supply-side, WFP’s partners are supporting the development of rural ZAMACE-certified warehouses to facilitate smallholder access to the exchange. Farmers’ organizations and individual smallholder farmers can offer their commodities on the exchange by depositing them in a certified warehouse and hiring a broker to represent them on the exchange.

Because WFP will not buy directly from farmers’ organizations in Zambia, the notion of participating and non-participating groups is not applicable. In theory, any farmers’ organization or smallholder farmer who can access ZAMACE is participating in P4P. However, ZAMACE-certified warehouses represent the only point of access to ZAMACE. Therefore, the impact assessment approach treats access to the warehouses as a proxy for access to ZAMACE.

Zambia will collect baseline and follow-up data from all farmers’ organizations that are receiving direct capacity building support to access ZAMACE-certified warehouses. It will also collect baseline and follow-up data from a sample of organizations with similar capacities but with limited access to the warehouses (either because they have not been informed of the ZAMACE opportunity, have not received training in how to prepare commodities for ZAMACE, or are too far

from a warehouse. It will also collect baseline and follow-up data from random samples of the members of these two groups of farmers' organizations. As the pilot progresses, analysis of the survey data will attempt to identify the impact of access to certified warehouses on sales behavior and incomes.

Perhaps the greatest risk to impact assessment in Zambia is the buying activity of the government. The government's Food Reserve Agency purchases maize directly from smallholder farmers at prices that are often well above the market. In 2010, it bought large quantities at prices about 50% above the prevailing market price. Consequently, few farmers' organizations were willing to test ZAMACE as a marketing platform. If the government continues this level of buying throughout the P4P pilot it may substantially limit P4P's ability to gain traction with smallholder farmers.

Tanzania

Tanzania's approach to P4P is more "typical." It procures from Savings and Credit Cooperatives (SACCOS) which collect commodities from farmers and pay them in advance. The programme has selected 26 SACCOS to participate in P4P. It selected non-participating SACCOS from wards adjacent to the participating SACCOS. It selected random samples of smallholder farmers from both groups – participating and non-participating SACCOS.

Substantial variation across participating SACCOS in Tanzania presents some challenges in identifying a similar group of non-participating SACCOS. Some participating SACCOS specialize in one commodity while others may deal in two or more. The heterogeneity in participating SACCOS makes it difficult to identify typical SACCOS' characteristics to guide selection of the non-participating group.

Rwanda

Rwanda's approach to procurement is also somewhat "typical" in that it buys from farmers' organizations and cooperatives. However, Rwanda is using a novel approach to selecting cooperatives from which to procure. The approach is particularly suited to impact assessment.

Rwanda obtained a list of all cooperatives in the country from the Rwanda Cooperative Agency. It then applied a set of selection criteria to identify a set of cooperatives eligible to participate in P4P. The country office then *randomly* selected a number of cooperatives as participating cooperatives and assigned the rest to the control group. The number of participating cooperatives selected will be consistent with P4P procurement targets.

Further developments, however, emphasize the challenges of impact assessment in a quasi-experimental setting. The Government of Rwanda has modeled its support for farmers' organizations on the P4P approach. It has begun to offer the same package of supply-side support and is also buying directly from organizations. The government's activity has essentially eliminated the difference between participating and non-participating farmers' organizations and farmers thus severely limiting the prospects for impact assessment.

Mozambique

Mozambique's P4P programme operates in three districts where it buys from the trading arm of an umbrella body (IKURU) which incorporates a number of farmer associations in one district, a farmer forum in another district, and two farmer federations in another district. Farmer federations

are district-level organizations which have farmer forums as members which, in turn, are composed of member farmer associations. The associations represent the lowest tier of the hierarchy and are small groups of 10-15 farmers.

This structure presents a number of challenges to selecting comparison groups. In particular:

- No organization equivalent to IKURU exists in the district in which it operates or in adjacent districts. It is therefore impossible to identify a similar group to serve as a comparison.
- Federations are apex organizations within their districts. Once again, it will be difficult to identify similar organizations in close geographic proximity to serve as a comparison group. Furthermore, the federations in adjacent districts are not comparable to the federations from which WFP will buy through P4P.

In these two districts, WFP will define the sampling frame as farmer associations. Those that regularly supply IKURU (in one district) and the district federation (in the other district) for P4P contracts are the participating farmers' organizations and those that do not are the non-participating organizations. Participating and non-participating associations are geographically dispersed across the regions and mixed together. The sampling frames of smallholder farmers are then the smallholder members of participating and non-participating associations.

Ethiopia may be another example worth including given the cu/pc scenario and the fact that we work through cus's which have a number of cu's under them from which they may or not receive produce. Also some of the control cus have moved into participating group now

Defining P4P Beneficiaries

Establishing accurate estimates of the number of smallholder/low income farmers benefiting from P4P is important in order to:

- Plan implementation and ensure an adequate capacity development strategy for farmers' organizations and smallholder/low income farmers.
- Determine the actual number of smallholder/low income farmers who have achieved the expected benefits associated with participation in P4P.

In the first case, estimates of expected beneficiaries contribute to planning. The Bill and Melinda Gates Foundation (BMGF) proposal derived estimates of expected beneficiaries from the number of smallholder/low income farmers expected to experience a very specific impact – average annual income increases of at least US\$50. It based these estimates on assumptions about expected procurement levels, average surpluses marketed by smallholder/low income farmers, and the net income derived from the sale of commodities. Estimates of expected beneficiaries helped define expectations and the criteria for assessing programme performance.

Individual P4P Country Implementation Plans (CIPs) determined country-specific targets for expected beneficiaries based on anticipated procurement levels and the average surpluses marketed

by smallholder/low income farmers. These estimates helped country offices plan their programme requirements and also established targets for monitoring performance.

In contrast to beneficiary estimates use for planning purposes, estimates of actual beneficiaries reflect the success of the programme in producing specific defined outcomes and impacts relative to smallholder/low income farmers.

The number of beneficiaries identified in planning documents and those who actually experience benefits are not necessarily the same. For example, incorrect assumptions used in planning or changing circumstances (e.g., procurement levels) in pilot countries may cause a discrepancy between the two numbers. Individual P4P pilot countries may also spread procurement so thinly across participants that it does not provide the sustained and meaningful level of procurement necessary to stimulate production, increase incomes, and produce beneficiaries. An implementation strategy that does not rationally match targeted beneficiaries to procurement levels may produce no actual beneficiaries. A poorly designed or executed P4P strategy may also fail to produce beneficiaries.

How Does P4P Define Beneficiaries?

Programme *participants* are smallholder/low income farmers who have the opportunity to sell to WFP through P4P. *Beneficiaries* are the subset of participants who obtain some benefit from participation. For monitoring and evaluation purposes, WFP defines beneficiaries relative to specific, pre-defined benefits and benefit levels. Obviously, not all participants will become beneficiaries. But well designed and executed programmes will transform the majority of participants into beneficiaries. Participants who become beneficiaries are direct beneficiaries in the sense that P4P explicitly targets them and they benefit directly from the engagement. However, farmers with no direct engagement with P4P may benefit indirectly from the programme's spillover effects. Such indirect beneficiaries may include farmers who benefit indirectly from higher farm gate prices brought about by increased competition in commodity markets.

Participants

All smallholder farmers who have the opportunity to sell to WFP through P4P. Usually the members of participating farmers' organizations.

Indirect beneficiaries are important, but for purposes of P4P's monitoring and evaluation, the most relevant definitions of beneficiaries relate directly to the expected outcomes and impacts of smallholder/low income farmers' participation in the programme. Purchase for Progress operates on the premise that WFP's procurement, along with supply-side support from partners, will enhance smallholders' access to more profitable markets, increase farm gate prices, and increase agricultural production as higher prices provide an incentive for investment in production. Increased production will – when tied to improved group marketing capacity, market development, and a favorable policy environment – often lead to the impact of increased household income. In monitoring and evaluation terms, P4P will produce a behavioral change or outcome (increased production) that will – in the right circumstances – generate the impact of increased household income.

Beneficiaries

Smallholder/low income farmers who realize a US \$50 increase in annual household income as a result of P4P.

Based on this rationale, WFP believes that the most relevant definition of a beneficiary is a smallholder/low income farmer who attains the ultimate desired impact of the programme – increased household income. Furthermore, the BMGF proposal sets a lower bound on the income increment necessary to qualify as a beneficiary. The proposal defines beneficiaries as smallholder/low income farmers who realize a US \$50 increase in annual household income as a result of P4P. Increased production – another principle intended outcome – directly implies increased income. Regardless of whether a household sells its increased production or not, it has a value. The measure of household income used to define beneficiaries will include the monetary value of all crops produced regardless of how the household used the crop (i.e., sold, consumed, fed to animals, used for seed).

The foregoing discussion should not imply that there are competing definitions of beneficiaries – P4P beneficiaries and BMGF beneficiaries. Instead, it should emphasize that P4P generates multiple outcomes and impacts, each of which provides useful information about different dimensions of programme performance.

Counting Beneficiaries

The exercise of estimating the number of actual beneficiaries is not trivial. It involves measuring complex concepts (production and income) with survey methods applied to population samples under difficult field conditions and within the context of a research design capable of attributing observed outcomes and impacts to the P4P intervention.

The number of beneficiaries is the number of smallholder/low income farmers who have increased their annual household income by at least US \$50 relative to the baseline and as a result of P4P. The latter part of this definition is critical. The programme does not necessarily expect to increase a household's income year after year. It aims instead to bring household income up to a specified level determined to be feasible in the programme planning phase (i.e., a US \$50 increase). The programme then expects to have developed markets sufficiently to sustain these increased incomes beyond P4P. Thus the reference point for assessing income change is always the baseline year.²⁰ To qualify as a beneficiary, a household must experience an increase in annual income of at least \$50 (relative to the baseline and non-P4P smallholder farmers).²¹

Other P4P Outcomes

P4P generates multiple outcomes and impacts, each of which provides useful information about different dimensions of programme performance. In addition to the actual beneficiaries that achieve an increase in household income, P4P will report on the following indicators (all disaggregated by sex):

- Average per farm marketable surplus of staple commodities produced by participating farmers.
- Average per farm quantity of staple commodities sold by participating farmers.
- Average (per smallholder farm) post-harvest losses of staple commodities as a percentage of annual production.

²⁰ If external shocks (e.g., droughts or floods) caused the baseline year to be unrepresentative of typical production levels, then the baseline may have to be adjusted.

²¹ In practice, either definition can be expressed on an annual basis or as an average over all programme years. Using averages over the period of engagement with P4P will mitigate the impacts of external shocks (e.g., floods and droughts) on production and incomes and thus on beneficiary counts.

The procedure for identifying beneficiaries must also attribute the income increase to the P4P intervention. Therefore, the beneficiary estimate must account for the measured (positive or negative) change in income of households that are not participating in P4P. This is standard procedure in impact assessment to attribute impacts to an intervention. Subtracting the observed change in non-participating households from the observed change in participating households²² corrects for factors other than the P4P intervention that may influence incomes. The relevant measure of income change is thus relative to *what would have happened without P4P*. Note that in extreme cases, a household that produced less than the previous year may be counted as a beneficiary if non-participating households experienced an even greater decline in production. In this case, P4P prevented a decline in incomes relative to that experienced by non-participating households.

In addition to comparing change to baseline and non-participating group values, procedures for estimating beneficiaries must also define income very broadly. Income includes not only households' cash income but also includes the monetary value of commodities retained for household consumption, animal feed, and seed as well as the value of any in-kind goods or services received by the household. Income must also be net of any expenses incurred in raising crops or livestock or running a business enterprise.

The biennial (every two years) surveys collect data necessary to identify beneficiaries. A beneficiary in year n is any household that satisfies the condition:

where: y = annual household income

\bar{y} = average (over households) annual household income

p and c denote P4P and non-P4P, respectively

n denotes year n of the P4P pilot and 0 represents the baseline year

The smallholder farmer data collection instrument collects data necessary to calculate annual household income.

When is a Participant not a Beneficiary?

A programme may fail to transform participants into beneficiaries for a number of reasons. These include:

1. **Poorly designed (scaled) programme** - The P4P development hypothesis states that bringing sufficient WFP demand to bear upon a targeted group of smallholder farmers for an appropriate period of time will trigger increases in production and subsequent sale of staple commodities. The magnitude and duration of the procurement stimulus required to provide the incentive to invest in production is an empirical question but there is likely some lower threshold. A programme that is inappropriately scaled (i.e., spreads its procurement too thinly over participants to attain the threshold) may fail to provide sufficient incentives

²² With both changes measured relative to the baseline for the relevant group.

to investing in production or market development. Thus, to transform participants into beneficiaries, a programme must match the number of participants to expected procurement.

2. **Weak implementation** - Even if the expected beneficiary and procurement targets imply a sufficient level of procurement from each farmers' organization/farmer to produce the desired supply and market development response, ineffective implementation may still inhibit the creation of beneficiaries. A programme may fail to provide the support farmers' organizations and farmers need to take advantage of WFP's demand by:
 - Failing to identify or effectively address the real capacity building needs of farmers' organizations or farmers,
 - Failing to address all of the constraints to market development and production improvement,
 - Selecting ineffective partners,
 - Failing to develop alternative markets thus creating dependency on WFP's procurement and threatening the sustainability of results, or
 - Falling victim to external risks.
3. **External shocks** - External shocks such as droughts and floods pose a constant threat to agricultural production in the developing world. Such shocks may cause production and incomes to fall below baseline levels. However, they do not necessarily reduce the estimated number of beneficiaries. Beneficiary estimates, like all impact measures, are appropriately measured relative to the counterfactual (i.e., what would have happened without P4P). Therefore, if P4P farmers experience a reduction in income relative to the baseline but non-P4P farmers experience an even greater reduction, then the P4P farmers fared better than they would have without P4P and may therefore be beneficiaries.
4. **Participants fail to engage with P4P** – Some members of participating farmers' organizations will not engage with P4P. They may not participate in capacity development activities, may not participate actively in the organization, and may not sell through the organization.

Frequency of Beneficiary Estimates

Measuring beneficiaries requires comparing participating farmers to non-participating farmers in order to attribute the observed benefit to P4P. Therefore, estimates of the number of beneficiaries are possible only for those years for which appropriate data are available for both groups. Only the countries selected for impact assessment will be able to calculate beneficiary numbers rigorously because only these countries will collect data from a control group. Non-impact assessment countries will be able to calculate income changes for participating farmers but will not be able to rigorously attribute the change to P4P.²³ In either case (impact assessment or non-impact assessment

²³ Even without the rigorous attribution to P4P, evidence from case studies and other sources may make a strong case for full or partial attribution.

countries) beneficiary counts will be available only in years three and five when the countries conduct follow-up surveys of P4P participants.

Measuring Household Income

Increasing smallholder household income is a key intended outcome of P4P. However, the proposal to the Bill and Melinda Gates Foundation, which is the primary P4P reference, does not clearly define what P4P intends to achieve in terms of income increases. In particular, Outcome 1 in the proposal states:

“By the end of the project, annual incomes of a target base of 350,000 smallholder farmers supporting 1.5 million people have increased by a projected average of US \$50.”

It is not entirely clear whether this refers to an average over participating farmers, the average over years per farmer, the average income increment over years of participation, or the average income increment per household by the end of the pilot.

Careful examination of projected farmer participation and total income increments (pages 28-29 of the BMGF proposal), however, suggests that the intended income target is to increase annual income per household-year of participation by an average of \$50. Thus, a smallholder farmer would be expected to attain an average \$50 increase in income *in each year he or she participated in P4P*.

The proposal is also unclear about the measure of income it expects P4P to influence. It usually refers to income in general but also refers to “annual crop income” and “farm income.” Discussions with WFP and BMGF personnel suggest that the proposal likely refers to increasing smallholder household income from all sources and from the sale of staple crops. However, because smallholder farmers will likely shift resources from other income-earning activities to increase production of staple crops, total household income from all sources is a more relevant measure for assessing the impacts of P4P on the welfare of smallholder farmers. The monitoring and evaluation system collects data to calculate total household income and also to disaggregate income by source (e.g., income from the sale of staple crops).

A greater challenge with relying on household income as a key outcome of P4P is the difficulty in obtaining accurate measures of income in a developing country context. Most smallholder farmers are not formally employed and keep few records of household finances which can encompass a variety of livelihood activities. Furthermore, many transactions (purchases and sales) may be non-monetary (e.g., barter, gifts, etc.). This situation makes accurate income measurement difficult for two reasons. First, the non-monetary nature of many transactions makes it necessary to estimate monetary equivalents – a process that introduces substantial measurement error and judgment. Second, the informal nature of income coupled with the lack of financial records means that farmers rely on recall when reporting income. Recall of such data is notoriously inaccurate and becomes even less accurate with the passage of time. The fact that many household income-earning enterprises are businesses exacerbates the measurement problem by requiring accurate measures of business expenses as well in order to attain a measure of net income.

The monitoring and evaluation system employs a number of strategies to address the income measurement challenge. These include:

- Providing extensive guidance on enumerator training. One element of the training focuses on developing effective probing techniques to estimate monetary equivalents of non-

monetary transactions and to improve the accuracy of respondents' recall of income and expenses.

- Using detailed questioning of agricultural production and marketing activities to enhance recall of income and expenses related to agriculture. The design team borrowed extensively from existing data collection instruments to design the questionnaire modules that ask about agricultural production, marketing, and other livelihood activities. Unfortunately, limited resources for primary data collection meant that questionnaires needed to be relatively compact thus constraining the depth of questioning.
- Timing data collection to maximize the accuracy of recall of key agricultural production and marketing data. The monitoring and evaluation system specifies that country offices administer the smallholder farmer survey soon after the main harvest and marketing season when households will have harvested their main crops and made most of their marketing decisions.

Evaluating Standard Local and Regional Procurement

Standard local and regional procurement (LRP) refers to WFP's standard approach to local procurement. With the LRP approach, WFP buys primarily through competitive tenders with large traders who have the capacity to reliably deliver large quantities of high quality commodities. Evaluation of LRP's impact on smallholder farmers is relevant because it represents a potential alternative to P4P as a way for WFP to structure its local procurement to the benefit of smallholder farmers. Even though WFP rarely procures directly from smallholder farmers or farmers' organizations through LRP, smallholders probably supply most of the commodities offered by traders and may thus benefit from WFP's procurement. There is good reason to believe that, under the right conditions, LRP may lead to increased farm gate prices for commodities, agricultural production, smallholder incomes, and investment in trading and food processing (Sserunkuuma & Associates Consult 2005).

To compare LRP and P4P, the monitoring and evaluation system needs to assess the relative impacts of LRP and P4P on market development and the welfare of smallholder farmers.²⁴ More specific criteria for comparing the performance of LRP with P4P include impacts on markets (positive and negative), market development impacts, agricultural development impacts, the number of smallholder farmers engaged, impacts on smallholder household income, and other measures of smallholder welfare. To contribute to learning in P4P, evaluation of LRP will need to answer the following questions.

- Does WFP's procurement through LRP affect market prices, aggregate production, market development, the number of smallholder farmers profitably engaged in markets, smallholder household income, and other measures of smallholder welfare?
- If so, under what conditions does it produce these impacts (e.g., level, regularity, and duration of procurement; political and policy environment; agricultural production environment; market environment; etc.)?

²⁴ The comparison also needs to consider the timeliness, efficiency, and safety of procurement but WFP's procurement database will provide those data.

- If so, how do those impacts compare to those generated by P4P and under what circumstances is one mode of procurement preferable to the other?

Answering these questions requires an assessment of the impacts of LRP and P4P. However, a number of issues make rigorous impact assessment of LRP difficult. These issues include:

- **Lack of a baseline** – WFP, and others, have been buying from larger traders in developing countries for many years. It is therefore difficult to determine baseline (i.e., pre-WFP procurement) conditions from which to measure impacts. Furthermore, since WFP's purchases may be indistinguishable from the purchases of other large buyers (i.e., demand is demand), a true baseline might have to be absent other similar sources of demand, or at least control for them.
- **Lack of a control group** – If LRP affects smallholder farmers' incomes it is likely to be because increased demand for commodities increases farm gate prices and aggregate production thereby increasing incomes. However, because the larger traders who supply WFP under LRP generally source their commodities nationally, the impacts are likely national in scope and do not accrue only to particular farmers or regions. If all, or most, farmers benefit from the impacts of LRP, then it is difficult to determine what would have happened without LRP.
- **Lack of records** – Even if there was reason to believe that the benefits of LRP accrued to a specific set of smallholder farmers (i.e., those who contributed to a batch of commodities sold to WFP), it will be difficult to determine how many farmers contributed commodities or the impact of the purchase on their welfare because traders rarely keep records of the individual farmers from whom they purchase commodities.²⁵
- **Difficulty predicting future LRP activity** – One possible approach to assessing the impact of LRP would be to identify a set of countries with little or no LRP activity but where WFP expects to substantially increase LRP procurement in the coming five years. Analysis of aggregate production, prices, market development, and farmer incomes as WFP ramps up procurement could then provide a before and after measure of the impact of LRP. However, it is difficult to predict the countries in which WFP is likely to substantially increase LRP activity.

The issues raised above represent significant barriers to assessing the impacts of LRP based on future procurement. An alternative approach is to attempt to draw some conclusions from the historical record of more than two decades of LRP experience and to augment the analysis with qualitative data from in-depth interviews with traders, processors, and smallholder farmers. The approach is retrospective and thus avoids many of the problems cited above. It is also likely to be cost effective because it relies largely on available secondary data rather than on extensive primary data collection.

²⁵ While it is possible that WFP could impose this level of record keeping on its suppliers, there is still little reason to believe that LRP benefits smallholder farmers who happen to contribute to a sale to WFP any more or less than any other smallholder farmer with access to markets.

The approach is data intensive and the availability and quality of data may ultimately render it infeasible. However, the basic data (i.e., WFP procurement, prices, and aggregate production) do appear to exist for a large number of countries. In 2008 WFP procured varying quantities of staple commodities from 74 developing countries. The organization's procurement database can provide detailed procurement statistics for the entire history of LRP procurement. FAOSTAT reports agricultural production, prices, and other data for 71 of these countries from 1961 through 2006.

In very broad strokes, the approach might proceed as follows:

Econometric Analysis of Price Effects of LRP – Conduct an econometric analysis of prices to determine whether LRP has affected price levels and price variability for agricultural commodities. Explanatory variables might include lagged prices, purchases by large buyers (e.g., WFP, other donors, government, etc.), and marketed surplus. The analysis would focus on determining whether the level of demand from WFP and others had affected commodity prices at the farm gate and whether price increases were persistent.

Econometric Analysis of Aggregate Output – Conduct an econometric analysis of aggregate agricultural output to determine whether LRP has affect agricultural production (i.e., whether there has been a supply response to increased prices). Explanatory variable might include lagged prices, marketed surplus, policy variables, weather variables, and determinants of demand (i.e., purchases by large buyers, population, urban population, income, etc.).

Qualitative Analysis – In-depth interviews with smallholder farmers, traders, and processors would focus on *how* LRP had affected smallholders, traders, and processors. For instance, it will be important to understand if smallholder farmers received higher prices and how they responded and if traders and processors increased their capacity in response to increased LRP activity, how reliant they are on WFP, and how and where they obtain the commodities they sell. Information on how and where traders and processors who sell to WFP source commodities will be crucial to estimating the number of smallholder farmers who contribute commodities to WFP through traders.

Seasonality

Many of the agriculture phenomena of interest to the monitoring and evaluation effort have distinct seasonal patterns. For example, smallholder farmers typically plant, harvest, and market their crops at specific times (seasons) throughout the year. Some countries have multiple agricultural seasons which may often concentrate on different crops. Furthermore, since subsistence farmers may depend on their own production for much of their household consumption, household welfare may also exhibit seasonal variation. The harvest season is typically a season of surplus while the period leading up to the harvest and prior to new crops being available for consumption is the lean season.

The seasonality of agricultural production influenced several aspects of the data collection process. For example, smallholder household welfare is an important intended outcome of P4P and food security is a key component of welfare. However, measures of household food security can be very sensitive to the season in which the data are collected. For example, households' consumption patterns (i.e., the quantity and types of food consumed) may be very different in the harvest season than in the lean season. Researchers' impressions of household food security may therefore depend on when they collect data.

WFP originally envisioned collecting data in both harvest and lean seasons in order to capture seasonal variation in food security and other measures of household welfare. However, limited

resources available for data collection coupled with the fact that many countries have multiple seasons in a calendar year (thus necessitating multiple rounds of annual data collection under the intended data collection strategy) rendered collecting data in both seasons impractical.

Given that resources and country office capacity precluded such frequent data collection the design team elected to time a single data collection effort (in years one, three, and five) for the end of the main harvest season when farmers' organizations and farmers could most accurately recall key agricultural production and marketing data. However, measures of food consumption (the food consumption score for example) collected at this time will likely overstate households' typical (throughout the year) food security situation. To deal with this potential bias, the monitoring and evaluation system includes alternative measures of household welfare with annual dimensions (i.e., income, assets, annual expenditures, number of months in which the household had to purchase food rather than rely on own production) that are not as sensitive to the timing of data collection.²⁶

The seasonality of agricultural production also influenced the way the design team structured elements of the data collection instruments. In particular, the instrument for smallholder farmers asked about production and marketing in each season separately. While this adds to the length and complexity of the interview, it maximizes the prospects for accurate and comprehensive recall of key agricultural production and marketing data and thus improves the quality of the data. It proved quite challenging, however, for the design team to devise a consistent and unambiguous way to refer to seasons that would generally apply to all countries and country offices will need to be particularly sensitive to this issue as data collection progresses.

Monitoring

Monitoring is the process of performing periodic assessments of programme performance. With P4P, monitoring involves collecting the data to update indicator values throughout the five-year pilot.²⁷ This section briefly describes the proposed monitoring approach, some of the decision points in the design, and the rationale for the selected approach.

Monitoring utilizes data from all of the sources mentioned previously in this report (i.e., surveys of farmers' organizations, farmers, and traders; case studies; market monitoring; procurement monitoring; and farmers' organization records). Data from these sources contribute to the P4P performance indicators which track programme performance. Country offices will collect these data according to the schedule outlined in Figure 8.

Countries will collect survey data from the same group of farmers' organizations and farmers in each survey. The resulting panel dataset will provide a rich source of information about the dynamic aspects of P4P's performance and will also provide the data necessary for monitoring.

²⁶ The smallholder farmer survey still collects data to calculate the food consumption score (FCS) used by WFP as a measure of food security. However, the FCS will reflect food consumption during the surplus season and may not reflect households' typical experience. Furthermore, it may not exhibit the degree or variation across households that it might if collected during the lean season.

²⁷ Most countries will collect survey data from farmers' organizations, farmers, and traders only in years one (baseline), three, and five. They will collect data from other sources throughout the pilot. See the Data Collection Schedule section on page 36 for more detail.

Maintaining Data Quality

The P4P pilot covers 21 countries on three continents.²⁸ Even under ideal conditions, maintaining the integrity of monitoring and evaluation data in this situation would be challenging. Variation in the capacities of country offices to manage the data collection process and cross-country differences in programme implementation exacerbate the challenges. The monitoring and evaluation system incorporates several features designed to minimize these difficulties and maintain data integrity and quality.

- The P4P Coordination Unit in Rome directs all global P4P monitoring and evaluation activities. Having a single unit responsible for monitoring and evaluation ensures a consistent overall approach and the broad perspective necessary to evaluate P4P performance across a variety of situations.
- The headquarters-directed approach includes a common set of data collection instruments to ensure the cross-country comparability of the basic data required to calculate values for global indicators. Since individual pilot countries tailor P4P implementation to their own situations, they may have somewhat different data needs for country-level monitoring and evaluation. The common data collection instruments allow countries to adapt the instruments to their own uses as long as they do not compromise the data required for monitoring and evaluation of the global programme.
- The monitoring and evaluation system also includes a detailed manual that describes common procedures for training enumerators and conducting field work. This should help improve the quality of data collection, especially in countries with fewer staff experienced in collection techniques. Also, the monitoring and evaluation system design team pretested the data collection instruments in four pilot countries and engaged P4P staff from three additional countries in those pretests. The pretests involved training in all aspects of data collection including enumerator training, translation, and organizing field work. The practical experience provided by the pretests strengthened the capacities of the seven countries that participated.
- The P4P Coordination Unit in Rome developed common analysis routines for basic data analysis. These routines will ensure that country offices conduct basic analyses and report results consistently.
- The P4P Coordination Unit in Rome will also provide ongoing support to country offices by email, telephone, and site visits. It will also draw on the expertise of country office and headquarters Vulnerability Assessment and Monitoring (VAM) Units to assist P4P programmes with data collection issues.

The P4P Coordination Unit is in the process of establishing an independent “data hub” to manage monitoring and evaluation data. The hub will provide technical assistance to country offices in data

²⁸ The P4P pilot countries are Afghanistan, Burkina Faso, El Salvador, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Malawi, Mali, Mozambique, Nicaragua, Sudan, Tanzania, Uganda, Zambia, Democratic Republic of Congo, Liberia, Rwanda, Sierra Leone, and Laos.

collection, management, and analysis. It will also physically manage the data and support analyses at the global level.

Country Office Capacity

Collecting meaningful monitoring and evaluation data, analyzing the data appropriately, and drawing insights from the analysis requires special training, skills, and resources committed to the effort. Some country P4P programmes have, or have access to, the requisite skills and resources while others do not. Therefore, the monitoring and evaluation system has to specify an approach that is feasible in the context of country office capacity.

The monitoring and evaluation system incorporates a number of features that accommodated potential capacity limitations. In particular, it takes the following steps to address capacity constraints.

- Developing data collection instruments that are as simple as possible given the objectives of P4P and the monitoring and evaluation system. In some cases this required a careful balancing act between ambitious objectives and feasibility. The requirement to capture a \$50 annual increase in household income, in particular, added significantly to the complexity of the smallholder farmer data collection instrument. However, the design team streamlined the instrument as much as was feasible.
- Scaled back the data collection effort by limiting surveys to years one, three, and five and requiring surveys of non-participating farmers' organizations and farmers only in the handful of countries selected for rigorous impact assessment.
- Providing extensive training materials to guide country offices in training enumerators and administering data collection activities.
- Enlisting the expertise of country-level VAM units when available to consult or assist with data collection, analysis, and interpretation.
- Supporting country offices from headquarters through teleconferences, email, and site visits.
- Coordinating translation of data collection instruments into the main languages (i.e., French and Spanish) to ensure consistent and accurate translation across countries.
- Coordinating programming of PDAs for countries that use them for data collection.
- Developing common, packaged data analysis routines.
- Developing detailed instructions for calculating and reporting indicator values.
- Provided an outline for baseline reports.
- The P4P Coordination Unit is currently establishing a “data hub” that will assist countries in data collection, manage quantitative and qualitative data, and conduct analyses of the data. The hub will relieve individual country offices and the Coordination Unit of much of the burden of managing and analyzing data.

These measures will help maintain the quality and utility of collected data; enhance country offices' ability to understand the data and draw lesson from the analysis; and support collaborative learning between country office staff, other stakeholders, the P4P Coordination Unit in Rome; and the Bill and Melinda Gates Foundation.

Evaluations

Purchase for Progress will undergo external mid-term and final evaluations – the mid-term evaluation in year three and the final evaluation in year five. The P4P Coordination Unit strongly suggests that the evaluators employ the existing data collection instruments and enumerators if primary data collection is required.

There are a number of reasons for this recommendation including:

- using the same data collection instruments will ensure that data on which the P4P programme is evaluating its own performance is consistent with the data used by external evaluators,
- using trained enumerators experienced with the data collection instruments will produce higher quality data than if evaluators use new enumerators, and
- using existing data collection instruments and experienced enumerators will use evaluators time and resources more efficiently.

Case Studies

The case studies are in-depth interviews with farmers' organizations, smallholder farmers, and traders designed to collect the detailed qualitative data necessary to understand how P4P affects these crucial actors. Country offices will conduct case study interviews biannually (twice yearly) with a selected group of participating and non-participating farmers' organizations and farmers and with groups of traders of different sizes and capacities. Biannual (twice yearly) interviews should be sufficient to assess how these actors interact with P4P.

The interviews will focus on developing a detailed qualitative understanding of how P4P works, or, if it does not work, why. General themes for questioning include:

- Has P4P opened up any new opportunities for participating farmers' organizations and farmers?
- How have participating farmers' organizations and farmers responded to the opportunities presented by P4P?
- Did farmers' organizations and farmers face any constraints that prevented or limited a response to P4P?
- Did P4P address any constraints that farmers' organizations or smallholder farmers face producing crops, storing crops, aggregating crops, accessing profitable markets, etc.?
- How did P4P address constraints or present opportunities and how did this affect respondents' businesses, farming practices, or welfare?

- How has P4P affected the relationship between farmers' organizations and farmers and traders? How, if at all, has this affected the trading sector and individual traders?

The detailed understanding of how P4P affects key participants and the contextual factors that influence the effect will also be important input for validating best practices.

Conducting effective case study interviews requires very well prepared interviewers. The interviews are not as structured as the interviews for collecting quantitative data. Instead, they are semi-structured interviews guided by general questioning themes. Interview guides provide the main themes and illustrative questions but the interviewer will have to be comfortable enough with the material and have enough knowledge of both P4P and monitoring and evaluation objectives to follow the natural flow of the conversation and formulate effective probes on the fly.

The section on “Integrating Qualitative Data” on page 34 provides additional detail on how to incorporate the quantitative data from the case studies in the monitoring and evaluation system. The *Purchase for Progress Monitoring Manual* provides detailed instructions on how to design, implement, and report the case studies. It also contains suggested case study interview guides.

ANNEX A

P4P Monitoring and Evaluation System Components

Report/document	Report/document description
Baseline Sampling Strategy	Describes the strategy for selecting farmers' organizations, farmers, and traders for data collection
Data Collection Manual	Provides a detailed description of all aspects of administering the surveys of farmers' organizations, farmers, and traders including complete descriptions of all questions and responses
Enumerator Training Manual	A training manual covering all aspects of training enumerators for data collection.
Farmer Livelihood Baseline Survey Instrument	Survey instrument for collecting baseline data from smallholder farmers.
Farmers' Organization Baseline Survey Instrument	Survey instrument for collecting baseline data from farmers' organizations.
Trader Baseline Survey Instrument	Survey instrument for collecting baseline data from traders.
Monitoring and Evaluation Manual	A manual describing how to conduct monitoring activities
Farmer Livelihood Follow-up Survey Instrument	Survey instrument for collecting follow-up data from smallholder farmers.
Farmers' Organization Follow-up Survey instrument	Survey instrument for collecting follow-up data from farmers' organizations.
Trader Follow-up Survey Instrument	Survey instrument for collecting follow-up data from traders.
Farmer Case Study Interview Guide	Interview guide for case studies with smallholder farmers.
Farmers' Organization Case Study Interview Guide	Interview guide for case studies with farmers' organizations.
Trader Case Study Interview Guide	Interview guide for case studies with traders.
Local and Regional Procurement Report	Report describing how to assess the impacts of standard local and regional procurement.

ANNEX B

Design Team Field Visits

The design team traveled to four P4P pilot countries in Africa (Zambia, Uganda, Malawi, and Burkina Faso) to conduct field visits with P4P programmes. The field visits focused on three objectives: 1) learning about the country’s programme and approach, 2) pretesting the draft data collection instruments and 2) validating the monitoring and evaluation approach as it applied to that country.²⁹ Table 2 summarizes the design team’s field activities.

Table 2. Pretest Schedule and Activities

Country	Dates of site visit	Stakeholder/partner interviews	Instruments pretested
Zambia	May 3 – May 16	<ul style="list-style-type: none"> • Zambia Agricultural Commodity Exchange (ZAMACE) • Production, Finance, and Technology (PROFIT) Project • Food Reserve Agency (FRA) • Famine Early Warning System (FEWS NET) 	<ul style="list-style-type: none"> • Trader
Uganda	May 17 – May 30	<ul style="list-style-type: none"> • LEAD Project • Uganda Commodity Exchange (UCE) • Bank of Uganda • Sasakawa Global 2000 • Ministry of Agriculture • Food and Agriculture Organization (FAO) 	<ul style="list-style-type: none"> • Farmers’ organization • Farmer • Trader
Malawi	May 31 – June 13	<ul style="list-style-type: none"> • Opportunity International Bank Malawi (OIBM) • Malawi Agricultural Commodity Exchange • National Association of Smallholder Farmers (NASFAM) 	<ul style="list-style-type: none"> • Farmers’ organization • Farmer
Burkina Faso	June 14 – June 17	<ul style="list-style-type: none"> • Fédération des Professionnels Agricoles du Burkina Faso. (Federation of Agricultural Professionals of BKF) • Comité Inter-professionnel des Céréales du Burkina. (Cereals cross-professional committee) 	<ul style="list-style-type: none"> • None

²⁹ Representatives from three additional countries (Mozambique, Tanzania, and Malawi) attended the field visits and contributed to, and learned from, the exercise.

Pretesting Data Collection Instruments and Procedures

The quality of data collected through structured interviews depends critically on quality data collection instrument, competent enumerators, and systematic field procedures. The pretests focused primarily on testing the data collection instruments but, because doing so required training enumerators and conducting field interviews, they also tested training and field procedures.

In each country where the design team tested data collection instruments, it began by training a group of enumerators to conduct interviews. Training required two to five days depending on the number of instruments tested in a particular country. Key steps of the training process included:

1. familiarizing enumerators with P4P and monitoring and evaluation objectives,
2. a comprehensive review of the relevant data collection instruments that explained the precise meaning of each question and response,
3. engaging enumerators in translation of the instruments into local language(s),
4. training enumerators in basic interviewing skills and practices, and
5. practicing administering the instruments in workshop and field settings.

Steps two and three in particular identified remaining deficiencies in the draft instruments. These included ambiguous wording, unclear instructions, questions and responses that were not relevant to the local situation, incomplete response categories, incorrect skip patterns, and cumbersome question sequencing. With their detailed local knowledge and experience, the local WFP staff and enumerators helped the design team correct these problems.

Enumerators, accompanied by WFP staff and design team members, then conducted interviews in the field to test the instruments' performance with intended respondents. Field pretests were an iterative process of field work (i.e., interviews) followed by debriefing and revision. The team repeated the process as time permitted or until the instruments performed acceptably. In practice, the design team rarely had time for more than two iterations of pretesting in a given country. However, the team conducted a sufficient number of tests of each instrument across the three countries in which it tested instruments.

The pretest interviews concentrated on learning how the instruments performed in terms of collecting reliable data. Enumerators paid particular attention to whether respondents understood and interpreted questions correctly and were able to provide meaningful answers. They observed, and noted, difficulties respondents experienced understanding or answering questions, the quality (i.e., accuracy and internal consistency) of responses, the time required for the interviews, and respondents' overall reaction to the interview. Debriefing sessions, conducted immediately after field work while the experience was fresh, explored problems enumerators encountered and modified the instruments or procedures to address common difficulties.

ANNEX C

Purchase for Progress Logframe Matrix ³⁰

Component	Indicator	Data source	Risks and assumptions
Impact: To facilitate increased agricultural production and sustained market engagement and thus increase incomes and livelihoods for participating smallholder/low income farmers, the majority of whom are women.	Participating smallholder/low income farmers' annual <u>household incomes</u> (relative to baseline and comparison groups, disaggregated by sex of household head)	Smallholder farmer household surveys	
	Number of participating smallholder/low income farmer's with average annual household income improvements of at least \$50.	Smallholder farmer household surveys	
Objective 1. To identify and share best practices for WFP, NGOs, governments, and agricultural market stakeholders to increase profitable smallholder/low income farmer engagement in markets.			
Outcome 1.1: WFP and other agricultural development stakeholders collaborate to identify procurement and market development best practices from P4P experience	Number of completed compilations (by WFP) of best practice programming and policy recommendations on pro-smallholder local procurement	Document review	Local procurement is an effective method for accomplishing development objectives without undue risk to WFP's and other stakeholders' core objectives.
Output 1.1.1: Mechanisms and procedures to collect and manage P4P performance data developed & functioning	Completed global level M&E system including M&E plan, implementation guidelines, M&E manual, and analysis and reporting routines/templates	Document review	WFP, and particularly the country offices, embrace the learning objective and have the capacity and funding necessary to support country-level M&E activities.
	Number of P4P pilot countries implementing M&E system (e.g., collecting data, producing required reports, etc.)	WFP P4P Unit records	
Output 1.1.2: Monitoring and evaluation results compiled, analyzed, and disseminated.	Percentage of required M&E reports delivered to, or developed by, P4P Unit (disaggregated by country/unit and report type)	WFP P4P Unit records	P4P Unit reviews and assimilates country office M&E reports and data
Output 1.1.3: Engagement of agricultural market stakeholders (e.g., governments, NGOs, partners, private sector, etc.) in dialogue to interpret findings and validate best practices facilitated	Average percentage of invited/expected organizations represented at event/meeting (disaggregated by event/meeting)	Meeting minutes or event attendance/participation records. Applicable events include country level action reviews, regional P4P meetings, global events, technical review panels, and lessons learnt events.	WFP is able to <u>engage</u> a sufficiently wide range of experts who <u>actively participate</u> in the learning process. Stakeholders are willing to participate in collaborative learning.

³⁰ The logframe matrix is, to some extent, a living document and may undergo minor revision during the rollout phase of the P4P pilot.

Component	Indicator	Data source	Risks and assumptions
Outcome 1.2: By the end of the project, agricultural development stakeholders (e.g., governments, NGOs, private sector, donors, etc.) have integrated smallholder/low income farmer-focused market development and procurement best practices into their operations, procedures, or policies	Number of participating stakeholders that have incorporated best procurement and market development recommendations into their operations, procedures, or policy documents	Document review, P4P Unit staff, country-level P4P staff. Document evidence that a stakeholder has incorporated a specific recommendation arising from the P4P pilot into its operations, procedures, or policy documents.	Other agricultural development stakeholders have a large enough presence and can effectively manage the risks associated with local procurement (i.e., not disrupt markets) and retain a focus on smallholder/low income farmers and women.
Output 1.2.1: Implications of lessons learned and best practices for programming or policy (including specific recommendations) documented and conveyed to agricultural market stakeholders and others	Number of publications, or other communications, produced by WFP that contain specific programming or policy recommendations (e.g., guidance to country offices, position papers, policy recommendations, etc.)	Review of documents and other communications (Country office quarterly reports, weekly not-for-the-record (NFR) papers from teleconferences between HQ and CO)	The appropriate stakeholders receive the message and are receptive to the policy recommendations arising from the P4P pilot.
	Number of meetings of agricultural development stakeholders at which either policy or programming proposals are tabled by P4P implementers	Records of WFP, P4P Unit, country offices, and implementing partners documenting formal presentation of P4P programming or policy recommendations at meetings/conferences with other stakeholders.	
<i>Activity:</i> Design, document, and implement a monitoring and evaluation system and plan for P4P including impact assessment models, baselines, data collection, sampling strategies, and training materials.			WFP is able to identify and engage the assistance of partners for data collection and develop the resources and capacities in country offices to manage the M&E process at the country level.
<i>Activity:</i> Develop and implement procedures to manage M&E data at both the country office and headquarters levels			Obtain sufficient funding to manage country-level M&E functions (e.g., data collection and analysis)
<i>Activity:</i> Develop and implement training programs for country office staff in M&E system management, implementation, analysis, and reporting			Training is effective
<i>Activity:</i> Develop standardized routines (SPSS syntax) and reporting templates for country-level analysis and reporting of M&E data			Country offices have the capacity and motivation to collect and analyze data and produce required reports

Component	Indicator	Data source	Risks and assumptions
<i>Activity: Articulate country-specific criteria for selecting participating farmers organizations</i>			<i>Criteria identify farmers' organizations that have the capacity to benefit from supply-side interventions and ultimately sell to WFP but not so advanced that they will not benefit from supply-side interventions.</i>
<i>Activity: Country offices and headquarters produce monthly and quarterly reports on P4P activities, issues, and lessons learned</i>			<i>P4P Unit and country offices have the capacity and motivation to collect and analyze data and produce required reports</i>
<i>Activity: Country offices and headquarters produce biannual M&E reports</i>			<i>Reports are a high enough priority given limited resources</i>
<i>Activity: Country offices and P4P Unit produce annual reports drawing out implications for programming and policy</i>			<i>Reports are a high enough priority given limited resources</i>
<i>Activity: P4P Unit collaborates with WFP Evaluation Unit to facilitate external mid-term and final evaluations of P4P pilot</i>			<i>P4P Unit obtains the resources to support evaluations (if necessary) and the Evaluation Unit engages the appropriate expertise to conduct the evaluations.</i>
<i>Activity: At mid-term and final evaluation points, conduct cost benefit/effectiveness analysis of P4P procurement modalities</i>			<i>Accounting and benefit data to support meaningful cost benefit analysis are available and WFP has access to the expertise necessary to conduct the analyses.</i>
<i>Activity: P4P Unit convenes annual global events in Rome to review P4P performance with country office staff, donors, partners, and other experts</i>			<i>The relevant individuals and organizations attend and participate in the events.</i>
<i>Activity: P4P Unit compiles materials to support review by Technical Review Panel</i>			<i>None.</i>
<i>Activity: P4P Unit convenes annual Technical Review Panel of experts in Rome to review M&E procedures and findings</i>			<i>Technical Review Panel members have the necessary expertise, interest, and sustained engagement in the learning process.</i>
<i>Activity: Country offices convene quarterly or biannual Action Reviews to draw out lessons learned (based on quarterly reports and analysis of M&E data) and validate P4P best practices</i>			<i>The relevant country-level partners attend and actively participate in the events.</i>
<i>Activity: Design and/or contribute to public forums to share knowledge about P4P best practices</i>			<i>Knowledgeable individuals participate in the forums.</i>
<i>Activity: Develop and distribute/disseminate market development and procurement best practices guidelines to WFP, agricultural stakeholders, and partners</i>			

Component	Indicator	Data source	Risks and assumptions
<i>Activity: Develop and distribute training materials for implementing market development and procurement best practices</i>			
<i>Activity: Design and execute cost-effective advocacy campaigns to promote adoption of best practices among agricultural market stakeholders</i>			
<i>Activity: Collaborate with the Institute for Development Studies (IDS) to develop and distribute/implement policy outreach and policy-level advocacy materials</i>			
<i>Activity: Develop and distribute policy briefs and position papers to highlight policy implications (including specific recommendations) of P4P market development and procurement best practices</i>			

Objective 2. To increase smallholder/low income farmers' capacities for agricultural production and market engagement in order to raise their income from agricultural markets.			
Outcome 2.1: By the end of the project, participating smallholder/low income farmers have increased their marketable surpluses of staple commodities.	Average per farm marketable surplus of staple commodities produced by smallholder members of participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity and sex of farmer)	Smallholder farmer household surveys	P4P is successful at building sustainable access to markets for smallholder/low income farmers at prices that reflect the cost of production.
	Average per farm quantity of staple commodities sold by participating smallholder/low income farmers (relative to baseline and comparison group, disaggregated by sex of household head)	Smallholder/low income farmer household surveys	
	Average (per smallholder farm) post-harvest losses of staple commodities as a percentage of annual production (relative to baseline and comparison group, disaggregated by commodity and sex of household head)	Smallholder/low income farmer household surveys	
Output 2.1.1: Smallholder/low income farmers trained in improved agricultural production inputs and practices	Percentage of participating farmers' organizations for which WFP has signed agreements with partners to improve agricultural productivity/production	Country office activity records	Training is effective, inputs are available, and farmers have sustainable markets at prices sufficient to encourage investment in agricultural production.
	Number of smallholder farmer members of participating farmers' organizations trained in improved agricultural productivity/production practices (disaggregated by sex of trainee)	Supply-side partner activity records	
Output 2.1.2: Participating smallholder/low income farmers trained in post harvest handling	Percentage of participating farmers' organizations for which WFP has signed agreements with partners to improve post-harvest handling facilities and practices	Country office activity records	Training is effective, farmers have the resources and incentives to put the training into practice, and implementation is adequate to reduce post-harvest losses.
	Number of smallholder farmer members of participating farmers' organizations trained in improved post-harvest handling and storage practices (disaggregated by sex of trainee)	Supply-side partner or WFP activity records	
Outcome 2.2: By the end of the project, participating smallholder/low income farmers' organizations have increased their capacity to aggregate and market their smallholder members' marketable surpluses of staple commodities	Average proportion of smallholder members' staple commodities sold through participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity and sex of registered farmers' organization member)	Smallholder/low income farmer household surveys Farmers' organization records and surveys	Smallholder farmers have increased their production of staple commodities and are choosing to sell more of their surpluses through the farmers' organization.
	Average (per registered member) quantity of staple commodities sold through participating farmers' organizations (relative to baseline and comparison group, disaggregated by sex of registered farmers' organization member)	Farmers' organization survey and records	

	Average (over participating smallholder farmers' organizations) price received for commodities as a percentage of the highest price in that locality during the marketing season	Farmers' organization records Secondary market data (source varies by country)	
Output 2.2.1: Participating smallholder/low income farmers' organization management staff trained in organizational management (e.g., governance, administration, financial)	Percentage of smallholder/low income farmers' organization management staff who have completed training in governance, administration, or financial management of farmers' organizations (disaggregated by sex)	Supply-side partner activity records	Training is effective, trainees implement lessons in running their organizations, smallholder farmers increase production and choose to sell their staple commodities through the farmers' organization.
Output 2.2.2: Participating smallholder/low income farmers' organizations trained in contracting	Number of farmers' organizations with at least one member of the management staff trained in organization management (i.e., governance, administration, or financial management of farmers' organizations)	Farmers' organization survey and records	Training is effective, trainees implement lessons in running their organizations, contract opportunities exist, smallholder farmers increase production and choose to sell their staple commodities through the farmers' organization.
	Average (over farmers' organizations) percentage of contracts successfully delivered. (relative to baseline and comparison group, disaggregated by country and primary reason for default)	Farmers' organization surveys and records	
	Number of participating smallholder/low income farmers' organizations qualified to participate in WFP competitive tenders (relative to baseline)	WFP procurement data and records	
Output 2.2.3: Stability and representativeness (sex and smallholders) of participating smallholder/low income farmers' organizations improved	Percentage of participating smallholder/low income farmers' organization members who are women (disaggregated by country)	Farmers' organization surveys and records	Improved representativeness leads to organizations that better respond to members' needs and are thus better able to effectively market members' commodities.
	Percentage of participating smallholder/low income farmers' organizations' elected leadership positions held by women (disaggregated by country)	Farmers' organization surveys and records	
	Number of participating smallholder/low income farmers' organization members who are smallholder farmers (disaggregated by country)	Farmers' organization surveys and records	
	Percentage change in total membership of participating smallholder/low income farmers' organizations (relative to baseline and disaggregated by sex of member).	Farmers' organization surveys and records	
Output 2.2.4: Mechanisms established to address participating smallholder/low income farmers' cash flow constraints	Number of participating smallholder/low income farmers' organizations with ability to offer their members some form of financing for crops at harvest (e.g., by pre-purchase, credit, access to warehouse receipt systems, or other full or partial pre-payment for crops) (relative to baseline and comparison group, disaggregated by type of financing)	Farmers' organization survey and records	Addressing cash flow constraints is sufficient to provide smallholder farmers greater flexibility in how they sell commodities and they then choose to sell those commodities through the farmers' organization.

	Number of participating farmers' organizations depositing commodities in a warehouse with a receipt system	Farmers' organization survey and records	
Outcome 2.3: By the end of the project, participating smallholder/low income farmers' organizations have increased access to markets for staple commodities	Average quantity of staple commodities sold by participating farmers' organizations (relative to baseline and comparison group, disaggregated by commodity)	Farmers' organization surveys and records	Smallholder farmer members increase production of staple commodities and choose to sell their surpluses through the farmers' organization.
	Average size of sale of staple commodities by participating smallholder/low income farmers' organization (relative to baseline and comparison group, disaggregated by commodity)	Farmers' organization surveys and records	
	Average number of different geographic markets sold into by participating smallholder/low income farmers' organizations (relative to baseline and comparison group)	Farmers' organization surveys and records	
Output 2.3.1: Partnerships for addressing identified constraints facing smallholder/low income farmers' organizations' access to markets established and monitored	Percentage of participating smallholder/low income farmers' organizations for which WFP has signed agreements with partners to provide market access support	WFP country office records	Partners are effective in working with farmers' organizations to address the identified constraints to market access.
Output 2.3.2: Availability of drying, cleaning, sorting, processing, and storage facilities available to participating smallholder farmers' organizations increased	Number of participating smallholder/low income farmers' organizations offering post-harvest handling services to their members (relative to baseline and comparison group, disaggregated by service)	Farmers' organization survey and records Smallholder farmer surveys	Markets exist for higher quality commodities, farmers' organizations lack the capacity to produce the quality demanded, and addressing constraints to drying, cleaning, sorting, processing, and storage is sufficient to meet quality standards.
	Number of participating farmers' organizations with access to warehouse storage capable of maintaining long-term quality of stored commodities.	Farmers' organization surveys and records Partner activity records Country office activity records	
Outcome 2.4: By the end of the project, the sale of staple commodities is contributing to improved welfare for households of participating smallholder/low income farmers	Average percentage contribution of sale of staple commodities to household incomes of participating smallholder/low income farmers (relative to baseline and comparison groups, disaggregated by sex of household head)	Smallholder/low income farmer household surveys	None
	Average food consumption score of participating smallholder/low income farmer households (relative to baseline and comparison groups, disaggregated by sex of farmers' organization member)	Smallholder/low income farmer household surveys	
	Average household asset score (HAS) of participating smallholder/low income farmer households (relative to baseline and comparison groups, disaggregated by country and sex of farmers' organization member)	Smallholder/low income farmer household surveys	

	Average annual household expenditure (food and non-food) by smallholder farmer households (relative to baseline and comparison groups, disaggregated by sex of household head)	Smallholder/low income farmer household surveys	
	Percentage of participating smallholder/low income farmers who are net sellers of staple commodities (i.e., produce more than they consume) (relative to baseline and comparison group, disaggregated by commodity)	Smallholder/low income farmer household surveys	
<i>Activity: Coordinate with partners to provide appropriate support (access to inputs and technical assistance) to increase productivity of smallholder/low income farmers.</i>			
<i>Activity: Collaborate with partners to provide training in post-harvest handling and storage practices.</i>			
<i>Activity: Identify and sign agreements with appropriate supply-side partners to meet identified gaps in the capacities of smallholder/low income farmers' organizations.</i>			
<i>Activity: Monitor partners' performance relative to agreements, desired P4P outputs, and the Bill and Melinda Gates Foundation activity/milestones</i>			
<i>Activity: Work with supply-side partners to facilitate access to credit for smallholder farmers</i>			
<i>Activity: Facilitate access to cleaning, drying, and storage facilities (e.g., partners rehabilitate or build warehouses, provide cleaning and drying equipment, or link farmers' organizations to certified warehouses, etc.)</i>			

Objective 3. To identify and implement best practices for increasing sales to WFP and others with a particular focus on smallholder/low income farmers.			
Outcome 3.1: The quantity of WFP's purchases from smallholder/low income farmer associations increases by 30% annually throughout the five-year P4P pilot phase	Quantity of food purchased annually by WFP from smallholder/low income farmers' organizations (disaggregated by commodity, procurement modality, and country)	WFP procurement records	Farmers have sufficient surpluses and WFP has sufficient need and capacity to support the targeted increase in procurement.
Output 3.1.1: A clear (country-specific) strategy for increasing procurement of staple commodities from smallholder/low income farmers documented	Number of P4P pilot countries with a documented plan for achieving the required growth increment	CIP, specific strategy for increasing local procurement to achieve the 30% growth target	External factors (i.e., production shocks, prices, etc.) do not curtail quantity available, WFP's need for staple commodities, or ability to procure locally without disrupting markets (i.e., local price is below IPP).
Output 3.1.2: Country offices' local procurement strategies explicitly document impacts on local markets and traders	Number of P4P pilot countries with documented local-specific decision rules to minimize/avoid market distortions	P4P country office records/documents	None.
	Number of P4P pilot countries producing timely market intelligence/impact reports	P4P country office records/documents	
Output 3.1.3: Country office staff trained in P4P procurement	Percentage of country offices with at least one staff member trained in some aspect of local procurement specific to P4P.	WFP country offices, P4P Unit records	Training is effective and addresses a relevant constraint to P4P procurement.
Output 3.1.4: WFP contracts for processed foods establish minimum requirements for smallholder/low income farmer content and means of verification	Average (over participating farmers' organizations) sales of staple commodities to processors. (measured annually and disaggregated by commodity and country)	WFP's P4P and Procurement Units	Processors represent a large enough market for commodities and farmers' organizations can provide adequate quality of commodities.
<i>Activity: Country offices design and regularly review P4P procurement strategy</i>			
<i>Activity: Country offices integrate purchases through P4P into food pipeline</i>			
<i>Activity: Develop standard format for direct and forward delivery contracts</i>			
<i>Activity: Conduct regular analyses of impacts of P4P procurement on local markets and traders</i>			
<i>Activity: Develop locally applicable decision rules to guide the decision on the timing and quantity of purchase from farmers organizations</i>			
<i>Activity: Develop guidance on price setting and contract negotiation for use by country offices.</i>			
<i>Activity: Develop materials and train P4P country office staff in P4P procurement (e.g., price setting, contract negotiation, quality assurance, etc.)</i>			
<i>Activity: WFP increases requirement for their suppliers of processed foods to procure from qualifying smallholder/low income farmers' organizations</i>			

<i>Activity: WFP assesses the potential for smallholder/low income farmers to contribute to WFP's processed foods needs</i>			
<i>Activity: Establish and apply clear criteria for selecting smallholder/low income farmers' organizations to participate in P4P</i>			

Objective 4. To transform WFP food purchase programmes so they better support sustainable small-scale production and address the root causes of hunger.			
Outcome 4.1: By 2013, WFP has transformed its programming, policies, rules, and regulations to incorporate a strategic focus on local procurement with a focus on smallholder/low income farmers	Financial regulations and procedures revised to incorporate pro-smallholder procurement	Review of WFP financial regulations	The P4P pilot concludes that an increased focus on local procurement delivers the desired development impacts and that risks to markets and WFP's core objectives are manageable.
	Job descriptions reflect needs/skills required to effectively manage local procurement	Review HR job descriptions for relevant positions (country directors and procurement, logistics, finance, programming staff). Job descriptions need to include managing/implementing P4P.	
	Program guidance manual revised to reflect a strategic approach to Local Procurement.	Document review	
	Number of P4P pilot countries in which risk management strategies explicitly acknowledge risks associated with pro-smallholder procurement	Document review	
	Percentage of PRROs, EMOPs, and country programmes that incorporate pro-smallholder local procurement as a programme component (disaggregated by country)	Review of documents	Projects have sufficient untied funding to buy under P4P
Output 4.1.1: WFP policies reflecting pro-smallholder procurement best practices endorsed by Executive Board	Percentage of pro local procurement policy proposals presented to WFP's Executive Board that are adopted.	P4P Unit documents and activity records	Local procurement serves WFP's needs and remains a priority for the organization.
Output 4.1.2: Integration and coordination across WFP operational units relevant to P4P implementation established	Percentage of required Steering Committee and Stakeholder group meetings convened.	P4P Unit records	Local procurement accepted by all relevant units.
Output 4.1.3: Country offices reliance on identified best procurement practices for local food procurement increased	Percentage of total annual procurement from local sources (disaggregated by supplier, i.e., trader, farmers' organization, etc.)	WFP procurement monitoring	WFP funding constraints (i.e., tied aid, timing of fund availability) and external factors (demand, availability, prices) do not constrain local procurement activities.
	Quantity of food procured locally (disaggregated by commodity, procurement modality, and country)	WFP procurement monitoring	
<i>Activity: Develop and package results of M&E and mid-term and final evaluations to illustrate impacts of P4P on WFP objectives</i>			
<i>Activity: P4P Unit convenes monthly meetings with Steering Committee and Stakeholder group</i>			
<i>Activity: Adapt existing WFP risk tool to manage risks to local markets and apply to assessing risk associated with P4P procurement.</i>			

<i>Activity: Train country office staff to manage pro-smallholder local procurement activities</i>			
<i>Activity: Country offices form steering committees to provide input on local procurement implementation</i>			

REFERENCES

1. Besculides, Melanie; Zaveri, Heather; Farris, Rosanne Farris, and Will, Julie. Identifying best practices for WISEWOMAN programs using mixed-methods evaluation. *Preventing Chronic Disease*. 2006 Jan.
2. Coulter, Jonathan; Walker, David J., and Hodges, Rick. Local and regional procurement of food aid in Africa: impact and policy issues. *Journal of Humanitarian Assistance*. 2007 Oct 28.
3. Lynch, Will. When to purchase food aid locally? *Frontline Issues in Nutrition Assistance: Hunger Report 2006*. Bread for the World Institute; 2006(Hunger Report).
4. Narma Consultancy. Final report on local food procurement - an analytical review, Nepal country case study. Kathmandu, Nepal.
5. Sserunkuuma & Associates Consult. Local and regional food procurement in Uganda: an analytical review. Kampala, Uganda; 2005 Jun.
6. Swedish International Development Cooperation Agency (Sida). A summary of the theory behind the LFA method: the logical framework approach. Stockholm, Sweden; 2004 Jan.
7. Tschirley, David and del Castillo, Anne Marie. Local and regional food aid procurement: an assessment of experience in Africa and elements of good donor practice. East Lansing, Michigan: Department of Agricultural Economics, Michigan State University(MSU International Development Working Paper (Paper No. 91)); 2007.
8. Wandschneider, Tiago and Hodges, Rick. Local food aid procurement in Uganda. 2005 Apr.
9. World Food Programme. Food procurement annual report: 2008. Rome, Italy: World Food Programme; 2009.
10. World Food Programme: Office of Evaluation and Monitoring. Monitoring and evaluation guidelines: how to design a results-oriented M&E strategy for development programmes. Rome, Italy.